HYDROLOGICAL PROCESSES
Hydrol. Process. **19**, 3177–3199 (2005)
Published online 10 May 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/hyp.5841

Simulation of water balance and forest treatment effects at the H.J. Andrews Experimental Forest

Scott R. Waichler,[1]* Beverley C. Wemple[2] and Mark S. Wigmosta[1]

[1] *Pacific Northwest National Laboratory, K9-36, PO Box 999, Richland, WA 99352, USA*
[2] *Department of Geography, University of Vermont, Burlington, VT, USA*

Abstract:

The distributed hydrology soil–vegetation model (DHSVM) was applied to the small watersheds WS1, 2, 3 in H.J. Andrews Experimental Forest, Oregon, and tested for skill in simulating observed forest treatment effects on streamflow. These watersheds, located in the rain–snow transition zone, underwent road and clearcut treatments during 1959–66 and subsequent natural regeneration. DHSVM was applied with 10 m and 1 h resolution to 1958–98, most of the period of record. Water balance for old-growth WS2 indicated that evapotranspiration and streamflow were unlikely to be the only loss terms, and groundwater recharge was included to account for about 12% of precipitation; this term was assumed zero in previous studies. Overall efficiency in simulating hourly streamflow exceeded 0·7, and mean annual error was less than 10%. Model skill decreased at the margins, with overprediction of low flows and underprediction of high flows. However, statistical analyses of simulated and observed peakflows yielded similar characterizations of treatment effects. Primary simulation weaknesses were snowpack accumulation, snowmelt under rain-on-snow conditions, and production of quickflow. This was the first test of DHSVM against observations of both control and treated watersheds in a classic paired-basin study involving a long time period of forest regrowth and hydrologic recovery. Copyright © 2005 John Wiley & Sons, Ltd.

KEY WORDS DHSVM; HJA; watershed modelling; forest treatment

INTRODUCTION

Defining impacts of forest practices on catchment hydrology is an important issue in natural resources management. Paired-basin studies involving measurement in a control watershed and one or more treated watersheds is the most direct and effective method for identifying hydrologic treatment effects if the overall period of record and duration of the treatment itself are long enough. Numerous paired-basin field studies have evaluated the effects of road construction and vegetation removal on basin hydrologic response at the H.J. Andrews Experimental Forest, Oregon (HJA), including annual water yield (Rothacher, 1970), peak streamflows (Harr and McCorison, 1979; Jones and Grant, 1996; Thomas and Megahan, 1998; Jones, 2000), and summer low flows (Rothacher, 1965). Most of these studies did find forest treatment effects on streamflow, with varying degrees of significance. The goal of this study is to evaluate how well the physically based distributed hydrology soil–vegetation model (DHSVM) can reproduce observed water balance and forest treatment effects, using the small watersheds WS1, 2, 3 at the HJA as the test case.

Two previous studies applied statistical techniques to the HJA dataset to investigate magnitude and persistence of forest treatment effects. Jones and Grant (1996) compiled and analysed peak streamflows for 34 years of record on WS1, 2, 3. Using analysis of variance to detect differences between treated and control basins, they found statistically significant differences in peak flow magnitudes following clearcut harvesting in WS1 that were detectable 22 years after harvesting. In WS3, where treatment was road construction followed

* Correspondence to: Scott R. Waichler, Pacific Northwest National Laboratory, K9-36, PO Box 999, Richland, WA 99352, USA.
E-mail: scott.waichler@pnl.gov

Received 14 July 2003
Accepted 10 September 2004

4 years later by harvesting 25% of basin area, statistically significant increases in peak flow magnitudes were detected only after forest harvesting occurred but were still detectable 25 years after harvest. Thomas and Megahan (1998) reanalysed these data using linear regression and presented new statistical models for the effects of treatment with time. Both studies documented similar magnitudes in peak flow increases following forest treatments, particularly for the smallest flow events. Thomas and Megahan (1998) did not find statistically significant treatment effects on WS3 more than 10 years after harvesting. They showed that treatment effects diminished with time, and concluded that flow increases were detectable only for flows less than the 2 year recurrence interval. These studies spurred an ongoing debate about the mechanisms responsible for changes in peak flows in these basins, the effects of forest clearing and roads on large floods, and the persistence of forest treatment effects through time.

Empirical studies provide limited ability to extrapolate to other settings, and much of the value and promise of process-based modelling is to provide a means of evaluating the impacts of land-use practices in areas lacking sufficient field study. Recent applications of DHSVM have evaluated the impacts of forest harvesting and road construction on watersheds in western Washington (Storck *et al*., 1998; Bowling *et al*., 2000; LaMarche and Lettenmaier, 2001) and the interior Columbia basin (VanShaar *et al*., 2002), and these studies ascribed significant predictive power to DHSVM for evaluation of hydrologic impacts from forest treatment. We modelled such impacts at an important research site in Oregon. HJA is one of the longest-running field sites in the Long Term Ecological Research (LTER) network and has produced a great deal of research on the hydrology of steep, forested catchments and the impacts of road building and timber harvest. Little published testing and improvement of process-based hydrologic models has been done at HJA, and no previous efforts have attempted to simulate basin hydrology at a fine temporal resolution over the entire period of record. Only two published studies and two theses have modelled the catchment hydrology with physically based models, and these did so on limited time spans and watersheds (Duan, 1996; Bredensteiner, 1998; Tague and Band, 2001b).

Most modelling studies of treatment effects use either a scenario method and compare land cover states (e.g. VanShaar *et al*., 2002)), or use a residuals method and look for trends in the residual time series, where the residuals are the difference between observed and simulated streamflows (e.g. Bowling *et al*., 2000). The scenario method involves differencing model outputs where the land-cover state in each scenario is held constant through time. Such simulations are operationally convenient and have the advantage of eliminating climate as a source of error in evaluating the treatment effect. However, scenario simulations may not adequately address model verification, usually span a short chronological period, and ignore the effect of forest regrowth on hydrologic recovery. In the residuals method, the simulated streamflow time series is produced by driving a constant land-cover state with observed meteorology over the same climate period as the observed streamflows. The residuals method minimizes the confounding effect of climate and takes advantage of the potentially long streamflow record from the treated watershed, but it does not give insight into the physical processes that might cause the change in hydrologic regime. This study incorporated aspects of both the scenario and residuals methods by simulating a long time period and contrasting control and treated conditions that actually occurred. This combination of characteristics permits model testing against empirical findings of treatment effects under real conditions involving forest regrowth and hydrologic recovery.

METHODS

We applied DHSVM to WS1, 2, 3 at the HJA to assess our ability to reproduce observed forest treatment effects on streamflow, including changes in peak flows that have been observed statistically. We simulated 41 years, most of the record, in all three watersheds at 10 m and 1 h resolution. Our analysis focused on WS1 and WS2 because these watersheds had the largest contrast in hydrologic regime. Four primary challenges in simulating these small watersheds emerged: (1) closing the water balance; (2) calibrating the model to achieve skill at reproducing both high and low flows in multiple watersheds; (3) reproducing observed forest

treatment effects that are subtle; (4) addressing the confounding effect that meteorology from different time periods had on treatment effects.

Site description

HJA is located in the Western Cascades province of Oregon (Figure 1). Its maritime climate is dominated by frontal systems from the Pacific Ocean during November–May and by high-pressure systems that produce warm, dry conditions for the rest of the year. Average annual precipitation at WS2 is 2300 mm, with a mean temperature of 9 °C. The small watersheds WS1, 2, 3 range from 0·6 to 1·0 km^2 in area, and have elevations ranging from 450 to 1000 m (Table I), placing them in the rain–snow transition zone. All three

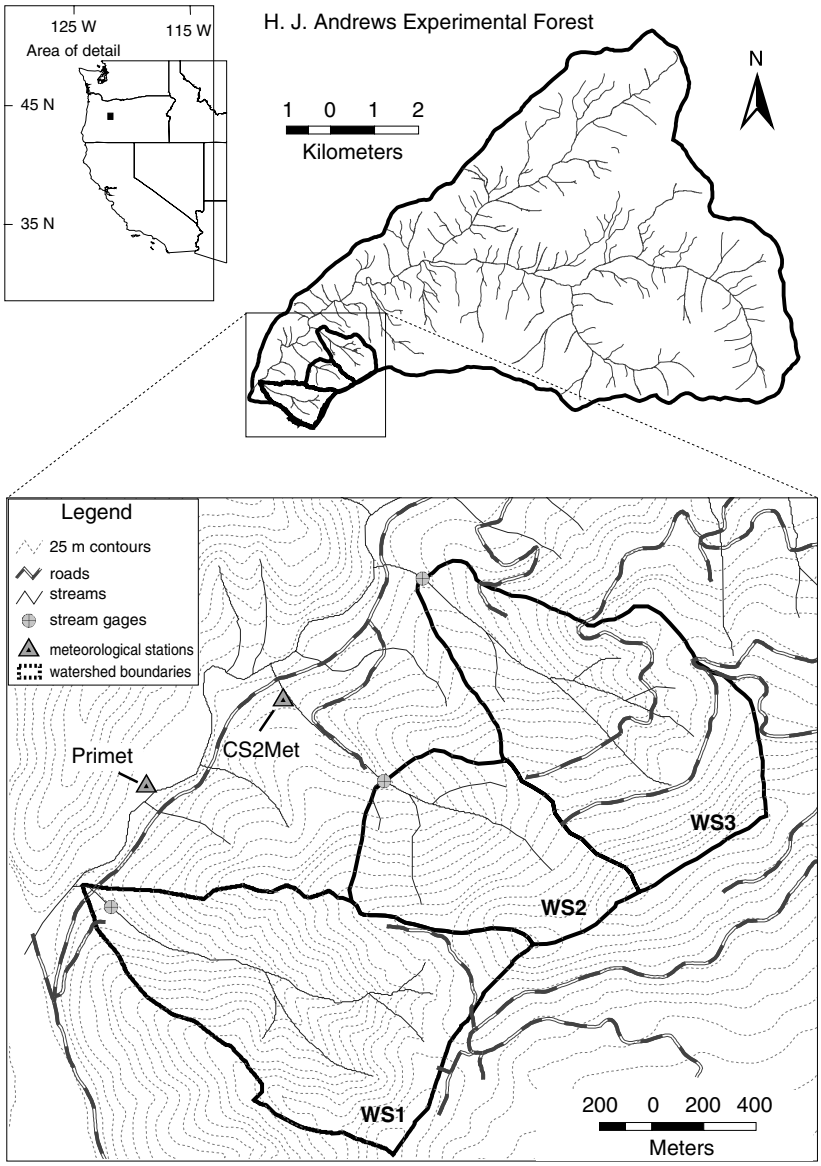

Figure 1. H.J. Andrews Experimental Forest, small watersheds WS1, 2, 3, and climate stations

Table I. HJA small watersheds WS1, 2, 3

Basin	Size (km^2)	Elevation (m)	Road density ($km\ km^{-2}$)	Treatment (change from old-growth forest)
WS1	0·96	460–990	0	Clearcut 1962–66; burned October 1966
WS2	0·60	530–1070	0	Untreated
WS3	1·01	490–1070	3·0	Roaded 1959; 25% patch cut 1963

watersheds were covered with old-growth Douglas fir–western hemlock forest before the forest treatments and the empirical study began in the late 1950s and early 1960s. WS2 has served as a control for the paired watershed experiment; its vegetation has remained old growth.

