Received: 30 December 2016 | Accepted: 30 May 2017
DOI: 10.1002/hyp.11248

WILEY

RESEARCH ARTICLE

Calibration of paired watersheds: Utility of moving sums in presence of externalities

Herbert Ssegane[1] | Devendra M. Amatya[2] | Augustine Muwamba[3] | George M. Chescheir[4] | Tim Appelboom[4] | E.W. Tollner[3] | Jami E. Nettles[5] | Mohamed A. Youssef[4] | François Birgand[4] | R.W. Skaggs[4]

[1] The Climate Corporation, 4 Cityplace Dr, St. Louis, MO, USA
[2] Center for Forested Wetlands Research, USDA-Forest Service, 3734 Hwy 402, Cordesville, SC 29434, USA
[3] University of Georgia, 597 D. W. Brooks Drive, Athens, GA 30602, USA
[4] Biological and Agricultural Engineering Department, North Carolina State University, Campus Box 7625, Raleigh, NC, USA
[5] Weyerhaeuser Company, P.O. Box 2288, Columbus, MS 39704-2288, USA

Correspondence
Herbert Ssegane, The Climate Corporation, 4 Cityplace Dr, St. Louis, MO, USA.
Email: herbert.ssegane@climate.com

Funding information
Weyerhaeuser, Catchlight Energy LLC, and U. S. Department of Energy's Bioenergy Technologies Office (BETO); Argonne National Lab

Abstract

Historically, paired watershed studies have been used to quantify the hydrological effects of land use and management practices by concurrently monitoring 2 similar watersheds during calibration (pretreatment) and post-treatment periods. This study characterizes seasonal water table and flow response to rainfall during the calibration period and tests a change detection technique of moving sums of recursive residuals (MOSUM) to select calibration periods for each control–treatment watershed pair when the regression coefficients for daily water table elevation were most stable to minimize regression model uncertainty. The control and treatment watersheds were 1 watershed of 3–4-year-old intensely managed loblolly pine (*Pinus taeda* L.) with natural understory, 1 watershed of 3–4-year-old loblolly pine intercropped with switchgrass (*Panicum virgatum*), 1 watershed of 14–15-year-old thinned loblolly pine with natural understory (control), and 1 watershed of switchgrass only. The study period spanned from 2009 to 2012. Silvicultural operational practices during this period acted as external factors, potentially shifting hydrologic calibration relationships between control and treatment watersheds. MOSUM results indicated significant changes in regression parameters due to silvicultural operations and were used to identify stable relationships for water table elevation. None of the calibration relationships developed using this method were significantly different from the classical calibration relationship based on published historical data. We attribute that to the similarity of historical and 2010–2012 leaf area index on control and treatment watersheds as moderated by the emergent vegetation. Although the MOSUM approach does not eliminate the need for true calibration data or replace the classic paired watershed approach, our results show that it may be an effective alternative approach when true data are unavailable, as it minimizes the impacts of external disturbances other than the treatment of interest.

KEYWORDS

bioenergy, block bootstrap resampling, geometric regression, model uncertainty, moving sums, recursive residuals, silvicultural operations, water table

1 | INTRODUCTION

By 2035, the use of biofuels is projected to triple compared to the 2011 baseline of 1.3 million barrels of oil equivalent per day (mboe/d), with advanced biofuels supplying 20% of this demand (Birol, 2014). Therefore, there is a greater need than ever to effectively optimize current land use practices in order to meet future biofuels production demands while limiting environmental impact. This requires studies to quantify effects of such transitions of land use change on local and regional hydrology and water quality (Georgescu & Lobell, 2010).

Vast areas of the southeastern U.S. coastal plain and Gulf Coast regions are covered by pine forests (*Pinus* spp.) planted primarily for timber production. A potential option for biofuel feedstock production within these planted pine systems is the intercropping of a perennial biofuel crop, such as switchgrass (*Panicum virgatum*) between rows of planted pine. However, sustainability of this system, including impact on water resources, must be quantified and compared to current forest management practices. In a traditional setting, pine is planted on bedded rows, and the space between each row is occupied by natural vegetation. Replacing the natural understory with switchgrass introduces a relatively uniform vegetation structure between pine rows and, thus, reduces intensity of runoff (Blanco-Canqui, Gantzer, Anderson, Alberts, & Thompson, 2004; Schmer, Liebig, Vogel, & Mitchell, 2011). Albaugh, Sucre, Leggett, Domec, and King (2012) studied the effects on water use and gross primary productivity of intercropping switchgrass in pine stands using 3-year (2009–2011) data from plot-scale experiments in an upper coastal plain site in North Carolina, USA. They reported an increase in water uptake (transpiration) by switchgrass during the peak growing season and a higher total annual evapotranspiration (ET) from the traditional pine stand than from the switchgrass. However, there have been no such extensive studies on a watershed-scale basis.

A number of studies (Amatya, Gilliam, Skaggs, Lebo, & Campbell, 1998; Amatya, Gregory, & Skaggs, 2000; Amatya & Skaggs, 2011; Amatya, Skaggs, Gilliam, & Hughes, 2003; McCarthy, Flewelling, & Skaggs, 1992; McCarthy, Skaggs, & Farnum, 1991; Skaggs, Breve, & Gilliam, 1994; Ssegane et al., 2013) have been conducted on pine plantations with pattern drainage in low gradient coastal North Carolina to evaluate effects of both silvicultural and water management practices on downstream water quantity and quality. Most of these studies were based on a classical paired watershed approach, where two neighbouring watersheds (one control and one treatment) were monitored concurrently during calibration (pretreatment) and post-treatment periods (Clausen & Spooner, 1993; Loftis, MacDonald, Streett, Iyer, & Bunte, 2001). During calibration, a statistically significant relationship was established between control and treatment watersheds such that any significant shift detected during treatment would be attributable to treatment effects (Clausen & Spooner, 1993; Loftis et al., 2001; Zégre, Skaugset, Som, McDonnell, & Ganio, 2010). The paired watershed approach also integrates the roles of forest cover, internal watershed behaviour, and climate variability to establish a "baseline" for reference (Zegre, 2008).

The paired watershed approach has been extensively used to assess effectiveness of conservation practices (Jokela & Casler, 2011; King, Smiley, Baker, & Fausey, 2008; Lemke et al., 2011; Tomer & Schilling, 2009), changes in water yield due to afforestation and harvesting or deforestation (Bliss & Comerford, 2002; Bosch & Hewlett, 1982; Bren & Lane, 2014), and best management practices for controlling sediment transport, nitrogen and phosphorus runoff and leachate (Jaynes et al., 2004). This approach continues to be used on low-order watersheds as the primary method for impact assessments (Bren & Lane, 2014) although its validity for predicting effects on large flooding events has been challenged (Alila, Kuraś, Schnorbus, & Hudson, 2009). The robustness of the paired watershed approach was examined by Bren and Lane (2014) to answer questions on (a) the gain in information as a function of the length of the calibration period, (b) the relative gain or loss of information when using daily or monthly or annual data, and (c) the effect of autocorrelation of residuals in calibration models. They recommend that the calibration equation development process with a set of data from the calibration period data should be examined over time until an efficient relationship with statistically significant regression parameters is obtained. It is important to note that structural changes due to natural disturbances or anthropogenic activities other than the treatment may yield erroneous inferences of treatment effects (e.g., Vogl and Lopes, 2009). This issue is similar to a question raised by Bren and Lane (2014), namely, how to obtain the most efficient calibration. Several studies have demonstrated significant changes in hydrology and nutrient concentrations and loads on forest sites due to silvicultural and water management operations. For example, Bliss and Comerford (2002) identified a decrease in water table immediately after harvesting and later an increase in water quantity for about 4 months following harvesting of flat woods in Florida. Arthur, Coltharp, and Brown (1998), using a paired watershed approach to quantify impacts associated with tree harvesting and best management practices on water yield and water quality, reported increased water yield in the year following clear cutting of woody species in a Kentucky forest. Xu et al. (2002) attributed the rise of water table elevation (WTE) following vegetation removal to reduced transpiration and showed that harvesting impacts on the hydrology of forested wetlands were more pronounced during the growing season. However, Laurén et al. (2009) demonstrated that uncertainty in pretreatment data and thus the calibration relationship of paired watershed studies may influence estimates of the magnitude and duration of the treatment effects. Their monitoring of phosphorous loads on two independently paired catchments in Finland before and after clear-cutting showed that small treatment effects may be masked by uncertainty in the pretreatment data.

