WATER RESOURCES RESEARCH, VOL. 45, W08416, doi:10.1029/2008WR007207, 2009

Forests and floods: A new paradigm sheds light on age-old controversies

Younes Alila,[1] Piotr K. Kuraś,[1,2] Markus Schnorbus,[3] and Robert Hudson[1]

Received 9 June 2008; revised 15 April 2009; accepted 30 April 2009; published 13 August 2009.

[1] The science of forests and floods is embroiled in conflict and is in urgent need of reevaluation in light of changing climates, insect epidemics, logging, and deforestation worldwide. Here we show how an inappropriate pairing of floods by meteorological input in analysis of covariance (ANCOVA) and analysis of variance (ANOVA), statistical tests used extensively for evaluating the effects of forest harvesting on floods smaller and larger than an average event, leads to incorrect estimates of changes in flood magnitude because neither the tests nor the pairing account for changes in flood frequency. We also illustrate how ANCOVA and ANOVA, originally designed for detecting changes in means, do not account for any forest harvesting induced change in variance and its critical effects on the frequency and magnitude of larger floods. The outcomes of numerous studies, which applied ANCOVA and ANOVA inappropriately, are based on logical fallacies and have contributed to an ever widening disparity between science, public perception, and often land-management policies for decades. We demonstrate how only an approach that pairs floods by similar frequency, well established in other disciplines, can evaluate the effects of forest harvesting on the inextricably linked magnitude and frequency of floods. We call for a reevaluation of past studies and the century-old, preconceived, and indefensible paradigm that shaped our scientific perception of the relation between forests, floods, and the biophysical environment.

Citation: Alila, Y., P. K. Kuraś, M. Schnorbus, and R. Hudson (2009), Forests and floods: A new paradigm sheds light on age-old controversies, *Water Resour. Res.*, *45*, W08416, doi:10.1029/2008WR007207.

1. Introduction

[2] ‘‘Forest hydrology has a sad history of being embroiled in controversies that never seem to get resolved’’ [*Dunne*, 1998, p. 795], a statement particularly true on the topic of forest harvesting, peak flows, and floods. These controversies have led to an ever increasing schism between science, public perception, and often management policies [*Calder*, 2002; *Kaimowitz*, 2004; *Calder et al.*, 2004; *Calder*, 2005; *Food and Agriculture Organization of the United Nations/Center for International Forestry Research* (*FAO/CIFOR*), 2005; *Laurance*, 2007; *Calder et al.*, 2007; *Bradshaw et al.*, 2009; *van Dijk et al.*, 2008], which has prompted repeated calls to educate the public and policy makers [*Calder*, 2004; *Forsyth*, 2005; *FAO/CIFOR*, 2005; *Jewitt*, 2005; *Calder and Aylward*, 2006]. Calls for questioning the science, however, have been rare, perhaps in part due to the political sensitivity of the topic (*DeWalle* [2003], echoed by *Stednick* [2008c]).

[3] Floods are a subset of the peak flow frequency distribution: A flood has a magnitude that exceeds channel capacity [*Leopold and Maddock*, 1954] and a return interval of 1–10 years or more depending on watershed physiographic and climatic characteristics [*Williams*, 1978]. Paired watershed experimental design continues to be the main reference for scientific studies examining the effects of forest harvesting on peak flows [*Robinson et al.*, 2003; *Best et al.*, 2003; *Andréassian*, 2004; *DeFries and Eshleman*, 2004; *Grant et al.*, 2008; *Bren*, 2008]. In these studies, control and treatment peak flows paired by meteorological input (referred to in this study as chronological pairing) have traditionally been evaluated with analysis of covariance (ANCOVA) [*Hewlett et al.*, 1969; *Hewlett*, 1982]. Only a few studies, however, have remained in accordance with the principles of ANCOVA [*Hewlett and Helvey*, 1970; *Harris*, 1977; *Troendle and King*, 1985; *Elder et al.*, 2006]. Over the past 4 decades, many have extended ANCOVA (regression analysis) to make inferences about treatment effect on flood events smaller and larger than an average peak flow [e.g., *Rothacher*, 1965, 1973; *Hornbeck*, 1973a; *Harr and McCorison*, 1979; *Harr et al.*, 1975; *Ziemer*, 1981; *Harr et al.*, 1982; *Verry et al.*, 1983; *Cheng*, 1989; *Riekerk*, 1989; *Wright et al.*, 1990; *Thomas and Megahan*, 1998; *Beschta et al.*, 2000; *Lewis et al.*, 2001; *Troendle et al.*, 2001; *Caissie et al.*, 2002; *Robinson et al.*, 2003; *Moore and Scott*, 2005; *Moore and Wondzell*, 2005; *Guillemette et al.*, 2005; *Stednick*, 2008a, 2008b]. This was accomplished by either interpreting a prevailing convergence of the preharvesting and postharvesting regression fits or by examining whether individual pairs of posttreatment peak flows fall within or outside the prediction limits of the

---

[1] Department of Forest Resources Management, University of British Columbia, Vancouver, British Columbia, Canada.

[2] Now at Northwest Hydraulic Consultants, Ltd., North Vancouver, British Columbia, Canada.

[3] Pacific Climate Impacts Consortium, University of Victoria, Victoria, British Columbia, Canada.

0043-1397/09/2008WR007207

pretreatment regression fit. The results of extended ANCOVA-based analyses provided apparent evidence in support of the hypothesis that forest harvesting affects the magnitude of small and medium peak flows but not necessarily that of the often vaguely defined larger floods. Apparent discomfort with these results led to the use of analysis of variance (ANOVA) [*Jones and Grant*, 1996; *Jones*, 2000]. Repeated analyses of the same data sets using both methods resulted in competing interpretations, often expressed with equal intensity [*Jones and Grant*, 1996; *Thomas and Megahan*, 1998; *Jones et al.*, 2000; *Beschta et al.*, 2000; *Jones and Grant*, 2001; *Thomas and Megahan*, 2001].

[4] Analytical procedures that compare control and treatment watershed responses corresponding to the same meteorological input (i.e., chronologically paired) have been the modus operandi of research methods in forest hydrology and a hallmark of the paired watershed study design. *Leopold* [1972] noted, however, that it has been notoriously difficult to make inferences about hydrological processes from paired watershed experiments, and similar sentiments continue to be echoed nearly 4 decades later [*Dunne*, 1998; *Eisenbies et al.*, 2007]. Despite the fact that pairing is an essential experimental design factor [*Beschta et al.*, 2000, p. 108], *Thomas and Megahan* [1998] expressed concerns that matching peak flows chronologically is a difficult and sometimes impossible task because storms in control and treatment watersheds do not always coincide in time, duration, intensity, or spatial extent. This issue is more pronounced in large than in small watersheds, and even more so in snow-dominated watersheds where events paired chronologically could be occurring 1–2 weeks apart [e.g., *Troendle and King*, 1985]. In addition, characteristics of soil moisture reservoirs and their effects on subsurface runoff travel time to streams can create lagged effects in streamflow response of the order of months in some forested watersheds [*Jones and Post*, 2004].

[5] Challenges in matching peak flow events by meteorological input are well recognized. The more critical problem with such pairing, however, is that it can conceal the true effects of forest harvesting on both the magnitude and frequency of peak flows. This flaw has not been identified previously in the literature. *Hewlett and Helvey* [1970, p. 779], in their analysis of peak flow data from the rain-dominated Coweeta Experimental Forest (North Carolina), appeared to have been once aware that changes in flood frequency cannot be addressed by chronological pairing when they stated that the "more difficult question concerns the frequency" and how this "might seriously affect flooding and flood damage downstream....Complete answers to old questions, particularly when surrounded by years of controversy, are slow in coming." While *Hewlett and Helvey* [1970] and *Leopold* [1981] recognized that the effects of forest harvesting on floods can be evaluated based on changes in magnitude or frequency, we contend that these two facets are inextricably linked and require a method that assesses both simultaneously.

[6] We have shaped our century-old and dominant scientific perception of the forests and floods relation by theoretical lines of reasoning and analytical methods that focus on changes in magnitude without invoking changes in frequency. As a consequence, we may have masked forest harvesting and deforestation effects on larger floods. Small changes in flood magnitude can be difficult or even impossible to detect, and yet they can lead to surprisingly large changes in flood return period. Our study is underpinned by the following fundamental construct: "Even modest increases in the magnitude of events in the tails of the distribution can have a very substantial impact on the expected return times of events of a given magnitude" [*Allen and Ingram*, 2002, p. 230]. We demonstrate how it is possible that the application of ANOVA and extended ANCOVA has for decades produced erroneous conclusions. Three main oversights make these methods inappropriate for quantifying the relation between forest harvesting and floods:

[7] 1. By not accounting for changes in frequency, these methods do not even reveal the correct changes in magnitude.

[8] 2. By decoupling magnitude and frequency, these methods fail to account for and preserve the all-important nonlinear and inverse relation between these two attributes.

[9] 3. The extension of these analyses, originally designed for means, does not account for a potential change in variance and its effect on the frequency and magnitude of floods.

[10] At the core of the problems associated with the chronological pairing of ANCOVA and ANOVA analyses is the fact that the return period of a flood event is defined by the ranking of peak flows in a control watershed [e.g., *Jones and Grant*, 1996; *Thomas and Megahan*, 1998; *Beschta et al.*, 2000; *Jones*, 2000; *Moore and Wondzell*, 2005]. Even if postharvest peak flows, indexed to the largest peak flows of the control watershed, do not increase relative to their expected preharvest levels under this definition, harvesting can still increase the frequency of the largest peak flows in the treatment watershed. This is because forests and forest practices and their effects on runoff production mechanisms can amplify some small- to medium-sized events in the control watershed to become among the largest peak flows on record following harvest in the treatment watershed. Chronological pairing fails to account for such changes in the rank order of events and is therefore not suitable for examining the two inextricably linked questions that should guide our investigation: What is the change in magnitude (frequency) for a peak flow with a frequency (magnitude) of interest? Such investigation requires a frequency-paired event analysis, where the watershed's hydrologic response of interest is the entire peak flow frequency distribution, as opposed to an individual peak flow event or a subsample of peak flows classified by event size or process-generating mechanism. As illustrated in Figure 1, to answer such questions we must first address the following: How do hydrologic processes at the watershed scale affect the frequency distribution? Does harvesting shift the mean only, or does it simultaneously affect the mean and variance? Does it change the form of the peak flow frequency distribution altogether?

[11] Frequency analysis, which involves fitting a frequency distribution for the sole purpose of extrapolating beyond an observed historic record, is a standard technique in the wider community of hydrological and climatological sciences [e.g., *Katz et al.*, 2002; *Meehl et al.*, 2000], but it is rarely used for examining paired watershed peak flows [e.g., *Sikka et al.*, 2003]. Flow duration curve analyses [*Klemeš*, 2000], better known among climatologists as "empirical ranking"

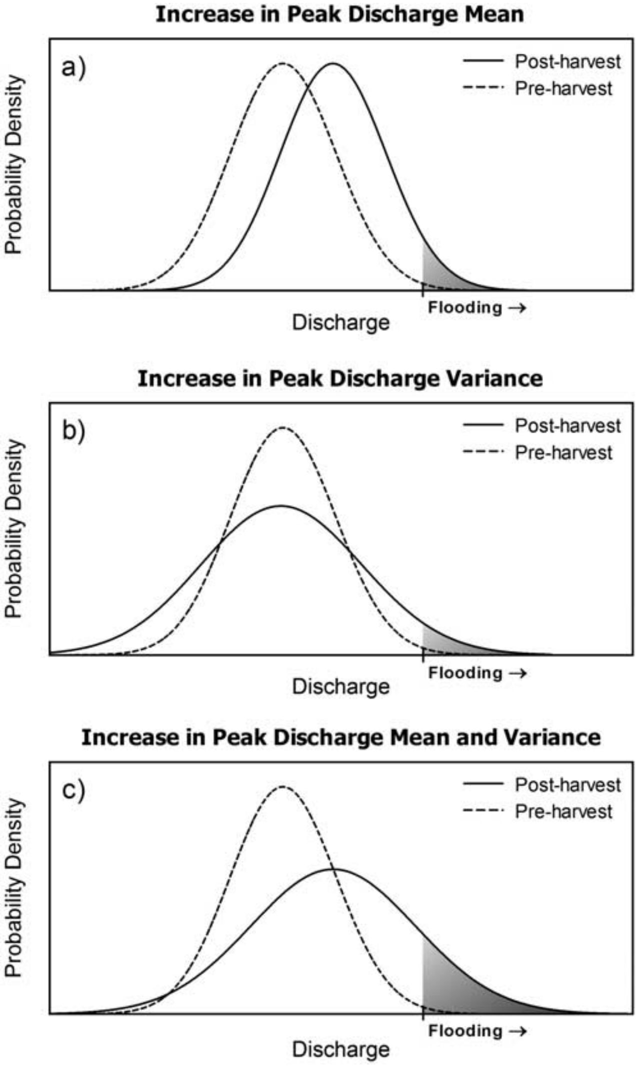

Figure 1. Schematic showing the effect on floods when (a) the mean peak flow increases, (b) the variance increases, and (c) both the mean and variance increase (the distribution shape can also change but is not depicted here). A frequency-paired method and a long enough peak flow record are absolute prerequisites to defining the threshold flood level unaffected by forests or forest harvesting.

methods [*Folland and Anderson*, 2002], which involve only a retrospective assessment of the relation between the magnitude and frequency using a historic record, has been widely used in hydrology and climatology [e.g., *Booth*, 1990; *Moscrip and Montgomery*, 1997; *Bonsal et al.*, 2001], and is not a novel concept in forest hydrology either. Flow duration curve analyses [e.g., *Lieberman and Hoover*, 1951; *McGuinness and Harrold*, 1971; *van Haveren*, 1988; *Burt and Swank*, 1992; *Swank and Vose*, 1994] or flow interval methods [e.g., *Troendle*, 1970; *Troendle and Olsen*, 1994; *Troendle et al.*, 2001] have occasionally been used for quantifying the effects of forest harvesting on other aspects of streamflow characteristics but rarely peak or flood flows. Perhaps many paired watershed studies published earlier did not have a sufficient record length to apply a frequency-paired analysis to address peak flow questions. Some of the more recent studies with a longer record, however, now provide the opportunity to apply the correct method.

[12] More recently, numerical models have been used for generating long-term data sets to enable the application of a frequency-paired analysis to quantify the effects of forest cover on peak flow regimes [e.g., *Schnorbus and Alila*, 2004; *Cuo et al.*, 2009]. Interestingly, however, changes in flood magnitude estimated by the two types of event pairing have been used, reported, and interpreted as being interchangeable not only in forest hydrology literature [e.g., *Verry et al.*, 1983; *Lewis et al.*, 2001; *Eisenbies et al.*, 2007; *Tonina et al.*, 2008] but, to our surprise, even in some urban [e.g., *Hollis*, 1975] and the more general hydrology literatures [e.g., *Brath and Montanari*, 2000; *Brath et al.*, 2006]. To our knowledge, there are no studies to date that compare the chronological and frequency-pairing methods for evaluating the effects of general land-use changes on flood flows. This study therefore has two far-reaching objectives. We first expose a set of intertwined and elusive, yet of the most fundamental construct, flaws in chronologically paired event analysis methods and illustrate how their outcomes may have reinforced a preconceived century-old scientifically indefensible paradigm of the forests and floods relation. We then demonstrate for the first time how frequency-paired event analysis, a scientific method well established in several other disciplines, can illuminate age-old forest hydrology controversies.

2. Study Sites

[13] In this study, we use peak flow data sets from two long-term paired watershed study sites in North America (one interior continental and one coastal maritime hydroclimate regime) to illustrate contrasts in the relation between forest harvesting and floods when derived from frequency-based versus conventional chronologically based peak flow analysis. Paired watershed sites in the United States provide the most complete, reliable, and heavily studied data of the effects of forestry on streamflow [*McCulloch and Robinson*, 1993]. Despite all of the other long-recognized limitations of the before-after control-impact (BACI) paired watershed design [*Underwood*, 1991, 1994; *Murtaugh*, 2000, 2002, 2003; *Grant et al.*, 2008], and although never sufficient in record length, these data sets can still be used to demonstrate how the current scientific perception of the relation between forests and floods has gone awry, as a result of the inappropriate type of event pairing.

2.1. Fool Creek of the Fraser Experimental Forest, Rocky Mountains of Colorado

[14] Fool Creek is a 289-ha treatment drainage in a paired watershed experiment at the Fraser Experimental Forest (FEF), located about 105 km northwest of Denver, Colorado [*Goodell*, 1958]. The watershed is characterized by steep slopes with predominantly northern aspects, ranging in elevation from 2896 to 3505 m above sea level (asl). Mean annual precipitation at the site is around 595 mm, 60–80% of which occurs as snow. The geology of the watershed is metamorphic, consisting of schist and gneiss derived from granite, subjected in the past to extensive glaciation. Soils are gravelly and low in fertility, except for deep alluvial soils adjacent to stream courses. Vegetation in the watershed consists of a dense mature stand (aged 250–350 years) of

lodgepole pine (*Pinus contorta*), Engelmann spruce (*Picea engelmannii*), and subalpine fir (*Abies lasiocarpa*). The upper 25% of the watershed consists of alpine terrain above the timberline. Calibration of the Fool Creek watershed began in 1943 and ended in 1954. The contiguous East St. Louis watershed (803 ha) was used as a control. Logging began in 1954 and was completed in 1956. The harvesting pattern consisted of alternating cut and leave strips of varying width (one, two, three, and six chains, where a chain equals 20.12 m) running normal to contours, with 40% of the watershed harvested (50% of the timbered area) and the forest left to regenerate naturally. Spur roads (14.2 km) built along contours were decommissioned after logging; culverts were removed on alternate roads, and all roads were grass seeded. The main haul road (5.3 km), however, is still open and subject to regular maintenance [*Alexander and Watkins*, 1977]. For more details, the reader is referred to *Goodell* [1958], *Alexander and Watkins* [1977], and *Troendle and King* [1985].

2.2. WS1 and WS3 of the H. J. Andrews Experimental Forest, Western Cascades, Oregon

[15] WS1 and WS3 are tributaries of Lookout Creek in the H. J. Andrews Experimental Forest (HJA) near Blue River, in the western Cascades of Oregon. WS1 is a 100-ha treatment watershed with an elevation range of 460–990 m asl. The watershed was 100% clear-cut harvested from 1962 to 1966 without the addition of roads. WS3 is a second neighboring treatment watershed (similar in size and characteristics) that was roaded and 25% patch-cut by 1963. At the time of harvesting, forests in both drainages mainly consisted of 100- to 500-year-old Douglas fir (*Pseudotsuga menziesii*), western hemlock (*Tsuga heterophylla*), and western red cedar (*Thuja plicata*) in closed-canopy stands. The contiguous WS2 watershed (60 ha) was used as a control for WS1 and WS3. Calibration of WS1 (WS3) started in 1955 and ended in 1961 (1958).

[16] Mean annual precipitation ranges from 2300 to more than 2500 mm at higher elevations, with over 80% falling between November and April. The three watersheds lie in the transient snow zone where precipitation tends to alternate between rain and snow. The watersheds are underlain by Tertiary and Quaternary volcanic rocks, primarily andesites and basalts, with some glacial deposits [*Sherrod and Smith*, 1989]. Soils are weakly developed with thick organic litter horizons, deeply weathered parent materials, and high stone content. Soil moisture storage and transfer is characterized by high porosity, infiltration rates, and percolation rates [*Dyrness*, 1969]. For more details, the reader is referred to *Jones and Grant* [1996], *Thomas and Megahan* [1998], and *Beschta et al.* [2000].

