WATER RESOURCES RESEARCH, VOL. 34, NO. 12, PAGES 3393–3403, DECEMBER 1998

Peak flow responses to clear-cutting and roads in small and large basins, western Cascades, Oregon: A second opinion

Robert B. Thomas
Sherwood, Oregon

Walter F. Megahan
National Council for Air and Stream Improvement, Port Townsend, Washington

Abstract. In this paper, we conduct a reanalysis of methods and data used by *Jones and Grant* [1996]. Data from three small watersheds (60–101 ha) and three pairs of large basins (60–600 km^2) in Oregon's western Cascades were used to evaluate effects of timber harvest and road construction on peak flows. We could not detect any effect of cutting on peak flows in one of the large basin pairs, and results were inconclusive in the other two large basin pairs. One small watershed was 100% clear-cut, a second was 31% patch-cut with 6% of the area affected by road construction, and a third was held as a long-term control. Peak flows were increased up to 90% for the smallest peak events on the clear-cut watershed and up to 40% for the smallest peak flows on the patch-cut and roaded watershed. Percentage treatment effects decreased as flow event size increased and were not detectable for flows with 2-year return intervals or greater on either treated watershed. Treatment effects decreased over time but were still found after 20 years on the clear-cut watershed but for only 10 years on the patch-cut and roaded watershed.

1. Introduction

Jones and Grant [1996] used data from three small watersheds (60–101 ha) and three pairs of large basins (60–600 km^2) in Oregon's western Cascades to evaluate effects of timber harvest and road construction on peak flows. Among other things, *Jones and Grant* [1996, p. 959] concluded that (1) "forest harvesting has increased peak discharges by as much as 50% in small basins and 100% in large basins"; (2) "the major mechanism responsible for these changes is the increased drainage efficiency of basins attributable to the integration of the road/patch clear-cut network with the preexisting stream channel network"; and (3) "the statistical analysis strongly suggests that the entire population of peak discharges is shifted upward by clear-cutting and roads; we see no reason to expect the biggest storms to behave differently from the rest of the population."

Harr [1979] reviewed the results of watershed studies to evaluate the effects of forest practices on streamflow at 11 different locations along the Pacific slope ranging from northern California to British Columbia. The most common cause of increased peak flows after timber cutting was wetter, more hydrologically responsive soils in the fall caused by decreased evapotranspiration losses. Less rainfall is needed to recharge soils under such conditions, resulting in large percentage increases in peak flows. Generally, storms are small during this time of year in the Pacific northwest, so the large relative flow increases are limited to the smaller flow events. Later in the fall as soil moisture differences become less important, the magnitude of peak flow differences becomes smaller or nonexistent. Other possible causes of peak flow increases from forest practices were identified, including soil compaction, forest road construction, and differences in snow accumulation and melt rates. In general, effects tend to decrease over time as forest stands regrow. Summarizing the results of the reported studies, *Harr* [1979, p. 34] concludes "Taken collectively, results of watershed studies indicate that size of peak flows may be increased, decreased, or remain unchanged after logging. Whether or not a change occurs depends on what part of the hydrologic system is altered, to what degree, and how permanent the alteration is." Subsequent studies on both small watersheds [*Harr*, 1980; *Ziemer*, 1981; *Harr et al*., 1982; *Harr*, 1986; *Golding*, 1987; *Wright et al*., 1990] and larger river basins [*Duncan*, 1986; *Storck et al*., 1995] suggest that Harr's 1979 conclusions still apply.

The results reported by *Jones and Grant* [1996] suggest much greater effects of forest practices on peak flows than reported elsewhere on both small watersheds and large basins. In this paper, we review the analytical methods used by Jones and Grant (hereafter J and G) and report on a reanalysis of their data.

2. Small Watershed Studies

The small watershed studies consisted of three drainages ranging in size from 60 to 101 ha on the H. J. Andrews Experimental Forest near Blue River in the western Cascades of Oregon. One watershed was maintained as a long-term control throughout the life of the study. After a calibration period (Table 1) the second watershed was 100% clear-cut and burned. Roads, comprising 6% of the watershed area, were constructed on the third watershed, followed by patch-cut logging and burning on three units covering an additional 25% of the watershed for a total clear-cut area of 31%.

J and G used hydrographs from control and treated watersheds paired by the same climatic event in both basins to assess treatment effects. Peak discharges were of primary interest, but storm volume, time of peak, and begin time were also

Paper number 98WR02500.
0043-1397/98/98WR-02500$09.00

Table 1. Time Periods Used for Regression Analysis of Treatment Effects

Watershed Pair	Treatment Used	Duration
CC/C	Calibration	Oct. 1955 to July 1962
CC/C	Clear-cut, burn	Aug. 1962 to Oct. 1966
CC/C	Recovery 1	Nov. 1966 to Dec. 1971
CC/C	Recovery 2	Jan. 1972 to Dec. 1976
CC/C	Recovery 3	Jan. 1977 to Dec. 1981
CC/C	Recovery 4	Jan. 1982 to June 1988
PCR/C	Calibration	Oct. 1955 to March 1959
PCR/C	Road, patch-cut, burn	April 1959 to Dec. 1963
PCR/C	Recovery 1	Jan. 1964 to Dec. 1968
PCR/C	Recovery 2	Jan. 1969 to Dec. 1973
PCR/C	Recovery 3	Jan. 1974 to Dec. 1978
PCR/C	Recovery 4	Jan. 1979 to Dec. 1983
PCR/C	Recovery 5	Jan. 1984 to June 1988

CC, clear-cut; PCR, patchcut and roads; C, control.

analyzed to assess changes in hydrograph shape in the small basins. The hydrograph data compilation, including selection of hydrograph characteristics and matching of peaks, was done using a computer algorithm. Although we have questions about the data compilation process, we used the data provided by J and G without change for our analyses.

2.1. Effects of Forest Harvest on Peak Flows

J and G [p. 960] rejected traditional analysis of covariance for assessing the effects of harvesting on peak discharges, stating "The widely used regression-intersection method for paired watershed studies violates the assumptions of normally distributed data and uniformly distributed residuals." Instead, they used differences in logarithms of matched peaks [*Eberhardt and Thomas*, 1991] as the response variable in analyses of variance (ANOVA). This variable, which is also the logarithm of the ratio of the matched peaks, was intended to adjust for possible nontreatment changes during the study period.

J and G used treated versus control watershed comparisons to test effects of harvesting over time by peak flow event size class. Data were grouped by pretreatment, treatment, and four or five posttreatment time periods. They used quartiles of ordered peaks sampled in the control basin to define peak flow size class boundaries. Because peak discharges have skewed frequency distributions, the "large" class contains flows with return periods as frequent as 0.4 years. While this ensures adequate samples for analysis, it conflicts with usual qualitative definitions of what constitutes a large flow given that bank-full flow occurs at somewhere between the 1.5- to 2.0-year flow return interval for most alluvial streams [*Richards*, 1982] and up to 11 to 100+ years for mountain streams [*Nolan et al.*, 1987].

For the clearcut watershed, J and G found all posttreatment periods showed statistically significant ($P < 0.05$) increases for the "all events" size group, for only the first posttreatment period for the "small events" group, and for none of the posttreatment periods for the "large events" group. For the patch-cut and roaded watershed (which had road and harvest treatment periods), there were no statistically significant effects from roads alone for any of the three size groups; there were statistically significant increases for all postlogging periods for the "all events" size group and statistically significant increases for the first two postlogging periods for both the "small" and "large events" groups.

