Journal of Hydrology 396 (2011) 148–157

Contents lists available at ScienceDirect

Journal of Hydrology

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

Load estimation with uncertainties from opportunistic sampling data – A semiparametric approach

You-Gan Wang [a,b,*], Petra Kuhnert [c], Brent Henderson [d]

[a] Centre for Applications in Natural Resource Mathematics (CARM), School of Mathematics and Physics, The University of Queensland, St. Lucia 4071, Australia
[b] CSIRO Division of Mathematics, Informatics and Statistics, Australia
[c] CSIRO Division of Mathematics, Informatics and Statistics, PO Box 120 Cleveland, QLD 4163, Australia
[d] CSIRO Division of Mathematics, Informatics and Statistics, GPO Box 664, Canberra ACT 2601, Australia

ARTICLE INFO

Article history:
Received 22 June 2010
Received in revised form 2 November 2010
Accepted 3 November 2010

This manuscript was handled by Geoff Syme, Editor-in-Chief

Keywords:
Biased sampling
Bootstrap
Load estimation
Rating curve
Standard error
Suspended sediment
Uncertainty

SUMMARY

We consider estimating the total load from frequent flow data but less frequent concentration data. There are numerous load estimation methods available, some of which are captured in various online tools. However, most estimators are subject to large biases statistically, and their associated uncertainties are often not reported. This makes interpretation difficult and the estimation of trends or determination of optimal sampling regimes impossible to assess. In this paper, we first propose two indices for measuring the extent of sampling bias, and then provide steps for obtaining reliable load estimates that minimizes the biases and makes use of informative predictive variables. The key step to this approach is in the development of an appropriate predictive model for concentration. This is achieved using a generalized rating-curve approach with additional predictors that capture unique features in the flow data, such as the concept of the first flush, the location of the event on the hydrograph (e.g. rise or fall) and the discounted flow. The latter may be thought of as a measure of constituent exhaustion occurring during flood events. Forming this additional information can significantly improve the predictability of concentration, and ultimately the precision with which the pollutant load is estimated. We also provide a measure of the standard error of the load estimate which incorporates model, spatial and/or temporal errors. This method also has the capacity to incorporate measurement error incurred through the sampling of flow. We illustrate this approach for two rivers delivering to the Great Barrier Reef, Queensland, Australia. One is a data set from the Burdekin River, and consists of the total suspended sediment (TSS) and nitrogen oxide (NO_x) and gauged flow for 1997. The other dataset is from the Tully River, for the period of July 2000 to June 2008. For NO_x Burdekin, the new estimates are very similar to the ratio estimates even when there is no relationship between the concentration and the flow. However, for the Tully dataset, by incorporating the additional predictive variables namely the discounted flow and flow phases (rising or recessing), we substantially improved the model fit, and thus the certainty with which the load is estimated.

1. Introduction

The estimation of riverine loads plays a key role in the management of receiving waters, lakes and lagoons throughout the world and in particular, the Great Barrier Reef (GBR) in Australia, where the load exported from coastal catchments has the potential to impact on reef health (Haynes and Michalek-Wagner, 2000; Furnas, 2003; Baker, 2003). Various policy initiatives and water quality improvement plans have been developed to guide management but these need to be underpinned by the reliable estimation of pollutant loads. Despite numerous load estimation methodologies appearing in the literature, little progress has been made in addressing the three key primary challenges associated with accurately quantifying and reporting loads: (i) adjusting for bias due to the way in which concentration and flow are sampled, (ii) incorporating measurement error and (iii) accounting for knowledge uncertainty through the accurate capture of system processes when developing predictive models for concentration.

There have been a number of comprehensive summaries of loads methodologies appearing in the literature (Gilroy et al., 1990; Cooper and Watts, 2002; Letcher et al., 2002; Cohn, 2005; Littlewood and Marsh, 2005). Methods have ranged from simple average and extrapolation estimators to ratio methods to more sophisticated rating curve approaches such as the seven-parameter model proposed by Cohn et al. (1992) which included seasonal and quadratic effects of log-flow in addition to a term that accounted for a long term trend and found an improved prediction for the

* Corresponding author at: Centre for Applications in Natural Resource Mathematics (CARM), School of Mathematics and Physics, The University of Queensland, St. Lucia 4071, Australia. Tel.: +61 7 3365 2311.
E-mail address: you-gan.wang@uq.edu.au (Y.-G. Wang).

0022-1694/$ - see front matter Crown Copyright © 2010 Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.jhydrol.2010.11.003

examples they trialled. These types of approaches are useful because they attempt to capture the system processes through the covariates in a model. For example, precipitation in GBR catchments occurs predominantly within a well-defined, summer wet season (November to April). The run-off and interflow associated with a wet season's initial, flow-inducing precipitation event tends to pick up unconsolidated, fine sedimentary material and nutrients that have accumulated on or just below the land surface of the catchment. These materials accumulate due to natural weathering, disturbance, anthropogenic activity (e.g. land cultivation) and biomass decay during the relatively long, intervening dry period between wet seasons (Wallace et al., 2008; Wang et al., 2009). These factors may cause the phenomenon of "first flush", i.e., the first significant channelised flow of the wet season which is generally accompanied by relatively high sediment and nutrient concentrations. One of the important challenges in modelling is being able to capture hydrological phenomena such as the concept of a "first flush" and processes such as "depletion". As wet seasons vary from year to year, simple linear or periodic time functions will not adequately capture information such as a "flush" and more novel covariates characterising these phenomena will need to be considered.

