Representative Sampling Location Criterion for Rivers

THOMAS G. SANDERS
Civil Engineering Department, Colorado State University, Fort Collins, Colorado 80523, USA.

Abstract

A quantitative criterion is presented for establishing zones of relatively complete mixing in a river. This zone would require only one sampling point in the lateral transect for obtaining representative water quality data and would be the ideal location for a permanent water quality monitoring station. The criterion based upon a two-way analysis of variance procedure at a 5 per cent level of significance is applied to water quality data collected from the Connecticut River Basin near Deerfield, Massachusetts, USA.

Notation

The following symbols are used in this paper:

F = constant from F tables
F calc = calculated F value
H_o = null hypothesis that the data are from the same parent population
H_l = alternate hypothesis that the data may not be from the same parent population
S_E = unaccounted variation
S_L = variation due to location
S_T = variation due to time
S_{TOT} = total variation
t = number of data collected at each sampling point j
x = sodium absorbance
$\bar{x}$ = mean water quality concentration from all data
x_{ij} = variable defining water quality concentration at location i and time j
$\bar{x}_{i.}$ = mean of data from each point on the lateral transect collected at the same time i
$\bar{x}_{.j}$ = mean of all data taken at the same point j in the lateral transect
y = sodium concentration in mg/ℓ
ℓ = number of sampling points
σ_E^2 = variance of errors
σ_L^2 = variance of data at different locations
σ_T^2 = variance of data at different times

Introduction

In order to effectively monitor water quality in a stream using only a single sampling point in the stream's cross section, the point being sampled should be within a zone of relatively complete mixing. That is to say that the concentration distribution of the water quality variable(s) being investigated is uniform within the cross section. Obviously for sediment this will not be the case, however, for many dissolved water quality variables there can exist zones within a stream which are relatively completely mixed and can be appropriate zones to sample when only one point is sampled. Therefore it is the purpose of this paper to develop a methodology for assessing quantitatively whether or not a zone in a stream is completely mixed and suitable for a representative sampling location.

The variation of water quality concentrations in a river's cross section can be sizeable. The concentration of phenols, for example, were reported to vary in the Mississippi River at St. Louis from 50 ppb at each shore to approximately zero at midstream (Haney and Schmidt, 1958). There was also significant variation of biochemical oxygen demand (BOD) and dissolved solids. Characklis *et al.* (1969) found large variations of chemical oxygen demand (COD) and orthophosphate in a river's cross section but much less variations of nitrate and ammonia in the same cross sections. Colston (1974) found in a study of urban runoff that the majority of all pollutants studied exhibited higher concentrations near the bottom of the stream than near the surface. Only the 5-day BOD and Kjeldahl nitrogen concentrations were the same at both locations. The rest of the pollutants including COD, total organic carbon (TOC), total solids (TS), suspended solids (SS) as well as conservative variables such as magnesium, manganese, lead and zinc had higer concentrations near the stream bottom. Calcium alone had a higher concentration at the surface.

For any water quality variable distributed in a river's cross section even if it is assumed to be completely mixed there will undoubtably be a variation in the measured concentrations due to the inherent randomness of the data. Therefore the question is how much variation in the concentrations is tolerable before multiple samples are required from different points in the lateral transect to estimate the quality of the river at that location.

Using an analysis of variance procedure, the water quality concentration variation within a river's cross section can be analysed to determine if the variation is large, requiring more than one sampling point in order to obtain a representative sample, or small, establishing a completely mixed zone. Analysis of variance procedures test whether a set of sample data can be considered to be drawn from the same parent population having identical means or different populations having different means. Once a level of significance which is a function of allowable variability of water quality is established by the monitoring agency, the question of significant variation can be quantitatively analysed. Simply stated in this application, the level of significance is a measure of the probability that the water quality data as a result of the analysis of variance test is found to be from a zone that is *not* completely mixed when in fact the zone is completely mixed, (in statistical jargon it is the probability of a Type I error). Therefore by increasing or decreasing the level of significance the monitoring agency is specifying an allowable level of variation of the water quality variable in a zone which can be considered completely mixed requiring only one sampling point. As long as the concentration variability in time is negligible relative to variation between locations, one-way analysis in variance is applicable. However, if the variablility of the water quality variable during the time interval when the samples are taken is significant, an additional dimension in the analysis of variance should be considered.

Under most circumstances, complete mixing will exist in the vertical direction, prior to being completely mixed in the lateral transect and the water quality variable concentrations will vary with time. The complete mixing in the vertical column assumes, of course, that the water quality variables are dissolved, the river is relatively shallow and not stratified, fast moving and a reasonable distance downstream from the outfall. Because the water quality variable may vary with lateral location and time a two-way analysis of variance would be required to assess the variability of the water quality concentrations in the lateral cross section. If the variability of the concentrations in time were not taken into account in the analysis, the possibility would exist that the variability of the water quality concentrations at different locations in the lateral transect may be the result of the time variability and not spatial variability.

