Received: 6 February 2020 | Accepted: 30 April 2020
DOI: 10.1002/hyp.13802

RESEARCH ARTICLE

WILEY

Forests and floods: Using field evidence to reconcile analysis methods

James C. Bathurst[1] | **Barry Fahey**[2] | **Andrés Iroumé**[3] | **Julia Jones**[4]

[1]School of Engineering, Cassie Building, Newcastle University, Newcastle upon Tyne, UK

[2]Manaaki Whenua–Landcare Research, Lincoln, New Zealand

[3]Facultad de Ciencias Forestales y Recursos Naturales, Universidad Austral de Chile, Valdivia, Chile

[4]College of Earth, Ocean, Atmospheric Science, Oregon State University, Corvallis, Oregon

Correspondence
James C. Bathurst, School of Engineering, Cassie Building, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK.
Email: james.bathurst@ncl.ac.uk

Abstract

The extent to which forests, relative to shorter vegetation, mitigate flood peak discharges remains controversial and relatively poorly researched, with only a few significant field studies. Considering the effect purely of change of vegetation cover, peak flow magnitude comparisons for paired catchments have suggested that forests do not mitigate large floods, whereas flood frequency comparisons have shown that forests mitigate frequencies over all magnitudes of flood. This study investigates the apparent inconsistency using field-based evidence from four contrasting field programmes at scales of 0.34–3.1 km^2. Repeated patterns are identified that provide strong evidence of real effects with physical explanations. Magnitude and frequency comparisons are both relevant to the impact of forests on peak discharges but address different questions. Both can show a convergence of response between forested and grassland/logged states at the highest recorded flows but the associated return periods may be quite variable and are subject to estimation uncertainty. For low to moderate events, the forested catchments have a lower peak magnitude for a given frequency than the grassland/logged catchments. Depending on antecedent soil saturation, a given storm may nevertheless generate peak discharges of the same magnitude for both catchment states but these peaks will have different return periods. The effect purely of change in vegetation cover may be modified by additional forestry interventions, such as road networks and drainage ditches which, by effectively increasing the drainage density, may increase peak flows for all event magnitudes. For all the sites, forest cover substantially reduces annual runoff.

KEYWORDS

catchment interventions, field evidence, flood frequency, flood magnitude, flood mitigation, forest hydrology

1 | INTRODUCTION

Despite global concern, the extent to which forests, relative to shorter vegetation, mitigate flood peak discharges remains controversial and relatively poorly researched. Resolution of this controversy is required, not only to satisfy scientific curiosity but also for obvious practical purposes such as assessing afforestation programmes to mitigate flooding, designing riparian infrastructure and investigating flood insurance claims. Additionally, there is a need to moderate the uncritical view among the public, governments and development agencies worldwide that forests prevent floods (e.g., Cambrian Wildwood, 2015; Confor, 2018). The main factors behind the controversy are disagreement over the means of quantifying the flood impact of forests and a relative lack of observational data, especially for large floods and for distinguishing between the effect of forests purely as a vegetation cover and the effects of management interventions, such as logging patterns and road networks. This paper addresses these points as follows.

1 Comparisons of peak flow magnitudes from paired catchments with and without forest cover for the same rainfall event (the so-called equal meteorology or chronological pairing method) suggest that forests (purely as a vegetation cover) can mitigate peak discharges for small to moderate floods but not extreme floods (e.g., Bathurst et al., 2011, 2011; Beschta, Pyles, Skaugset, & Surfleet, 2000; Thomas & Megahan, 1998). The higher annual evapotranspiration of forested catchments creates, on average, larger soil moisture deficits and therefore, during a storm event, the soil absorbs more of the rainfall that would otherwise contribute to flood runoff. For large events, this buffering effect is overwhelmed, and peak discharges are little affected by vegetation cover (e.g., Soulsby, Dick, Scheliga, & Tetzlaff, 2017). However, Alila and associates (Alila, Kuraś, Schnorbus, & Hudson, 2009; Green & Alila, 2012; Kuraś, Alila, & Weiler, 2012) have criticized chronological pairing for its irrelevance to flood risk and because the dependency of the storm response on individual catchment characteristics and conditions such as antecedent soil moisture content renders the method inexact: there may be very little mitigating effect if the forested catchment is already saturated, even for low to moderate events. They propose instead the comparison of flood frequency curves (so-called frequency pairing). Frequency pairing automatically incorporates the effect of varying soil moisture and other catchment conditions because it considers the full hierarchy of flood peaks derived as catchment conditions and any other influences vary over time. Alila and associates show on this basis that, in contradiction to the above, forest logging significantly increases the frequency and magnitude of peak discharges relative to the unlogged state and this effect increases with increasing peak magnitude. Therefore, and for the first time, this paper tests the hypothesis that the two, apparently conflicting, approaches can be reconciled and that they are both of value.

2 The value of the methods turns on the questions which they address. Green and Alila (2012) propose the question: what is the change in magnitude (frequency) for an event of a specific frequency (magnitude) of interest? Clearly this is addressed by frequency pairing. Flood frequency curves are a standard means of characterizing a catchment response and link flood magnitude to flood frequency. Frequency pairing therefore addresses public concerns that loss of forest cover increases the frequency of floods. It is also very relevant to engineering projects which design for an event with a specific exceedance frequency, such as the 100-year flood. Similarly, channel stability, as defined by channel geometry, is understood to vary with the flood frequency regime. However, there is another question: would the peak discharge have been as big if the forest cover had not been removed? This cannot be dismissed as irrelevant, because it is one that is asked by affected citizens (e.g., "But many whose properties were pulverized asked if clear-cutting boosted the magnitude of the flood." [Fowler, 2018]), who might for example wish to lodge a claim for damages against a forest company. There is a general public belief that removal of forest cover increases discharge peaks (e.g., Cambrian Wildwood, 2015; Confor, 2018). Affected citizens may not therefore be so interested in (or may not understand) details of flood frequency; they are more likely to be concerned about the magnitude of a specific event. It is the responsibility of the hydrologist, therefore, at least to examine the question and to determine how well it can be answered. Here, frequency pairing may be less relevant because it provides an overall long-term characterisation of flood response and does not comment on individual storms. By contrast, chronological pairing has nothing to say about frequency or risk but does allow comparisons on a specific storm-by-storm basis. An important aim of the paper is therefore to determine exactly what information can be derived from chronological pairing and whether this is useful in answering the question of whether a particular peak discharge would have occurred, or would have been as big, if a forest cover had been in place.

3 Stratford et al. (2017) find that the strongest support for a peak flow mitigation effect from forest cover comes from modelling studies and that the results of field studies are more conflicted. Carrick et al. (2018) similarly note an increasing reliance on modelling studies and a lack of direct field evidence. The few field studies that do exist concentrate mostly on the Pacific Northwest of North America (e.g., Alila et al., 2009; Beschta et al., 2000; Green & Alila, 2012; Jones & Grant, 1996; Kuraś et al., 2012; Thomas & Megahan, 1998). Chronological pairing and frequency pairing are therefore applied to four research catchments from across the world, providing new field evidence and greatly expanding on the previous geographically limited studies.

4 One impediment to field studies is the rarity of the larger flood peaks, combined with the lack of stationarity in conditions (e.g., in vegetation cover or climate) over the long periods between occurrences of large floods (e.g., Jones & Grant, 2001; Yu & Alila, 2019). This makes it difficult to assemble a flood peak series for an individual catchment that both exhibits stationary conditions and is long enough for statistically robust analysis of the larger flood peaks. This study therefore follows Lewis, Reid, and Thomas (2010) and Green and Alila (2012) in noting apparent trends in the field data, regardless of statistical significance, and conducting metastudies to investigate whether such trends have been measured repeatedly, although individually appearing to be statistically insignificant. The aim is to combine the records of the four catchments to determine if forest impacts on the few largest peak discharges in the records show similar behaviour, thus overcoming the limitations of record length. This goes beyond simply increasing the sample size of rare events for statistical analysis by pooling samples from multiple catchments. A repeated pattern across the catchments, explainable by physical reasoning, would provide strong evidence of a real effect, irrespective of statistical significance.

5 Stratford et al. (2017) note the need for more investigation of contextual factors, including the impacts of forest management practices (such as drainage ditching and road networks), compared with the impact of purely forest cover itself. Road and ditch networks may act similarly by, in effect, extending the stream network and increasing drainage efficiency. Both practices have thus been found to increase flood peaks (e.g., Jones & Grant, 1996; Robinson, 1998). La Marche and Lettenmaier (2001) suggest that road impacts increase with flood return period, while vegetation cover impacts decrease. Previous studies of the flood impact of different forest management practices have been carried out on a case-by-case basis. By combining catchments with a range of practices, this study offers a more integrated view distinguishing the effect of forest cover on its own (apparent from the repeated patterns above) from the effects of the individual practices (apparent from distinctive deviations from these patterns).

Overall, the paper uses a new high-quality data set to make a first attempt at reconciling the analysis methods and, in so doing, presents a new conceptual model of the impact of forests and forest management interventions on peak discharge magnitude and frequency distributions. The emphasis is on floods driven by rainfall rather than snowmelt.

2 | FIELD SITES

The sites, from the four corners of the Earth, have been the subject of long-running research programmes on the impacts of both afforestation and logging and represent a range of forest management practices (Figure 1, Table 1). The extensive data availability in each case is complemented by the authors' detailed knowledge of the sites.

2.1 | Wark Forest, UK

The site lies in the headwaters of the River Irthing in northwest England, northeast of Carlisle. For the period 2003–2012, it paired the largely forested (Sitka Spruce plantation) Coalburn catchment with the largely peat grassland Flothers catchment. Coalburn, on its own, is the UK's longest-running forest research catchment (since 1967). Precipitation was measured by storage gauge and tipping bucket raingauge at the outlet of the Coalburn catchment. Discharge was measured by a two-stage weir for the Coalburn catchment and by stage records and a rating curve based on spot discharge gaugings by current meter for the Flothers catchment. At the time of the study period, there was only one minor track in each catchment. The principal management feature is the drainage ditch network that was cut in the Coalburn catchment before plantation in 1972. In its early years, the network increased both the annual runoff and the storm peak discharges compared with the pre-existing grassland condition. Subsequently, the ditches have partly filled with debris but a small part of the network still appears to affect storm flow response. Further details are given by Archer and Newson (2002), Bathurst et al. (2018), Birkinshaw, Bathurst, and Robinson (2014) and Robinson (1998).

