Ecotoxicology (2009) 18:824–828
DOI 10.1007/s10646-009-0325-2

Effects of measurement error on the strength of concentration-response relationships in aquatic toxicology

Derek L. Sonderegger · Haonan Wang · Yao Huang · William H. Clements

Accepted: 9 May 2009 / Published online: 4 June 2009

Abstract The effect that measurement error of predictor variables has on regression inference is well known in the statistical literature. However, the influence of measurement error on the ability to quantify relationships between chemical stressors and biological responses has received little attention in ecotoxicology. We present a common data-collection scenario and demonstrate that the relationship between explanatory and response variables is consistently underestimated when measurement error is ignored. A straightforward extension of the regression calibration method is to use a nonparametric method to smooth the predictor variable with respect to another covariate (e.g., time) and using the smoothed predictor to estimate the response variable. We conducted a simulation study to compare the effectiveness of the proposed method to the naive analysis that ignores measurement error. We conclude that the method satisfactorily addresses the problem when measurement error is moderate to large, and does not result in a noticeable loss of power in the case where measurement error is absent.

Keywords Arkansas River, CO · Ephemeroptera · Measurement error · Nonparametric function smoothing · Regression calibration

D. L. Sonderegger (✉) · H. Wang · Y. Huang
Department of Statistics, Colorado State University, Fort Collins, CO, USA
e-mail: sondereg@stat.colostate.edu

W. H. Clements
Department of Fish, Wildlife and Conservation Biology, Colorado State University, Fort Collins, CO, USA

Introduction

In ecological risk assessment, it is common to take a measurement of contaminant concentration and use that single observation to represent exposure that an organism would experience over some duration of time. This is a legitimate practice if there is very little temporal or spatial variation in contaminant concentrations, however, that is rarely the case. While spatial variation in contaminant concentrations resulting from patchy distributions in the field may be quantified by taking replicate samples, experimental designs rarely account for significant temporal variation. Because of natural variation in stream discharge, temperature, pH and other physicochemical characteristics, water contaminant concentrations often show significant temporal variation. (Clements 2004) reported significant seasonal variation in heavy metal concentrations associated with stream discharge. Researchers measuring diel (24 h) cycles of heavy metals have reported that concentrations of Zn can increase by a factor of five from afternoon minimum values to early morning maximum levels (Nimick et al. 2003). While every attempt can be made to collect water quality samples at the same time every year, natural fluctuations will cause the samples to be taken at different points of the daily and annual cycles.

Measurement error is a commonly studied topic in statistics (Fuller 1987; Carroll et al. 2006; Cheng and van Ness 1999). The most well known result is that for simple linear regression, uncertainty in measuring the predictor variable leads to a substantial underestimation of the relationship between the predictor and response variable. Consequently, if measurement errors are present but unaccounted for in a statistical model, the resulting inference will be less likely to detect an association between contaminants and responses and the magnitude of the association between contaminants and responses will typically be underestimated.

(Nimick et al. 2003) recommends modifying traditional field sampling methods in order to account for measurement error but this presupposes that the error mechanism is known and that the increased sampling is practical. (Yuan 2007) provides a post hoc method of including measurement error in data analysis. For the method we describe, the sampling design is created with measurement error in mind, and the additional fieldwork is not too burdensome.

Methods

Data used for this analysis were collected from the Arkansas River, a metal-polluted stream located in central Colorado. Elevated concentrations of heavy metals (Cd, Cu, and Zn) were consistently reported downstream of Leadville, CO and frequently exceeded acutely toxic levels (Clements 2004). From 1989 to 2004 water chemistry, habitat quality, and abundance of macroinvertebrates were measured at several monitoring stations upstream and downstream from metal sources. In 1991, after 2 years of monitoring, state and federal agencies began a comprehensive restoration effort to improve water quality in the river.