We established this experiment in early 2009 to quantify the watershed-scale effects of managing pine forests intercropped with switchgrass on poorly drained soils in Carteret County, North Carolina. We observed the hydrology (i.e., WTE and flow) of treatment watersheds using a paired watershed approach. Accordingly, this study required initial silvicultural operations in order to establish switchgrass growth. These silvicultural operations included harvesting, site preparation (bedding, shearing, and raking), and herbaceous broadleaf control as described in the plot-scale study by Albaugh et al. (2012). All or some of these activities may have impacted the soil hydraulic properties and, thus, impacted the WTE, drainage outflow, and nutrient and sediment dynamics (Albaugh et al., 2012). For example, Skaggs, Amatya, Chescheir, and Blanton (2006) found 20–30 times higher effective hydraulic conductivity for the top 90 cm of the Deloss fine sandy loam soil in a matured plantation forest when compared to the data published prior to harvest by the NRCS Soil Survey at the Carteret County, North Carolina study site. Skaggs et al. (2006) observed that harvest did not appear to affect those values, but site preparation for regeneration, including bedding, reduced the effective hydraulic conductivity to values typically assumed for these soil series. Developing a calibration regression relationship with data collected from the control and treatment watersheds undergoing such disturbances violates the assumption of the paired watershed approach.

The specific objective of this study was to quantify the timing of hydrologic shifts in regression relationships of control and treatment watersheds due to silvicultural management operations and to develop calibration relationships that minimize the effects of external factors. We achieved this by minimizing uncertainty in the regression parameters based on their structural stability over time. Our study also presents the response of WTE to these operations. We collected data for the pretreatment calibration period from 2009 to 2012, during which time silvicultural operations were conducted. These practices may have acted as external factors, affecting pretreatment hydrologic relationships between control and treatment watersheds. A change detection technique of moving sums (MOSUMs) of recursive residuals (Bauer & Hackl, 1978; Chu, Hornik, & Kaun, 1995; Zeileis et al., 2013) was used to minimize the effects of these external factors on the uncertainty of parameters of the calibration models. Utilizing the MOSUMS approach, pretreatment regression models were developed using a subset of the 2009–2012 data, where the effect of external factors was minimal and the regression coefficients were concurrently stable and statistically significant. Use of the terms "structural stability" or "temporal stability" of regression coefficients in this study refers to non-significant changes in the regression coefficients (intercept and slope) over the calibration period due to a shift in the relationship between control and treatment watersheds. This shift may be caused by externalities other than the operational treatment used for switchgrass establishment.

2 | METHODS AND MATERIALS

2.1 | Site description

Our study site, Carteret 7, located in Carteret County, North Carolina (Latitude of 34.8220 and Longitude of −76.6680), was established by Weyerhaeuser Company and Catchlight Energy LLC, a joint venture between Chevron and Weyerhaeuser Company, on land owned and managed by Weyerhaeuser Company. The site consists of four small experimental watersheds (Figure 1: D0 = 27.5 ha, D1 = 26.3 ha, D2 = 25.9 ha, and D3 = 27.1 ha), each with four parallel ditches for enhanced drainage due to the low topographic gradients. The D1, D2, and D3 watersheds have been used for previous studies, but watershed D0 (north of D1) was established in 2009 specifically for this study. These watersheds were surrounded by forested land to the north, south, and west and by agricultural land to the east. Mean annual rainfall over a 21-year period was 1517 mm with a 10–20% annual variation, due to hurricanes and tropical storms (Amatya & Skaggs, 2011) occurring mostly during the summer–fall growing season defined as a period from May to October.

The natural soil surface gradient from the north boundary of plot D0 to the south boundary of plot D3 is less than 0.3 m over the 1,600 m distance for a slope of less than 0.025% (Figure 1). The low gradient continues for over 5 km both north and south of the borders of the research site. Consequently, lateral subsurface flow under natural conditions is very low. Subsurface drainage from the watersheds is driven by a system of parallel 1.2 to 1.5 m deep drainage ditches spaced 100 m

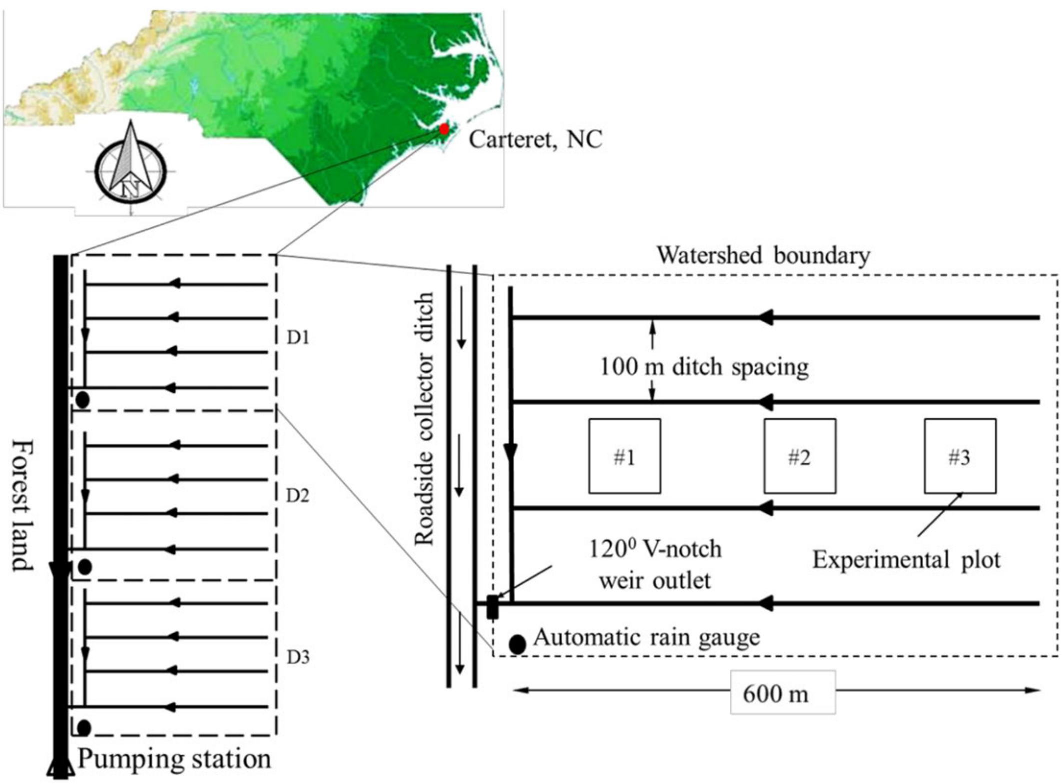

FIGURE 1 The location and layout of artificially drained pine-forest experimental watersheds with monitoring stations in Carteret County, North Carolina. Water table elevation was monitored in plots #1 and #3. Watershed D0 is north of watershed D1 with drainage configuration similar to D1–D3. The ground surface elevation is 2.35, 2.28, 2.33, and 2.23 m for D0, D1, D2, and D3, respectively