Two-way Analysis of Variance

In a two-way analysis of variance the assumptions are made that there is a variability of the data in time and location, and that the magnitude of the respective variabilities are independent and additive.
That is,

$$S_{TOT} = S_E + S_T + S_L \tag{1}$$

where

$S_{TOT} = \sum_{i=1}^{t} \sum_{j=1}^{\ell} (x_{ij} - \bar{x})^2$, total variation

$S_T = \sum_{j=1}^{\ell} \sum_{i=1}^{t} (x_{i.} - \bar{x})^2$, variation to time

$S_L = \sum_{i=1}^{t} \sum_{j=1}^{\ell} (\bar{x}_{.j} - \bar{x})^2$, variation due to location

S_E = unaccounted random variation

$\bar{x}$ = mean water quality concentration from all data

x_{ij} = water quality concentration at location i and time j

$\bar{x}_{i.} = \frac{1}{\ell} \sum_{j=1}^{\ell} x_{ij}$, mean of data from each point in the lateral transect collected at the same time i.

$\bar{x}_{.j} = \frac{1}{t} \sum_{i=1}^{t} x_{ij}$, mean of all data taken at the same point j in the lateral transect.

t = number of data collected at each sampling point i.

ℓ = number of sampling points.

TABLE 1
AN ANOVA TABLE USED IN A TWO-WAY ANALYSIS OF VARIANCE

	Degrees of Freedom	Sum of Squares	Mean Sum Sq.	F calc
Rows	$t - 1$	S_T	$S_T/t - 1$	$\frac{S_T(\ell - 1)}{S_E}$
Columns	$\ell - 1$	S_L	$S_L/\ell - 1$	$\frac{S_L(t - 1)}{S_E}$
Error	$(t - 1)(\ell - 1)$	$S_E = S_{TOT} - S_T - S_L$	$S_E/(t - 1)(\ell - 1)$	
Total	$(t)(\ell) - 1$			

The mean squared deviation with respect to time and location divided by the mean squared error are independent random variables having an F distribution with $(t - 1)$, $(t - 1)(\ell - 1)$, and $(\ell - 1)$, $(t - 1)(\ell - 1)$ degrees of freedom respectively (Guttman and Wilks, 1965). Therefore, using an Anova table (Table 1) and the null hypothesis H_o, that the water quality variable concentrations sampled at different times and locations are taken from the same population having identical mean and variance can be tested against the alternate hypothesis H_1, that the variable concentrations sampled at different locations and at different times in a river's cross section are drawn from different populations having different means. The hypotheses which test the equality of the variances are:

$$H_o: \frac{\sigma_L^2}{\sigma_E^2} = 1, \frac{\sigma_T^2}{\sigma_E^2} = 1 \tag{2}$$

$$H_1: \frac{\sigma_L^2}{\sigma_E^2} \neq 1, \frac{\sigma_T^2}{\sigma_E^2} \neq 1 \tag{3}$$

where:

H_o is the null hypothesis that the data are from the same parent population, having identical means

H_1 is the alternate hypothesis that the data may not be from the same parent population, having different means.

$\sigma_L^2 = S_L/\ell - 1$, variance due to location

$\sigma_T^2 = S_T/t - 1$, variance due to time.

$\sigma_L^2 = S_E/(\ell - 1)(t-1)$, random variation

The sequence of the calculations to test the validity of the null hypothesis is illustrated in Table 1. Of primary importance is the calculation of S_E, the unaccounted variability. It is the difference between the total variation and the variations in time and location. The calculated ratio of the mean spatial variability and mean unaccounted variability as well as the ratio of the mean temporal variation and mean unaccounted variability designated in the table as F calc. are estimates of the limiting F constants.

If the F calc. were less than the F constant which is a function of the degrees of freedom (this is a function of the number of data and level of significance and is found in most statistics tables) the ratio of the variances equals one and the null hypothesis is valid (Equation 2).

However, if the F calc. exceeds the F constant, the ratio of the variances is not equal to one and the null hypothesis may not be valid. Therefore, the water quality concentrations probably would not have been derived from the same parent population and would be a function of the location or time or both.

Application of the Criterion

In an application using two-way analysis of variance as a method for assessing spatial variation of a water quality variable in a river's cross section, water quality data were collected and analysed from a straight section of the Connecticut River in the state of Massachusetts, U.S.A. (Sanders, 1974). Grab samples were collected every five minutes for a period of 2,5 h at six different sampling locations in the river's cross section. Separated by approximately 37 m in the lateral direction, the sampling points were approximately 3 700 m downstream from the nearest outfall (Deerfield Massachusetts Sewage Treatment Plant).