3. Methods

3.1. Overview

[17] The effect of forest harvesting on peak flow regimes was assessed by comparing a peak flow sample observed from a watershed following harvest (posttreatment sample) with a sample of peak flows expected to occur from the same watershed during the same period in the absence of harvesting (expected posttreatment sample). Chronological pairing assesses treatment effects on the magnitude of peak flows by comparing events derived from the same rainstorm (for pluvial regimes) or from the same annual snowmelt freshet (for nival regimes). Frequency-based pairing, on the other hand, assesses treatment effects on the magnitude of peak flows by comparing events that have the same historic probability of occurrence. As the expected peak flows are not observed, they must be modeled in some fashion. Expected posttreatment peak flows in both approaches are predicted using the pretreatment calibration regression established from each individual study (treatment peak flow regressed on paired-control peak flow). Confidence intervals for the chronologically paired assessment were derived directly from the predictive uncertainty of the calibration equation. Confidence intervals for the frequency-paired assessment are a combination of this predictive uncertainty and quantile sampling uncertainty, both estimated via Monte Carlo simulation.

[18] Observed posttreatment peak flows were adjusted to remove the effects of forest regrowth by including a recovery trend with time to a regression fit applied to chronologically paired observed and expected posttreatment discharges. We illustrate the difference in outcome between chronological and frequency-based pairing using recovery-adjusted peak flows in part to eliminate, though not always completely, the confounding factor of forest regrowth. For instance, if a watershed returns to preharvest conditions in 20 years, paired watershed data cannot reveal the effects of harvesting on a 50-year event with or without the removal of forest regrowth effects from the time series of peak flows. The purpose of adjusting peak flows to offset recovery is also to allow the effects of harvesting during the most critical period prior to any substantial forest regrowth to be evaluated, as well as the effects of other longer-lasting forest land-use changes such as deforestation. A brief account of the analysis and discussion of recovery unadjusted data sets is then presented in section 4.4.

3.2. Estimation of Predicted Discharge

[19] Peak flow magnitudes expected in the absence of forest harvesting (hereinafter referred to as "expected discharge") were predicted from simple linear regression relations between chronologically matched peak discharges in treatment and control watersheds. Regressions were constructed based on observed data at both watersheds during the pretreatment period, where the expected discharge from the treatment watershed, $\hat{Y}_i$, was estimated by

$$\hat{Y}_i = b_0 + b_1 X_i, \tag{1}$$

where X_i is the matched discharge from the control watershed. We assume that any climatic changes have not affected equation (1); however, this is unknown. Fool Creek analyses evaluate annual maximum posttreatment peak flow data sets for 1957–1985 [*Troendle and King*, 1985] and 1957–2004 [*Elder et al.*, 2006], where the pretreatment period spans 9 years (1943–1952). HJA analyses utilize peak flow data sets used by *Jones and Grant* [1996] and *Thomas and Megahan* [1998], where peak flow matching was based on matching storm events (resulting in multiple matched peaks per year). The pretreatment period spans 1955–1962 (77 events) and 1955–1959 (47 events) for WS1 and WS3, respectively, with the posttreatment period spanning 1966–1988 (93 events) and 1964–1988 (91 events) for WS1 and WS3, respectively. For some control basin discharge, X_0, measured after treatment, the predicted value of the individual observation $\hat{Y}_0$ has confidence limits that are determined analytically [*Draper and Smith*, 1998]. For HJA estimates, variables and statistics based on discharge transformed by the natural logarithm were used, and confidence limits were then back-transformed. For Fool Creek estimates, however, variables and statistics based on the actual untransformed discharge were used. The calibration regression equation (1) may not capture well the relation between control and treatment watersheds during the most extreme flood events because of short preharvest time period. The effects of this on our estimated harvesting-induced changes in magnitude and frequency of floods, if any, remain unknown.

3.3. Flow Duration Curve Analysis

[20] Flow duration curve analysis assesses the change in magnitude for a given probability, or conversely, the change in probability for a given magnitude. Given the peak discharge random variable Y, the pth quantile y_p is the peak discharge magnitude with cumulative probability p:

$$F_Y(y_p) = p, \tag{2}$$

where F_Y is the cumulative distribution function (cdf) of Y. Given a sample of observed discharge values Y_i of sample size n (i.e., chronological events for $i = 1, 2, \ldots, n$), values can be ranked such that $Y_{(j)}$ is the jth largest value in the sample, where $Y_{(1)} > Y_{(2)} > \ldots > Y_{(n)}$ (i.e., ranked events for $j = 1, 2, \ldots, n$). An estimate of the exceedance probability, $1-p$, for ranked event $Y_{(j)}$ is obtained by

$$1 - F_Y\left[Y_{(j)}\right] = \frac{j - 0.40}{n + 0.2}, \tag{3}$$

where the right-hand side of (3) is the approximately quantile-unbiased Cunnane plotting position [*Stedinger et al.*, 1993]. With p estimated from (3), the discharge event of rank j, $Y_{(j)}$, is an empirical estimate of the pth quantile y_p. Fool Creek analyses are based on the annual maximum series (AMS) while the analyses of HJA watersheds are based on partial-duration peak flow series (PDS). All frequencies and return periods estimated in this study are historic values and do not represent the future. We estimate how moderate to extreme events are changing as a result of forest harvesting via flow duration curves to avoid the more challenging assumptions associated with the use of a frequency distribution. Such empirical ranking methods have a long tradition in the climate and hydrology fields [*Folland and Anderson*, 2002, and references therein]. As we discuss later, however, ranking methods can still introduce uncertainties caused by sampling variability.

[21] Confidence limits of the ranked expected discharge sample were estimated by considering both the predictive uncertainty of the regression estimate of an individual observation and the sampling uncertainty of the quantile estimator. Although the predictive uncertainty of linear regression may be calculated analytically, the additional step of ranking the estimates precludes direct application. Instead, Monte Carlo simulation was used to determine the variance of the ranked estimate, $\hat{Y}_{(j)}$, and was structured as follows: (1) Random errors, e_i, were introduced for each estimate of $\hat{Y}_i$ for a given i by randomly sampling from a t distribution with $n-2$ degrees of freedom; (2) updated discharge estimates were calculated as $\tilde{Y}_i = \hat{Y}_i + e_i$; (3) the updated estimates, $\tilde{Y}_i$, were ranked and back-transformed in the case of WS1 and WS3, producing ranked estimates $\tilde{Y}_{(j)}$; (4) steps 1–3 were repeated for $m(= 10{,}000)$ iterations; (5) the variance, $Var[\tilde{Y}_{(j)}]$, and mean, $\bar{\tilde{Y}}_{(j)}$, for each rank, j, was estimated from the m random samples. If we consider the independent control discharge to be a series of monotonically decreasing values, then $X_{(j)} = X_j$ and, from (1), $\hat{Y}_{(j)} = \hat{Y}_j$; however, after accounting for random error, it is often the case that $\tilde{Y}_{(j)} \neq \tilde{Y}_j$. Consequently, after repeated Monte Carlo sampling we find that $\hat{Y}_{(j)}$ is not necessarily equal to $\bar{\tilde{Y}}_{(j)}$ for each rank j, such that calculation of $\hat{Y}_{(j)}$ strictly from (1) introduces a slight downward (upward) bias for large (small) return periods (as further discussed in section 4.2 regarding Figure 4). Therefore $\bar{\tilde{Y}}_{(j)}$ was used to provide an unbiased estimate of the peak discharge quantile corresponding to the event of rank j. Monte Carlo analysis also suggested that $\tilde{Y}_{(j)}$ is approximately normally distributed for all but the largest and smallest j, and for illustrative purposes, we assume that $\tilde{Y}_{(j)}$ is normally distributed with mean $\bar{\tilde{Y}}_{(j)}$ and variance $Var_1[\tilde{Y}_{(j)}]$ for all j.

[22] Monte Carlo simulation was used to assess the additional uncertainty imposed by sampling variability upon the quantile estimates at each rank, j. To facilitate sampling of empirical quantile values at the sample extremes, we fit parametric frequency distributions to the $\bar{\tilde{Y}}_{(j)}$ series. The generalized extreme value (GEV) distribution was used for the Fool Creek AMS, whereas the general Poisson-Pareto (GP) distribution was used for WS1 and WS3 PDS. Steps were as follows: (1) Fit chosen distribution to expected discharge sample, $\bar{\tilde{Y}}_{(j)}$, of size n; (2) randomly sample a discharge series of size n from the distribution using parameters specified by step 1; (3) reestimate distribution parameters; (4) using exceedance probabilities from equations (2) and (3), reestimate quantile values for each rank j; (5) repeat steps 1–4 for $m(= 10{,}000)$ iterations; (6) estimate variance for each quantile directly from m random samples for each rank, j. This second Monte Carlo analysis indicated that quantile estimates are normally distributed [*Stedinger et al.*, 1993] for this particular sample size, with mean $\bar{\tilde{Y}}_{(j)}$ and variance $Var_2[\tilde{Y}_{(j)}]$. Considering both sources of uncertainty, the expected series quantile estimate at rank j was therefore assumed to be normally distributed as $\tilde{Y}_{(j)} \sim \mathrm{N}\{\bar{\tilde{Y}}_{(j)}, Var_1[\tilde{Y}_{(j)}] + Var_2[\tilde{Y}_{(j)}]\}$, with $\alpha/2$ two-sided confidence limits given by

$$\begin{aligned} &\bar{\tilde{Y}}_{(j)} - z_{1-\alpha/2}\sqrt{Var_1\left[\tilde{Y}_{(j)}\right] + Var_2\left[\tilde{Y}_{(j)}\right]} \quad \mathit{to} \\ &\quad \bar{\tilde{Y}}_{(j)} + z_{1-\alpha/2}\sqrt{Var_1\left[\tilde{Y}_j\right] + Var_2\left[\tilde{Y}_{(j)}\right]}. \end{aligned} \tag{4}$$

3.4. Statistical Versus Hydrological and Practical Significance

[23] The general science literature has been said to have become so smitten with statistical hypothesis testing that science without statistics has become almost inconceivable and unparalleled [e.g., *Johnson*, 1999; *Anderson et al.*,

2000; *Guthery et al.*, 2001]. Therefore we plotted α-level confidence intervals around prelogging peak flow frequency relations and tabulated p values associated with differences between prelogging and postlogging sample peak flow frequency distributions. However, we also decided to report, interpret, and discuss all evidence (or lack thereof) provided by our data sets, i.e., effects of forest harvesting on peak flows regardless of their "statistical significance." Instead, we point out when appropriate how a particular physical interpretation of an effect may be hampered by sampling variability. We argue that failure to find an effect via null-hypothesis testing does not imply that there is none, but may be a result of low statistical power; this is especially true with ANOVA and ANCOVA when used with data that are not temporally independent [*Wilcox*, 1995]. Significance testing suffers from severe deficiencies at the statistical and philosophical levels [*Elliott and Brook*, 2007]. First, the tradition of assigning an arbitrary threshold probability ($p < 0.05$) above which one concludes absence of an effect and below which there is suddenly a scientifically rigorous conclusion of a difference has come under a crescendo of criticisms from many quarters [e.g., *Cohen*, 1994; *Guthery et al.*, 2001; *Burnham and Anderson*, 2002]. Second, the consequences of dismissing an effect based on statistical significance with no consideration for either physical or practical significance can be far reaching, an argument continuously emphasized across disciplines [e.g., *Klemeš*, 1974; *Kirk*, 1996; *Johnson*, 1999]. The third argument against the overreliance on statistical significance testing, and perhaps most critically pertinent to our study, is the fact that small effects (e.g., changes in flood magnitude) detected in a first step of an evaluation can translate into large effects (e.g., changes in return period) in a subsequent step of the same evaluation [*Yoccoz*, 1991, p. 107]. The overreliance on the null-hypothesis testing in experimental and observational studies has led to conflicting and internally contradicting research results, has a debilitating effect on the progress of science, and impairs the usefulness of research as a means of solving practical problems [*Schmidt*, 1996]. Null-hypothesis testing is falling out of favor in the general sciences [*Elliott and Brook*, 2007, and references therein]. Nonetheless, we still report the results of our tests but without touting our findings as important solely on the basis of their p or α values, as it is often the case in the forest hydrology literature [e.g., *Thomas and Megahan*, 1998; *Beschta et al.*, 2000]. Science is a search for themes that have the potential to collapse the chaos of nature into a set of simple, explanatory models [*Cohen and Stewart*, 1994]. "The search may be aided by the statistical hypothesis, but it starts and ends with the research hypothesis" [*Guthery et al.*, 2001, p. 382].

3.5. Adjusting for Hydrologic Recovery

[24] The nonparametric Spearman rank correlation and Mann-Whitney split-sample tests of "raw" observed peak discharge series suggest that each of these can be considered stationary. Both tests fail to detect a trend or location difference at a significance level of 0.05 for all unadjusted for recovery time series. However, we also tested for recovery trends in the observed data assuming linearity [*Thomas and Megahan*, 1998] and found the time trend coefficient to be statistically significant ($p < 0.05$) for all data sets. Consistent with our stand on the interpretation of null-hypothesis testing outcomes, we decided to adjust the observed discharges to remove the recovery trends by

$$Y_i' = Y_i + b_2 t_i \tag{5a}$$

$$\log_e Y_i' = \log_e Y_i + b_2 t_i, \tag{5b}$$

where equations (5a) and (5b) apply to FEF and HJA treatment watersheds, respectively, and recovery-adjusted analyses compare $Y'_{(j)}$ to $\tilde{Y}_{(j)}$. Our detrending uses a simple linear model and is still based on chronological event pairing. We assume that this has not unduly affected our estimated changes in magnitude and frequency, especially for hydroclimate regimes with slow recovery of forest harvesting effects as in the case of FEF (Fool Creek) and HJA (WS3). However, this can only be ascertained through a more sophisticated hydrologic recovery model as we further illustrate in section 4.4.

4. Results

4.1. Frequency and Chronological Pairing Can Lead to Contrasting Outcomes

[25] Trees grow slowly at Fool Creek due to the dry, cold, and high-elevation environment. Hence hydrologic recovery is slow, creating an opportunity to evaluate treatment effect on events larger than the mean annual flood with 48 years of posttreatment data (1957–2004). Previous analyses at Fool Creek estimated that full hydrologic recovery of the flow regime should occur within 60–80 years after logging [*Troendle and King*, 1985; *Elder et al.*, 2006]. To illustrate the value of a longer observation period, our analysis of Fool Creek was first conducted with the initial 27 years [*Troendle and King*, 1985], and subsequently with the entire 48 years [*Elder et al.*, 2006] of average daily posttreatment peak flow data.

[26] Using the frequency-paired analysis, treatment appears to have caused a change in the overall characteristics of the frequency distribution (Table 1). This results from a 30% (30%) upward shift in mean and a 30% (23%) reduction in variance for the 27-year (48-year) data set. A cross comparison of observed and expected probability density functions (pdf) for both the 27- and 48-year analyses clearly illustrates how the frequency of events larger than the mean has increased (Figure 2). The cdf for the first 27 years of record (Figure 2a) indicates that with the exception of the largest observation, treatment has shifted all peak flows upward. As discussed further in section 4.2, the lack of treatment effect on the largest flood on record cannot be interpreted in terms of physically meaningful effects without considering several sources of sampling uncertainty. This includes the empirical return period estimate of the largest few flood observations [*Stedinger et al.*, 1993], a case in point that is reinforced with the addition of 21 years of posttreatment data.

[27] The cdf for the entire 48 years of record (Figure 2b) indicates that treatment has again shifted all peak flows upward with the exception of the same largest observation on record. A comparison of Figures 2a and 2b indicates how sensitive the frequency of the largest events is to sample size. The top two events are the same in Figures 2a and 2b,

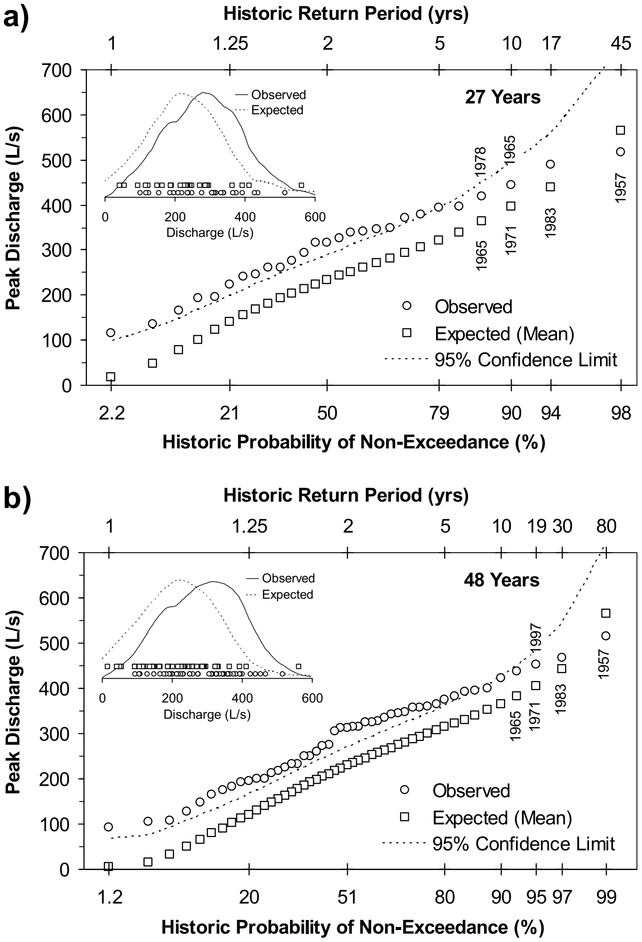

Figure 2. Flow duration curve analysis for observed and mean expected recovery adjusted daily peak flows at Fool Creek (a) 27 years posttreatment (1957–1983), and (b) 48 years posttreatment (1957–2004). Ninety-five percent confidence limits are estimated from the combined predictive uncertainty of the pretreatment calibration regression and sampling variability of the plotting position estimate; inset charts illustrate treatment-induced shifts in probability density functions (pdf's).

yet the flood event at which the postharvest cdf falls below the preharvest level has a return period in the range of 17 to 45 years (30 to 80 years) for the 27-year (48-year) time series analyses. These empirical results imply that if an additional 20 years of data were available, the maintained treatment effect could again be extended for events with return periods larger than 30 years, but only if the hydrology of the watershed has not fully or substantially recovered. Extending the observation period beyond full recovery will obviously not add new information in regard to the effects of treatment on larger events. Changes in the characteristics of the frequency distribution at Fool Creek have resulted in increasing the magnitude across the entire range of historic peak flow frequencies.

[28] Increases in the magnitude and frequency of floods across a wide range of return periods in the Fool Creek analysis could be attributable to the following (some explanations, however, are conjectural and need to be tested in future work). First, *Troendle and King* [1985] found that the difference in peak snow water equivalent is on average 9% higher in the treatment (Fool Creek) than in the control watershed (East St. Louis Creek). The authors also concluded that peak flows were positively correlated with 1 April snow water equivalent. We found this to be a moderate correlation ($r^2 \sim 0.5$), however, which suggests that there are more dominant processes at work than just snow accumulation, such as the energy involved in snowmelt. Canopy snow interception and soil storage capacities, therefore, may not be used as the only rationale to support claims such as, "during the largest rain or snowmelt events the soils and vegetative canopy will have little additional storage capacity, and under these conditions much of the rainfall or snowmelt will be converted to runoff regardless of the amount or type of vegetative cover" [*MacDonald and Stednick*, 2003, p. 13]. Second, an upper limit of the energy causing melt at Fool Creek may not have been reached within the first 50 posttreatment years. In the dry and cold climate of Fool Creek, there is a low chance that the upper limit of energy would be reached prior to the disappearance of snow. Third, empirical cdfs in snow-dominated watersheds have a relatively mild slope, which may result in larger events being more easily affected by even modest increases in energy. Fourth, higher posttreatment energy levels from all sources collectively produce higher melt rates and therefore higher runoff in stream channels, particularly when these energy increases are more synchronized from various elevation bands and aspects within the same drainage. Last but not least, although 72% of roads at Fool were decommissioned immediately after logging, the

Table 1. Relative Change in Sample Statistics and Statistical Testing for Recovery-Adjusted Peak Flows[a]

	Watershed			
	WS1 (1966–1988)	WS3 (1964–1988)	Fool Creek (1957–1983)	Fool Creek (1957–2004)
Statistic				
Mean (%)	34%	28%	30%	30%
Variance (%)	−4%	33%	−30%	−23%
Sample size (n)	93	91	27	48
Test				
Two sample K-S (D)	0.4194[b]	0.3407[b]	0.3333	0.3125[c]
Wilcoxon signed-rank (Z)	8.3644[b]	8.2819[b]	4.5047[b]	5.9847[b]

[a] Relative to mean expected samples. H_0: observed sample derives from the same population as the mean expected sample.
[b] Significant at $p < 0.05$.
[c] Significant at $p < 0.01$.