The goal of this long-term study was to use the relationship between metal concentration and macroinvertebrate community structure to assess effectiveness of remediation and resulting improvements in water quality. Because the stream receives a mixture of heavy metals (Cd, Cu, Zn), we used cumulative criterion units (CCU) to quantify metal contamination (Clements et al. 2000). CCU is defined as the ratio of the measured metal concentration to the hardness adjusted chronic criterion concentration, summed for each metal. In this research, the abundance of metal-sensitive mayflies (Ephemeroptera:Heptageniidae) was used as an indicator of stream health. We used a square-root transformation of mayfly abundances to stabilize the variance. In this paper we report data from station AR1 (Clements 2004) collected from 1989 to 2004. During the spring and fall of each year, 5 replicate macroinvertebrate samples were collected along with a single water sample for analysis of heavy metals. Because metal concentrations were represented by a single water sample on each sampling occasion, and because those measurements show considerable diel and seasonal variability (Clements, unpublished data), significant measurement error may exist in these data.

Because of seasonal variation in metal contamination and invertebrate colonization from an upstream source, spring and fall mayfly densities show surprisingly little correlation. To more clearly demonstrate the method used, in this paper we analyze only the fall mayfly samples. Spring CCU values tend to be higher and are more variable than the fall values and that any yearly averaging necessitates a statistical modification similar to the method being proposed.

Due to the remediation efforts above the Arkansas River, we expected a long-term decreasing trend in CCU, but the exact shape of this relationship is unknown. We modeled this relationship in two ways. First, we examined only the fall CCU values and fit a non-parametric function with no constraints on its shape. Second, created a model that incorporates both the spring and fall water quality measurements and our knowledge that a strong remediation effort occurred in the summer of 1991. The remediation is modeled by allowing a discontinuity in the smoothing function, but unfortunately this leaves only 4 data points to estimate the curve before remediation. Due to this data scarcity, we restrict our smoothing function to a flat line over this region.

We refer to the regression of mayfly counts on the raw CCU values as the *naive* estimator of the slope parameter, the regression onto the smoothed fall CCU values as the *de-noised* estimator of the slope parameter, and the regression onto the smoothed CCU values where we used both fall and spring CCU values and allow a discontinuity in 1991 as the *sophisticated de-noised* model.

Mathematical details

Consider the simple regression model

$$Y = \alpha + \beta x + \varepsilon \quad \text{where} \quad \varepsilon \sim N(0, \sigma^2) \tag{1}$$

Suppose that we observe the data $\{X_i, Y_i\}$ where $X_i = x_i + \delta_i$ and $\delta_i \sim N(0, \tau^2)$ for i in $\{1,\ldots,n\}$. In addition, auxiliary information (such as time) is available such that $x = g(t)$. It is reasonable to use the observed data $\{X_i, t_i\}$ to estimate the true values of the covariate $\{x_i\}$. The observed $\{Y_i\}$ can then be regressed onto these de-noised estimates $\{\hat{x}_i\}$.

Regression calibration is typically done by performing a linear regression of $\{X_i\}$ onto $\{t_i\}$ in order to estimate $\{\hat{x}_i\}$, but more sophisticated modeling methods could also be used. Nonparametric smoothers are often more convenient because the researcher doesn't have worry about imposing a functional form on the relationship, only restrictions on the continuity or smoothness. Because the relationship between $\{X_i\}$ and $\{t_i\}$ is typically not of interest, the difficulty of interpretation of the nonparametric smoother is not an issue. The smoother could also incorporate system knowledge to impose physical constrains on the prediction (e.g. the smoothed values must be positive, the relationship form is known over an interval but is unknown over the rest).

The two most common approaches to finding the de-noised or 'smoothed' version of $\{X_i\}$ using nonparametric function estimation are kernel estimation (Fan and Gijbels 1996) and regression splines (Green and Silverman 1993; Ruppert et al. 2003). Although both approaches are appropriate, preliminary analyses showed similar results and due to computational advantages, we restrict our discussion to regression splines.

(Cai et al. 2000) introduced the methodology of regressing the response onto the smoothed predictor using a wavelet smoother and (Cui et al. 2002) extended these ideas to the kernel regression smoother. Both papers demonstrated the asymptotic normality and consistency of the slope parameter of the de-noised variable.