J and G also devised an index to measure the relative changes in flow for the ANOVA categories by dividing each of the category means by the pretreatment means and representing them as a percent change in peak discharge. This approach is incorrect because their indices represent percent increases in differences in the logarithms (logarithms of the ratios) of matched peaks and not in the discharges themselves. The indices give an inflated measure of actual changes in peak discharge depending on the absolute and relative magnitudes of the means.

The ANOVA model used by J and G can be expressed as

$$\log_e y_{ij} - \log_e x_{ij} = A_{ij} + e_{ij} \tag{1}$$

or rewritten as

$$\log_e y_{ij} = A_{ij} + \log_e x_{ij} + e_{ij} \tag{2}$$

where $\log_e y_{ij}$ and $\log_e x_{ij}$ represent the natural logarithms of the peak discharges in the ith treatment period and the jth flow event size category for the treated and control watersheds, respectively, A_{ij} are the cell means, and e_{ij} is random error. Equation (2) can be considered a special case of the regression model with a coefficient B_{ij} for $\log_e x_{ij}$ equal to 1. By dropping the flow event size categories and estimating a coefficient B_i for the peak flow variable $\log_e x_i$, (2) becomes the traditional analysis of covariance (ANCOVA) model with a simple linear regression for each treatment period:

$$\log_e y_i = A_i + B_i \log_e x_i + e_i. \tag{3}$$

This model is less restrictive than the J and G ANOVA, requires no additional assumptions, and is more appropriate for testing the effects of forest management activities on average peak flows. It also avoids having to assign arbitrary peak flow size classes and permits the use of additional covariates as needed.

The data sets for each treatment pair were divided into prelogging and postlogging recovery time categories of ~5 years each as shown in Table 1. There were four recovery periods for the clear-cut watershed and five for the patch-cut and roaded watershed. We applied the ANCOVA model using the data provided by J and G regressing the treated peak discharges on the matched control peaks. We used natural logarithms of the peaks since tests indicated that they would more closely meet regression assumptions.

Regressions were developed for each of the posttreatment time periods on both treated watersheds (Table 2). The fits were excellent for all 11 regressions with no indications that they should not be used to model mean response. R^2 values for the regressions averaged just under 95% with the lowest at 88.5% and all p values were zero to three decimal places. Bonferroni's adjustment [*Miller*, 1980] was used to give an overall significance level of 0.05 for the 11 regressions. There was no indication of lack of fit for any regression. Scatterplots and plots of residuals against the fitted values indicated conformity with assumptions of linearity and constant variance. Finally, we tested the slopes' values for each of the regressions to ensure that the B_i values were different from 1 and thus confirm that (3) is more appropriate than (2). Except for recovery period 2 for the clear-cut watershed, slopes for all 11 recovery period regressions are statistically different from 1.0 (using Bonferroni's adjustment for an overall significance of

Table 2. Statistics for Regressions of $\log_e$ Discharge in the Treated Watershed on Matched $\log_e$ Discharge of the Paired Control Basin for Pretreatment and Recovery Time Periods in the Small Basins

				Standard Error				
Basins	Period	Intercept	Slope	Intercept	Slope	r^2	s	n
CC/C	Pretreatment	0.694	0.998	0.053	0.025	95	0.168	77
CC/C	R1	0.749	0.849	0.103	0.043	88	0.256	52
CC/C	R2	0.876	0.962	0.064	0.033	94	0.174	58
CC/C	R3	0.699	0.874	0.078	0.035	92	0.197	54
CC/C	R4	0.725	0.890	0.062	0.028	94	0.171	67
PCR/C	Pretreatment	0.330	0.912	0.056	0.026	96	0.158	47
PCR/C	R1	0.430	0.854	0.056	0.024	96	0.159	55
PCR/C	R2	0.477	0.897	0.052	0.022	97	0.147	53
PCR/C	R3	0.471	0.906	0.044	0.021	97	0.117	56
PCR/C	R4	0.390	0.885	0.051	0.023	96	0.147	60
PCR/C	R5	0.398	0.902	0.082	0.033	94	0.158	46

CC, clear-cut; PCR, Patch-cut and roads; C, control. All regressions were significant with p values of 0.000. Original units are in cubic meters per second.

0.05), indicating that ANCOVA is the appropriate model. The regression results indicate that this method of analysis meets all statistical assumptions.

The pretreatment regressions for the clear-cut and patch-cut and road watersheds were compared to each of their respective recovery regressions to assess changes in relationships between the peak discharges over time (Tables 3 and 4). A given recovery regression was considered to be different from its pretreatment regression if at least one of the tests for a difference between slopes or intercepts was significant. Because of different sample sizes, t tests were employed to compare regression coefficients using individual regression variance estimates. *Satterthwait*'s [1946] method was used to estimate degrees of freedom, and Bonferroni's technique was used to adjust for multiple testing.

All four recovery models for the clear-cut watershed were distinct from the pretreatment model; the test for both regression coefficients was significant except the slope for the second recovery period. On the $\log_e$ scale the treatment tended to rotate the lines clockwise around the upper right corner, increasing intercepts and decreasing slopes (Figures 1a–1d). The lines for subsequent recovery periods tended to return toward the pretreatment state, but the regression for the fourth recovery period still tested different at the end of the study.

Figures 1a–1d suggest that the individual postlogging flow values at the high end of the scale tend to fall within the 95% prediction levels for individual values for the prelogging regression. We define the highest individual postlogging flow that exceeded the 95% prediction level for individual values for the prelogging regression as the maximum detectable flow increase (MDFI). MDFI values for recovery periods 1, 2, 3, and 4 amounted to 0.42, 0.28, 0.36, and 0.21 $m^3\ s^{-1}$, respectively, on the basis of control watershed flows. A total of 39 years of data are available for the control watershed. Tests showed that the lognormal distribution was appropriate to define the annual maximum flow series for the data. We made a flow frequency analysis on the $\log_e$ transformed values using the *Hazen* [1930] plotting position formula. A locally weighted robust regression (lowess) fit to the data using a smoothing factor of 0.2 was used to define flow probabilities [*Cleveland*, 1985]. On the basis of the flow frequency analysis, the flow return interval for the four MDFI values from the clear-cut watershed amounted to about 1.9, 1.3, 1.6, and 1.1 years for recovery periods 1, 2, 3, and 4, respectively.

For the patch-cut and roaded basin, ANCOVA showed that treatment effects could be detected only during the first two recovery periods, as manifested by statistical differences in the regression intercepts. Plots of the data support this result (Figures 2a and 2b). Rotation of the regression lines for this basin pair is less obvious than for the clear-cut basin but is apparent, especially during the first recovery period. The largest flow increases occur during the first recovery period. Increases during the second period are reduced as the treated regression line approaches the pretreatment regression. The treated and

Table 3. Comparisons of Differences Between Intercepts and Slopes of Pretreatment and Posttreatment Regressions for Clear-Cut and Control Watershed Pairs

Comparison	Mean Differences	Variance	t value	DF	p value
Intercept R1	0.585	0.007	7.21	91	0.000*
Intercept R2	0.309	0.006	4.06	108	0.000*
Intrecept R3	0.447	0.005	6.13	106	0.000*
Intercept R4	0.415	0.004	6.52	139	0.000*
Slope R1	−0.149	0.003	−2.96	83	0.004*
Slope R2	−0.036	0.002	−0.86	112	0.391
Slope R3	−0.124	0.002	−2.87	101	0.005*
Slope R4	−0.108	0.001	−2.87	136	0.004*

DF, degrees of freedom. Regressions are in $\log_e$ values. Asterisks denote significant comparisons ($P < 0.0063$ after Bonferroni adjustment for "experimentwise" $P < 0.05$).

