Environment Protection Engineering

Vol. 5 1979 No. 2

L. F. BLIVEN*, F. A. KOEHLER,* L. F. HORNEY,*
M. R. OVERCASH,* F. J. HUMENIK*

AREAWIDE ASSESSMENT OF RURAL STREAM WATER QUALITY**

PART II

DATA ANALYSIS TECHNIQUE

Cost-effective water quality programs to improve surface water quality must rely mostly on data documenting water quality limitations and sources contributing to the problem. Adequate areawide assessment of rural nonpoint source impact on stream concentrations and loads is one of the difficult problems confronting water quality agencies. Probability (random) sampling was reviewed in Part I of this paper series because it can provide an unbiased data set for assessment of areawide water quality. The feasibility of employing this monitoring strategy was demonstrated by field monitoring following a model, statistically-based grab sampling plan. This paper utilizes these data to demonstrate analysis techniques which are useful in quantifying areawide water quality.

The technigues include comparison of simple statistical parameters, analysis of variance, and regression analysis. These techniques were employed to investigate general water quality, comparison between physiographic areas, and water-quality land-use relationships, respectively.

1. INTRODUCTION

Areawide assessment of surface water quality can be achieved by employing probability (random) sampling. In the first paper [5] of this series, the general methodology for designing a random sampling plan was reviewed and a model sampling plan was developed to assess rural water quality on a watershed basis. The plan included sampling in both space and time because nonpoint sources have spatial variability due to physiographic and land-use variations and temporal variability due to seasonal and weather variations.

The feasibility of implementing the model statistical sampling plan was tested by conducting an areawide assessment or rural water quality in a 12,900 km^2 watershed. In order to reduce travel time and costs, the sample universe was limited to about 25% of the total watershed. Four physiographic, land-use areas were identified and sampling

* Department of Biological and Agricultural Engineering, North Carolina State University, Raleigh, North Carolina, USA.

** Paper Number 5788 of the Journal Series of the North Carolina Agricultural Experiment Station, Raleigh, North Carolina.

sites were selected from each to permit comparisons between the areas. The sites were grab sampled at a constant rate of twice per 28 days, but the actual sampling days for each site were chosen by a restricted random sampling.

The objectives of this paper are to demonstrate some data analysis techniques that can be employed to assess areawide water quality and to present an interpretation of the impact of rural nonpoint sources on stream water quality in the study watershed. The methods discussed are analyses of simple statistical parameters, analysis of variance, and regression analysis. The data base for this demonstration is the data collected during the 2-year Chowan feasibility study with a complete 18 months of data at all sites. The data were analyzed to gain information concerning

1) general water quality,

2) comparison between physiographic areas, and

3) water quality land-use relationships. These analyses were conducted to answer three questions concerning rural water quality in the study watershed. How does rural stream water quality impacted by nonpoint sources compare to water quality criteria? Does the average water quality from different physiographic land use areas of the watershed vary significantly? What is the impact of land use activities and seasonal variation on stream water quality? These questions must be answered in order to

1) document that a rural water quality problem exists,

2) identify the primary areas of concern, and

3) suggest probable cause and effect relationships. These are basic questions which confront all agencies responsible for developing technically justifiable and cost-effective programs to maintain or improve areawide water quality.

The data analysis techniques employed in this paper are common statistical methods which have been presented in-depth in numerous statistical texts such as Statistical Methods [8] and Sampling Techniques [2]. Practical application of these techniques is often efficiently accomplished with the aid of a computer and appropriate software systems. The computations reported here were performed using the „Statistical Analysis System", (SAS) [3] whichis a statistical analysis software system.

2. GENERAL BASIN SUMMARY

Simple statistical parameters, such as the mean value, maximum and minimum values, and coefficient of variation, provide useful information about a data set. The mean value is often compared to values from other sources while the minimum and maximum values provide information about the worst possible conditions that were encountered. The coefficient of variation (CV), the standard deviation divided by the mean, indicates the relative variability of a parameter.

The physical and chemical grab sample measurements for 18 months of data from all sites during the feasibility study are summarized in table 1. All chemical constituents are reported as the elemental form. The parameters summarized in table 1 are water yield (WYD), average downstream current velocity (VELOCITY), stream temperature (TEMP), dissolved oxygen (DO), conductivity (COND), chloride (Cl), carbon oxygen demand (COD), total organic carbon (TOC), total phosphorus (TP), nitrate plus nitrite nitrogen (NO_3-N), total Kjeldahl nitrogen (TKN), ammonia (NH_3-N), and suspended solids (SS) concentrations. Mean values for the four areas were relatively uniform throughout the river basin, but the coefficient of variation was typically greater than 30 percent for all measured parameters except pH which had an average value of $8^0/_0$. In many cases the CV approached 100%. These large CV values indicated enormous naturally occurring variations within each area due to both spatial and temporal factors and demonstrated the large variation in individual sample values.

