Received: 1 October 2020 | Revised: 2 June 2021 | Accepted: 24 June 2021

DOI: 10.1002/hyp.14300

RESEARCH AND OBSERVATORY CATCHMENTS: THE LEGACY AND THE FUTURE

WILEY

A catchment water balance assessment of an abrupt shift in evapotranspiration at the Hubbard Brook Experimental Forest, New Hampshire, USA

Mark B. Green[1,2] | Scott W. Bailey[2] | John L. Campbell[2] | Kevin J. McGuire[3] | Amey S. Bailey[2] | Timothy J. Fahey[4] | Nina Lany[2] | David Zietlow[2]

[1]Department of Earth, Environmental, and Planetary Sciences, Case Western Reserve University, Cleveland, Ohio

[2]Northern Research Station, US Department of Agriculture Forest Service, North Woodstock, New Hampshire

[3]Department of Forest Resources and Environmental Conservation and Virginia Water Resources Research Center, Virginia Polytechnic Institute and State University, Blacksburg, Virginia

[4]Department of Natural Resources & the Environment, Cornell University, Ithaca, New York

Correspondence
Mark B. Green, Department of Earth, Environmental, and Planetary Sciences, Case Western Reserve University, Cleveland, OH, USA.
Email: mbg78@case.edu

Funding information
USDA Forest Service, Northern Research Station, Madison, WI, Grant/Award Numbers: 1114804, 1014507

Abstract

Small catchments have served as sentinels of forest ecosystem responses to changes in air quality and climate. The Hubbard Brook Experimental Forest in New Hampshire has been tracking catchment water budgets and their controls – meteorology and vegetation – since 1956. Water budgets in four reference catchments indicated an approximately 30% increase in the evapotranspiration (ET) as estimated by the difference between precipitation (P) and runoff (RO) starting in 2010 and continuing through 2019. We analyzed the annual water budgets, cumulative deviations of the daily P, RO and water budget residual (WBR $=$ P $-$ RO), potential ET (PET) and indicators of subsurface storage to gain greater insight into this shift in the water budgets. The PET and the subsurface storage indicators suggest that this change in WBR was primarily due to increasing ET. While multiple long-term hydrological and micrometeorological data sets were used to detect and investigate this change in ET, additional measurements of groundwater storage and soil moisture would enable better estimation of ET within the catchment water balance. Increasing the breadth of long-term measurements across small gauged catchments allows them to serve as more effective sentinels of substantial hydrologic changes like the ET increase that we observed.

1 | INTRODUCTION

Small, gauged catchments are useful for evaluating impacts of changing environmental conditions on ecosystem functioning. Their size (typically <100 ha) allows estimation of catchment water and solute budgets, providing a method for determining ecosystem-scale responses to natural and human perturbations. For example, catchments across the eastern United States were instrumental in documenting how increased acid deposition in the mid-to-late 20th century disrupted acid-sensitive ecological and biogeochemical processes (Bailey et al., 1996; Likens et al., 1996). Small catchments are similarly well-suited to detect changes in water budgets as the hydrologic cycle intensifies due to climate change (Creed et al., 2014).

Climate change is altering precipitation patterns (O'Gorman & Schneider, 2009; Trenberth, 2011; Huang et al., 2017), snowpacks (Zeng et al., 2018) and humidity (Byrne & O'Gorman, 2018), with consequences for evapotranspiration (Kramer et al., 2015; McVicar et al., 2012; Vadeboncoeur et al., 2018) and stream flow (Young et al., 2019). Soil water and groundwater changes in response to climate change are less common due to a lack of long-term data for tracking trends. Further, catchment heterogeneity (e.g., subsurface, vegetation, snowpack) makes catchment-scale long-term storage trends difficult to quantify. Documenting the interactions of hydrological processes and their drivers as they change is important to improving hydrologic models used to forecast the implications of climate change on water resources.

At the Hubbard Brook Experimental Forest (HBEF) in New Hampshire, USA, the hydrology and meteorology of small, forested catchments have been monitored since 1955. Early catchment studies at the HBEF utilized a paired catchment approach to evaluate experimental manipulations designed to identify how forests and their management regulates the quantity and quality of water supplies (Hornbeck et al., 1997). Continuous, long-term monitoring of the hydrologic cycle has produced insights into how climate change has altered snow hydrology (Campbell et al., 2010), evapotranspiration (Vadeboncoeur et al., 2018) and stream flow (Campbell et al., 2011). Shorter-term studies addressing stream source water tracing (Benettin et al., 2015; Fuss et al., 2016; Hogan & Blum, 2003; Hooper & Shoemaker, 1986), groundwater controls on streamflow generation (Detty & McGuire, 2010; Gannon et al., 2014), soil hydrology (Gannon et al., 2017; Germann et al., 1986), water vapour dynamics (Green et al., 2015), evapotranspiration impact on streamflow recession (Federer, 1973) and stream network expansion/contraction (Jensen et al., 2017) have led to a rich understanding of the hydrology of the HBEF catchments, making them effective sentinels of a changing environment.

This study documents recent changes in the water balance at the HBEF. Analysis of associated datasets suggest that an increase in evapotranspiration (ET) caused this response.

2 | METHODS

2.1 | Site

The HBEF is located in the White Mountain National Forest in central New Hampshire, USA (43°56′N, 71°45′W; Figure 1). Four HBEF reference catchments with at least 40 years of hydrometeorological data are the focus of this analysis; two south-facing (W3 and W6) and two north-facing (W7 and W8). We focused on the reference catchments because our interest was in how water budgets responded to ambient environmental changes, not those due to experimental manipulation. These catchments drain forested mountainsides, ranging in elevation from 530 to 905 m. Their areas range from 13.2 at W6 to 77.4 ha at W7 while average slopes range from 16 to 17 degrees. Bedrock is a slowly weathering high-grade crystalline schist, quartzite and calc-silicate granulite of the Silurian Rangeley Formation (Burton et al., 2002). Soils at the HBEF are mostly Spodosols, developed in glacial till and reworked till and glaciofluvial deposits of sandy loam to loamy sand textures. Glacial drift is thin and interspersed with exposed bedrock along ridgelines and portions of the stream network; thickness ranges up to 10 m. Spatial variation in soil development reflects the subsurface hydrology (Bailey et al., 2014). Generally, the less weathered C horizon starts at about 0.7 m, with few roots penetrating the upper portion of the C horizon. The climate at the HBEF is humid temperate with a mean annual temperature of 5.5°C and monthly averages ranging from −9 to 19°C (Bailey et al., 2003; Campbell et al., 2021).

The forest biomass is dominated by the northern hardwood species sugar maple (*Acer saccharum* Marshall), yellow birch (*Betua alleghaniensis* Britton) and American beech (*Fagus grandifolia* Ehrh.) on deeper and better drained soils, with red spruce (*Picea rubens* Sarg.), balsam fir (*Abies balsamea* Mill.) and eastern hemlock (*Tsuga canadensis* (L.) Carrière) on shallow and wetter soils (Siccama et al., 2007). The HBEF was selectively harvested from the 1880s to 1910s, primarily for spruce by loggers using axes and horses. In 1920, the land was transferred to the White Mountain National Forest, with designation of the experimental forest in 1955. The reference catchments have not been actively managed, but have been affected by atmospheric deposition and periodic natural disturbances (e.g., wind, ice and snow storms, insect defoliations). A complete forest inventory performed in 1965, 1977 and every 5 years thereafter, shows that forest biomass accumulated at W6 until about 1983, and has remained at near steady state since then (Battles et al., 2014). The other reference catchments have limited vegetation monitoring, but are part of a larger valley-wide forest monitoring programme sampled decadally. Sampling from 1995–2005 showed that overall biomass was at steady state at the valley-wide scale while there was a small increase in red spruce, and a decrease in birch species (van Doorn et al., 2011).

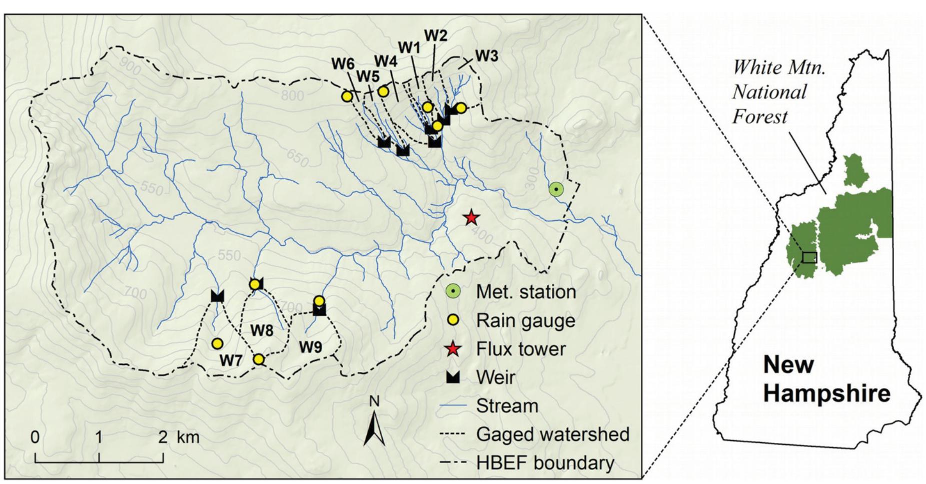

FIGURE 1 Map of the Hubbard Brook Experimental Forest

2.2 | Data

Precipitation and stream discharge have been measured at the HBEF reference catchments considered in the present study starting in the mid-1950s to late-1960s (Campbell et al., 2021). Precipitation (P) was estimated at the catchment-scale by spatial interpolation of measurements from multiple rain gauges (Green et al., 2018; USDA Forest Service, 2020a). Gauge sites were initially instrumented with standard cumulative rain gauges that were shielded (Alter type), located in maintained forest openings and measured weekly (Leonard & Reinhart, 1963). A subset of sites also included a Belfort weighing rain gauge used to distribute weekly precipitation into individual days (Federer, 1990; Green et al., 2018; Leonard & Reinhart, 1963). Since June 2016, 15-minute precipitation has been measured with a shielded (Alter type) ETI NOAH IV digital weighing rain gauge, summed for each day. Stream stage was measured continuously at a v-notch weir, and W6, W7 and W8 have an additional flume installed for measuring high flows. Stage was converted to discharge (Q) using theoretical stage-discharge relationships (Bailey et al., 2003; Reinhart & Pierce, 1964; See et al., 2020) and the daily runoff (RO) was estimated by summing the Q measurements and dividing by the catchment area (USDA Forest Service, 2020a, 2020b).