Table 4. Comparisons of Differences Between Intercepts and Slopes of Pretreatment and Posttreatment Regressions for Patch-cut and Control Watershed Pairs

Comparison	Mean Difference	Variance	*t* value	DF	*p* value
Intercept R1	0.3084	0.0039	4.96	91	0.000*
Intercept R2	0.0236	0.0036	3.39	85	0.001*
Intercept R3	0.1628	0.0036	2.70	87	0.008
Intercept R4	0.1584	0.0039	2.57	92	0.012
Intercept R5	0.1060	0.0044	1.61	88	0.112
Slope R1	−0.0582	0.0012	−1.66	95	0.100
Slope R2	−0.0155	0.0011	−0.46	90	0.646
Slope R3	−0.0060	0.0011	−0.18	90	0.857
Slope R4	−0.0274	0.0012	−0.80	96	0.214
Slope R5	−0.0106	0.0018	−0.25	83	0.857

Regressions are in $\log_e$ values. Asterisks denote significant comparisons ($P < 0.005$ after Bonferroni adjustment for "experimentwise" $P < 0.05$).

control regression lines continued to converge in the three subsequent recovery periods (Figures 2c–2e). For the first two recovery periods when statistically significant peak flow increases occurred, MDFI values (based on control watershed flows) amounted to 0.23 $m^3\ s^{-1}$ in recovery period 1 and 0.39 $m^3\ s^{-1}$ in recovery period 2. Flow return intervals for the MDFI values were 1.2 and 1.8 years in recovery periods 1 and 2, respectively.

The plotted regressions as well as the low magnitude of the MDFI values indicate that flow increases are inversely propor-

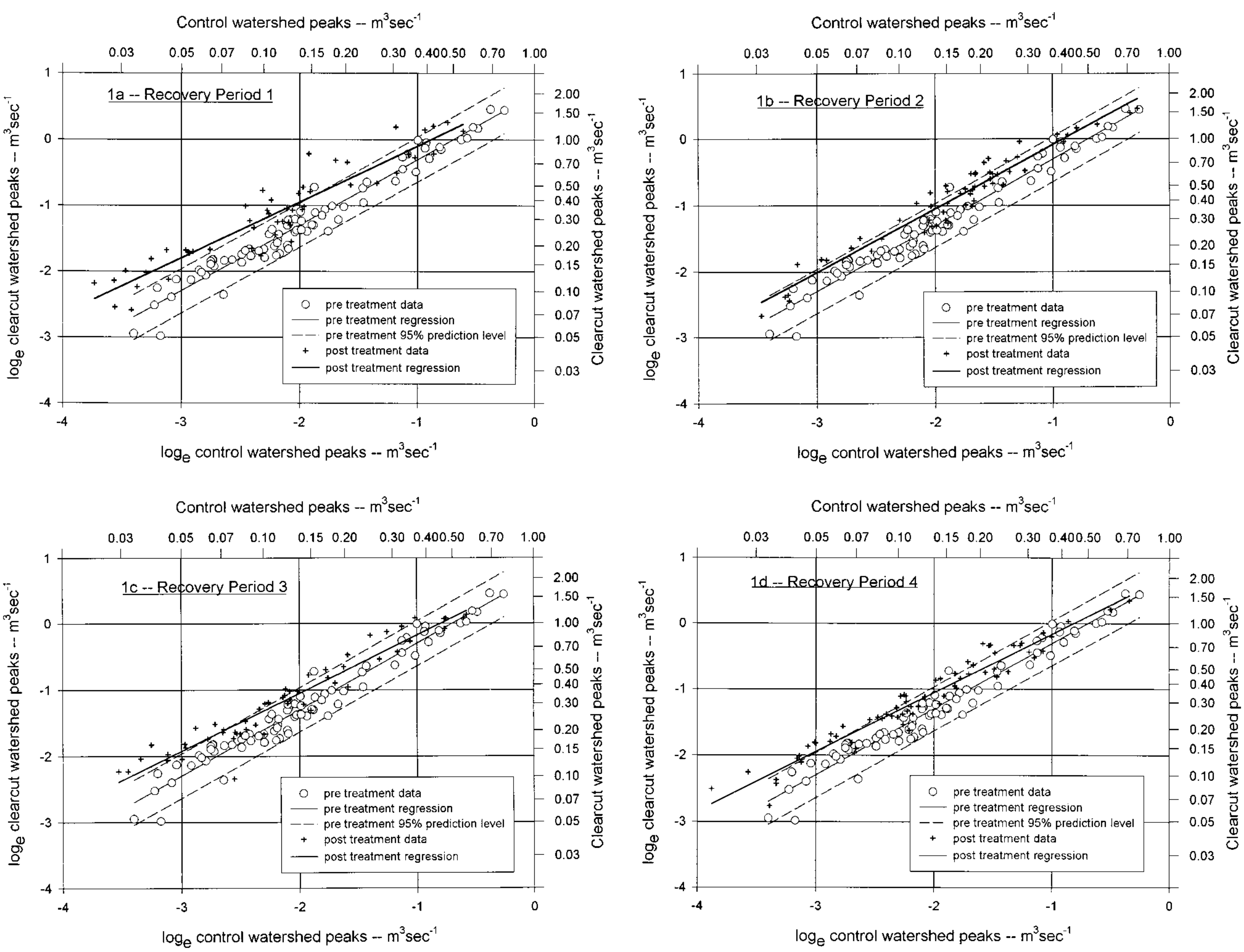

Figure 1. Regression line and the 95% individual value prediction limits for the calibration period compared to the regression line for each of the four postlogging time recovery periods for peak flows on the clear-cut logged watershed compared to the unlogged control watershed. Posttreatment regressions are statistically significant ($P < 0.05$) from the calibration regression for all four of the 5-year-long posttreatment recovery periods.