In order to illustrate general relationships, watershed mean concentration values were calculated and compared to water quality criteria proposed by the Environmental Protection Agency (EPA) [3] and to typical values for secondary treated municipal wastewater in the United States.

The 50 States have adopted individual water quality criteria based upon EPA guidelines and local natural or background conditions. Although North Carolina and Virginia regulations are somewhat different than the EPA guidelines, the EPA values are used for the following discussion of the rural stream water quality. Thus these values are presented only to provide a range of source values and not as a representation of point source impacted water quality.

Oxygen supply was measured as DO concentration and the potential oxygen demand was assessed by TOC and COD concentrations because dissolved oxygen is necessary to maintain fish and other aquatic life. The mean DO levels of the four areas of the Chowan Basin were greater than the regulatory criteria minimum level of 5 mg/dm^3; however, infrequent violations were measured. These violations usually occurred during low flow conditions in the warmer months and thus represented groundwater contributions rather than depletion due to biological or organic stresses. The watershed average TOC and COD concentrations were about 10 and 25 mg/dm^3, respectively, while the maximum recorded concentrations were about 40 and 90 mg/dm^3, respectively. The COD concentration was typically about 2.5 times greater than the TOC concentration. EPA has not established water quality criteria for either TOC or COD, but typical COD concentrations for raw and secondary treated municipal wastewater are 450 and 50 mg/dm^3.

Nitrogen and phosphorus are basic elements for primary productivity, and thus relative concentrations are often employed to assess the potential of algal production. Additionally, EPA has proposed a drinking water quality criterion of 10 mg/dm^3 for nitrate based upon human health considerations. While the maximum measured NO_3-N concentration (9.0 mg/dm^3) of the rural stream water approached this limit, almost all samples had concentrations much less than the proposed standard. For the study areas, the mean TKN and NO_3-N concentration was about 1.5 mg/dm^3. Secondary treated municipal wastewater typically has a total nitrogen concentration of 25 mg/dm^3.

To prevent biological nuisances, EPA suggested water quality criteria for TP is 0.1 mg/ /dm^3 for flowing streams not discharging directly into lakes or impoundments, 0.05 mg/dm^3

Grab sampling measurement summary,
Komplikacja wyników prób wyrywkowych,

Variable	Units	Number of samples	Mean	Minimum value	Maximum value	C.V.* %
		Forested Piedmont				
WYD	mm/day	144	1.28	0.0	31.1	280
VELOCITY	m/s	144	0.20	0.0	0.85	85
TEMP	°C	144	14.5	1.5	24.0	39
DO	mg/dm^3	142	8.56	1.70	13.9	30
pH		70	6.14	5.00	7.20	8
COND	μmhos/cm	144	49.7	10.0	110	38
Cl		144	5.46	1.50	55.3	109
COD		72	23.5	<4.0	87.0	59
TOC		72	8.78	<1.0	30.0	65
TP	mg/dm^3	144	0.12	0.02	1.15	150
NO_3-N		144	0.04	<0.01	0.24	100
TKN		144	1.12	0.11	4.19	74
NH_3-N		50	0.03	<0.01	0.29	167
SS		74	8.87	0.33	94.3	182
		Agricultural Piedmont				
WYD	mm/day	112	1.58	0.0	34.1	278
VELOCITY	m/s	112	0.17	0.0	1.11	106
TEMP	°C	112	14.9	1.0	26.0	42
DO	mg/dm^3	105	9.08	5.50	13.6	24
pH		55	6.26	4.80	7.60	8
COND	μmhos/cm	114	51.2	19.0	220	50
Cl		112	4.20	2.10	29.3	73
COD		56	17.8	<4.0	75.0	76
TOC		56	6.98	<1.0	31.0	84
TP	mg/dm^3	112	0.10	0.02	0.50	80
NO_3-N		112	0.11	<0.01	0.87	100
TKN		112	1.00	<0.04	3.73	77
NH_3-N		42	0.02	<0.01	0.14	200
SS		53	6.82	0.33	30.2	98

* C.V. = Standard Deviation/Mean

in any stream at the point where it enters a lake or reservoir, and 0.025 mg/dm^3 within a lake or reservoir. Because all four areas of the Chowan Basin, including the forested Piedmont, had a mean TP concentration greater than 0.1 mg/dm^3, the natural and the rural nonpoint source impacted stream phosphorus levels in the Chowan Basin appeared to be greater than the proposed water quality criteria. Nevertheless, it should be noted that the rural stream maximum TP concentration (1.5 mg/dm^3) was only about one-tenth the average TP concentration of 10 mg/dm^3 for secondary treated municipal wastewater.