apart. According to drainage theory, the midpoint between parallel drains of equal elevation can be treated as a no-flow boundary, so the watersheds are divided along the midpoints of the fields separating the parallel ditches. The elevations of all of the drainage ditches across all of the watersheds are nearly the same; however, it is possible that a small amount of water seeps from one watershed to the other due to uneven boundary conditions at the surface of the fields. This could occur when the water tables are close to the bottom of the ditches. However, in these conditions, drainage rates would be low due to low flow gradients and lower hydraulic conductivity values of the deeper soil layers. Seepage rates would be negligible compared to the overall subsurface rates to the ditches (Amatya, Skaggs, & Gregory, 1996).

Surface drainage does not occur on the watersheds due to the microtopography created by the 25 cm raised beds for the pine trees that are parallel to the ditches. Since seepage rates between watersheds are negligible, the water tables measured in fields located two parallel drainage ditches away from the watershed boundaries would not be affected by the adjacent watersheds. Fipps and Skaggs (1986) showed that the water table in a field located two parallel drains away from a water source (such as an unlined irrigation canal) would not be affected by seepage from the source. It is important to note that the water table gradients and drainage intensities in the Fipps and Skaggs (1986) study were much greater than those observed on our watersheds.

2.2 | Weather and hydrologic data

The four experimental watersheds (Figure 1) were instrumented to measure and record weir stage (cm), water table depth (cm), water quality, and rainfall (mm). Total rainfall on each watershed was collected with automatic tipping bucket rain gauges calibrated by manual gauge measurements in an open area. Air and soil temperatures (°C), relative humidity (%), wind speed (km hr^{-1}), and solar radiation (W m^{-2}) were measured and recorded on a 30-min interval using an onsite weather station. An anemometer and a relative humidity sensor were stationed at 3.4 m and 2.3 m above ground, respectively. The 30-min weather data were integrated to obtain a daily average, which was then used to calculate the Penman-Monteith-based potential ET (Monteith, 1965) for a standard grass reference (REF-ET) for the site.

An adjustable height, 120° V-notch weir, located at the outlet of each watershed, measured drainage outflow by continuously recording water levels (stage) upstream and downstream of the weir. The bottom of the V-notch was approximately 100 cm below the average soil surface elevation for each watershed. Automatic stage recorders were installed to make stage measurements every 12 min upstream and downstream of the water level control structure, and a V-notch weir was set about 0.3 m above the bottom of outlet ditch. In 1990, a pump was installed at the roadside collector ditch downstream of all watershed outlets in order to minimize weir submergence during large events (Amatya et al., 1996). Flow rates (m^3 s^{-1}) were computed using discharge-stage relationships for non-submerged and submerged weir conditions (Brater & King, 1976). Twelve-minute flow rates were integrated to obtain daily totals normalized by watershed area (mm day^{-1}). Calculated daily flow values greater than the capacity of the downstream culvert during large storm events were capped at 45 mm day^{-1}, the approximate culvert drainage capacity (Amatya & Skaggs, 2011). Such data were excluded from analysis because of uncertainty due to highly submerged conditions (Amatya et al., 1998; USGS, 1997). WTE was continuously recorded on an hourly basis at the front (east side) and back (west side) small experimental plots of each watershed (Figure 1). The average of back and front WTE was assumed as the representative WTE for each watershed.

2.3 | Silvicultural operational and management practices

The control watershed, D2 (loblolly pine of 14–15-year-old stands), was thinned (50% reduction in basal area) in early 2009 (Figure 2). Watersheds D0 and D1 were clear-cut in April of 2009, and 85% of D3 was harvested by November of 2009 when wet conditions halted operations.

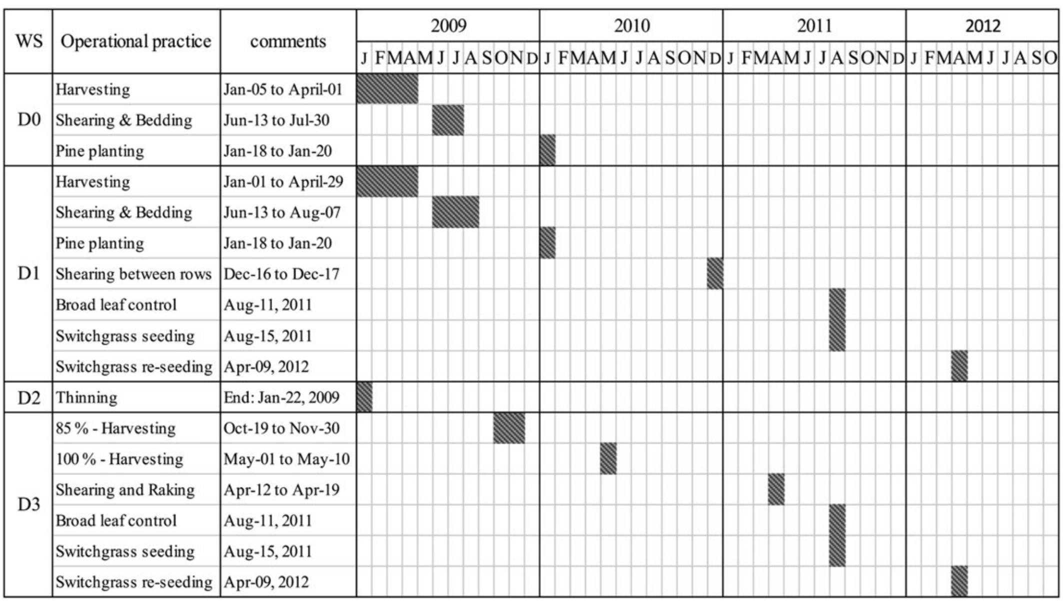

FIGURE 2 Timeline of operational management practices. WS stands for watershed

Harvest was complete by May of 2010. Pine planting on D0 and D1 was completed by January of 2010. Watershed conditions at D0 and D1 remained the same until the December 2010 when the understory between pine rows in D1 was sheared in preparation for switchgrass planting. Switchgrass seed was broadcast on D1 and D3 in August of 2011 but did not germinate well due to excessively wet conditions. However, the second phase of switchgrass seed broadcast, which occurred between March and April of 2012, had much better germination in the intercropped D1 watershed, but coverage at D3 was still only about 15 to 20%. Other management practices included broadleaf control on D1 and D3 in August of 2011. Treatment effects of switchgrass growth on D1 and D3 through 2015 and the young pine and natural understory growth on watershed D0 were not addressed in this paper.