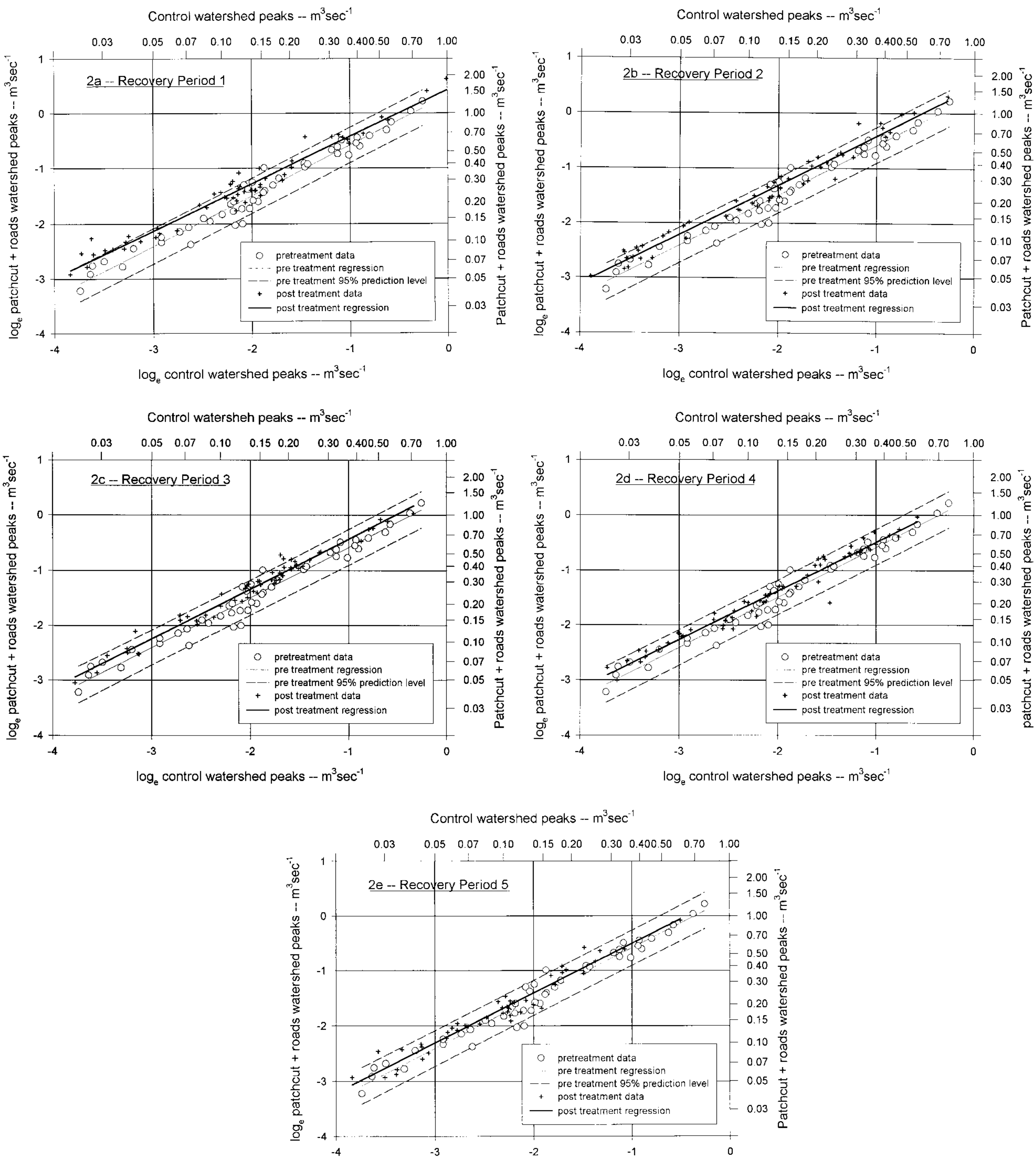

Figure 2. Regression line and the 95% individual value prediction limits for the calibration period compared to the regression line for each of the five postlogging time recovery periods for peak flows on the patch-cut and roaded watershed compared to the unlogged control watershed. Posttreatment regressions are statistically significant ($P < 0.05$) from the calibration regression for the first two of the five 5-year-long posttreatment recovery periods.

tional to flow. We can illustrate the relative magnitude of mean flow changes as the percentage increase between the treated and calibration regression lines relative to the calibration regression line as a function of flow on the control watershed. All regressions were back-transformed into discharge units and corrected for bias using the "smearing" procedure [*Duan*, 1983]. The largest bias correction was <2%. Separate relationships were developed for each of the statistically significant recovery periods on each of the study watersheds (Figures 3a and 3b). Data are plotted over the maximum range of data sampled on the control watershed for the duration of statistically significant watershed recovery (20 years for the clear-cut

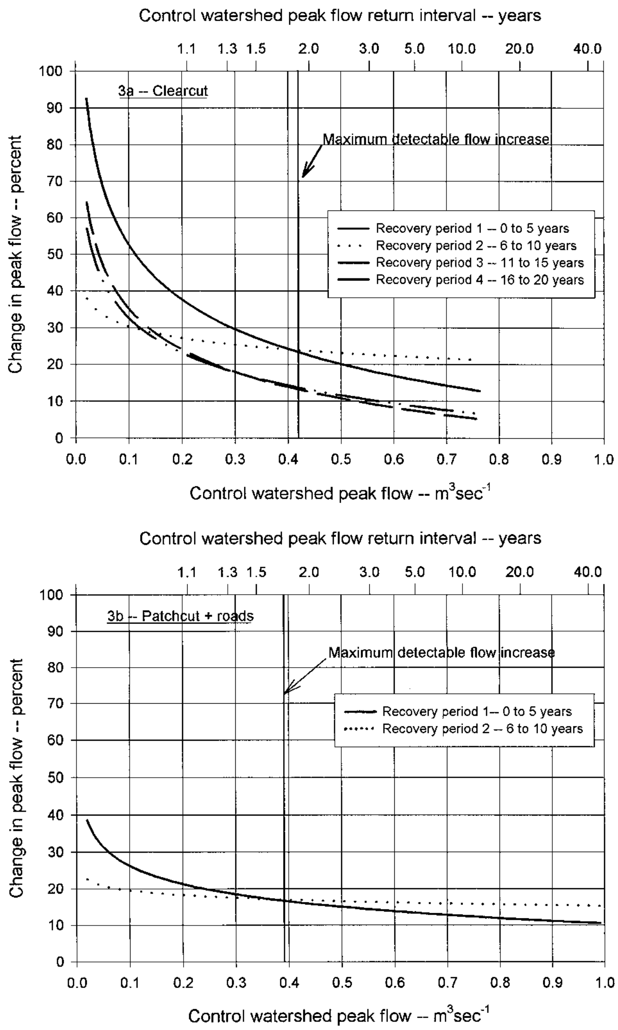

Figure 3. Percent increase in peak flow on the (a) clear-cut and (b) patch-cut and roaded watersheds in relation to peak flow and peak flow return interval on the control basin for all statistically significant recovery periods on each study watershed. The highest maximum detectable flow increase for any of the statistically significant recovery periods is shown for each of the study watersheds.

watershed and 10 years for the patch-cut and roaded watershed). The greatest flow increases occur on the clear-cut watershed during recovery period 1 (Figure 3a) and amount to about 90% at the smallest flow levels. Percent changes in flow decrease rapidly as flows increase over the range of data. At the point of maximum detectable flow the percent increase value amounts to about 23% in recovery period 1. Percent change relationships generally tend to decrease over time for the different recovery periods but reflect differences in both elevation and slope because of variance inherent in the regression equations. General trends were similar for the patch-cut and roaded watershed, but flow increases were considerably less with maximum values of about 40% in recovery period 1 as compared to the 90% increase on the clear-cut watershed. Again, increases are inversely proportional to flow rates with increases amounting to about 17% at the point of maximum detectable flow increase.

To remove time as a categorical variable, we combined data for all recovery periods and developed a model for each treated basin by regressing the peak discharge from the treated basin on the matching control peak and the time in years since the end of timber harvest activities. The time predictor variable tests the hypothesis that flow increases are maximum immediately after treatment and decrease over time as the forest stand regrows. The recovery regression for the clear-cut basin is given by

$$\log_e Q_1 = 0.812 + 0.884 \log_e Q_2 - 0.00628t \tag{4}$$

in which Q_1 and Q_2 are the matched peak discharges (in m^3 s^{-1}) in the clear-cut and control watersheds, respectively, and t is the time since treatment in years. All regression coefficients are highly significant ($P = 0.004$), the R^2 is 92.1%, and the standard error of estimate is 0.201. The recovery regression equation for the patch-cut and roaded watershed pair is

$$\log_e Q_3 = 0.506 + 0.886 \log_e Q_2 - 0.00630t \tag{5}$$

where Q_2 and t are defined above and Q_3 is the peak discharge in the patch-cut and roaded watershed. Again, all regression coefficients are highly significant ($P = 0.000$). The R^2 for this model is 96.4% and its standard error of estimate is 0.146. Residual plots are reasonable for both regressions.