The mean chloride concentration of the rural streams was 6.5 mg/dm^3 compared to a mean Cl concentration of 100 ± 75 mg/dm^3 for secondary treated wastewater.

Table 1

June 1975–November 1976
czerwiec 1975–listopad 1976

Variable	Units	Number of samples	Mean	Minimum value	Maximum value	C.V.* %
		Well-drained coastal plain				
WYD	mm/day	140	2.18	0.0	75.6	340
VELOCITY	m/s	140	0.07	0.0	0.56	114
TEMP	°C	140	14.9	0.05	26.0	44
DO	mg/dm^3	129	7.00	1.40	13.6	35
pH		67	5.93	4.80	7.40	11
COND	μmhos/cm	140	60.1	25.0	190	45
Cl		140	8.25	2.44	51.4	58
COD		70	26.1	<4.00	68.0	52
TOC		70	10.5	<1.00	33.0	59
TP	mg/dm^3	140	0.12	0.02	0.60	83
NO_3-N		140	0.75	<0.01	8.96	187
TKN		140	1.10	0.30	4.77	62
NH_3-N		41	0.07	<0.01	1.03	245
SS		56	8.68	0.55	167	260
		Poorly-drained coastal plain				
WYD	mm/day	136	0.43	0.0	7.62	44
VELOCITY	m/s	136	0.11	0.0	0.50	82
TEMP	°C	136	15.6	2.0	26.0	41
DO	mg/dm^3	136	7.13	0.20	13.1	38
pH		71	6.01	4.70	7.60	10
COND	μmhos/cm	136	80.4	1.30	300	50
Cl		136	9.37	2.63	20.6	31
COD		68	26.6	<4.0	91.0	71
TOC		68	10.7	<1.0	36.0	77
TP	mg/dm^3	136	0.22	0.03	1.65	109
NO_3-N		136	0.53	<0.01	1.69	158
TKN		136	1.18	<0.04	5.25	62
NH_3-N		68	0.17	<0.01	1.30	188
SS		68	14.4	0.8	75.0	97

In summary, the mean nutrient concentrations for the four physiographic areas were relatively uniform throughout the watershed and were substantially less than secondary treated municipal wastewater. Large standard deviations with respect to mean values for many parameters in all areas were indicative of large variation which occur in rural areas. The low DO concentrations and the elevated TP concentrations in the forested Piedmont area demonstrated the need to establish water quality criteria with respect to regional background conditions. The regional background concentrations also provide baseline conditions upon which an assessment of nonpoint source impact can be measured.

Instream data summary,
Charakterystyka cieku,

Variable	Units	Mean of sites	Minimum site average	Maximum site average	C. V. %
		Forested Piedmont — 4 sits			
WYD	mm/day	1.28	0.85	1.15	23
VELOCITY	m/s	0.20	0.17	0.29	30
TEMP	°C	14.5	14.2	14.7	1
DO	mg/dm^3	8.56	8.40	8.81	2
pH		6.14	5.99	6.22	2
COND	μmhos/cm	49.7	45.4	56.1	10
Cl		5.46	4.23	6.93	21
COD		23.5	17.1	29.8	23
TOC		8.78	6.74	10.5	18
TP	mg/dm^3	0.12	0.10	0.14	17
NO_3-N		0.04	0.02	0.06	50
TKN		1.12	1.02	1.19	6
SS		8.87	6.15	12.1	21
		Agricultural Piedmont — 3 sites			
WYD	mm/day	1.58	1.32	1.75	15
VELOCITY	m/s	0.17	0.13	0.20	24
TEMP	°C	14.9	14.5	15.7	4
DO	mg/dm^3	9.08	8.93	9.21	2
pH		6.27	6.17	6.37	2
COND	μmhos/cm	51.2	42.2	68.0	29
Cl		4.20	3.51	5.08	19
COD		17.8	13.9	20.7	20
TOC		6.98	6.05	8.32	17
TP	mg/dm^3	0.10	0.10	0.11	4
NO_3-N		0.11	0.08	0.17	45
TKN		1.00	0.92	1.13	11
SS		6.82	4.26	11.0	54

3. COMPARISON BETWEEN AREAS

Analysis of variance (ANOVA) is a statistical technique employed to test for the equality of several means simultaneously. It splits the total variation of a data set into meaningful components which measure different sources of variation. The classification of observations on the basis of a single criterion is a one-way ANOVA. Similarly data classified by two criteria result in a two-way ANOVA. For more than two criteria, the analysis of variance is referred to as a multi-way ANOVA.

