Multivariate Regression Analysis of Nutrient/Flow Relationships

Leslie L. Shoemaker
Roland C.Steiner
James A. Smith

DRAFT

ICPRB Working Paper

May 1984

INTERSTATE COMMISSION ON THE POTOMAC RIVER BASIN

1055 First Street
Rockville, Md. 20850
(301) 340-2661

The purpose of this study is to assess the feasibility of using multivariate regression analysis to predict nutrient loads and concentrations at the fall line of the Potomac River at Chain Bridge. Nutrient observations at Chain Bridge are insufficient input data for most water quality models and various at-site regression techniques have been used in the past [NVPDC, 1983., HydroQual, 1982., Hydroscience, 1976] to reconstruct continuous records. Because of the detailed long term flow data available for upstream stations at major tributaries in the Potomac River basin successful application of this additional information could greatly improve current methodology for estimating nutrient data.

Initial examination of this proposed methodology was begun in a cooperative project with the Quality of Water Branch of the USGS. Ten years of daily flow data are used from January 1, 1974 to December 31, 1983. Four upstream USGS flow stations are chosen to represent the streamflow of the Potomac River above the fall line. Flow data from the USGS station near Washington, D.C. is used to represent Chain Bridge flow for the calculation of loads. The stations (see Table 1) account for 88% of the drainage area. Several methods are available to adjust for the missing area including the regression constant, a 'dummy' station, and area weighted adjustment of the Goose Creek flow.

Table 1: Flow Data (cfs) January 1, 1974 - December 31, 1983

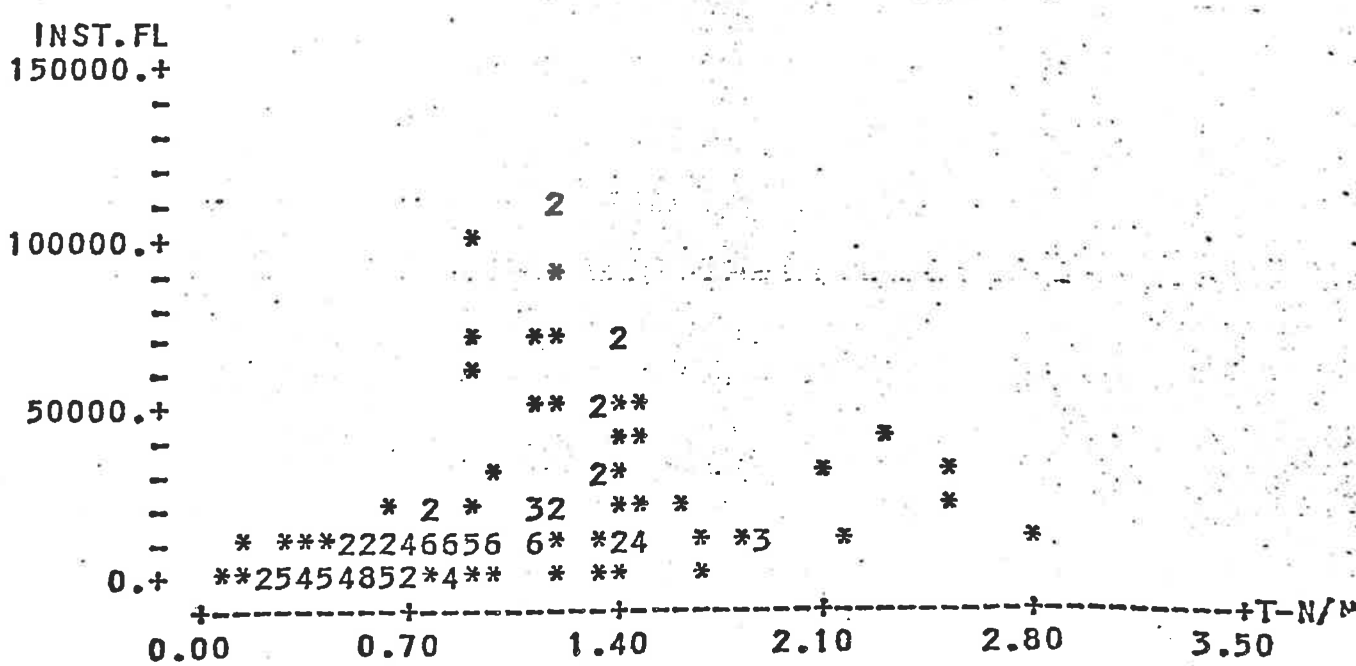
Station	USGS #	Drainage Area sq. mi.	% of Total
Potomac River at Shepherdstown, WV	1618000	5936	51
Shenandoah River at Millville, WV	1636500	3040	26
Monocacy River at Jug Bridge, MD	1643000	817	7
Goose Creek near Leesburg, VA	1644000	332	3
Potomac River near Washington, D.C. Little Falls Pumping St.	1646500	11570	100

The nutrient data used is from the National Stream Quality Accounting Network (NASQAN) station at Chain Bridge for the same 10 year time period as the flow data. Six nutrient species are used in this preliminary analysis including dissolved and total ammonia, dissolved and total nitrite/nitrate, and dissolved and total phosphorus. Regression of nutrient concentration and instantaneous flow at Chain Bridge showed very poor adjusted coefficient of determination (r^2) ranging from 0 to 30%.