2.4 | Structural stability of calibration relationships

To minimize effects of silvicultural operations as external factors on the uncertainty of calibration regression models and provide a more reliable calibration with adequate length, as noted by Bren and Lane (2014), a change detection technique of MOSUMs of recursive residuals (Bauer & Hackl, 1978; Chu et al., 1995) was used to select the longest calibration periods for each control–treatment watershed pair (D1 vs. D2 and D3 vs. D2). There were no management operations on D0 after pine planting. However, a MOSUM test was carried out for the D0 versus D2 calibration model as a reference for testing for false identification of structural breaks. The MOSUM is a variant of the cumulative sums (CUSUM) method (Brown, Durbin, & Evans, 1975). Both methods test the null hypothesis that regression coefficients of a linear regression are constant over time against an alternative hypothesis that coefficients change over time due to extraneous factors. The CUSUM and MOSUM tests have been applied to detect temporal changes in areas of land use and land cover, hydrology, and national economic indicators (de Jong, Verbesselt, Zeileis, & Schaepman, 2013; Ghosh, 2009; Olmo, Pilbeam, & Pouliot, 2011; Tiwari, Shahbaz, & Islam, 2012; Verbesselt, Hyndman, Newnham, & Culvenor, 2010; Verbesselt, Zeileis, & Herold, 2012; Vogl & Lopes, 2009; Webb, Kathuria, & Turner, 2012). For this study, we used the MOSUM of recursive residuals (defined by Equations 1 to 4) because it is more sensitive to temporarily unstable parameters than CUSUM. Also, cumulated sums calculated with the MOSUM approach become less sensitive as the number of residuals becomes larger (Chu et al., 1995).

$$M_r = \sum_{i=r-w+1}^{r} \frac{W_r}{\sigma}; \quad r = p+w, \ldots, N, \tag{1}$$

$$W_r = \frac{y_r - x_r' b_{r-1}}{\sqrt{\left(1 + x_r'\left(X_{r-1}' X_{r-1}\right)^{-1} x_r\right)}}, \tag{2}$$

$$\sigma^2 = \frac{1}{N-p} \sum_{i=p+1}^{N} \left(W_i - \overline{W}\right), \tag{3}$$

$$\overline{W} = \frac{1}{N-p} \sum_{i=p+1}^{N} W_i\ , \tag{4}$$

where M_r is the rth MOSUM of recursive residuals with a data window size of w, p is the total number of regression coefficients, N is the total number of data samples, W_r is the rth recursive residual, y_r is the rth observation of the response variable, x_r' is the rth row vector of the explanatory variables, b_{r-1} is the ordinary least squares estimate of parameter b using data before the rth time step, σ^2 and $\overline{W}$ are variance and mean estimates of the recursive residual W_i, respectively. The X_{r-1} and X'_{r-1} are the $[(n-1) \times p]$ regressor matrix and its transpose using data before the rth time step.

The MOSUM test for change detection was implemented in an R-software environment using the "strucchange" package (Zeileis et al., 2013), which follows a three-step procedure. The first step checks for existence of structural change based on the assumption that variability of the MOSUMs of recursive residuals under structural stability follows a Brownian motion (a random walk) with an expected mean of zero. Therefore, if the MOSUM crosses the 95% confidence boundary, then structural change is detected. For details on the technical basis and the asymptotic function of the 95% confidence boundary, refer to Zeileis et al. (2013). When structural change is detected in the first step, then steps two and three determine the number and location of change points, also known as break points or break dates. Break points and corresponding 95% confidence intervals are estimated based on methods developed by Bai (1994, 1997) and Bai and Perron (1998, 2003a, 2003b) and implemented by Zeileis et al. (2013). The second step determines the number of break points by minimizing the Bayesian information criterion. However, one can predefine the maximum number of break points for a given time series. For this analysis, we arbitrarily assigned four break points to correspond to the potential effects of clear-cutting, shearing and bedding, pine planting, and planting of switchgrass. The third step iteratively determines the location of break points by minimizing the regression sums of squares.

Based on this procedural implementation of MOSUM and the fact that analyses are made on MOSUMs, the location where the MOSUM crosses the 95% confidence boundary is not always the location of the break points. When the MOSUM returns inside the boundary, it does not mean the relationship has regained the previous structural stability.

Finally, it is important to note that the strength of the linear relationship, size of the moving window, and number of predetermined break points all influence the location of the break points. We used a 30-day data window to calculate MOSUMs of recursive residuals, and a 5% (α = 0.05) significance level to detect structural changes in regression coefficients. Only the WTE data were used to determine pairwise control–treatment calibration periods for both daily WTE and daily flow data, because for these low-gradient artificially drained coastal plain soils, the depth of the water table is the main driver of the hydrology.

2.5 | Regression analysis

Due to serial correlations inherent in a daily time series, regression relationships between control and treatment watersheds were determined using a resampling technique that accounts for serial correlation in daily time series and geometric mean regression (Efron & Tibshirani, 1994; Elkinton et al., 1996; Plotnick, 1989). The resampling technique is called block bootstrap (Khaliq, Ouarda, Gachon, Sushama, & St-Hilaire, 2009; Kundzewicz & Robson, 2004;

Politis, 2003). In this approach, the original data were resampled in predetermined blocks for a large number of times to estimate regression coefficients. This method incorporates the effects of serial correlations higher than the first-order dependencies. Sørensen et al. (2009) used a block length of 14 days. Their choice of greater lengths gave similar results. For this analysis, block bootstrapping was performed using a time series function "tsboot" with a fixed block length of 50 days and 10,000 bootstrap resamples in R (R Development Core Team, 2015). Use of 20-, 30-, and 100-day block lengths gave similar results.

Geometric mean regression was used to estimate regression coefficients for each bootstrap resample. To ensure replicability of results, the same arbitrary number of 4711 was used to seed the random number generator. Geometric mean regression, also known as the reduced major axis regression, is suited for paired watershed analysis because it assumes errors are associated with both dependent (treatment watershed) and independent (control watershed) variables (Friedman, Bohonak, & Levine, 2013). Ten thousand bootstrap resamples were used to estimate regression coefficients and corresponding confidence intervals.

3 | RESULTS

3.1 | Water table elevation response to rainfall

Rainfall during 2009 and 2012 was relatively similar and well distributed throughout each year with some large events in the fall. Rainfall distributions in 2010 and 2011 were also similar, and most of the larger events occurred in the fall and the winter. The driest year was 2011, which had an annual rainfall of 1,181.7 mm and a net precipitation (difference between rainfall and REF-ET) deficit of 53 mm. The wettest year was 2010, which had an annual rainfall of 1,420.6 mm and a net precipitation surplus of 275.7 mm. Analysis of rainfall measured at each watershed shows similar rainfall distributions across all watersheds each year with slight differences in actual rainfall amount. These annual differences were similar to the rainfall observed at the weather station. However, on average, there was a general negative linear spatial trend of average annual rainfall from D0 with the highest rainfall (1,348.6 mm), to D3 with the lowest rainfall (1,234.2 mm).

Seasonal climatic and vegetation dynamics affected WTE (Figures 3– 5), such that the WTE dropped below 1.5 m (above sea level) in the summer, with occasional large rainfall events that temporarily raised the water table to the soil surface with ponding. For example, a large area of D3 had ponded conditions as a result of large events with daily rainfall exceeding 150 mm in late September 2010 and 100 mm in mid-October 2011 on wet antecedent conditions. Vegetation effects were reflected by higher WTE on all three treatment watersheds (young pine or emergent vegetation with a shallow root system and less evaporative demand) than the control watershed (D2: 14- to 15-year pine with deep root system and high evaporative demand) during the growing season. However, water table response to large storm events, characterized by a rise to the soil surface, was similar among all watersheds; an observation consistent with previous studies at this site (Amatya & Skaggs, 2011). The average difference of WTE between D0 and D2 was 17.3 ± 0.9 cm (±95% CI), 22.5 ± 0.8 cm between D1 and D2, and 9.1 ± 0.6 cm between D3 and D2. The differences between control and treatment watersheds for the 2009 to 2012 pretreatment calibration period were significantly greater than the differences observed using the historical data (Ssegane et al., 2013), except between D0 and D2 (because D0 was established in 2009).