Micrometeorological measurements from a long-term weather station at the HBEF headquarters were used to estimate potential ET (PET) from 1989 to 2020. Mean daily values were calculated from hourly measurements. Air temperature (T_a) and relative humidity (RH) were measured with a Campbell Scientific Inc. 201 T_a and RH sensor housed in a Gill radiation shield. Wind speed was measured with a Met One 3-cup anemometer from 1981 to 2003, and then with an RM Young model 05103 anemometer from 2003 to 2020. Solar radiation (R_s) was measured with a LiCor pyranometer from 1981 to 2018, and with an Apogee pyranometer from 2018 to 2020. Solar radiation was converted to total available energy as net radiation (R_n) minus ground heat flux (G) by establishing a linear relationship between R_s and $R_n - G$ measured at a flux tower located within the HBEF (AmeriFlux site HBK; Kelsey et al., 2019). The R_n was measured with a Kipp and Zonen net radiometer from 2016 to 2020. Ground heat flux was measured with four Huske Flux soil heat flux plates which were averaged to provide daily values. The relationship between R_s and $R_n - G$ was established on a monthly basis and applied to the entire R_s record (Supplemental methods).

We used two recent data sets that measured aspects of subsurface water storage. Soil water volumetric content was measured from 2011 to 2017 with Decagon 5TM combination temperature and volumetric water content probes, logged hourly at 14 sites along elevation gradients on the south- and north-facing slopes (Groffman, 2019; Wilson et al., 2020). The probes were installed in the top 5–10 cm of soil and thus mostly indicated moisture in the organic and uppermost mineral horizons. Shallow groundwater levels were measured with a network of wells that were installed across W3 and screened near the soil C horizon or at bedrock in shallow soils if no C horizon was present (Detty & McGuire, 2010; Gannon et al., 2014). Data were recorded with Odyssey Water Level Loggers or Hobo Level Loggers every 10 min. The network was established in 2006 with 31 wells and there are currently over 100 wells; however, only a small portion of the network collected data at any given time (McGuire et al., 2019).

2.3 | Data analysis

2.3.1 | Annual water budgets

We calculated catchment water budgets on a daily and June 1 water-year basis. The water budget was formulated as:

$$\mathrm{P} - \mathrm{RO} = \mathrm{WBR} = \mathrm{ET} + \Delta\mathrm{SWE} + \Delta\mathrm{SS} + \mathrm{L}, \tag{1}$$

where P is precipitation, RO is runoff, WBR is the water budget residual, ET is evapotranspiration, ΔSWE is change in snowpack, ΔSS is change in subsurface storage (soil water and groundwater) and L is loss to deep seepage, all terms in the units of mm of water per time. By choosing a water year that starts in the spring (i.e., June 1), typical for eastern US catchment studies (Kelly et al., 2016; Likens, 2013; Lynch & Corbett, 1990; Patric & Reinhart, 1971), $\Delta\mathrm{SWE} = 0$ between years. Water loss to deep seepage or leakage is negligible, which is supported by a comparison of 32 small research catchments which showed that the HBEF had amongst the lowest estimates of deep seepage (Verry, 2003). The water budget was estimated for each catchment individually and a combined water budget for all four catchments was calculated by area-weighting P, RO and WBR. We also estimated the water balance using a 7-year moving window to minimize any interannual ΔSS. The 7-year period was chosen based on the climate and vegetation characteristics at the HBEF compared to the suggested optimal integration window suggested by Han et al. (2020). Since ET and ΔSS are not directly measured in our catchments, we analyzed other variables to gain insight into the contribution of each to changes in the WBR.

2.3.2 | Cumulative water budgets

We explored the dynamics of the sub-annual water balance and longer-term variation using a cumulative deviation analysis. The technique provided a continuous picture of the water budget and its deviations from its typical status. Cumulative sum analyzes are effective at identifying abrupt changes in time series (Hawkins & Olwell, 2012; Hinkley, 1971) and have been used in the classical mass curve approach to assess the quality of hydrological monitoring (Searcy & Hardison, 1960) and water resources analysis such as reservoir operations (Klemeš, 1979). Cumulative sum techniques have been useful in identifying inflection points in water quality time series (Briceño et al., 2014; Cluis, 1983; Regier et al., 2019) and analyzing hydrologic system responses to land cover and climate change (Nijzink et al., 2016; Smail et al., 2019).

The water budget variables analyzed were P, RO and WBR. For each variable, the cumulative sum was calculated over the full time series on a daily time step. An ordinary least squares regression line was fit to the time series; the slope of the model was taken as the long-term central tendency. The residuals, then, were visualized and analyzed to assess inflection points and multi-year deviations from the long-term central tendency.

2.3.3 | Potential evapotranspiration

PET was calculated to assess a major driver of the ET at the HBEF since it is considered an energy limited ecosystem (Creed et al., 2014). Daily PET was estimated with the Penman-Monteith method (Allen et al., 1998):

$$\mathrm{ET} = \frac{\Delta(R_n - G) + \rho_a c_p \frac{(e_s - e_a)}{r_a}}{\lambda\left(\Delta + \gamma\left(1 + \frac{r_c}{r_a}\right)\right)}, \tag{2}$$

where ET is in mm/day, Δ is the slope of the saturated vapour pressure versus temperature curve (kPa/C), R_n is net radiation (MJ/m^2/day), G is ground heat flux (MJ/m^2/day), ρ_a is the density of air (kg/m^3), c_p is the heat capacity of air (MJ/kg C), e_s is the saturated vapour pressure (kPa), e_a is the actual vapour pressure (kPa), r_a is the aerodynamic resistance (d/m), r_c is the canopy resistance (d/m), γ is the psychrometric constant (kPa/C) and λ is the latent heat of vapourization (MJ/kg). The e_s, Δ, γ and r_a were estimated using standard methods (Allen et al., 1998). We assumed air pressure to be constantly 95 kPa since we do not have long-term air pressure measurements. The r_c was estimated on a monthly basis by rearranging Equation 2 for r_c and calculating it with measurements from the flux tower (Xu et al., 2020, Figure S1). The monthly r_c was scaled by the annual leaf area index (LAI) by multiplying annual normalized LAI by the r_c values within that year. Annual LAI of the forest immediately west of W6 was measured by collecting leaf litter with a network of 51 collectors positioned randomly in four plots along the elevation gradient (Fahey & Cleavitt, 2021). Leaf litter was sorted and tallied by species and leaf numbers multiplied by average area per leaf to estimate LAI. The LAI was normalized by dividing the annual values by the long-term median. The LAI data were collected from 1993 to 2017; thus, the long-term median was used for the years 1989–1992 and 2018–2019 in order to match the length of the micrometeorology record. The presence of a trend in LAI was assessed using the Kendall (1938) correlation and the linear slope of the trend was quantified with the Sen (1968) estimator.

Growing season ET was modelled by comparing the summed PET and P during the period when the forest canopy was fully intact. If growing season P was greater than PET, the ET estimate was equal to P. Otherwise, the growing season ET estimate was equal to the PET. This simple model was consistent with the suggestion from Kirchner and Allen (2020) that most of the P during the growing season at the HBEF becomes ET. The canopy was considered leafed-out when the phenological state, which was assessed weekly and interpolated to daily values, was at 3 or higher on a 0–4 scale where 4 is the peak growing season canopy (Richardson et al., 2006). Generally, the canopy leafed-out from June to September.

2.3.4 | Subsurface storage

The difference in year-to-year subsurface storage contributes to the WBR, but is difficult to assess at the catchment scale. We used measurements of low stream flows, soil moisture and shallow groundwater as indices to suggest possible ΔSS. Annual low stream flows were used as an indicator of the catchment water storage trend (e.g., Brutsaert, 2008; McNamara et al., 2011). We calculated the 5th percentile specific discharge from the 5-minute discharge data to provide an indicator of catchment storage per water year. This simple approach was chosen over the storage-discharge relationship which relies on specific conditions to assess stream recession characteristics (Brutsaert, 2008). Year-to-year difference in shallow soil moisture for the Hubbard Brook basin was estimated by normalizing the moisture at each of the 14 sites by its post-snowmelt value, which is assumed to be approximately the field capacity (Wilson et al., 2020). The mean normalized soil moisture per water year was calculated from the hourly data from all sites.

We calculated the year-to-year difference in water table height in shallow wells by isolating the May and June water table measurements per year and describing the central tendency and spread of the water table during that period each year. We further estimated the amount of ΔSS that could arise from the interannual water table change in the shallow groundwater zone by multiplying the change in median water table height by an assumed specific yield of 0.35 mm/mm.

Total storage deviation for the area covering the HBEF was also evaluated using remotely sensed data from the Gravity Recovery and Climate Experiment (GRACE) mission (Landerer et al., 2020). We used the Mass Concentration blocks produced by the NASA Jet Propulsion Laboratory to visualize monthly storage deviations from the 2004 to 2009 mean.

3 | RESULTS

3.1 | Annual water budgets

The annual water budgets showed consistently higher P from water year 2003 to 2019 (Figure 2). For the area-weighted average P of all four catchments (1970–2019), only four of the sixteen years in the 2003–2019 period had lower annual P than the long-term mean of 1488 mm/year: 1432 mm/year in 2004, 1481 mm/year in 2012, 1450 mm/year in 2013 and 1392 mm/year in 2014. The mean annual P prior to 2003 was 1428 mm/year and after 2003 it was 1606 mm/year. Annual RO followed a similar pattern as P, except a step increase was not apparent (Figure 2(b),(e)). The water years 2003–2011 were a contiguous period of higher than normal RO, but the years after 2011 were closer to the long-term mean. The WBR deviated starting in the late 2000s (Figure 2(c),(f)). Catchments on the north-facing slope (W7 and W8) showed increases starting in approximately 2010. The WBR trends on the south facing catchments were more complicated, with notable drops from 2009 to 2013 in W3 and a 2009 drop in W6. After these drops, the WBR increased in recent years (2015–2019).