Multivariate regression analysis of nutrient concentration with the four upstream flow stations showed substantial improvement in r^2. In order to refine the model, methodology was developed to determine travel time from each of the upstream gages to Chain Bridge for varying flow regimes. A routing model had been developed by Steiner et al. [1984] based on USGS time of travel studies [Taylor et al., 1984] and was modified for this purpose.

The new relationships between lagged flow data and nutrient concentration and loads are examined in greater detail for nitrite/nitrate for a 3 year subset of the initial data. Total nitrite/nitrate is chosen since the correlation with at site flow is low (r^2=10.5%), but not as problematic as volatile constituents such as ammonia (see Figure 1).

Figure 1: Chain Bridge Flow (cfs) vs. Total Nitrite/Nitrate (mg/l)



A first cut approach is adopted in the preliminary analysis of nitrogen data in the Potomac River Basin. The starting data base consists of lagged (particle travel time) flow in the sub-basins:

- Potomac to Shepherdstown
- Shenandoah River
- Monocacy River
- Goose Creek
- Remaining local drainage to Chain Bridge (D.C.)

and sampling data of total Nitrate/Nitrite at Chain Bridge for three years beginning with 1979. In this first cut, linear regression analysis is used only to investigate the underlying relationships between the flow and nitrogen data. In addition, single harmonic Fourier analysis is applied to the nitrogen data in order to explain some of the variation.

Equations relating load and concentration of nitrogen at Chain Bridge to streamflow at various gaging sites are derived using multivariate linear regression. In the example described below two streamflow variables, representing upper basin flow and lower basin flow, are used. The "upper basin" is taken to be the Appalachian portion of the basin above Harpers Ferry. Upper basin flow is computed from the Potomac River gage at Shepherdstown and the Shenandoah River gage at Millville. The "lower basin" is taken to be the Piedmont portion of the basin. Lower basin flow is computed from the Monocacy River gage at Jug

Bridge and the Goose Creek gage at Leesburg. As described below, splitting flow into "upper" and "lower" basin components yields better prediction results for nitrogen at Chain Bridge than by simply using Chain Bridge streamflow. This result can be interpreted as follows: splitting streamflow into Appalachian and Piedmont components allows the model to distinguish the effects of land-use, soil type, and geology on nitrogen inputs to the river. For example, high streamflow from the Monocacy River with its heavy agricultural land use, erodible soils, and close proximity to the Potomac estuary has a much greater impact on nutrient loads to the estuary than high streamflow from the forested South Branch Potomac River in West Virginia. More refined partitions of streamflow contributions could provide even better estimates of Chain Bridge nutrient loads than those obtained by "upper basin"-"lower basin" splits.

The regression equation relating Chain Bridge nitrogen loads to upper basin flow and lower basin flow is the following:

$$NL = 1.0*FU + 1.8*FL ,$$

where,

NL = Chain Bridge nitrogen load in tons/day,

FU = Upper basin flow in cfs,

FL = Lower basin flow in cfs.

The adjusted coefficient of determination (r^2) is 90%. By comparison, the adjusted coefficient of determination for the regression of Chain Bridge nitrogen load versus Chain Bridge streamflow is 60%.

The regression equation relating Chain Bridge nitrogen concentration to upper basin and lower basin flow is the following:

$$NC = 1.6*XU + 3.4*XL$$

where,

NC = Chain Bridge nitrogen concentration in mg/l,

XU = fraction of total flow from upper basin,

XL = fraction of total flow from lower basin.

The adjusted coefficient of determination (r^2) is 23%. By contrast, the adjusted coefficient of determination obtained from regressing Chain Bridge nitrogen concentration against Chain Bridge flow is 10.5%.

Seasonality is an important feature of both nitrogen and streamflow data. To predict nitrogen load and concentration from streamflow, it is necessary to determine whether seasonality is also an important feature in the relationship between nitrogen and streamflow. Preliminary analysis indicates that the relationship between nitrogen concentration and streamflow has a strong seasonal component. This seasonality may reflect the change in flow regime during the course of a year. For example, Spring storm runoff from saturated land provides very different pathways for nitrogen to enter river channels than ground water contributions to river channels during Summer baseflow periods. Partitioning flow between ground water and storm flow sources is one approach that will be taken to improve estimates of nitrogen concentration at Chain Bridge. Refinement of the estimates of travel times of nutrients should also improve estimates of Chain Bridge nutrient loads and concentration.

References

HydroQual, Inc. Calibration and Verification of A Mathematical Model of the Eutrophication of the Potomac Estuary. 1982

Hydroscience, Inc. Water Quality Anaysis of the Potomac River. Report to the Interstate Commission on the Potomac River Basin. 1976

Northern Virginia Planning District Commission. Chesapeake Bay Basin Model, A Final Report. US EPA Chesapeake Bay Program. Annapolis, MD. 1983.

Steiner, R.C., Spielmann, B.A., Sheer, D.P. Development of Response Plans for Toxic Spills. Paper presented at ASCE, Hydraulics Division Specialty Conference, August 1984.

Taylor, K.R., James, R.W.Jr., and Helinsky, B.M., Traveltime and Dispersion in the Potomac River, Cumberland, Maryland, to Washington, D.C., Open-File Report 83-861, U.S. Geological Survey, Towson, Maryland, 1984.