2.2 | The Glendhu experimental catchment study, New Zealand

The site lies in the headwaters of the Waipori River in South Island, New Zealand, west of Dunedin. Established in 1979, it is New Zealand's longest-running catchment study. It pairs the partially forested (Monterey Pine [*Pinus radiata*] plantation) GH2 catchment with the tussock grassland GH1catchment. For most of the study

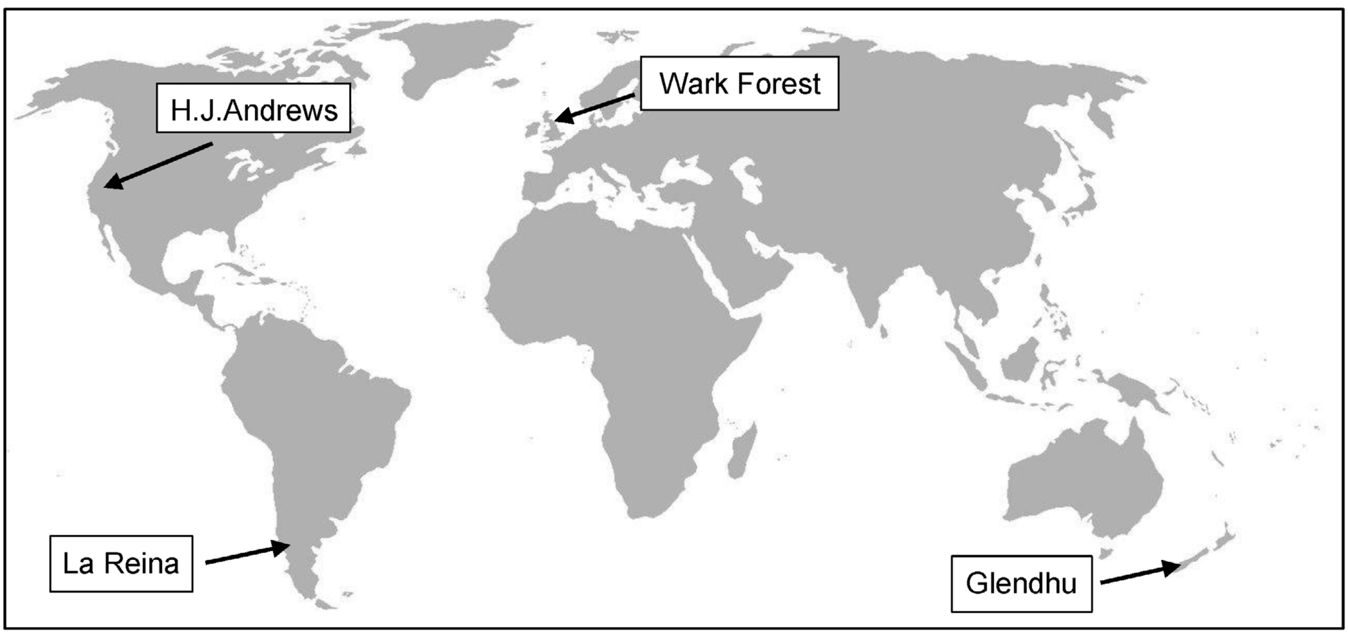

FIGURE 1 Location of the field sites

TABLE 1 Details of the field sites

Site	Latitude	Köppen-Geiger climate classificat-ion[a]	Mean annual precipitat-ion *(relevant period or region in italics)* (mm)	Paired catchments	Area (ha)	Altitude range (m above sea-level)	Vegetation	Management	Data record
Wark Forest, UK	55° north	Cfb	1,672 *(2003–12)*	Coalburn	150	275–330	90% forest, 10% peat grassland	Initially grassland; 1972 ditched; 1973 planted	1967-present
				Flothers	140	275–325	36% forest, 64% peat grassland	1995 forest planted, rough sheep grazing	2003–2012, with gaps
Glendhu, New Zealand	46° south	Cfb	1,330 *(1980–2013)*	GH1	216	460–680	100% tussock grassland		1980–2013
				GH2	310	460–680	75% forest, 25% tussock grassland[b]	1981 contour ripped; 1982 planted; 1991 canopy closure; 2014 logged	1980–2013
H.J. Andrews, Oregon, USA	44° north	Csb	2,300–2,500 *(western Cascades)*	WS1	96	439–1,027	100% logged and reforested	1962–66 logged	1955–1988
				WS2	60	545–1,079	100% forest		1955–1988
				WS3	101	476–1,080	25% logged and reforested	1959 roads; 1963 logged	1955–1988
La Reina, Chile	40° south	Cfb	2,576 *(1997–2018)*	—	34.4	35–225	79.4% forest, 20.6% roads and riparian vegetation	1977 forest planted; 1999/2000 logged; 2000 reforested; 2017/18 logged	1997–2018, with gaps

[a]Peel, Finlayson, and McMahon (2007): C = temperate, s = dry summer, f = without dry season and b = warm summer.
[b]Forested proportion revised from 67% in Fahey and Payne (2017).

period, precipitation was measured by tipping bucket raingauges (one near each catchment outlet) and by storage gauges. Discharge was monitored using broad-crested, 120° v-notch weirs. The only roads of any significance are the access roads essentially confined to the catchment boundaries. Two specific management features stand out. Contour ripping was carried out before GH2 was planted in 1982, having the effect of increasing the annual baseflow as a proportion of the total flow. Then, throughout the study, a woody scrub, mānuka (*Leptospermum scoparium*), has been gradually encroaching upon the tussock grassland catchment. By 2015, the total scrub cover (of several types) was 39.5%, of which mānuka accounted for 28.4%. Mānuka has higher evapotranspiration rates than tussock grassland and its spread has measurably reduced the annual runoff from GH1. Harvesting of the trees in the GH2 catchment began in March 2014 and was completed in early 2018. Further details are given by Fahey and Payne (2017) and Fahey, McNeill, and Payne (2018).

2.3 | H.J. Andrews experimental forest, Oregon, USA

The site lies in the headwaters of the Willamette River in the Cascade Range of Oregon, USA, east of Eugene. It pairs a control forested catchment (WS2, largely old growth Douglas Fir) with two catchments (WS1 and WS3) that have undergone logging and subsequent regrowth. Data records begin in 1952. Precipitation was measured by storage gauge and recording gauge and may take the form of rain or snow. Discharges were measured by trapezoidal flumes. The principal management features concern roads and the proportion of catchment logged. Absence of roads in WS1 (logging was by skyline suspension) contrasts with the network of 2.7 km of roads in WS3, constructed in 1959 with an approximate density of 2.7 km km^{-2}. WS1 was 100% clear-cut during 1962–1966 and the residue was burned in 1966. WS3 was only partially (25%) clear-cut, in summer 1963, and the resulting patch cuts were burned. Both WS1 and WS3 were replanted after burning, with Douglas Fir. Further details are given by Jones and Grant (1996) and https://andrewsforest.oregonstate.edu/research/infrastructure/watersheds.

2.4 | La Reina, Chile

The site lies in the Rio Bueno catchment in the Coastal Mountain Range of Los Lagos Region, Chile, west of Osorno. Monitored from 1997 to 2018, it is Chile's longest-running forest research catchment. La Reina is a single catchment which has undergone two plantation forest rotations. It was planted in 1977 with Monterey Pine (*Pinus radiata*) over 79.4% of its area, the remainder being covered with roads (an approximate density of 1.2 km km^{-2}) and riparian vegetation, including deciduous trees. Precipitation was measured with a tipping bucket raingauge and discharge was monitored with a flume. The roads lie mostly along the catchment boundary and incorporate only one channel crossing. The plantation was logged from October 1999 to March 2000 and in June–July 2000 was replanted with *Eucalyptus nitens* (42.1% of the catchment area) and *Pinus radiata* (37.3%). At the beginning of 2017, a new cycle of harvesting began and by the end of March 2018 the total area covered with *Eucalyptus nitens* had been clearcut. The principal management feature is that logging was carried out with rubber-tyred skidders in areas with gentle slopes and by cable logging on the steeper slopes. Further details are given by Birkinshaw, Bathurst, Iroumé, and Palacios (2011), Iroumé, Mayen, and Huber (2006) and Iroumé, Palacios, Bathurst, and Huber (2010).

3 | METHODOLOGY

The sites are first assessed for stationarity of conditions (by considering precipitation trends and by double mass curve analysis) and for their compliance with the conventionally expected forest impact of reduced annual runoff (by considering the rainfall-runoff relationships). Chronological and frequency pairing comparisons then analyse the forest impact on peak discharges.

Chronological pairing compares peak discharges from different catchments or different catchment states when paired by the same or an equal rain event. For the paired catchments, the peak discharges are plotted against each other while, for La Reina, peak discharge is plotted against storm rainfall, distinguishing between the pre- and post-logging periods.

Flood frequency curves were prepared by ranking the peaks in size order and estimating exceedance probability using the Gringorten (1963) relationship.

$$P(x) = (r - 0.44)/(N + 0.12), \quad (1)$$

where $P(x)$ is the probability of a peak discharge equalling or exceeding a magnitude x in any given year, r is the ranking where $r = 1$ is the largest peak discharge and N is the number of years in the data record. The return period $T(x)$ for an annual maximum series (or recurrence interval for a partial duration series) was calculated for each peak discharge as $1/P(x)$. Generalized Extreme Value (GEV) distributions were fitted to the annual maximum series using L-moments.