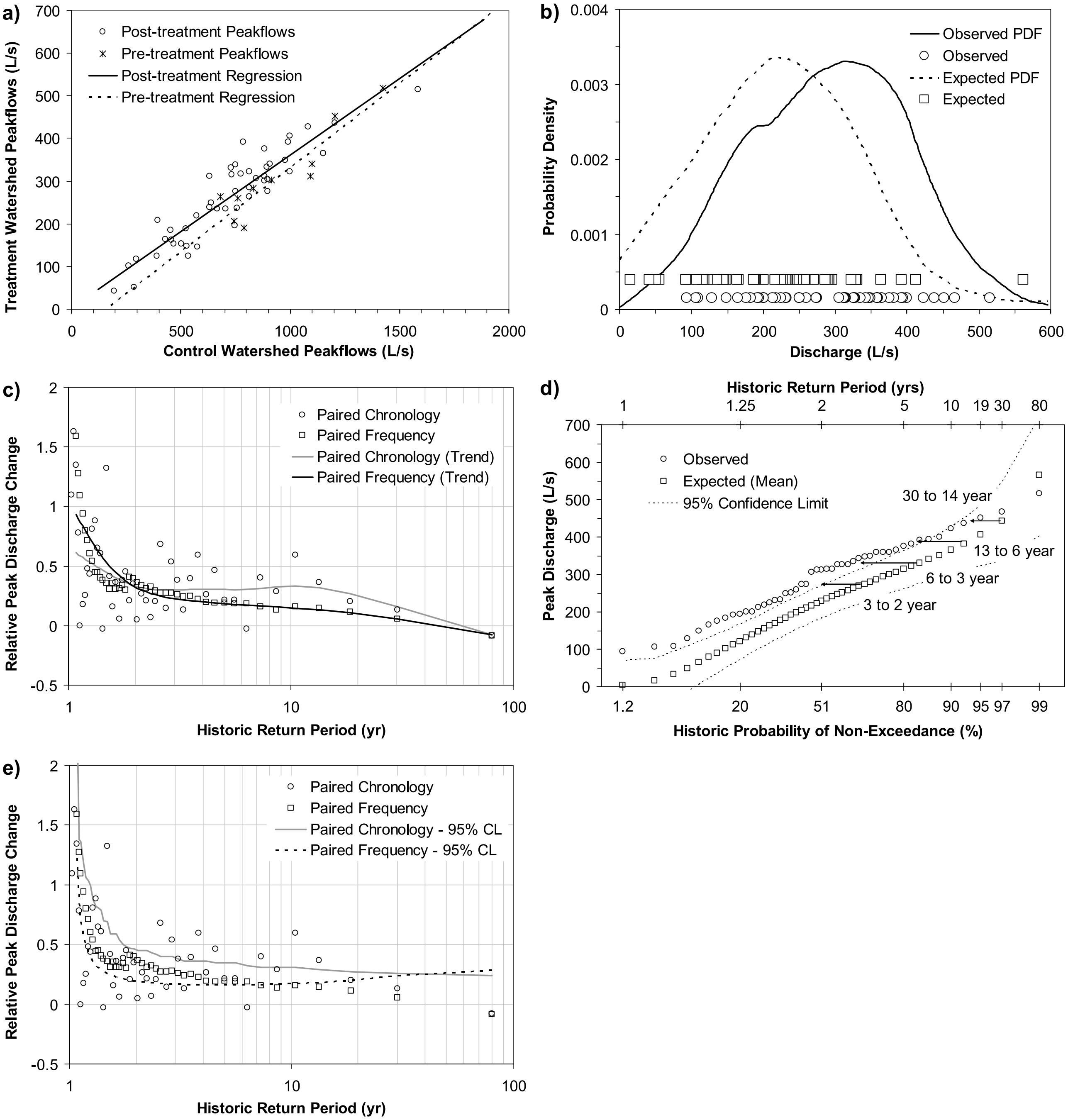

Figure 3. Contrasts between chronology-paired and frequency-paired recovery adjusted peak flow analyses illustrating (a) convergence of regression lines in traditional analysis of covariance (ANCOVA), versus (b) an overall shift in the frequency of events (pdf), (c) similar trends in plot of ΔQp versus return period, (d) changes in event frequency, and (e) statistical power of frequency-paired analysis.

remaining roads and other disturbances associated with forest harvesting may intercept runoff, which results in altering natural flow pathways to the stream network. As will be illustrated in the subsequent analysis of two data sets from HJA, roads, for instance, may have the potential to consistently increase the magnitude of peak flows over a wider range of return periods. This pattern of peak flow increases across the full range of event frequencies at Fool Creek has also been reproduced with numerical modeling of tree removal with no roads and explained by alteration of snow accumulation, snowmelt energetics, and runoff synchronization [*Schnorbus and Alila*, 2004].

[29] Figure 3 shows the difference between forest harvesting and flood relations when derived from chronologically paired versus frequency-paired event analyses. The extension of ANCOVA to events smaller and larger than an average reaffirms the perception of a rapidly vanishing treatment effect with event size, as represented by the convergence of the two regressions derived from the Fool Creek data set (Figure 3a). Such a perception was theorized using regression analysis by *Swindel and Douglass* [1984]

and supported by an overwhelming number of empirical studies (summarized by *Beschta et al.* [2000]; *Eaton and Church* [2001]; *Lewis et al.* [2001]; *Scherer and Pike* [2003]; *MacDonald and Stednick* [2003]; *Calder* [2005]; *Moore and Wondzell* [2005]; *Guillemette et al.* [2005]; *Eisenbies et al.* [2007]; *Grant et al.* [2008]). The frequency-paired approach, however, provides a much different perspective; namely, all peak flows save the largest event were shifted upward and the largest peak flows on the observed record became more frequent (Figure 3b).

[30] *Lull and Reinhart* [1972, p. 84] stated, "Changing rates of snow melt by manipulating forest cover may either increase or decrease snow-melt peaks, depending on watershed conditions, including land-use diversification, and weather sequences." This is a commonly cited line of interpretation on treatment response to harvesting [e.g., *Verry et al.*, 1983; *Harr*, 1986; *Storck et al.*, 1998; *Andréassian*, 2004; *Moore and Wondzell*, 2005; *Chang*, 2006; R. D. Harr, Effects of timber harvest on streamflow in the rain-dominated portion of the Pacific Southwest, paper presented at the Timber Harvest Workshop, Forest Service, U.S. Department of Agriculture, Portland, Oreg., 1979]. While this interpretation is physically defensible if peak flow response to treatment is defined using chronological pairing (Figures 3a and 3c), it becomes irrelevant and possibly misleading to the research question at hand when peak flows are paired by equal frequency (Figures 3d and 3e). In addition, even in cases when both methods of pairing may lead to similar decreasing trends in treatment effect versus event size (Figures 3c and 3e), only the frequency-paired analysis reveals that even smaller changes in magnitude with high type I error probabilities (commonly concluded and dismissed as lacking statistical significance) can translate into surprising changes in return period (Figure 3d). The upward shift in mean with no increase in variance appears to have caused a 3-year event to become a 2-year, a 6-year to become a 3-year, a 13-year to become a 6-year, a 19-year to become a 9-year, and a 30-year to become a 14-year event (Figure 3d). While such an intriguing pattern substantiates the hypothesis that all observed peak flows roughly doubled in expected frequency regardless of event magnitude, it also translates into an increasing change in return period with increasing magnitude. Surprisingly, this trend reveals a perception opposite to that projected by a decreasing change in peak flow magnitude with increasing return period. This new insight stems from our evaluation of the effects of forest harvesting on floods and cannot be revealed by ANOVA or ANCOVA.

[31] We may have identified here the most unfortunate drawback of chronologically paired event analyses. Extended ANCOVA conceals the true forests and floods relation. What extended ANCOVA does not reveal about the effects of forests on floods (that is, the change in frequency) is more relevant than what we believed it actually revealed (that is, the change in magnitude albeit the wrong one) in many studies spanning several decades. This is a direct consequence of the inverse and highly nonlinear relation between magnitude and frequency that cannot be accommodated by a regression fit. Such a relation forces this trend to continue or until the vertical difference between the frequency relations of Figure 3d vanishes completely. Larger samples of pretreatment and postharvest simulated peak flows in snow-dominated catchments showed how such a "no-effect" threshold is not reached even at the 50- and 100-year flood events for a wide range of peak flow metrics (hourly, daily, or weekly average annual maxima) [*Schnorbus and Alila*, 2004, Figure 9; *Brooks et al.*, 2003, Figure 6.2]. Thus a possible "no-effect" threshold return period at Fool Creek can only be ascertained by a longer record of peak flows during an unrecovered flow regime (as discussed further in section 4.2). Although they differ in governing physical processes, our findings are analogous to the statistical paradigm that helped shape the scientific perception of the effects of climate change on weather [*Katz*, 1993]. There is a strong nonlinear relation between changes in mean and changes in the probability of extremes [*Mearns et al.*, 1984; *Wigley*, 1985], where small changes in mean or in events larger than the mean can translate to large changes in return period [*Allen and Ingram*, 2002; *Schaeffer et al.*, 2005].

4.2. Uncertainties Around the Upper Tails of Sample Frequency Distributions

[32] In the absence of convincing statistical evidence, some strongly argue for decision making based on the most hydrologically plausible view point, even if the hydrologic information is only of a qualitative nature [e.g., *Klemeš*, 1974]. Nevertheless, the vertical difference between the upper tails of the preharvest and postharvest sample cdfs or the lack of treatment effect for the largest observation on record (e.g., Figures 2 and 3) cannot be interpreted in terms of physically meaningful effects without considering several sources of uncertainty.

[33] First, for any sample size there are uncertainties associated with the estimation of return periods for the few largest observations [*Stedinger et al.*, 1993] but especially the largest observation [*Folland and Anderson*, 2002]. Our analysis of 27 and 48 years of observed record at Fool Creek is consistent with the analysis of 100 years of simulated peak flows in a similar snow-dominated watershed by *Schnorbus and Alila* [2004]. Both show how the largest observation is also the only one that fell to or below the preharvest flood frequency curve. The exact return period of the largest flood event in a sample of any size cannot be determined by an empirical plotting position equation. As such, the plotting position equation could cause a premature convergence of the two preharvest and postharvest cdfs. We contend that a negative change in the magnitude of the largest few events could be real but could also be an artifact of (1) a mismatch in event return periods (i.e., estimate of return period at outer tails is uncertain) and/or (2) uncertainties in estimated expected discharges by a regression model developed based on a short sample of peak flows.

[34] Second, the relatively brief posttreatment observation period may preclude the occurrence of larger return period events, which would have better defined the upper tail of the observed and expected pdf's. In such case, the quick convergence of the two cdfs could be an artifact of sample size. Conclusions on the effect on a 50-year flood at Fool Creek, for instance, cannot be drawn without extending the observation period before substantial peak flow regime recovery, despite the apparent lack of evidence of treatment effect beyond the 30-year threshold flood (Figure 3d).

[35] Third, even if preharvest and postharvest cdfs converge or intersect around a certain range of small to medium

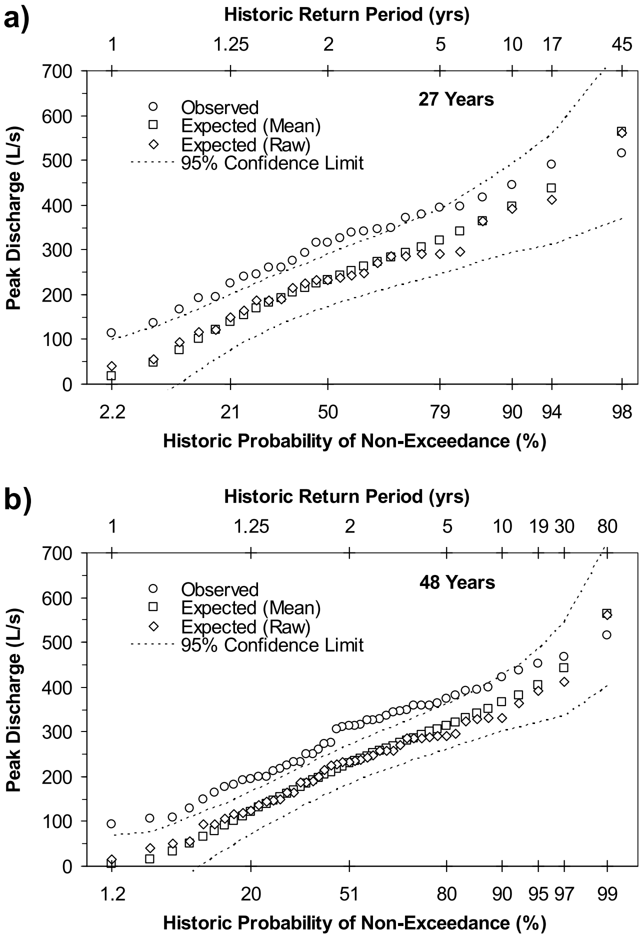

Figure 4. Flow duration curve analysis for observed, expected (raw), and mean expected recovery adjusted daily peak flows at Fool Creek (a) 27 years posttreatment (1957–1983) and (b) 48 years posttreatment (1957–2004). Ninety-five percent confidence limits are estimated from the combined predictive uncertainty of the pretreatment calibration regression and sampling variability of the plotting position estimate.

return periods, they may start to move farther apart at a higher range of return periods (as manifested by our Figure 5b of WS3 and Figure 9 of *Schnorbus and Alila* [2004]). Breaks in the slopes of empirical flood frequency curves are common and have long been attributed to possible threshold response processes, i.e., mixtures in the flood generation mechanism [*Potter*, 1958; *Archer*, 1989; *Alila and Mtiraoui*, 2002]. The simple linear regression model used to predict expected peak flow magnitudes may not be sophisticated enough to capture such dynamics. This model could introduce errors in predicting larger events, particularly if they extend beyond the range of the pretreatment calibration data [*Hornbeck*, 1973b], as is the case for the largest events observed at Fool Creek and WS3. In this case, the convergence (or lack thereof) of the observed and expected cdfs may be an artifact of the pretreatment calibration model.

[36] Fourth, the cdf of the expected peak flows may have been affected by a loss of variability associated with the use of the pretreatment calibration equation. This loss of variability inflates treatment effects more strongly for the few observations in the upper tail. In our study, we corrected for the loss of variability created by the use of the calibration equation to predict the “raw” expected posttreatment peak flows. At Fool Creek, for instance, our correction shifted the expected cdf to bring the two cdfs closer to each other but to a large extent only beyond a return period of 10 (20) years for the 27-year (48-year) data sets (Figure 4). Correction for the loss of variability has revealed the obvious: The outer tails of a frequency distribution are sensitive to even small changes in variance. This observation underscores the importance of the relation between forest harvesting and the variability of peak flows and its effects on the magnitude and frequency of flood flows, which has been overlooked by the use of chronologically paired event analyses over the last several decades (as we further elaborate in section 5). Despite uncertainties around the upper tails of sample frequency distributions, we feel that for the same sample size our Fool Creek results illustrate how frequency-paired analysis is more revealing about the effect of forest harvesting on larger floods than chronologically paired analysis methods.

4.3. Harvesting Can Affect Large Floods More So Than Small and Medium Floods

[37] Data sets of the H. J. Andrews Experimental Forest, Oregon, provide a unique opportunity to shed light on the relative effects of roads. This is because WS1 was 100% clear-cut harvested without the addition of roads while WS3 was roaded and patch-cut to only 25% harvest level. Using a frequency-paired analysis, treatment appears to have caused a change in the overall characteristics of the instantaneous peak flow distribution (Table 1) for the 23 (25) posttreatment years at WS1 (WS3). This is a result of a 34% (27%) upward shift in mean and a 4% reduction (34% increase) in variance of the WS1 (WS3) data set. A qualitative cross comparison of observed and expected pdf’s in Figure 5 illustrates how the frequency of events larger than the mean has increased. Treatment has shifted all peak flows upward at both watersheds, except the largest observation on record at WS1. The combined change in mean and variance at WS1 and WS3 appears to have roughly doubled the expected frequency of 3- to 5-year and 5- to 40-year events, respectively. Results at WS1 (harvesting without roads) support the hypothesis that the suppression of evapotranspiration and changes in snow accumulation and melt following the removal of deep-rooted conifers led to increasing deep subsurface moisture storage [*Adams et al.*, 1991; *Harr*, 1977]. Our results do not unconditionally support the contention that forest removal without roads has no effect on floods beyond a 15-year return period, despite the apparent lack of evidence of treatment effect on the largest flood observation on record (Figure 5a). As discussed earlier, this could be an artifact of the uncertainty induced by the plotting position equation [*Stedinger et al.*, 1993] in combination with the confounding effects of a small sample size. In fact, while the largest storms on record occurred at WS3 in 1964 and 1965 immediately after logging (Figure 5b), WS1 was still being logged at that time (logging occurred between 1962 and 1966) and thus peak flows from those years were not included in the WS1 time series. This omission affects the whole time series and the effect of treatment at WS1 on the largest floods therefore remains somewhat inconclusive.

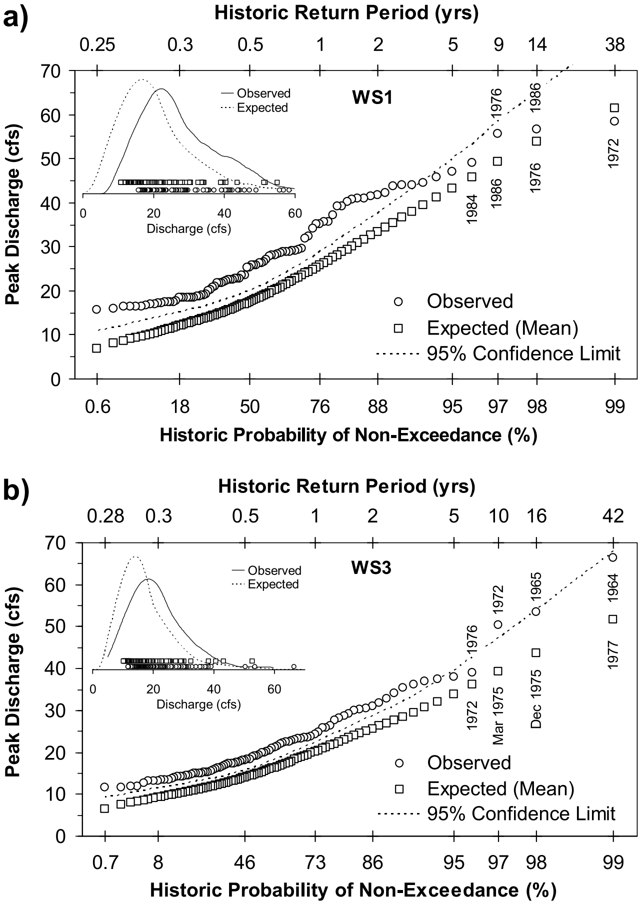

Figure 5. Flow duration curve analysis for observed and mean expected instantaneous recovery adjusted peak flows at (a) WS1 23 years posttreatment (1966–1988), and (b) WS3 25 years posttreatment (1964–1988). Ninety-five percent confidence limits are estimated from the combined predictive uncertainty of the pretreatment calibration regression and sampling variability of the plotting position estimate; inset charts illustrate treatment-induced shifts in pdf's.