The greater average WTE difference between D1 and D2 was due to treatment in D1: clear-cutting in 2009, shearing of the understory in December 2010, and switchgrass seed broadcast in August 2010 and April of 2012 (Figure 2). The average WTE difference between D3 and D2 was the least since harvesting at D3 was later (between November 2009 and May 2010) than at D1. Another reason for this difference was the high variability between the WTE of the front and back plots of D3 (soil heterogeneity) compared to similar plots for D0, D1, and D2. A similar trend was evident in the maximum single-day difference between the control and treatment watersheds. The largest difference between D1 and D2 was 86.3 cm on June 28, 2009, during shearing and bedding. The largest difference between D3 and D2 (41.0 cm) was observed on August 28, 2011, after shearing and raking of D3 for switchgrass seed broadcasting (Figure 2).

3.2 | Effects of harvesting on seasonal WTE and flow

Prior to the complete harvest of D1, the D1 WTE was higher than D2 (Table 1; January to April of 2009), consistent with earlier studies

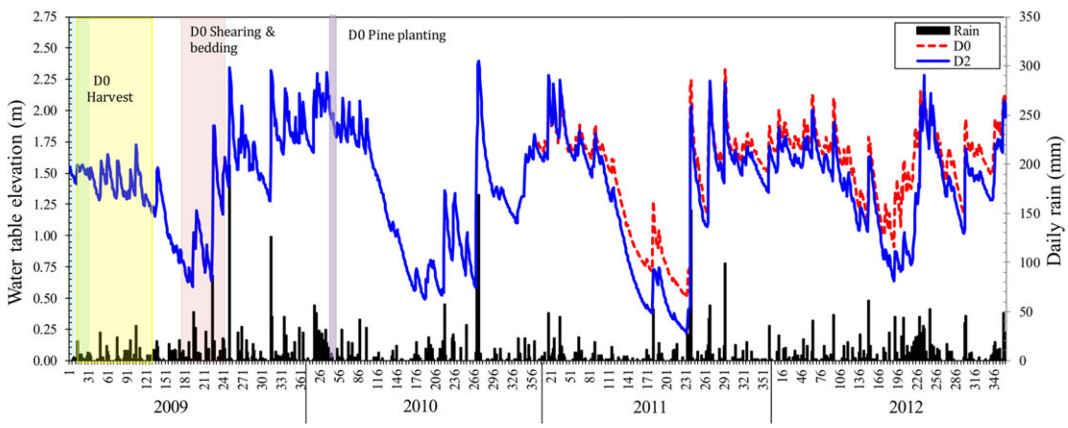

FIGURE 3 Daily average water table elevation for D0 and D2. Daily rain is represented by the average of rain on watersheds D0, D1, D2, and D3

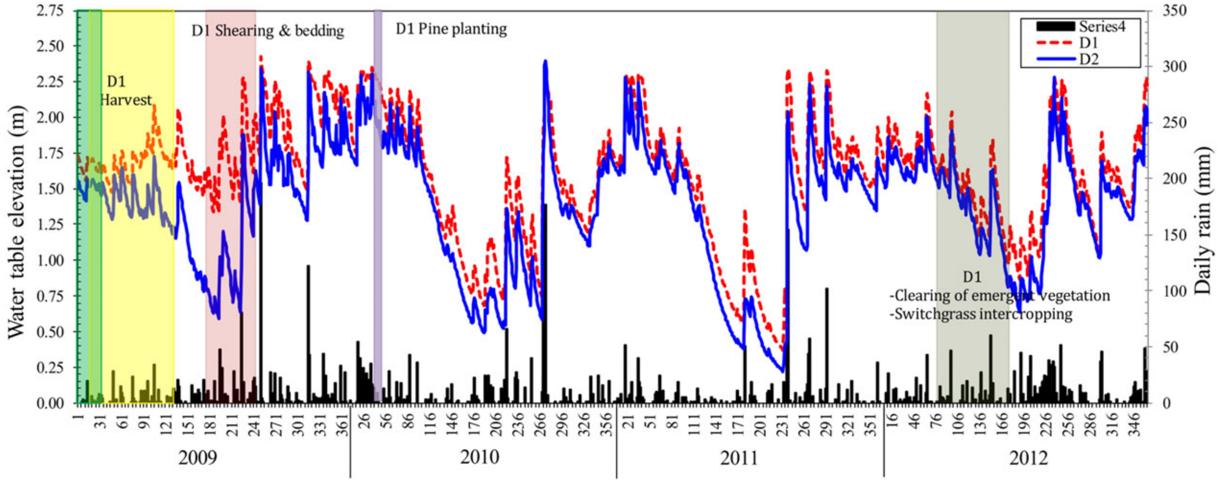

FIGURE 4 Daily average water table elevation for D1 and D2. Daily rain is average of rain on watersheds D0, D1, D2, and D3

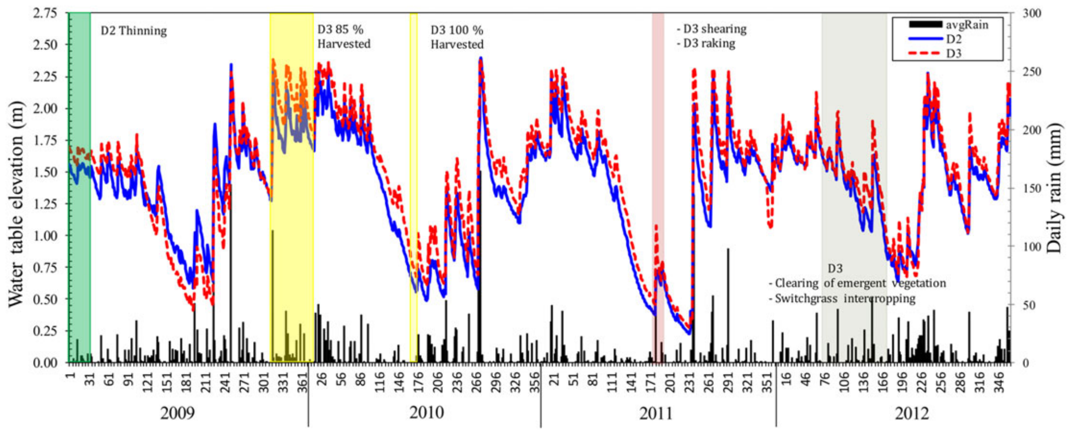

FIGURE 5 Daily average water table elevation for D3 and D2. Daily rain is represented by the average of rain on watersheds D0, D1, D2, and D3

(Amatya & Skaggs, 2011; Ssegane et al., 2013), whereas D2 flow was greater than D1 flow (Table 1). However, after harvest of D1 (2010 to 2012), the difference in D1 and D2 flow was mostly positive (D1 flow greater than D2 flow) during the growing season and negative (D2 flow greater than D1 flow) during the dormant season. The negative differences in flow during the dormant season, dominated by high water tables and low ET (which minimally affects flow), are attributable to intrinsic differences in watershed soils (Blanton, Skaggs, Amatya, & Chescheir, 1998) and microtopography, including possible lateral seepage from D1 to a roadside drainage ditch at the north boarder of D1 depicted by preharvest behaviour (D2 flow greater than D1 flow). The large positive differences between D1 and D2 during the growing season (specifically from August to October) may be attributed to diminishing ET rates of the emergent vegetation at D1 (Sampson, Amatya, Lawson, & Skaggs, 2011), yet transpiration rates of D2 (relatively old pine with less areal coverage of understory) were still high, and thus, higher flows occurred on D1. Exceptions to this trend include September and October of 2010 and August and September of 2012. A possible explanation of these exceptions is illustrated in Figure 4. The water table on D1 was consistently shallower than on D2. During the dormant season, the difference in water table became less due to the above stated reasons so D1 generally produced less flow than D2. During the growing season, the water table was considerably deeper in D2, and the flow from D2 was less compared to D1. Very large rainfall events in September of 2010 and October of 2011 compensated for the difference in WTE, and the WTE almost reached the soil surface for both D1 and D2. Under these conditions, the role of vegetation became less dominant than other watershed characteristics.