Description of DHSVM and model inputs

The DHSVM is a process-based, distributed parameter hydrologic model designed for simulating runoff processes in forested, mountainous environments. Specification of vegetation and soil types is done at the resolution of a digital elevation model (DEM) grid. Elevation data of the DEM are used to simulate topographic controls on absorbed shortwave radiation, precipitation, air temperature and downslope water movement. The model simulates canopy interception, evaporation, transpiration, snow accumulation and melt in the canopy and on the ground, vertical unsaturated water flow, and lateral saturated groundwater flow. The structure of the DHSVM is described by Wigmosta *et al*. (1994, 2002).

Input to the DHSVM includes: grids of surface elevation, soil type, soil thickness, and vegetation type; tables of soil and vegetation biophysical parameter values keyed to type; tables of grid locations occupied by stream channels and road segments; tables of stream and road parameter values; and time series of the meteorological variables air temperature, precipitation, wind speed, relative humidity, solar radiation and downward longwave radiation from one or more stations. All raw data were obtained from the Forest Science Data Bank (FSDB, www.fsl.orst.edu/lter). Watersheds and channel networks were defined with a 10 m DEM, stream gauge locations, standard ArcInfo algorithms, and manual editing. Soil maps and parameters were taken from data measured by Dyrness (1969) and compiled by Bredensteiner (1998). Road culvert locations were obtained from Wemple (1998). Meteorologic data were obtained from the primary meteorology station PRIMET and another station located just below WS2, CS2MET (Figure 1) (Henshaw *et al*., 1998). Hourly data from PRIMET were used for water year (WY)80–WY98. Hourly meteorology values for WY58–WY79 were generated from daily CS2MET data. Further information regarding spatial and meteorology inputs can be found in Waichler *et al*. (2002) and Waichler and Wigmosta (2003) respectively.

WS1,2,3 have had varying vegetation states: WS2 is entirely old-growth, WS1 is entirely regrown forest, and WS3 has patches of regrown forest inside an old-growth matrix. We incorporated vegetation regrowth through simple equations of leaf area index (LAI) and overstory vegetation height H_o. Overstory and understory LAI were set equal to old-growth values (8·5, 0·5) during the pretreatment period, then declined linearly during the treatment period to 0·01, then recovered nonlinearly after treatment ended. Overstory height was reduced to 1·0 m and understory height was assumed constant at 0·5 m, maintaining the canopy higher than the understory as required by the current model structure. Regrowing overstory leaf area LAI_o was simulated with an exponential growth function developed by Richards (1959) and applied to HJA by Duan (1996).

$$LAI_o = LAI_{max}(1 - e^{-B_1 t}) \tag{1}$$

where $LAI_{max} = 8{\cdot}5$, t is time (fractional years), and $B_1 = 0{\cdot}065$. Regrowing understory leaf area LAI_u was simulated with a modified version of Equation (1) to take into account the shading effect of the overstory:

$$LAI_u = LAI_{max}(1 - e^{-B_1 t}) - 0{\cdot}15\ LAI_o \tag{2}$$

where $LAI_{max} = 2{\cdot}0$ and $B_1 = 1{\cdot}0$. Regrowing overstory height was simulated with

$$H_o = 1{\cdot}0 + H_{max}(1 - e^{-B_1 t}) \tag{3}$$

where $H_{max} = 59$ m and $B_1 = 0{\cdot}05$. Physiological properties were set equal for both vegetation types.

In addition to trying to reproduce historical conditions, we also simulated alternative treated/untreated scenarios for WS1, where the treatment was 100% clearcut imposed at dates different from historical reality. For these simulations, regrowth from a clearcut state began on 1 October 1957 and 1 October 1979 for the periods WY58–WY79 and WY80–WY98 respectively. These periods were selected to coincide with the different sources of meteorology data mentioned previously. The scenario of no treatment was defined as constant old-growth conditions.

Water balance and calibration

The first task of any interannual watershed modelling effort is to obtain a reasonable water balance at long time scales. At the HJA, most researchers have assumed that the water balance in the small watersheds can be described as

$$P = Q + ET + \Delta S \tag{4}$$

where P is precipitation, Q is streamflow (runoff), ET is evapotranspiration, and ΔS is change in soil moisture. For the small watersheds at HJA, P and Q are assumed to be adequately measured, and Q is the primary validation variable available for modellers. Most HJA researchers have assumed that the concrete weirs at the WS1, 2, 3 gauging stations, having bases installed into the regolith, force all groundwater flow to the surface for in-channel measurement, and hence Equation (4) lacks a groundwater recharge term. However, it is possible that subsurface flow exits the small watersheds as either shallow groundwater flow near the gauges, or as deep groundwater flow with a significant vertical component, as within a larger scale flow system. On an average annual basis, $\Delta S = 0$, so the feasibility of Equation (4) depends on the magnitude of ET required to complete the balance.

ET computed from $P - Q$ on an annual basis can be compared with empirical estimates from similar environments. Using transpiration estimates based on sap flux measurements at HJA (Moore *et al*., 2004) and wet canopy evaporation estimates based on eddy-flux measurements at a similar environment in Washington (Link *et al*., 2004), an upper limit for annual ET was estimated as 740 mm (Table II). These previous studies observed ET for parts of 1999 and 2000, years that had typical weather, as evaluated from records of monthly air temperature and precipitation during 1979–2002 at HJA. The benchmark of 740 mm year^{-1} is considered an upper bound estimate because the daily wet canopy evaporation rates are potential values, and because the transpiration rates of Moore *et al*. (2004) were measured in a riparian zone and did not take into account the drier hillslopes. In almost every year at HJA, implied ET, i.e. $P - Q$, was greater than this independent estimate (Figure 2). Either the independent estimate is too low, or groundwater flux/recharge G is another loss term in the local water balance. One possibility for underestimating ET is neglecting the role of duff, logs, and soil in holding and releasing water. Harmon and Sexton (1995) found that logs alone could store and release about 45 mm year^{-1} at HJA. Nijssen *et al*. (1997) found evaporation from the moss and litter layer in the BOREAS experiment to be very significant, albeit with a much more open canopy and less canopy ET compared with an environment like HJA. However, the importance of groundwater flux in catchment water balance has also been demonstrated. Confounding subsurface flow paths in carefully monitored low-order basins have been found in rock types that contain far fewer fractures than the volcanic rock at the HJA (e.g. Anderson *et al*., 1997).

DHSVM was calibrated by trying to match observed annual, monthly, and hourly Q and annual, monthly, and daily ET as informed by Table II. 'Hard' calibration based on error minimization at particular time steps was done for streamflow but not for ET, because the limited data available did not support it. Instead, ET rates from HJA (Moore *et al*., 2004) and a similar environment in Wind River, Washington (Link *et al*., 2004)

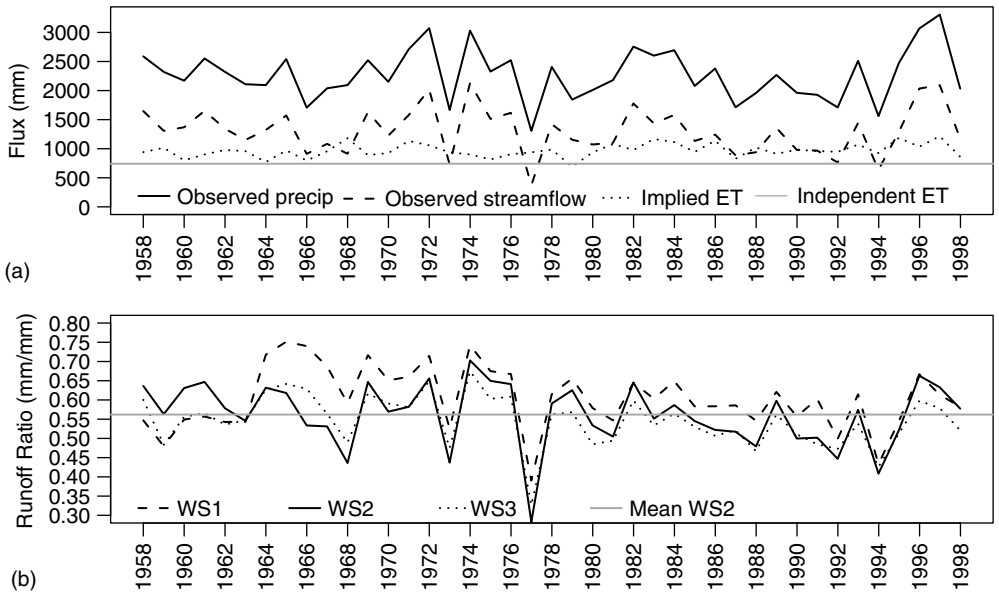

Figure 2. Annual fluxes. (a) Observed precipitation and streamflow and estimated evapotranspiration. Implied ET equals difference between precipitation and streamflow and assumes no change in storage or deep groundwater recharge. Independent ET equals estimate of ET using limited sap flux and eddy flux data from HJA and Wind River, Washington (Link *et al*., 2004; Moore *et al*., 2004). (b) Runoff ratio, streamflow/precipitation

Table II. Independent estimates of daily HJA evapotranspiration. Rates for canopy evaporation (from interception) are from eddy flux measurements at the Wind River site (Link *et al*., 2004). Rates for transpiration are based on sap flux measurements at WS1, 2 (Moore *et al*., 2004). Mean daily ET is estimated as a weighted average of canopy evaporation (occuring on wet days), and transpiration (dry days), and is conservatively high, since on many wet days evaporation from interception takes up only part of the day

Month	Average number of wet days	Average number of dry days	Mean daily flux (mm)		
			Canopy evaporation (on wet days)	Transpiration (on dry days)	Mean daily ET
Oct	12	19	3	1	1·8
Nov	19	11	2	1	1·6
Dec	20	11	1	0·5	0·8
Jan	20	11	1	0·5	0·8
Feb	18	10	2	0·5	1·5
Mar	19	12	3	1	2·2
Apr	18	12	4	1·5	3·0
May	13	18	5	2	3·3
Jun	9	21	6	2·5	3·6
Jul	4	27	5	2	2·4
Aug	5	26	4	1·5	1·9
Sep	8	22	3	1	1·5
Annual total	165	200			742

were used in a 'soft' calibration that (1) limited maximum ET to approximately 3 mm day^{-1} and (2) achieved monthly ET that approximated the synthesis estimates in Table II. As stated above, the years 1999 and 2000 observed in these studies had fairly average climatology at HJA and Wind River. Simulated groundwater

recharge G was used as a remainder term to complete the water balance after matching the target streamflow and ET fluxes as well as possible. Consideration of monthly P, Q, and ET led to the hypothesis that deep groundwater flux was roughly proportional to precipitation, i.e. it mostly occurred during the wet-season months. Perhaps the recharge is most active during the wet season through transient and saturation-dependent formation of preferential flow into the bedrock, as found by Sidle *et al*. (2000) in Japanese hillslopes. We accounted for the apparent seasonality of recharge by scaling unit hydraulic gradient in Darcy's law by the thickness of the saturated soil, so that recharge was proportional to water table height.