For the paired catchments with records from before and after the relevant intervention (i.e., Glendhu and H.J. Andrews), and using the peak discharge data from the chronological pairings, linear calibration equations were derived for the pre-intervention period as:

$$Q_{\text{intervention}} = a\,Q_{\text{control}} + b, \quad (2)$$

where $Q_{\text{intervention}}$ is the peak discharge for the catchment due to undergo intervention, Q_{control} is the discharge for the control catchment that remains unaffected by the intervention and a and b are coefficients. (The equations are shown later in Figure 5.) Frequency curves were then determined for the affected catchment for the post-intervention period using both the measured data and a data series calculated from the post-intervention data of the control catchment using Equation (2). The latter represents the response that would be expected from the catchment if it had not undergone intervention, assuming otherwise stationary conditions between the pre- and post-intervention periods. Comparison of the two curves indicates the extent to which the intervention has altered the flood frequency curve.

For the post-intervention period, the fitted trendlines in the chronological pairing are not necessarily linear. The form is selected to characterize the data pattern on a visually representative basis and is not proposed as a quantitative model.

4 | ANALYSIS AND RESULTS

4.1 | Stationarity of conditions

Wark Forest, Glendhu and La Reina all show slight increasing trends in annual precipitation over the analysis periods, while H.J. Andrews shows a slight decreasing trend (Figure 2). Relative to the other sites, the Wark Forest trend is exaggerated by the shorter period of record and by the unusually wet summer of 2012. However, double mass curve analysis for cumulative annual runoff against precipitation shows constant relationships for the Wark Forest paired catchments for the entire period (Figure 3). In other words, there were no time-related deviations that might otherwise have biased the peak discharge comparison between the catchments.

Double mass curve analysis for the other sites shows the expected trends of increased runoff in a catchment post-logging and a decreasing runoff following plantation and during forest growth (Figure 3). The analysis for the H.J. Andrews sites plots cumulative runoff against cumulative runoff, rather than cumulative precipitation, as this distinguishes the patterns a little more clearly.

4.2 | Annual runoff

Figure 4 compares the rainfall–runoff relationships for the forested and grassland or logged catchments, according to the calendar year for Glendhu and the water year for the others (1 October–30 September in the UK and USA and 1 April–31 March in Chile). To emphasize the maximum difference between the vegetation covers, the Wark Forest data are represented by an equivalent 100% grassland catchment and an equivalent 100% forest catchment calculated from the Coalburn and Flothers data for their seven years of simultaneous availability during 2003–2012 (Bathurst et al., 2018). The Reina data are represented by the series for forest cover (1997–2000 and 2010–17) and the series for the post-logging periods, including early plantation growth (2000–2009 and 2017–2018). The Glendhu data are limited to the period 1991–2004 between canopy closure in the GH2 catchment and the development of extensive mānuka scrub cover in GH1 and the forest catchment data are recalculated for an

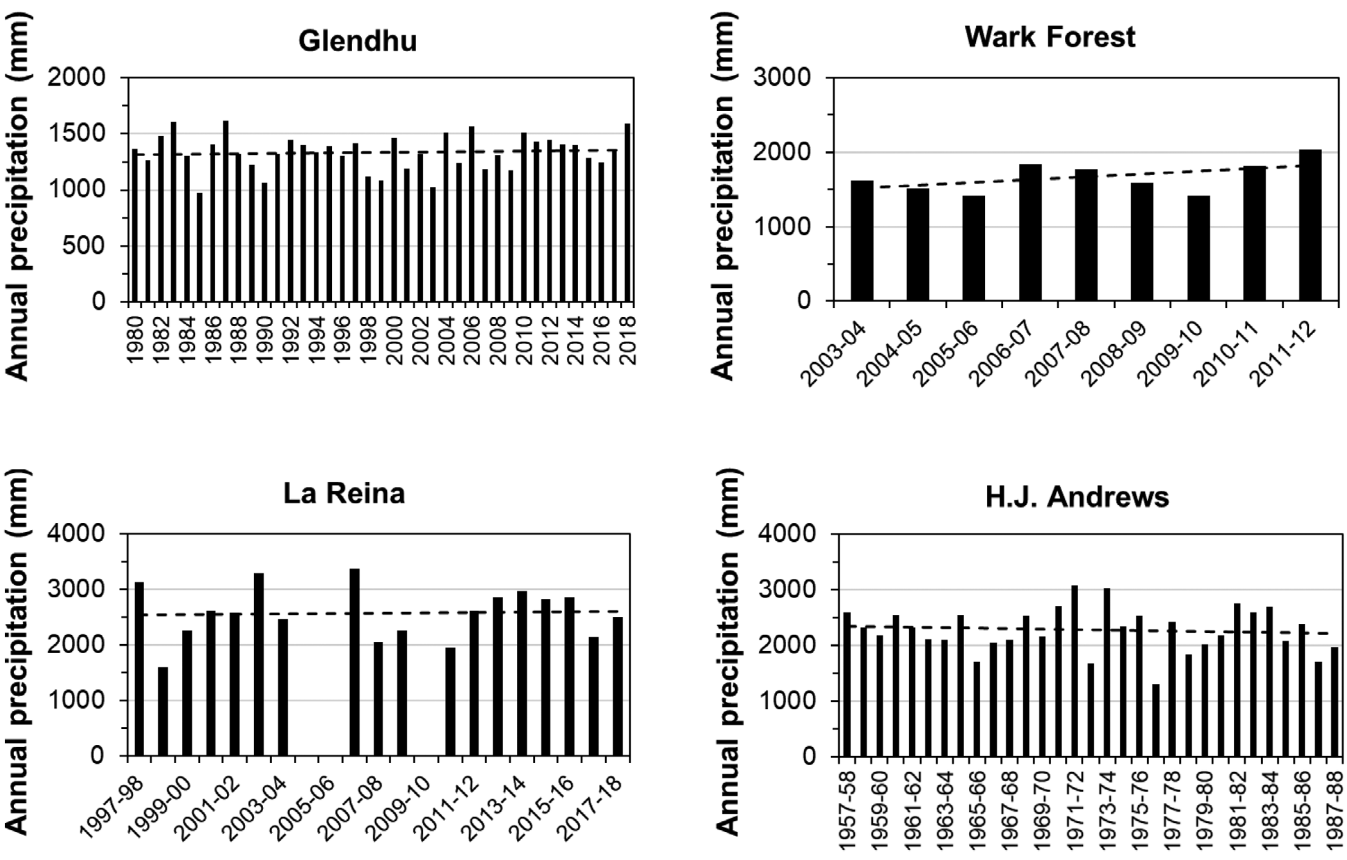

FIGURE 2 Annual precipitation trends (dashed lines) for the analysis periods at the field sites

