United States Environmental Protection Agency | Office of Water Washington, D.C. 20460 | 841-F-93-009 September 1993

EPA

Paired Watershed Study Design

INTRODUCTION

The purpose of this fact sheet is to describe the paired watershed approach for conducting nonpoint source (NPS) water quality studies. The basic approach requires a minimum of two watersheds - control and treatment - and two periods of study - calibration and treatment. The control watershed accounts for year-to-year or seasonal climate variations, and the management practices remain the same during the study. The treatment watershed has a change in management at some point during the study. During the calibration period, the two watersheds are treated identically and paired water quality data are collected (Table 1). Such paired data could be annual means or totals, or for shorter studies (<5 yr), the observations could be seasonal, monthly, weekly, or event-based. During the treatment period, one watershed is treated with a best management practice (BMP) while the control watershed remains in the original management (Table 1). The treated watershed should be selected randomly by such means as a coin toss. The reverse of this schedule is possible for certain BMPs; the treatment period could precede the calibration period. For example, the study could begin with two watersheds in two different treatments, such as "BMP" and "no BMP". Later both watersheds could be managed identically to calibrate them. Since no calibration exists before the treatment occurs, this reversed design is considered risky.

Table 1. Schedule of BMP implementation.

	Watershed	
Period	Control	Treated
Calibration	no BMP	no BMP
Treatment	no BMP	BMP

The basis of the paired watershed approach is that there is a quantifiable relationship between paired water quality data for the two watersheds, and that this relationship is valid until a major change is made in one of the watersheds. At that time, a new relationship will exist. This basis does not require that the quality of runoff be statistically the same for the two watersheds; but rather that the relationship between paired observations of water quality remains the same over time except for the influence of the BMP. Often, in fact, the analysis of paired observations indicates that the water quality is different between the paired watersheds. This difference further substantiates the need to use a paired watershed approach because the technique does not assume that the two watersheds are the same; it does assume that the two watersheds respond in a predictable manner together.

EXAMPLE

To illustrate the paired watershed approach, data taken from a study in Vermont will be used. The purpose of the study was to compare changes in field runoff (cm) due to conversion of conventional tillage to conservation tillage.

Selection of Watersheds

1. Watersheds should be similar in size, slope, location, soils, and land cover.
2. Watersheds should be small enough to obtain uniform treatment over the entire watershed.
3. Watershed outlets should have a stable channel and cross section for discharge monitoring, and should not leak at the outlet.
4. Each watershed should be in the same land cover for a number of years prior to the study so that they are at a steady-state.

Advantages

1. Climate and hydrologic differences over years are statistically controlled.
2. Can attribute water quality changes to a treatment.
3. Control watershed eliminates need to measure all components causing change.
4. Watersheds need not be identical.
5. Study can be completed in shorter time frame than trend studies.
6. Cause-effect relationships can be indicated.

Disadvantages

1. Response to treatment likely to be gradual over time which influences the variance.
2. Study vulnerable to catastrophes such as hurricanes.
3. Shortened calibration may result in serially correlated data.
4. Variances between time periods may not be equal due to drastic treatment.
5. Minimal change in the control watershed is permitted.
6. Requires similar watersheds in close proximity.

The west watershed was the control and was 1.46 hectares (ha) in area. The east watershed was the treatment field and was 1.10 ha. Conventional tillage was moldboard plow whereas conservation tillage was a single disk harrow. The calibration period was one year during which 49 paired observations of storm runoff were made. The treatment period was three years during which 114 paired observations of runoff were made. Data were log-transformed to approach normality based upon the Wilks-Shapiro (W) statistic. The equality of variances between periods was tested using the F-test. Residual plots were examined to check for independence of errors. The statistical package SAS® was used for all analyses.

CALIBRATION

The relationship between watersheds during the calibration period is described by a simple linear regression (Figure 1)

between the paired observations, taking the form:

$$treated_i = b_o + b_1(control_i) + e \quad (1)$$

where *treated* and *control* represent flow, water quality concentration, or mass values for the appropriate watershed, b_o and b_1 are regression coefficients representing the regression intercept and slope, respectively, and e is the residual error.