The above seasonal shift in hydrologic behaviour, particularly the flow, was possibly further complicated by seasonal variation of climatic variables, effects of site preparation for switchgrass seed broadcasting, and uncertainty in estimated flows during short periods of weir submergence and backflow, mostly on D3. For example, the weir outlets at all four watersheds were submerged (downstream stage greater than the V-notch of the weir) from September 27 to October 04 of 2010 and January 26 of 2011 with backflow on D3 (January 26, 2011) due to failure of the downstream pump installed to minimize such submergences on upstream weir outlets. The total incidences of submergence and backflow on all watersheds occurred 0.80% and 0.04% of the time, respectively, based on 12-min stage data during this time. D3 flow did not immediately increase compared to D2 after 85% of harvesting on D3 in November of 2009 because it was during the

TABLE 1 Monthly differences in water table elevation (WTE) and flow of D1 (3–4-year young pine to be intercropped with switchgrass) and D2 (14–15-year pine; control watershed)

Preharvest/post-harvest of D1								
		Difference, [D1–D2]		D2		Difference, [D1–D2]		D2
Year	Month	WTE cm	Flow mm	Rain mm	Month	WTE cm	Flow mm	Rain mm
2009	Nov	27	−1	209	May	47	11	81
	Dec	27	24	180	Jun	66	0	100
	Jan	15	−8	53	Jul	77	0	117
	Feb	16	−10	69	Aug	58	25	244
	Mar	23	−19	48	Sep	32	3	287
	Apr	42	−4	88	Oct	27	4	75
Post-harvest of D1								
2010	Nov	9	0	45	May	17	0	33
	Dec	11	−3	82	Jun	22	0	47
	Jan	17	−43	177	Jul	26	0	111
	Feb	24	15	111	Aug	31	0	175
	Mar	15	1	138	Sep	21	−7	421
	Apr	14	−4	47	Oct	13	−13	25
2011	Nov	15	1	68	May	19	0	17
	Dec	13	−1	37	Jun	22	0	67
	Jan	13	−9	147	Jul	28	0	50
	Feb	18	−4	99	Aug	24	41	256
	Mar	9	−4	83	Sep	30	14	185
	Apr	12	−2	38	Oct	23	13	131
2012	Nov	12	−1	20	May	16	0	130
	Dec	14	7	159	Jun	19	1	52
	Jan	13	1	62	Jul	25	0	172
	Feb	10	−1	65	Aug	25	−6	318
	Mar	10	−2	84	Sep	15	−4	132
	Apr	10	1	96	Oct	10	1	110

Note. The left and right columns represent months during the dormant (November to April) and growing (May to October) season, respectively. Shaded rows are months with high occurrences of weir submergence (downstream weir stage above the V-notch of the weir). Months of November and December for a given year are out of order and are not sequential with Jan–Apr. They represent part of the dormant season for that year (Jan–Apr and Nov–Dec).

TABLE 2 Monthly differences in water table elevation (WTE) and flow of D3 (clear-cut for planting switchgrass only) and D2 (14–15-year pine; control watershed)

Preharvest/post-harvest of D3								
		Difference, [D3–D2]		D2		Difference, [D3–D2]		D2
Year	Month	WTE cm	Flow mm	Rain mm	Month	WTE cm	Flow mm	Rain mm
2009	Nov	15	−4	209	May	−14	0	81
	Dec	23	−2	180	Jun	−19	0	100
	Jan	13	−1	53	Jul	−23	0	117
	Feb	15	−1	69	Aug	−24	−1	244
	Mar	14	−1	48	Sep	−3	−5	287
	Apr	12	−1	88	Oct	4	0	75
Post-harvest of D3								
2010	Nov	14	0	45	May	24	0	33
	Dec	14	−3	82	Jun	19	0	47
	Jan	16	−109	177	Jul	18	0	111
	Feb	18	−40	111	Aug	16	0	175
	Mar	11	−17	138	Sep	19	−9	421
	Apr	11	−1	47	Oct	16	−17	25
2011	Nov	9	6	68	May	18	0	17
	Dec	−8	3	37	Jun	12	0	67
	Jan	10	−22	147	Jul	6	0	50
	Feb	17	−5	99	Aug	9	8	256
	Mar	13	−4	83	Sep	22	18	185
	Apr	17	0	38	Oct	17	9	131
2012	Nov	10	6	20	May	17	3	130
	Dec	9	13	159	Jun	12	4	52
	Jan	8	−2	62	Jul	5	0	172
	Feb	1	2	65	Aug	0	20	318
	Mar	4	3	84	Sep	8	3	132
	Apr	5	−6	96	Oct	10	7	110

Note. The left and right columns represent months during the dormant (November–April) and growing (May–October) season, respectively. Light-shaded rows are months with high occurrences of weir submergence (downstream weir stage above the V-notch of the weir), and dark-shaded rows are months when backflow (downstream weir stage greater than upstream weir stage) occurred on D3. Months of November and December for a given year are out of order and are not sequential with Jan–Apr. They represent part of the dormant season for that year (Jan–Apr and Nov–Dec).

dormant season, and observed differences in D2 and D3 flows were similar to preharvest conditions Table 2; (D2 flow > D3 flow). Note that prior to harvest, D3 was a 35-year-old pine stand compared to D2 that was a 12-year-old stand thinned in early 2009. Effects of vigorous emerging vegetation with increased leaf area index (LAI), particularly in 2010, a year after harvest of D1 and D3 may have also resulted in some negative differences in some months of the growing season. Sampson et al. (2011) noted that in coastal loblolly pine stands, herbaceous and arborescent species can dominate the site LAI for some years after a harvest.

3.3 | MOSUM change detection in pairwise calibration periods

Movement of MOSUM outside the 95% confidence bounds (Figure 6; horizontal dotted lines) is indicative of a structural break in the stability of regression coefficients. The points where the MOSUM crosses the confidence bounds may not represent actual points when structural changes occurred, because of the MOSUMs. These points are represented by the vertical dotted lines known as "break points" or "break dates." There were no operational management practices on D0 after pine planting (Figure 2). and the MOSUM (Figure 6a) shows no structural break. The D0 and D2 relationships are based on data from January 01, 2011, to March 31, 2012 (456 days: over one annual cycle).