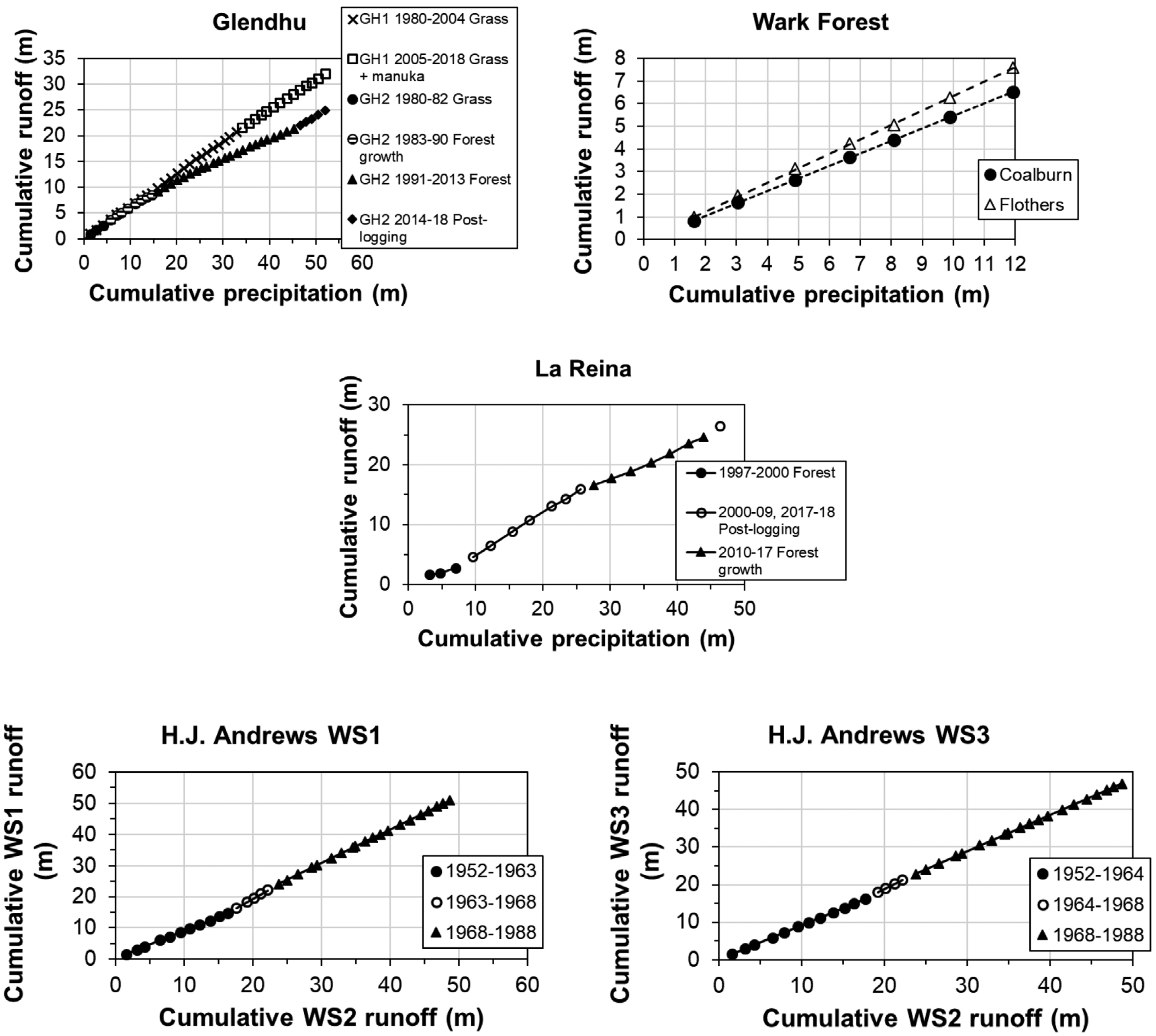

FIGURE 3 Double mass curves for cumulative annual runoff for the field sites

equivalent 100% forest cover. The H.J. Andrews Forest data are represented by the forested control WS2 catchment and the logged WS1 catchment for the period 1963–1972. For the WS1 catchment, this combines that part of the logging period in which annual runoff shows a clear increased response (1963–1966) and the following six-year period (1966–1972) (selected to represent the logged catchment with minimum effect from the new plantation and including the high rainfall year of 1971–1972). In general, the runoff increases linearly with rainfall for the given data range, albeit with scatter, and shows the conventionally expected reduction for the forested state compared with the grassland/logged state. Perhaps fortuitously, the 100% grassland/logged relationships for all four sites align, providing a convenient basis for comparing the effects of forest cover. There is a rough alignment (again perhaps fortuitous) of the forested relationships for Wark Forest, Glendhu and H.J. Andrews while the Reina site shows a separate much bigger runoff reduction. However, subsets of the Glendhu and Reina data sets form a transition between the two forest alignments. For the Glendhu site there is a tendency (but not in every year) for the trees to increase their water use and therefore increasingly reduce runoff as they grow (most apparent for those years that have a similar rainfall). For La Reina, the opposite occurs during the latter half of the 2010–2017 period: the trees increasingly reduce their water use and runoff increases as they age (again most apparent for years of similar rainfall). The reason for this pattern is not yet clear but it has been observed in other catchments in Chile and may reflect a reduced water use in mature, compared with young, trees. The two forest alignments therefore define approximate bounds on the study data for the forested catchments. The fitted lines in Figure 4 show the

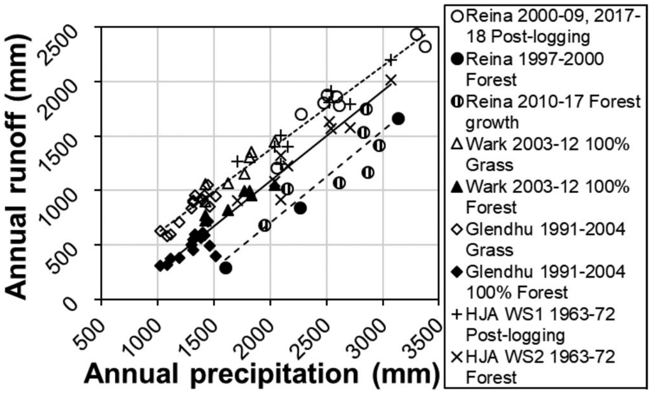

FIGURE 4 Comparison of annual runoff with annual rainfall for the four sites. Data are separated into three alignments characterized by fitted linear trendlines, as described in the text

three alignments (including the one for the grassland/logged catchments) to be roughly parallel, so that, for annual rainfalls of 1,500 and 3,000 mm, forest runoff as a proportion of grassland/logged runoff is 29 and 72% respectively for the lower forest alignment and 68 and 89% respectively for the upper forest alignment (noting, though, that the lines are provided for visual guidance and are not proposed as a quantitative model).

4.3 | Comparison of peak discharges by chronological pairing

Ideally the chronological pairing would be based on annual maximum peak discharges but only the Glendhu site has an annual maximum time series long enough to define a statistically credible pattern. For the other catchments, some form of partial duration or peak-over-threshold series is employed to ensure sufficient data points.

For the Glendhu site, Figure 5a compares the annual maximum flood peaks determined for the grassland control catchment GH1 with the corresponding peaks for the forested catchment GH2 (i.e., for the same event). The data show significant scatter for 1980–1986, that is, the 1980–1981 pre-plantation period plus 5 years following the December 1981 contour ripping but when the plantation had little impact on annual evapotranspiration (as shown in Figure 3). Overall, though, the gradient (coefficient *a*) in the regression of GH2 on GH1 in Equation (2) is similar to that of the line of equality. (For the period 1980–1986, $a = 0.932$, $b = -0.0159$, $R^2 = 0.493$; with the addition of the post-logging data of 2014–2018, $a = 1.046$, $b = -0.343$, $R^2 = 0.597$.) That line is therefore adopted as the pre-intervention calibration equation (i.e., Equation [2]). Also, on this basis, the contour ripping does not appear to have significantly affected the peak discharges. The post-plantation data are divided into two periods, the first beginning at canopy closure in 1991 and the second defined by the extensive (greater than 20%) spread of mānuka scrub across the grassland catchment. In general the GH2 peaks are lower than their GH1 counterparts. The polynomial curve fitted to the 2005–2013 data is considered a more accurate visual representation than a linear line.

There is no obvious effect of the mānuka scrub on the peak flows in the grassland catchment but, to check in more detail, mean storm peak discharges were calculated according to class for the periods 1993–2001 (mānuka encroachment less than 20%) and 2005–2013 (mānuka cover averaging 26%) (Table 2). Except for the largest storm class (in which the earlier period is poorly represented), there is little difference in mean peak discharges between the two periods, suggesting that the mānuka has had no discernible effect on peak discharges, at least for an encroachment of no more than 26% of the catchment area.

The Wark Forest comparison (Figure 5b) is based on a 6.83-year partial duration series of 40 flood events, corresponding to a threshold discharge of 0.183 $m^3\ s^{-1}\ km^{-2}$ for the Coalburn catchment (Bathurst et al., 2018). In this case it is the forested, not the grassland, catchment that produces the higher peak discharges for a given rainfall event. It is thought that part of the Coalburn catchment's ditch network supports a flashier runoff response, while runoff in the Flothers catchment is impeded by near-surface and surface storage of water and the greater flow resistance of the vegetation layer (Bathurst et al., 2018). This is especially true in the summer when the grass is taller and the groundwater levels are lower. The responses are closer in the winter, when the ground is close to saturation and the grass dies back. Figure 5b therefore contrasts December–April (when the two catchments exhibit a similar response) and May–November (when the forested catchment peaks range significantly higher than the corresponding grassland catchment peaks). Peak discharge moderation for low to moderate events is thus a function of soil moisture content. Significantly, though, both periods still show convergence of response for the biggest events. Polynomial curves provide good visual representation of the data variations for both periods. The ditch effect means that the Wark Forest data do not show the effect of purely forest cover on peak discharge response. Nevertheless, the overall similarity of the response pattern to the other sites (i.e., a greater range of differences in the paired catchment responses at lower discharges and convergence at the larger discharges) suggests that the site still provides a valid illustration of the effect of land use change on peak discharge response.

For La Reina, paired peak discharges and rainfall depths were identified for the pre-logging (1997–2000), immediate post-logging (2000–2004) and later post-logging (2006–2017) conditions. Rainfall events were selected according to a threshold rainfall depth of 5 mm and a separation of at least 5 hours without rain from another period of rain. For the sake of clarity only the first two periods are compared in Figure 5c and logarithmic scales are used to highlight the differences at low to moderate events. The fitted trendlines are therefore power laws. On average, the volume of precipitation from the individual rainstorms that generated the peak flows was not significantly different (*t*-statistic at a 95% level) between the two periods, thereby supporting the hypothesis that the evident increase in average peak flows results from the loss of forest cover. Both the pre- and post-logging periods are able to produce similar peak discharges for all