The significant negative time variable in (4) and (5) indicates that a treatment effect occurred in both cases with an initial maximum effect that decays over time. Both fitted equations have essentially the same coefficients for Q_2 and t, indicating that both treated watersheds respond similarly to changes in flow on the control watershed and that both watersheds are recovering at about the same rate with time. The primary difference in the recovery models is in the intercepts; the intercept for the 100% clear-cut watershed is considerably larger than that for the 31% patch-cut and roaded watershed, suggesting that treatment effects were much more pronounced on the clear-cut watershed.

2.2. Effects of Forest Harvest on Hydrograph Characteristics

J and G used categorical analysis to evaluate effects of logging and roads on peak discharge, volume, begin time, and time of peak on differences between corresponding values for matched hydrographs. We intended to reevaluate the work and did a preliminary check of the data. While time differences between comparable characteristics were generally small (of the order of fractions of hours or hours), a substantial number of differences were very large. For example, start time of hydrograph rise in watershed 3 ranged from 5.7 days before to 4 days after the beginning of the matched hydrograph rise in the control basin. Times of peaks in watershed 1 ranged from 1 day before to 3.3 days after the watershed 2 peak.

These basins are small, adjacent, and subject to essentially identical storm inputs. Therefore the same features on paired hydrographs are likely to occur within an hour or so of each other. Differences of 5 or more days in start times of hydrographs and from 1 to 3 days in times of matched peaks are difficult to understand and suggest that at least in some cases, measurements were made on misidentified features. J and G used a computer algorithm to select and measure hydrograph properties. Designing a reliable algorithm to select features from an essentially infinite number of possible shapes is a major undertaking. Our experience in selecting such features "by hand" indicates that the rigid application of selection rules frequently produces unforeseen situations forcing their modification. In some cases the problems cannot be resolved, and storms must be discarded. Peaks are readily identified, but it is much more difficult to select starting and ending times of hydrographs and to deal with multiple hydrographs.

Because of the likely presence of mismeasured values for hydrograph components, we did not finish our analysis of differences in hydrograph characteristics. J and G used nonparametric analysis to reduce the effects of possible rogue measurements. Although these methods are forgiving for "true" outliers, they do not replace the need for care in measurement. Also, for this problem it is more useful to analyze changes in mean response. In spite of the magnitude of the task, each storm should be reviewed by a technician to ensure that the quantities are correct before analysis.

The peak flow analyses presented earlier are also dependent on computer selection of storm peaks. Because peaks are more easily identified and distributions could be normalized with $\log_e$ transformations, we felt these data were suitable for analysis. However, it is a possibility that these results were also affected by some mismeasured values.

2.3. Effects of Road Construction

Referring to the patch-cut and roaded watershed study results, and J and G [p. 972] report "discharges increased by 50% in the first 5 years after treatment and were 25 to 40% higher than pretreatment up to 25 years later" and state, "The major mechanism responsible for these changes is the increased drainage efficiency of basins attributable to the integration of the road/patch clear-cut network with the preexisting stream channel network." However, other than the first 5-year response, we find little to support these conclusions. One factor that might suggest a roading effect in our data is the fact that slopes for the two significantly different recovery period regressions on the patch-cut and roaded watershed were not different from the pretreatment slope ($P > 0.05$). This causes a smaller reduction in the change in relative flow rates with increasing peak flows on the patch-cut and roaded watershed (Figure 3b) as compared to the clear-cut watershed (Figure 3a). The implication is that road cutslopes tend to intercept more subsurface flow for larger storm events and/or that "integration of the road/patch clear-cut network with the preexisting stream channel network" is more efficient for larger storm events.

Although it is possible that such effects do occur, it is important to consider the limitations of such a conclusion given the available data. First, we can detect no treatment effect on the patch-cut and roaded watershed after 10 years. One would expect road cutslope subsurface flow interception effects and the increased efficiency of the drainage network caused by roads to last longer than 10 years. Second, we did find a nonsignificant regression slope for the second recovery period on the clear-cut watershed. It is possible that the nonsignificant regression slopes for the first two recovery periods on the patch-cut and roaded watershed resulted from normal statistical variance. Third, the experimental design does not warrant such a broad, sweeping conclusion. The results are based on an unreplicated experiment, and the pretreatment calibration period was only about 3.5 years, a very short time to serve as a basis to evaluate changes over the next 29 years. Fourth, in our opinion, the purported changes in peak flows are not supported by changes in hydrograph characteristics, given that statistical analysis of hydrograph characteristics (using either parametric or nonparametric techniques) is inappropriate because of serious questions about the reliability of the data. Fifth, other factors such as differences in watershed characteristics or effects of scouring of the channel to bedrock by massive debris flows in the patch-cut and roaded basin in 1964 could account for some or all of the "road" effects. Channel scouring would remove large woody debris, leading to reduced flow resistance, which in turn would tend to decrease time of concentration and increase peak flows.

3. Large Watershed Studies

For their large-basin analyses, J and G used sets of matched storm peak discharges for three pairs of adjacent basins on the western slope of the Cascade range in Oregon. The Breitenbush and North Santiam Rivers comprised the northernmost basin pair, consisted of 280 and 559 km^2, respectively, and had 191 storm peaks. The second pair, Blue River and Lookout Creek with 119 and 62 km^2, respectively, are centrally located (Lookout Creek contains the three small experimental watersheds) and had 148 matched storm peaks. Salmon Creek and the North Fork of the Middle Fork of the Willamette River make up the southernmost pair, contain 313 and 637 km^2, respectively, and had 171 peaks.

Apart from basin size, the large-basin data differ from those for the small experimental watersheds (and most other studies) in two major respects. Timber harvesting was not based on an experimental plan but occurred more or less continuously from 1932 to 1991 according to economic factors. As a result, there are no well defined pretreatment and posttreatment periods. In addition, all six basins have been cut, so there are no true controls. The basin in each pair having the least percentage area logged at the end of the data collection period was used as the control. For all three pairs the "control" basin was more heavily cut for part of the study period. Lookout Creek, the control for the Blue River–Lookout Creek pair, was more heavily cut for about three quarters of the period.

J and G used differences in peak discharges and cut areas in each basin pair to measure changes in mean response of storm peaks. This model ultimately depends on relationships between mean peak response and cut totals in each individual basin. That is, if the expected storm peak response in a given basin, $E(y)$, is affected by the cumulative percentage of the basin area cut, x, there is some function G such that

$$E(y) = G(x) \tag{6}$$

However, many factors affect peak flows in a basin besides area cut, so relationships given by (6) are difficult to detect. To reduce variance, J and G evaluated the three large-basin pairs using linear regressions of differences in matched unit area peaks on corresponding differences in cumulative percentages of cut areas between paired basins. However, we question their finding of valid relationships between these variables. Our reanalysis of their large-basin data is discussed in sections 3.1–3.4.

3.1. Response Variable

J and G used differences in unit area peak discharges for the response variable rather than differences in logarithms. This is problematic for two reasons. The residuals from J and G's model reject Anderson-Darling normality tests [*Minitab Incorporated*, 1995] for all three basin pairs. Each individual test was carried out at the $0.05/3 \approx 0.017$ level in accord with the

Bonferroni adjustment to ensure an overall significance of 0.05 [*Miller*, 1981]. Although removal of two or three outlying points from any of the data sets could alter the tests, visual assessment of the residual plots shows sudden slope changes in the point patterns that corroborate the test results. Also, using differences in logarithms accords with the recommendation of *Eberhardt and Thomas* [1991] sanctioned by J and G, a procedure used by them to measure changes in their small watershed analyses.