A researcher cannot simply use the de-noised version of an explanatory variable in subsequent analysis and inference without adjustment. While it is possible to derive the asymptotic distribution of $\hat{\beta}$ in certain instances, in general, bootstrap methods are easily used to calculate desired confidence intervals (CI). Because observations were collected at specific time points, the simple method of re-sampling the vectors $\{Y_i, X_i, t_i\}$ does not work. Instead, a bootstrap sample is created by independently re-sampling estimates of the measurement errors $d_i = x_i - \hat{x}_i$ and process errors $e_i = y_i - \hat{y}_i$ and then adding those errors to the estimated$\hat{X}_i$, $\hat{Y}_i$ values. To be explicit, for the ith observation of the bootstrap sample, two indices are randomly selected (say j,k) and the bootstrap observation is $\{\hat{X}_i + d_j,\ \hat{Y}_i + e_k,\ t_i\}$. We provide R code to perform the bootstrap procedure in Appendix.

Simulation study

To demonstrate the effect of ignoring measurement error, we examine three examples of the measurement error model (1) where $\alpha = 0$, $\beta = 1$, and $g(t) = (1 - t)^2$ for $t \in [0, 1]$. In the first case δ is small (attenuation factor $\lambda \approx 0.81$) reflecting an instance where measurement error should not have a substantial effect. The second case shows $\delta = \sigma$ ($\lambda \approx 0.29$) and the third case has $\delta = 2\sigma$ ($\lambda \approx 0.17$). We ran 2000 simulations for each case. For the de-noising procedure, each simulation inference was based on 500 bootstrap samples. The output of this simulation is shown in Table 1.

When measurement error was the same or greater than the response error, the de-noising procedure was clearly superior to the naive estimator. The bias of the naive estimator was quite large and the confidence interval coverage rates (percent of CIs that contain the true parameter value) were far from the desired 95% rate. The de-noising procedure handled the measurement error reasonably well in that the observed bias is quite small. The observed coverage rates were close to the desired 95% level.

In the case where the measurement error standard deviation δ was small, the de-noising procedure did not provide any benefit over standard linear model procedure; however, the procedure did not perform substantially worse. Neither procedure had appreciable bias, the average lengths of the 95% CIs were roughly equivalent, and coverage rates were close to the desired 95%.

Results and discussion

Metal concentrations (as CCU) decreased over time as a result of remediation activities in the Arkansas River (Fig. 1). Standard errors and confidence intervals for the naive estimator were based on assumed asymptotic normality of the error terms. Confidence intervals for the both

Table 1 Simulation results comparing the naive estimator that ignores measurement error with the proposed de-noising procedure for a variety of parameter combinations

Parameters			Ignoring measurement error			De-noising procedure		
Sigma	Delta	N	Bias	Length	Coverage	Bias	Length	Coverage
0.2	0.02	30	−0.006	0.480	0.952	−0.002	0.441	0.934
		60	−0.002	0.340	0.950	0.003	0.326	0.939
		120	−0.005	0.240	0.951	0.000	0.235	0.946
	0.2	30	−0.281	0.538	0.442	0.004	0.671	0.943
		60	−0.293	0.372	0.125	0.003	0.476	0.940
		120	−0.301	0.262	0.004	0.006	0.340	0.946
	0.4	30	−0.611	0.488	0.002	0.030	1.631	0.940
		60	−0.628	0.329	0.000	0.010	0.885	0.944
		120	−0.635	0.228	0.000	0.003	0.568	0.939

The bias column is difference between the mean estimate and the true parameter value

Length is the average length of the resulting 95% CI

Coverage is the proportion of simulations whose 95% CI contained the true parameter value