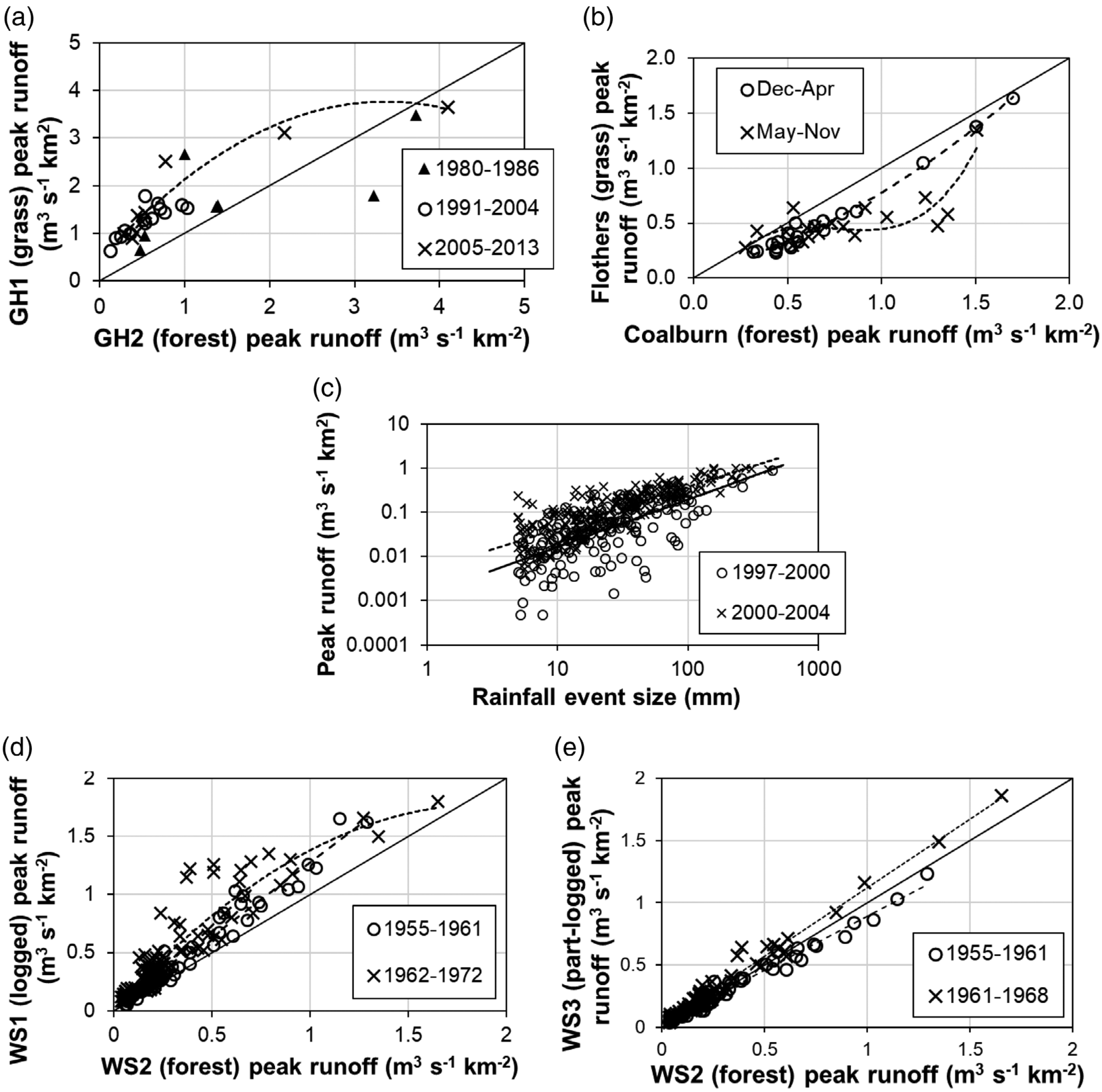

FIGURE 5 Comparison by chronological pairing. (a) Glendhu annual peak runoffs for 1980–1986 (no or negligible forest impact), 1991–2004 (mānuka cover <20%) and 2005–2013 (mānuka cover averaged 26%). Dashed polynomial trendline fitted to the data for 2005–2013. (b) Wark Forest peak runoffs, contrasting the periods December–April and May–November. Dashed lines are fitted polynomial trendlines. (c) La Reina peak discharge versus rainfall depth, comparing pre-logging (1997–2000) and immediate post-logging (2000–2004) periods. Trendlines are fitted power laws, solid for the first period and dashed for the second. (d) H.J. Andrews WS1 and WS2 peak runoffs for 1955–1961 (pre-logging) and 1962–1972 (WS1 logging and post-logging). Dashed lines are fitted trendlines, linear for the first period and polynomial for the second. (e) H.J. Andrews WS3 and WS2 peak runoffs for 1955-March 1961 (pre-logging and road construction) and September 1961–1968 (pre-logging with roads and WS3 logging and post-logging). Dashed lines are fitted linear trendlines. Except for La Reina, solid line is line of equality

event sizes, indicating no mitigating effect from the forest cover. However, the pre-logging period also shows peak discharges ranging to lower levels than the post-logging period for small to moderate events, indicating that a mitigating effect can occur, presumably when the event antecedent soil moisture conditions permit. Because of the logarithmic scale, though, the apparent visual convergence of the two data sets at the largest events may be misleading. Two reasons are advanced for the notably large variation in runoff response for a given rainfall, spanning over two orders of magnitude. The first is the selection method for the events. All events that correspond to the rainfall

Storm class $m^3 s^{-1} km^{-2}$	Number of storms in period		Mean peak discharge $m^3 s^{-1} km^{-2}$	
	1993–2001	2005–2013	1993–2001	2005–2013
0.2–0.5	150	125	0.296 (±0.081)	0.314 (± 0.083)
0.5–1	39	61	0.665 (±0.130)	0.671 (±0.139)
1–1.5	25	12	1.140 (± 0.112)	1.110 (± 0.094)
>1.5	4	4	1.610 (± 0.113)	2.54 (± 0.681)

Note: Standard deviations are in brackets.

TABLE 2 Comparison of mean storm peak discharges for the Glendhu tussock grassland catchment (GH1) for the periods 1993–2001 (mānuka cover <20%) and 2005–2013 (mānuka cover averaging 26%)

event selection criteria are used, rather more than might be obtained from a conventional partial duration or annual maximum selection process. Consequently, there are some small storms with low peaks for which there is no correspondence in the data sets for the other catchments. A second reason might be that the soils at La Reina are drier, owing to the greater forest evapotranspiration at that catchment compared with the others (Figure 4). Consequently, there is potential for a greater reduction in peak discharges.

A sequence of paired peak discharge data was available for the H.J. Andrews catchments (Jones & Grant, 1996) for 1955–1988, comprising pre- and post-logging periods (https://andrewsforest.oregonstate.edu/sites/default/files/lter/data/studies/hf07/hf07fmt.htm).

The peaks were selected by an algorithm that required a certain rate of rise of the hydrograph. Events were separated by a return to a threshold low flow and hence were considered to be independent. The data refer to the period between late autumn (October/November) and the following early spring (April/May). For the natural conditions before logging (1955–1961), there is a clear linear relationship between WS1 and WS2 (control) (Figure 5d). For both the WS1 logging period (1962–1966) and the following six-year period (1967–1972) (selected to represent the 100% logged WS1 catchment with minimum effect from the new plantation), removal of the forest cover results in increased peak discharges for low to moderate floods. As these two periods show very similar responses they are plotted as a single period. As with the Wark Forest and Reina catchments, though, there is a range of impacts, with some data points remaining aligned with the pre-logging relationship. There is no evidence of a forest effect at the highest discharges, where the pre- and post-logging relationships converge. The pre-logging relationship is well represented by a linear line (a = 1.279, b = −0.0026, R^2 = 0.966) while the deviations of the post-logging relationship (i.e., greater width of data scatter at lower relative to higher discharges) are more accurately illustrated by a polynomial curve.

Comparison of WS3 and WS2 was initially carried out for the periods of pre-logging without roads (1955–1958), pre-logging with roads in WS3 (1959–1962) and the combination of the WS3 logging period (25% in three patches during summer 1963) with the five following years (1963–1968). Within the 1959–1962 period, though, the data up to March 1961 coincide with the relationship for 1955–1958 while the data from September 1961 coincide with the relationship for 1963–1968. (Observed sequences of channel fill and scour may explain this separation, as discussed later.) Comparison was therefore drawn between the two periods 1955-March 1961 (i.e., pre-logging without roads and with the first 2 years of WS3 roads) and September 1961–1968 (i.e., the third year of pre-logging with WS3 roads plus WS3 logging and post-logging) (Figure 5e). (There are no data for April–August 1961.) Both periods show a clear linear relationship between the catchments, the second (a = 1.102, b = 0.0199, R^2 = 0.978) delivering higher WS3 specific peak discharges for given WS2 discharges than the first (a = 0.871, b = 0.0231, R^2 = 0.978). There is a small increase in the range of WS3 discharges for a given WS2 discharge at around 0.5 $m^3 s^{-1} km^{-2}$ (the two most noticeable points being in the early post-logging period) but there is no tendency between the two relationships to converge at the larger peak discharges.

4.4 | Comparison of peak discharges by frequency pairing

Flood frequency curves were prepared using the same peak discharge data as for the chronological pairing, apart from some modifications for the H.J. Andrews sites explained below. Following convention, the abscissa of the frequency diagram is labelled "return period" for analysis based on an annual maximum series and "recurrence interval" for analysis based on a partial duration series.

For the Glendhu catchment, as noted above, the line of equality in Figure 5a is adopted as the pre-intervention calibration equation. The post-plantation frequency curve comparison is then directly between the curves for GH1 and GH2 for the period 1991–2013 (Figure 6a), where GH1 represents the expected behaviour of GH2 without afforestation. The curves show a convergence of the responses at the higher discharges while, at the lower discharges, the grassland catchment GH1 has a higher flood magnitude for a given flood frequency than the forested catchment GH2 or, alternatively, a lower return period for a given flood magnitude. Mathematically, if, in chronological pairing, the largest peak discharge for each catchment is for the same event and the discharge values are similar, the flood frequency curves based on ranking of the same data must converge at the largest flow. However, the return period at which convergence apparently occurs is dependent on the length of the record. The larger the number of years N in Equation (1), the larger becomes the return period $T(x)$ for the first ranked peak discharge in the data series. The simple rank-based, annual maximum flood frequency curve is therefore complemented by a fitted GEV distribution for which the shape is determined not just by the largest point but by weighting across all the data points.