As an example, an ANOVA was conducted using the feasibile study data. The purpose of this ANOVA was to investigate possible differences in long-term average (ari-

Table 2

June 1975—November 1976
czerwiec 1975—listopad 1976

Variable	Units	Mean of sites	Minimum site average	Maximum site average	C.V. %
		Well-drained coastal plain — 4 sites			
WYD	mm/day	2.18	0.48	3.77	84
VELOCITY	m/s	0.07	0.05	0.09	29
TEMP	°C	14.9	14.5	15.3	2
DO	mg/dm^3	7.00	5.87	7.87	13
pH		5.93	5.61	6.08	4
COND	µmhos/cm	60.1	54.5	63.0	6
Cl		8.25	6.36	9.88	18
COD		26.1	16.9	32.6	28
TOC		10.5	7.3	13.3	27
TP	mg/dm^3	0.12	0.07	0.14	25
NO_3-N		0.75	0.06	2.20	131
TKN		1.10	0.94	1.28	15
SS		8.7	3.5	18.4	75
		Poorly-drained coastal — 4 sites			
WYD	mm/day	0.43	0.30	0.52	23
VELOCITY	m/s	0.11	0.06	0.13	27
TEMP	°C	15.6	14.8	16.2	4
DO	mg/dm^3	7.13	5.56	8.38	19
pH		6.01	5.82	6.07	2
COND	µmhos/cm	80.4	67.8	89.5	14
Cl		9.37	7.19	11.36	18
COD		26.6	16.8	40.5	37
TOC		10.7	7.6	18.3	46
TP	mg/dm^3	0.22	0.13	0.41	59
NO_3-N		0.53	0.03	1.17	108
TKN		1.18	0.84	1.33	19
SS		44.4	6.9	28.0	67

thmetic) instream concentrations for stream water in the four areas. Long-term instream conditions for the 15 streams are summarized by areas in table 2.

Mean values for the 15 sites were used to evaluate possible physical and chemical differences between streams in the Piedmont and Coastal Plain, agricultural Piedmont (AP) and forested Piedmont (FP), and poorly-drained Coastal Plain (PDCP) and well-drained Coastal Plain (WDCP). These comparisons were made by a one-way analysis of variance, and the results are summarized in table 3.

The ANOVA showed significant ($P < 0.10$) differences existed among areas for some of the physical and chemical parameters. The mean water velocity was approximately 100% greater in the Piedmont than in the Coastal Plain. Although temperature differen-

Table 3

Analysis of instream data, June 1975–November 1976

Analiza danych dla cieku, VI 1975–XI 1976

Variable	Model r^2	Probability of significant difference		
		Piedmont vs. Coastal plain	AP vs. FP	PDCP vs. WDCP
WYD	0.37	$P > 0.10$	$P > 0.10$	$0.01 < P < 0.05$*
VELOCITY	0.70	$P < 0.01$**	$P > 0.10$	$P > 0.10$
TEMP	0.49	$0.05 < P < 0.10$	$P > 0.10$	$0.05 < P < 0.10$
DO	0.59	$P < 0.01$**	$P > 0.10$	$P > 0.10$
pH	0.54	$P < 0.01$**	$P > 0.10$	$P > 0.10$
COND	0.71	$P < 0.01$**	$P > 0.10$	$P > 0.10$
Cl	0.75	$P < 0.01$**	$P > 0.10$	$P > 0.10$
COD	0.23	$P > 0.10$	$P > 0.10$	$P > 0.10$
TOC	0.24	$P > 0.10$	$P > 0.10$	$P > 0.10$
TP	0.40	$P > 0.10$	$P > 0.10$	$0.05 < P < 0.10$
NO_3–N	0.25	$0.05 < P < 0.10$	$P > 0.10$	$P > 0.10$
TKN	0.17	$P > 0.10$	$P > 0.10$	$P > 0.10$
SS	0.19	$P > 0.10$	$P > 0.10$	$P > 0.10$

* Significant difference

** Highly significant difference

ces were established between both the Piedmont vs. Coastal Plain and the PDCP vs. WDCP areas, they were not considered meaningful on a general water quality basis because the variations were less than 5%. Thus the 25% greater dissolved oxygen level in the Piedmont compared to the Coastal Plain probably reflected increased aeration association with higher velocities in the Piedmont streams. The lower COND level (28%) and Cl concentration (45%) in the Piedmont compared to the Coastal Plain probably were due to decreased atmospheric inputs with increasing distance from the ocean as recorded in precipitation studies by GAMBELL and FISHER [4].