The analysis of variance should indicate that concentration variations are significantly different at a specified level of significance in the lateral cross section as complete mixing would probably not occur 3,7 km downstream from the outfall (Sanders *et al.*, 1977). The outfall was located within 6,1 m of the western bank of the river and subsequently the water quality variable concentration downstream should be higher at the west bank than at other points in the lateral transect. The flow of the upstream outfall was insignificant compared to the average flow of the Connecticut River, 0,5 ℓ/s and 6 352 ℓ/s respectively.

Sodium was chosen as the water quality variable to assess the spatial variability in the river's cross section. It is generally not considered a pollutant, but it is conservative and ubiquitous both naturally and in sewage. Most significant is its ability to be detected with high precision at concentrations of less than 1 mg/ℓ with the Perkin Elmer atomic absorption spectrophotometer Model 303. In addition the sodium concentration in the Connecticut River has been reported upstream from the outfall in the range of 10 mg/ℓ, easily detected by the spectrophotometer.

The depth of the Connecticut River at the predetermined sampling location, on the day of the sampling was approximately 1,22 m which eliminated the necessity for using the depth samplers. Each person sampled approximately 0,46 m from the water surface: this being the length of an extended arm from the fingers to the elbow. The sampling points (boats) were on a line perpendicular to the western bank. The distances from the western bank measured using a sextant are listed in Table 2. The absorbances which are linearly related to the concentration of the sodium for each of the 186 grab samples are also listed in Table 2.

Referring to Table 2, it is apparent that the concentration of sodium near the western shore of the river is higher than that in the middle or near the eastern shore. It is also apparent that the sodium absorbance appears to be related to the time of the sampling. The absorbance at every location tended to decrease and then increase during the sampling period. The time dependency, whether it was actual temporal variation or the result of minor electronic variability of the instrument, could not be ignored. Thus it was necessary to use the two-way analysis of variance to assess the spatial variation in the lateral direction.

As indicated by the comparison of the F calc. to the F constant at a 5 per cent level of significance in the completed Anova Table (Table 3), there is substantial variation of the sodium absorbances both with time (rows) and location (columns). Since by definition of two-way analysis of variance, the variation with time is independent and does not affect the variation with location, the results indicate that the cross section is not completely mixed in the lateral direction: more than one sampling point is needed.

TABLE 2
THE ABSORBANCE WHICH IS LINEARLY RELATED TO THE SODIUM CONCENTRATION* OF THE 31 GRAB SAMPLES AT EACH OF SIX SAMPLING POINTS ON THE CONNECTICUT RIVER. LOCATION 1 IS LOCATED NEAR THE WESTERN SHORE OF THE RIVER AND LOCATION 6 NEAR THE EASTERN SHORE. THE SAMPLES WERE TAKEN AT FIVE MINUTE INTERVALS.

	Location					
Sample Number	1	2	3	4	5	6
1	0,324 2	0,292 4	0,297 6	0,311 6	0,309 8	0,308 0
2	0,266 0	0,238 8	0,226 9	0,229 9	0,226 2	2,230 6
3	0,264 4	0,238 1	0,236 6	0,235 1	0,237 3	0,238 1
4	0,299 3	0,277 4	0,279 0	0,274 1	0,272 5	0,272 5
5	0,299 3	0,267 6	0,267 6	0,272 0	0,266 8	0,274 1
6	0,263 6	0,247 2	0,238 1	0,238 1	0,230 6	0,236 6
7	0,272 5	0,253 4	0,247 2	0,242 6	0,238 8	0,240 3
8	0,272 5	0,244 1	0,244 1	0,245 7	0,250 3	0,244 1
9	0,266 0	0,248 7	0,245 7	0,250 3	0,251 8	0,242 6
10	0,262 8	0,237 3	0,247 2	0,238 1	0,242 6	0,242 6
11	0,266 0	0,245 7	0,244 1	0,244 1	0,238 1	0,247 2
12	0,269 2	0,245 7	0,241 1	0,241 1	0,247 1	0,251 8
13	0,262 8	0,244 1	0,240 3	0,245 7	0,242 6	0,244 1
14	0,255 7	0,241 1	0,248 7	0,250 3	0,244 1	0,247 2
15	0,262 8	0,238 1	0,237 3	0,238 1	0,223 3	0,224 8
16	0,259 6	0,241 1	0,238 1	0,245 7	0,245 7	0,239 6
17	0,259 6	0,249 6	0,244 9	0,239 6	0,242 6	0,247 2
18	0,263 6	0,244 1	0,237 3	0,235 1	0,235 1	0,224 0
19	0,254 9	0,245 1	0,238 1	0,233 6	0,244 1	0,244 1
20	0,253 4	0,233 6	0,221 8	0,230 6	0,233 6	0,221 8
21	0,254 9	0,229 1	0,224 8	0,235 1	0,240 3	0,233 6
22	0,270 8	0,246 4	0,229 1	0,229 1	0,235 8	0,233 6
23	0,259 6	0,245 7	0,226 2	0,230 6	0,231 4	0,233 6
24	0,245 7	0,226 2	0,233 6	0,224 8	0,235 1	0,241 1
25	0,242 6	0,233 6	0,222 6	0,217 5	0,318 2	0,230 6
26	0,240 3	0,221 8	0,220 4	0,221 1	0,226 2	0,231 2
27	0,222 6	0,207 6	0,210 4	0,201 3	0,200 7	0,213 2
28	0,221 8	0,195 4	0,191 1	0,187 8	0,187 1	0,195 2
29	0,267 6	0,247 2	0,242 6	0,248 0	0,239 6	0,242 5
30	0,276 5	0,248 7	0,246 4	0,245 7	0,245 7	0,251 8
31	0,259 6	0,244 1	0,238 1	0,233 6	0,238 1	0,244 9
Mean	0,263 2	0,241 9	0,238 9	0,238 9	0,239 4	0,241 3
Distance From West Bank (m)	47,2	75,3	105,8	144,5	181,4	249,9