Three important questions must be answered prior to shifting from the calibration period to the treatment period: a) is there a significant relationship between the paired watersheds for all parameters of interest, b) has the calibration period continued for a sufficient length of time, and c) are the residual errors about the regression smaller than the expected BMP effect?

Regression significance. The significance of the relationship between paired observations is tested using analysis of variance (ANOVA). The test assumes that the regression residuals: are normally distributed, have equal variances between treatments, and are independent.

Hand calculations to test for the significance of the relationship are shown in Snedecor and Cochran (1980, p. 157) (Table 2). The values for Table 2 are calculated from:

$$S_y^2 = \Sigma Y_i^2 - \frac{(\Sigma Y_i)^2}{n} \quad (2)$$

$$S_x^2 = \Sigma X_i^2 - \frac{(\Sigma X_i)^2}{n} \quad (3)$$

$$S_{xy} = \Sigma X_i Y_i - \frac{(\Sigma X_i)(\Sigma Y_i)}{n} \quad (4)$$

$$S_{yx}^2 = \frac{S_y^2 - (S_{xy})^2/S_x^2}{n - 2} \quad (5)$$

Also, the regression coefficients and coefficient of determination are determined from:

$$b_1 = \frac{S_{xy}}{S_x^2} \quad (6)$$

$$b_o = \bar{Y} - b_1\bar{X} \quad (7)$$

$$r^2 = \frac{(S_{xy})^2/S_x^2}{S_y^2} \quad (8)$$

Table 2. Analysis of variance for linear regression.

Source	Degrees of freedom	Sum of squares	Mean squares	F
regression	1	$(S_{xy})^2/S_x^2$	$(S_{xy})^2/S_x^2$	$[(S_{xy})^2/S_x^2]/S_{yx}^2$
residual	n-2	$S_y^2-(S_{xy})^2/S_x^2$	S_{yx}^2	
total	n-1	S_y^2		

In order to perform the calculations by hand, initially calculate: ΣX_i, ΣY_i, $\Sigma X_i Y_i$, ΣX_i^2, ΣY_i^2, $\overline{X}$, $\overline{Y}$. The mean squares (MS) are determined by dividing the sum of squares by the degrees of freedom (df).

For the example above, the following was calculated by hand: ΣX_i = -123.403, ΣY_i = -180.704, $\Sigma X_i Y_i$ = 533.553, ΣX_i^2 = 381.713, ΣY_i^2 = 814.847, $\overline{X}$ = -2.518 (10^X =0.003041 cm), and $\overline{Y}$ = -3.688 (10^Y = 0.000205 cm). Therefore, S_y^2 = 148.441, S_{xy} = 78.463, S_x^2 = 70.933, and S_{yx}^2 = 1.312. Using SAS, the appropriate program is listed below. This program was used to generate Table 3.

SAS PC Program

```
data flow;
  title 'Total Flow (cm)';
  infile 'fname.dat';
  input flow1 flow2;
logflow1=log10(flow1);
logflow2=log10(flow2);
Proc reg;
  Model logflow2=logflow1
        / P CLM;
run;
```

The resulting F statistic for this example would indicate that the regression relationship adequately explains a significant amount ($p<0.001$) of the variation in paired flow data.

Calibration duration. The ratio between the residual variance (mean squares) (S_{yx}^2) for the regression and the smallest worthwhile difference (d) is used to determine if a sufficient sample

Table 3. Analysis of variance for regression of treatment watershed runoff on control watershed runoff.

Source	df	MS	F	p
model	1	86.79	66.17	0.0001
error	47	1.31		
total	48			

has been taken to detect that difference, from:

$$\frac{S_{yx}^2}{d^2} = \frac{n_1 n_2}{n_1+n_2}\left\{\frac{1}{F\left(1+\frac{F}{n_1+n_2-2}\right)}\right\} \qquad (9)$$