[38] The addition of roads to partial clear-cutting in WS3 appears to have produced a different peak flow response (a pronounced increase in variance) from clear-cutting alone in WS1. Roads at WS3 have increased the stream network density by 21–50%, possibly speeding the watershed response by converting subsurface flow pathways to surface flow pathways [*Wemple et al.*, 1996]. In addition, large road-related debris flows in 1964 may have enhanced road and cutting effects in WS3 due to possible channel scouring to bedrock [*Jones and Grant*, 1996]. The results of our analysis at WS3 support these physical lines of reasoning, which may explain why roads either directly (rerouting subsurface flows to the stream network) and/or indirectly (scouring of streams by debris flows) have contributed to increasing peak flow variability.

[39] Treatment effect fluctuates with increasing return period for WS1 and WS3 (Figure 5), which could be an indication of treatment affecting peak flows in various ways through diverse physical processes (driven by rain, rain on snow, or snowmelt and occurring in the winter, summer, or fall). As discussed earlier, irregularities in the slopes of empirical cdfs have previously been attributed in part to a mixture of frequency distributions describing different hydroclimate flood-generation mechanisms [e.g., *Potter*, 1958; *Alila and Mtiraoui*, 2002]. Posttreatment processes at WS3 may have contributed to sustained increases in the largest peak flows on record as illustrated by the divergent preharvest and postharvest cdfs in Figure 5b. Notwithstanding the sampling uncertainty in the upper tails of frequency distributions discussed in section 4.2, this is an incidence where physical justifiability may play a significant role [*Klemeš*, 1974]. The sensitivity of the upper tail of a frequency distribution to a combined increase in mean (27%) and variance (34%) may have created the opportunity to observe such an intriguing trend. This divergent trend is a new insight that stems out of our evaluation of the effects of forest harvesting and cannot be revealed by ANOVA or ANCOVA. This further substantiates how larger events can be more sensitive to a simultaneous increase in mean and variance (WS3) than in mean alone (WS1) [*Katz and Brown*, 1992].

4.4. How Long Do Changes in Frequency Persist After Harvesting?

[40] Whether peak flow time series are adjusted or unadjusted for recovery, one must use a frequency-paired approach. Flow duration curve analysis of a single historic time series of peak flows, changing in time because of forest regrowth, however, cannot be used to properly assess hydrologic recovery. Many should be skeptical about comparisons of probability statements based on flow duration curves since the probabilistic interpretation of time series unadjusted for recovery is compromised in proportion to the departure of the underlying data from stationarity. As mentioned previously, nonparametric statistical tests of "raw" observed peak discharge series suggest that each of these can be considered stationary, but forest regrowth, slow as it might be, must have an effect on the peak flow regime. For illustrative purposes only, we show our flow duration curve analysis of the recovery-unadjusted peak flow data sets without making any claims about its scientific defensibility. Treatment continues to cause a change in the overall characteristics of the frequency distribution of unadjusted peak flows for the 23 (25) years of posttreatment data at WS1 (WS3) (Table 2). Treatment also shifts peak flow observations upwards for all return periods at WS3 (Figure 6b), but only up to the 5-year return period event at WS1 (Figure 6a). Treatment also continues to cause a change in the overall characteristics of the frequency distribution of unadjusted peak flows at Fool Creek (Table 2) and shifts peak flows upward for all return periods apart from the largest two observations (Figure 7).

[41] These results are still revealing despite being tainted by peak flow time series that are not perfectly stationary. This is especially the case when the largest observations on the postharvest cdf are evenly spread throughout the postlogging period. At Fool Creek, for instance, the fourth largest event on the postharvest cdf occurs in 1958 in the 27-year, but in 1997 in the 48-year time series analyses (Figures 7a and 7b) with a small yet similar change in magnitude, suggesting a lack of recovery even 41 years after treatment. Differences between Figures 4 and 7 support the hypothesis that forest

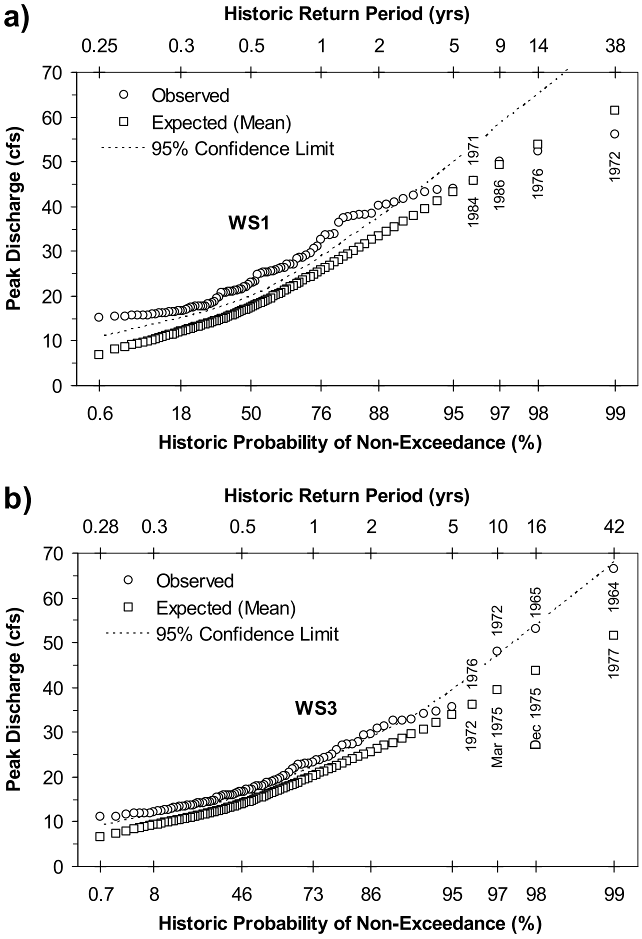

Figure 6. Flow duration curve analysis for observed and expected recovery unadjusted peak flows at (a) WS1 23 years posttreatment (1966–1988), and (b) WS3 25 years posttreatment (1964–1988). Ninety-five percent confidence limits are estimated from the combined predictive uncertainty of pretreatment calibration regression and sampling variability of the plotting position estimate; inset charts illustrate treatment-induced shifts in pdf's.

regrowth at Fool Creek has mitigated some of the changes in flood magnitude but that the high elevation, cold and dry climate, and snow environment may have caused changes in flood frequency to persist 50 years after logging. Roughly, a 10-year has still become a 5-year and a 30-year has become a 20-year event (Figure 7). Similarly, some regrowth of vegetation at WS1 [*Halpern*, 1989] appears to have mitigated the effects on the magnitude of peak flows (Figures 5a and 6a). At WS3, a 2-year has still become a 1-year and a 40-year has become a 15-year (Figures 5b and 6b). The lack of recovery of road effects at WS3 [*Gucinski and Furniss*, 2000] would be expected to cause the changes in historic return periods to persist for decades after logging or until the roads are decommissioned. Such contentions, however, may only be ascertained with a longer record that captures floods in the frequency range of the 1964 and 1965 events at WS1 and WS3. The state of flow regime recovery at any point in time after logging or the complete return of a watershed to preharvest conditions can only be fully investigated using multiple time series of peak flows, each representing a different but static stage of forest growth in a watershed driven by the same long-term climate input. Such an experiment is feasible only through a numerical modeling exercise.

5. Discussion

5.1. Decades of Science Misguided by the Irrelevant Research Hypotheses

[42] The few earlier studies that complied with the strict rules of ANCOVA, directly or indirectly, cautioned not to extend the method to events larger than a mean [*Hewlett and Helvey*, 1970; *Troendle and King*, 1985; *Harris*, 1977]. *Harris* [1977], for instance, categorically noted that it is invalid to use ANCOVA for anything other than quantifying the effect of forest harvesting on an average response. The extension of ANCOVA, however, was later indoctrinated on the grounds that (1) analysis of the mean of a naturally variable watershed response masks the importance of individual events, and (2) postharvest peak flow time series are nonstationary [*Beschta*, 1978, p. 1015]. Ironically, the same

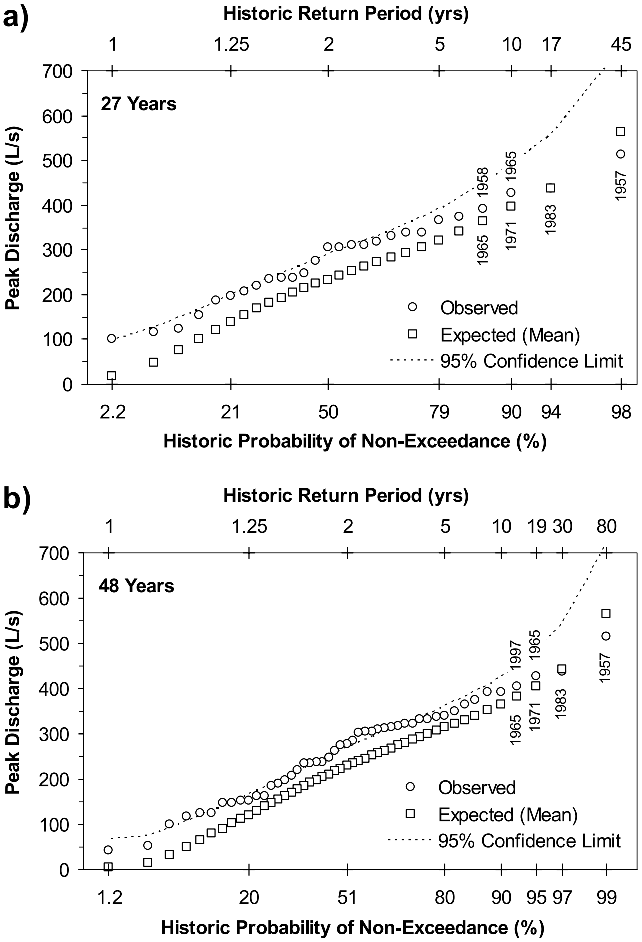

Figure 7. Flow duration curve analysis for observed and mean expected unadjusted daily peak flows at Fool Creek (a) 27 years posttreatment (1957–1983), and (b) 48 years posttreatment (1957–2004). Ninety-five percent confidence limits are estimated from the combined predictive uncertainty of the pretreatment calibration regression and sampling variability of the plotting position estimate.

Table 2. Relative Change in Sample Statistics and Statistical Testing for Raw (Recovery-Unadjusted) Peak Flows[a]

	Watershed			
	WS1 (1966–1988)	WS3 (1964–1988)	Fool Creek (1957–1983)	Fool Creek (1957–2004)
Statistic				
Mean (%)	25%	19%	19%	17%
Variance (%)	−15%	28%	−29%	−18%
Sample size (n)	93	91	27	48
Test				
Two sample K-S (D)	0.3763[b]	0.2527[c]	0.2593	0.2292
Wilcoxon signed-rank (Z)	8.1728[b]	8.278[b]	4.1924[b]	5.6257[b]

[a] Relative to mean expected samples. H_0: observed sample derives from the same population as the mean expected sample. Significant at $p < 0.05$.
[b] Significant at $p < 0.01$.
[c] Significant at $p < 0.001$.

rationale can be used against the extension of an analysis meant only for the comparison of means: (1) Watershed response is mainly stochastic, but the evaluation of individual events, as in, for example, the extended ANCOVA of *Beschta et al.* [2000] and *Moore and Scott* [2005], treats this response in part as a deterministic process. “If we are interested in identifying extremes in a collection of parameters, the focus of statistical analysis must shift from individual values to the group as a whole” [*Link and Sauer*, 1996, p. 1633]. (2) If there is no intention of extrapolating beyond the observed record, an evaluation of historic frequencies affected by harvesting using empirical ranking methods of recovery-adjusted peak flows is well established in several disciplines, and has been reported to be suitable for providing “initial, reasonably self-consistent estimates of percentiles or return periods” [*Folland and Anderson*, 2002, p. 2955]. If extrapolation is necessary, however, there are frequency analysis techniques to accommodate nonstationarity caused by forest regrowth [e.g., *Katz et al.*, 2002; *El Adlouni et al.*, 2007].

[43] It is arguably obvious that forest harvesting alters processes generating peak flows in ways that change not only event magnitude but also its rank order. It is neither obvious nor possible, however, to quantify and physically explain a forest-harvest-induced change in the magnitude and frequency of a specific flood event, or identify the “no-effect” threshold flood, outside the framework of a frequency distribution due to the highly nonlinear linkage between magnitude and frequency. The prevailing reductionist and deterministic approach to our research continues to shape the way we define, attempt to understand, and communicate hydrology and how it is affected by forest harvesting and deforestation through reasoning around individual processes [e.g., *Gilmour et al.*, 1987; *van Dijk et al.*, 2008] and the examination of single peak flow events [e.g., *Beschta et al.*, 2000; *Moore and Scott*, 2005] one at a time. *Eaton and Church* [2001, p. 33], for instance, state, “Certainly, to identify the physical processes responsible for particular hydrological behaviour, it is necessary to analyze individual events in detail.” We caution that the physical cause behind a difference in magnitude for a chronologically paired peak flow event can be misleading because the frequency (magnitude) of a specific flood event depends not only on its own magnitude (frequency) but also on the magnitude (frequency) of all other flood events in the entire observed record. Contrary to the dominant yet irrelevant research hypothesis conventionally tested by chronological pairing, our purely stochastic forests and floods research hypothesis, evaluated via frequency-paired event analysis, constrains us to compare two events of equal frequency that do not necessarily belong to the same category of meteorological events from a physical standpoint. On an earlier draft of this manuscript, one anonymous referee made the comment that this does not sound like an advantage from the viewpoint of understanding the effects of deforestation on runoff processes. We argue that research hypotheses are not meant to be adopted just for convenience.

[44] *Hewlett and Helvey* [1970] once recognized that the combination of hydrometeorological processes producing peak flow responses (similar by return period) at the outlet of each watershed is of importance in physically explaining treatment effect on the flood response of a watershed. Paired preharvest and postharvest floods may be derived from completely dissimilar antecedent hydrometeorological conditions. In snow and transient snow regimes, a difference in the frequency of such floods may be caused by altered antecedent snowpack conditions and snow surface energetics. In rain regimes, a difference in the frequency of such floods may be caused by changes in antecedent soil moisture storage and storm input. A frequency-paired event analysis may therefore lead to new ways of perceiving forest and flood relations not only in snow and transient snow hydroclimates but even in the least expected case of forest cover removal or changes with and without roads in some pure rain-dominated regimes.

[45] *Hewlett and Helvey* [1970], for instance, never ruled out a strong connection between tree felling without roads and floods in the Coweeta paired watershed experiment of the southern Appalachian Mountains of western North Carolina (a rain-dominated study site with no pronounced annual wet and dry periods, where large floods can occur any time of the year [*Eisenbies et al.*, 2007], and where forests cause persistent year-round soil moisture deficits [*Hornbeck*, 1973a]). *Hewlett and Helvey* [1970, p. 779] acknowledged that chronological pairing addresses “only part of the complex role of land use on floods.” They left the more complex part of the question open because they were once convinced that the only way to answer it is by invoking changes to the frequency of floods, when they stated

> The second and more difficult question concerns the frequency with which a combination of hydrologic depth factors, antecedent water storage, precipitation intensity, and forest removal might seriously affect

> flooding and flood damage downstream. . .. It appears that even under the humid climate of the southern Appalachians moisture deficits on forested slopes have a persistent year-round effect on quick flow, if not on storm peaks. . .. Therefore forest vegetation plays a definite role as a practical factor in downstream flooding even where cutting does not disturb the soil's infiltration capacity.

[46] Ironically, *Hewlett and Helvey* [1970] continues to be dubiously cited as one of the benchmark studies in support of the current scientific perception as best illustrated by this quote: "Studies in America [*Hewlett and Helvey*, 1970], and south Africa [*Hewlett and Bosch*, 1984] were amongst some of the first to question the importance of the link between forest conversion and flooding" [*FAO-CIFOR*, 2005, p. 5]. Flow duration curve analysis using some 25 years of data revealed how the conversion of two mixed oak and hickory-covered watersheds to white pine stands, also at the rain-dominated Coweeta, decreased peak flows across the full range of historic frequencies up to and approximately the 25-year event (*Swank and Vose* [1994]; also cited by *Riedel et al.* [2005]). To use the terminology of *Kuhn* [1970], such a rare frequency-based "anomaly" in decades of published "normal" science (misguided by the wrong type of research hypotheses as a result of chronological pairing) begs the question, Could this pattern extend to larger return periods if we have a longer simulated or observed record that captures the effect of forests on larger floods?

5.2. The Lost Dimension of Frequency in the Science of Forests and Floods

[47] ANCOVA and ANOVA have fueled the scientific debate on a topic historically recognized as being a "high-level political arena" in the United States [*Hirt*, 1994] and globally [*Ives*, 2006, p. 187]. ANOVA [e.g., *Jones and Grant*, 1996] and extended ANCOVA [e.g., *Thomas and Megahan*, 1998; *Beschta et al.*, 2000; *Moore and Wondzell*, 2005; *Moore and Scott*, 2005] type analyses are both based on the same inappropriate chronological pairing. Neither method was designed to reveal changes in flood frequency or leads to correct changes in flood magnitude (because magnitude and frequency are decoupled). Published paired watershed studies rarely invoke changes in frequency, and since reported changes in magnitude are based on chronological pairing, frequency changes cannot be inferred indirectly. Frequencies invoked in a few studies [*Thomas and Megahan*, 1998; *Beschta et al.*, 2000; *Lewis et al.*, 2001; *Moore and Wondzell*, 2005] had used inappropriate surrogates of event size since return period estimates were based on the ranking of control watershed peak flows.

[48] Our results show how changes in frequency are an essential, yet long-ignored dimension in the analysis of the relation between forest harvesting and floods. Of equal importance, our study illustrates conceptually and empirically how a change in flood magnitude, estimated by the chronologically paired analyses of ANOVA and extended ANCOVA, is incorrect and can be misleading. Our frequency-paired event analyses at Fool Creek and WS3 project a scientific perception that is inconsistent with the outcomes of most previous studies utilizing extended ANCOVA (summarized in several reviews such as *Troendle et al.* [2001], *MacDonald and Stednick* [2003], *Eaton and Church* [2001], *Scherer and Pike* [2003], *Moore and Wondzell* [2005], *Troendle et al.* [2006], and *Grant et al.* [2008]). Our analysis has demonstrated for the first time that a snow-dominated paired watershed study revealed a substantial change in frequency to events as large as 13-year (becoming a 6-year), 19-year (becoming a 9-year) and 30-year (becoming a 14-year) floods. Although our analysis at Fool, with only 48 years of record, introduces some sampling uncertainty, larger samples of preharvest and postharvest (with no roads) simulated peak flows in snow-dominated catchments showed how this changing pattern of event return period may not actually end at the 30-year but can extend to the 50-year and 100-year events [*Schnorbus and Alila*, 2004, Figure 9; *Brooks et al.*, 2003, Figure 6.2]. Forest harvesting at Fool Creek shifted all peak flows upward even though only the mean of the frequency distribution increased, while variance decreased. If harvesting increased the variance at Fool Creek, the changes to larger floods could have been more substantive (as we illustrated for WS3). It is possible that changes to the frequency distribution of peak flows at Fool Creek are predominantly caused by timber removal because, as mentioned earlier, all spur roads (72% of the entire road network) were decommissioned immediately after logging [*Alexander and Watkins*, 1977].