The objective function used in calibration was a combination score based on equal weighting of bias and baseline-adjusted efficiency E'_1, which compares the model predictions with an alternative predictor comprised of 12 monthly means of the observations (see Appendix). Adjustment of the following set of parameters was done in a trial-and-error process: lateral (K_l) and vertical (K_v) hydraulic conductivity, leakage conductivity for groundwater recharge (K_g), and rain and snow temperature thresholds. K_l and K_v were scaled uniformly to maintain relative differences across soil types. K_g was constant across soil types.

The selection of watersheds and time periods on which to base the calibration started with the control watershed (WS2) during the pretreatment period (WY58–WY62), as this would be the situation of a typical forest manager wanting to predict impacts of future harvesting. However, the streamflow response at WS2 during WY58–WY62 was found to be somewhat anomalous in comparison with other years, with the runoff ratio during that time markedly higher than WS1 and WS3 (Figure 2b). Variable evapotranspiration was unlikely to be the cause of such a large difference, because the forest type was the same across all three basins. If ET was not significantly different between the watersheds during this time, then it seems plausible that regolith permeability and G could explain the difference. However, it is difficult to confirm that G varies between basins because the pretreatment period was relatively brief, and, after 1962, changing land surface influenced the water balance of the treated watersheds and obscured any evidence for different groundwater recharge rates. The runoff ratio of WS1 clearly increased after harvest, and stayed higher than the others for a long time, but all three watersheds converged again around 1994. During the large flood of February 1996, the gauge at WS3 was destroyed in a debris flow and, beginning in that year, the runoff ratio for WS3 is distinctly lower than the others, indicating that there is more groundwater flow after the flood, perhaps due to the new weir construction or the scoured condition of the channel bed upstream.

To evaluate whether the higher runoff at WS2 during the pretreatment period was statistically significant, annual streamflow was regressed on annual precipitation for the WY58–WY62 and WY58–WY98 periods. The WY58–WY62 regression line was within the 95% confidence band surrounding the WY58–WY98 regression line. We inferred from this that high streamflow during the pretreatment period was not significantly different from other years. From a practical standpoint, however, we had to incorporate other time periods to achieve a reasonable calibration for the entire period of record; therefore, two later periods were selected as being representative of varied climatic conditions and having observed hourly meteorology input: WY80–WY83 and WY94–WY98. We also included WS1 during WY58–WY62 in the calibration process. No calibration was done with data from the regrowth period of WS1, and WS3 was not used at all. The key assumption of the paired watershed study is that these basins have similar climate, soils, and vegetation. After one accounts for any minor differences in these characteristics with the model input, the output should be as good for WS1 and WS3 as it is for WS2. However, there is also the possibility that the representation of the regrowing vegetation is inadequate, the effects of which would be difficult to distinguish from an inadequate representation of other model characteristics, such as soil properties and groundwater recharge.

Peakflow selection and statistical tests

Two peakflow datasets were defined for the historical WS1 versus WS2 analysis. The first set comprised observed peakflows in WS1 and WS2 from the Jones and Grant (1996) events, plus additional observed events that we identified by hydrograph inspection that met the minimum flow standard of 0·03 $m^3\ s^{-1}$ mentioned in Jones and Grant (1996). The purpose of using this first dataset was to see whether the conclusions of

Thomas and Megahan (1998) would remain the same with different samples and hourly peaks as opposed to instantaneous peaks, and to provide a benchmark for comparing the simulations. The second set comprised simulated peakflows in WS1 and WS2. Some observed peakflows had no corresponding peakflows in the model output, so the second set was limited to those high-flow events where peaks were evident in both the observed and simulated flows. Sample sizes for Jones and Grant (1996), our augmented observed dataset, and our simulated set were 308, 329, and 308 respectively.

Two peakflow datasets were also defined for the hypothetical treated WS1 versus untreated WS1 scenario analysis. The first set comprised simulated peakflows from WY58–WY79, and the second set comprised simulated peakflows from WY80–WY98. In both sets, peakflows were based on the same population of Jones and Grant (1996) events (plus our additions) that was used in the historical analysis.

For continuity with previous studies, and to characterize a continuous relationship over a range in flow magnitude, the regression methods of Thomas and Megahan (1998), as revised by Thomas (personal communication, 2002), and two new regression methods were applied to simulated peakflows. We wanted to determine whether model output was of sufficient quality to draw the same conclusions as those based on observations in the previous empirical studies. The first part of the analysis focused on the comparison of historical conditions in WS1 and WS2. The first statistical model was a simple linear regression of hourly peakflow in WS1 on hourly peakflow in WS2, as in Eq. 3 of Thomas and Megahan (1998):

$$y_i = A_i + B_i x_i + e_i \tag{5}$$

where y is $\log_e$ of WS1 peakflow ($m^3\ s^{-1}$), x is $\log_e$ of WS2 peakflow ($m^3\ s^{-1}$), A and B are coefficients, e is error, and i is treatment period. Equation (5) was applied to the same four recovery periods used by Jones and Grant (1996) and Thomas and Megahan (1998) (November 1966–December 1971, January 1972–December 1976, January 1982–June 1988) and all regression models were significant ($p < 0{\cdot}0001$).

Thomas and Megahan (1998) used the following criteria to determine whether the relationship between WS1 and WS2 peakflows in a recovery period was different from that of the pretreatment period. If the slope B, intercept A, or both were different, then they deemed the recovery period to be different from pretreatment. Significance level for slope and intercept terms was the Bonferroni adjustment of 'experimentwise' error, $0{\cdot}05/8 = 0{\cdot}006\,25$. Subsequently, R.B. Thomas (personal communication, 2002) identified a problem with flow units in the original analysis and devised an improved method for testing the similarity of pretreatment and recovery periods. In the new method, slope terms are tested first. If the slope of a recovery period is different from the pretreatment, then the recovery period is judged to be different. Otherwise, analysis continues to test the intercepts as follows. Data from recovery periods where slope is not different from the pretreatment, plus the pretreatment data, are included in fitting a constant-slope regression of the form

$$y_i = C_1 G_1 + C_2 G_2 + \cdots + C_k G_k + C_{k+1} x_i + e_i \tag{6}$$

where y_i, x_i, and e_i are as before, $C_{j=1\ldots k}$ are coefficients estimated by regression, $G_{j=1\ldots k}$ are indicator coefficients identifying each period, with a value of one for treatment period j and a value of zero for others, and k is the number of included recovery periods plus one. The final step of the revised method is to test the differences between recovery period and pretreatment period intercepts C_j resulting from Equation (6), i.e. the vertical separation between constant-slope regression lines.

The second regression model included time as a continuous variable, providing one equation for the whole experiment, as in Thomas and Megahan (1998: equation (4)):

$$y_i = A + B x_i + D t + e_i \tag{7}$$

where t is the time since treatment ended (fractional years) and D is another coefficient. The significance level for the time term was set at 0·05.

The second part of the statistical analysis compared two scenarios in the same watershed, treated WS1 and untreated WS1. Here, Equation (5) was used with $y = \log_e$ (peakflows from treated WS1) and $x = \log_e$ (peakflows from untreated WS1). Equation (7) was applied similarly. In the WS1 scenarios, all years were recovery years, and these were assigned to four periods with lengths similar to Thomas and Megahan (1998). Since the watershed was the same in both scenarios, the regression line would have a unit slope and zero intercept if there were no treatment effect. Unlike Thomas and Megahan (1998), where the focus was on finding differences between the regressions of the recovery periods and the pretreatment period, the test of interest here was whether the slope was different from one and the intercept different from zero.

To check whether underlying assumptions for regression were met, plots of residuals versus control, histograms of residuals, and quantile–quantile plots with normal probability were made for all regressions, and inspection of these indicated no problems.

RESULTS

Results are presented in four sections corresponding to the major themes of the paper: water balance closure, overall simulation of streamflow, simulation of peakflows, and impact of forcing data (meteorology) on the predictions. In each section, simulations of historical reality are described first, followed by the scenario simulations. The scenario model runs used the same chronologically correct meteorology input as the historical simulation, but shifted the forest treatment in time.

Water balance

The simulated annual water balance for WS1, 2, 3 is shown in Table III. Both this study and Jones (2000) neglected increases of precipitation with elevation in deriving the local water balance. Lapse rates estimated with the PRISM climate model (Daly *et al.*, 1996) as applied to multiple HJA sites with a range of elevations are positive and result in a 13% increase in mean WS1 precipitation compared with the data from the base climate station CS2MET. Implied ET calculated from base precipitation and streamflow is already higher than estimates based on measured ET (Table II, Figure 2), so the existence of positive lapse rates for precipitation lends further support for the existence of groundwater recharge as a significant term in the water balance. Undercatch in precipitation measurements would further increase the amount of precipitation that must be accounted for. Simulated evapotranspiration was highest in the untreated, old-growth WS2, and lowest in WS1, which experienced the greatest reduction in vegetation. Simulated groundwater recharge was 12–13% of precipitation.

The simulated mean monthly water balance (Figure 3b) contrasts with the results of Jones (2000) (Figure 3a), who assumes Equation (4) and derives estimates of ET using the Thornthwaite method for potential evapotranspiration. The cumulative difference between monthly fluxes $P - Q - \mathrm{ET} - G$ should be close to zero at the end of the average WY. However, in the previously reported water balance there is a final error of about 470 mm, or 21% of P. The total annual ET in that water balance is also about 470 mm, whereas the simulated ET in this study is 630 mm. G was included in our model to provide a plausible mechanism for the loss that cannot be explained by ET. We also found a different seasonality for ET; maximum rates

Table III. Major water balance components, WY58–WY98. Fluxes are noted as simulated or observed. Runoff ratio $R = (Q_{\mathrm{sim}}/P_{\mathrm{obs}})$, $Q_{\mathrm{error}} = [(Q_{\mathrm{sim}} - Q_{\mathrm{obs}})/Q_{\mathrm{obs}}] \times 100$

Watershed	P_{obs} (mm year^{-1})	Q_{sim} (mm year^{-1})	$\mathrm{ET}_{\mathrm{sim}}$ (mm year^{-1})	G_{sim} (mm year^{-1})	R	Q_{error} (%)
WS2	2177	1272	628	277	0·58	−2·6
WS1	2177	1391	536	250	0·64	−0·1
WS3	2177	1287	599	291	0·59	4·2