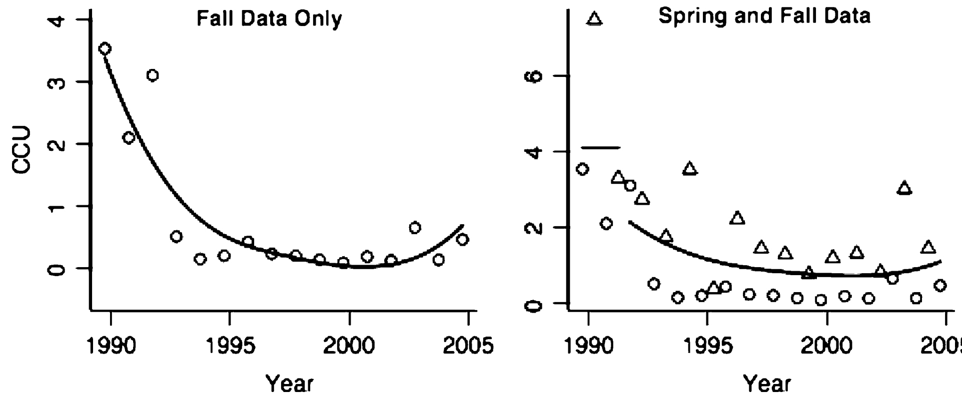

Fig. 1 *Left* CCU vs. time along with the smoothing function. *Right* CCU vs. time for both spring (*triangles*) and fall (*circles*) data. The smoothing function was forced to be flat until the remediation and then a smoother was fit to the remaining data

Table 2 Slope parameter estimates and corresponding 95% confidence intervals for the naive versus the de-noised estimators for the Arkansas River field data

	$\hat{\beta}$	95% CI
Naive estimator	−2.14	(−3.21, −1.07)
De-noised estimator	−2.88	(−4.06, −1.95)
Sophisticated de-noised estimator	−2.45	(−3.74, −1.55)

de-noised estimators were based on $n = 10{,}000$ bootstrap samples. The confidence interval lengths for the naive and de-noised estimators are similar, indicating the small loss of power associated with using the more complicated estimator (Table 2). The most important difference is that the naive estimator has a much smaller magnitude than either of the de-noised estimators. The boundaries of the CI for the de-noised estimator have a substantially larger magnitude than those of the naive estimator. These results indicate that by ignoring measurement error, scientists risk underestimating the relationship between the abundance of metal-sensitive mayflies and heavy metal pollution.

Conclusions

The relationship between chemical concentrations and biological responses is an integral component of ecological risk assessment. Thus, any factor that consistently affects the nature of this relationship has the potential to fundamentally alter our understanding of how chemicals impact ecosystems. (Nimick et al. 2003) suggested that because of temporal variation in contaminant concentrations, it might be necessary to modify traditional field sampling protocols in aquatic ecosystems. We agree with this recommendation, but feel the potential effects of unmeasured temporal variation may be considerably more insidious. Our results suggest that temporal variation in contaminant concentrations introduces significant bias into the concentration-response relationship. This bias typically results in an underestimation of the strength of the relationship between contaminants and biological responses. Field data from the Arkansas River showed that the slope estimates ($\hat{\beta}$) of the relationship between abundance of mayflies and metal concentration increased in magnitude by approximately 14–34% when we accounted for measurement error. While this systematic bias is relatively small if the measurement error is small compared to the overall variability, it increases as measurement errors increase. Our simulation results indicated that the naive estimator consistently provided more biased estimates of slope parameters and misleading CIs as the amount of measurement error increased. While the lengths of the CIs for the de-noised estimator were longer in high measurement error cases, the estimator was effectively unbiased and provided CIs that contained the true parameter value at the desired 95% rate.

Although this paper has only illustrated the 1-dimensional case, this procedure can easily be extended to the multivariate case in several ways. First, Eq. (1) could include other covariates that do not have measurement error. Second, the smoothed variable could be a function of 2 or more auxiliary variables. Third, the auxiliary variable could be used to smooth several covariates.