More importantly, untransformed differences in peak flows are not likely to be good measures of expected changes in mean peak discharge. The untransformed difference model used by J and G implies that differences in unit area peak discharges caused by cutting are the same regardless of storm size. This is probably not true; peaks depend heavily on the size of a storm, and changes would be expected to depend on storm magnitude as well. Because differences in logarithms are also the logarithms of the ratios, differences in the logarithms measure changes in the treated basin as a proportion of the control basin. Therefore differences in logarithms of unit area storm peaks are more likely to be independent of storm size than untransformed differences.

3.2. Predictor Variable

J and G used differences in cumulative percentages of cut areas in the two basins for matched storms as the predictor variate in their regression model. Both basins in all three pairs had been harvested, and the basin with the heavier cut changed at least once for each pair during the study period. Basins with the least percentage cut in each pair at the end of the study were used as controls and defined the direction of differencing for the response and predictor variables for the entire study period.

Differences in cumulative percentages harvested in the basin pairs were small. For the Salmon-Willamette and Breitenbush-Santiam basin pairs the range in differences never exceeded 5%. Such a small range in the predictor variable makes it difficult to show harvesting effects if they exist. The range in differences in the Blue River–Lookout Creek basin pair was larger: from −12.8 to 2.7%. Negative differences indicate times when the Lookout Creek control basin was more heavily cut than the "treated" Blue River basin. This condition existed for about the first three quarters of the study period in this basin pair.

Another concern with the predictor variate is that it does not account for basin recovery over time. Once an area is cut, it is included in the predictor whether the logging was just completed or occurred decades earlier. Therefore the variate measures the maximum effect of harvesting even though the effects of harvesting are known to diminish over time. Similarly, the predictor does not account for spatial patterns of the cut areas; a logged hectare miles away in the upper reaches of the basin is treated identically to one just above the gaging station. Any effects of harvesting on water delivery depend not only on the area cut but also on the locations of the cut areas in the basin and the resulting changes in flow routing.

3.3. J and G Model

Let y_1 and y_2 be the matched peak discharges for the treated and control basins, respectively. (In this section, y_1 and y_2 can be the peaks or their logarithms; in section 3.4, logarithms will be used explicitly.) Then if x_1 and x_2 are the corresponding cumulative percentages of areas logged in a basin pair, J and G's model can be written as

$$y_1 - y_2 = A + B(x_1 - x_2) + \varepsilon \tag{7}$$

with parameters A and B and random error ε with zero expectation. This model suggests that the function G in (6) is linear, so that for the individual basins in a pair,

$$y_1 = a_1 + b_1 x_1 + \varepsilon_1 \tag{8a}$$

$$y_2 = a_2 + b_2 x_2 + \varepsilon_2 \tag{8b}$$

with similar definitions of parameters and random errors. Subtracting (8b) from (8a) gives

$$y_1 - y_2 = a_1 - a_2 + b_1 x_1 - b_2 x_2 + \varepsilon_1 - \varepsilon_2, \tag{9}$$

Adding zero to the right side of (9) in the form $b_1 x_2 - b_1 x_2$ and rearranging gives

$$y_1 - y_2 = A + b_1(x_1 - x_2) + x_2(b_1 - b_2) + \varepsilon \tag{10}$$

in which $A = a_1 - a_2$ and $\varepsilon = \varepsilon_1 - \varepsilon_2$ (ε_1 and ε_2 are correlated so ε can have a smaller variance than either component). Equation (10) has the form of J and G's model (equation (7)) only when the third term on the right is zero; that is, if there is a true control ($x_2 = 0$) or if slopes in both linear functions (equations (8a) and (8b)) are identical ($b_2 = b_2$). Because $x_2 > 0$ in all of J and G's data sets, their model can hold only if the slopes in each pair are equal. This is a strong assumption without hydrologic support and difficult to verify because of high variation.

Consider the subset of cases in which cutting histories are linear functions of time t in years since the start of the study. Then $x_1 = c_1 t$ and $x_2 = c_2 t$ define cutting histories in the basins with slopes $c_1 > c_2 \geq 0$. (Note that for this subset of cases the control basin has a smaller percentage area cut than the treated basin at any time after the beginning of the study.) Substituting in (9) and setting $A = a_1 - a_2$ and $\varepsilon = \varepsilon_1 - \varepsilon_2$:

$$y_1 - y_2 = A + b_1 c_1 t - b_2 c_2 t + \varepsilon \tag{11}$$

But $x_1 - x_2 = (c_1 - c_2)t$, so $t = (x_1 - x_2)/(c_1 - c_2)$, which, substituted in (11), gives

$$y_1 - y_2 = A + \frac{b_1 c_1 - b_2 c_2}{c_1 - c_2}(x_1 - x_2) + \varepsilon \tag{12}$$

Therefore, in the simpler situation in which cutting histories are linear functions of time, J and G's model is valid (i.e., there are no "extra" terms) with slope:

$$B = \frac{b_1 c_1 - b_2 c_2}{c_1 - c_2} \tag{13}$$

If there is a true control ($c_2 = 0$), then, again, $B = b_1$. However, if there has been some cutting in the control basin ($c_2 \neq 0$), the value of B is complicated and surprising. The numerator can be positive or negative depending on the relative magnitudes of the quantities so that B can also be above or below zero. The magnitude of B is then largely controlled by the size of the differences between c_1 and c_2. In particular, positive values of B do not ensure positive values of b_1 and b_2.

3.4. Alternative Model

The structural problems of J and G's model can be removed by regressing the logarithm of the treated basin peaks on the

Table 5. Statistics for Regressing Logarithms of Peak Flows in "Treated" Basins on Logarithms of Peak Flows in "Control" Basins and Percent Differences in Cut Areas

Basin Pair	Regression Coefficients (Standard Errors)			n	R^2	F Statistic (p Value)
	Constant	lnConPk	CumDiff			
Blue River-Lookout Creek	0.175 (0.037)	0.729 (0.398)	0.014 (0.004)	148	0.70	11.50 (0.001)
Salmon Creek–Willamette River	−0.115 (0.053)	0.991 (0.037)	0.049 (0.016)	171	0.81	9.73 (0.002)
Breitenbush River–Santiam River	0.229 (0.058)	0.802 (0.039)	0.014 (0.015)	191	0.69	0.83 (0.362)

lnConPK, peak flows in control basins, CumDiff, percent difference in cut areas. F statistics and p values measure statistical significance of adding the CumDiff predictor; the p values should be compared to $0.05/3 \approx 0.0167$ using the Bonferroni adjustment for three tests. Only the Breitenbush-Santiam pair is not significant. The *Wetz* [1964] method was used to assess "usefulness" of the regressions. Requiring the range of CumDiff to be at least 2 times the estimated regression error entails an F statistic of $4 \times F_{1-.0167,1,n-3} \approx 23$. The cumulative difference term meets the Wetz criterion for none of the three regressions.

logarithm of the control basin peaks and the difference in percentages of areas cut (J. Lewis, personal communication, 1997). This differs from J and G's model by using the control peak as a measure of storm size and allowing it to have a coefficient other than 1. Changing to explicit notation for logarithms, the revised model can be written

$$\ln y_1 = A + C \ln y_2 + B(x_1 - x_2) + \varepsilon \tag{14}$$

This model is consistent with the mean response of the logarithms of the peaks in each basin being linear functions of the cut areas (equations (8a) and (8b)). This can be seen by adding zero to the right side of (8a) expressed as the difference between the two sides of (8b) after multiplication by b_1/b_2. After collecting terms and comparing to (14) it is clear that

$$\begin{aligned} A &= a_1 - (b_1/b_2)a_2, & C &= b_1/b_2, \\ B &= b_1, & \varepsilon &= \varepsilon_1 - (b_1/b_2)\varepsilon_2 \end{aligned} \tag{15}$$

That is, the model defined in (14) can be expressed in terms of the coefficients and errors in the individual linear models indicated in (8a) and (8b) without additional terms as occurs with J and G's model.