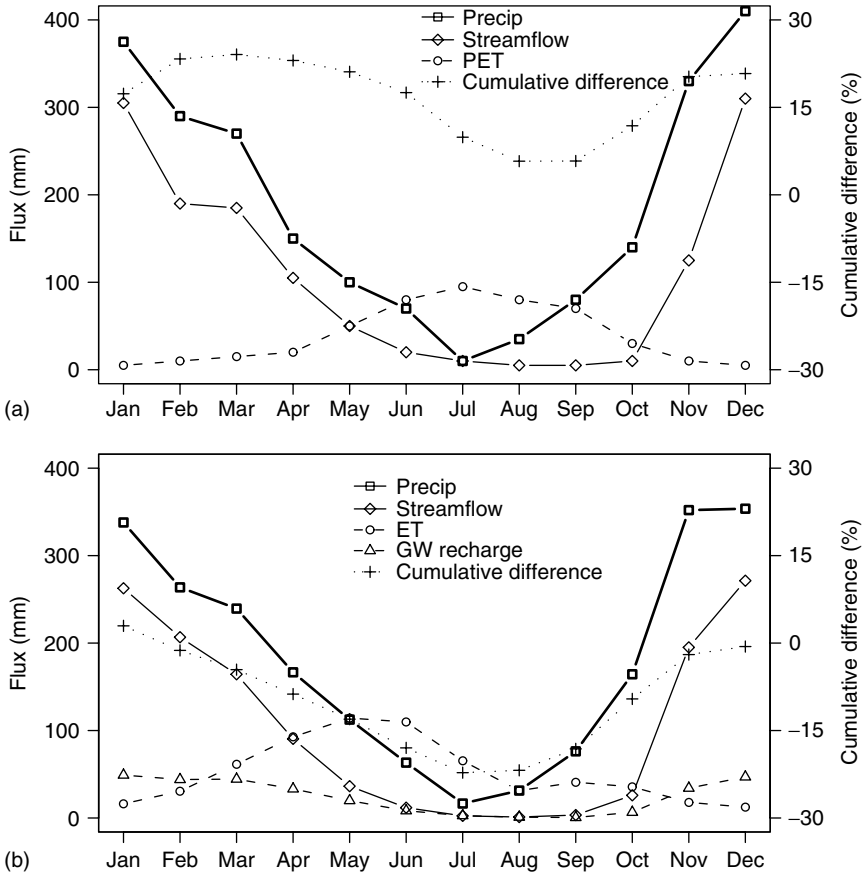

Figure 3. Monthly water balance, WS2. (a) Water balance as reported in Jones (2000), based on WY52–WY96. (b) Water balance from this study, based on WY58–WY98. 'Cumulative difference' is cumulative sum of $(P - Q - ET - G)$, expressed as a percentage of precipitation. Cumulative difference should be zero at end of year assuming no long-term change in storage

are reached in the spring, when soil moisture is greatest, rather than later in the summer when meteorological conditions are more favourable but soil drought is limiting for transpiration and there is little wet canopy evaporation taking place (Figure 3).

Comparing the WS1 scenarios of treatment/no treatment, ET decreased with treatment during all months except August and September, with the largest absolute reductions during the period of maximum transpiration, spring and early summer. The greatest differences in predicted streamflow were during May, June, and November.

All streamflows

The simulated overall 'hydrologic regime' of WS2, quantified using the methods of Post and Jones (2001), matched the real watershed characteristics fairly well (Table IV). The main weakness of the model was underpredicting baseflow and overpredicting quickflow, where the simulated and observed hydrographs were separated into components using the algorithm described in Post and Jones (2001). Streamflow modelling skill was generally highest for WS2 and lowest for WS3 (Table V). Interannual streamflow variation was surprisingly difficult to reproduce, with mean errors for annual streamflow ranging from −8 to +9%, and mean absolute errors ranging from 3 to 9%. Maximum efficiency in simulating hourly streamflow was obtained for WS2 during one of the calibration periods ($E_2 = 0{\cdot}856$; see Appendix for definition of model skill statistics).

Table IV. Hydrologic regime of WS2, as characterized using the methods of Post and Jones (2001). Observed and simulated are based on WY58–WY98. Slope terms are from simple linear regression of annual flow on annual precipitation. CV is coefficient of variation, standard deviation/mean

Flux	Observed	Simulated
Mean annual streamflow (mm year^{-1})	1306	1272
Mean annual baseflow (%)	40	36
Mean annual quickflow (%)	60	64
Slope, streamflow	0·86	0·87
Slope, baseflow	0·33	0·34
Slope, quickflow	0·56	0·53
CV, daily streamflow	1·78	1·90
CV, annual streamflow	0·31	0·30
CV, daily runoff ratio	4·65	5·02
CV, annual runoff ratio	0·15	0·14

Table V. Streamflow modelling skill for time periods[a]

Watershed	Period	MEA (%)	MAEA (%)	E_2	E'_1	d'_1
All years						
WS2	WY58–WY98	−2·3	5·8	0·807	0·506	0·763
WS1	WY58–WY98	−0·2	5·5	0·789	0·516	0·753
WS3	WY58–WY98	1·1	5·2	0·738	0·395	0·736
Calibration						
WS2	WY94–WY98	−2·1	3·9	0·826	0·594	0·803
WS2	WY80–WY83	−3	6·8	0·856	0·537	0·777
WS2	WY58–WY62	−7·9	7·7	0·794	0·476	0·750
WS1	WY58–WY62	9·2	9·4	0·783	0·461	0·730
Meteorology datasets						
WS2	WY80–WY98	0·8	6·0	0·800	0·517	0·770
WS2	WY58–WY79	−4·9	5·7	0·811	0·495	0·756
WS1	WY80–WY98	1·4	5·1	0·804	0·550	0·773
WS1	WY58–WY79	−1·6	5·9	0·778	0·486	0·734
WS3	WY80–WY95	5·7	8·0	0·664	0·339	0·720
WS3	WY58–WY79	−2·3	3·4	0·771	0·423	0·743

[a] MEA: mean error of annual streamflow; MAEA: mean absolute error of annual streamflow; E_2: efficiency; E'_1: first-degree efficiency based on monthly means; d'_1: first-degree modified index of agreement, based on monthly means (see Appendix).

For E_2, a value less than about 0·6 is typically considered a poor fit, whereas a value greater than about 0·8 is considered a very good fit. Many hydrologists modelling streamflow are satisfied with a value of around 0·7 at a daily time step (e.g. Wilcox *et al*., 1990). The drop-off in model skill for other years and other watersheds was not severe except for WS3 during WY80–WY95, which had $E_2 = 0{\cdot}664$. Mean annual streamflow errors were less under the WY58–WY79 meteorology dataset versus the WY80–WY98 dataset, but among hourly efficiency results there was no pattern. The highest goodness-of-fit in a single year was obtained for WS2 in WY87, which was not part of the calibration, with $E_2 = 0{\cdot}920$ (Figure 4). Peaks and winter baseflow were reproduced well, but spring baseflow was underpredicted owing to a lack of shallow groundwater storage in DHSVM. The worst goodness-of-fit occurred with WS3 in WY93, with $E_2 = 0{\cdot}069$. Several WS3 peakflow events were substantially overpredicted, and others were substantially underpredicted. During the high runoff and peakflow years of WY96 and WY97, E_2 for WS1 was 0·831 and 0·833 respectively.

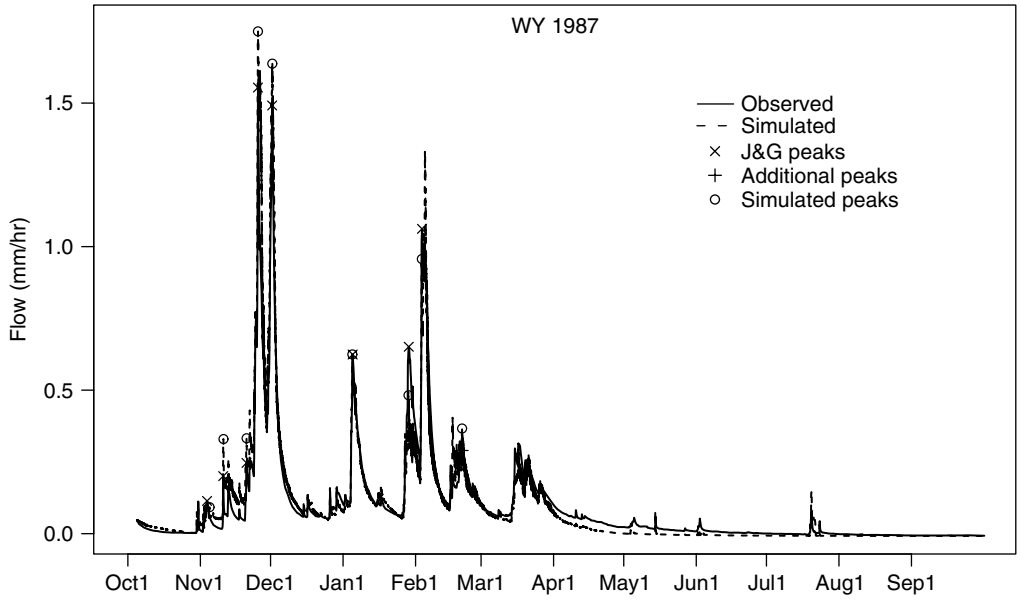

Figure 4. Hydrograph, WS2, WY87. Highest goodness-of-fit among all watershed–water year combinations. $E_2 = 0{\cdot}920$, $E'_1 = 0{\cdot}753$, $d'_1 = 0{\cdot}881$, J&G = Jones and Grant (1996)

The goodness-of-fit obtained at the hourly time scale here compares favourably with the results of the only other published study of modelling the small watersheds over an entire year (Tague and Band, 2001a,b). In that study, the model RHESSys was run at a daily time step and calibrated on WS2 over just 1 year, WY63, and verified on WS3 over a pretreatment year, 1959. The resulting calibration and verification values for E_2 were 0·77 and 0·7 respectively. Our study involved greater challenge for hydrologic modelling: hourly time step, essentially the entire period of record, and transient vegetation. At a daily time step over WY58–WY98, E_2 was 0·825, 0·827, and 0·763 for WS1, 2, 3 respectively.

Peakflows

Calibration was focused on the overall streamflow record, and, not surprisingly, model skill was lower for peakflows and individual categories of flow magnitude (Table VI). The model tended to overpredict the smallest flows and underpredict the largest flows. In general, model skill was higher for WS2 than for WS1. When considering only peakflows, the smallest events were overpredicted by 39% in WS1 but only by 2% in WS2. The largest peakflows were underpredicted by 36% in WS1 and by 27% in WS2. For all other size classes, predicted peak flows were within 15% of observed values in WS2 and within 30% of observed values in WS1.

Some of the shortcomings of the model application with respect to peakflow simulation are demonstrated in the February 1996 flood (Figure 5). During this large rain-on-snow event, described for nearby locations by Marks *et al*. (1998), most of the observed snowpack at the elevation range of WS1, 2, 3 was melted off during a 2 day period, but the simulated decrease in snow water content for this period was only about 10 mm. In addition to the lack of snowmelt, response times to rainfall inputs were too long, indicating that the model did not route water fast enough to the stream channels.