[49] The data sets of WS1 and WS3 have been analyzed by three different research groups without consensus on treatment effect for "larger" events [*Jones and Grant*, 1996; *Thomas and Megahan*, 1998; *Beschta et al.*, 2000; *Jones*, 2000; *Jones and Grant*, 2001; *Thomas and Megahan*, 2001]. *Jones and Grant* [1996] analyzed the same data sets using ANOVA and concluded the following: (1) "forest harvesting has increased peak discharges by as much as 50% in small 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 pre-existing 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). *Thomas and Megahan* [1998] reanalyzed the same data sets with extended ANCOVA and reacted: (1) "Jones and Grant (1996) conclusions were not supported by their study results" [*Thomas and Megahan*, 2001, p. 181], and (2) "Our re-analysis of small basin responses showed a rapid drop in relative peak flow increases for increasing peak flows with no statistically significant increases detectable above less than bankfull levels (2-year return interval) on either study watershed [*Thomas and Megahan*, 1998, Figures 3a and 3b]" [*Thomas and Megahan*, 2001, p. 182]. *Beschta et al.* [2000] also reanalyzed the same data sets with an extension of ANCOVA and concluded, "results also indicate that peak flow increases are not evident for events greater than a 5-yr return interval for watersheds WS1 and WS3" (p. 117).

[50] Our frequency-paired analysis presented in Figure 5b reveals that the entire sample of peak discharges was indeed shifted upward by clear-cutting and roads and that the biggest floods (events with approximate return intervals of 10–40 years) can even be more affected than the small and medium floods (1-year to 5-year events). *Verry et al.* [1983] used frequency analysis of a shorter paired watershed data set and revealed how rain-induced 10-year peak flows were affected more than the 2-year peak flows after the removal of aspen from a 23.2-ha watershed at the Marcell Experi-

mental Forest, Minnesota. *La Marche and Lettenmaier* [2001] also came to similar conclusions using a frequency analysis of model-simulated preharvest and postharvest (with roads) peak flows in a 149-km^2 dominantly transient snow subcatchment of the Deschutes River, Washington. For a 25-km^2 snow-dominated catchment in British Columbia, Canada, *Schnorbus and Alila* [2004] revealed through a longer record of simulated peak flows how preharvest and postharvest (with no roads) frequency curves diverge and how the presumed "no-effect" threshold flood is not reached even at the 50- and 100-year return periods. We emphasize, however, that the same contention cannot be supported by the ANOVA analysis of *Jones and Grant* [1996] for the same argument used by *Thomas and Meghan* [2001, p. 182], namely, "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." This clearly illustrates how the large preharvest events that were either unaffected or reduced by harvesting were used under chronological pairing to support the claim that logging does not affect the observed flood flows on record. Under frequency-pairing, those same large preharvest events, by competing for rank in the treated watershed with preharvest medium-sized events that were amplified by logging to a similar magnitude, now contribute to the support of our opposite assertion, namely, forest harvesting can affect the observed flood flows on record (even if we disagreed on their exact return periods).

[51] Our analysis of recovery-adjusted peak flows in WS3, which are representative of the period immediately following treatment, shows that logging and roads roughly change a 2-year event into a 1-year event, but could also be changing a 40-year event into a 15-year event (Figure 5b). This trend calls for further scrutiny since it could continue for larger events if more data were collected without altering the current state of roads. Roads cause more permanent physical changes to watersheds, and the effects are not expected to recover at the same rate as forest stands [*Gucinski and Furniss*, 2000]. Our flow duration curve analysis of the borderline stationary time series of recovery-unadjusted peak flows points to the possibility that "hydrologic recovery" at WS3 and Fool may have mitigated some of the effect on the magnitude of floods, but the effect on event frequencies may persist for decades after harvest. *Thomas and Megahan* [1998], for instance, predicted with extended ANCOVA that statistically, WS3 had recovered within the first 10 years following treatment. It appears as if "statistical insignificance," often used as a reason for dismissing the impact on floods, is not as scientifically or practically relevant in an evaluation that focuses only on a change in magnitude, as in ANOVA and extended ANCOVA. This underscores the importance of changes in frequency in the investigation of the relation between forest harvesting and floods. It also clearly illustrates how profound the implications of overlooking changes in frequency can be as small changes in flood magnitude can translate into large changes in their return periods and the larger the flood event the more substantial is the change in event return period (e.g., Figures 3d and 5b). Our study points to the urgent need for more research due to the possible profound implications on larger watersheds either as a result of conventional forest harvesting and deforestation practices or as a consequence of the current massive and unprecedented forest epidemics that have ravaged the landscape in North America from New Mexico to British Columbia [*Struck*, 2006; *Robbins*, 2008].

5.3. The Prevailing Scientific Perception Is Not Right: It Is Not Even Wrong

[52] The convergence of flood frequency models appears to have been confused with the convergence of two regressions in extended ANCOVA. The use of the latter has resulted in forest hydrologists always maintaining that forest harvesting does not affect large events [e.g., *Troendle and Stednick*, 1999; *Troendle et al.*, 2001; *Calder*, 2005, Figure 2.6; *MacDonald and Stednick*, 2003; *Troendle et al.*, 2006]. A common line of reasoning used in the literature in support of such a contention is provided by *Troendle et al.* [2006, p. 12], which incidentally refers to our Fool Creek: ". . .In the case of Fool Creek, peak flow increased an average of 20 percent (see Figure 4). However, the 3 largest peaks of the posttreatment period from 1967–1998 were not significantly increased." In such a line of reasoning, a "large" event is again defined based on the ranking of peak flows from a control watershed. This is analogous to a flood-control engineer notifying a community living in the floodplain of a treatment watershed to evacuate only as a result of a large event in a neighboring control watershed; however, a medium-sized event in the control could correspond to one of the largest events in the treatment, and unexpectedly flood the community because harvesting has changed the frequency of larger floods. We contend that the convergence of two conceptual regression lines imposed by chronological pairing is a nontestable hypothesis; and while it may be physically meaningful, it does not necessarily translate into the presupposed convergence of two flood frequency relations, and is therefore misleading. *Shermer* [2006] states, "In science, if an idea is not falsifiable, it is not that it is wrong, it is that we can not determine if it is wrong, and thus it is not even wrong." Our case against the flawed concept of chronological pairing and associated extended ANCOVA and ANOVA methods is a reminder of Pauli's proverb: "This isn't right. It's not even wrong."

[53] *Harr* [1986, p. 1096] raised serious concerns about interpreting the convergence of two regression fits to mean that forest harvesting does not affect larger floods, emphatically categorizing this type of interpretation as "irrelevant." Convergence nevertheless continues to be used in support of the contention that the often ambiguously defined larger floods are not affected, while forest harvesting may increase small- and medium-sized peak flows [e.g., *Thomas and Megahan*, 2001; *Troendle et al.*, 2001; *MacDonald and Stednick*, 2003; *Calder*, 2005; *Eisenbies et al.*, 2007; *Stednick*, 2008a, 2008b]. The incorrect interpretation of convergence has forced a "no-effect" return period threshold to be as small as a 2-year [*Thomas and Megahan*, 2001, p. 182; *MacDonald and Stednick*, 2003, p. 13], 5-year [*Beschta et al.*, 2000, p. 117], and 10-year [*Calder*, 2005, Figure 2.6] event. The convergence of two regression lines has emerged as a universal and seemingly irrefutable phenomenon since only a few studies have led to divergent fits [e.g., *Harr et al.*, 1979; *King*, 1989], some of which were questioned on various counts [e.g., *Harr et al.*, 1979; *Wright et al.*, 1990] or considered "suspect" [e.g., *Grant et al.*, 2008, p. 14]. The results of frequency analyses in our

study illustrate how such convergence is indeed irrelevant: Even when the postharvest events chronologically paired with the largest preharvest events in the control watershed are not increased, or are reduced, forest harvesting can still increase the frequency and magnitude of peak flows that have become some of the largest on record in the treatment watershed (contrast our Figures 3a with 3b, and Figures 3c with 3d for Fool Creek, and our Figures 5b and 6b with Figures 2 and 3b of *Thomas and Megahan* [1998], and Figure 6 of *Jones and Grant* [1996] for WS3).

[54] *Harr*'s [1986] diagnosis of a problem with the extended ANCOVA is correct but the author's remedy of applying a chronologically paired event analysis to only the largest events on record, which are dominated by a rain-on-snow regime, is inappropriate. In fact, a separate evaluation of any subsample of peak flows classified by size or generation mechanism outside the framework of the frequency distribution may be viewed as leading to the logical fallacy of composition [e.g., *Harr*, 1986; *Jones and Grant*, 1996; *Jones*, 2000]. The same could be said of any preconceived conceptual line of reasoning around an individual peak flow event [e.g., *Meyer*, 1928; *Zinke*, 1965; *Jeffrey*, 1970] or an actual analysis of individual chronologically paired events [e.g., *Swanson and Hillman*, 1977; *Storck et al.*, 1998; *Whitaker et al.*, 2002; *Ranzi et al.*, 2002] outside the framework of a frequency distribution. We agree with Harr's diagnosis, but would like to raise the ante: The inappropriate type of pairing of peak flows in ANOVA and extended ANCOVA studies may have led to a paradigm best described as being systematically misdirected by logical fallacies.

5.4. Chronological Pairing: A Paradigm for Low Power and Misleading Measures of Effect Variability

[55] Historically, a seemingly increased "variability" of peak flows with a "negligible" change in mean, detected under chronological pairing, has been interpreted to imply that the effect of logging is simply "unpredictable" (e.g., R. D. Harr (presented paper, 1979); also quoted by *Hewlett* [1982, p. 557]). Conceptually, however, a change (or lack thereof) in the mean of a frequency distribution is not the only proxy of changes in the magnitude and frequency of larger events [*Schaeffer et al.*, 2005]. This fact has been overlooked because we have always reasoned around chronological pairing and as such placed exclusive focus on changes in flood magnitude and not frequency.

[56] Studies from contrasting hydroclimate regimes in the United States and Japan had concluded, albeit using chronological pairing, that forest harvesting increased the variability of peak flows [e.g., *Hewlett and Helvey*, 1970; *Nakano*, 1971; *Hornbeck*, 1973a; *Hewlett*, 1982; R. D. Harr, presented paper, 1979]. The high variability of watershed response to harvesting under the conceptual framework of chronological pairing has puzzled forest hydrologists for decades [e.g., *Hewlett and Helvey*, 1970; *Swindel and Douglass*, 1984; *Troendle and King*, 1985; *Harr*, 1986; *Andréassian*, 2004; *Eisenbies et al.*, 2007]. *Eisenbies et al.* [2007, p. 81], for instance, report on *Hewlett and Helvey* [1970]:

> *Hewlett and Helvey* [1970] observed a 22% increase in stormflow volume at Coweeta during two separate events that approached the 7-day, 100-year return period for rainfall, seemingly to imply that forestry affects larger events; however, they did not draw this conclusion. They were less certain about the effects of forest clearing on peak flow, except that the variability of larger peakflows increased. The residuals from the two events represented the largest positive and negative deviations from the regression line.

[57] *Andréassian* [2004, p. 12] reports on *Troendle and King* [1985] and incidentally on forest harvesting effects at Fool Creek: "...while the hydrological impact of treatment remains constantly positive for the annual flow over the 30 years, it becomes negative in some years for the flood flow and especially for the flood peak (i.e. in these years, the effect of cutting the forest was to decrease the flood intensity!)." We emphasize that a decrease in magnitude under the chronological pairing framework while physically meaningful is not necessarily equivalent to a decrease in magnitude under the frequency pairing domain.

[58] In many studies, the variability of chronologically paired peak flow response to harvesting was viewed as a nuisance and was suppressed by logarithmic transformation of the data from the outset [e.g., *Thomas*, 1990; *Wright et al.*, 1990; *Jones and Grant*, 1996; *Thomas and Megahan*, 1998; *Beschta et al.*, 2000; *Jones*, 2000; *Lewis et al.*, 2001; *Guillemette et al.*, 2005]. This was often required to obtain a tighter posttreatment relation between the expected and observed peak flows in extended ANCOVA, and to satisfy the basic assumptions of normality and homogeneity of variance in both ANOVA and ANCOVA analyses.

[59] Our frequency-paired event analyses illustrate how the difference in preharvest and postharvest variability of peak flows at WS3 appears to be a critical, if not the most essential, aspect of the investigation and may be a keystone to any physical understanding and prediction of the effects of forest harvesting on larger floods. It is vital to point out, however, that chronological pairing must not be used to quantify a change in the variance of the frequency distribution of peak flows. A chronologically paired difference in peak flow magnitude between control and treatment watersheds will always be more variable after harvest, because the rank order of a peak flow event changes even if treatment increased only the mean of the frequency distribution. In this case, such an increase in variability is simply an artifact of the inappropriate type of pairing and is not equivalent to a forest-harvest-induced increase in variance of the frequency distribution of peak flows. For instance, although chronological pairing revealed that forest harvesting at Fool Creek caused a substantial increase in the variability of the difference in magnitude between control and treatment peak flow responses (Figure 3a), only the mean of the frequency distribution of peak flows had increased, while the variance had actually decreased (Table 1). In addition, Figure 3c of Fool Creek illustrates how the relation between return period and the relative change in peak flow shows substantially less scatter in the frequency-paired domain. The frequency-paired analysis exhibits a stronger relation than the chronologically paired analysis (Figure 3c), which makes the former statistically more powerful in detecting peak flow changes. This is further corroborated by the differences in the width of the two confidence bands and in the number of points lying within and outside these limits (Figure 3e). Since chronologically paired analysis has dominated research methods in forest hydrology, it may explain why hydrologic responses of forested watersheds to silvicultural practices have been, and continue to

be, characterized as "highly variable, and for the most part unpredictable" (*Hibbert* [1967, p. 535], echoed by *Eisenbies et al.* [2007]; R. D. Harr (presented paper, 1979), echoed by *Hewlett* [1982, p. 557]; *Leopold* [1972], echoed by *Jones* [2000]). *Thomas and Megahan* [1998] dismissed their relations between forest harvesting and peak flows in larger watersheds, and those of *Jones and Grant* [1996], in part because of a lack of statistical significance and predictive capability. Both studies, however, overlooked the fact that chronological pairing introduces an artificial level of effects variability which results in reducing statistical power. The time has come to end the unprecedented state of confusion in the science of forests and floods caused by the inappropriate type of event pairing.

5.5. Floods in a Changing Land Use: Variability Is More Important Than Averages

[60] Ecologists have recognized that a change in variance caused by an environmental disturbance cannot be detected by a BACI design [*Underwood*, 1991, 1994]. Nevertheless, they have realized the importance of flow variability to the integrity of stream aquatic ecosystems in science [*Poff et al.*, 1997] and management [*Landres et al.*, 1999]. The climate change community was also long ago introduced to a new statistical paradigm that helped shape the current perception of the relation between climate change and weather extremes, which is underpinned by the tenet "variability is more important than averages" [*Katz*, 1993; *Katz and Brown*, 1992]. In forest hydrology, on the other hand, little attention has been paid to possible changes in the variability of peak flows, and most important, no attention has been paid to its implications on the magnitude and frequency of larger floods. This lack of attention is evident in most experimental paired watershed and some numerical modeling studies [e.g., *Storck et al.*, 1998; *Whitaker et al.*, 2002; *Waichler et al.*, 2005; *Tonina et al.*, 2008] and is, at least in part, a consequence of our deep-rooted focus on regression rather than probability distribution models and on changes in magnitude rather than frequency. *Hewlett* [1982, p. 553], for instance, summarizes a landmark paired watershed study in Japan:

> Some excellent data was collected in Japan in an effort to relate stormflow volumes and peaks to clearcutting [*Nakano*, 1971]. . ..The net result seems to be a large increase in the variance of peaks and volumes. . .. These data would be worth reanalyzing by more advanced regression techniques, for they appear to have been collected carefully over many years.

[61] *Archer* [2007, p. 5] documented in novel ways how an increase in flow variability in the rain-dominated Plynlimon paired watershed experiment in mid-Wales is related to forest land use and to our knowledge, for the first time in forest hydrology literature, points out, albeit in qualitative ways, what most would consider in hindsight an obvious but critical linkage:

> Whilst flow variability (or flashiness) is related to the frequency and magnitude of flood flow, it is an important property in its own right with respect to influence on sediment transport. . .channel morphology. . .and river ecology.

It is clear that knowledge about how forest land use can affect the overall characteristics of flood frequency distributions (not just mean and variance) is critical to any assessment of the societal and environmental impacts of forestry and deforestation.

[62] Despite the drawbacks, paired watershed studies could be a good source of data for quantifying treatment effect on variance, but may only continue to be useful if observation periods are extended to collect more extremes with concurrent suppression of posttreatment vegetative growth. The latter option, to our knowledge, has rarely been considered in such experimental designs for either regulatory or environmental reasons. Potential threshold processes or nonlinearity in the response of a watershed [*Eisenbies et al.*, 2007] underscores the need for longer pretreatment observation period. This would allow for a better characterization of the nature of the relation between peak flow responses in control and treatment watersheds. A longer posttreatment observation period during which treatment has a sustained effect is also necessary [*Hornbeck*, 1973b]. If treatment increases variability, then the observed posttreatment cdf may shift upward at the upper tail. This makes it possible for the observed and expected cdfs to diverge or move farther apart at higher return periods, as appeared to be revealed by our analysis of WS3.

5.6. Convenience of Chronological Pairing Stifled the Progress of Science

[63] *Hewlett* [1982, p. 557] reviewed the state of science using small paired watershed studies of the day and despite disparate scale mismatch, widely extrapolated their outcomes to larger river basins. *Hewlett* [1982], nonetheless, had been for decades, and continues to be, a commonly cited manifesto for advocating worldwide policies that are based on a "no evidence" of a relation between forests and floods in large basins [e.g., *Hamilton*, 1985, 1987, 1988, 1992, 2008]. *Hamilton* [2008, p. 6], for instance, states,

> *Hewlett* [1982] reviewed the evidence from watershed research worldwide and reported that no cause and effect was demonstrated between forest cutting in the headwaters and floods in the lower basin. No conflicting information has been published since, more than 20 years later.

This is, however, a consequence of the preponderance of chronological pairing outcomes in the literature, which continues to create a sense of complacency and derail the progress of science on this topic in small and large basins, as evidenced by the following [*Thomas and Megahan*, 2001, p. 182]:

> . . . 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

and [*Beschta et al.*, 2000, p. 118]

> . . . it does not appear that the hypothesis of large increases in flood-size peak flows as a result of past and current forest land management practices should rank high on the list of future research questions.

[64] It is clear how chronological pairing has impaired our ability to extrapolate the outcomes of small paired watershed studies to the more relevant larger multiple land-use watersheds. This has led to commonly repeated dogmatic statements such as the following [*Calder et al.*, 2007, p. 945]:

> Now forest hydrologists generally agree that, although forests mitigate floods at the local scale and for small to medium-size flood events, there is no evidence of significant benefit at larger scales and for larger events.

Such a dominant perception, however, has been shaped by evaluation methods and conceptual lines of reasoning that focus solely on a change in magnitude with no consideration for any changes in flood frequency. Forest cover has no effect on the frequency of a flood only when a watershed starts responding as a parking lot in accordance with the rational method. For decades, Hewlett has challenged the underlying assumptions behind such ''engineering hydrologic orthodoxy'' [*Jackson et al.*, 2005, p. 2095] for forested watersheds, yet it appears that over 4 decades of ANOVA and extended ANCOVA applications in forest hydrology research are based on some of the same assumptions.