The average instream NO_3 concentration of the Piedmont streams (0.07 mg/dm^3) was substantially less than the Coastal Plain value (0.64 mg/dm^3). However, the NO_3—N model had a small r^2 (0.25) because the Coastal Plain streams had a large inherent variation in mean concentrations (0.03 to 2.20 mg/dm^3). The mean TP concentration in the AP was about three times the value of 0.07 mg/dm^3 recorded for the EP. The COD, TOC, TKN, and SS concentrations did not show any significant ($P > 0.10$) variation among the four study areas.

In summary, an analyses of variance employing the feasibility study data for 15 sites in the four areas indicated that the dominant variation was between the Piedmont and Coastal Plain with relatively minor variations occurring between (1) the agricultura Piedmont and forested Piedmont, and (2) the poorly-drained and well-drained Coastal Plain. The water quality differences between the Piedmont and Coastal Plain were judged to be primarily the result of naturally occurring physiographic variations in (1) basin characteristics, such as soil type and stream hydraulics and (2) ocean proximity.

The Chowan River watershed drains into an estuary with a flushing time scale on the order of one to three months that has in several recent years had excessive algal blooms. To minimize the algal problem, nutrient concentrations need to be minimized. The results of this analysis showed not only that background TP concentrations were elevated but also that the agricultural Piedmont had significantly higher TP concentrations than background. Further, the nitrate values were significantly higher in the Coastal Plain than in the Piedmont. This information is directive in delineating where effort should be expended to reduce stream nutrient concentrations. However, it should be recognized that background conditions may be different in the Coastal Plain and further data is required to assess the impact of rural nonpoint sources on stream water NO_3—N concentrations in this physiographic region.

4. WATER-QUALITY LAND-USE RELATIONSHIPS

Regression is a versatile mathematical technique for data analysis. It can be employed to determine if a dependent variable depends upon an independent variable, and if so, the relationship is often used as a predictive tool. Regression can be used to determine the shape of a curve and to test theories about cause and effect. Finally, it may be used to estimate the error in a parameter in an experiment after adjustments have been made for the effect of a related variable.

In order to predict present and future rural water quality, models are often employed which relate water quality to simple land-use factors. These models are probably best supported by a one-year water-quality land-use study in the United States of 928 sites sampled monthly reported by OMERNIK [6]. Studies found increased nitrogen and phosphorus concentrations and loadings with increased agricultural activity.

Land-use and seasonal effects on water and nutrient yields from the rural Chowan sites are assessed by a regression model of the form:

$$Y = Y_s e \tag{1}$$

where Y is the yield; Y_s is a seasonal coefficient; B is an attenuation coefficient; and F is the percent forested area of the subbasin. A semi-log model was chosen rather than a linear model since it seemed intuitively to be a better representation of the physical system. A semi-log model implies that the yield changes by a constant proportional factor for a constant percentage change in forested area whereas a linear model implies that the yield changes by a constant amount for a constant percentage change in forested area. Data from grab sampling at the four forested Piedmont sites, plus all eight Coastal Plain sites, for two years (11/74–11/76) as well as from the three agricultural sites for 18 months (6/74–11/76) were used for this analysis. A year was composed of four, 3-month seasons representing winter, spring, summer, and fall; winter corresponding

Table 4

Yield model

Model przyrostu zanieczyszczeń wskutek spływów

Variable	Season	V_s	B	r^2	Land use	Season
WYD (mm/day)	Winter	1.20±0.14	0.0235	0.52	$P < 0.01$**	$P < 0.01$**
	Spring	0.87±0.58	0.0235			
	Summer	0.23±0.18	0.0235			
	Fall	0.16±0.05	0.0235			
TKN (kg/km²·day)	Winter	97±9	0.0248	0.53	$P < 0.01$**	$P < 0.01$**
	Spring	57±34	0.0248			
	Summer	29±24	0.0248			
	Fall	8±1	0.0248			
NO_3-N (kg/km²·day)	Winter	236±56	0.0667	0.51	$P < 0.01$**	$P < 0.01$**
	Spring	190±151	0.0667			
	Summer	35±23	0.0667			
	Fall	12±6	0.0667			
TP (kg/km²·day)	Winter	11±3	0.0288	0.48	$P < 0.01$**	$P < 0.01$**
	Spring	16±14	0.0288			
	Summer	3±2	0.0288			
	Fall	2±1	0.0288			