An additional factor inhibiting model skill in the pre-WY80 peakflow events is the assumed uniform hourly distribution of precipitation over a given day. Two groups of top 10 peakflows were determined for WS1, by magnitude of observed flow and by magnitude of the difference between observed and simulated flow. Four events, including the February 1996 flood, were in both groups. Most were rain-on-snow events prior to

Table VI. Streamflow modelling skill for flow magnitude categories[a]

Flow magnitude		Augmented Jones and Grant (1996) events			All hourly flows		
Category	mm h^{-1}	Bias	d_1	n	Bias	d_1	n
WS1							
1	<0·396	1·387	0·200	23	1·130	0·762	$3·2 \times 10^5$
2	0·396–0·756	0·917	0·301	76	0·942	0·355	$2·5 \times 10^4$
3	0·756–1·260	0·837	0·339	91	0·926	0·326	8255
4	1·260–3·290	0·707	0·457	113	0·856	0·493	4370
5	>3·290	0·636	0·269	34	0·659	0·320	538
WS2							
1	<0·396	1·018	0·462	73	0·974	0·769	$3·2 \times 10^5$
2	0·396–0·756	1·021	0·314	102	0·961	0·392	$2·2 \times 10^4$
3	0·756–1·260	0·988	0·451	86	1·042	0·364	8685
4	1·260–3·290	0·857	0·590	66	0·943	0·510	2858
5	>3·290	0·735	0·329	10	0·711	0·340	249

[a] Categories 1–3 are same as small, small-to-medium, and medium-to-large categories respectively in Jones and Grant (1996); categories 4 and 5 are large events in Jones and Grant (1996). Augmented set of Jones and Grant (1996) events includes their events plus some additional peaks from the dataset. Bias = (mean simulated)/(mean observed); d_1: index of agreement based on grand mean ($\overline{O'} = \overline{O}$ in Equation (A.4)); n: sample size.

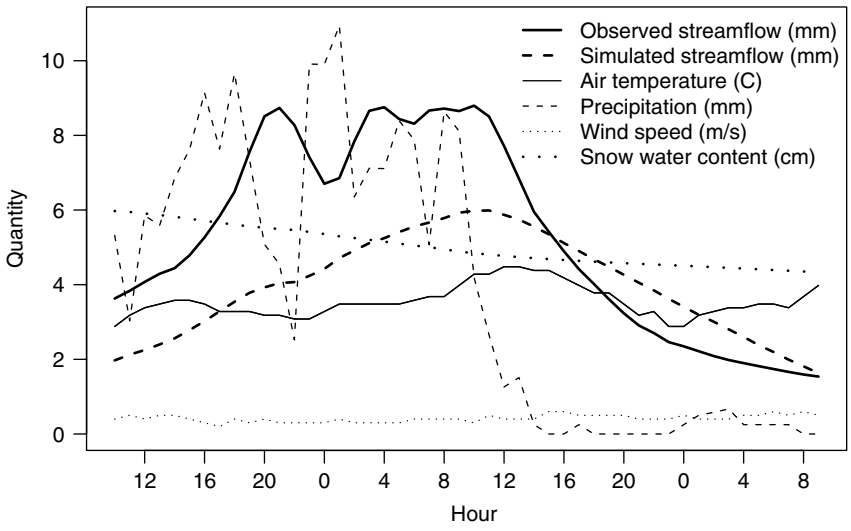

Figure 5. Rain-on-snow flood event, 6–8 February 1996, WS1. Observed peakflow is 2·49 m^3 s^{-1} (8·8 mm h^{-1})

WY80. A too-persistent snowpack was inferred as the primary problem in the simulations of most of them. Uniform rainfall over the day and delayed quickflow response also lowered model skill.

The relative change in streamflow magnitudes between scenarios was also investigated. As expected, the largest streamflow increases in response to treatment were in the low flow range. Mean hourly flow below the 50th percentile increased by 84% in WS1 and 15% in WS3, but mean hourly flow above the 90th percentile increased by only 1·7% in WS1 and not at all in WS3. For the 10 largest peakflows, WS1 decreased by 1·5% under treatment and WS3 increased by 1·1%.

Treatment effects on peakflows

Statistical results involving observed peakflows data (Figure 6) were practically identical to those obtained by Thomas and Megahan (1998), confirming that changing the sample size and using hourly instead of instantaneous peakflows did not significantly change the outcome. Slope and intercept values resulting from simple linear regression of log(WS2) on log(WS1) peakflows were similar, and all recovery periods were significantly different from the pretreatment period. Slopes for periods R1 and R4 were different from the pretreatment period, but slopes for R2 and R3 were not, leading to further tests on their intercepts. These were found to be different from pretreatment.

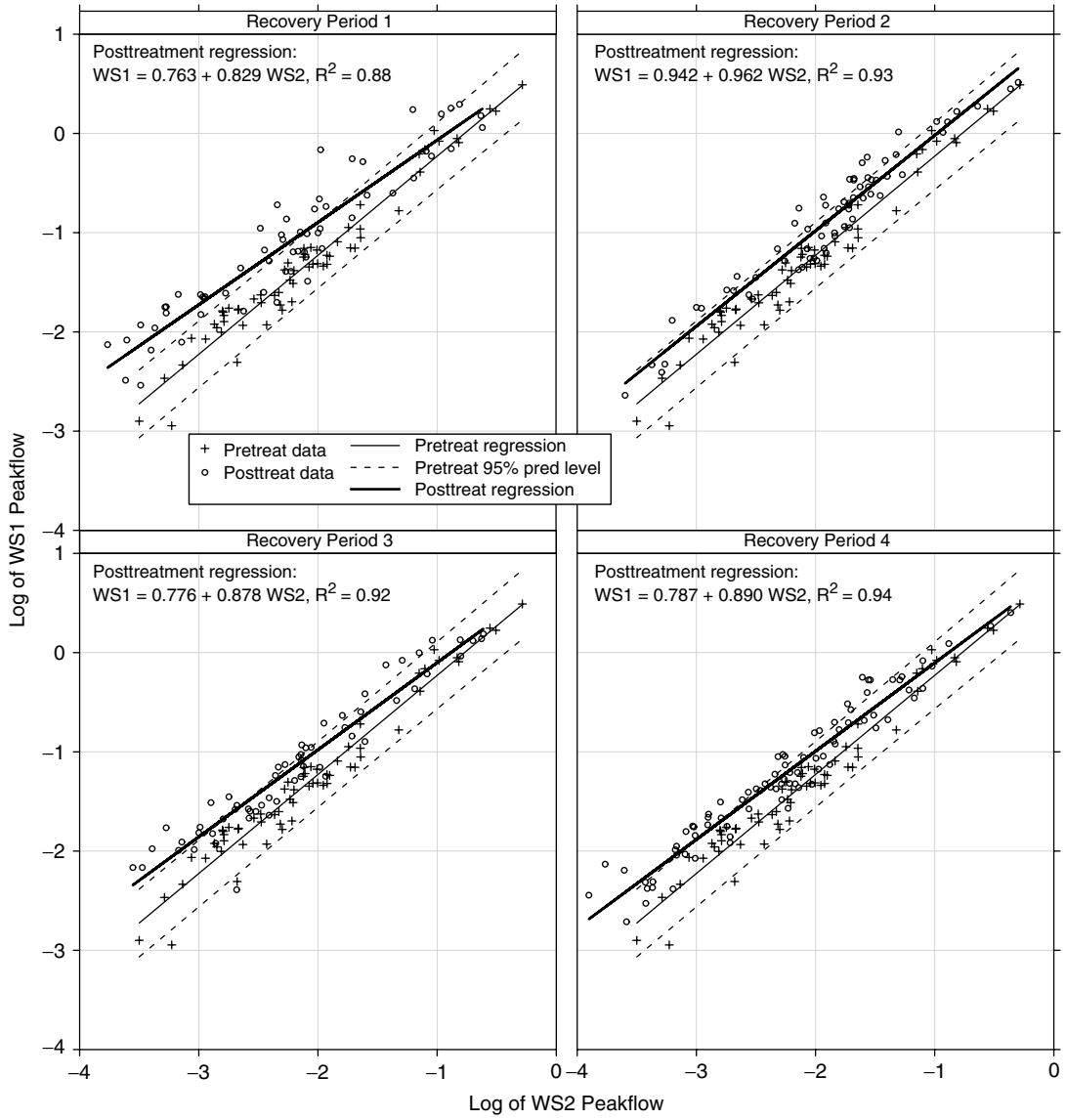

Figure 6. Observed WS1 versus observed WS2, augmented set of Jones and Grant (1996) peaks after WY57. Regression line and 95% individual prediction limits for $\log_e$(WS1) on $\log_e$(WS2) during the four recovery periods. Pretreatment regression was WS1 = 0·769 + 0·998WS2, $R^2 = 0{\cdot}95$, and all regressions were significant ($p \leq 0{\cdot}0001$). Methods after Thomas and Megahan (1998)

The same analysis of simulated peakflows had somewhat different statistical outcomes. There was more scatter in the simulated peakflows and, therefore, prediction intervals were wider (Figure 7). Unlike the observed dataset, all recovery period slopes were found to be not different from pretreatment. However, intercepts were different for periods R1, R3, R4, leaving only period R2 as not different from pretreatment and having a different final interpretation from the observed dataset. The regressions of simulated peakflows were also similar to the observations, with the exception of the highest flows during R1, for having post-treatment lines that did not cross the pretreatment line. Simulations and observations both indicated a positive treatment effect for most periods and flow rates.