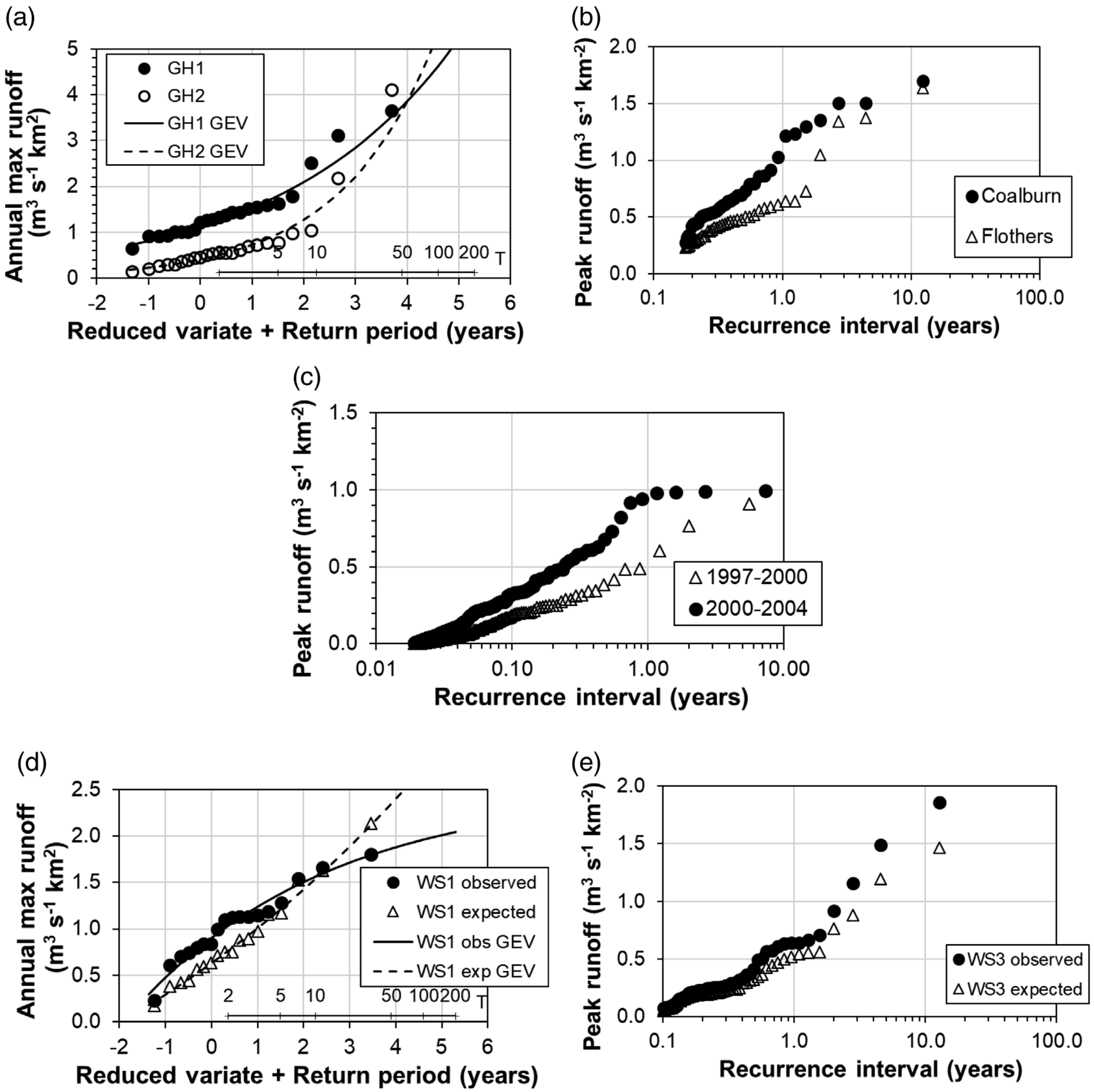

FIGURE 6 Comparison by flood frequency pairing. (a) Glendhu rank-based annual maximum frequency curves and fitted GEV frequency distributions. (b) Wark Forest rank-based frequency curves. (c) La Reina rank-based frequency curves for the pre- and post-logging periods. (d). H.J. Andrews WS1 rank-based annual maximum frequency curves for the observed (logged) and expected (forested) states of the catchment and fitted GEV frequency distributions. (e) H.J. Andrews WS3 rank-based frequency curves for the observed (roads and partial logging) and expected (forested) states of the catchment

For Wark Forest, there are no pre- and post-intervention periods and the existing catchments are simply compared with each other (Figure 6b). For the smaller recurrence intervals, a given runoff magnitude occurs more frequently (with a lower return period) in the ditched forested (Coalburn) than in the grassland (Flothers) catchment. Comparison of the pre- and post-logging frequency curves for La Reina shows a similar pattern except that the post-logging curve lies above the pre-logging curve (Figure 6c). In both cases, convergence of the rank-based frequency curves occurs at the higher runoff values (although subject to the uncertainties associated with these values).

For H.J. Andrews, the post-logging increase in flood peak discharges evident for the WS1 catchment in Figure 5d was maintained, albeit in more subdued form, to at least 1979–1980 (e.g., Jones & Grant, 1996, their Table 2; Thomas & Megahan, 1998, their Table 2).

For a sounder statistical basis, rank-based frequency curves were therefore derived for the 18 logging and post-logging years 1962–1963 to 1979–1980 (Figure 6d). The period is long enough for the curves to be determined for an annual maximum series rather than the partial duration series, enabling GEV distributions to be fitted. For the smaller return periods, peak discharge magnitudes for a given return period are larger for the observed (logged) than the expected (forested) states but both the rank-based curves and the fitted distributions converge at the largest peak discharges, subject to the associated uncertainties.

Rank-based frequency curves for the WS3 catchment were derived for the period 1961–1962 to 1967–1968, comparing the observed (with roads and partial logging) and expected (as if forested without a noticeable road impact) behaviours (Figure 6e). The peak discharge magnitudes for a given return period remain higher for the observed than for the expected state throughout and there is no convergence at the highest discharges.

5 | DISCUSSION

5.1 | Suitability of the four catchments as a basis for a metastudy

While there is significant year-to-year variation in precipitation, there is little long-term trend over the period of record to bias the analysis (Figure 2). Runoff trends are instead much more significantly related to the specific catchment interventions (Figure 3). Figure 4 confirms that the catchments, whatever their management interventions, display the conventionally expected behaviour of decreased annual runoff for a forest cover relative to a grass cover or logged state (e.g., Andréassian, 2004; Bosch & Hewlett, 1982; Zhang et al., 2017). The enhanced nature of this behaviour for the Reina catchment may be related to the high water demand of the exotic tree species, especially over the dry summer period, and a high soil water retention capacity (Huber, Iroumé, & Bathurst, 2008). The consistency of annual conditions and response establishes the credibility of the combined catchments as a firm foundation for a metastudy of forest impacts on flood response. At the same time, the choice of sites enables the consistency or otherwise of the flood response to be tested for a range of interventions: afforestation at Glendhu, logging and roads at H.J. Andrews, mature forest and ditch network at Wark Forest and afforestation and logging at La Reina.

5.2 | Reconciliation of chronological and frequency pairing for analysing forest impact on peak discharges

The controversies that require resolving are whether forests (purely as a vegetation cover) can mitigate peak discharges for large events and whether, in addressing this matter, both the chronological and the frequency pairing methods contribute relevant evidence.

Considering first the impact of forest cover on peak discharge, the chronological pairings of forested and grassland or logged catchments show remarkably similar results for all four field sites (Figure 5). The Glendhu, La Reina and H.J. Andrews WS1/WS2 sites are the most relevant for the effect of forests purely as a vegetation cover but Wark Forest illustrates the same principles in the response to land use differences. For the low to moderate events, the effect of an absence of forest cover ranges from a significant increase (decrease in the case of Wark Forest) in peak discharges for a given event (more than doubling) to no effect at all. The variation is assumed to be due to differences in catchment conditions, especially soil moisture content. If the paired catchments are both saturated, the peak responses will be much the same, despite the differences in vegetation. Wark Forest (Figure 5b) shows this dependency can vary seasonally. For the largest events, all three paired catchments show a convergence of response (Figure 5a,b, d). For La Reina (Figure 5c) this latter pattern is less clear; modelling by Birkinshaw et al. (2011) using 1,000 years of synthetic rainfall data suggests that the range of response remains constant in absolute terms but decreases as a percentage of the event discharge, so indicating a relative rather than an absolute convergence.

The rank-based frequency pairing similarly presents a consistent pattern (Figure 6). For the small to moderate floods, the forested catchment has a lower peak magnitude for a given flood frequency (or a longer return period for a given peak magnitude) than the grassland/logged catchment. (Again the inverse applies to Wark Forest.) For the largest floods, the two curves converge.

The overall agreement between the sites provides confidence in the metastudy approach. It is also reinforced by the one case where there is no convergence, namely, H.J. Andrews WS1/WS3, as agreed by both pairing methods (Figures 5e and 6e). In that case, the road construction may have had the effect of changing the drainage density and thus flow paths (i.e., a hydraulic effect), which would influence the response of events of all magnitudes. Channel scour (discussed later) may also have enhanced the effect. By contrast, the difference in the other cases is assumed to depend on the forest creating a soil moisture deficit to absorb part of the storm rainfall, a hydrological effect that becomes increasingly irrelevant at the larger events.

The value of chronological pairing turns on its relevance to the question of whether a particular peak discharge would have occurred, or would have been as big, if a forest cover had been in place. For low to moderate events, the method shows that forest cover may mitigate peak discharge response for a given rainfall event but also that it may not; the effect depends on not only the vegetation cover but also other factors such as soil moisture, soil depth and snow cover. In other words, it is difficult to give a categorical answer. The results, however, do give some indication of the potential mitigation effect for the most favourable circumstances. For larger events, at which factors such as soil moisture become less relevant, chronological pairing carries a clearer message: for a given rainfall input, vegetation cover has little impact and, in the absence of any other factors, the responses of the paired catchments become very similar. A conceptual model of the pattern is proposed in Figure 7a. Chronological pairing

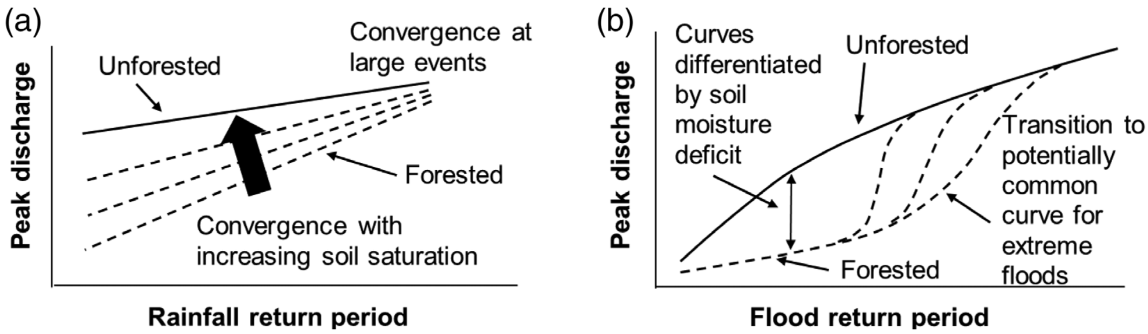

FIGURE 7 Conceptual comparison of peak discharges for forested and unforested catchments as a function of (a) rainfall event return period (chronological pairing) and (b) flood return period (frequency pairing)

therefore addresses the question but its message runs counter to the public perception of forest impact and offers little comfort to a flood-affected citizen wishing to claim for damages resulting from forest logging. If the largest peak discharges are relatively unaffected by vegetation cover, there is no basis for a claim. For the more moderate events, the claimant would have to go to a lot of trouble to obtain data on catchment conditions, such as soil moisture content, to determine if a forest cover would have provided a mitigating effect. Only if the forestry intervention in some way induced a hydraulic effect maintained at all peak discharges might a claim be successful.