This model was fitted to the three large-basin data sets (Table 5). The three regressions have high R^2 values and are significant (all have p values of zero to three decimal places) because of the high correlation between matched peaks. However, the point of interest in this model is whether or not the term expressing differences in cut areas is significant, that is, if differences in percentage of areas harvested offer substantial improvement in predictive capability of the model. F statistics and p values are given for testing whether or not differences in percent of areas cut improve the prediction of mean peak responses in the treated watersheds. This variate was not significant for the Breitenbush-Santiam River pair comparing the p values against $0.0167 = 0.05/3$ to assure a Bonferroni experimental error of 0.05 for the three tests. Sets of residuals from this model passed the Anderson-Darling normality test except for Breitenbush-Santiam.

Because the F statistics are small even for the two rejected basin pairs, the "usefulness" of these relationships is questionable. Regressions are often used for prediction if the computed F statistic is greater than the F value for a selected significance level and appropriate degrees of freedom. Statistical significance, however, does not imply "hydrological significance" in the sense of having an acceptable level of predictive capability relative to error in the regression. A method for identifying "useful" as well as significant regressions was presented by *Wetz* [1964] and *Box and Wetz* [1973]. Their method requires the calculated F statistic not only to be greater than the F value for the appropriate significance level and degrees of freedom but to be greater by a multiple established for particular problems. *Ellerton* [1978] applied the procedure to a subset of terms as required here.

The *Wetz* [1964] assessment is subjective in the same sense as selecting a significance level is subjective and should depend on the context of the test being made. The method requires an evaluation of the magnitude of the error in prediction relative to the range of prediction. A minimal criterion might be to require the error to be less than half the range of prediction; that is, for the range to be at least twice the error. In this case the multiple is 4 for a "large" number of observations. Therefore, if we require that a regression have a predictive range at least twice as large as the estimated error, the regression F statistic should be at least 4 times the F value denoting statistical significance. *Draper and Smith* [1981] have summarized the procedure and present tables that help select appropriate multiples for a range of conditions. They recommend that the multiple be "at least" 4 or 5 and state that most problems require the range of the predicted values be at least 3 times the error, which results in a multiple of 6 for large data sets.

Applying the *Wetz* [1964] criterion to these three alternative regressions at the least stringent level recommended by *Draper and Smith* [1981] involves multiplying the theoretical F values for each test by 4. The F values are ~5.8 for each test (again, using the Bonferroni adjustment for the three tests to ensure an overall significance level of 0.05), so the F statistic would need to exceed $4 \times F_{1-0.0167,1,n-3} \approx 23$. The largest F statistic, for the Blue River–Lookout Creek basin pair, is 11.5, which is less than half of this value. Furthermore, this is the least rigorous requirement; if a multiple of 6 is required as recommended by Draper and Smith, the level to exceed would be ~35. By this criterion the difference term in these regressions cannot be considered useful for purposes of prediction.

4. Discussion and Conclusions

4.1. Small watershed studies

The index values defined by J and G to represent percentage changes in peak flows in small basins are not correct. Contrary to J and G's contention, the use of regression analysis (ANCOVA) to evaluate the effects of logging on the small basins is valid and appropriate. Our analyses show statistically significant increases in peak flows for all four recovery periods (a total of 20 years) on the clear-cut watershed, but only for the first two recovery periods (a total of 10 years) on the patch-cut and roaded basin. Peak flow increases expressed as a percentage change from control watershed, flows showed inverse trends with flow rate. Maximum peak flow increases occurred during the first recovery period (first 5 years) after disturbance on both treated watersheds. Maximum peak flow increases on the clearcut watershed ranged from about 90% for the smallest flow events to about 25% at the point of maximum detectable flow increase. On the patch-cut and roaded watershed, maximum peak flow increases amounted to about 40% for the smallest flow events, declining to about 15% at the point of the maximum detectable flow increase level. All flow increases are occurring at less than bank-full levels (assuming a 2-year return interval for bank-full flow) with most of the increases occurring at levels much below bank-full. There is a statistically significant, exponential decrease in treatment effects over time on both clear-cut and patch-cut and roaded watersheds. The rate of decrease in treatment effects over time is about equal on both treated watersheds. On the basis of our assessment of the small watershed treatment effects, we can find little support for concluding that forest roads had an inordinate effect on peak flows on the patch-cut and roaded watershed.

4.2. Large Watershed Studies

Obtaining large matched basin pairs suitable for studying the long-term effects of forest harvesting is difficult because of patterns of land ownership and widespread timber cutting and road building. At least some timber harvesting and road building has occurred in most large basins, so it is difficult to find a true control and it is rarely possible to direct harvesting activities for research purposes. The three large-basin data sets used by J and G reflect this problem. Although J and G detected relationships between cutting and peak flows in these basins, we believe the data do not support their conclusions. Our reasons lie in problems with the variables, with the model employed to assess changes, and in the criteria used to indicate the predictive capability of the models.

The variables used by J and G in their study are problematic. The response variable (differences in matched peaks) implies that expected changes in mean peak discharge are independent of storm size. We believe this is very unlikely and suggest that the difference in logarithms (i.e., the logarithm of the ratios) is a more reasonable approach since it measures change as a proportion of storm size in the control basin. It is also the variable used by J and G in their small basin study. The predictor variable does not account for the effect of location in the basin or the ameliorative effects of time since harvesting. We also question their assignment of a control when both basins have been more heavily cut for some storms in the study period.

However, the structure of J and G's large-basin model is our primary concern. If linear relationships exist between mean peak response and percentage area harvested in individual basins, the difference model will not show them unless there is a true uncut control basin or if slopes of the relationships in both basins are identical. Even in a simplified case where cutting is a linear function of time, a positive slope in J and G's model does not ensure positive slopes in the linear relationships for the individual basins. A model formed by regressing the "treated" peak on the "control" peak and the difference in percentage of cut area does not have these problems. We used this model with J and G's data and found the Breitenbush-Santiam basin pair not statistically significant. Also, the other two basin pairs did not pass minimal requirements for a regression to be "useful" for prediction according to the criterion developed by *Wetz* [1964].

We realize the difficulty of obtaining good information on the effects of harvesting activities on large forested basins. It is essentially impossible to plan such studies, so the available data will always be observational. Such studies require particularly strong results before coming to conclusions that conflict with numerous earlier studies. While applauding J and G's attempt to analyze the available data, we feel their conclusions [p. 959] that "landscape scale forest harvesting has produced detectable changes in peak discharges in basins ranging up to 600 km^2 in the western Cascades" and that "forest harvesting has increased peak discharges by as much as . . . 100% in large basins" are not appropriate.

Of course, our objections and analyses do not prove that there is no relationship between forest harvesting and mean response of peak discharge. Rather, they raise questions about whether or not there is sufficient support in J and G's study to conclude that forest harvesting and road building affect peak discharges in large basins. We have raised important questions about the variables used, the form of the model, and whether or not sufficient evidence has been marshaled to accept the results of their study. Data sets containing information on the effects of forest harvesting and road building on peak discharges are necessarily observational. For this reason it is important that evidence be very strong before concluding that there is a relationship, not to mention using it in making predictions concerning future changes. This is especially true considering that the results contradict the accepted understanding of numerous previous studies.