Model: $Y = Y_s e^{-BF}$

** Highly significant

to December, January, and February. Results indicated that models representing Y_s as a function of season and geoclimatic area were not significantly better than models representing Y_s as a function of season alone. Therefore, the simpler models which relate Y_s only to season are summarized in table 4 for water yield (WYD) and TKN, NO_3—N and TP yields. This analysis indicated a highly significant ($P < 0.01$) impact of both season and percent forested area on yields. A watershed typically had about 350% greater water yields during the winter and spring than during the summer and fall. The impact of land use on water yield was a 26% increase for each 10% decrease in forested area. The nutrient yield models demonstrated relationships similar to the water yield model.

The land-use and seasonal effects on volume average concentration were also assessed by a model of the form:

$$C = C_s e^{-BF} \tag{2}$$

where C is the flow weighted average concentration and C_s is a seasonal constant. The results of fitting TKN, NO_3—N, and TP concentrations to this model are summarized in table 5. In general, the models displayed significant relationships but had lower r^2 values than the yield models. TKN and TP concentration variations were attributable to seasonal but not land-use effects while NO_3—N concentrations were a function of both season and land use.

The model sampling plan was designed primarily to compare the water quality from four land-use physiographic areas. Thus the regression analyses relating yield and concentrations to land use and season resulted in relatively low correlation coefficients due to variation among the four areas. The yield models indicated increased yields during the winter and spring due primarily to increased water yield associated with rainfall and lower evapotranspiration during these seasons. The seasonal rainfall pattern was fairly uniform across the watershed so it dominated the yield models and thus provided higher correlation coefficients than were observed for the concentration models which were not dependent upon the seasonal rainfall distribution.

Table 5

Volume average concentration model
Model średnich stężeń objętościowych

Variable	Season	C_s	B	r^2	Land use	Season
TKN (mg/dm^3)	Winter	1.20±0.02	0.0013	0.44	$P > 0.10$	$P < 0.01$**
	Spring	1.05±0.12	0.0013			
	Summer	1.65±0.26	0.0013			
	Fall	0.78±0.15	0.0013			
NO_3–N (mg/dm^3)	Winter	2.88±0.36	0.0432	0.33	$P < 0.01$**	$P < 0.05$*
	Spring	2.72±0.78	0.0432			
	Summer	2.76±0.64	0.0432			
	Fall	1.08±0.28	0.0432			
TP (mg/dm^3)	Winter	0.135±0.017	0.0052	0.20	$P > 0.10$	$P < 0.01$**
	Spring	0.208±0.096	0.0052			
	Summer	0.194±0.010	0.0052			
	Fall	0.196±0.036	0.0052			

Model: $C = C_s e^{-BF}$
* Significant
** Highly significant

5. SUMMARY AND CONCLUSIONS

Probability sampling has broad application potential for areawide assessment of surface water quality. A model sampling plan was developed for assessing areawide rural water quality to illustrate probability sampling concepts. The feasibility of conducting a model random sampling plan was demonstrated by

(a) utilizing the proposed sample site definition in conjunction with available maps to identify sample units within the sampling universe of a study watershed, and

(b) conducting a field sampling program to obtain areawide rural water quality data from the study watershed during a two-year period according to a random grab sampling schedule.

Statistical data analysis methods helpful in areawide water quality assessment and planning were illustrated. Common statistical parameters such as the mean, maximum, minimum and coefficient of variation for stream samples were compared to proposed or existing water quality criteria. The importance of measuring local background or natural conditions in assessing rural nonpoint sources and overall stream quality should not be overlooked. For example, the total phosphorus concentrations in the background area of the feasibility study watershed were greater than proposed EPA stream criteria.

Another problem encountered by water quality planners is to identify areas which have the greatest impact on water quality. Analysis of variance is a mathematical tool which can utilize data from several different identifiable areas to determine whether significant differences in mean water quality exist among the areas considering the variation of the sites within each area. This analysis technique showed that major water quality differences in the study watershed were between the Piedmont and the Coastal Plain regions with minor variations within these regions. In particular, elevated nitrate concentrations were recorded for the Coastal Plain and elevated total phosphorus concentrations were recorded in the agricultural Piedmont. This information is directive for future research into detailed cause and effect relationships, and the best allocation of resources to improve water quality.