The first break point between the WTE of D1 and D2 (between March 26–28, 2009; Figure 6b) coincides with harvest of D1 (Figures 2 and 6). The second break point (between October 1–10, 2009) occurred 1 month after shearing and bedding on D1. The third break point (between February 17 to March 3, 2010) was 1 month after pine planting on D1, and the fourth break point (between August 24–30,

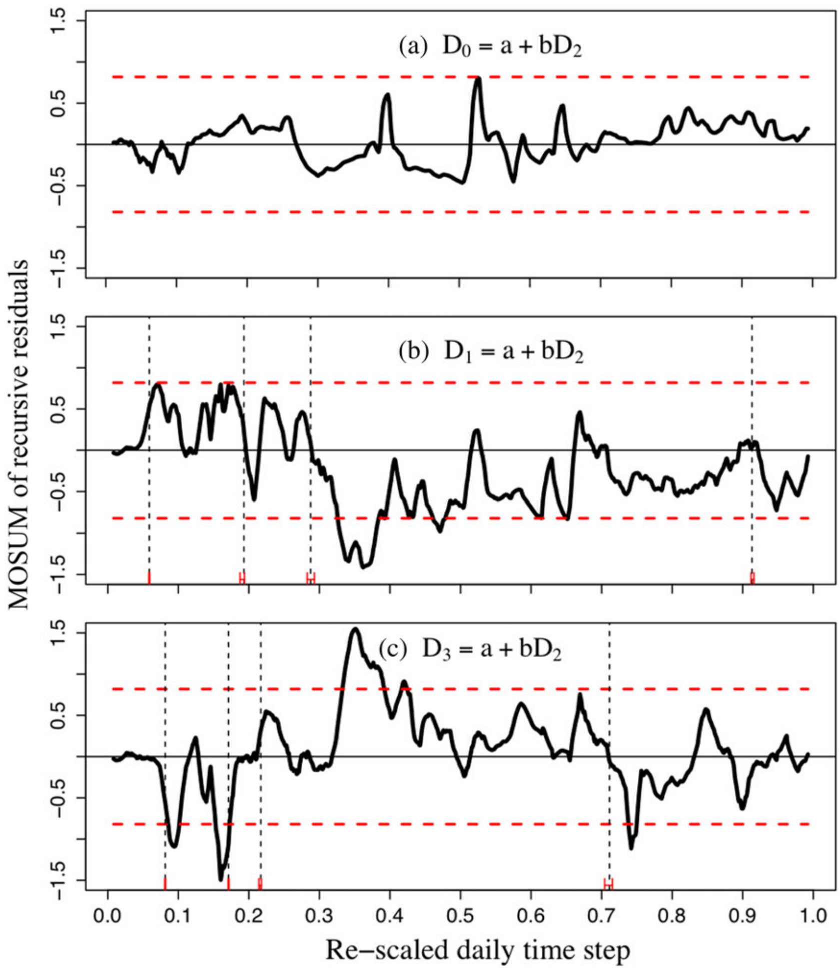

FIGURE 6 Graphs of the moving sums (MOSUMs) of recursive residuals for the linear relationships between the water table elevation of (a) D0 and D2, (b) D1 and D2, and (c) D3 and D2. A shift of the MOSUM outside the 95% confidence intervals (long horizontal dotted lines) is indicative of a structural break in the linear relationship. The vertical dotted lines are estimated break points (break dates). The corresponding small horizontal lines that cross each break date are the respective 95% confidence intervals for each break date. Because the analysis is on moving sums, the location where the MOSUM cross the 95% confidence boundary is not always the location of the break points. Also, when the MOSUM return inside the boundary, it does not mean the relationship has regained the previous structural stability. No structural break on D0–D2, but there are structural breaks on D1–D2 and D3–D2 water table elevation linear regression models

2012) was about 3 months after the second rebroadcast of switchgrass seed on D1. As a result, the actual calibration period for D1 and D2 WTE and flow relationships with a minimal operational disturbance started 1 month after pine planting on D1 (March 1, 2010). This calibration period continued, as preparations for the second phase of switchgrass seed broadcasting were made. Broadcasting occurred on March 31, 2012 (making the calibration period 762 days altogether). This period of minimal disturbance occurred between pine planting in January 2010 and the final site preparation for switchgrass sowing in May 2012 on the intercropped site. This period spanned the very wet periods of September 2010 and August 2011, the very dry periods of spring–summer of 2010 and 2011, and a period of average rainfall in 2010. Work by Bren and Lane (2014) demonstrated that good calibration (Nash-Sutcliffe efficiency >0.8) by simple linear models could be achieved after 100 days of data. This calibration period defined by the MOSUM method captures over two annual cycles with associated seasonal variations.

The first break point for D3 and D2 WTE linear relationship (between April 28–30, 2009) occurred 2 months after 50% thinning of D2 (Figures 2 and 6c) and coincided with the start of the growing season after thinning. The second break point (between September 6–8, 2009) was 1 month before 85% harvesting of D3. This break point is not associated with any known prior silvicultural management activity. The third break point (between November 9–14, 2009) coincided with the 85% harvesting of D3, and the fourth break point (between October 26 to November 10, 2011) is about 2 months after the first seeding of switchgrass on D3 (August 15, 2011). Harvesting of the final 15% of trees on D3 in May of 2010 did not significantly alter the D3 and D2 WTE relationship due to emergent vegetation on the previously harvested 85% portion of the watershed. Therefore, the calibration period for D3 and D2 was between the end of 85% harvesting of D3 (December 1, 2009), and prior to preparations for the first phase of switchgrass planting (July 31, 2011), or 608 days altogether.

3.4 | Pairwise calibration with stable regression relationships

The coefficient of determination (R^2), Nash-Sutcliffe efficiency (*NSE*), and the root mean squared error (RMSE) were indicative of strong, quantifiable, and predictable relationships between hydrologic responses of control and treatment watersheds (Table 3). For R^2 ($0 \le R^2 \le 1.0$) and *NSE* ($-\infty \le NSE \le 1.0$), a value of 1.0 is indicative of an optimal model. Therefore, the calibration equations showed the two hydrologic responses of the control watershed were strong predictors of similar responses at treatment watersheds and, thus, significant predictors of treatment effects using a paired watershed approach.

All of the data from the period after harvesting to prior to switchgrass planting (2009–2012) yielded significantly different regression equations with relatively weaker but not significantly different regression statistics (lower R^2, and *NSE*, and higher RMSE) for D1 and D2

TABLE 3 Results based on block bootstrapped geometric mean regression analysis and 10,000 bootstrap resamples

Paired	Data	Intercept[a] A (Al, Au)	Slope[a] B (Bl, Bu)	R^2	*NSE*	RMSE
WTE0 vs. WTE2	All data[b]	0.38 (0.33, 0.45)	0.83 (0.78, 0.86)	0.99	0.99	0.04
WTE1 vs. WTE2	All data	0.32 (0.20, 0.59)	0.94 (0.80, 1.00)	0.88	0.87	0.16
MOSUM data	0.24 (0.18, 0.29)	0.95 (0.91, 0.99)	0.97	0.97	0.08	
WTE3 vs. WTE2	All data	0.11 (0.03, 0.19)	1.01 (0.94,1.07)	0.97	0.97	0.09
MOSUM data	0.14 (0.07, 0.21)	1.01 (0.97,1.05)	0.99	0.99	0.06	
Q0 vs. Q2	All data[b]	0.31 (0.20, 0.44)	1.24 (1.13, 1.38)	0.92	0.92	0.78
Q1 vs. Q2	All data	0.07 (−0.07, 0.20)	0.94 (0.78, 1.30)	0.84	0.83	1.07
	MOSUM data	−0.02 (−0.13, 0.04)	1.10 (0.83, 1.58)	0.83	0.82	0.99
Q3 vs. Q2	All data	0.03 (−0.02, 0.09)	0.85 (0.76, 1.05)	0.91	0.91	0.68
	MOSUM data	−0.02 (−0.07, 0.05)	0.77 (0.72, 0.82)	0.91	0.91	0.69

Note. MOSUM = moving sum; NSE = Nash-Sutcliffe efficiency; RMSE = root mean squared error; WTE = water table elevation.