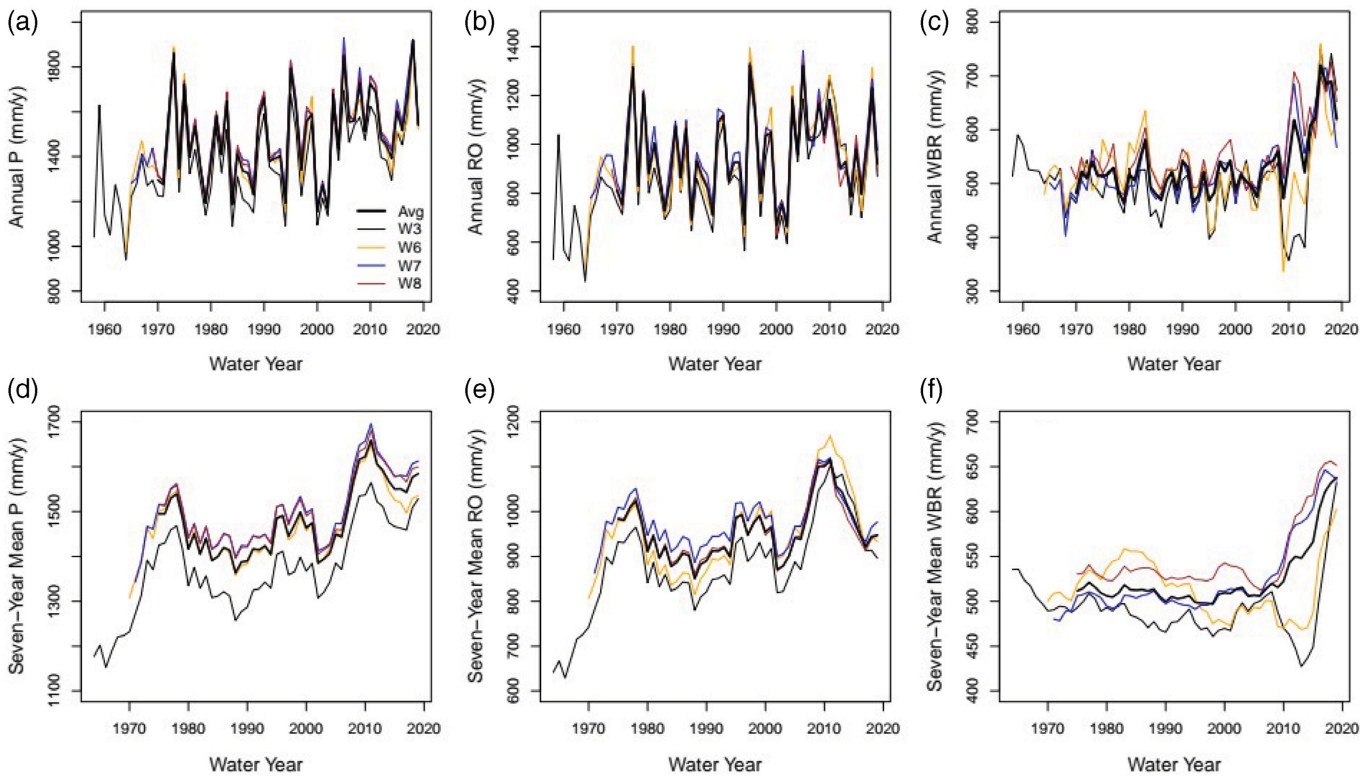

FIGURE 2 Annual precipitation (a), runoff (b), and water balance residual (c) in W3, W6, W7, and W8 over their measured record, including the area-weighted flux. Also shown is the seven-year average annual precipitation (d), runoff (e), and water balance residual (f)

TABLE 1 Slopes of the single mass curve for precipitation (P), runoff (RO), and the water balance residual (WBR) across the reference catchments at Hubbard Brook

Catchment	Period of record	Slope P (mm/year)	Slope RO (mm/year)	Slope WBR (mm/year)
W3	1958–2019	1367	880	487
W6	1964–2019	1449	947	512
W7	1965–2019	1493	977	517
W8	1969–2019	1494	954	540

Note: The slopes are for the entire record for each catchment, which varies across catchments. Thus, differences between catchments are due to site differences and the time period that their record covers.

Across all four catchments, the mean annual WBR prior to 2000 was 509 mm/year and increased to 667 mm/year from 2015 to 2019. The seven-year moving window water budgets showed a pattern that was similar to the annual water budgets. The WBR dip in the 2008–2013 period in the south-facing catchments was more visible with the inter-annual variation smoothed out (Figure 2(f)).

3.2 | Cumulative water budget deviations

The slopes of the cumulative summed P, RO and WBR approximated the long-term central tendency for each and are shown in Table 1. Amongst the four catchments, precipitation ranged from 1367 to 1494 mm/year, RO ranged from 880 to 977 mm/year and the WBR ranged from 487 to 540 mm/year.

Cumulative deviations in the four catchments showed clearer hydrologic changes than the annual sums. Precipitation and RO generally followed similar patterns across the four catchments, with apparent drying from the beginning of the record until February 1972, followed by transitions from drying to wetting occurring in March 1989, July 1995 and July 2003 (Figure 3). All four catchments received 2000–2700 mm of additional P between July 2003 and June 2020 with a pause in wetting between November 2012 and October 2016. Runoff deviations tracked P for most of the record in all four catchments, except since November 2012, when RO increased relative to P in W3 and W6, and decreased relative to P in W7 and W8.

The cumulative deviation of the WBR varied more across catchments than P and RO, except the 2015–2020 period showed an abrupt increase in all four. The W3 cumulative WBR deviation was relatively stable until an approximately 500 mm decrease between

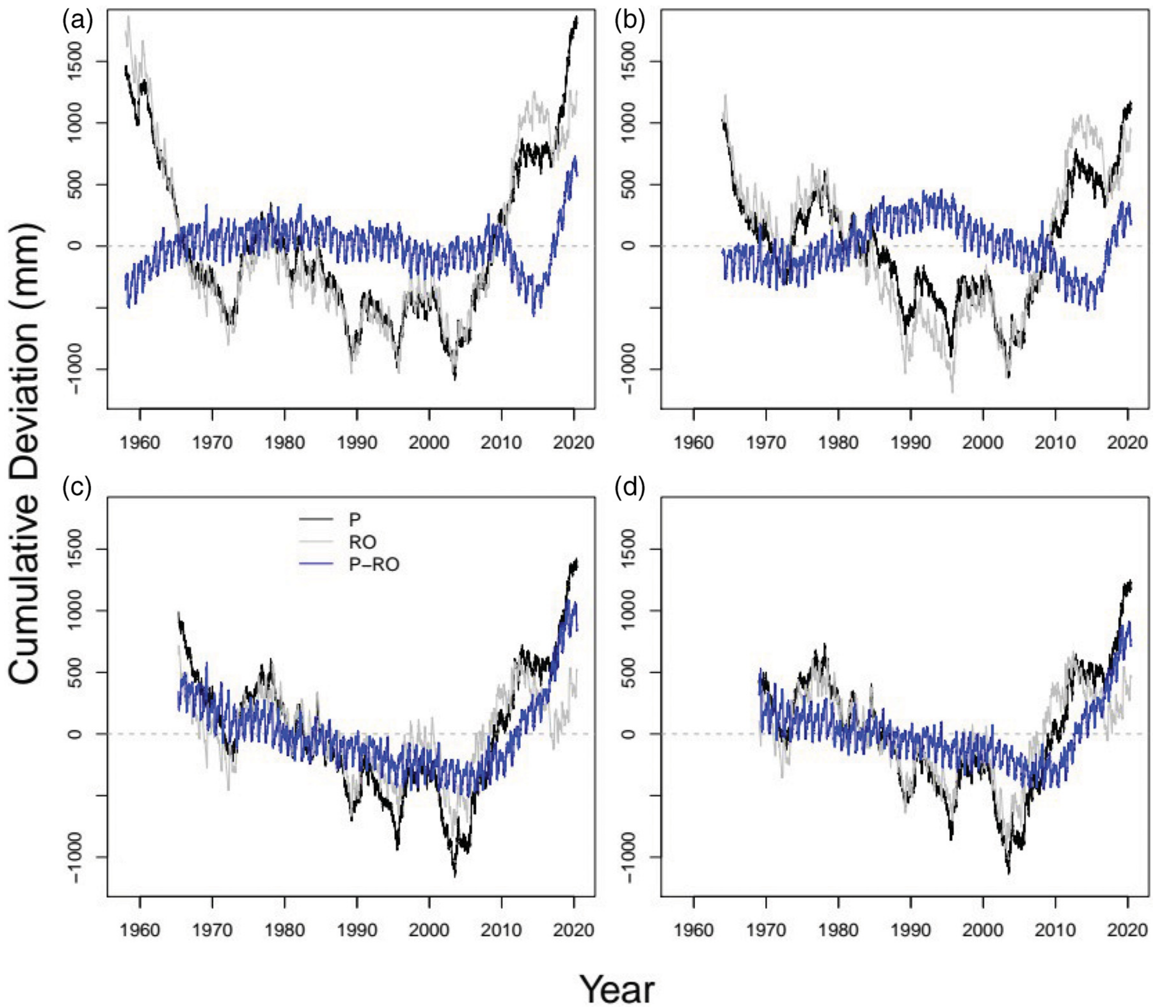

FIGURE 3 Cumulative deviations of precipitation (P), runoff (RO), and the water budget residual (P-RO) across the measured record at (a) W3, (b) W6, (c) W7, and (d) W8. The dashed reference line at 0 indicates the long-term central tendency, reported in Table 1

2009 and 2014 which reversed direction as a 1000 mm increase that continued through May 2020 (Figure 3(a)). The W6 cumulative WBR deviation was stable from 1963 before increasing from 1975 to 1992 followed by a long-term decrease until 2015 when it reversed upwards by about 800 mm (Figure 3(b)). Both W7 and W8 showed around 600–800 mm decreases in the cumulative WBR deviation from the beginning of their record until around 2008–2010 and then reversed for a 1200–1500 mm increase from 2010 to 2020.

3.3 | Potential ET and subsurface storage

Meteorological variables that drive PET all showed systematic changes from water years 1990 to 2019. Air T generally increased during this time period with an annual low in 1993 and high in 2015. (Figure 4(a)). Wind speed increased from 1990 to 2000 and then decreased through 2019 (Figure 4(b)). Net radiation minus ground heat flux decreased from 1990 to 2010 and then increased through 2019 (Figure 4(c)). Vapour pressure deficit increased from 1990 to 2019 with a notable low period from 2003 to 2010. Leaf area index showed the disturbance and recovery from the 1998 ice storm (Rhoads et al., 2002), followed by a subtle increase from 2000 to 2017 (Kendall $\tau = 0.31$, $p = 0.08$, Sen slope $= 0.044$ per year), although the LAI at the end of the time series was similar to the pre-ice storm LAI (Figure 5). The result of this trend was a 0.74 increase in LAI from 2000 to 2017. Mean annual PET was 771 mm/year, varying interannually around the mean from 1990 to 2010 (with a notable peak in the 2001 drought year) and then showed a step increase from water year 2011 to 2019 (Figure 6(a)). The PET cumulative deviation showed less systematic change from 1990 to 2010 followed by a cumulative increase of about 500 mm from 2011 to 2019 (Figure 6(b)).