[65] *Hewlett and Helvey* [1970] once warned that downstream flooding might be caused by an increase not in peak flows, but in runoff volume due to logging headwaters; therefore the combination of increased stormflow volumes and increased amounts of sediment deposited in channels can increase the frequency with which streamflow exceeds channel capacity. Some 40 years later, an increase in lowland flood risk caused by possible geomorphological consequences of upland forest land use turned into what could be interpreted as ''just a semantic'' argument [e.g., *van Dijk et al.*, 2008, p. 3]. In addition, peak flow magnitude attenuation via desynchronization and hydraulic routing is more often the only mechanism invoked in relation to the cumulative downstream flooding effects of forest land-use change in headwaters [*Hewlett*, 1982; *Calder*, 2005; *MacDonald and Coe*, 2007; *Grant et al.*, 2008] (see also http://www.worldagroforestry.org/water/downloads/bca_bruijnzeel.pdf). As a direct consequence of reasoning from chronological pairing, however, mechanisms such as flow and sediment routing and the consequential downstream geomorphic changes, while physically relevant, do not invoke the critically linked changes in flood frequency and therefore could be misleading. In most hydroclimate regimes, the slopes of the cdfs of larger streams are milder than those of their headwater tributaries [*Archer*, 1989; *Blöschl and Sivapalan*, 1997]. This may cause a small increase in the magnitude of the former to translate into a large decrease (increase) in the return period (frequency) of floods. Larger watersheds with mildly sloped peak flow frequency curves may be susceptible to even larger changes in return period as the magnitude of peak flow increases. The relevance of such hypotheses can only be brought to light and tested using the right analytical framework: a frequency-paired approach.

[66] Chronological pairing has also formed, and continues to form, the basis of evaluating the effects of forest harvesting on physical, chemical, and biological watershed response variables [e.g., *Beschta*, 1978; *Hicks et al.*, 1991; *Beschta et al.*, 2000; *Martin et al.*, 2000; *Swank et al.*, 2001; *Dhakal and Sidle*, 2004, 2008]. It is also frequently used to investigate ''hydrologic recovery,'' i.e., the return of a watershed to pretreatment conditions [e.g., *Thomas*, 1990; *Thomas and Megahan*, 1998]. This inappropriate type of pairing may therefore have clouded our view of the more general relation between forest land use and the biophysical environment. *Dhakal and Sidle* [2008], *Johnson et al.* [2007], and *Johnson and Edwards* [2008], for instance, disagreed on how forest harvesting affects landslides but appear to have overlooked the most inconspicuous vice of pairing event by storm input. Their arguments and counterarguments on the effect of forest harvesting on landslides are still based on methods that do not reveal changes in event frequency and, equally important, do not guarantee correct changes in event magnitude. This recent exchange between *Dhakal and Sidle* [2008], *Johnson et al.* [2007], and *Johnson and Edwards* [2008] on forest harvesting and landslide relations is reminiscent of many other, often intense, exchanges on forest harvesting and peak flows (*Jones and Grant* [1996] versus *Thomas and Megahan* [1998]; *Beschta et al.* [2000]; *Harris* [1977] versus *Beschta* [1978]; *Harr* [1986] versus *Hess* [1984]; *Rothacher* [1973]; *Harr and McCorison* [1979]; *Harr et al.* [1975, 1979] versus *Wright et al.* [1990]). It is indeed ''nice to know why and how highly competent scientists can repeatedly come up with completely contradictory results [and interpretations]'' [*Quesenberry et al.*, 2005, p. 1121]. Although in a completely different field, we found a striking analogy between the essence of our article and that of *Quesenberry et al.* [2005]. Relentless disagreements between scientists have motivated their work in the field of stem cell research. As it turned out, a certain aspect of stem cell research was based on the logical fallacy of an irrelevant conclusion. The persistent disagreement between forest hydrologists, centered around chronologically paired event analysis methods, is a compelling testimony of how forest hydrology has been in a state of growing crisis and is now ripe for a change in paradigm (as defined by *Kuhn* [1970]). What has been turned into an enigmatic topic (as labeled by *Eisenbies et al.* [2007]) must now be addressed with appropriate scientific methods.

6. Summary and Conclusion

[67] A resolution of how to quantify the effects of rapidly reversed deforestation on floods, which are of the kind represented by the small paired watersheds of this study, may not be viewed by some as important, for instance, to the more challenging question of the effects of massive quasi-permanent deforestation on floods in large tropical mountain basins. The outcomes of these small paired watershed studies, however, have been and continue to be used to advocate policies related to land-use management over larger basins in hydroclimate regimes worldwide, including the tropics (e.g., read *Lull and Reinhart* [1972], *Hewlett* [1982], *Hamilton and King* [1983], *Hamilton* [1985, 1990], *Bruijnzeel* [1990], *Eaton and Church* [2001], *Food and Agriculture Organization of the United Nations* (*FAO*) [2003], *Kaimowitz* [2004], *MacDonald and Stednick* [2003], *FAO-CIFOR* [2005], *Calder* [2005], *Troendle et al.* [2006], *Calder et al.* [2007], *Grant et al.* [2008], *Eisenbies et al.* [2007], *Kochenderfer et al.* [2007], *van Dijk et al.* [2008], and *Hamilton* [2008] (see also http://www.worldagroforestry.org/water/downloads/bca_bruijnzeel.pdf)). It is important therefore to expose the caveats of small paired watershed studies, if only to put an end to (1) using their potentially erroneous outcomes to influence land-use policy, and (2) their outcomes stifling the progress of science.

[68] Our study does not directly examine the effects of forests on the more ''catastrophic'' floods because of the challenges involved in inferring a 100-year flood from a

25-year record. Nevertheless, the long-held belief that forests or forest harvesting do not affect larger floods may not be reconcilable with the shift in mean and potential increase in variability already acknowledged in prior paired watershed studies, but never linked to the magnitude and frequency of larger floods. Perhaps it is not by coincidence that a correlation between forests, forest harvesting, deforestation, and floods was found in the few extraordinary studies that had not applied a chronologically paired event analysis [*Anderson and Hobba*, 1959; *Swank and Vose*, 1994; *Schnorbus and Alila*, 2004; *La Marche and Lettenmaier*, 2001; *Bradshaw et al.*, 2007, 2009; *Lin and Wei*, 2008].

[69] The repeated application of ANOVA and extended ANCOVA, which we show here to be inappropriate, has led researchers to continue to reaffirm erroneous conclusions of previous work. ANOVA and extended ANCOVA may have been accepted in past studies because their findings also reinforce theoretical preconceptions, some even predating our experimentation with paired watershed studies [e.g., *Meyer*, 1928; *Zinke*, 1965; *Jeffrey*, 1970; *Swindel and Douglass*, 1984, Figures 1 and 2; *MacDonald and Stednick*, 2003, p. 13; *Calder*, 2005, Figure 2.6] that reason exclusively around a change in flood magnitude (albeit the incorrect one), while overlooking the potentially serious consequences of small changes in magnitude on flood frequency. *DeWalle* [2003, p. 1255] states

> We have long believed that forest cover by itself only can play a limited role in controlling peak flows due to extreme events. That is, an extreme rain event, spawned perhaps by a hurricane, would produce the same peak flows with or without forest cover, assuming all other conditions, especially soil conditions, were maintained. Thus, it came as no surprise that forest management experiments generally showed that harvesting of trees with minimum soil disturbance did not appear to lead to major increases in annual peak flows or floods, although specific attention was often not given to extreme events.

[70] The classical argument invoked to support the claim that there is "no evidence" that forests affect larger flood events, repeatedly echoed in elite journals [e.g., *Calder et al.*, 2007; *van Dijk et al.*, 2008] and influential United Nations policy documents [e.g., *FAO-CIFOR*, 2005; *Hamilton*, 2008] is that during larger floods, the processes of interception and evaporation by vegetative cover are overwhelmed by precipitation and therefore the effect of forests "would be expected to be most significant for small storms and least significant for the largest storms" [*Calder*, 2005, p. 47]. This assertion (which appears to use the unqualified and nonstatistical term "significant") is also based on chronological pairing and indexes flood size to storm input, potentially concealing the effects of forest harvesting and deforestation on the frequency, magnitude, duration, and volume of floods.

[71] *Hewlett* [1982, p. 546] states,

> [H]ydrologists have understandably been confused by the difficulties inherent in describing the nature and frequency of floods to laymen, who are apt to have little patience with probability statements. . ..But among ourselves we must drop back to rigorous language in order to discuss and trade information about land-use causes and flood effects.

Ironically, decades of published paired watershed studies rarely invoked changes in flood frequency. Could genuine efforts to reach out to a skeptical public have been based on an irrelevant conclusion, and led to reaffirming the dogma that forests do not affect large floods? The convergence or divergence of two peak flow frequency distributions, and not two regression models, is important, but even more important is how changes in magnitude translate into changes in frequency. Frequency is a more relevant surrogate of flood risk, especially if small changes in magnitude (within measurement or modeling accuracies) can lead to large changes in return period. Let there be no confusion that unless we reinstate the dimensions of frequency and variability in the science of forest hydrology and address them head-on, we fail to fully explore the true relation between forests, floods, and the biophysical environment.

[72] The reported changes in flood frequency and magnitude for our study watersheds should only be extended to other basins with care, as it is anticipated that they are strongly dependent upon sample size, basin physiography, climate, and forestry practices. It is therefore necessary to determine the extent to which our reported changes in flood magnitude and frequency can be generalized to other basins, by conducting similar studies in different hydroclimate regimes. Our ritualistic adherence to the moribund concept of null-hypothesis significance testing on the one side, and the frequent use of the layman term "significance" on the other, added fuel and more ambiguity to the so-called "debate" on forests, deforestation, forest harvesting, and floods. The multiple-working hypotheses approach to statistical inference based on information-theoretic and Bayesian methods, increasingly used in science literature, offers compelling advantages to the study of forest and flood relations [*Burnham and Anderson*, 2002; *Stephens et al.*, 2005; *Lukacs et al.*, 2007; *Bradshaw et al.*, 2007; *Elliott and Brook*, 2007]. Long-term and expensive paired watershed studies have been, and continue to be, defended as the approach to research methods in forest hydrology [*Hewlett et al.*, 1969; *Stednick*, 2008c]; however, these are plagued with caveats [*Grant et al.*, 2008; *Bren*, 2008] and not ideal for investigating floods, even if one were able to suppress tree growth effectively for the purpose of collecting longer-term data of treatment effects. Paired watershed experiments are essentially black-box studies [*Bruijnzeel*, 2004]. Our reductionist and deterministic small-scale process-based studies carried out in research watersheds still suffer from a lack of connection to the watershed-scale response [*Dunne*, 1998; *Sivapalan*, 2003]. Ironically, the results of such studies continue to be used to defend the erroneous outcomes of paired watershed experiments. This may explain in part why our extrapolations to the larger and operationally more relevant spatial and temporal scales are often left to arm waving [*Benda and Miller*, 2001]. *Kimmins et al.* [2005] eloquently articulated why the reductionist small-scale "jigsaw puzzle" science in forestry (hydrology in our case), which in their own words leads to "jigsaw" puzzle policy, sometimes disappoint or even fail us. Chronological pairing has contributed to an entire community being misdirected by the irrelevant hydrological research hypotheses, flawed statistical methods, and their misleading outcomes for over 50 years.

[73] The public still believes unquestioningly that forests reduce flood risk [*Nisbet*, 2002; *Mortimer and Visser*, 2004; *FAO-CIFOR*, 2005]. In contrast, a long-standing "policy-,

and socioeconomically-oriented generation of forest hydrologic 'myth busters'" (as referred to by *Bruijnzeel* [2005] and *Hamilton* [1990]) continues to dispel such a perception as a "myth" concocted by popular press, environmentalists, and conservation agencies attempt to hoodwink the public (as claimed by *Hamilton* [1985], *Kaimowitz* [2004], *FAO-CIFOR* [2005], and *Calder and Aylward* [2006]). Some recently raised questions are now more legitimate than ever: Is there more truth to the public perception than has been acknowledged by the scientific community [*Bruijnzeel*, 2005, p. 14]? Is there more to the forestry and floods issue than a "red herring" raised by antidevelopment factions [*DeWalle*, 2003, p. 1256]?

[74] While it remains possible that forests, deforestation, or forest harvesting have no effects beyond some case-specific threshold-flood response, such a hypothesis cannot be substantiated by ANOVA and ANCOVA, or supported by a preconceived "theory" that is also based on the inappropriate type of event pairing. While large floods may not appear to increase much in magnitude, they may occur more frequently as a result of forest harvesting or deforestation. Any method or theoretical line of reasoning based solely on evaluating changes in flood magnitude, without invoking changes in frequency, is likely to conceal the true relation between forests and flood risk. In view of this mounting evidence against our entrenched and potentially inappropriate forest and flood paradigm, supported by a set of conceptual and empirical arguments long recognized in the wider hydrology, ecology, and climate change literatures, we urge the forest hydrology community to reexamine their conclusions from previous studies. Sample size and the confounding effects of hydrologic recovery may continue to pose challenges in evaluating the effects of forest harvesting and deforestation on floods using paired watershed studies. Any effort to address such challenges, however, must be based on the correct methods, particularly when using hydrologic numerical models in which the same watershed is used as its own control, and the forested and harvested land use scenarios are driven by the same long-term climate input [e.g., *Schnorbus and Alila*, 2004; *Cuo et al.*, 2009].

[75] Chronological pairing has dominated research methods in forest hydrology, and work is also urgently needed to examine whether this has distorted our perspective on the effects of forests on the general biophysical environment. If the past is any indication, the future of science on this topic may be rife with challenges, in part because flood frequency analysis also has a persistent history of being misused in the name of "scientific rigor" [*Klemeš*, 1974, 1986, 2000]. Although our study used well-established ranking methods that are not constrained by the same assumptions implied in fitting a frequency distribution model, further scrutiny of our claims using ours or other data sets may prove that we were also not immune to being misdirected by sampling variability. Sampling uncertainties notwithstanding, however, two wrongs do not make a right. Our dominant scientific perception of the forests and floods relation remains largely indefensible. We maintain that only a frequency-paired approach may be used to evaluate the relation between forests and floods. While almost certainly at the expense of new challenges, the frequency-paired approach has the potential for illuminating many previous confusions, misconceptions, uncertainties, and lack of consensus on the science of forests and floods. Our literature is full of arguments and counterarguments on both sides of this debate that have served to only distract us further from the real question at hand. Moving forest hydrology forward starts with unbinding oneself from erroneous precedents, not the least of which is advocating land-use policy in the name of the indefensible science.

[76] **Acknowledgments.** We thank four referees for their especially detailed and helpful comments on earlier drafts. We acknowledge feedback from and discussions with all colleagues, students, and friends with whom we shared our ideas and manuscript before and during submission, but this certainly does not mean that any of them will necessarily agree with our writing style, inferences, interpretations, and conjectures. We thank Chris Folland, Paul Pilon, Taha Ouarda, Ali Fares, Sarah L. Lewis, Peter Mikes, Kabir Rasouli, Kim Green, Drew Brayshaw, Glynnis Horel, Pat Teti, Rita Winkler, Steve Chatwin, Allan Chapman, John Innes, Bruce Larson, Cindy Prescott, Paul Wood, Gary Bull, Dave Tait, Val LeMay, Dave Spittlehouse, and John Stednick for discussions. We thank, Jeff McDonnell, Jos Beckers, Jaime Cathcart, Scott Weston, Dan Bewley, Philip Marren, Tim Blair, Jerry Maedel, and Jannette Bij de Vaate for comments. We thank Dan Moore for his assistance and feedback on estimating the uncertainty in the predicted flood frequency curve. We thank Julia Jones and Gordon Grant for generously agreeing to provide us a copy of their data. H. J. Andrews's data originated from Forest Science Data Bank (http://www.fsl.orst.edu/lter), a partnership between the Department of Forest Science, Oregon State University, and the U.S. Forest Service Pacific Northwest Research Station, Corvallis, Oregon. Fool Creek data originated from the Rocky Mountain Research Station Web site (http://www.fs.fed.us/rm/). We acknowledge U.S. agencies and staff for long-term monitoring and for making data available to the wider science community. Our research was supported by the National Science and Engineering Research Council of Canada, Forest Science Program of British Columbia, British Columbia Ministry of Forests and Range, British Columbia Ministry of the Environment, Forest Practices Board of British Columbia, and Natural Resources Canada (Mountain Pine Beetle Initiative).

References

Adams, P. W., A. L. Flint, and R. L. Fredriksen (1991), Long-term patterns in soil moisture and revegetation after clearcut of a Douglas-fir forest in Oregon, *For. Ecol. Manage.*, *41*, 249–263, doi:10.1016/0378-1127(91)90107-7.

Alexander, R. R., and R. K. Watkins (1977), The Fraser Experimental Forest, Colorado, *Gen. Tech. Rep., 40*, Rocky Mt. For. and Range Exp. Stn., For. Serv., U.S. Dep. of Agric., Fort Collins., Colo.

Alila, Y., and A. Mtiraoui (2002), Implications of heterogeneous flood frequency distributions on traditional stream discharge prediction techniques, *Hydrol. Processes*, *16*, 1065–1084, doi:10.1002/hyp.346.

Allen, M. R., and W. J. Ingram (2002), Constraint on future changes in climate and the hydrologic cycle, *Nature*, *419*, 224–232, doi:10.1038/nature01092.

Anderson, D. R., K. P. Burnham, and W. L. Thompson (2000), Null hypothesis testing: Problems, prevalence and an alternative, *J. Wildl. Manage.*, *64*, 912–923, doi:10.2307/3803199.

Anderson, H. W., and R. L. Hobba (1959), Forest and floods in the northwestern United States., *Publ. 48*, pp. 30–39, Int. Assoc. of Hydrol. Sci., Gentbrugge, Belgium.

Andréassian, V. (2004), Water and forests: From historical controversy to scientific debate, *J. Hydrol.*, *291*, 1–27, doi:10.1016/j.jhydrol.2003.12.015.

Archer, D. R. (1989), Flood wave attenuation due to channel and floodplain storage and effects on flood frequency, in *Floods: Hydrological, Sedimentological and Geomorphological Implications*, pp. 37–46, John Wiley, New York.

Archer, D. R. (2007), The use of flow variability analysis to assess the impact of land use change on the paired Plynlimon catchments, mid-Wales, *J. Hydrol.*, *347*, 487–496, doi:10.1016/jhydrol.2007.09.036.

Benda, L., and D. Miller (2001), Beyond arm waving: Thinking critically at large scales, *Watershed Manage. Counc. Networker*, *10*, 1–16.

Beschta, R. L. (1978), Long-term patterns of sediment production following road construction and logging in the Oregon Coast Range, *Water Resour. Res.*, *14*, 1011–1016, doi:10.1029/WR014i006p01011.

Beschta, R. L., M. R. Pyles, A. E. Skaugset, and C. G. Surfleet (2000), Peak flow responses to forest practices in the western cascades of Oregon, USA, *J. Hydrol.*, *233*, 102–120, doi:10.1016/S0022-1694(00)00231-6.

Best, A., L. Zhang, T. McMahon, A. Western, and R. A. Vertessy (2003), A critical review of paired catchment studies with reference to seasonal flows and climate variability, *Tech. Rep., 25/03*, Commonw. Sci. and Ind. Res. Organ. (CSIRO) Land and Water, Canberra, Australia.

Blöschl, G., and M. Sivapalan (1997), Process controls on regional flood frequency: Coefficient of variation and basin scale, *Water Resour. Res.*, *33*, 2967–2980, doi:10.1029/97WR00568.

Bonsal, B. R., X. Zhang, L. A. Vincent, and W. D. Hogg (2001), Characteristics of daily and extreme temperature over Canada, *J. Clim.*, *14*, 1959–1976, doi:10.1175/1520-0442(2001)014<1959:CODAET>2.0.CO;2.

Booth, D. (1990), Stream channel incision following drainage basin urbanization, *Water Resour. Bull.*, *26*, 407–417.

Bradshaw, C. J. A., N. S. Sodhi, K. S.-H. Peh, and B. W. Brook (2007), Global evidence that deforestation amplifies flood risk and severity in the developing world, *Global Change Biol.*, *13*, 2379–2395, doi:10.1111/j.1365-2486.2007.01446.x.

Bradshaw, C. J. A., B. W. Brook, K. S.-H. Peh, and N. S. Sodhi (2009), Flooding policy makers with evidence to save forests, *Ambio*, *38*, 125–126, doi:10.1579/0044-7447-38.2.125.