where S_{yx}^2 is the estimated residual variance about the regression, d^2 is the square of the smallest worthwhile difference, n_1 and n_2 are the numbers of observations in the calibration and treatment periods ($n_1 = n_2$ for this calculation because n_2 is not known yet), and F is the table value ($p=0.05$) for the variance ratio at 1 and $n_1 + n_2 - 3$ df. The difference (d) is selected based on experience and would vary with project expectations. If the left side of the equation is greater than the right side of the equation, then there are an insufficient number of samples taken to detect the difference. For the example, S_{yx}^2 was 1.312 (from Table 3), $n_1 = n_2$ was 49, and F was 3.94. A ten percent change from the mean was considered a worthwhile difference; therefore, d = 0.10 * $\overline{X}$ = 0.10 * log 0.003041 cm and S_{yx}^2/d^2 = 20.7. The right side of Equation (9) = 6.0; since 20.7 is greater than 6.0, there

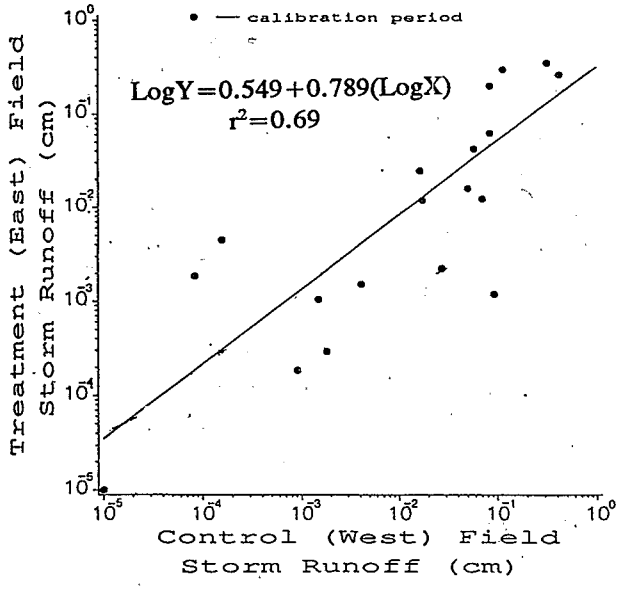

Figure 1. Calibration period regression.

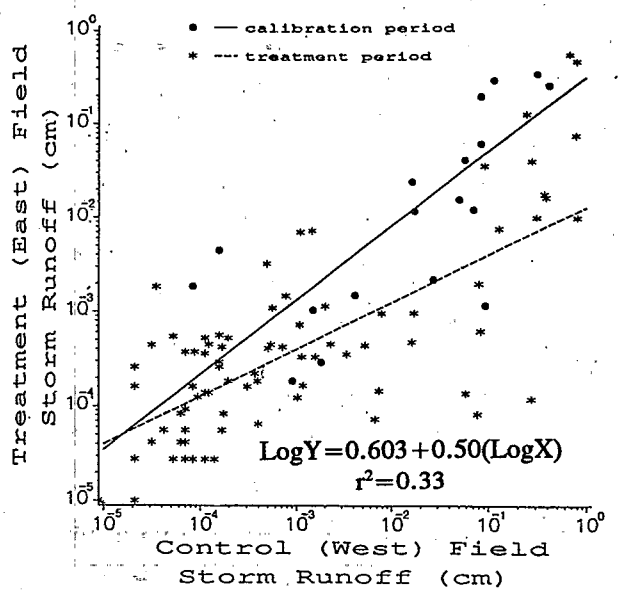

Figure 2. Treatment and calibration period regressions.

was an insufficient number of observations to detect a 10% change in discharge. There were enough samples to detect a 20% change in discharge ($S^2/d^2 = 5.2$).

Residual errors. The confidence bands for the regression equation allow determining the level of change needed to have a significant treatment effect. Thus, how far away from the calibration regression must the treatment data be to be significantly different? Confidence bands for the regression are determined from:

$$CI = \pm(t)(S_{yx})\sqrt{\frac{1}{n}+\frac{(X_i-\overline{X})^2}{S_x^2}} \qquad (10)$$

where CI is the confidence interval, S_{yx} is the square root of S_{yx}^2, n and S_x^2 have been previously defined, t is Student's 't', and X_i is the value at the point of comparison to compare to the mean on the regression line. Confidence limits can be generated in SAS by adding / P CLM to the MODEL statement (see page 4).