The growing season length ranged from 110 to 148 days, showing substantial swings from year to year (Figure 7(a)). The south-facing site had a longer growing season than the north face (median of 139 days on the south face compared to 131 days on the north face). The growing season precipitation was similar on the two aspects, varying interannually with a series of high years from 2008 to 2015

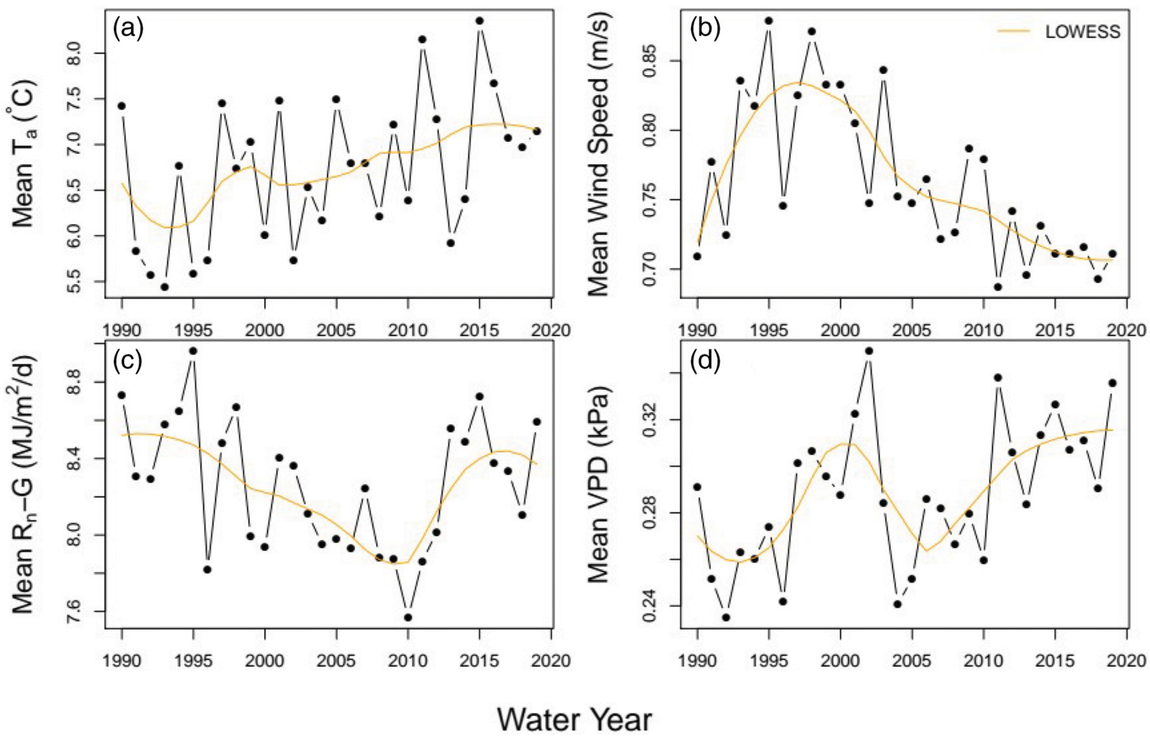

FIGURE 4 Time series of four meteorological variables used to calculate potential evapotranspiration: (a) air temperature, (b) wind speed, (c) net radiation minus soil heat flux, and (d) vapour pressure deficit

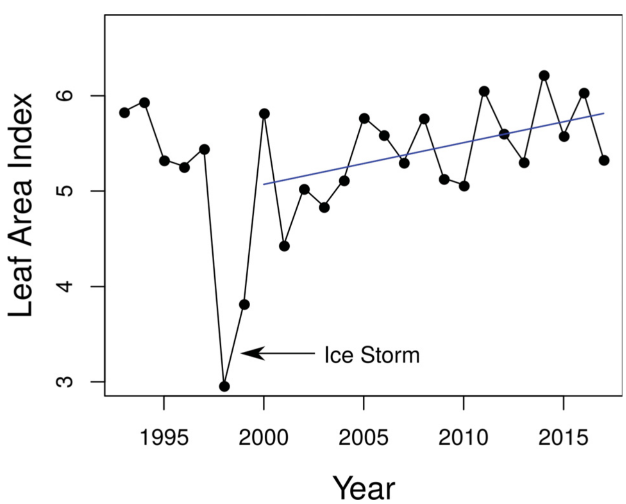

FIGURE 5 Mean annual leaf area index measured in from four broadleaf forest stands located west of W6. An ice storm that damaged the canopy in 1998 is noted. The blue line shows a best-fit line (Sen regression) for the 2000–2017 period following the ice storm

(Figure 7(b)). Growing season PET was stable and under 500 mm from 1990 to 2005 with a two-year peak in the drought years of 2000 and 2001, and then increased during the 2005–2019 period from near 500 to 600 mm (Figure 7(c)). The estimated growing season ET was consistently below 500 mm from 1990 to 2009, and then increased in stepwise fashion to about 540 mm with a notable low year of 2016 (Figure 7(d)).

Subsurface storage indicators increased from early in the records to 2010 followed by a decrease to the present (Figure 8). Low flows across the catchments increased in the south-facing catchments (W3 and W6) during the 2005–2015 period, but the north-facing catchments (W7 and W8) showed less evidence of this increase because the entire record had generally higher low flows. Soil moisture showed a clear downward trend from water year 2011 to 2017 with the lowest value occurring in 2016 (Figure 9). A similar decrease was apparent in the May and June groundwater table data with 2016 and 2018 being low years (Figure 10). These interannual changes in water table height were equivalent to a mean ΔSS from groundwater of −5.6 mm from 2013 to 2019 (range from 98 mm decrease from water year 2015 to 2016 to a 123 mm increase from water year 2018 to 2019). The total water storage deviation estimated by GRACE showed an intra-annual range of about 200 mm with little visible evidence of monotonic changes during the 2010–2019 period of catchment WBR increases (Figure S3).

4 | DISCUSSION

The increase in the WBR since 2010 across all four catchments is unprecedented in the 64-year HBEF record. Here, we demonstrate that this hydrologic change is real and not a measurement artefact. Then, we interpret the results and suggest that this change is mostly due to increasing ET rather than change in catchment storage. Finally, we discuss potential drivers that could cause increasing ET and account for the WBR change.

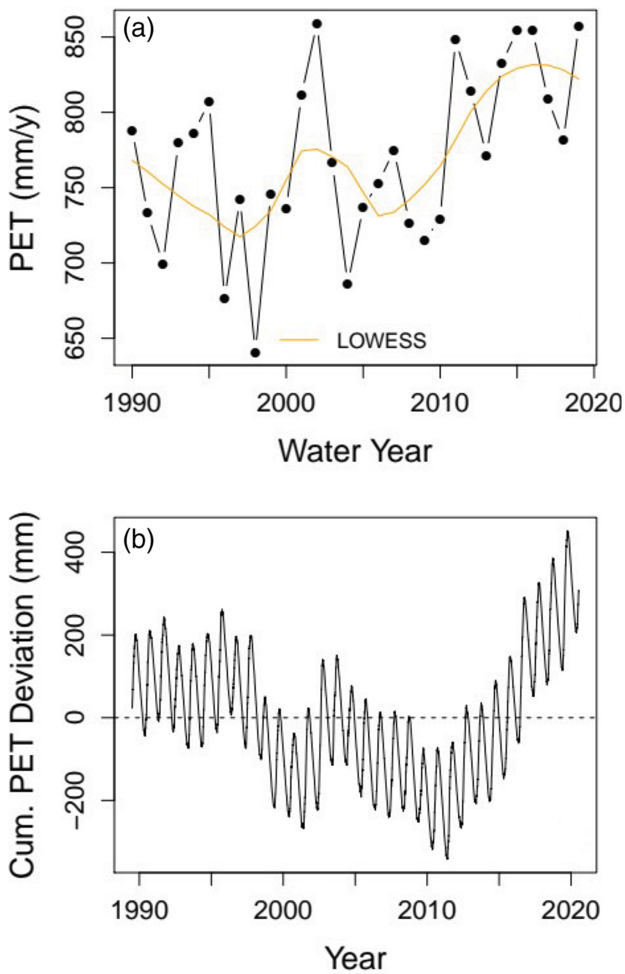

FIGURE 6 Annual potential evapotranspiration (PET) at the Hubbard Brook Experimental Forest long-term comprehensive met station, as estimated by the Penman-Monteith method. Panel A shows the sum per water year with a LOWESS smoother to illustrate the trend, and panel B shows the cumulative sum of the deviation from typical PET (766 mm/year indicated by the dashed line)

5 | HYDROLOGIC MEASUREMENTS DURING THE WBR INCREASE

Discharge and P measurements at the HBEF were modernized starting in 2012, and changes in P monitoring may have contributed to increases in the annual WBR. Analog P instruments were decommissioned in 2016 and replaced with digital weighing P gauges. Those gauges catch 2%–7% more P on average, based on a comparison of daily P estimates from co-located instruments between 2011 and 2013. This amount could increase the annual WBR by about 12% per year starting in June 2016 or cumulatively 252 mm from June 2016 to May 2020. This increase was substantially less than the observed WBR increase, which averaged 667 mm/year over that time, and therefore cannot explain the difference. In June 2016, the P network was also reduced from 23 to 9 P gauges and the spatial interpolation method was changed from the Thiessen polygon to the inverse distance weighting method (Green et al., 2018; USDA Forest Service, 2020a, 2020b). This transition caused a subtle decrease in P estimates for W6 and no systematic change in P estimates for W3, W7 and W8 (Supplemental results). The impact of the changed P estimate for W6 was a persistent 23 mm/year decrease in WBR, which would cause lower WBR estimates in W6. Thus, the impact would be the opposite of the increases in WBR that we observed and counteract the increase in WBR caused by the P gauge change. The net impact of P monitoring changes on W3, W7 and W8 would be increases in WBR after June 2016, but this increase was only about 12% of the increase in WRB (see the Supplemental results for additional details).

Discharge measurements over the entire record have involved measuring the height of water passing over a V-notch weir with a float and pulley placed in a stilling well. The data loggers used to record the water stage changed from analog to digital in 2012. Our independent hook-gauge and a second visual float/pulley stage measurements remained the same (e.g., Yanai et al., 2015), and used to adjust any minor biases in the recorded stage. We also note that the increase in WBR is not a step increase, as might be expected if this signal were due to a change in measurement methods, but rather a gradual increase over time. Thus, we conclude that the WBR increase was not due to any changes in discharge measurement techniques.

6 | EVIDENCE OF CHANGING STORAGE OR INCREASING EVAPOTRANSPIRATION

According to our water balance equation (Equation 1) and beyond any changes associated with P measurements, the additional increase in WBR was due to either an increase in ET or ΔSS or some combination of these two. Here, we examine whether ET or ΔSS was responsible for the WBR change by breaking the WBR change into three distinct periods of this hydrologic change: (1) a 2003–2009 wetting period, (2) a 2010–2014 pause in the wetting with differing catchment responses on the north and south aspects and (3) a 2015–2020 increase in WBR across all catchments (Table 2).

The catchments rewetted from 2003 to 2009 after the 2001 to 2002 dry period, resulting in favourable conditions for net accumulation of subsurface storage. Precipitation from 2003 to 2009 was high (mean area-weighted P = 1615 mm/year) and PET was relatively low (mean PET = 812 mm/year); thus, there was excess precipitation available to recharge groundwater or generate runoff. The R_n-G and *VPD* were particularly low during this period, driving the low PET (Figure 5). In response, the annual RO was the highest on record and the cumulative RO deviation went from being 1000 mm cumulatively lower than normal to 500–1000 mm higher than normal (Figures 2 and 3). The low flows in W3, W6 and W7 during this period also increased, reaching the highest on record in W3 and W6 in 2010 (Figure 8).