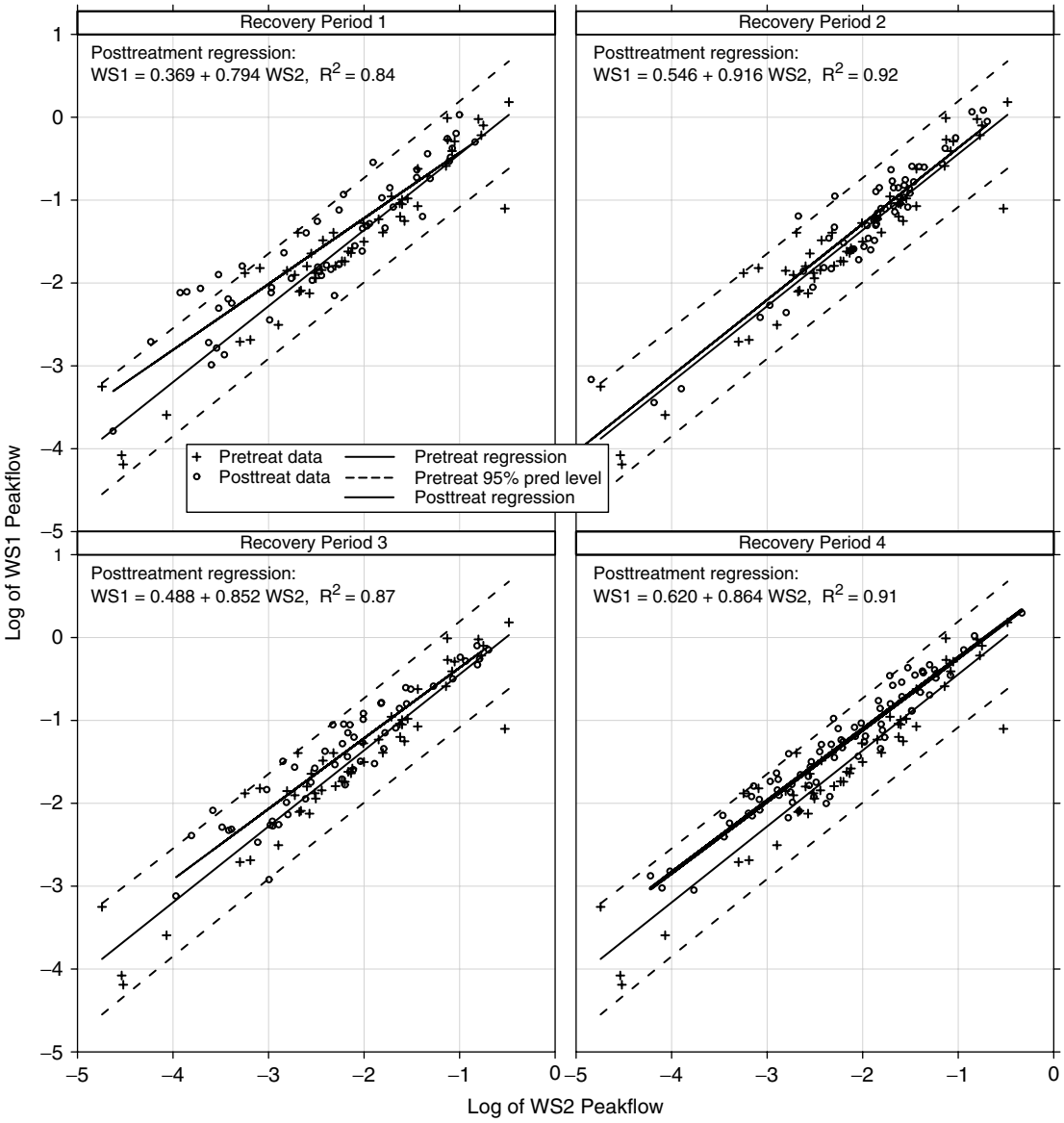

Figure 7. Simulated WS1 versus simulated WS2, augmented Jones and Grant (1996) peaks after WY57. Regression line and 95% individual prediction limits for $\log_e$(WS1) on $\log_e$(WS2) during the four recovery periods. Pretreatment regression was WS1 = 0·473 + 0·917WS2, $R^2 = 0{\cdot}90$, and all regressions were significant ($p \leq 0{\cdot}0001$)

Both observed and simulated peakflows also had an inverse relationship of treatment effect to event size (Figure 8). However, the changes in peakflow due to treatment were mostly lower with the simulations, and the relative ordering of the recovery periods was different. The peakflow change from treatment was also negative at high flows for the simulated period R1, indicating that the model predicts peakflows during the largest events to be greater under old-growth conditions than under a very young forest with low leaf area.

When WS1-treated peakflows were regressed on control WS2 peakflows and time as a continuous variable for all events from November 1966 to September 1979 (Equation (7)), the observed dataset had a negative coefficient for time ($-0{\cdot}0106$, $p = 0{\cdot}035$), confirming that the treatment effect in the real system diminishes with time. The time coefficient from the regression of simulated peakflows was also negative, but not significant ($-0{\cdot}0093$, $p = 0{\cdot}193$).

Comparison of simulated WS1 peakflows from scenario model runs eliminated watershed as a confounding factor and provided additional information about the model's representation of treatment effects. All regressions of treated WS1 on untreated WS1 were significant (Figure 9). All slopes were different from unity and all intercepts were different from zero, indicating that DHSVM output contained treatment effects for all recovery periods. In the plots, the 95% confidence band on the regression line goes below the 1 : 1 line representing no treatment, implying a negative treatment effect at high flow magnitudes. However, this crossing and slope less than unity are partly due to the influence of small flow magnitude points that are well above the 1 : 1 line. The treatment effect clearly declines with time and flow magnitude in all subsets, especially when comparing R1 and R4. In a regression of treated WS1 on untreated WS1 and continuous time, the time term was significant and negative, as with the observed peakflows dataset.

Effect of time period

The time period used to define peakflow datasets and meteorology input to DHSVM also impacted predicted treatment effects. The augmented Jones and Grant (1996) dataset of WS1 versus WS2 did not indicate a monotonic recovery trajectory, as seen in the out-of-order placement of recovery period R2 in Figure 8. The R2 period contained fewer small peakflows than the other recovery periods, and the increase under treatment was less for that period than R3 or R4. The comparison of simulated peakflows showed a similar effect, so we inferred that the climate that produced the selected peakflow events was the source of most of the discrepancy rather than an unknown shift in the watershed land cover or hydrologic response. A similar ordering of recovery period effects is seen in the scenarios based on WS1 alone (Figure 10). Quality of

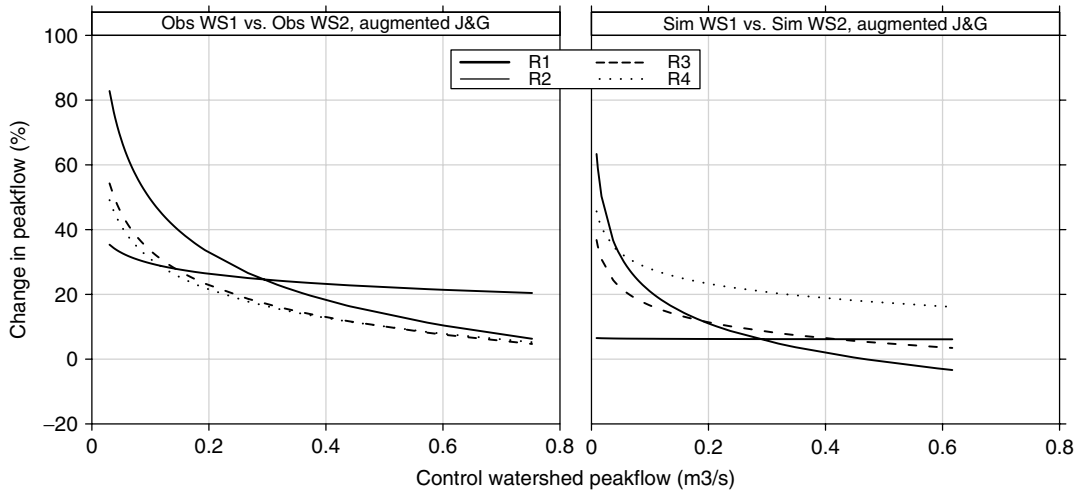

Figure 8. Percentage increase in WS1 peak flow versus WS2 peak flow during all recovery periods. J&G = Jones and Grant (1996)

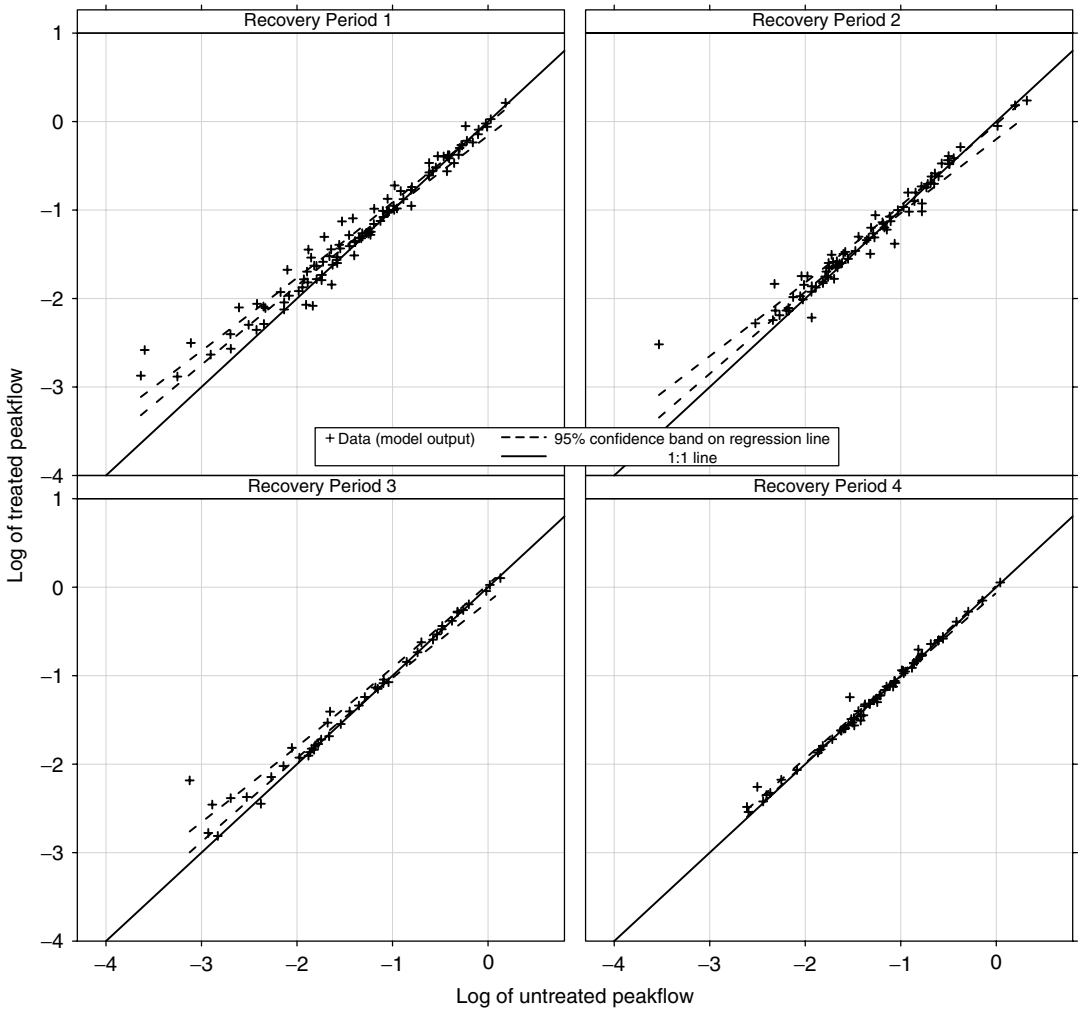

Figure 9. Regression of treated on untreated $\log_e$(peakflows) from scenarios with WS1. The 1 : 1 line corresponds to no treatment. Peakflows correspond to the observed events of the Jones and Grant (1996) dataset from both the WY58–WY79 and WY80–WY98 meteorology periods. Simulated peakflows were assigned to recovery periods as follows. R1: WY58–WY62 and WY80–WY84. R2: WY63–WY67 and WY85–WY89. R3: WY68–WY72 and WY90–WY94. R4: WY73–WY79 and WY95–WY98. Confidence band is for the regression *line*, not individual points

meteorology input to DHSVM probably had a secondary impact on simulation of treatment effects. The hourly meteorology input to DHSVM was disaggregated from daily data for the first period, whereas the meteorology input for the second period was hourly data. The lower quality of meteorology input probably contributed to the less consistent response to the simulated forest treatment in the first meteorology period. Results for simulations on the second meteorology period concur with expectations, showing a systematic decline in treatment effects over the four recovery periods.