TREATMENT

At the end of the treatment period the significance of the effect of the BMP is determined using analysis of covariance (ANCOVA). The analysis is actually a series of steps determining a) the significance of the treatment regression equation, b) the significance of the overall regression which combines the calibration and treatment period data, c) the difference between the slopes of the calibration and treatment regressions, and d) the difference between the intercepts of the calibration and treatment regressions. The

Table 4. Analysis of covariance for comparing regression lines.

Source	df	S_x^2	S_{xy}	S_y^2	b_1	df	SS	MS	F
Within									
Calibration	n_1-1	Eq.(3)	Eq.(4)	Eq.(2)	Eq.(6)	n_1-2	$S_y^2-(S_{xy})^2/S_x^2$	Eq.(5)	--
Treatment	n_2-1	Eq.(3)	Eq.(4)	Eq.(2)	Eq.(6)	n_2-2	" "	Eq.(5)	--
				Pooled	Error	Σ	Σ	SS/df	
Slopes	n_1+n_2-2	Σ	Σ	Σ	Eq.(6)	n_1+n_2-3	$S_y^2-(S_{xy})^2/S_x^2$	Eq.(5)	
				Slope difference		1	Slope SS - Error SS		MS/Error MS
						1	Combined SS - Slope SS		MS/Slope MS
Intercepts	n_1+n_2-1	combined data				n_1+n_2-2	$S_y^2-(S_{xy})^2/S_x^2$		

analysis can be computed by hand as shown in Table 4 (Snedecor and Cochran, 1980, p. 386). In order to perform the calculations by hand, the following are determined for the example treatment data: ΣX_i = -358.14, ΣY_i = -416.05, $\Sigma X_i Y_i$ = 1408.37, ΣX_i^2 = 1352.54, ΣY_i^2 = 1653.43, $\overline{X}$ = -3.1416, $\overline{Y}$ = -3.650, and n = 114. Therefore, S_y^2 = 135.00, S_{xy} = 101.32, and S_x^2 = 227.43. The ANCOVA is completed for the example in Table 5. The summations symbol(Σ) in Table 4 is used to signify the addition of the column entries above it.

Since the slopes were found to be different, the differences in intercepts do not have any real meaning and do not need to be calculated. That is, if slopes are different, intercepts will usually be different. However, the calculation for the test of intercepts is presented to show the method. The combined data are determined by summing the ΣX_i, ΣY_i, $\Sigma X_i Y_i$, ΣX_i^2, and ΣY_i^2 values for both the calibration and treatment periods and calculating new values for S_y^2, S_{xy}, and S_x^2. The calculation of F for the intercept uses the slope MS in the denominator. The F for the slope test uses the error MS in the denominator. A significant difference in intercepts but not slopes indicates an overall parallel shift in the regression equation.

Using SAS, an example program is listed below. This program contains both a test of the treatment regression in the PROC REG statement and a test comparing the regression lines in the PROC GLM statement.

SAS PC Program

```
Proc reg;
  model logflow2=logflow1;
run;
Proc glm;
  class period;
  model logflow2=logflow1 period
        logflow1*period;
run;
```

The treatment period regression was found to be significant based on the analysis of variance for regression (Table 7).

Table 5. Example analysis of covariance for comparing regression lines.

Source	df	S_x^2	S_{xy}	S_y^2	b_1	df	SS	MS	F
Within									
Calibration	48	70.933	78.463	148.441	1.106	47	61.650	1.3117	
Treatment	113	227.430	101.315	135.000	0.445	112	89.866	0.8024	
				Error		159	151.516	0.9529	
Slopes	161	298.363	179.778	283.441	0.603	160	175.116	1.0945	
				Slope difference		1	23.600	23.600	24.77***
						1	5.8453	5.8453	5.34*
Intercepts	162	311.671	178.762	283.492		161	180.961		

*** indicates significance at p=0.001
* indicates significance at p=0.05

Table 7. ANOVA for regression of treatment watershed runoff on control watershed runoff for the treatment period.