The 2010–2014 period was a transitional period with lower precipitation and a turning point from lower to higher PET. The mean P during this period was 1545 mm/year and mean PET was 874 mm/year; thus, there was less excess precipitation and RO was driven by a

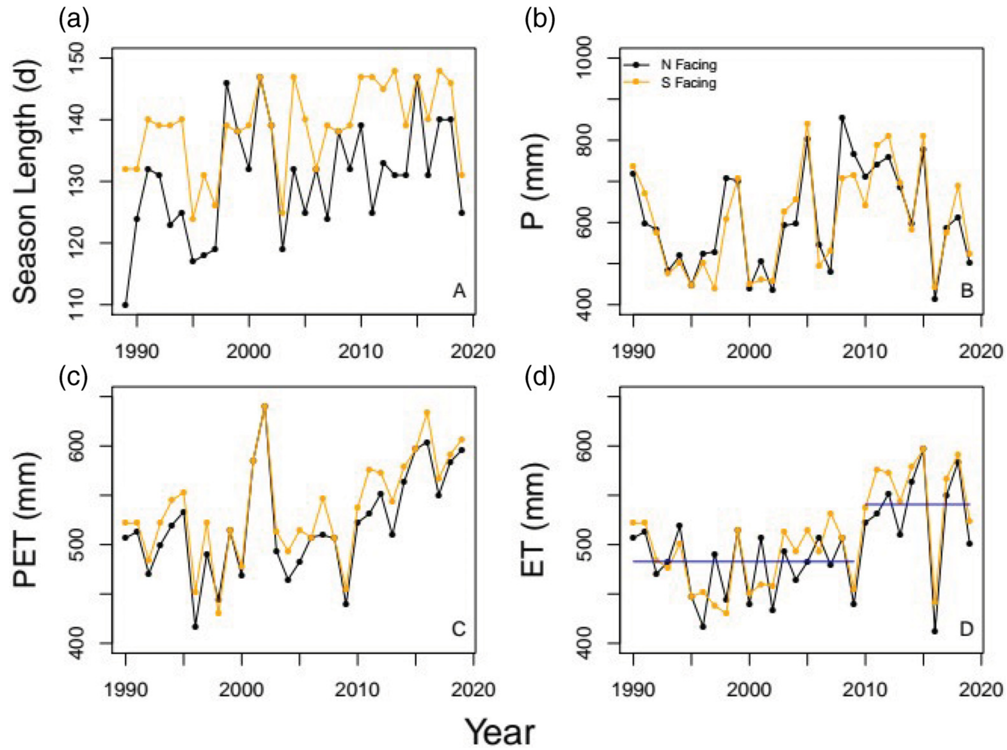

FIGURE 7 Growing season (a) length based on canopy phenology measurements, (b) precipitation (P) during the growing season, (c) potential evapotranspiration (PET) during the growing season, and (d) a simple estimate of growing season evapotranspiration (ET) assuming a balance between PET and P. North and south facing sites are compared; only phenology and precipitation data are collected on the two aspects. The median ET from 1990 to 2009 (median = 483 mm/year) and from 2010 to 2019 (median = 541 mm/year) are shown as solid blue lines in panel D

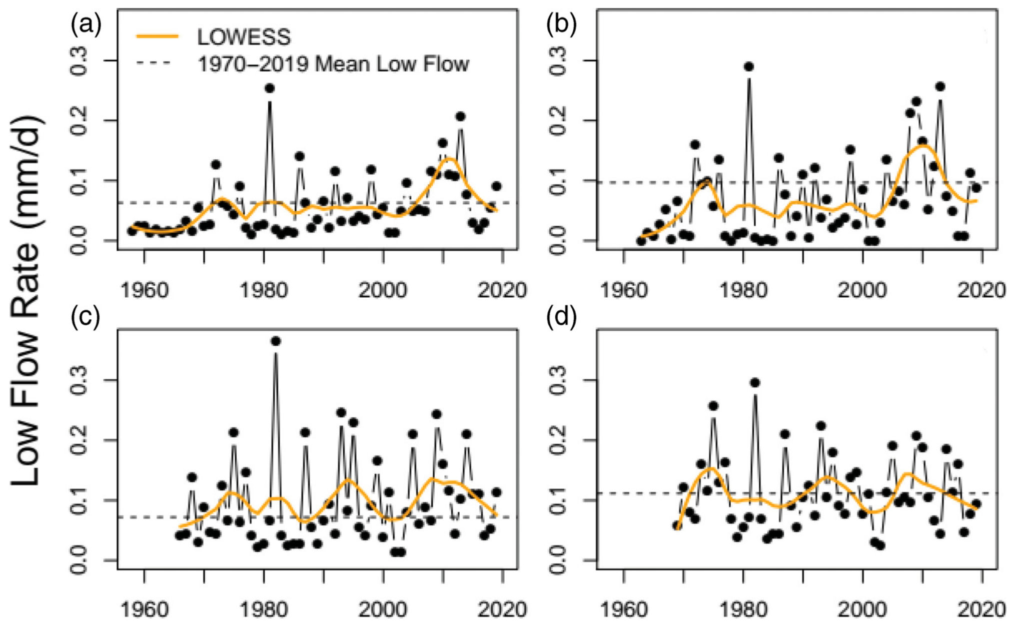

FIGURE 8 Time series of 5th percentile flow rates for (a) Watershed 3, (b) Watershed 6, (c) Watershed 7, and (d) Watershed 8. A LOWESS curve is shown to clarify multiple-year patterns

combination of ET and residual storage from the previous period. These conditions seem to have caused differential responses in the catchments on the two topographic aspects. In the case of W3 (and somewhat in W6), this resulted in a transition from high to low flows in 2010 to below-average low flows by 2014 (Figure 8) and low WBR from 2010 to 2014. In the case of W7 and W8, this period marked the beginning of increases in WBR (Figures 2 and 3). We hypothesize that the south-facing catchments were able to store more of the excess P from 2003 to 2009 and release it more slowly than the north-facing catchments. The south-facing catchments have a relatively narrow band of bedrock outcrops along their upper divides, whereas bedrock outcrops are more prevalent over a broader area in the upper portions of the north-facing catchments. The deepest glacial drift deposits that have been found in the south-facing catchments are in the upper portion of the catchments, in positions that are dominated by shallow bedrock in the north-facing catchments. This contrast suggests a difference in 'fill-and-spill' storage behaviour (cf. Tromp-van Meerveld & McDonnell, 2006) by topographic aspect with potentially greater storage in deeper deposits on the south facing compared to the north facing aspects of the Hubbard Brook valley.

Using a water isotope-calibrated hydrologic model for W3, Benettin et al. (2015) estimated deep groundwater storage of 760 mm, which suggests groundwater contributions to streamflow from glacial drift that is at least 3 m in depth. Additionally, we expect that ET may have started to increase in 2010 as evidenced by the increasing WBR in W7 and W8. However, we do not have sufficient data on the north-facing aspect to know the presence or magnitude of PET or LAI changes that might explain the slightly different WBR response.

After this 2010–2014 transitional period, all four catchments demonstrated drastic increases in WBR from water years 2015 to 2019, which we hypothesize was due to increased ET. This period had relatively high P (mean P = 1650 mm/year) and high PET (mean PET = 902 mm/year). The high PET during this period was due to warm T_a, high VPD and R_n relative to the beginning of the wetting in 2003 (Figure 4), resulting in an inflection point upwards in PET beginning in 2011 (Figure 6(b)). During this period, low flows were average or below average (Figure 8), and subsurface storage indicators were low (Figures 9 and 10), suggesting stable or decreasing catchment storage. The WBR of 667 mm/year during this period was 31% higher than the 1970–2002 area-weighted mean of 509 mm/year, which was likely due to high ET.

The overall picture for the 2003–2020 period is multiple years of increasing catchment storage as precipitation increased (Figures 3 and Figure S3), followed by years of high PET (Figure 6(b)), particularly during the growing season (Figure 7(c)), resulting in higher ET.

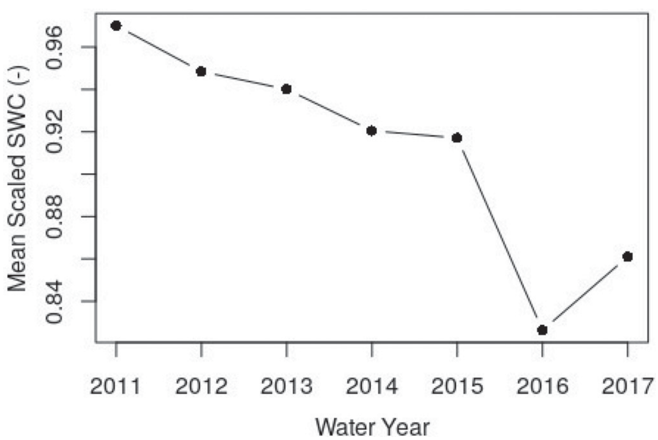

FIGURE 9 Mean annual scaled soil moisture across 14 locations within the HBEF valley

7 | POTENTIAL DRIVERS OF INCREASING EVAPOTRANSPIRATION

We propose that the increase in ET from 2010 to 2020 above any change associated with precipitation measurements was caused by both increased PET and reduced forest canopy resistance (r_c). The HBEF is generally considered energy limited because annual P > PET (Creed et al., 2014; Jones et al., 2012; Vadeboncoeur et al., 2018), so we might expect ET to track PET closely. The increase in growing season PET from about 500 mm/year to near 600 mm/year between 2007 and 2019 (Figure 7(c)) has similar timing and but slightly lower magnitude than the increase in WBR (Figure 2(f)). Also, we observed a positive cumulative PET deviation of about 500 mm from 2011 to 2020 (Figure 6(b)), compared to the approximately 800–1300 of cumulative WBR that occurred (Figure 3). If we recalculate PET without the LAI data, and instead use the long-term mean, the positive

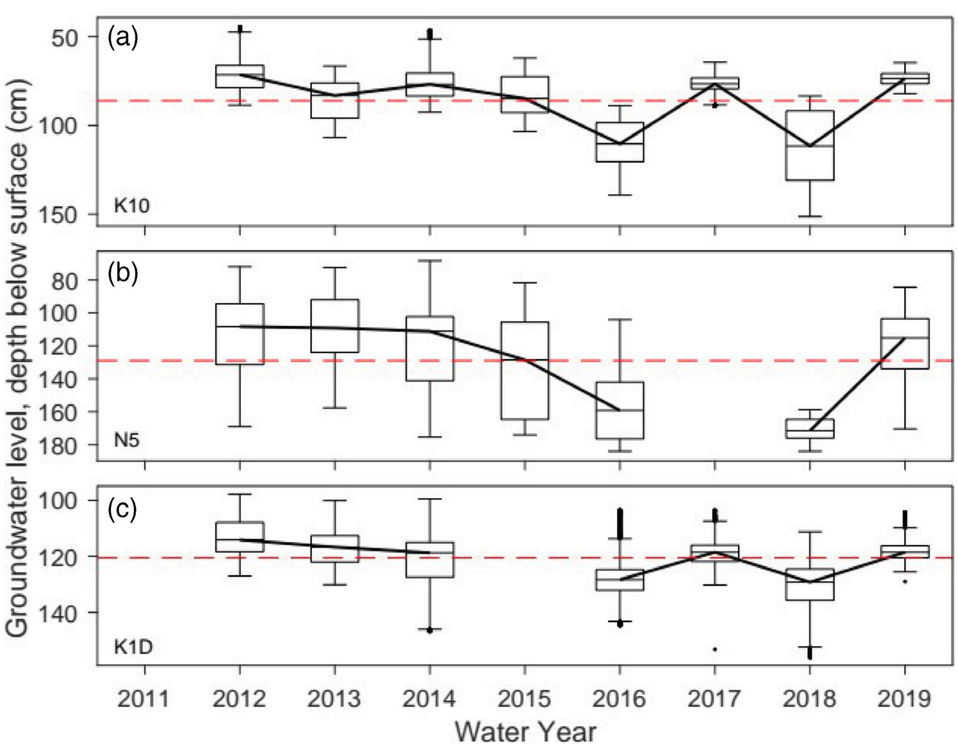

FIGURE 10 Boxplots of water table heights in May and June from 2012 to 2019 at sites (a) K10, (b) N5, and (c) K1D within the W3 catchment. The horizontal dashed line represents the multi-year mean of the median water table

