STATUS OF THE NIAGARA RIVER POINT SOURCE DISCHARGE INFORMATION: SAMPLING DESIGN AND ESTIMATION OF LOADING

A. H. EL-SHAARAWI
National Water Research Institute, Burlington, Ontario, L7R 4A6, Canada

and

D. J. WILLIAMS
Inland Waters Directorate, Ontario, Canada

(Received May 1989)

Abstract. Two basic requirements, low bias and high precision, are necessary for generating reliable estimates for the load from point and nonpoint sources of pollution. Biases and low precision can be the result of using a bad sampling design and/or an inadequate method of estimation. The effects of biases can be reduced at the design stage prior to the data collection or at the data analysis stage. This paper discusses the statistical issues involved in generating adequate load estimations using recently published point source discharge data from the Niagara River to illustrate these issues.

Introduction

Toxic pollutants are discharged to the Niagara River from a number of municipal and industrial point sources. According to the 1984 Report of the Niagara River Toxics Committee (NRTC, 1984), ninety-five percent of the total point source load of EPA priority pollutants was contributed by 37 of the 188 known discharges. Two reports summarizing the most recent (1985/86) loading data were subsequently released in August and September 1987 by the New York State Department of Environmental Conservation (DEC) and the Ontario Ministry of the Environment (MOE) respectively (McMahon, 1987; MOE, 1987). Both reports identified major reductions between 1981/82 and 1985/86 in both the organic and inorganic priority pollutant loads from these same facilities. The DEC Report, in particular, cited a number of causative factors for these reductions including plant closings, process shutdowns, the completion of wastewater treatment plants, and the successful implementation of remedial programs.

Closer scrutiny, however, suggests that there are also statistical considerations related to (1) the sampling design used to generate the data and (2) the method used for computing the loads in both these reports, that were overlooked and cause some concern about the reliability of the reported load estimates. The statistical factors which limit the usefulness of the data include: (a) the presence of systematic error; (b) the low level of precision; and (c) the unavailability of a measure of uncertainty in the data (either for individual facilities or the total load estimate).
The objectives of this paper are to discuss the roles of the statistical issues involved in designing an efficient monitoring program capable of reliably estimating the loads to the river and detecting real year to year changes in these load estimates. Data from the DEC and MOE reports are used to illustrate the issues involved. In addition, a method for sampling the effluents is proposed which is useful for measuring both the total load to the river and the loads from individual major discharges.

Requirements for a Reliable Load Estimate

A reliable estimate of the load must satisfy a number of requirements which need to be kept in mind prior to planning the data collection and during analysis and interpretation of the data. Briefly, these are: (1) absence of systematic error (bias), (2) a high level of precision and (3) some estimate of the uncertainty surrounding the load estimate generated from the data. The role of each requirement is discussed briefly below with examples.

1. Bias

It is important that load estimates be free of bias. Bias will lead to over- or underestimation of the load. One of the main difficulties is not knowing the direction, magnitude and/or even the causes of the bias. The sources of bias include: (1) contribution from unknown discharges (will cause an underestimate of unknown magnitude), (2) no data for known discharges (an estimate of these loads can perhaps, be attempted using external information), (3) lack of knowledge or complete ignorance about the nature and structure of the variability in both concentrations and flows, (4) changes in sampling and analytical techniques, and (5) the method used to estimate the loads.

Tables I to IV illustrate the effect of ignoring the data structure when estimating the loading.

Table I presents measurements of flow and total phosphorus concentration (TP) taken by MOE over a consecutive three day period on four separate occasions at the Niagara Falls WWTP. The data indicate large differences in TP concentrations between the different samples and hence, a sampling plan which does not take this into account will lead to a biased estimate for the load. Samples taken on each of the three days within each of the four sampling occasions were treated as replicates in a one way analysis of variance (ANOVA) to evaluate the significance of differences among the concentrations on different occasions. Table II shows that when the four separate sampling occasions are compared, both the TP concentrations and TP loads exhibit strong significant differences.

Table III presents the monthly average total priority pollutant loads for the ten most significant U.S. discharges to the Niagara River for 1985/86 (McMahon, 1987). A similar analysis to that noted above was carried out by dividing the 12 months of data into four groups of three months each. The results, provided in Table IV,