WATER RESOURCES RESEARCH, VOL. 37, NO. 1, PAGES 181–183, JANUARY 2001

Reply

R. B. Thomas
Sherwood, Oregon

W. F. Megahan
National Council for Air and Stream Improvement, Sequim, Washington

1. Introduction

The effect of forest practices on streamflow is an important issue, especially with respect to downstream flooding on river basins in the Pacific Northwest. *Jones and Grant* [1996] (hereinafter referred to as J&G) presented results of several studies on small and large watersheds that they said implied large impacts of forest practices on peak flows. They concluded: (1) "forest harvesting has increased peak discharges by as much as 50% in small basins and 100% in large basins," (p. 959) (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," (p. 972) 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" (p. 972).

Together these statements could be taken to imply that forest practices increase floods in both small and large basins by as much as 100% and that much of the increase can be attributed to forest roads. We felt that J&G's conclusions were not supported by their study results. Owing to strong public interest in the effects of forest practices on flood peaks and the call by federal and state agencies for "science-based" forest management decisions, we reanalyzed the J&G data [*Thomas and Megahan*, 1998] (hereinafter referred to as T&M) to evaluate their work.

The comments by *Jones and Grant* [this issue] indicate they believe our paper "confirms and, in some cases, strengthens the conclusions" of their original paper. In this response we explain why it does not.

2. Small Basins

J&G used analysis of variance (ANOVA) on log ratios of treated versus control peak discharges to test treatment effects on storm subsets in small basins by size, time period, and season. Although J&G developed a categorical variable for event type, they present no analysis to evaluate the effects of different types of events. In contrast, T&M regressed logarithms of peak discharge in a treated basin on the corresponding control basin peaks for time periods similar to those used by J&G. J&G are critical that we did not also analyze the effects of event type and season. However, this was beyond the scope of our purpose to respond to their three conclusions stated above. Also, we were concerned about likely interactions among event size, season, and type. Although J&G recognized this problem [*Jones and Grant*, 1996, p. 963], they made no attempt to account for possible interactions in their analyses. Therefore the results of the analyses for the separate effects of event size and season, reported in Tables 2 and 3 of their paper, may well be misleading.

Paper number 2000WR900277.
0043-1397/01/2000WR900277$09.00

J&G established four event size categories with peak return periods of <0.125, 0.125–0.2, 0.2–0.4, and 0.4–100 years and designated them as "small," "small to medium," "medium to large," and "large," respectively. This categorization was one of our major disagreements with J&G's original work. While it ensures adequate sample sizes in each event size class for purposes of their ANOVA, reference to such categories can be very misleading. Most hydrologists would not characterize a 0.4 year return interval event as "large." Therefore, including such small events in the "large" ANOVA category produces less "heterogeneity" than J&G imply. Our approach avoids this pitfall by expressing responses in the treated basins as continuous functions of the corresponding control peaks. The good fits of our regressions suggest that our analysis of covariance (ANCOVA) model satisfactorily represents the relationship between the responses and the peaks for all types and sizes of events.

To evaluate the effects of time, we used the same time classification as J&G. However, we also developed regressions with a continuous time variable to show that treatment effects decrease exponentially over time (*Thomas and Megahan* [1996] equations (4) and (5) plus discussion), something that J&G were unable to do using their categorical approach.

J&G state in their comment that differences in outcome of the small-basin analyses are a "simple consequence" of the Bonferroni multiple comparison procedures used with our regression versus their ANOVA analysis. However, the differences are less simple than they imply. We agree with their statement that five comparisons to ensure an experimentwise error of 0.05 requires an individual test level of 0.005 for regression and 0.01 for ANOVA. However, the "experiments" being protected are not the same. The five regression tests cover all storm sizes so that the experiment includes all tests required for one basin pair. The five ANOVA tests, however, are needed for each peak size class so that the experiment (to be protected at the 0.05 level) in that case includes only one size class. We believe that the tests using all storm sizes for one station pair forms the more reasonable unit and that our test gives a better measure for comparison.

3. Effects of Cutting on Large Events

J&G continue to assert that the entire population of flow event sizes responds similarly to forest cutting in spite of their own analyses and ours. We agree with J&G that peak flows on small basins increase from ~40 to 90 percent but only for the smallest events. J&G called storms "large" if they had return periods between 0.4 and 100 years. They could not detect statistically significant increases on the clear-cut watershed for this category of events for any time period [*Jones and Grant*, 1996, Table 2]. J&G did show statistically significant increases in their "large" flow category for the 25% clear-cut and roaded watershed but only for the first two periods after treatment. J&G provided plots of treatment effect by event sizes and highlighted the 10 highest events for each small watershed [*Jones and Grant*, 1996, Figures 5 and 6]. There is no indication of treatment effects on these plots for the large events, a fact J&G fail to mention in their paper. Our reanalysis of small-basin responses showed a rapid drop in relative peak flow increases for increasing peak flows with no statistically significant increases detectable above slightly less than bankfull levels (2-year return interval) on either study watershed [*Thomas and Megahan*, 1998, Figures 3a and 3b].

We also agree with *Jones and Grant*'s [this issue] statement that "statistical analysis of extreme flood events from long-term records will always be inconclusive" because of small sample sizes. However, both the J&G and T&M analyses show that effects of larger events, including events of up to about a 40-year return interval, are not detectable on the small basins. Given the clear trends that are shown, we maintain that large sample sizes of extreme events are not needed to make inferences about their relative effects.