Brath, A., and A. Montanari (2000), The effects of the spatial variability of soil infiltration capacity in distributed flood modelling, *Hydrol. Processes*, *14*, 2779–2794, doi:10.1002/1099-1085(20001030)14:15<2779::AID-HYP91>3.0.CO;2-X.

Brath, A., A. Montanari, and G. Moretti (2006), Assessing the effect of flood frequency of land use change via hydrological simulation (with uncertainty), *J. Hydrol.*, *324*, 141–153, doi:10.1016/j.jhydrol.2005.10.001.

Bren, L. J. (2008), Paired watershed projects in Australia: Long-term studies and short-term politics, *Eos Trans. AGU*, *89*(53), Fall Meet. Suppl., Abstract H12A-06.

Brooks, K. N., P. F. Ffolliott, H. M. Gregersen, and L. F. DeBano (2003), *Hydrology and the Management of Watersheds*, Blackwell, Ames, Iowa.

Bruijnzeel, L. A. (1990), *Hydrology of Moist Tropical Forests and Effects of Conversion: A State of Knowledge Review*, Div. of Water Sci., U.N. Educ., Sci., and Cult. Organ., Paris.

Bruijnzeel, L. A. (2004), Hydrological functions of tropical forests: Not seeing the soil for the trees?, *Agric. Ecosyst. Environ.*, *104*, 185–228, doi:10.1016/j.agee.2004.01.015.

Bruijnzeel, L. A. (2005), Land use and land cover effects on runoff processes: Forest harvesting and road construction, in *Encyclopedia of Hydrological Sciences*, edited by M. G. Anderson, chap. 119, pp. 1813–1830, John Wiley, Hoboken, N. J.

Burnham, K. P., and D. R. Anderson (2002), *Model Selection and Multimodal Inference: A Practical Information-Theoretic Approach*, 2nd ed., Springer, New York.

Burt, T. P., and W. T. Swank (1992), Flow frequency responses to hardwood-to-grass conversion and subsequent succession, *Hydrol. Processes*, *6*, 179–188, doi:10.1002/hyp.3360060206.

Caissie, D., S. Jolicoeur, M. Bouchard, and E. Poncet (2002), Comparison of streamflow between pre and post timber harvesting in Catamaran Brook (Canada), *J. Hydrol.*, *258*, 232–248, doi:10.1016/S0022-1694(01)00576-5.

Calder, I. R. (2002), Forest and hydrological services: Reconciling public and science perceptions, *Land Use Water Resour. Res.*, *2*, 1–12.

Calder, I. R. (2004), Forests and water—Closing the gap between public and science perceptions, *Water Sci. Technol.*, *49*, 39–53.

Calder, I. R. (2005), *Blue Revolution: Integrated Land and Water Resource Management*, Earthscan, London.

Calder, I. R., and B. Aylward (2006), Forest and floods: Moving to an evidence-based approach to watershed and integrated flood management, *Int. Water Resour. Assoc.*, *31*, 1–13.

Calder, I. R., et al. (2004), Forest and water policies: The need to reconcile public and science perceptions, *Geol. Acta*, *2*, 157–166.

Calder, I. R., J. Smyle, and B. Aylward (2007), Debate over flood-proofing effects of planting forests, *Nature*, *450*, 945, doi:10.1038/450945b.

Chang, M. (2006), Forest and streamflow, in *Forest Hydrology: An Introduction to Water and Forests*, chap. 10, pp. 191–236, CRC Press, Boca Raton, Fla.

Cheng, J. D. (1989), Streamflow changes after clear-cut logging of a pine beetle-infested watershed in southern British Columbia, Canada, *Water Resour. Res.*, *25*, 449–456, doi:10.1029/WR025i003p00449.

Cohen, J. (1994), The earth is round ($p < .05$), *Am. Psychol.*, *45*, 1304–1312, doi:10.1037/0003-066X.45.12.1304.

Cohen, J., and I. Stewart (1994), *The Collapse of Chaos: Discovering Simplicity in a Complex World*, Penguin, New York.

Cuo, L., D. P. Lettenmaier, M. Alberti, and J. E. Richey (2009), Effects of a century of land cover and climate change on the hydrology of the Puget Sound basin, *Hydrol. Processes*, *23*, 907–933, doi:10.1002/hyp.7228.

DeFries, R., and K. N. Eshleman (2004), Land-use change and hydrologic processes: A major focus for the future, *Hydrol. Processes*, *18*, 2183–2186, doi:10.1002/hyp.5584.

DeWalle, D. R. (2003), Forest hydrology revisited, *Hydrol. Processes*, *17*, 1255–1256, doi:10.1002/hyp.5115.

Dhakal, A. S., and R. Sidle (2004), Pore water pressure assessment in a forested watershed: Simulations and distributed field measurements related to forest practices, *Water Resour. Res.*, *40*, W02405, doi:10.1029/2003WR002017.

Dhakal, A. S., and R. Sidle (2008), Discussion on “Ground-water response to forest harvesting: Implications for hillslope stability” by Johnson, A. C., Edwards, R. T. and Erhardt, R, *J. Am. Water Resour. Assoc.*, *44*, 1–7, doi:10.1111/j.1752-1688.2008.00168.x.

Draper, N. R., and H. Smith (1998), *Applied Regression Analysis*, John Wiley, New York.

Dunne, T. (1998), Critical data requirements for prediction of erosion and sedimentation in mountain drainage basins, *J. Am. Water Resour. Assoc.*, *34*, 795–808, doi:10.1111/j.1752-1688.1998.tb01516.x.

Dyrness, C. T. (1969), Hydrologic properties of soils on three small watersheds in the western Cascades of Oregon, *Res. Notes PNW-111*, For. Serv., U.S. Dep. of Agric., Corvallis, Oreg.

Eaton, B., and M. Church (2001), Hydrological effects of forest harvest in the Pacific Northwest, *Riparian Decis. Tool Tech. Rep. 3*, 56 pp., Dep. of Geogr., Univ. of B. C., Vancouver, B. C., Canada.

Eisenbies, M. H., W. M. Aust, J. A. Burger, and M. B. Adams (2007), Forest operations, extreme events, and considerations for hydrologic modeling in the Appalachians—A review, *For. Ecol. Manage.*, *242*, 77–98, doi:10.1016/j.foreco.2007.01.051.

El Adlouni, S., T. B. M. J. Ouarda, X. Zhang, R. Roy, and B. Bobée (2007), Generalized maximum likelihood estimators for the nonstationary generalized extreme value model, *Water Resour. Res.*, *43*, W03410, doi:10.1029/2005WR004545.

Elder, K., L. Porth, and C. A. Troendle (2006), The effect of timber harvest on the Fool Creek watershed after five decades, *Eos Trans. AGU*, *87*(52), Fall Meet. Suppl., Abstract B21F-01.

Elliott, L. P., and B. W. Brook (2007), Revisiting Chamberlain: Multiple working hypotheses for the 21st Century, *BioScience*, *57*, 608–614, doi:10.1641/B570708.

Folland, C., and C. Anderson (2002), Estimating changing extremes using empirical ranking methods, *J. Clim.*, *15*, 2954–2960, doi:10.1175/1520-0442(2002)015<2954:ECEUER>2.0.CO;2.

Food and Agriculture Organization of the United Nations (FAO) (2003), *The State of the World's Forests*, Rome.

Food and Agriculture Organization of the United Nations/Center for International Forestry Research (FAO/CIFOR) (2005), *Forests and Floods: Drowning in Fiction or Thriving on Facts?*, Bangkok.

Forsyth, T. (2005), Land use impacts on water resources—Science, social and political factors, in *Encyclopedia of Hydrological Sciences*, edited by M. G. Anderson, chap. 187, pp. 2911–2924, John Wiley, Hoboken, N. J.

Gilmour, D. A., M. Bonell, and D. S. Cassells (1987), The effects of forestation on soil hydraulic properties in the Middle Hills of Nepal: A preliminary assessment, *Mt. Res. Dev.*, *7*, 239–249, doi:10.2307/3673199.

Goodell, B. C. (1958), A preliminary report on the first year's effects of timber harvesting on water yield from a Colorado watershed, *Pap. 36*, Rocky Mt. For. and Range Exp. Stn., For. Serv., U.S. Dep. of Agric., Fort Collins., Colo.

Grant, G. E., S. L. Lewis, F. J. Swanson, J. H. Cissel, and J. J. McDonnell (2008), Effects of forest practices on peak flows and consequent channel response: A state-of-science report for western Oregon and Washington, *Gen. Tech. Rep. PNW-GTR-760*, Pac. Northwest Res., Stn., U.S. Dep. of Agric., Portland, Oreg.

Gucinski, H., and M. Furniss (2000), Forest roads: A synthesis of scientific information, For. Serv., U.S. Dep. of Agric., Washington, D. C. (Available at www.fs.fed.us/news/roads/science.pdf)

Guillemette, F., A. P. Plamondon, M. Prevost, and D. Levesque (2005), Rainfall generated stormflow response to clearcutting a boreal forest: Peak flow comparison with 50 worldwide basin studies, *J. Hydrol.*, *302*, 137–153, doi:10.1016/j.jhydrol.2004.06.043.

Guthery, F. S., J. J. Lusk, and M. J. Peterson (2001), The fall of the null hypothesis: Liabilities and opportunities, *J. Wildl. Manage.*, *65*, 379–384, doi:10.2307/3803089.

Halpern, C. B. (1989), Early successional patterns of forest species: Interactions of life history traits and disturbance, *Ecology*, *70*, 704–720, doi:10.2307/1940221.

Hamilton, L. S. (1985), Overcoming myths about soil and water impacts of tropical forest land uses, in *Soil Erosion and Conservation*, edited by S. A. El-Swaify, W. C. Moldenhauer, and A. Lo, pp. 680–690, Soil and Water Conserv. Soc., Ankeny, Iowa.

Hamilton, L. S. (1987), What are the impacts of Himalayan deforestation on the Ganges-Brahmaputra lowlands and delta? Assumptions and facts, *Mt. Res. Dev.*, *7*, 256–263, doi:10.2307/3673202.

Hamilton, L. S. (1988), The recent Bangladesh flood disaster was not caused by deforestation alone, *Environ. Conserv.*, *15*, 369–370.

Hamilton, L. S. (1990), Foreword to "Hydrology of moist tropical forests and effects of conversion: A state of knowledge review" by L. A. Bruijnzeel, Div. of Water Sci., U.N. Educ., Sci., and Cult. Organ., Paris.

Hamilton, L. S. (1992), The protective role of mountain forests, *GeoJournal*, *27*, 13–22, doi:10.1007/BF00150632.

Hamilton, L. S. (2008), Forests and water: A thematic study prepared in the framework of the Global Forest Resources Assessment 2005, *FAO For. Pap., 155*, U.N. Food and Agric. Organ., Rome.

Hamilton, L. S., and P. N. King (1983), *Tropical Forested Watersheds: Hydrologic and Soils Response to Major Uses or Conversions*, 168 pp., Westview, Boulder, Colo.

Harr, R. D. (1977), Water flux in soil and subsoil in a steep forested slope, *J. Hydrol.*, *33*, 37–58, doi:10.1016/0022-1694(77)90097-X.

Harr, R. D. (1986), Effects of clear-cutting on rain-on-snow runoff in western Oregon: A new look at old studies, *Water Resour. Res.*, *22*, 383–392, doi:10.1029/WR022i007p01095.

Harr, R. D., and F. M. McCorison (1979), Initial effects of clear-cut logging on size and timing of peak flows in a small watershed in western Oregon, *Water Resour. Res.*, *15*, 90–94, doi:10.1029/WR015i001p00090.

Harr, R. D., W. C. Harper, J. T. Krygier, and F. S. Hsieh (1975), Changes in storm hydrographs after road building and clear-cutting in the Oregon Coast Range, *Water Resour. Res.*, *11*, 436–444, doi:10.1029/WR011i003p00436.

Harr, R. D., R. L. Fredriksen, and J. Rothacher (1979), Changes in streamflow following timber harvest in southwestern Oregon, *Res. Pap. PNW-249*, Pac. Northwest For. and Range Exp. Stn., For. Serv., U.S. Dep. of Agric., Portland, Oreg.

Harr, R. D., A. Levno, and R. Mersereau (1982), Streamflow changes after logging 130-year-old Douglas fir in two small watersheds, *Water Resour. Res.*, *18*, 637–644, doi:10.1029/WR018i003p00637.

Harris, D. D. (1977), Hydrologic changes after logging in two small Oregon coastal watersheds, *U.S. Geol. Surv. Water Supply Pap., 2037*, 31 pp.

Hess, S. (1984), Timber harvesting and flooding, *J. Soil Water Conserv.*, *39*, 115–117.

Hewlett, J. D. (1982), Forests and floods in the light of recent investigations, in *Proceedings of the Canadian Hydrology Symposium*, pp. 543–560, Nat. Res. Counc. of Can., Ottawa, Ont., Canada.

Hewlett, J. D., and J. M. Bosch (1984), The dependence of storm flows on rainfall intensity and vegetal cover in South Africa, *J. Hydrol.*, *75*, 365–381, doi:10.1016/0022-1694(84)90060-X.

Hewlett, J. D., and J. D. Helvey (1970), Effects of forest clear-felling on the storm hydrograph, *Water Resour. Res.*, *6*, 768–782, doi:10.1029/WR006i003p00768.

Hewlett, J. D., H. W. Lull, and K. G. Reinhart (1969), In defence of experimental watersheds, *Water Resour. Res.*, *5*, 306–316, doi:10.1029/WR005i001p00306.

Hibbert, A. D. (1967), Forest treatment effects on water yield, in *International Symposium on Forest Hydrology*, edited by W. E. Sopper and H. W. Lull, pp. 527–543, Pergamon, Oxford, U. K.

Hicks, B. J., R. L. Beschta, and R. D. Harr (1991), Long-term changes in streamflow following logging in western Oregon and associated fisheries implications, *Water Resour. Bull.*, *27*, 217–226.

Hirt, P. W. (1994), *A Conspiracy of Optimism: Management of the National Forests Since World War Two*, Univ. of Nebr. Press, Lincoln.

Hollis, G. E. (1975), The effects of urbanization on floods of different recurrence interval, *Water Resour. Res.*, *11*, 431–435, doi:10.1029/WR011i003p00431.

Hornbeck, J. W. (1973a), Storm flow from hardwood-forested and cleared watersheds in New Hampshire, *Water Resour. Res.*, *9*, 346–354, doi:10.1029/WR009i002p00346.

Hornbeck, J. W. (1973b), The problem of extreme events in paired watershed studies, *Res. Note NE-175*, For. Serv., U.S. Dep. of Agric., Upper Darby, Pa.

Ives, J. D. (2006), Comment on "Forests and floods: Drowning in fiction or thriving on facts?", *Mt. Res. Dev.*, *26*, 187–188, doi:10.1659/0276-4741(2006)26[187:FAFDIF]2.0.CO;2.

Jackson, C. R., W. T. Swank, and R. Olszewski (2005), Hydrologist: John D. Hewlett (1922–2004), *Hydrol. Processes*, *19*, 2093–2095, doi:10.1002/hyp.5885.

Jeffrey, W. W. (1970), Hydrology of land use, in *Handbook on the Principles of Hydrology*, edited by D. M. Gray, pp. 13.1–13.57, Water Inf. Cent., Port Washington, N. Y.

Jewitt, G. (2005), Water and forests, in *Encyclopedia of Hydrological Sciences*, edited by M. G. Anderson, chap. 186, pp. 2895–2910, John Wiley, Hoboken, N. J.

Johnson, A. C., and R. T. Edwards (2008), Reply to discussion on "Ground-water response to forest harvesting: Implications for hillslope stability" by A. C. Johnson, R. T. Edwards, and R. Erhardt, *J. Am. Water Resour. Assoc.*, *44*(4), 1062–1065.

Johnson, A. C., R. T. Edwards, and R. Erhardt (2007), Ground-water response to forest harvest: Implications for hillslope stability, *J. Am. Water Resour. Assoc.*, *43*, 134–147.

Johnson, D. H. (1999), The insignificance of statistical significance testing, *J. Wildl. Manage.*, *63*, 763–772, doi:10.2307/3802789.

Jones, J. A. (2000), Hydrologic processes and peak discharge response to forest removal, regrowth, and roads in 10 small experimental basins, western Cascades, Oregon, *Water Resour. Res.*, *36*, 2621–2642, doi:10.1029/2000WR900105.

Jones, J. A., and G. E. Grant (1996), Peak flow responses to clear-cutting and roads in small and large basins, western Cascades, Oregon, *Water Resour. Res.*, *32*, 959–974, doi:10.1029/95WR03493.

Jones, J. A., and G. E. Grant (2001), Comment on "Peak flow responses to clear-cutting and roads in small and large basins, western Cascades, Oregon'', *Water Resour. Res.*, *37*, 179–180, doi:10.1029/2000WR900115.

Jones, J. A., and D. A. Post (2004), Seasonal and successional streamflow response to forest cutting and regrowth in the northwest and eastern United States, *Water Resour. Res.*, *40*, W05203, doi:10.1029/2003WR002952.

Jones, J. A., F. J. Swanson, B. C. Wemple, and K. U. Snyder (2000), Effects of roads on hydrology, geomorphology and disturbance patches in stream networks, *Conserv. Biol.*, *14*, 76–85, doi:10.1046/j.1523-1739.2000.99083.x.

Kaimowitz, D. (2004), Useful myths and intractable truths: The politics of the link between forests and water in Central America, in *Forests, Water and People in the Humid Tropics*, edited by M. Bonell and L. A. Bruijnzeel, pp. 86–98, Cambridge Univ. Press, Cambridge, U. K.

Katz, R. W. (1993), Towards a statistical paradigm for climate change, *Clim. Res.*, *2*, 167–175, doi:10.3354/cr002167.

Katz, R. W., and B. G. Brown (1992), Extreme events in changing climate: Variability is more important than averages, *Clim. Change*, *21*, 289–302, doi:10.1007/BF00139728.

Katz, R. W., M. B. Parlange, and P. Naveau (2002), Statistics of extremes in hydrology, *Adv. Water Resour.*, *25*, 1287–1304, doi:10.1016/S0309-1708(02)00056-8.

Kimmins, J. P., C. Welham, B. Seeley, M. Meitner, R. Rempel, and T. Sullivan (2005), Science in forestry: Why does it sometimes disappoint or even fail us?, *For. Chron.*, *81*, 723–734.

King, J. G. (1989), Streamflow responses to road building and harvesting: A comparison with the equivalent cut area procedure, *U.S. Res. Pap. INT-401*, 13 pp., Intermtn. Res. Stn., For. Serv., U.S. Dep. of Agric., Ogden, Utah.

Kirk, R. E. (1996), Practical significance: A concept whose time has come, *Educ. Psychol. Meas.*, *56*, 746–759, doi:10.1177/0013164496056005002.

Klemeš, V. (1974), Some problems in pure and stochastic hydrology, in *Proceedings of a Symposium on Statistical Hydrology*, *Misc. Publ.*, *1275*, pp. 2–15, U.S. Dep. of Agric., Washington, D. C.

Klemeš, V. (1986), Dilettantism in hydrology: Transition or destiny?, *Water Resour. Res.*, *22*, 177S–188S, doi:10.1029/WR022i09Sp0177S.

Klemeš, V. (2000), Tall tales about tails of hydrological distributions 1 and 2, *J. Hydrol. Eng.*, *5*, 227–239, doi:10.1061/(ASCE)1084-0699(2000)5:3(227).

Kochenderfer, J. N., M. B. Adams, G. W. Miller, and D. J. Helvey (2007), Factors affecting large peakflows on Appalachian watersheds: Lessons from the Fernow Experimental Forest, *Res. Pap. NRS-3*, 24 pp., Northern Res. Stn., For. Serv., U.S. Dep. of Agric., Newtown Square, Pa.