Proper uncertainty analysis for loads estimation is also of interest to take account of different sources of errors such as model error and spatial and temporal variations. The recent work of Rustomji and Wilkinson (2008) makes use of the bootstrap technique to resample the residuals and place confidence intervals around estimates of load based on a non-linear regression approach that includes flow terms as covariates. This has been a significant advance from the standard approaches used to investigate uncertainty which make use of simulation techniques that investigate the variability between loads estimation methods and various sampling regimes (Guo et al., 2002; Etchells et al., 2005).

In this paper, we attempt to develop a general regression estimation procedure that provides reliable load estimates with minimal bias and an associated measure of uncertainty. To account for sampling bias, the prediction of flow and concentration is performed at the same regular time intervals and correlation is introduced into the modelling process to account for serial dependence where required.

In Section 2 we will briefly introduce the two datasets from the Tully and Burdekin rivers in North Queensland, Australia. In Section 3 we begin with a discussion of the key bias issues related to the sampling of loads and knowledge of system processes followed by statistical models for predicting the concentrations and an outline of estimation procedures that take into account these sampling biases. In this section, we also outline the generalized rating-curve approach that captures knowledge uncertainty through incorporating of hydrologically meaningful covariates and in particular, we define a new predictive variable, discounted flow (DF), which can mimic the recovery and exhaustion process commonly seen in temporal measures of concentration and flow. The analysis results are given in Section 4 followed by Section 5, a discussion of the method and how it can be generally applied to rivers in other catchments in the GBR.

For the readers convenience, we have listed our notation in Appendix A.

2. Materials and methods

Water quality monitoring in the Great Barrier Reef Catchment Area including the Wet Tropics region (sunch as Tully) and Dry Tropics region (such as Burdekin) is of high priority for the Australian and Queensland Governments to understand possible threats from agricultural runoff along the northern and central Queensland coastline. Recent reef-rescue initiative is to provide a reduction in sediment, nutrient and pesticide loads at the end of each catchment (Bainbridge et al., 2009) which requires reliable estimation of the load and its associated uncertainty. We will use two datasets from the Great Barrier Reef catchments to demonstrate potential bias and other modelling issues, one dataset is from the Tully River (small flow) and the other is from the Burdekin Catchment (large flow). Knowledge of the uncertainties for each program will assist in modelling and future target setting activities.

Samples were collected using a bucket from the centre of the channel flow where possible; otherwise samples were collected from the edge of the waterway (see Mitchell et al., 2006). For suspended sediment loads, two aliquots from each sample were vacuum-filtered onto pre-weighed, polycarbonate membrane filters. The filters were stored at room temperature in pre-washed glass vials, dried overnight at 60 °C and re-weighed, using an analytical-grade balance accurate to 0.001 mg. Standard wet chemical methods were used for the analyses of dissolved inorganic nutrients implemented on a Skalar segmented-flow analyser. Samples stored frozen for analyses were thawed immediately before analysis. Suspended sediment (mg/L) was calculated as the total mass of filter-collected material per unit volume of river water (see Furnas et al., 1995).

The Tully River (catchment area of 1475 km^2) in North Queensland, Australia is a small, fast flowing tropical river that delivers to the Great Barrier Reef lagoon. The river flows initially westward to Koombooloomba Dam and changes direction below Koombooloomba Dam from north to a south easterly direction to Tully from where it flows eastwards to the Coral Sea. The climate is wet tropical. Riparian vegetation has undergone considerable modification. Upper reaches are pristine, but lower sections are substantially modified. The catchment hydrology and barrier effect are undergoing some change from natural.

The flow record spans 8 years and covers the period 1 July 2000 to 16 April 2008, at irregular time spacing (ranging from 0.16 to 34 h, a mean of 1.32 and a median of 0.66 h). The sample sizes for the water quality varies from year to year. For some periods flow is measured at intervals of a few days, but for much of the year the flow measurements are only approximately monthly. The Burdekin catchment is the second largest catchment draining to the GBR lagoon occupying an area of 130,000 km^2. It is the largest catchment in terms of mean gauged annual discharge and the land use within the catchment is dominated by cattle grazing. Development of the catchment by European settlers began in the mid-1800s with the introduction of sheep and cattle grazing and the commencement of alluvial mining. It is generally accepted that post settlement activities such as these would have increased the annual average flux of sediment to the GBR lagoon (see e.g., Belperio, 1979) and in recent years trace-element analysis of coral cores has provided evidence in support of that proposition (McCulloch et al., 2003). Much of the catchment lies upstream of the Burdekin Falls. We will use the data collected in 1997.

Water samples were collected at Inkerman Bridge by the Australian Institute of Marine Science (AIMS) as part of their riverine monitoring program for the purpose of calculating annual loads, while flow data were recorded by a Queensland Department of Natural Resources and Water (NRW) gauge located at Clare, which resides approximately 20 km upstream from the sampling site where water samples were taken for obtaining concentrations of TSS and NO_x in the laboratory. Sampling sites selection was based on flow contribution, site access, the availability of sampling volunteers and the presence of NRW gauging stations. The flow data were collected to best represent the discharge from this large catchment as other contributories between these two locations are small.