TABLE 2 Annual precipitation (P), runoff (RO), and water budget residuals (WBR) for the four reference catchments and the area-weighted average of the four catchments during the 2003–2019 period of abrupt hydrologic change

	Precipitation (mm/year)					Runoff (mm/year)					WBR (mm/year)				
Water year	W3	W6	W7	W8	Avg	W3	W6	W7	W8	Avg	W3	W6	W7	W8	Avg
2003	1610	1691	1697	1702	1679	1127	1239	1196	1198	1184	484	451	501	504	495
2004	1383	1435	1446	1447	1432	929	984	957	970	957	455	451	489	477	475
2005	1695	1791	1928	1871	1849	1185	1267	1383	1352	1322	509	524	544	518	527
2006	1464	1555	1585	1577	1554	940	1050	1049	1025	1018	523	505	537	552	536
2007	1542	1610	1558	1560	1559	1020	1077	991	1032	1016	521	533	567	527	542
2008	1576	1654	1797	1760	1727	1035	1151	1226	1190	1168	540	502	571	570	559
2009	1474	1527	1522	1520	1511	1089	1191	1012	1004	1039	385	336	510	516	472
2010	1625	1726	1761	1755	1727	1268	1285	1144	1150	1183	357	441	616	605	544
2011	1579	1680	1720	1721	1687	1178	1159	1035	1014	1069	400	521	685	707	618
2012	1397	1473	1506	1512	1481	991	989	897	827	902	406	484	609	684	579
2013	1383	1436	1463	1483	1450	1002	975	907	894	929	381	462	556	588	521
2014	1294	1326	1427	1432	1392	711	778	823	790	785	582	548	604	642	607
2015	1513	1512	1651	1620	1602	883	895	1005	1036	980	630	617	646	585	622
2016	1451	1487	1541	1535	1515	700	728	830	836	796	751	759	711	699	719
2017	1595	1563	1743	1662	1673	924	933	1029	992	988	671	630	714	670	685
2018	1923	1903	1918	1924	1920	1182	1314	1267	1199	1230	741	589	651	725	689
2019	1541	1522	1543	1539	1540	868	906	976	887	920	673	616	567	652	620

cumulative deviation between 2011 and 2020 is about 250 mm (not shown), indicating that the 500 mm of PET that we can explain is approximately half from atmospheric demand (available energy and VPD) and half from LAI increasing. The remaining 300–800 mm of unexplained WBR, may be due to other factors related to r_c, such as forest community composition, structure or tree physiology change. Previous work has demonstrated how ET modeled with the Penman-Monteith equation (Equation 2) in cool temperate forests is highly sensitive to r_c (Beven, 1979). Specific physiological controls would be those that control photosynthetic rates (e.g., foliar nitrogen; Niinemets, 1997) or other processes that impact stomatal function (e.g., xylem embolism; Domec et al., 2004). The WBR at the HBEF has shown sensitivity to vegetation changes in the past. Green et al. (2013) showed a multi-year WBR increase after a soil amendment with calcium silicate that altered canopy albedo, which is related to the maximum photosynthetic rate (Ollinger et al., 2008). Transpiration rates could be increasing as forest soils in the northeastern United States experience increases in soil pH and Ca, and associated decreases in Al, as atmospheric acid deposition rates have declined (Hazlett et al., 2020; Lawrence et al., 2015). Interception rates may also be increasing as red spruce increases its presence in our region over recent decades (Foster & D'Amato, 2015; Kosiba et al., 2018) and the number of days per year that deciduous trees have leaves has increased (Groffman et al., 2012). The area weighted WBR (WBR') compared to annual PET suggests that the WBR':PET was high in water years 2016, 2017 and 2018 (Figure 11), indicating a possible reduction in r_c over these years. We hypothesize that a reduction in r_c in the 2010–2020 period has led to increases in ET. The cause of

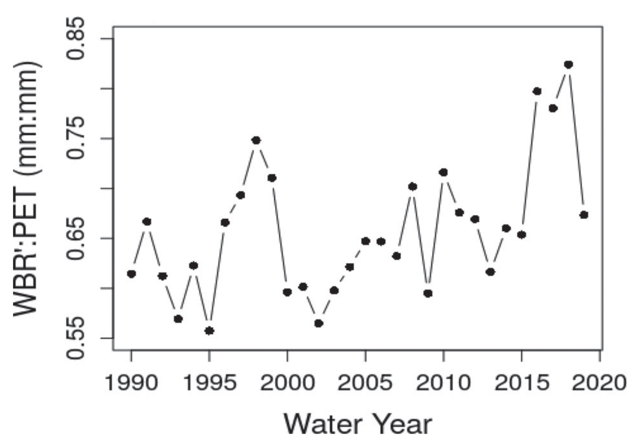

FIGURE 11 Time series of the area weighted water budget residual (WBR') to potential evapotranspiration (PET) ratio from water year 1990 to 2019

decreased r_c will require further synthesis of existing vegetation data from HBEF as well as more characterization of physiological changes that might have occurred in the forest.

Beyond the increase in ET, some of the WBR increase is likely due to ΔSS. Previous studies in the northeastern United States have documented increases in storage (along the order of 100 mm) based on GRACE and discharge analyzes (Thomas et al., 2016; Thomas & Famiglietti, 2019). More comprehensive observations of ΔSS are needed to be more confident as we attribute changes in WBR to different processes. The HBEF had previously documented minimal loss to deep seepage (L) on annual water budgets (Verry, 2003). Our study suggests that ΔSS was a small part of the annual water budgets, based on dynamics in soil moisture from the uppermost part of the soil profile and variation in shallow water table position. However, preliminary ground penetrating radar surveys, confirmed by soil pits and monitoring well installations, show that portions of these study catchments have much deeper and more permeable glacial deposits than previously recognized. The extent of T perching of the shallow groundwater system in the soil zone, and dynamics in its recharge to deeper groundwater storage in areas of thicker glacial deposits, are poorly known. Better understanding of the groundwater system, with improved estimates of ΔSS, would allow us to more accurately estimate the magnitude of changes in ET. Recent studies have highlighted the importance in estimating ΔSS across many catchments and concluded that while ΔSS may be small in some contexts, it is not negligible and is related to catchment characteristics (Han et al., 2020; Rice & Emanuel, 2019). As ET is changing with increasing atmospheric CO_2 and T_a, small catchment studies will benefit from more comprehensive measurements, including greater attention to ΔSS and subsurface characterization, to help isolate hydrologic changes and their likely causes. While we have a wealth of hydrological, meteorological, geological and vegetation measurements at the HBEF, we still lack enough information to explain differences in PET and hydrogeology across topographic aspects. Better measurements of ΔSS at the HBEF, particularly of the deeper groundwater storage, will help refine our understanding of ΔSS and improve our estimation of catchment-scale ET.

While we strive for more comprehensive monitoring of the catchments at the HBEF, the existing long-term data enabled documentation of this hydrologic change, highlighting the importance of long-term catchment studies. It is only in the context of long-term observations that we can see how the WBR values from 2015 to 2019 are highly unusual and warrant attention. If this is an approximately 30% increase in ET as we hypothesize, we need to understand its cause more definitively and determine the spatial extent of its occurrence. Increases in ET have been documented and are expected with climate change (Brutsaert, 2017; Gaertner et al., 2019; Pascolini-Campbell et al., 2021; Tsuruta et al., 2020; Younger et al., 2020). However, observed changes have been inconsistent and show both increases and decreases within the northeastern US forest region (Vadeboncoeur et al., 2018) even though at continental scales mostly increases have been observed (Szilagyi et al., 2001; Walter et al., 2004). Small catchments with comprehensive hydrological records are vital for documenting changes in ET and accurately attributing those changes to specific drivers.

ACKNOWLEDGEMENTS

We thank two anonymous reviewers who helped improve this manuscript. We also thank Cindy Wood and Natalie Cleavitt for help producing the leaf area index data, and Ian Halm, Hannah Vollmer and Gabe Winant for help producing the hydrologic data. The Hubbard Brook Experimental Forest is operated and maintained by the USDA Forest Service, Northern Research Station, Madison, WI (1014507, 1114804).

DATA AVAILABILITY STATEMENT

Precipitation and discharge data that support the findings of this study are openly available in the Environmental Data Initiative at https://doi.org/10.6073/pasta/87584eda806dd5a480423b6bfefec577 and https://doi.org/10.6073/pasta/c64ad38eef4f56d9e34749f166f64caa. Meteorological, leaf area index, canopy phenology, soil moisture, and water table data that support the findings of this study are openly available in the Environmental Data Initiative at https://doi.org/10.6073/pasta/7486a33ab8549c262233ad3e4a8b42a3, https://doi.org/10.6073/pasta/3958640a5f5ed3af7b5e40a5cc710b40, https://doi.org/10.6073/pasta/3511ed3f4a50ee86fddb3fbf8b42ccd5, https://doi.org/10.6073/pasta/f2c18a955c24eadaec1fa0d915a7b527, https://doi.org/10.6073/pasta/e7c793b98b895de2bb5e505f9ff5e0cb, https://doi.org/10.6073/pasta/e7c793b98b895de2bb5e505f9ff5e0cb, https://doi.org/10.6073/pasta/e6ca833db8b6a4931ab9fafb91191d38, and https://doi.org/10.6073/pasta/a7b6b61df98b65244eba64d8bc391582.

ORCID

Mark B. Green https://orcid.org/0000-0002-7415-7209
Scott W. Bailey https://orcid.org/0000-0002-9160-156X
John L. Campbell https://orcid.org/0000-0003-4956-1696

REFERENCES

Allen, R. G., Pereira, L. S., Raes, D., & Smith, M. (1998). Crop evapotranspiration-guidelines for computing crop water requirements-FAO irrigation and drainage paper 56, Fao, Rome, 300(9), D05109.

Bailey, A. S., Hornbeck, J. W., Campbell, J. L., & Eagar, C. (2003). Hydrometeorological database for Hubbard Brook Experimental Forest: 1955-2000. *U.S. Forest Service General Technical Report-305*, 36.