Kuhn, T. S. (1970), *The Structure of Scientific Revolutions*, 2nd ed., Univ. of Chicago Press, Chicago, Ill.

La Marche, J. L., and D. P. Lettenmaier (2001), Effects of forest roads on flood flows in the Deschutes River, Washington, *Earth Surf. Processes Landforms*, *26*, 115–134, doi:10.1002/1096-9837(200102)26:22005115::AID-ESP166>3.0.CO;2-O.

Landres, P. B., P. Morgan, and F. J. Swanson (1999), Overview of the use of natural variability concepts in managing ecological systems, *Ecol. Appl.*, *9*, 1179–1188.

Laurance, W. F. (2007), Forests and floods, *Nature*, *449*, 409–410, doi:10.1038/449409a.

Leopold, L. B. (1972), Hydrologic research on instrumented watersheds, *IASH Publ.*, *97*, 134–150.

Leopold, L. B. (1981), The topology of impacts, in *Cumulative Effects of Forest Management on California Watersheds*, *Spec. Publ., 3268*, 65 pp., Div. of Agric., Sci., Univ. of Calif., Berkeley.

Leopold, L. B., and T. J. Maddock (1954), *The Flood Control Controversy*, Ronald Press, New York.

Lewis, J., S. R. Mori, E. T. Keppeler, and R. R. Ziemer (2001), Impacts of logging on storm peak flows, flow volumes and suspended sediment loads in Caspar Creek, California, in *Land Use and Watersheds: Human Influence on Hydrology and Geomorphology in Urban and Forest Areas*, *Water Sci. Appl. Ser.*, vol. 2, edited by M. S. Wigmosta and S. J. Burges, pp. 85–125, AGU, Washington, D. C.

Lieberman, J. A., and M. D. Hoover (1951), Stream-flow frequency changes on Coweeta experimental watersheds, *Eos Trans. AGU*, *32*(1), 73–76.

Lin, Y., and X. Wei (2008), The impact of large-scale forest harvesting on hydrology in the Willow watershed of Central British Columbia, *J. Hydrol.*, *359*, 141–149, doi:10.1016/j.jhydrol.2008.06.023.

Link, W. A., and J. R. Sauer (1996), Extremes in ecology: Avoiding the misleading effects of sampling variation in summary analyses, *Ecology*, *77*, 1633–1640, doi:10.2307/2265557.

Lukacs, P. M., W. L. Thompson, W. L. Kendall, W. R. Gould, P. F. Doherty, K. P. Burnham, and D. R. Anderson (2007), Concerns regarding a call for pluralism of information theory and hypothesis testing, *J. Appl. Ecol.*, *44*, 456–460, doi:10.1111/j.1365-2664.2006.01267.x.

Lull, H. W., and K. G. Reinhart (1972), Forests and floods in the eastern United States, *Res. Pap. NE-226*, For. Serv., U.S. Dep. of Agric., Washington, D. C.

MacDonald, L. H., and D. Coe (2007), Influence of headwater streams on downstream reaches in forested areas, *For. Sci.*, *53*, 148–168.

MacDonald, L. H., and J. Stednick (2003), Forests and water: A state of-the-art review for Colorado, *Completion Rep., 196*, 65 pp., Colo. Water Resour. Res. Inst., Fort Collins.

Martin, C. W., J. W. Hornbeck, G. E. Likens, and D. C. Buso (2000), Impacts of intensive harvesting on hydrology and nutrient dynamics of northern hardwood forests, *Can. J. Fish. Aquat. Sci.*, *57*, suppl. 2, 19–20, doi:10.1139/cjfas-57-S2-19.

McCulloch, J., and M. Robinson (1993), History of forest hydrology, *J. Hydrol.*, *150*, 189–216, doi:10.1016/0022-1694(93)90111-L.

McGuinness, J. L., and L. L. Harrold (1971), Reforestation influences on small watershed streamflow, *Water Resour. Res.*, *7*, 845–852, doi:10.1029/WR007i004p00845.

Mearns, L. O., R. W. Katz, and S. H. Schneider (1984), Extreme high-temperature events: Changes in their probabilities with changes in mean temperature, *J. Clim. Appl. Meteorol.*, *23*, 1601–1613, doi:10.1175/1520-0450(1984)02320051601:EHTECI>2.0.CO;2.

Meehl, G., et al. (2000), An introduction to trends in extreme weather and climate events: Observations, socioeconomic impacts, terrestrial ecological impacts, and model projections, *Bull. Am. Meteorol. Soc.*, *81*, 413–416, doi:10.1175/1520-0477(2000)08120050413:AITTIE>2.3.CO;2.

Meyer, A. F. (1928), *The Elements of Hydrology*, 2nd ed., 522 pp., John Wiley, Hoboken, N. J.

Moore, R. D., and D. F. Scott (2005), Camp Creek revisited: Streamflow changes following salvage harvesting in a medium-sized, snowmelt-dominated catchment, *Can. Water Resour. J.*, *30*, 331–344.

Moore, R. D., and S. M. Wondzell (2005), Physical hydrology and the effects of forest harvesting in the Pacific Northwest: A review, *J. Am. Water Resour. Assoc.*, *41*, 763–784.

Mortimer, M. J., and R. J. M. Visser (2004), Timber harvesting and flooding: Emerging legal risks and potential mitigations, *South. J. Appl. For.*, *28*, 69–75.

Moscrip, A. L., and D. R. Montgomery (1997), Urbanization, flood frequency, and salmon abundance in Puget lowland streams, *J. Am. Water Resour. Assoc.*, *33*, 1289–1297, doi:10.1111/j.1752-1688.1997.tb03553.x.

Murtaugh, P. A. (2000), Paired intervention analysis in ecology, *J. Agric. Biol. Environ. Stat.*, *5*, 280–292, doi:10.2307/1400454.

Murtaugh, P. A. (2002), On rejection rates of paired intervention analysis, *Ecology*, *83*, 1752–1761.

Murtaugh, P. A. (2003), On rejection rates of paired intervention analysis, Reply, *Ecology*, *84*, 2799–2802, doi:10.1890/03-3022.

Nakano, H. (1971), Effect on streamflow of forest cutting and change in regrowth on cut-over area (in Japanese with English summary), *Bull. Gov. For. Exp. Sta.*, *240*, 1–251.

Nisbet, T. (2002), Implications of climate change: Soil and water, in *Climate Change and UK Forests*, *Bull.*, *125*, edited by M. S. J. Broadmeadow, pp. 53–67, For. Comm., Edinburgh, U. K.

Poff, N. L., et al. (1997), The natural flow regime, *BioScience*, *47*, 769–784, doi:10.2307/1313099.

Potter, W. D. (1958), Upper and lower frequency curves for peak rates of runoff, *Eos Trans. AGU*, *39*, 100–105.

Quesenberry, P. J., G. Dooner, M. Dooner, and A. Mehrdad (2005), Developmental biology: Ignoratio elenchi: Red herrings in stem cell research, *Science*, *308*, 1121–1122, doi:10.1126/science.1104432.

Ranzi, R., M. Bochicchio, and B. Bacchi (2002), Effects of floods of recent afforestation and urbanization in the Mella River (Italian Alps), *Hydrol. Earth Syst. Sci.*, *6*, 239–253.

Riedel, M. S., E. S. Verry, and K. N. Brooks (2005), Impacts of land use conversion on bankfull discharge and mass wasting, *J. Environ. Manage.*, *76*, 326–337, doi:10.1016/j.jenvman.2005.02.002.

Riekerk, H. (1989), Influence of silvicultural practices on the hydrology of pine flatwoods in Florida, *Water Resour. Res.*, *25*, 713–719, doi:10.1029/WR025i004p00713.

Robbins, J. (2008), Bark beetles kill millions of acres of trees in west, *N.Y. Times*, 18 Nov.

Robinson, M., et al. (2003), Studies of the impact of forests on peak flows and baseflows: A European perspective, *For. Ecol. Manage.*, *186*, 85–97, doi:10.1016/S0378-1127(03)00238-X.

Rothacher, J. (1965), Streamflow from small watersheds on the western slope of the Cascade Range of Oregon, *Water Resour. Res.*, *1*, 125–134, doi:10.1029/WR001i001p00125.

Rothacher, J. (1973), Does harvest in west slope Douglas-fir increase peak flow in small forest streams?, *Res. Pap. PNW-163*, 13 pp., Pac. Northwest For. and Range Exp. Stn., For. Serv., U.S. Dep. of Agric., Portland, Oreg.

Schaeffer, M., F. M. Selten, and J. D. Opsteegh (2005), Shifts in means are not a proxy for changes in extreme winter temperatures in climate projections, *Clim. Dyn.*, *25*, 51–63, doi:10.1007/s00382-004-0495-9.

Scherer, R., and G. P. Pike (2003), Effects of forest management activities on streamflow in the Okanagan Basin, British Columbia: Outcomes of a literature review and workshop, *FORREX Ser. 9*, For. Res. Ext. Partnership (FORREX), Kamloops, B. C., Canada.

Schmidt, F. L. (1996), Statistical significance testing and cumulative knowledge in psychology: Implications for training of researchers, *Psychol. Methods*, *1*, 115–129, doi:10.1037/1082-989X.1.2.115.

Schnorbus, M., and Y. Alila (2004), Forest harvesting impacts on the peak flow regime in the Columbia Mountains of southeastern British Columbia: An investigation using long-term numerical modeling, *Water Resour. Res.*, *40*, W05205, doi:10.1029/2003WR002918.

Shermer, M. (2006), Wronger than wrong: Not all wrong theories are equal, *Sci. Am.*, *295*(5), 19.

Sherrod, D. R., and J. G. Smith (1989), Preliminary map of upper Eocene to Holocene volcanic and related rocks of the Cascade Range, Oregon, *U.S. Geol. Surv. Open File Rep., 89-14*, 17 pp.

Sikka, A. K., J. S. Samra, V. N. Sharda, P. Samraj, and V. Lakshmanan (2003), Low flow and high flow responses to converting natural grassland into bluegum (*Eucalyptus globulus*) in Nilgiris watersheds of south India, *J. Hydrol.*, *270*, 12–26, doi:10.1016/S0022-1694(02)00172-5.

Sivapalan, M. (2003), Process complexity at hillslope scale, process simplicity at the watershed scale: Is there a connection?, *Hydrol. Processes*, *17*, 1037–1041, doi:10.1002/hyp.5109.

Stedinger, J. R., R. M. Vogel, and E. Foufoula-Georgiou (1993), Frequency analysis of extreme events, in *Handbook of Hydrology*, edited by D. R. Maidment, chap. 18, pp. 18.1–18.66, McGraw-Hill, New York.

Stednick, J. D. (2008a), Effects of timber harvesting on streamflow in the Alsea watershed study, in *Hydrological and Biological Responses to Forest Practices: The Alsea Watershed Study*, edited by J. D. Stednick, chap. 2, pp. 19–36, Springer, New York.

Stednick, J. D. (2008b), Long-term streamflow changes following timber harvesting, in *Hydrological and Biological Responses to Forest Practices: The Alsea Watershed Study*, edited by J. D. Stednick, chap. 10, pp. 139–155, Springer, New York.

Stednick, J. D. (2008c), Research opportunities in hydrology and biology in future water studies, in *Hydrological and Biological Responses to Forest*

Practices: The Alsea Watershed Study, edited by J. D. Stednick, chap. 17, pp. 307–314, Springer, New York.

Stephens, P. A., S. W. Buskirk, and G. D. Hayward (2005), Information theory and hypothesis testing: A call for pluralism, *J. Appl. Ecol.*, *42*, 4–12, doi:10.1111/j.1365-2664.2005.01002.x.

Storck, P., L. Bowling, P. Wetherbee, and D. Lettenmaier (1998), Application of a GIS-based distributed hydrology model for prediction of forest harvest effects on peak stream flow in the Pacific Northwest, *Hydrol. Processes*, *12*, 889–904, doi:10.1002/(SICI)1099-1085(199805)12:62005889::AID-HYP661>3.0.CO;2-P.

Struck, D. (2006), "Rapid warming" spreads havoc in Canada's forests: Tiny beetles destroying pines, *Wash. Post*, p. A1, 1 March.

Swank, W. T., and J. M. Vose (1994), Long-term hydrologic and stream chemistry responses of southern Appalachian catchments following conversion from mixed hardwoods to white pine, in *Hydrologie kleiner Einzugsgebiete: Gedenkschrift Hans M. Keller*, edited by R. Landolt, pp. 164–172, Schweiz. Ges. fur Hydrol. und Limnol., Bern, Switzerland.

Swank, W. T., J. M. Vose, and K. J. Elliott (2001), Long-term hydrologic and water quality responses following commercial clearcutting of mixed hardwoods on a southern Appalachian catchment, *For. Ecol. Manage.*, *143*, 163–178, doi:10.1016/S0378-1127(00)00515-6.

Swanson, R. H., and G. R. Hillman (1977), Predicted increased water yield after clear-cutting verified in west-central Alberta, *Inf. Rep. NOR-X-198*, Can. Dep. of Fish. and Environ., Can For. Serv., North. For. Cent., Edmonton, Alberta, Canada.

Swindel, B. F., and J. E. Douglass (1984), Describing and testing nonlinear treatment effects in paired watershed experiments, *For. Sci.*, *30*, 305–313.

Thomas, R. B. (1990), Problems in determining the return of a watershed to pretreatment conditions: Techniques applied to a study at Caspar Creek, California, *Water Resour. Res.*, *26*, 2079–2087.

Thomas, R. B., and W. F. Megahan (1998), Peak flow responses to clear-cutting and roads in small and large basins, western Cascades, Oregon: A second opinion, *Water Resour. Res.*, *34*, 3393–3403, doi:10.1029/98WR02500.

Thomas, R. B., and W. F. Megahan (2001), Reply to comment on "Peak flow responses to clear-cutting and roads in small and large basins, western Cascades, Oregon", *Water Resour. Res.*, *37*, 180–183, doi:10.1029/2000WR900277.

Tonina, D., C. Luce, B. Rieman, J. M. Buffington, P. Goodwin, S. R. Clayton, S. M. Ali, J. J. Barry, and C. Berenbrock (2008), Hydrological response to timber harvest in northern Idaho: Implications for channel scour and persistence of salmonids, *Hydrol. Processes*, *22*, 3223–3235, doi:10.1002/hyp.6918.

Troendle, C. A. (1970), Flow interval method for analyzing timber harvesting effects on streamflow regimen, *Water Resour. Res.*, *6*, 328–336, doi:10.1029/WR006i001p00328.

Troendle, C. A., and R. M. King (1985), The effect of timber harvest on the Fool Creek watershed, 30 years later, *Water Resour. Res.*, *21*, 1915–1922, doi:10.1029/WR021i012p01915.

Troendle, C. A., and W. K. Olsen (1994), Potential effects of timber harvest and water management on streamflow dynamics and sediment transport, in *Sustainable Ecological Systems: Implementing an Ecological Approach to Land Management*, *Gen. Tech. Rep., RM-247*, pp. 34–41, For. Serv., U.S. Dep. of Agric., Flagstaff, Ariz.

Troendle, C. A., and J. D. Stednick (1999), Discussion on "Effects of basin scale timber harvest on water yield and peak streamflow" by Timothy A. Burton, *J. Am. Water Resour. Assoc.*, *35*, 177–181, doi:10.1111/j.1752-1688.1999.tb05462.x.

Troendle, C. A., M. S. Wilcox, G. S. Bevenger, and L. S. Porth (2001), The Coon Creek water yield augmentation project: Implementation of timber harvesting technology to increase streamflow, *For. Ecol. Manage.*, *143*, 179–187, doi:10.1016/S0378-1127(00)00516-8.

Troendle, C. A., L. H. MacDonald, and C. H. Luce (2006), Fuels management and water yield, in *Cumulative Watershed Effects of Fuels Management: A Western Synthesis*, general technical report, Rocky Mt. Res. Stn., U.S. Dep. of Agric., Fort Collins, Colo.

Underwood, A. J. (1991), Beyond BACI: Experimental designs for detecting human environmental impacts on temporal variations in natural populations, *Aust. J. Mar. Freshwater Res.*, *42*, 569–587, doi:10.1071/MF9910569.

Underwood, A. J. (1994), On beyond BACI: Sampling designs that might reliably detect environmental disturbances, *Ecol. Appl.*, *4*, 3–15, doi:10.2307/1942110.

van Dijk, A. I. J. M., M. van Noordwijk, I. R. Calder, S. L. A. Bruijnzeel, J. Schellekens, and N. A. Chappell (2008), Forest-flood relation still tenuous—Comment on "Global evidence that deforestation amplifies flood risk and severity in the developing world" by C. J. A. Bradshaw, N. S. Sodhi, K. S.-H. Peh and B. W. Brook, *Global Change Biol.*, *15*, 110–115, doi:10.1111/j.1365-2486.2008.01708.x.

van Haveren, B. P. (1988), A reevaluation of the Wagon Wheel Gap Forest Watershed Experiment, *For. Sci.*, *34*, 208–214.

Verry, E. S., J. R. Lewis, and K. N. Brooks (1983), Aspen clearcutting increases snowmelt and storm flow peaks in north central Minnesota, *Water Resour. Bull.*, *19*, 59–67.

Waichler, S. R., B. C. Wemple, and M. S. Wigmosta (2005), Simulation of water balance and forest treatment effects at the H.J. Andrews Experimental Forest, *Hydrol. Processes*, *19*(16), 3177–3199, doi:10.1002/hyp.5841.

Wemple, B. C., J. A. Jones, and G. E. Grant (1996), Channel network extension by logging roads in two basins, Western Cascades, Oregon, *J. Am. Water Resour. Assoc.*, *32*, 1195–1207, doi:10.1111/j.1752-1688.1996.tb03490.x.

Whitaker, A., Y. Alila, J. Beckers, and D. Toews (2002), Evaluating peak flow sensitivity to clear-cutting in different elevation bands of a snowmelt-dominated mountainous catchment, *Water Resour. Res.*, *38*(9), 1172, doi:10.1029/2001WR000514.

Wigley, T. M. L. (1985), Impact of extreme events, *Nature*, *316*, 106–107.

Wilcox, R. R. (1995), ANOVA: A paradigm for low power and misleading measures of effect size?, *Rev. Educ. Res.*, *65*, 51–77.

Williams, G. P. (1978), Bank-full discharge of rivers, *Water Resour. Res.*, *14*, 1141–1480, doi:10.1029/WR014i006p01141.

Wright, K. A., K. J. H. Sendek, R. M. Rice, and R. B. Thomas (1990), Logging effects on streamflow: Storm runoff at Caspar Creek in northwestern California, *Water Resour. Res.*, *26*, 1657–1667.

Yoccoz, N. G. (1991), Use, overuse, and misuse of significance tests in evolutionary biology and ecology, *Bull. Ecol. Soc. Am.*, *72*, 106–111.

Ziemer, R. R. (1981), Storm flow response to road building and partial cutting in small streams of northern California, *Water Resour. Res.*, *17*, 907–917, doi:10.1029/WR017i004p00907.

Zinke, P. J. (1965), Influence of land use on floods in relation to runoff and peak flows, in *Proceedings: State Board of Forestry Meeting, February 26, 1965—Relationship of Land Use to 1964–1965 Flood*, pp. 47–52, Div. of For., Calif. Dep. of Conserv., Sacramento.

Y. Alila and R. Hudson, Department of Forest Resources Management, University of British Columbia, 2045-2424 Main Mall, Vancouver, BC V6T 1Z4, Canada. (younes.alila@ubc.ca; rob.hudson@telus.net)

P. K. Kuraś, Northwest Hydraulic Consultants, Ltd., 30 Gostick Place, North Vancouver, BC V7M 3G3, Canada. (peterkuras@gmail.com)

M. Schnorbus, Pacific Climate Impacts Consortium, University of Victoria, P.O. Box 1700, Station CSC, Victoria, BC V8W 2Y2, Canada. (mschnorb@uvic.ca)