DISCUSSION

Applying a physically based model forces the hydrologist to consider all of the major terms in the water balance explicitly. Direct measurements of ET that can guide modelling have only recently become available

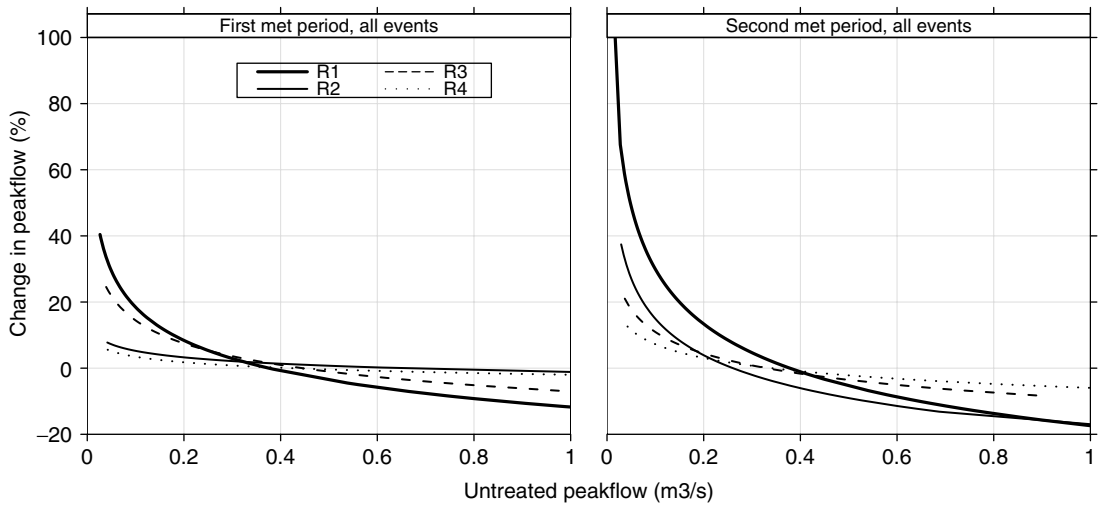

Figure 10. Percentage increase in WS1 treatment scenario peak flow versus untreated peak flow. First meteorology period is WY58–WY79 and second meteorology period is WY80–WY98. To increase the sample size and reduce the confounding effect of variable climate across periods, an algorithm was used to select the peakflow sets for this figure. Total sample size for all recovery periods was increased from 299 (augmented Jones and Grant (1996)) to 898

at HJA, primarily for transpiration via observations of sap flux. More work is needed for evaporation from interception storage and soils. Perhaps new microwave methods may make catchment-scale measurements of total ET possible (Parlange *et al*., 2001). Vegetation regrowth was prescribed simply and independently of local soil, terrain, and microclimate properties. Data on key properties such as LAI are needed for better model input and to guide a dynamic simulation of vegetation. From simulation of ET based on available data, we infer that an additional loss term of groundwater recharge accounts for about 12% of mean annual precipitation, using a conservative estimate of *P*. The monthly water balance suggests that the required additional loss is most active during the winter rainy season. We attribute this additional loss to groundwater flux out of the watersheds. Although we assumed a uniform groundwater recharge conductivity across soil types, the possibility of significant differences is suggested by the varying runoff ratios of the small watersheds during years when landcovers were the same. Rothacher *et al*. (1967) were the first to point out significant differences in runoff on area–depth basis among WS1, 2, 3, and suggested that variable 'deep seepage' might be one cause.

Model calibration focused on the overall streamflow record, and tradeoffs in producing baseflow and quickflow resulted in a positive bias for low flows and negative bias for high flows. To improve baseflow simulation, we are currently developing a new approach for groundwater movement that will offer a deeper flowpath while still maintaining a reasonable number of parameters. To improve peakflow simulation, the primary need is to improve the parameterization of snow functions and routing of 'quickflow'.

The underprediction of peakflows, especially during rain-on-snow events, led us to evaluate the simulated snow hydrology and its relationship to canopy characteristics more closely. To clarify the effects of canopy growth (via increasing LAI and height) on simulated hydrology during rain-on-snow conditions, DHSVM was run with six scenarios of constant vegetation properties representing canopy ages of 1, 5, 10, 20, 40 years and old growth. A single grid cell at elevation 719 m (close to the average WS1 elevation of 716 m) was simulated for the period 1 October 1995–2 October 1996 with each canopy age scenario, and the states and fluxes around the time of the major rain-on-snow event in 1996 were compared (Table VII). As LAI increased, shortwave radiation reaching the ground (or snow) surface decreased, and interception capacity increased. Increasing height and LAI led to greater decay of windspeed through the canopy and, therefore, less aerodynamic conductance to the surface. The decrease in both shortwave radiation and turbulent transfer

Table VII. Comparison of simulated energy and mass variables under canopies of different forest ages. Values in top part of table are means during a period that included snowpack build-up followed by a large rain-on-snow event, 23 Jan to 10 Feb 1996. Values in bottom part of table correspond to the rain-on-snow event, 4 Feb to 10 Feb[a]

	1 year	5 years	10 years	20 years	40 years	Old growth
Canopy LAI	0·53	2·4	4·1	6·2	7·9	8·5
Canopy height (m)	3·9	14	24	38	52	60
Mean values, 23 Jan to 10 Feb 1996						
Shortwave radiation to LL (W m^{-2})	58·1	36·9	24·1	14·2	9·3	7·9
Wind speed, LL (m s^{-1})	0·13	0·08	0·07	0·06	0·06	0·06
Aerodynamic conductance, LL (mm s^{-1})	1·53	1·01	0·90	0·82	0·78	0·76
Intercepted SWE (mm)	1·3	8·3	10·2	10·9	11·4	11·6
Intercepted rain (mm)	0·1	0·6	0·9	1·3	1·7	1·8
Snowpack SWE (mm)	0·2	5·4	18·6	28·4	33·3	35·1
Water flow to ground (mm h^{-1})	1·05	1·04	1·04	1·04	1·03	1·02
Other values, 4 Feb to 10 Feb						
Mean water flow to ground (mm h^{-1})	1·83	1·85	2·04	2·13	2·12	2·11
Initial intercepted SWE (mm)	1·9	5·6	7·5	8·8	8·8	8·8
Final intercepted SWE (mm)	0	0	0	0	0	0
Initial snowpack SWE (mm)	0·4	0·5	30·6	45·2	50·3	52·2
Final snowpack SWE (mm)	0	0	0	0	5·2	8·3

[a] LL: lower layer (snowpack or understory vegetation); SWE: snow water equivalent. Water flow to ground surface includes direct precipitation and snowmelt. WS1 canopy age was 30 years in 1996.

resulted in more persistent snowpacks for older canopies. Even though older canopy intercepted a greater proportion of snowfall that could then melt more rapidly than the ground snowpack, the more important effect of older canopy on snow hydrology was the greater sheltering of the ground snowpack from melt. This resulted in older canopy having more snowpack available for melting at the start of the rain-on-snow event. Total water flux to the ground surface (from melting snow and rainfall) was similar for all canopy ages, but snowmelt comprised a much larger proportion of this during the peakflow period for the older age classes. All age classes experienced complete melting of snow intercepted in the canopy, but the 40 year and old-growth scenarios did not have complete melting of the snowpack on the ground.

These results indicate reasonable differences between age class results, even though snowmelt occurred too slowly overall. Considering all the peakflows generated from rain-on-snow events during the simulated history, snowpack was generally too persistent for both mature and young canopies and there was a lack of sensible and latent heat transfer to drive melting. Here again we found tradeoffs, in trying to produce enough aerodynamic conductance for adequate turbulent transfer to the snowpack during winter without driving too much ET from the canopy at other times. Mean windspeed and aerodynamic conductance G_A at ground level both decreased with canopy age, but for all ages were quite low, with G_A ranging from 1·5 to 0·8 mm s^{-1} during the period of snow accumulation followed by rapid depletion during the rain-on-snow storm (23 January to 10 February). Mean observed windspeed at the PRIMET meteorological station during the time of rapid melting (4–6 February) was only 0·42 m s^{-1} (Figure 5) even though this storm was very windy over a large area of Oregon and Washington. If windspeeds were, in fact, systematically greater than the data indicate and the model input reflected that, resulting in higher G_A, then other aspects of evapotranspiration parameterization would require downward adjustment to keep ET fluxes in the appropriate ranges.

At the basin scale, these differences caused by canopy age resulted in some simulated treated peakflows being less than the untreated ones, mostly at medium-to-large flows. Greater simulated snowmelt during rain-on-snow events in the old-growth (untreated) case was caused primarily by (1) a larger area having snowpack prior to the storm and (2) more intercepted snow in the canopy available for rapid melt. Both

treatment and no-treatment cases had a ripe snowpack in these events, so throughfall falling on the ground snowpack and snowmelt flowed directly to the soil surface. Because the simulated rate of snowmelt was mostly independent of the snowpack water content, the key difference between canopy states over the short time scales of peak runoff generation was the extent rather than thickness of the snowpack—there was a larger area with snowpack in the old-growth case than in the treated case. In contrast to this model behaviour, Harr and McCorison (1979) found in an HJA plot study that a clearcut tends to have more snowpack than an old-growth plot, and that the snowpack retards water movement in rain-on-snow events, giving the same result of treated peakflow being less than untreated peakflow, but for different reasons. This conflict suggests a need for more understanding and further model development of terrain- and canopy-dependent snow accumulation and melt in forests.

Simulation of 'quickflow' in DHSVM also needs to be improved. Runoff generation at the stream channel needs to occur more rapidly in response to rainfall and snowmelt on the hillslope. DHSVM's current grid-based design is a weakness from the standpoint of representing the hillslope as a continuum that can translate excess water rapidly downhill, but future development will seek to overcome this limitation and represent quickflow more effectively.

Although simulated streamflow error was significant and increased with flow rate, the overall statistical conclusions based on model output were similar to those based on observations. Peakflow increases had an inverse trend with flow rate, similar to the findings of Thomas and Megahan (1998), but some negative treatment effects (peakflow decreases with treatment) were predicted at the highest flow rates. Simulated treatment effects were somewhat smaller than previous empirical findings, but were statistically detectable and tended to decline with time. The scenario approach based on using one watershed under two different land-cover trajectories yielded the clearest signal treatment effects. Simulating both control and treated watersheds over a long time period and comparing their flows directly resulted in a weaker but still reasonable comparison to the paired watershed data and previous studies based on it.

This study addressed the impacts of canopy removal and did not test DHSVM on observed treatment effects associated with forest roads. The treatment effect identified by Thomas and Megahan (1998) in WS3 was roughly proportional to the extent of harvest in that watershed, and no clear signal in the basin streamflow due to roads alone was identified. More retrospective testing of DHSVM in watersheds having only roads as the treatment for an adequate period of time would be desirable.

Finally, the results underscored the practical difficulty of defining peakflow impacts in watershed streamflow data even when a strong treatment such as 100% clear-cutting has been applied. The trajectory of observed peakflow response as defined for the recovery periods R1–R4 was not monotonic, and the simulations supported the inference that variable climate, expressed through meteorology conditions that drive storms, was to blame rather than some unknown change in land cover or hydrologic process. Climate was also more important than quality of hourly meteorology input in causing a similar response in the model scenarios.