Bailey, S. W., Brousseau, P. A., McGuire, K. J., & Ross, D. S. (2014). Influence of landscape position and transient water table on soil development and carbon distribution in a steep, headwater catchment. *Geoderma*, *226–227*, 279–289.

Bailey, S. W., Hornbeck, J. W., Driscoll, C. T., & Gaudette, H. E. (1996). Calcium inputs and transport in a base-poor forest ecosystem as interpreted by Sr isotopes. *Water Resources Research*, *32*(3), 707–719.

Battles, J. J., Fahey, T. J., Driscoll, C. T., Blum, J. D., & Johnson, C. E. (2014). Restoring soil calcium reverses forest decline. *Environmental Science & Technology Letters*, *1*, 15–19.

Benettin, P., Bailey, S. W., Campbell, J. L., Green, M. B., Rinaldo, A., Likens, G. E., McGuire, K. J., & Botter, G. (2015). Linking water age and solute dynamics in streamflow at the Hubbard Brook Experimental Forest, NH, USA. *Water Resources Research*, *51*(11), 9256–9272.

Beven, K. (1979). A sensitivity analysis of the Penman-Monteith actual evapotranspiration estimates. *Journal of Hydrology*, *44*(3–4), 169–190.

Briceño, H., Miller, G., & Davis, S. E. (2014). Relating Freshwater Flow with Estuarine Water Quality in the Southern Everglades Mangrove Ecotone. *Wetlands*, *34*, 101–111.

Brutsaert, W. (2008). Long-term groundwater storage trends estimated from streamflow records: Climatic perspective. *Water Resources Research*, *44*(2), W02409.

Brutsaert, W. (2017). Global land surface evaporation trend during the past half century: Corroboration by Clausius-Clapeyron scaling. *Advances in Water Resources*, *106*, 3–5.

Burton, W. C., Walsh, G. J., & Armstrong, T. R. (2002). *Bedrock geologic map of the Hubbard Brook Experimental Forest, Grafton County, New Hampshire*, U.S. Geological Survey Open-File Report 00-045, Reston, VA: U.S. Department of the Interior.

Byrne, M. P., & O'Gorman, P. A. (2018). Trends in continental temperature and humidity directly linked to ocean warming. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(19), 4863–4868.

Campbell, J. L., Driscoll, C. T., Pourmokhtarian, A., & Hayhoe, K. (2011). Streamflow responses to past and projected future changes in climate at the Hubbard Brook Experimental Forest, New Hampshire, United States. *Water Resources Research*, *47*(2), W02514.

Campbell, J. L., Ollinger, S. V., Flerchinger, G. N., Wicklein, H., Hayhoe, K., & Bailey, A. S. (2010). Past and projected future changes in snowpack and soil frost at the Hubbard Brook Experimental Forest, New Hampshire, USA. *Hydrological Processes*, *24*(17), 2465–2480.

Campbell, J. L., Rustad, L. E., Bailey, S. W., Bernhardt, E. S., Driscoll, C. T., Green, M. B., et al. (2021). Watershed studies at the Hubbard Brook Experimental Forest: Building on a long legacy of research with new approaches and sources of data. *Hydrological Processes*, *35*(1), e14016.

Cluis, D. A. (1983). Visual techniques for the detection of water quality trends: Double-mass curves and cusum functions. *Environmental Monitoring and Assessment*, *3*, 173–184.

Creed, I. F., Spargo, A. T., Jones, J. A., Buttle, J. M., Adams, M. B., Beall, F. D., … Green, M. B. (2014). Changing forest water yields in response to climate warming: Results from long-term experimental watershed sites across North America. *Global Change Biology*, *20*(10), 3191–3208.

Detty, J. M., & McGuire, K. J. (2010). Threshold changes in storm runoff generation at a till-mantled headwater catchment. *Water Resources Research*, *46*(7), W07525.

Domec, J. C., Warren, J. M., Meinzer, F. C., Brooks, J. R., & Coulombe, R. (2004). Native root xylem embolism and stomatal closure in stands of Douglas-fir and ponderosa pine: Mitigation by hydraulic redistribution. *Oecologia*, *141*(1), 7–16.

Fahey, T., & Cleavitt, N. (2021). Hubbard Brook Experimental Forest: Leaf Area Index (LAI) Bear Brook Watershed (West of Watershed 6) ver 1. *Environmental Data Initiative*. https://doi.org/10.6073/pasta/e6ca833db8b6a4931ab9fafb91191d38 .

Federer, C. A. (1973). Forest transpiration greatly speeds streamflow recession. *Water Resources Research*, *9*(6), 1599–1604.

Federer, C. A. (1990). Change, persistence and error in thirty years of hydrometeorological data at Hubbard Brook. Climate variability and ecosystem response. USDA For. Serv. Gen. Tech. Rep. SE-65, 3–12.

Foster, J. R., & D'Amato, A. W. (2015). Montane forest ecotones moved downslope in northeastern USA in spite of warming between 1984 and 2011. *Global Change Biology*, *21*(12), 4497–4507.

Fuss, C. B., Driscoll, C. T., Green, M. B., & Groffman, P. M. (2016). Hydrologic flowpaths during snowmelt in forested headwater catchments under differing winter climatic and soil frost regimes. *Hydrological Processes*, *30*(24), 4617–4632.

Gaertner, B. A., Zegre, N., Warner, T., Fernandez, R., He, Y., & Merriam, E. R. (2019). Climate, forest growing season, and evapotranspiration changes in the central Appalachian Mountains, USA. *Science of the Total Environment*, *650*, 1371–1381.

Gannon, J. P., Bailey, S. W., & McGuire, K. J. (2014). Organizing groundwater regimes and response thresholds by soils: A framework for understanding runoff generation in a headwater catchment. *Water Resources Research*, *50*(11), 8403–8419.

Gannon, J. P., McGuire, K. J., Bailey, S. W., Bourgault, R. R., & Ross, D. S. (2017). Lateral water flux in the unsaturated zone: A mechanism for the formation of spatial soil heterogeneity in a headwater catchment. *Hydrological Processes*, *31*(20), 3568–3579.

Germann, P. F., Pierce, R. S., & Beven, K. (1986). Kinematic wave approximation to the initiation of subsurface storm flow in a sloping forest soil. *Advances in Water Resources*, *9*(2), 70–76.

Green, M. B., Bailey, A. S., Bailey, S. W., Battles, J. J., Campbell, J. L., Driscoll, C. T., Fahey, T. J., Lepine, L. C., Likens, G. E., Ollinger, S. V., & Schaberg, P. G. (2013). Decreased water flowing from a forest amended with calcium silicate. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(15), 5999–6003.

Green, M. B., Campbell, J. L., Yanai, R. D., Bailey, S. W., Bailey, A. S., Grant, N., Halm, I., Kelsey, E. P., & Rustad, L. E. (2018). Downsizing a long-term precipitation network: Using a quantitative approach to inform difficult decisions. *PLoS One*, *13*(5), e0195966.

Green, M. B., Laursen, B. K., Campbell, J. L., McGuire, K. J., & Kelsey, E. P. (2015). Stable water isotopes suggest sub-canopy water recycling in a northern forested catchment. *Hydrological Processes*, *29*(25), 5193–5202.

Groffman, P. (2019). Snow depth, soil frost depth and snow water content along an elevation gradient at the Hubbard Brook Experimental Forest vers 5. *Environmental Data Initiative*. https://doi.org/10.6073/pasta/3958640a5f5ed3af7b5e40a5cc710b40

Groffman, P. M., Rustad, L. E., Templer, P. H., Campbell, J. L., Christenson, L. M., Lany, N. K., … Driscoll, C. T. (2012). Long-term integrated studies show complex and surprising effects of climate change in the northern hardwood forest. *Bioscience*, *62*(12), 1056–1066.

Han, J., Yang, Y., Roderick, M. L., McVicar, T. R., Yang, D., Zhang, S., & Beck, H. E. (2020). Assessing the steady-state assumption in water balance calculation across global catchments. *Water Resources Research*, *56*, e2020WR027392. https://doi.org/10.1029/2020WR027392

Hawkins, D. M., & Olwell, D. H. (2012). *Cumulative sum charts and charting for quality improvement*. Springer Science & Business Media.

Hazlett, P., Emilson, C., Lawrence, G., Fernandez, I., Ouimet, R., & Bailey, S. (2020). Reversal of forest soil acidification in the northeastern United States and Eastern Canada: Site and soil factors contributing to recovery. *Soil Systems*, *4*(3), 54.

Hinkley, D. V. (1971). Inference about the change-point from cumulative sum tests. *Biometrika*, *58*, 509–523.

Hogan, J. F., & Blum, J. D. (2003). Tracing hydrologic flow paths in a small forested watershed using variations in 87Sr/86Sr,[Ca]/[Sr], [Ba]/[Sr] and δ18O. *Water Resources Research*, *39*(10), 1282.

Hooper, R. P., & Shoemaker, C. A. (1986). A comparison of chemical and isotopic hydrograph separation. *Water Resources Research*, *22*(10), 1444–1454.

Hornbeck, J. W., Martin, C. W., & Eagar, C. (1997). Summary of water yield experiments at Hubbard Brook Experimental Forest, New Hampshire. *Canadian Journal of Forest Research*, *27*, 2043–2052.

Huang, H., Winter, J. M., Osterberg, E. C., Horton, R. M., & Beckage, B. (2017). Total and extreme precipitation changes over the northeastern United States. *Journal of Hydrometeorology*, *18*(6), 1783–1798.

Jensen, C. K., McGuire, K. J., & Prince, P. S. (2017). Headwater stream length dynamics across four physiographic provinces of the Appalachian Highlands. *Hydrological Processes*, *31*(19), 3350–3363.

Jones, J. A., Creed, I. F., Hatcher, K. L., Warren, R. J., Adams, M. B., Benson, M. H., … Clow, D. W. (2012). Ecosystem processes and human influences regulate streamflow response to climate change at long-term ecological research sites. *Bioscience*, *62*(4), 390–404.

Kelly, C. N., McGuire, K. J., Miniat, C. F., & Vose, J. M. (2016). Streamflow response to increasing precipitation extremes altered by forest management. *Geophysical Research Letters*, *43*(8), 3727–3736.

Kelsey, E., Green, M., & Evans, D. (2019). Hubbard Brook Experimental Forest: Flux Tower Data ver 2. *Environmental Data Initiative*. https://doi.org/10.6073/pasta/19bb57896cbdb686f93049b9b2d95c09

Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, *30* (1/2), 81–93.

Kirchner, J. W., & Allen, S. T. (2020). Seasonal partitioning of precipitation between streamflow and evapotranspiration, inferred from end-member splitting analysis. *Hydrology & Earth System Sciences*, *24*(1), 17–39.

Klemeš, V. (1979). Storage mass-curve analysis in a systems-analytic perspective. *Water Resources Research*, *15*, 359–370.

Kosiba, A. M., Schaberg, P. G., Rayback, S. A., & Hawley, G. J. (2018). The surprising recovery of red spruce growth shows links to decreased acid deposition and elevated temperature. *Science of the Total Environment*, *637*, 1480–1491.