4. Large Basins

J&G state that their large-basin analyses differ from ours due to differences in choices of model, transformation, and significance level. We disagree with this characterization because such choices are critical. The model is the central feature of a formal statistical test and is determined in part by the transformations selected. If it does not correctly represent the situation being investigated, a significant result has no meaning.

T&M's large-basin model differs from J&G's model in at least two ways: transformation to logarithms and alteration of the model to include a second explanatory variable. Logarithms are justified by the residuals for this model being normally distributed and thereby better approximating assumptions required by the formal tests. More critically, logarithms measure changes in peak flows as proportions of storm size, a more realistic hypothesis than the untransformed differences used by J&G. The second variable (differences in cutting between treated and "control" basins) in T&M's model allows a true assessment of the effects of cutting on peak discharge that is not compromised by the problems in J&G's model. The statistical significance of tests on the coefficients of this variable are marginally significant statistically for only two of the three station pairs, using a 5% significance level for the three-test "experiment."

When developing predictive models, investigators should ensure that a model has acceptable predictive power. We assessed this criterion using a method by *Box and Wetz* [1973] explained by *Draper and Smith* [1981] and found that our model had little predictive power at the least rigorous level. We concur with *Draper and Smith* [1966, p. 64], who state "... that a 'statistically significant' regression ... does not necessarily mean that the equation is useful for predictive purposes. Unless the range of values predicted by the fitted equation is considerably greater than the size of the random error, prediction will often be of no value even though a "significant" F-value is obtained, since the equation will be 'fitted to the errors' only."

J&G contend our use of the *Box and Wertz* [1973] procedure amounts to changing the significance level from 0.05 to 0.0001, but we do not agree. The significance level sets a probability for not rejecting a true hypothesis and is selected to protect against making type 1 errors in particular cases. If the statistical hypothesis is rejected, an assessment of the model's predictive capability should then be made. Changing the significance level to 0.0001 alters the original hypothesis test, which we did not intend to do. These are separate and sequential questions.

As we showed in our paper, J&G's large-basin model cannot measure the effects of different cutting areas in the matched basins except in certain specific, indeterminable, and highly unlikely situations. Therefore significance of this variable does not imply a useable relationship between harvesting and peak discharges. In spite of this problem, J&G then proceed to extrapolate their model from ~50 to 550% beyond the ranges of differences in cut areas contained in their data sets. While our large-basin model allowed valid use of the difference-in-cutting variable, it was significant at low levels only in the Blue River/Lookout Creek and Salmon Creek/Willamette River basin pairs. Further analysis indicated that our model should not be used for prediction. Still, in their comment, J&G extrapolated our model to event sizes from ~40 to 200% (in log space) above the range of events in our data sets.

Because of the invalid variable in J&G's large-basin model and the lack of predictive capability in T&M's model, it is our opinion that using either model for prediction is unwarranted. Also, extrapolating such questionable models so far above the ranges of measured data is certainly not justified.

5. Effects of Roads

J&G suggest in their comment that roads have an influence on peak flows because the effects of forest harvesting are proportionally greater when roads are present, according to their data. However, they then acknowledge that other factors could cause the differences. We do not deny that roads may increase peak flows in some situations. However there is also evidence from both empirical watershed studies [*Springer and Coltharp*, 1980; *King and Tennyson*, 1984] and modeling studies (M. Wigmosta, personal communication, 2000) that roads can decrease peak flows as well. J&G (p. 972) state that "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" but provide no proof.

6. Conclusions

We agree with J&G that timber harvest practices can increase peak flows by up to ~100% but only for the smallest events. J&G present no convincing evidence in their original paper or in their comment to support their contention that all event sizes including large floods react similarly. If anything, they present evidence that the opposite is true. We still believe that their large watershed studies are inconclusive and that they present no data to prove their claim that forest roads are the major cause of flow increases.

References

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 Statistics, U. of Wisconsin, Madison, 1973.

Draper, N. R., and H. Smith, *Applied Regression Analysis*, 407 pp., John Wiley, New York, 1966.

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

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.

Jones, J. A., and G. E. Grant, Comment on "Peak flow responses to clear-cutting and roads in small and large basins, western Cascades, Oregon: A second opinion" by R. B. Thomas and W. F. Megahan, *Water Resour. Res.*, this issue.

King, J. G., and L. C. Tennyson, Alteration of streamflow characteristics following road construction in north central Idaho, *Water Resour. Res.*, *20*(8), 1159–1163, 1984.

Springer, E. P., and G. B. Coltharp, Effects of logging roads on storm hydrographs, in *Symposium on Watershed Management*, pp. 228–239, Am. Soc. Civ. Eng., New York, 1980.

Thomas, R. B., and W. F. Megahan, Peak flow responses to clear-cutting and roads in small and large basins, western Cascades, Oregon: A second opinion, *Water Resour. Res.*, *34*(12), 3393–3403, 1998.

W. F. Megahan, Forest Watershed Program, NCASI, 555 Graywolf Ridge Road, Sequim, WA 98382. (megawalt@olympus.net)

R. B. Thomas, 14851 S. E. Michael Court, Sherwood, OR 97140.

(Received February 1, 2000; revised August 11, 2000; accepted September 5, 2000.)