CONCLUSIONS

We identified and addressed a fundamental problem in previously published water balances for the HJA small watersheds. On a mean annual basis, Q and ET do not explain about 12% of precipitation, using a conservative estimate of P. This portion of the water balance was attributed to groundwater recharge, and modelling indicates it is most active during the winter wet season.

Overall, the model output agreed well with the observed overall 'hydrologic regime' and streamflows at hourly and annual time scales. High efficiency ($E_2 > 0{\cdot}7$) and reasonably low bias in streamflow modelling were achieved at an hourly time step over almost the entire period of record at HJA, and compared favourably to previous simulation efforts. However, the model overpredicted low flows and underpredicted high flows. Inadequate storage and release groundwater as baseflow was responsible for the low flows problem. Inadequate snowmelt generation and quickflow translation down the hillslope were responsible for the high flows problem.

Although too-slow runoff generation led to attenuated peakflows in the simulations and degraded statistical results concerning treatment effects, the overall sensitivity of the model to observed forest treatments was confirmed. The simulations also highlighted the difficulty of predicting treatment effects in different climates, and supported the inference that climate rather than watershed change was responsible for anomalies in recovery paths.

ACKNOWLEDGEMENTS

Financial support was provided by the National Council for Air and Stream Improvement, Battelle-Pacific Northwest Division, and Pacific Northwest National Laboratory. We thank Don Henshaw/OSU for his assistance in obtaining climate and streamflow data, and Robert B. Thomas for his assistance with implementing the statistical methods used on peakflows. Data sets were provided by the Forest Science Data Bank, a partnership between the Department of Forest Science, Oregon State University, and the US Forest Service Pacific Northwest Research Station, Corvallis, Oregon.

APPENDIX: STATISTICS FOR EVALUATING GOODNESS-OF-FIT

The Nash and Sutcliffe (1970) efficiency E_2 casts the mean of the observations as a benchmark for the model:

$$E_2 = 1{\cdot}0 - \frac{\sum_{i=1}^{N}(O_i - P_i)^2}{\sum_{i=1}^{N}(O_i - \overline{O})^2}. \tag{A.1}$$

where N is the number of time steps, O_i is the observed value at time step i, P_i is the predicted value at time step i, and $\overline{O}$ is the mean of the observations. Values of E_2 are always less than R^2.

Two first-degree goodness-of-fit measures from Legates and McCabe (1999) use absolute values of differences instead of squares. A further discrimination can be made by using a baseline mean that involves seasonal or other categorical variation inherent in the data. Here, the baseline mean for streamflow was defined as the mean for each month of the year, where the mean is taken across all years in the simulation period. Avoidance of squaring and use of baseline mean instead of the grand mean provides more stringent tests of model skill. The baseline-adjusted, first-degree efficiency is

$$E_1' = 1{\cdot}0 - \frac{\sum_{i=1}^{N}|O_i - P_i|}{\sum_{i=1}^{N}|O_i - \overline{O'}|} \tag{A.2}$$

where $\overline{O'}$ is the baseline mean of the observations. All of the above measures of efficiency have a possible range of $-\infty$ to 1·0. When efficiency is zero, the model is no better or worse than the observed mean as a predictor. The closer the baseline mean is to the individual observations, the lower the efficiency is likely to be. The objective for calibration was to maximize the combination score S:

$$S = (1 - |1 - B|) + E_1' \tag{A.3}$$

where B is bias, defined as the mean ratio of hourly simulated to observed streamflow.

The baseline-adjusted modified index of agreement

$$d_1' = 1{\cdot}0 - \frac{\sum_{i=1}^{N} |O_i - P_i|}{\sum_{i=1}^{N} (|P_i - \overline{O'}| + |O_i - \overline{O'}|)} \tag{A.4}$$

has the advantage of having the same range as the more familiar R^2, 0 to 1·0.

REFERENCES

Anderson SP, Dietrich WE, Montgomery DR, Torres R, Conrad ME, Loague K. 1997. Subsurface flow paths in a steep, unchanneled catchment. *Water Resources Research* **33**(12): 2653–2673.

Bowling LC, Storck P, Lettenmaier DP. 2000. Hydrologic effects of logging in western Washington, United States. *Water Resources Research* **36**(11): 3223–3240.

Bredensteiner KC. 1998. *An investigation of vegetation–hydrology interactions in watershed 1 at the H.J. Andrews Experimental Forest*. Master's thesis, Oregon State University.

Daly C, Neilson RP, Phillips DL. 1996. A statistical–topographic model for mapping climatological precipitation over mountainous terrain. *Journal of Applied Meteorology* **33**: 140–158.

Duan J. 1996. *A coupled hydrologic–geomorphic model for evaluating effects of vegetation change of watersheds*. PhD thesis, Oregon State University.

Dyrness CT. 1969. *Hydrologic properties of soils on three small watersheds in the western Cascades of Oregon*. Pacific Northwest Research Note, Forest and Range Experiment Station, US Department of Agriculture, Portland, OR.

Harmon ME, Sexton J. 1995. Water balance of conifer logs in early stages of decomposition. *Plant and Soil* **172**: 141–152.

Harr RD, McCorison FM. 1979. Initial effects of clearcut logging on size and timing of peak flows in a small watershed in western Oregon. *Water Resources Research* **15**(1): 90–94.

Henshaw DL, Bierlmaier FA, Hammond HE. 1998. The H.J. Andrews climatological field measurement program. In *Data and information management in the ecological sciences: a resource guide*, Michener WK, Porter JH, Stafford SG (eds). LTER Network Office: University of Mexico; 117–122. http://www.ecoinformatics.org/guide/henshaw1.fv2.htm [22 February 2005].

Jones JA. 2000. Hydrologic processes and peak discharge response to forest removal, regrowth, and roads in 10 small experimental basins, western Cascades, Oregon. *Water Resources Research* **36**(9): 2621–2642.

Jones JA, Grant GE. 1996. Peak flow responses to clear-cutting and roads in small and large basins, western Cascades, Oregon. *Water Resources Research* **32**(4): 959–974.

LaMarche JL, Lettenmaier DP. 2001. Effects of forest roads on flood flows in the Deschutes River, Washington. *Earth Surface Processes and Landforms* **26**: 115–134.

Legates DR, McCabe GJ. 1999. Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. *Water Resources Research* **35**(1): 233–241.

Link TE, Flerchinger GN, Unsworth MH, Marks D. 2004. Simulation of water and energy fluxes in an old-growth seasonal temperate rainforest using the simultaneous heat and water (SHAW) model. *Journal of Hydrometeorology* **5**(3): 443–457.

Marks D, Kimball J, Tingey D, Link T. 1998. The sensitivity of snowmelt processes to climate conditions and forest cover during rain-on-snow: a case study of the 1996 Pacific Northwest flood. *Hydrological Processes* **12**: 1569–1587.

Moore GW, Bond BJ, Jones JA, Phillips N, Meinzer FC. 2004. Structural and compositional controls on transpiration in 40- and 450-year-old riparian forests in Western Oregon, USA. *Tree Physiology* **24**: 481–491.

Nash JE, Sutcliffe JV. 1970. River flow forecasting through conceptual models, Part 1—a discussion of principles. *Journal of Hydrology* **10**: 282–290.

Nijssen B, Haddeland I, Lettenmaier D. 1997. Point evaluation of a surface hydrology model for BOREAS. *Journal of Geophysical Research* **102**(29): 367–378.

Parlange MB, Pahlow M, Bou-Zeid E, Kumar V. 2001. Land–atmosphere coupling over complex hilly terrain: case studies. In *Chapman Conference on State-of-the-Art in Hillslope Hydrology*, Band L McDonnell J (eds). Sunriver, OR. American Geophysical Union.

Post DA, Jones JA. 2001. Hydrologic regimes of forested, mountainous, headwater basins in New Hampshire, North Carolina, Oregon, and Puerto Rico. *Advances in Water Resources* **24**: 1195–1210.

Richards FJ. 1959. A flexible growth function for empirical use. *Journal of Experimental Biology* **10**: 290–300.

Rothacher J. 1965. Streamflow from small watersheds on the western slope of the Cascade Range of Oregon. *Water Resources Research* **1**(1): 125–134.

Rothacher J. 1970. Increases in water yield following clear-cut logging in the Pacific Northwest. *Water Resources Research* **6**(2): 653–658.

Rothacher J, Dyrness CT, Fredriksen RL. 1967. *Hydrologic and related characteristics of three small watersheds in the Oregon Cascades*. Pacific Northwest Forest and Range Experiment Station, Forest Service, US Department of Agriculture, Corvallis, OR.

Sidle RC, Tsuboyama Y, Noguchi S, Hosoda I, Fujieda M, Shimizu T. 2000. Stormflow generation in steep forested headwaters: a linked hydrogeomorphic paradigm. *Hydrological Processes* **14**: 369–385.

Storck PA, Bowling L, Wetherbee P, Lettenmaier DP. 1998. Application of a GIS-based distributed hydrologic model for the prediction of forest harvest effects of peak flow in the Pacific Northwest. *Hydrological Processes* **12**: 889–904.

Tague CL, Band LE. 2001a. Evaluating explicit and implicit routing for watershed hydro-ecological models of forest hydrology at the small catchment scale. *Hydrological Processes* **15**: 1415–1439.

Tague CL, Band LE. 2001b. Simulating the impact of road construction and forest harvesting on hydrologic response. *Earth Surface Processes and Landforms* **26**: 135–151.

Thomas RB, Megahan WF. 1998. Peak flow responses to clear-cutting and roads in small and large basins, western Cascades, Oregon: a second opinion. *Water Resources Research* **34**(12): 3393–3403.

VanShaar JR, Haddeland I, Lettenmaier DP. 2002. Effects of land-cover changes on the hydrological response of interior Columbia River basin forested catchments. *Hydrological Processes* **16**: 2499–2520.

Waichler SR, Wigmosta MS. 2003. Development of hourly meteorological values from daily data and significance to hydrological modeling at H.J. Andrews Experimental Forest. *Journal of Hydrometeorology* **4**(2): 251–263.

Waichler SR, Wigmosta MS, Wemple BC. 2002. *Modeling the effects of forest treatments on streamflow at the H.J. Andrews Experimental Forest*. PNWD-3180, Battelle-Pacific Northwest Division, Richland, WA.

Wemple BC. 1998. *Investigations of runoff production and sedimentation on forest roads*. PhD thesis, Oregon State University.

Wigmosta MS, Vail LW, Lettenmaier DP. 1994. A distributed hydrology–vegetation model for complex terrain. *Water Resources Research* **30**(6): 1665–1679.

Wigmosta MS, Nijssen B, Storck P. 2002. The distributed hydrology soil vegetation model. *Mathematical Models of Small Watershed Hydrology and Applications*, Singh VP Frevert D (eds) Water Resources Publications LLC: 7–42.

Wilcox BP, Rawls WJ, Brakensiek DL, Wight JR. 1990. Predicting runoff from rangeland catchments: a comparison of two models. *Water Resources Research* **26**(10): 2401–2410.