Frequency pairing considers the change in magnitude (frequency) for an event of a specific frequency (magnitude) of interest. Figure 7b suggests a possible conceptual model for the impact of forest cover on flood frequency. For all but the largest floods on record, the greater (on average) soil moisture deficits under forest cover mean that peak discharges of given return periods are smaller in forested than in non-forested catchments (and are not necessarily from the same event in the two cases). This in itself is helpful for engineering design, impact assessments, public education and other aspects of catchment management. However, it is an overall long-term context, which neither describes the relative responses for a given event nor accounts for the distinctive patterns of floods of different origins (e.g., different seasonal rainfalls, snowmelt or rain-on-snow). For the largest floods, the effect of the soil moisture deficit becomes increasingly insignificant. The flood frequency curves converge through a transition zone and at the highest magnitudes and return periods are hypothesized to follow a common curve.

The two pairing methods can be reconciled in several ways. Chronological pairing shows, on a storm-by-storm basis, that it is quite possible (at least in flow regimes dominated by rainfall) for forests to have no mitigating effect on magnitude (depending on the soil saturation). In such cases, the discharge peaks would be of the same magnitude for the forested and non-forested states but would correspondingly (Figure 7b) be of different frequency. Further, as noted earlier, for mathematical reasons, frequency pairing must show convergence if the chronological pairing shows convergence. Equally, where chronological pairing does not show convergence, neither does frequency pairing and this pattern has a plausible physical explanation. However, the correspondence of convergence between the methods depends on the largest discharge peaks on record (i.e., the first-ranked in the flood series) showing convergence. So far, according to the available data, this has always been the case, both for the rainfall regime sites of this study and the snowmelt regime sites of Green and Alila (2012) (although those authors explain that they do not consider the effect to be significant). The metastudy approach suggests that such repeated patterns, explainable by physical reasoning, provide strong evidence of a real effect, regardless of statistical significance. Nevertheless, the possibility remains open that, with more extensive data bases, the largest peaks on record may sometime differ between the paired catchments.

Whatever the potential mitigating effects may be, both chronological and frequency pairing show that such effects are likely to be more obvious at small to moderate events and to diminish at the larger events (at least for rainfall-dominated flow regimes). The difficulties of obtaining a long enough data record for statistical reliability and of measuring the higher flows accurately, though, create significant uncertainty in quantifying both the magnitudes and the return periods of the convergence zone. Green & Alila (2012, their Figure 2) found this return period to increase with record length, raising the possibility that convergence may not occur within any practical range of flood flows. In addition, the confidence limits for fitted flood frequency distributions begin to expand significantly once the return period exceeds around half the length of the record on which the curve is based and even more so as it exceeds the length of the record altogether (e.g., Linsley, Kohler, & Paulhus, 1975). The accuracy of the fitted distributions for the higher discharges (at which convergence is modelled) must therefore be questionable. Indeed, the projections of the GEV curves beyond the highest measured discharges in Figure 6a, d are not physically logical if the only difference between the paired catchments is vegetation cover, as they show the extreme peaks to be higher in the forested catchment than in the grassland or logged catchment. It is not clear what the curve should be for the range of extreme peak discharges but extrapolation of the forest curve derived for lower discharges would be likely to overestimate the 100-year flood while extrapolation of the grassland/logged curve might underestimate it.

It remains unclear if there is a characteristic flood return period at which the responses of forested and non-forested catchments converge. First estimates from Figure 6 indicate a range of return periods from around 50 years for Glendhu to roughly 10 years for the other three sites. In addition to the influence of record length, catchment characteristics may play a determining role. For example, catchments with deeper soils or greater forest evapotranspirations (e.g., La Reina in Figure 4) may maintain separate frequency curves for forested and non-forested states to the very largest floods (as perhaps indicated by Birkinshaw et al.'s (2011) 1,000-year simulations for La Reina). Invoking the power of the metastudy, the fact that all the paired catchments (except for WS1/WS3) show convergence of response for data records of no more than a decade or two, suggests nevertheless that the flood events required for convergence to occur may in many cases be relatively common (with return periods of perhaps 5–20 years) rather than relatively rare.

For this study, the major differences in response between the paired catchments or catchment states at each site are explainable by plausible physical reasons related to the forestry interventions. More subtle differences in the relative patterns between the sites will depend on site characteristics, such as soil properties, the significance of groundwater response, the vegetation type and the climate. It was beyond the scope of this study to investigate such dependencies but they appear to have relatively little impact on the broad response pattern established for the type of catchment examined here. For a wider range of sites, though, they may hold greater significance (e.g., Cosandey et al., 2005).

5.3 | Impact of forest management practice

Considering the effect of vegetation cover only (without any other interventions), there is unanimity of response across the catchments. Relative to the logged or grassland state, evapotranspiration is higher in forested catchments. At the annual scale runoff is therefore reduced quite substantially, especially in the drier years (Figure 4). At the event scale, as discussed in the previous section and Figure 7, forest cover may mitigate peak discharge magnitude and frequency at low to moderate floods but has less effect at the largest events when any buffer of soil moisture deficit is overwhelmed by the amount of rainfall (e.g., Figure 5d).

This pattern is clear where the contrast is between 100% and zero forest cover. For the 25% logging in H.J. Andrews WS3, the same pattern appears to be present but only to a minor degree (e.g., the small perturbations evident for WS2 discharges around 0.5 $m^3 s^{-1} km^{-2}$ in Figures 5e and 6e). Similarly the spread of mānuka in the grassland Glendhu GH1 catchment seems to have had little impact on peak discharges, at least up to a catchment cover of 26%. This suggests that the threshold of percentage change in vegetation cover needed to cause a measurable impact on peak discharges is no smaller than the 20% change considered necessary to cause a measurable change in annual runoff (Bosch & Hewlett, 1982; Stednick, 1996) and may be larger.

The altered response in H.J. Andrews WS3 apparent in Figures 5e and 6e occurs in 1961, 2 years after the road construction in that catchment and 2 years before the 25% forest logging. Whether it is due to the road construction is therefore not entirely clear. Jones and Grant (1996) report that a landslide deposited material in the channel in 1960 and this, along with the large amount of woody debris present in the channel, may possibly have temporarily reduced the channel efficiency, counteracting the effect of the roads. Channel efficiency would then have greatly increased in December 1961, when a kilometre of channel length was scoured by a debris flow, and again during similar events in December 1964 and January 1965. On the other hand, the change is uncharacteristically short-lived. Data for the 1970s (not shown here) indicate a return towards the pre-intervention period (noted also by Thomas and Megahan (1998)). Possibly this indicates a restocking of the channel with woody and other debris. Roads, through reduced surface infiltration and associated drainage works, may effectively extend the natural stream network and increase drainage efficiency. Unlike the purely vegetation effect which becomes irrelevant at large rainfall events, this effect, along with that of channel scouring, remains relevant for the full range of events. The patterns shown in Figures 5e and 6e, where there is no convergence for large events, therefore suggest a minor forest logging effect (with a potential for slightly increased peaks at small events) which disappears at larger events, superimposed on a road and channel scour effect (increased peaks) which applies to all events.

For Wark Forest, Figures 5b and 6b suggest that some of the ditches in the Coalburn catchment can still deliver a higher peak discharge than the Flothers grassland catchment during periods of reduced soil saturation and summer grass growth, for small to moderate events. The effect, though, seems to be as much to do with conditions in the Flothers catchment as with the ditch effect in the Coalburn catchment. Consequently, the ditch effect does not apply for the full range of discharges (as might otherwise be expected) but is overcome at larger events or at times of soil saturation when the drainage efficiency of the Flothers matches that of Coalburn.

The hydrological effect of purely vegetation change on flood peak discharges can thus be dominated by the hydraulic effect of changes in drainage efficiency associated with roads, ditching and channel scouring, especially at large events. Other potential hydrological effects, not distinguishable in the data presented here but not necessarily inactive at the sites, may arise from differences in accumulated snowpack between forested and logged/grassland catchments (e.g., Green & Alila, 2012; Jennings & Jones, 2015; Jones & Perkins, 2010) and from alterations in soil permeability linked to ditching and contour ripping.

6 | CONCLUSIONS

Combining the data of four catchment studies has extended the geographical range of research into the impacts of forests on flood peak discharges and provided encouraging new field evidence to support the hypothesis that the chronological and frequency pairing methods for analysing the impacts can be reconciled. The study establishes an approach for comparing the methods and proposes a conceptual model for the impact of forests on discharge peaks (Figure 7) which other researchers can test. Central to the study has been a careful interpretation of the data for each catchment, based on physical reasoning and the authors' detailed understanding of their catchments.

1 Repeated patterns across the four study sites demonstrate consistency of response between field studies and provide strong evidence of real effects with physical explanations.
2 Both chronological and frequency pairing are relevant methods for determining the impact of forests on flood peak discharges but they address different, complementary questions. Frequency pairing provides an overall long-term characterization of the relationship between frequency and magnitude of discharge peaks; catchments with and without interventions have clearly differentiated relationships. Chronological pairing provides storm-by-storm commentary on peak magnitudes; relative responses between catchments with and without interventions may be highly variable, as a function also of catchment conditions.
3 Most of the available data refer to low to moderate floods. In this range, relative to a grassland or logged catchment, a forested catchment has a lower peak discharge magnitude for a given flood frequency or a larger return period for a given peak discharge magnitude. Within this overall pattern, though, the effect of the forest cover on a storm-by-storm basis depends on catchment conditions, especially the soil antecedent moisture content. Forest cover may mitigate the flood peak discharge (potentially by 50% or more). Equally, a given rainstorm may generate peaks of the same magnitude, but different return periods, for different vegetation covers, meaning no mitigation of peak discharge magnitude. More colloquially, forests do not prevent floods but they can make them less frequent.
4 For the largest events on record, both frequency and chronological pairing show a convergence of response, which suggests an increasing irrelevance of the vegetation cover.
5 The flood frequency, above which forest cover loses its potential mitigating effect, seems likely to vary between catchments, with moderate return periods (5–20 years) in some cases but more extreme values possible in others.
6 Catchment interventions other than purely change of vegetation cover can modify the above responses. Increased peak flows due to road networks, drainage works and channel scouring may be apparent for all events, with no convergence of response at large events. Ditch networks in the forested catchment may invert the relative magnitude of peak discharges otherwise expected between forested and grassland catchments.
7 For all the sites, whatever the management interventions, forest cover substantially reduces annual runoff by comparison with the grassland/logged state, especially in drier years.