4.3. Future Research

Numerous studies show that forest cutting and road construction affect the hydrologic function of forest slopes. However, many factors such as the location and type of timber cutting, road location, road design, watershed characteristics, antecedent storm conditions, and the nature of the storm event all influence how well hydrologic changes manifest by forest cutting and roads synchronize with the natural watershed hydrograph and thus either augment or reduce peak flows. Given the complex nature of the effects of forest cutting and roads on streamflow, it is not surprising that the literature provides mixed messages about peak flow responses, including increases, no change, and decreases. It is even more difficult to detect the effects of forest cutting and road construction from historical data in large watersheds as was shown in this study, because of the lack of physical and statistical controls.

So how can forest managers forecast the effects of their activities on peak flows? First, we need more studies to better understand runoff processes from forested slopes with and without cutting and road effects with an emphasis on the role of macropores [*Ziemer*, 1992; *Chen and Wagenet*, 1992]. Process studies should be nested within carefully controlled small watershed studies to integrate watershed scale responses. At the small-watershed scale it should be possible to "switch" road effects on and off by alternating between outsloped and insloped road drainage design over time. Insloping maximizes delivery of road runoff to streams, whereas outsloping delivers runoff to the slope below the road and thus minimizes effects. Process studies of this sort should be coupled with the development and validation of physically based, distributed hydrologic models in order to forecast the effects of forest cutting and roading activities on a given watershed. Recent advances in such simulation model development [*Bowling and Lettenmaier*, 1997] are a start in this direction. Once such models have been validated against measured results from controlled small-watershed studies of roading and cutting effects, they should provide a viable means for evaluating timber harvest effects in large basins as well.

Acknowledgments. We commend Julia Jones and Gordon Grant for their efforts to understand the effects of forest cutting on peak flows in large and small watersheds. They expended considerable effort to compile and analyze a massive data set. We also want to thank Jones and Grant for meeting with us on two occasions for detailed discussions of their paper, for generously agreeing to provide us copies of their data, and for patiently responding to numerous questions. We also thank reviewers including Robert Beschta, Terry Cundy, Jack Lewis, Ray Rice, Robert Ziemer, and two anonymous individuals.

References

Bowling, L. C., and D. P. Lettenmaier, Evaluation of the effects of forest roads on streamflow in Hard and Ware Creeks, Washington, *Timber Fish Wild Water Resour. Ser.*, *Tech. Resp. 155*, 189 pp., Washington Dep. of Nat. Resour. Olympia, 1997.

Box, G. E. P., and J. M. Wetz, Criteria for judging adequacy of estimation by an approximating response function, *Tech. Rep. 9*, 99 pp., Dep. of Stat. Univ. of Wis., Madison, 1973.

Chen, C., and R. J. Wagenet, Simulation of water and chemicals in macropore soils, 1, Representation of the equivalent macropore influence and its effect on soilwater flow, *J. Hydrol.*, *130*, 105–126, 1992.

Cleveland, W. S., *The Elements of Graphing Data*, 323 pp., Wadsworth, Belmont, Calif., 1985.

Draper, N. R., and H. Smith, *Applied Regression Analysis*, 2nd ed., 709 pp., John Wiley, New York, 1981.

Duan, N., Smearing estimate: A nonparametric retransformation method, *JASA J. Am. Stat. Assoc.*, *78*(383), 605–610, 1983.

Duncan, S. H., Peak stream discharge during thirty years of sustained yield timber management in two fifth order watersheds in Washington state, *Northwest Sci.*, *60*(4), 258–264, 1986.

Eberhardt, L. L., and J. T. Thomas, Designing environmental field studies, *Ecol. Monogr.*, *61*(1), 53–73, March, 1991.

Ellerton, R. R. W., Is the regression equation adequate?—A generalization, *Technometrics*, *20*(3), 313–315, 1978.

Golding, D. L., Changes in streamflow following timber harvest of a coastal British Columbia watershed, in *Forest Hydrology and Watershed Management*, *IAHS Publ.*, *167*, 509–517, 1987.

Harr, R. D., Effects of timber harvest on streamflow in the rain-dominated portion of the Pacific northwest, paper presented at *Workshop on Scheduling Timber Harvest for Hydrologic Concerns*, U.S. For. Serv. Pac. Northwest Reg., Portland, Oreg., 1979.

Harr, R. D., Streamflow after patch logging in small drainages within the Bull Run municipal watershed, Oregon, *Res. Pap. PNW-268*, U.S. For. Serv., Pac. Northwest For. and Range Exp. Stn., Portland, Oreg., 1980.

Harr, R. D., Effects of clearcutting on rain-on-snow runoff in western Oregon: A new look at old studies, *Water Resour. Res.*, *22*(7), 1095–1100, 1986.

Harr, R. D., A. Levno, and R. Mersereau, Streamflow changes after logging 130-year-old Douglas Fir in two small watersheds, *Water Resour. Res.*, *18*(3), 637–644, 1982.

Hazen, A., *Flood Flows: A Study of Frequencies and Magnitudes*, John Wiley, New York, 1930.

Jones, J. A., and G. E. Grant, Peak flow responses to clear-cutting and roads in small and large basins, western Cascades, Oregon, *Water Resour. Res.*, *32*(4), 959–974, 1996.

Miller, R. G., Jr., *Simultaneous Statistical Inference*, 2nd ed., 299 pp., Springer-Verlag, New York, 1981.

Minitab Incorporated, *Minitab Reference Manual Release 10Xtra*, State College, Pa., 1995.

Nolan, K. M., T. M. Lisle, and H. M. Kelsey, Bankfull discharge and sediment transport in northwestern California, in *Erosion and Sedimentation in the Pacific Rim*, edited by R. L. Beschta, et al., *IAHS Publ.*, *165*, 439–450, 1987.

Richards, K., *Rivers Form and Process in Allivial Channels*, Methuen, New York, 1982.

Satterthwait, F. E., An approximate distribution of estimates of variance components, *Biometrics*, *2*, 110–114, 1946.

Storck, P., D. P. Lettenmaier, B. A. Connelly, and T. W. Cundy, Implications of forest practices on downstream flooding, *Timber Fish Wild. Ser.*, *Rep. TFW-SH20-96-001*, 100 pp., Phase II final report, Washington Dep. of Nat. Resour., Olympia, 1995.

Wetz, J. M., Criteria for judging adequacy of estimation by an approximating response function, Ph.D thesis, Dep. of Stat. Univ. of Wis., Madison, 1964.

Wright, K. A., K. H. Sendek, R. M. Rice, and R. B. Thomas, Logging effects on streamflow: Storm runoff at Casper Creek in northwestern California, *Water Resour. Res.*, *26*(7), 1657–1668, 1990.

Ziemer, R. R., Storm flow response to road building and partial cutting in small streams of northern California, *Water Resour. Res.*, *17*(4), 907–917, 1981.

Ziemer, R. R., Effect of logging on subsurface pipeflow and erosion: Coastal northern California, USA, in *Erosin, Debris Flows and Environment in Mountain Regions*, *IAHS Publ.*, *209*, 187–197, 1992.

W. F. Megahan, National Council for Air and Stream Improvement, 615 W Street, Port Townsend, WA 98368. (e-mail: megawalt@olympus.net)

R. B. Thomas, 14851 Southeast Michael Court, Sherwood, OR 97140. (e-mail: bobt@hevanet.com)

(Received January 13, 1998; revised July 17, 1998;
accepted July 27, 1998.)