ARTICLE IN PRESS

Environmental Modelling & Software xxx (2014) 1–11

Contents lists available at ScienceDirect

Environmental Modelling & Software

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

A Bayesian network incorporating observation error to predict phosphorus and chlorophyll *a* in Saginaw Bay

YoonKyung Cha [a], Craig A. Stow [b,*]

[a] *Cooperative Institute for Limnology and Ecosystems Research, University of Michigan, Ann Arbor, MI, USA*
[b] *NOAA Great Lakes Environmental Research Laboratory, Ann Arbor, MI 48108, USA*

ARTICLE INFO

Article history:
Received 20 August 2013
Received in revised form
10 February 2014
Accepted 11 February 2014
Available online xxx

Keywords:
Phosphorus targets
Water quality criteria
Dreissenid invasion
Bayesian hierarchical modeling
Observation error
Bayesian network
Saginaw Bay

ABSTRACT

Empirical relationships between lake chlorophyll *a* and total phosphorus concentrations are widely used to develop predictive models. These models are often estimated using sample averages as implicit surrogates for unknown lake-wide means, a practice than can result in biased parameter estimation and inaccurate predictive uncertainty. We develop a Bayesian network model based on empirical chlorophyll-phosphorus relationships for Saginaw Bay, an embayment on Lake Huron. The model treats the means as unknown parameters, and includes structure to accommodate the observation error associated with estimating those means. Compared with results from an analogous simple model using sample averages, the observation error model has a lower predictive uncertainty and predicts lower chlorophyll and phosphorus concentrations under contemporary lake conditions. These models will be useful to guide pending decision-making pursuant to the 2012 Great Lakes Water Quality Agreement.

Published by Elsevier Ltd.

1. Introduction

Bayesian networks provide an extremely useful framework when a main goal of model development is to support environmental decision-making under uncertainty (Kelly et al., 2013; Landuyt et al., 2013). Our knowledge of the processes that drive environmental systems is always incomplete, although acknowledging uncertainty is often regarded as an admission of ignorance and an impediment to effective decision-making. Quantified uncertainty, however, provides decision-makers with information that, when used appropriately, can improve the decision process (Reckhow, 1994).

Measurement, or more generally observation, error is an uncertainty source that has not been widely considered in Bayesian networks to date. Observation error occurs when recorded values of the independent or dependent variables in a model are only approximate representations of the "true" but unknown quantities they are assumed to represent. While, in principle, models can be constructed to explicitly incorporate this error (Fuller, 1987), in practice this is rarely done.

For example, models are often estimated using sample averages as surrogates for unknown population means. In the aquatic sciences correlations between sample average lake chlorophyll *a* and total phosphorus concentrations have been used for many years to estimate simple linear regression models (Dillon and Rigler, 1974). These models have, in turn, been used for a wide variety of inference, including comparisons among lakes of estimated slope parameters, examination of food web effects on the chlorophyll-total phosphorus relationship, and the prediction of lake responses to changes in phosphorus inputs (Stow and Cha, 2013).

However, using sample averages for the response variable in a regression model, as surrogates for the "true", unknown means, poses a likely violation of the assumption that all observations are of equal variance, particularly if the averages are calculated using differing sample sizes. Similarly, using sample averages as surrogates for means for predictor variable values violates the assumption that the independent variables are observed without error (Fuller, 1987). Consequently, in linear models, slope estimators are biased toward zero, while in nonlinear models the degree and direction of bias depends on the specific model form and the location in both sample and parameter space (Stow and Reckhow, 1996; Carroll, 2006).

These problems are often unrecognized by scientists in the aquatic sciences and other environmental disciplines. This lack of

* Corresponding author.
E-mail addresses: ykcha@umich.edu (Y. Cha), craig.stow@noaa.gov (C.A. Stow).

http://dx.doi.org/10.1016/j.envsoft.2014.02.010
1364-8152/Published by Elsevier Ltd.