Despite the striking similarities in the catchment responses which underlie the proposed model of Figure 7, wider confirmation of the model is required, through extension to a larger number of catchments (with a range of characteristics and management interventions), to catchment areas larger than the few square kilometres of this study and to tropical zones. Particular work is needed to determine if it is possible to identify a characteristic flood return period at which the responses of forested and non-forested catchments converge.

ACKNOWLEDGEMENTS

The authors acknowledge the efforts of the teams that collected the field data over many years, in good weather and bad. The data for Coalburn were supplied by the Environment Agency for England. The data for the H.J. Andrews sites were provided by the H.J. Andrews Experimental Forest research program, funded by the National Science Foundation's Long-Term Ecological Research Program (DEB 1440409), the US Forest Service Pacific Northwest Research Station and Oregon State University. Stimulating conversations with Andras Bardossy and Chris Kilsby greatly helped the paper development. This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

James C. Bathurst https://orcid.org/0000-0002-5650-2259
Barry Fahey https://orcid.org/0000-0003-4568-9297
Andrés Iroumé https://orcid.org/0000-0001-8148-1309

REFERENCES

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

Andréassian, V. (2004). Waters and forests: from historical controversy to scientific debate. *Journal of Hydrology*, *291*, 1–27.

Archer, D., & Newson, M. (2002). The use of indices of flow variability in assessing the hydrological and instream habitat impacts of upland afforestation and drainage. *Journal of Hydrology*, *268*, 244–258.

Bathurst, J., Birkinshaw, S., Johnson, H., Kenny, A., Napier, A., Raven, S., … Stroud, R. (2018). Runoff, flood peaks and proportional response in a combined nested and paired forest plantation/peat grassland catchment. *Journal of Hydrology*, *564*, 916–927.

Bathurst, J. C., Birkinshaw, S. J., Cisneros, F., Fallas, J., Iroumé, A., Iturraspe, R., … Sarandón, R. (2011). Forest impact on floods due to extreme rainfall and snowmelt in four Latin American environments 2: Model analysis. *Journal of Hydrology*, *400*, 292–304.

Bathurst, J. C., Iroumé, A., Cisneros, F., Fallas, J., Iturraspe, R., Gaviño Novillo, M., … Ramirez, M. (2011). Forest impact on floods due to extreme rainfall and snowmelt in four Latin American environments 1: Field data analysis. *Journal of Hydrology*, *400*, 281–291.

Beschta, R. L., Pyles, M. R., Skaugset, A. E., & Surfleet, C. G. (2000). Peakflow responses to forest practices in the western cascades of Oregon, USA. *Journal of Hydrology*, *233*, 102–120.

Birkinshaw, S. J., Bathurst, J. C., Iroumé, A., & Palacios, H. (2011). The effect of forest cover on peak flow and sediment discharge—an integrated field and modelling study in central-southern Chile. *Hydrological Processes*, *25*(8), 1284–1297.

Birkinshaw, S. J., Bathurst, J. C., & Robinson, M. (2014). 45 years of non-stationary hydrology over a forest plantation growth cycle, Coalburn catchment, Northern England. *Journal of Hydrology*, *519*, 559–573.

Bosch, J. M., & Hewlett, J. D. (1982). A review of catchment experiments to determine the effect of vegetation changes on water yield and evapotranspiration. *Journal of Hydrology*, *55*, 3–23.

Cambrian Wildwood, 2015. https://www.cambrianwildwood.org/deforestation-and-flooding.

Carrick, J., Abdul Rahim, M. S. A. B., Adjei, C., Kalee, H. H. H. A., Banks, S. J., Bolam, F. C., … Stewart, G. (2018). Is planting trees the solution to reducing flood risks? *Journal of Flood Risk Management*, *12*, 1–10. https://doi.org/10.1111/jfr3.12484

Confor, 2018. http://www.confor.org.uk/news/latest-news/mp-calls-for-200-million-new-trees-by-2020-to-stop-floods/.

Cosandey, C., Andréassian, V., Martin, C., Didon-Lescot, J. F., Lavabre, J., Folton, N., … Richard, D. (2005). The hydrological impact of the mediterranean forest: a review of French research. *Journal of Hydrology*, *301*, 235–249.

Fahey, B., McNeill, S., & Payne, J. (2018). The effect of mānuka encroachment on water yield from the tussock catchment, Glendhu experimental study, upland East Otago, New Zealand. *Journal of Hydrology (NZ)*, *57*(2), 81–89.

Fahey, B., & Payne, J. (2017). The Glendhu experimental catchment study, upland East Otago, New Zealand: 34 years of hydrological observations on the afforestation of tussock grasslands. *Hydrological Processes*, *31*, 2921–2934. https://doi.org/10.1002/hyp.11234

Fowler, S. (2018). Of forests and floods: Devastatingly high water raises clear-cut questions. CBC news, June 18, 2018. Retrieved from https://www.cbc.ca/news/canada/new-brunswick/forests-and-floods-clear-cutting-new-brunswick-floods-1.4703225.

Green, K. C., & Alila, Y. (2012). A paradigm shift in understanding and quantifying the effects of forest harvesting on floods in snow environments. *Water Resources Research*, *48*(10), W10503. https://doi.org/10.1029/2012WR012449

Gringorten, I. I. (1963). A plotting rule for extreme probability paper. *Journal of Geophysical Research*, *68*(3), 813–814.

Huber, A., Iroumé, A., & Bathurst, J. (2008). Effect of *Pinus radiata* plantations on water balance in Chile. *Hydrological Processes*, *22*, 142–148.

Iroumé, A., Mayen, O., & Huber, A. (2006). Runoff and peak flow responses to timber harvest and forest age in southern Chile. *Hydrological Processes*, *20*, 37–50.

Iroumé, A., Palacios, H., Bathurst, J., & Huber, A. (2010). Runoff and peakflows after clearcutting and the establishment of a new plantation in an experimental catchment, southern Chile. *Bosque*, *31*(2), 117–128 (In Spanish).

Jennings, K., & Jones, J. A. (2015). Precipitation-snowmelt timing and snowmelt augmentation of large peak flow events, western Cascades, Oregon. *Water Resources Research*, *51*(9), 7649–7661.

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

Jones, J. A., & Grant, G. E. (2001). Comment on "Peak flow responses to clear-cutting and roads in small and large basins, western Cascades, Oregon: A second opinion" by R. B. Thomas and W. F. Megahan. *Water Resources Research*, *37*(1), 175–178.

Jones, J. A., & Perkins, R. M. (2010). Extreme flood sensitivity to snow and forest harvest, western Cascades, Oregon, United States. *Water Resources Research*, *46*(12), W12512. https://doi.org/10.1029/2009WR008632

Kuraś, P. K., Alila, Y., & Weiler, M. (2012). Forest harvesting effects on the magnitude and frequency of peak flows can increase with return period. *Water Resources Research*, *48*(1), W01544. https://doi.org/10.1029/2011WR010705

La Marche, J. L., & Lettenmaier, D. P. (2001). Effects of forest roads on flood flows in the Deschutes River, Washington. *Earth Surface Processes and Landforms*, *26*, 115–134.

Lewis, J., Reid, L. M., & Thomas, R. B. (2010). Comment on "Forest and floods: A new paradigm sheds light on age-old controversies" by Younes Alila et al. *Water Resources Research*, *46*, W05801. https://doi.org/10.1029/2009WR008766

Linsley, R. K., Kohler, M. A., & Paulhus, J. L. H. (1975). *Hydrology for Engineers* (2nd ed.). New York, NY: McGraw-Hill.

Peel, M. C., Finlayson, B. L., & McMahon, T. A. (2007). Updated world map of the Köppen-Geiger climate classification. *Hydrology and Earth System Sciences*, *11*(5), 1633–1644.

Robinson, M. (1998). 30 years of forest hydrology changes at Coalburn: water balance and extreme flows. *Hydrology and Earth System Sciences*, *2*(2), 233–238.

Soulsby, C., Dick, J., Scheliga, B., & Tetzlaff, D. (2017). Taming the flood—How far can we go with trees? *Hydrological Processes*, *31*(17), 3122–3126. https://doi.org/10.1002/hyp.11226

Stednick, J. D. (1996). Monitoring the effects of timber harvest on annual water yield. *Journal of Hydrology*, *176*, 79–95.

Stratford, C., Miller, J., House, A., Old, G., Acreman, M., Dueñas-Lopez, M. A., … Tickner, D. (2017). Do trees in UK-relevant river catchments influence fluvial flood peaks? Wallingford, UK, NERC/Centre for Ecology & Hydrology, 46pp. (CEH Project no. NEC06063).

Thomas, R. B., & Megahan, W. F. (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.

Yu, X. J., & Alila, Y. (2019). Nonstationary frequency pairing reveals a highly sensitive peak flow regime to harvesting across a wide range of return periods. *Forest Ecology and Management*, *444*, 187–206.

Zhang, M., Liu, N., Harper, R., Li, Q., Liu, K., Wei, X., … Liu, S. (2017). A global review on hydrological responses to forest change across multiple spatial scales: Importance of scale, climate, forest type and hydrological regime. *Journal of Hydrology*, *546*, 44–59.

How to cite this article: Bathurst JC, Fahey B, Iroumé A, Jones J. Forests and floods: Using field evidence to reconcile analysis methods. *Hydrological Processes*. 2020;34:3295–3310. https://doi.org/10.1002/hyp.13802