Source	df	MS	F	p
model	1	45.13	56.25	0.0001
error	112	0.80		
total	113			

Table 8. ANCOVA for comparing calibration and treatment regressions.

Source	df	MS	F	p
model	3	43.99	46.17	0.001
error	159	0.95		
overall	1	103.09	108.18	0.0001
intercept	1	5.47	5.74	0.0178
slope	1	23.42	24.58	0.0001

The analysis of covariance obtained in SAS output summarizes the significance of the overall model, compares the two regression equations, the regression intercepts, and slopes (Table 8). The ANCOVA indicates that the overall treatment and calibration regressions were significantly different, and that the slopes and intercepts of the equations also were different. The difference in slopes is evident in Figure 2. The slight differences in F values between the hand calculation method and the SAS output are due to rounding errors.

DISPLAYING AND INTERPRETING RESULTS

The most common methods for displaying the results include a bivariate plot of paired observations together with the calibration and treatment regression equations (Figure 2). Another useful graph is a plot of deviations ($y_{observed}$ - $y_{predicted}$) as a function of time during the treatment. The predicted values are obtained from the calibration regression equation. For the example, the plot of deviations indicates that for most paired observations, the observed value was less than that predicted by the calibration regression equation. Results should be provided of mean values for each period and each watershed. The overall results due to the treatment can be expressed as the % change based on the mean predicted and observed values. For the example, there was a 64 % reduction in mean runoff due to the treatment (Table 9).

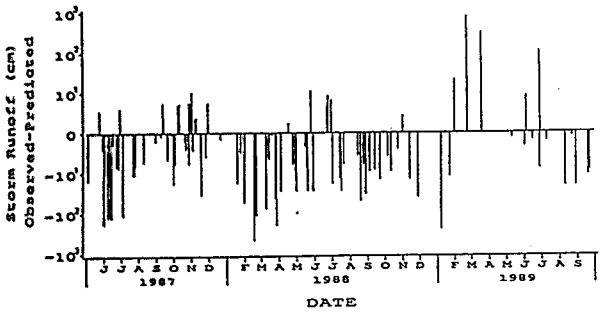

Figure 3. Observed deviations from predicted discharge.

Table 9. Mean values by period and watershed.

	Runoff (cm) x 10^{-2}	
Calibration		
Control	0.30	
Treatment	1.63	
Treatment		
Control	0.08	
Treatment	0.04	
Predicted	0.11	-64%

FURTHER READING

Bernstein, B.B. 1983. An optimum sampling design and power tests for environmental biologists. J. Environ. Mgmt. 16:35-43.

Hewlett, J.D. and L. Pienaar. 1973. Design and analysis of the catchment experiment. In Proc. Symp. Use of Small Watersheds. E.H. White (Ed.). Univ. Kentucky.

Green, R. H. 1979. Sampling design and statistical methods for environmental biologists. New York: John Wiley and Sons.

Kovner, J.L. and T.C. Evans. 1954. A method for determining the minimum duration of watershed experiments. Trans. AGU. 35(4):608-612.

Ponce, S.L. 1980. Statistical methods commonly used in water quality data analysis. WSDG Technical Paper WSDG-TP-00001. USDA Forest Service. Fort Collins, CO 80524.

Reinhart, K.G. 1967. Watershed calibration methods. In Proc. Intern. Symp. on Forest Hydrology. W.E. Sopper and H.W. Lull. (Eds.) Pergamon Press. Oxford. P.715-723.

SAS Institute, Inc. 1986. SAS system for linear models. Cary, NC 27511.

Snedecor, G.W. and W.G. Cochran. 1980. Statistical methods. 7th Ed. The Iowa State University Press. Ames, Iowa.

Wilm, H.G. 1949. How long should experimental watersheds be calibrated? Amer. Geophs. Union Trans. Part II. 618-622.

ACKNOWLEDGEMENT

This project was supported by U.S. EPA Office of Wetlands, Oceans, and Watersheds under EPA Contract No. 68-C9-0013. This fact sheet was prepared by Dr. John C. Clausen, University of Connecticut and Dr. Jean Spooner, North Carolina State University, and reviewed by Mr. Steve Dressing, U.S. EPA .