Regression analyses can be used to investigate water quality versus land use relationships. Increasing agricultural activities were shown to be associated with increasing nutrient levels in a study of 928 U. S. nonpoint source watersheds by OMERNIK [6]. Analysis of data from the 15 sites in the Chowan watershed supported this general conclusion, but the correlation coefficients for both studies were rather low ($0.2 < r^2 < 0.7$). The large variability of field data must be recognized by water quality planning agencies that use regression analysis results to predict water quality trends based on changing land use patterns or that use mathematical models to predict existing water quality rather than employing field measurements to evaluate existing conditions.

In summary, quantification of nonpoint sources water quality must address both spatial and temperal variability. Further, background or natural conditions also vary so these must be established locally in order to determine the NPS pollution impact. Water quality agencies must recognize and accept this natural variability and design monitoring strategies which can account for it. Probability sampling represents a technically valid, unbiased method to quantify areawide water quality, and traditional statistical data analysis techniques provide a valuable mechanism for data interpretation.*

---

* The project was supported with supplemental funds from the Environmental Protection Agency Grant No. R803328, ,,Pollution From Rural Land Runoff," and the Water Resources Research Institute of the University of North Carolina in cooperation with the North Carolina Agricultural Experiment Station, and the North Carolina Cooperative Extension Service.

ACKNOWLEDGEMENTS

The portion of total project effort reported herein could not have been possible without the support work of the total research team including D. W. Hayne, Institute of Statistics; D. H. Howells, Water Resources Research Institute of North Carolina; J. W. Gilliam, Soil Science Department, North Carolina State University.

REFERENCES

[1] BARR A. J., GOODNIGHT J. H., SALL J. P., HELWIG J. T., *A users guide to SAS*, SAS Institute, Inc., P. O. Box 10522, Raleigh N. C., USA 1976.

[2] COCHRAN W. G., *Sampling techniques*, 2nd edition, John Wiley Sons, New York, USA 1963.

[3] EPA, Pre-publication Copy-quality Criteria for Water, Washington, D. C. 1976.

[4] GAMBELL A. W., FISHER D. W., *Chemical composition of rainfall in eastern North Carolina and south-eastern Virginia*, U. S., Geol. Survey Water-Supply Paper 1535-K, 1966.

[5] HUMENIK F. J., OVERCASH M. R., HORNEY L. F., KOEHLER F. A., BLIVEN L. F., *Areawide assessment of rural stream water quality*, Part I. *Statistical sampling methodology*, Environ. Prot. Engr. Vol. 5 (1979), No. 1.

[6] OMERNIK J. M., *Nonpoint source-stream nutrient level relationships: a national study*, U. S. Environmental Protection Agency, APA-600/3-77-105, 1977.

[7] RUBIN A. R., *Some important considerations for land application of wastewater*, Water Resources Research Inst. of North Carolina, N. C. State University, Raleigh, N. C. USA 1977.

[8] SNEDECOR G. W., COCHRAN W. G., *Statistical methods*, 6th edition, Iowa State University Press, Ames, Iowa, USA 1974.

OKREŚLENIE WPŁYWU ZANIECZYSZCZEŃ OBSZAROWYCH NA JAKOŚĆ WODY W ZLEWNIACH ROLNICZYCH

Część II

METODYKA ANALIZY WYNIKÓW

Ekonomiczne efektywne programy ochrony wód powierzchniowych muszą opierać się na prawidłowej analizie danych dotyczących rozmieszczenia źródeł zanieczyszczeń z uwzględnieniem wszystkich innych czynników wpływających na jakość wód. Jednym z najtrudniejszych problemów dla władz terenowych jest ustalenie wpływu źródeł zanieczyszczeń niepunktowych na stężenia i ładunki zanieczyszczeń w ciekach o rolniczym charakterze zlewni.

Metoda poboru prób wg teorii prawdopodobieństwa, przedstawiona w części I niniejszej pracy, daje obiektywne spojrzenie na obszarowe źródła zanieczyszczeń. Przedstawiono przykład zastosowania statystycznego modelu wyrywkowego poboru prób w warunkach terenowych. W niniejszym artykule przedstawiono technikę analizy wyników pomiarów terenowych pod kątem ilościowego określenia wpływu zanieczyszczeń obszarowych na jakość wód. Metodyka uwzględnia porównanie prostych parametrów statystycznych, analizę wariancji i analizę regresji. Technika ta umożliwia kolejno przedstawienie ogólnego stanu czystości wód powierzchniowych, porównanie różnych rejonów fizjograficznych i zależności stanu czystości wód od wykorzystania terenu.