The de-noising procedure's success is based on the smoother being a consistent estimator of $\{x_i\}$. For both the spline and local polynomial regression smoothers, all that is necessary is for the measurement errors to have mean 0. If that is not the case, an appropriate adjustment to the modeling of $g(t)$ must be made.

The procedure suggested in this paper is applicable in a large number of situations and is relatively easy to implement. There appears to be little cost in inferential power when measurement error is small, and reduces bias in parameter estimates when measurement error is moderate to large. As such, there is little reason not to use such a procedure in situations where appropriate covariates are available.

Acknowledgments We are grateful for the comments and insights made by two anonymous referees. This research was supported by the

US Environmental Protection Agency STAR Program, grant #R832441, the National Science Foundation, IGERT grant #DGE-0221595, and National Science Foundation grant #DMS-0706761. This manuscript has not been formally reviewed by the EPA or NSF. The views expressed here are solely those of the authors. The EPA does not endorse any products or commercial services mentioned in this report.

Appendix: R code for the bootstrap procedures used to calculate confidence intervals for the de-noising procedure.

```
# Assuming X,Y,time vectors are defined
library('SemiPar') # for the smoothing function 'spm'
n < - length(X) # number of observations
# Find the smoothed CCU
model.spm < - spm(X ~ f(time));
smoothed.X < - model.spm$fit$fitted;
# Fit the regression of Y onto smoothed X
model.dn < - lm(Y ~ smoothed.X);
Y.hat < - model.dn$fitted;
# Find the estimated values of epsilons and deltas
e < - Y - model.dn$fitted.values;
d < - X - smoothed.X;
# Bootstrap from e,d
bootstrap.slope < - rep(NA, 1000);
for(i in 1:10000){
  index1 < - sample(1:n, n, replace = T);
  index2 < - sample(1:n, n, replace = T);
  X.boot < - smoothed.X + d[index1];
  Y.boot < - Y.hat + e[index2];
  model.spm < - spm(X.boot ~ f(time));
  smoothed.X < - model.spm$fit$fitted;
  model < - lm(Y ~ smoothed.X);
  bootstrap.slope[i] < - model$coefficients;
}
quantile(bootstrap.slope, p=c(0.025, 0.975));
```

References

Cai Z, Naik PA, Tsai CL (2000) De-noised least squares estimators: an application to estimating advertisement effectiveness. Statist Sinica 10:1231–1243

Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM (2006) Measurement error in nonlinear models: a modern perspective. Chapman & Hall/CRC, Boca Raton, FL

Cheng C-L, van Ness JW (1999) Statistical regression with measurement error. Oxford University Press Inc., New York

Clements WH (2004) Small-scale experiments support causal relationships between metal contamination and macroinvertebrate community responses. Ecol Appl 14:954–967. doi:10.1890/03-5009

Clements WH, Carlisle DM, Lazorchak JM, Johnson PC (2000) Heavy metals structure benthic communities in Colorado mountain streams. Ecol Appl 10:626–638. doi:10.1890/1051-0761(2000)010[0626:HMSBCI]2.0.CO;2

Cui H, He X, Zhu L (2002) On regression estimators with de-noised variables. Statist Sinica 12:1191–1205

Fan J, Gijbels I (1996) Local polynomial modeling and its applications. Chapman & Hall, London

Fuller W (1987) Measurement error models. John Wiley and Sons, Inc, New York, NY

Green PJ, Silverman BW (1993) Nonparametric regression and generalized linear models: a roughness penalty approach. Chapman & Hall, New York, NY

Nimick DA, Gammons CH, Cleasby TE, Madison JP, Skaar D, Brick CM (2003) Diel cycles in dissolved metal concentrations in streams: occurrence and possible causes. Water Resour Res 39:1247–1263. doi:10.1029/2002WR001571

Ruppert D, Wand MP, Carroll RJ (2003) Semiparametric regression. Cambridge University Press, New York

Yuan LL (2007) Effects of measurement error on inferences of environmental conditions. J N Am Benthol Soc 26:152–163. doi:10.1899/0887-3593(2007)26[152:EOMEOI]2.0.CO;2