Kramer, R. J., Bounoua, L., Zhang, P., Wolfe, R. E., Huntington, T. G., Imhoff, M. L., Thome, K., & Noyce, G. L. (2015). Evapotranspiration trends over the eastern United States during the 20th century. *Hydrology*, *2*, 93–111.

Landerer, F. W., Flechtner, F. M., Save, H., Webb, F. H., Bandikova, T., Bertiger, W. I., … Yuan, D. N. (2020). Extending the global mass change data record: GRACE Follow-On instrument and science data performance. *Geophysical Research Letters*, *47*(12), e2020GL088306.

Lawrence, G. B., Hazlett, P. W., Fernandez, I. J., Ouimet, R., Bailey, S. W., Shortle, W. C., Smith, K. T., & Antidormi, M. R. (2015). Declining acidic deposition begins reversal of forest-soil acidification in the northeastern US and eastern Canada. *Environmental Science & Technology*, *49*(22), 13103–13111.

Leonard, R. E., & Reinhart, K. G. (1963). Some observations on precipitation measurement on forested experimental watersheds. Research Note NE-6. Upper Darby, PA: US Department of Agriculture, Forest Service, Northeastern Forest Experiment Station. 1–4, 6.

Likens, G. E. (2013). *Biogeochemistry of a forested ecosystem*. Springer Science & Business Media.

Likens, G. E., Driscoll, C. T., & Buso, D. C. (1996). Long-term effects of acid rain: response and recovery of a forest ecosystem. *Science*, *272*(5259), 244–246.

Lynch, J. A., & Corbett, E. S. (1990). Evaluation of best management practices for controlling nonpoint pollution from silvicultural operations. *Journal of the American Water Resources Association*, *26*(1), 41–52.

McGuire, K., Bailey, S., & Gannon, J. (2019). Water level recordings from wells in Watershed 3 at the Hubbard Brook Experimental Forest, 2007 - present ver 3. *Environmental Data Initiative*. https://doi.org/10.6073/pasta/a7b6b61df98b65244eba64d8bc391582

McNamara, J. P., Tetzlaff, D., Bishop, K., Soulsby, C., Seyfried, M., Peters, N. E., Aulenbach, B. T., & Hooper, R. (2011). Storage as a metric of catchment comparison. *Hydrological Processes*, *25*(21), 3364–3371.

McVicar, T. R., Roderick, M. L., Donohue, R. J., Li, L. T., Van Niel, T. G., Thomas, A., … Mescherskaya, A. V. (2012). Global review and synthesis of trends in observed terrestrial near-surface wind speeds: Implications for evaporation. *Journal of Hydrology*, *416*, 182–205.

Niinemets, Ü. (1997). Role of foliar nitrogen in light harvesting and shade tolerance of four temperate deciduous woody species. *Functional Ecology*, *11*(4), 518–531.

Nijzink, R., Hutton, C., Pechlivanidis, I., Capell, R., Arheimer, B., Freer, J., Han, D., Wagener, T., McGuire, K., Savenije, H., & Hrachowitz, M. (2016). The evolution of root-zone moisture capacities after deforestation: A step towards hydrological predictions under change? *Hydrology and Earth System Sciences*, *20*, 4775–4799.

O'Gorman, P. A., & Schneider, T. (2009). The physical basis for increases in precipitation extremes in simulations of 21st-century climate change. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(35), 14773–14777.

Ollinger, S. V., Richardson, A. D., Martin, M. E., Hollinger, D. Y., Frolking, S. E., Reich, P. B., … Smith, M. L. (2008). Canopy nitrogen, carbon assimilation, and albedo in temperate and boreal forests: Functional relations and potential climate feedbacks. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(49), 19336–19341.

Pascolini-Campbell, M., Reager, J. T., Chandanpurkar, H. A., & Rodell, M. (2021). A 10 per cent increase in global land evapotranspiration from 2003 to 2019. *Nature*, *593*(7860), 543–547.

Patric, J. H., & Reinhart, K. G. (1971). Hydrologic effects of deforesting two mountain watersheds in West Virginia. *Water Resources Research*, *7*(5), 1182–1188.

Regier, P., Briceño, H., & Boyer, J. N. (2019). Analyzing and comparing complex environmental time series using a cumulative sums approach. *MethodsX*, *6*, 779–787.

Reinhart, K. G., & Pierce, R. S. (1964). *Stream-gaging stations for research on small watersheds (No. 268)*. US Department of Agriculture.

Rhoads, A. G., Hamburg, S. P., Fahey, T. J., Siccama, T. G., Hane, E. N., Battles, J., Cogbill, C., Randall, J., & Wilson, G. (2002). Effects of an intense ice storm on the structure of a northern hardwood forest. *Canadian Journal of Forest Research*, *32*(10), 1763–1775.

Rice, J. S., & Emanuel, R. E. (2019). Ecohydrology of interannual changes in watershed storage. *Water Resources Research*, *55*, 8238–8251. https://doi.org/10.1029/2019WR025164

Richardson, A. D., Bailey, A. S., Denny, E. G., Martin, C. W., & O'Keefe, J. (2006). Phenology of a northern hardwood forest canopy. *Global Change Biology*, *12*(7), 1174–1188.

Searcy, J. K., & Hardison, C.H. (1960). Double-mass curves. Geological Survey Water Supply Paper 1541-B. United States Government Printing Office, Washington.

See, C. R., Green, M. B., Yanai, R. D., Bailey, A. S., Campbell, J. L., & Hayward, J. (2020). Quantifying uncertainty in annual runoff due to missing data. *PeerJ*, *8*, e9531.

Sen, P. K. (1968). Estimates of the regression coefficient based on Kendall's tau. *Journal of the American Statistical Association*, *63*(324), 1379–1389.

Siccama, T. G., Fahey, T. J., Johnson, C. E., Sherry, T. W., Denny, E. G., Girdler, E. B., Likens, G. E., & Schwarz, P. A. (2007). Population and biomass dynamics of trees in a northern hardwood forest at Hubbard Brook. *Canadian Journal of Forest Research*, *37*, 737–749.

Smail, R. A., Pruitt, A. H., Mitchell, P. D., & Colquhoun, J. B. (2019). Cumulative deviation from moving mean precipitation as a proxy for groundwater level variation in Wisconsin. *Journal of Hydrology X*, *5*, 100045.

Szilagyi, J., Katul, G. G., & Parlange, M. B. (2001). Evapotranspiration intensifies over the conterminous United States. *Journal of Water Resources Planning and Management*, *127*(6), 354–362.

Thomas, B. F., & Famiglietti, J. S. (2019). Identifying climate-induced groundwater depletion in GRACE observations. *Scientific Reports*, *9*(1), 1–9.

Thomas, B. F., Landerer, F. W., Wiese, D. N., & Famiglietti, J. S. (2016). A comparison of watershed storage trends over the eastern and upper Midwestern regions of the United States, 2003–2015. *Water Resources Research*, *52*(8), 6335–6347.

Trenberth, K. E. (2011). Changes in precipitation with climate change. *Climate Research*, *47*(1–2), 123–138.

Tromp-van Meerveld, H. J., & McDonnell, J. J. (2006). Threshold relations in subsurface stormflow: 2. The fill and spill hypothesis. *Water Resources Research*, *42*(2), W02411.

Tsuruta, K., Kosugi, Y., Katsuyama, M., Kosugi, K. I., Suzuki, M., & Tani, M. (2020). Long-term effects of evapotranspiration on the flow duration curve in a coniferous plantation forest over 40 years. *Hydrological Research Letters*, *14*(1), 1–8.

USDA Forest Service, Northern Research Station. (2020a). Hubbard Brook Experimental Forest: Total Daily Precipitation by Watershed, 1956 - present ver 12. *Environmental Data Initiative*. https://doi.org/10.6073/pasta/3d093a09034c4863b3834f5db58fc0b2 .

USDA Forest Service, Northern Research Station. (2020b). Hubbard Brook Experimental Forest: Daily Streamflow by Watershed, 1956 - present ver 11. *Environmental Data Initiative*. https://doi.org/10.6073/pasta/d10220595119b502fe2ac14833fa4b9b

Vadeboncoeur, M. A., Green, M. B., Asbjornsen, H., Campbell, J. L., Adams, M. B., Boyer, E. W., Burns, D. A., Fernandez, I. J., Mitchell, M. J., & Shanley, J. B. (2018). Systematic variation in evapotranspiration trends and drivers across the northeastern United States. *Hydrological Processes*, *32*(23), 3547–3560.

Van Doorn, N. S., Battles, J. J., Fahey, T. J., Siccama, T. G., & Schwarz, P. A. (2011). Links between biomass and tree demography in a northern hardwood forest: a decade of stability and change in Hubbard Brook Valley, New Hampshire. *Canadian Journal of Forest Research*, *41*(7), 1369–1379.

Verry, E. S. (2003). Estimating ground water yield in small research basins. *Groundwater*, *41*(7), 1001–1004.
Walter, M. T., Wilks, D. S., Parlange, J. Y., & Schneider, R. L. (2004). Increasing evapotranspiration from the conterminous United States. *Journal of Hydrometeorology*, *5*(3), 405–408.
Wilson, G., Green, M., Brown, J., Campbell, J., Groffman, P., Duran, J., & Morse, J. (2020). (in revision). Snowpack affects soil microclimate throughout the year. *Climatic Change*, *163*(2), 705–722.
Xu, H., Xiao, J., & Zhang, Z. (2020). Heatwave effects on gross primary production of northern mid-latitude ecosystems. *Environmental Research Letters*, *15*(7), 074027.
Yanai, R. D., Tokuchi, N., Campbell, J. L., Green, M. B., Matsuzaki, E., Laseter, S. N., Brown, C. L., Bailey, A. S., Lyons, P., Levine, C. R., Buso, D. C., Likens, G. E., Knoepp, J. D., & Fukushima, K. (2015). Sources of uncertainty in estimating stream solute export from headwater catchments at three sites. *Hydrological Processes*, *29*(7), 1793–1805.
Young, D., Zégre, N., Edwards, P., & Fernandez, R. (2019). Assessing streamflow sensitivity of forested headwater catchments to disturbance and climate change in the central Appalachian Mountains region, USA. *Science of the Total Environment*, *694*, 133382.
Younger, S. E., Jackson, C. R., & Rasmussen, T. C. (2020). Relationships among forest type, watershed characteristics, and watershed ET in rural basins of the Southeastern US. *Journal of Hydrology*, *591*, 125316.
Zeng, X., Broxton, P., & Dawson, N. (2018). Snowpack change from 1982 to 2016 over conterminous United States. *Geophysical Research Letters*, *45*(23), 12–940.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Green, M. B., Bailey, S. W., Campbell, J. L., McGuire, K. J., Bailey, A. S., Fahey, T. J., Lany, N., & Zietlow, D. (2021). A catchment water balance assessment of an abrupt shift in evapotranspiration at the Hubbard Brook Experimental Forest, New Hampshire, USA. *Hydrological Processes*, *35*(8), e14300. https://doi.org/10.1002/hyp.14300