Water Research 115 (2017) 138–148

Contents lists available at ScienceDirect

Water Research

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

Bayesian inference of uncertainty in freshwater quality caused by low-resolution monitoring

Tobias Krueger

IRI THESys, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany

ARTICLE INFO

Article history:
Received 14 September 2016
Received in revised form
11 January 2017
Accepted 26 February 2017
Available online 27 February 2017

Keywords:
Water Framework Directive
Multinomial model
Phosphorus
Nitrogen
Oxygen
Tamar

ABSTRACT

Regulatory, low temporal resolution monitoring of freshwater quality does not fully capture the frequency distributions of the requisite parameters, particularly those that are highly skewed and heavy-tailed. Hence the summary statistics ultimately compared to environmental standards are uncertain. Quantifying this uncertainty is crucial for robust water quality assessment and possible remediation, but requires strong assumptions. This paper compares three ways to model the missing data needed to fully characterise a frequency distribution in a Bayesian framework using multi-year/multi-location orthophosphate (arithmetic mean standard), dissolved oxygen (DO; 10th percentile standard) and ammonia (90th percentile standard) data from the Tamar catchment in Southwest England. First, fitting an assumed parametric model of the frequency distribution (lognormal or Weibull), there is appreciable uncertainty around the "best" model fit. Second, Bayesian Model Averaging is more general in accommodating cases where the data are ambiguous with regard to the best model, but does not take into account possibly missing data. Third, a quasi-nonparametric multinomial model of the monitoring process that places some weight on those missing data yields wider and heavier-tailed frequency distributions. One-at-a-time sensitivity analysis suggests that the multinomial model for mean orthophosphate is sensitive to the choice of support range and the prior weights given to the missing data. Sensitivity is lower for 10th percentile DO and 90th percentile ammonia. The resultant probability densities of ecological status under the EU Water Framework Directive span several status classes, meaning ecological status is more uncertain than previously acknowledged. For orthophosphate, the regulatory, empirical determination of ecological status is not only overly precise but also biased.

1. Introduction

Freshwater quality parameters, such as phosphorus, nitrogen and oxygen concentrations, are routinely monitored by environmental regulators to assess the status of surface waters and inform water resources management. In Europe, the legislative driver is currently the Water Framework Directive (WFD; 2000/60/EC). In the US, it is the Clean Water Act (33 U.S.C. ch. 26) through the Impaired Waters and Total Maximum Daily Load Program. The compliance monitoring is typically done at a low temporal resolution, which in the UK, for example, is fortnightly or monthly, so that no more than 12–26 samples per year are available. This sampling pattern does not fully capture the frequency distributions of the parameters, particularly those that are highly skewed and heavy-tailed, such as phosphorus which is characterised by short-term extremes. Johnes (2007) demonstrated this effect for daily data of discharge, suspended solids and total phosphorus, showing how the upper tails of the empirical frequency distributions contracted progressively as the data were sub-sampled to weekly and then monthly resolution. Ferrant et al. (2013) showed based on sub-sampling a 10-min nitrate-N dataset that a fortnightly monitoring scheme would have missed all extreme concentration values. This error is partly due to the operational realities of sample collection, which usually prevent sampling during heavy rainfall events and other extreme conditions that are highly relevant for pollutant mobilisation and transport. The sampling error, which of course remains unknown outside of sub-sampling studies, translates into uncertainty about the statistical moment or percentile which is ultimately compared to an environmental standard or objective. Skeffington et al. (2015) demonstrated based on sub-sampling hourly dissolved oxygen and total reactive phosphorus data how the uncertainty of the WFD classification and the risk of misclassification increased progressively when moving to weekly and

E-mail address: tobias.krueger@hu-berlin.de.

http://dx.doi.org/10.1016/j.watres.2017.02.061

monthly resolution. Despite advances in high-resolution monitoring for research purposes (Campbell et al., 2015; Jordan et al., 2005; Outram et al., 2014), the limitations of the regulatory monitoring are likely to remain that way (Johnes, 2007). There is thus an imperative to understand and work explicitly with the uncertainties associated with these data.

For robust water quality assessment and possible remediation, a quantification of water quality uncertainty becomes crucial (Skeffington et al., 2015). The problem, however, is that this type of uncertainty quantification requires consideration of the data that could have been sampled but have not, an oxymoron really, which will always rely on assumptions. The objective of this paper is to investigate which assumptions about the possibly missing data might reasonably be made and with what consequences by comparing three Bayesian approaches that quantify the uncertainty caused by data limitations probabilistically. The starting point and first approach studied here is Bayesian inference of an assumed parametric model of the frequency distribution from the available data (Gelman et al., 2013). The pivotal assumption here is the parametric model, and the problem becomes one of verifying this. I have not found a straightforward application of this approach to a water quality problem, though Carstensen (2007) fitted lognormal distributions to marine nutrient concentration data, albeit using classic Maximum Likelihood instead of Bayesian inference. Bayesian water quality studies that did infer frequency distributions from the data explicitly augmented this procedure with some process modelling (Patil and Deng, 2011; Qian and Reckhow, 2007).