Abbreviations

i	index of riverine TP concentration
n	number of riverine TP concentration samples
k	index of lake TP concentration
m	number of lake TP concentration samples
h	index of lake Chl concentration
l	number of lake Chl concentration samples
j	index of year
J	number of years
x	log-transformed riverine TP variable
y	log-transformed lake TP variable
z	log-transformed lake Chl variable
a	intercept parameter of the simple model
b	slope parameter of the simple model
A	annual sample averages

Greek letters

μ_x, μ_y,μ_z	true unknown annual means
σ_x, σ_y, σ_z	standard deviation of individual concentration samples
α	intercept parameter of the observation error model
β	slope parameter of the observation error model
ε, ε^*	model error term
σ_{yo}, σ_{zo}, σ_{ys}, σ_{zs}	standard deviation of model errors
μ	common mean of mean annual riverine TP
σ	standard deviation of mean annual riverine TP

recognition may occur because, in the past, computational limitations made programming models to accommodate observation error difficult, thus, the effects of these assumption violations were not emphasized in applied statistics classes. Additionally, for many applications the practical consequences of violating these assumptions may be minor. However, in applications where precise quantification is important, such as prediction for environmental decision-making, building models that include structure to explicitly incorporate observation error may be useful. Advances in computational power and software now make constructing such models possible.

Herein, we develop a Bayesian network that incorporates observation error in both the predictor and response variables to develop a model that links tributary phosphorus concentration, lake phosphorus concentration, and lake chlorophyll *a* concentration. The sub-model linking lake chlorophyll and phosphorus concentrations is based on the Dillon and Rigler (1974) prototype. We compare results from this observation error model (OEM) to

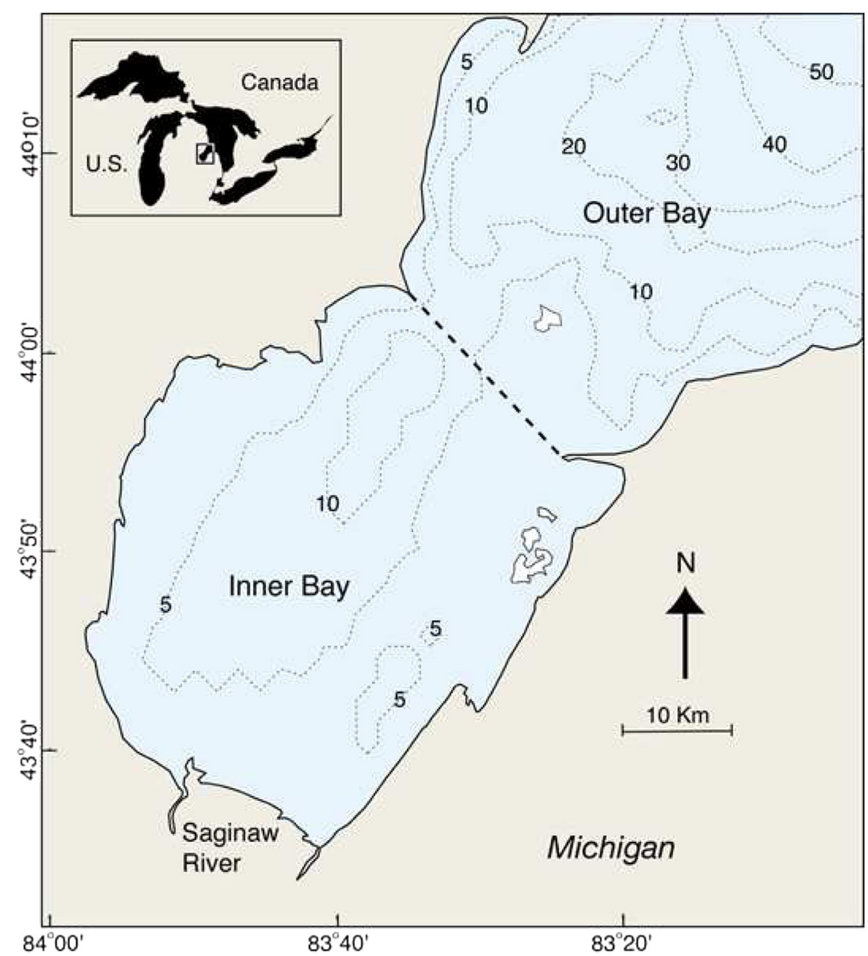

Fig. 1. Map of the Canadian/United States Laurentian Great Lakes. Inset depicts Saginaw Bay, located on the southwestern side of Lake Huron. Contours depict depth (m).

Please cite this article in press as: Cha, Y., Stow, C.A., A Bayesian network incorporating observation error to predict phosphorus and chlorophyll *a* in Saginaw Bay, Environmental Modelling & Software (2014), http://dx.doi.org/10.1016/j.envsoft.2014.02.010