Badania, przeprowadzone na obszarze 12 900 km^2 zlewni, uwzględniają celowość ograniczania podróży służbowych i zmierzają do ustalenia niezbędnych parametrów na podstawie wycinka zlewni, w tym przypadku 25% powierzchni. Na podstawie uzyskanych wyników wyprowadzono wykładnicze równania przyrostu zanieczyszczeń w cieku wskutek spływów powierzchniowych.

DIE BESTIMMUNG DES EINFLUSSES VON FLÄCHENARTIGEN VERUNREINIGUNGEN AUF DIE WASSERQUALITÄT IN LANDWIRTSCHAFTLICHEN EINZUGSGEBIETEN

Teil II
METHODIK DER ERGEBNISANALYSE

Wirtschaftlich effektive Programme des Gewässerschutzes müssen sich auf einer regelmßüigen Analyse der zerstreuten Schmutzquellen stützen unter Bezugnahme anderer, die Wasserqualität beeinflussenden Faktoren. Für die Gewässerschutz-Behörden ist es sehr schwer, den Einfluss der scharartigen schmutzlquellen auf die Konzentration und Belastung der Gewässer im Einzugsgebiet landwirtschaftlichen Charakters zu erfassen.

Eine Probenahme gemäß den Regeln der Wahrscheinlichkeitstheorie (vergl. den ersten Teil dieses Berichtes), gibt einen objektiven Überblick aller flächenartigen Verschmutzungen. Ein Beispiel der Anwendung eines statistischen Ausschnittmodells der Probenahme im Gelände wird dargestellt.

In diesem Teil werden die im Gelände durchgeführte Messungen analysiert, ebenso der Einfluss von flächenartigen Verschmutzungen auf die Wasserqualität. Diese Methode zieht einfache statistische Größen, die Varianzanlyse und die Regressionsrechnung in Betracht. Dieses Vor gehen ermöglicht eine laufende Darstellung des allgemeinen Standes der Gewässergüte, den Vergleich von verschiedenen physiografischen Regionen und die Abhängigkeit der Wasser-Güteklassen von der Art der Landnutzung.

Diese Methode, die an einem Einzugsgebiet von 12 900 km² Fläche erprobt wurde, berücksichtigt die Zweckmäßigkeit und die Begrenzung von Dienstreisen und ist auf die Festlegung der unentbehrlichen Faktoren nur anhand eines Teiles des Einzugsgebiets abgezielt – in diesem Fall war das 1/4 der gesamten Fläche. Auf Grund die ser Ergebnisse, wurde eine exponentielle Gleichung der Belastungs-Zuwachsrate infolge von Oberflächenabflüssen aufgestellt.

ОПРЕДЕЛЕНИЕ ВЛИЯНИЯ АРЕАЛЬНЫХ ЗАГРЯЗНЕНИЙ НА КАЧЕСТВО ВОДЫ НА СЕЛЬСКОХОЗЯЙСТВЕННОЙ ВОДОСБОРНОЙ ПЛОЩАДИ

Часть II
МЕТОДИКА АНАЛИЗА РЕЗУЛЬТАТОВ

Экономически эффективные программы защиты поверхностных вод должны опираться на правильный анализ данных, касающихся размещения источников загрязнений с учетом всех других факторов, влияющих на качество воды. Одной из наиболее трудных проблем для местных органов управления водным хозяйством является определение влияния источников неточесных загрязнений на концентрацию и заряды загрязнений в потоках на территории водосборной площади, носящей сельскохозяйственный характер.

Метод отбора проб по теории вероятностей, представленный в I-й части настоящей работы, дает объективный взгляд на ареальные источники загрязнений. Приведен пример применения статистической модели выборочного отбора проб в полевых условиях. В статье представлена техника анализа результатов полевых измерений с точки зрения количественного определения влияния ареальных загрязнений на качество воды. В методике учитывается сопоставление простых статистических параметров, анализ дисперсии и анализ регрессии. Эта техника позволяет представить общее состояние чистоты поверхностных вод, а затем сравнение разных физиографических районов и зависимости состояния чистоты воды от использования местности.

Методология, проверенная на территории 12 900 км² водосборного бассейна, учитывает целесообразность ограничения поездок по командировке и стремится определить необходимые параметры на основе участка водосборной площади, составляющего 25% этой площади. По полученным результатам выведены экспоненциальные уравнения приращения загрязнений в потоке в результате поверхностного стока.