It will generally be more robust to average over multiple hypotheses of the frequency distribution, known as Bayesian Model Averaging (BMA; Hoeting et al., 1999), which is the second approach analysed in this paper. BMA has, to my knowledge, not been applied to frequency distributions in a water quality context, but there are applications in other fields. Conigliani (2010) used BMA in a clinical cost-effectiveness context to average lognormal, gamma, Weibull and inverse normal models of highly skewed and heavy-tailed patient cost data. In BMA, individual model results are weighted by the model likelihood. However, the model likelihood is still conditional on the available data and not those that have not been sampled. In the case of insufficient sampling, BMA will thus generally under-estimate the true uncertainty.

BMA will be compared with a third approach, Bayesian inference of a quasi-nonparametric model, here the multinomial model of the sampling process (Aitkin, 2010), which does reflect the analyst's prior ignorance of the shape of the frequency distribution. Under a special case of prior that again neglects possibly missing data, the multinomial model is known as Bayesian bootstrap (Rubin, 1981). The Bayesian bootstrap, like the classic bootstrap (Hirsch et al., 2015), considers the available data representative of the population, which may again be unjustified for small samples from skewed and heavy-tailed frequency distributions (Conigliani, 2010). Under the more general multinomial model, as will be seen, the "prior weight" over the support range of the water quality parameter becomes the pivotal assumption. It will be discussed how this assumption can be made in practice. While the multinomial model can deal with any type of summary statistics, including percentiles and means, for percentile standards the quasi-nonparametric binomial model is a more parsimonious choice (McBride and Ellis, 2001; Smith et al., 2001; Solow and Gaines, 1995). Hence, the multinomial model results will be briefly checked for consistency with the binomial model in this paper.

The structure of the paper is as follows. Section 2 describes the methods of Bayesian inference for an assumed parametric model, BMA and the quasi-nonparametric multinomial and binomial models, and how these will be analysed and compared using data from the Tamar catchment in Southwest England. Section 3 presents the results of the analysis by comparing the summary statistics and associated uncertainty distributions resulting from the three approaches for typical moments and percentiles of three selected water quality parameters. The summary statistics will be evaluated against existing water quality standards to illustrate the impact of uncertainty on the assessment of surface waters. A sensitivity analysis of the multinomial model will be carried out. Section 4 discusses the limitations and benefits of the individual approaches, suggests how their assumptions may be best made in practice and draws out common lessons. Section 5 concludes with some general implications.

2. Methods

When assessing a water quality parameter we want to make inference about a population $\boldsymbol{Y}$, i.e. the instances of the parameter in a time window (e.g. a year), using a sample $\boldsymbol{y} = (y_1, \ldots, y_n)$ of size n drawn from the population. We are interested in summary statistics of the population, such as the arithmetic mean and percentiles. In this paper, I compare three methods of estimating these summary statistics probabilistically. Notes on mathematical notation: vectors are in bold face throughout this paper; generic parameter vectors are denoted by $\boldsymbol{\theta}$; super-script $[t]$ denotes the tth realisation of a quantity from a Monte Carlo sample.

2.1. Bayesian inference of assumed parametric model

The description follows Aitkin (2010). In Bayesian theory, assuming a parametric model of the population $f(\boldsymbol{y}|\boldsymbol{\theta})$, the posterior probability distribution of the model parameters $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ is the prior probability distribution $\pi(\boldsymbol{\theta})$ updated by the likelihood function $L(\boldsymbol{\theta}|\boldsymbol{y})$ through Bayes rule:

$$\pi(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{L(\boldsymbol{\theta}|\boldsymbol{y})\pi(\boldsymbol{\theta})}{\int L(\boldsymbol{\theta}|\boldsymbol{y})\ \pi(\boldsymbol{\theta})\ d\boldsymbol{\theta}}. \tag{1}$$

The likelihood function is the probability of the observed data as a function of the model parameters given measurement precision δ, which is generally considered high relative to the variability in the data:

$$L(\boldsymbol{\theta}|\boldsymbol{y}) = \left[\prod_{i=1}^{n} f(y_i|\boldsymbol{\theta})\right]\delta^n. \tag{2}$$

Bayesian theory requires that we express any prior information as a probability distribution, although this can be non-informative relative to the information in the data. From the posterior distribution of model parameters, the desired summary statistics of $\boldsymbol{Y}$ (e.g. arithmetic mean and percentiles) can be calculated, generally by simulation, in special cases analytically. Here, I compare two parametric models of the frequency distributions of water quality parameters, the lognormal and the Weibull distribution, which were chosen for their flexible and complementary behaviour (Conigliani, 2010). The lognormal model is right-skewed whereas the Weibull model may be left-skewed, right-skewed or symmetrical, and may thus approximate the normal distribution without negative support. The models are sensible choices for water quality data that cannot be negative, are right-skewed (lognormal) with possibly heavy tails (Weibull) like orthophosphate-phosphorus and ammonia-nitrogen, or occasionally left-skewed (Weibull) like dissolved oxygen. For other data, other models may be chosen based on our theoretical understanding of their behaviour. However, our past experience of what might be suitable distributional forms may be influenced by the very same sample deficiencies that we try to