*The absorbances, x, can be converted to sodium concentration in mg/ℓ, y, with the equation $y = 4{,}4 + 37{,}5\ x$.

TABLE 3
THE ANOVA TABLE OF A TWO-WAY ANALYSIS OF VARIANCE OF SODIUM ABSORBANCE FROM SIX LATERAL SAMPLING LOCATIONS IN THE CONNECTICUT RIVER

	Degrees of freedom	Sum of Sq. deviation	Mean Sum of Sq. deviation	F calc	F constant
Rows	30	0,065	0,002	79,475	1,54
Columns	5	0,014	0,003	103,194	2,27
Error	150	0,004	0,000		
Total	185	0,083			

The mean absorbance was the highest near the western shore (Location 1) and decreased moving east to the middle of the river (Location 4). However, moving from the middle of the river to the eastern most shore the sodium absorbance increased. It was assumed that the sewage treatment plant located on the east bank 4,8 km upstream had such a low flow, less than 0,18 ℓ/s, that it would not affect the downstream concentration. Apparently, the assumption was unfounded.

Although the difference of the sodium concentration in the lateral transverse is negligible, the maximum and minimum concentrations were 16,6 and 11,4 mg/ℓ respectively, it is, nevertheless, statistically significant. Other water quality variables which include total organic carbon, chlorides, BOD, etc., quite probably do vary in the cross section as well and would be of sufficient magnitude that the concentration differences may not be negligible.

Conclusions

In order to obtain a representative mean water quality variable concentration at this particular sampling location on the Connecticut River, more than one sampling point would be required in the river's cross section. Water quality variable concentrations at just the midpoint of the river would not suffice as the samples would not take into account the sewage effluents. Similarly, water quality samples at either shore would not be representative as the effluent concentrations of the sewage treatment plants would have a significant effect on the magnitude of the water quality concentrations. Therefore, it would appear that sampling at the third points of the lateral transverse would be the minimum number of sample locations to adequately define mean concentrations representative of the river's cross section at a location which is not within a zone of complete mixing.

Acknowledgements

This investigation was supported by research Grant WR-A041-MASS as authorized under the Water Resources Act of 1964 from the Office of Water Resources Research Department of the Interior and a National Science Foundation Grant PRA-7913073. The author is indebted to Donald Dean Adrian and Robert C. Ward for their counsel during this project.

References

CHARACKLIS, W.G., GAUDET, F.J., ROE, F.L. and BEDIENT, P.B. (1969) Maximum Utilization of Water Resources in a Planned Community, Stormwater runoff Quality: Data Collection, Reduction and Analysis Office of Research and Development, U.S. Environmental Protection Agency, Cincinnati, Ohio, EPA-600/2-79-050b, July, 99 pp.

COLSTON, N.V. (1974) Characterization of Urban Land Runoff. Paper presented at ASCE National Water Resources Conference, Los Angeles, California, January, 29 pp.

GUTTMAN, I. and WILKS, S.S. (1965) *Introductory Engineering Statistics*, John Wiley and Sons, Inc., New York.

HANEY, P.D. and SCHMIDT, J. (1958) Representative Sampling and Analytical Methods in Stream Studies *Sewage and Industrial Wastes* **30** (6) 812-820.

SANDERS, T.G. (1974) Rational Design Criteria for a River Quality Monitoring Network. Submitted to the Graduate School of the University of Massachusetts in partial fulfilment of the requirements for the Degree of Doctor of Philosophy, August, 229 pp.

SANDERS, T.G., ADRIAN, D.D. and JOYCE, J.M. (1977) Mixing Length for Representative Water Quality Sampling. *JWPCF* **49** 2467-2478.