[a]Letters l and u refer to the lower and upper 95% bootstrapped confidence intervals, respectively.

[b]Water table elevation (WTE, m) and flow (Q, mm) relationships between watersheds D0 and D2 are based on all data because no structural instability was detected using the MOSUM approach.

WTE, and D3 and D2 flow compared to use of MOSUM-derived data (Table 3). For the WTE relationships between D1 and D2, the slopes were not significantly different from the slopes of the MOSUM data, but the intercepts were significantly different. The intercepts of all flow equations were non-significant because their 95% confidence intervals included a zero intercept. The daily flow relationships using all data for D3 and D2 compared to MOSUM data were significantly different because of differences in the slopes (Table 3). There were no significant differences between regression equations for D3 and D2 WTE and D1 and D2 flow.

Overall, the RMSE of the MOSUM calibration equations were smaller than those based on all data. Therefore, the statistically significant differences in WTE and flow calibration relationships based on all data compared to MOSUM-based data will potentially result in significantly different treatment effects, specifically as it pertains to the quantity and duration of treatment effect.

4 | DISCUSSION

4.1 | Uncertainty of calibration relationships

The paired calibration relationships (Table 3) were developed using a block bootstrap geometric mean regression and data measured during periods with non-significant shifts in the regression coefficients due to operational practices. Compared to ordinary least-squares regression, the geometric mean regression minimizes effects of data uncertainty on regression coefficients by simultaneously minimizing both the explanatory (x-errors) and response (y-errors) errors (Plotnick, 1989). However, neither method is structured to detect significant temporal shifts in regression coefficients or the use of the MOSUM test. Figure 7 is a sample comparison of the uncertainty of regression coefficients presented as a frequency distribution obtained by using

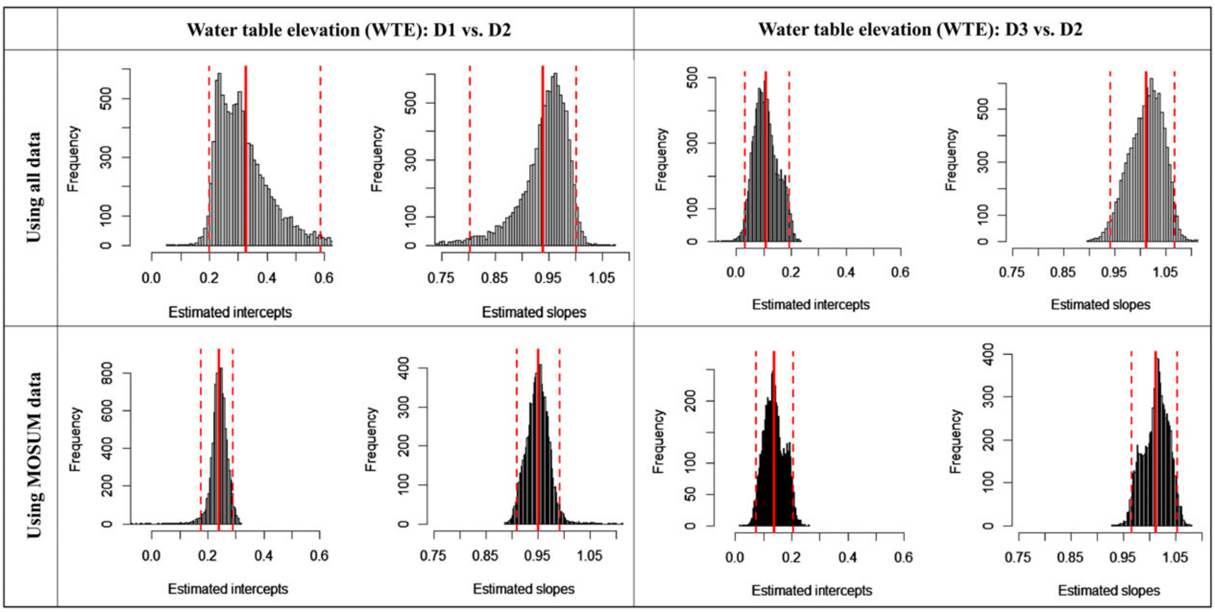

FIGURE 7 Differences in the data result in the uncertainty of regression coefficients (all data: top plots versus bottom data that is temporally stable as determined by moving sums of recursive residuals: MOSUM). The dotted vertical lines represent the 95% CI whereas the solid line is the estimated coefficient (mean of 10,000 bootstrap resamples). Uncertainty is represented by the width of the 95% CI. A small width is indicative of less uncertainty

10,000 bootstrap resamples when MOSUM data (period with stable regression coefficients) was used compared to use of all 2009–2012 data. The top two graphs in Figure 7 are frequency distributions of the intercept and slope (means shown as solid vertical lines for all data 2009–2012), whereas the bottom graphs correspond to frequency distribution of the coefficients when MOSUM data (2010–2012) was used.

The slope of the MOSUM data, 0.95 (95% CI [0.91, 0.99]) is not significantly different from the slope of all data, 0.94 (95% CI [0.80, 1.00]; Figure 7). However, the intercepts are significantly different (Figure 7: compare 0.24 [0.18, 0.29] and 0.32 [0.20, 0.59]), and thus, the two calibration equations are significantly different. Although the slopes of the two different periods are not significantly different, the uncertainty of the slope of all data was greater than the uncertainty of MOSUM data (Figure 7). The model uncertainty is also reflected in the model fit performance statistics of *NSE* (0.97 vs. 0.87; Table 3) and RMSE (0.084 m vs. 0.162 m). Similar uncertainty trends were observed in daily WTE and flow calibration equations between D3 and D2. For example, calibration relationships with a similar slope but statistically different intercepts will consistently over-predict or under-predict the expected values during the post-treatment period.

4.2 | Comparison to classic approach using historical calibration data

Field monitoring on watersheds D1, D2, and D3 at the study site started in late 1987, and the data were reported in several studies (e.g., Amatya et al., 1996; 1998; Amatya et al., 2000; Amatya et al., 2003; Amatya & Skaggs, 2011; McCarthy et al., 1991; McCarthy et al., 1992; Skaggs et al., 1994). Those studies provide the chronology of management activities on the three watersheds over the past 25 years. Therefore, a "true-calibration" period was established using data from 1988 to 1990 between watersheds D1 and D2, when both watersheds were in 14- to 15-year-old mature pine stands without any other external disturbances. This calibration relationship had previously been used to quantify effects of various water management and silvicultural operations using a paired watershed approach (Amatya, Skaggs, Blanton, & Gilliam, 2006; Amatya et al., 1996). The 2007–2008 data were recently used to quantify effects of site preparation on water quality (Muwamba et al., 2015). Figure 8 compares the classical calibration relationship between D1 and D2 by using historical data (1988–1990) and the MOSUM-derived calibration relationship developed in this study (2010–2012) after minimizing impacts of external factors. Statistical analysis (analysis of covariance of the two regression lines) indicated that the two relationships were not significantly different at a threshold 0.05 level of significance (Figure 8).

MOSUM-based pretreatment calibration relationships are unique in that they use watershed response data that is most closely tied to the treatment period. The MOSUM approach provides a statistical technique to test this assumption in case external factors shift this consistent relationship. Clausen and Spooner (1993) state that the paired watershed design assumes a consistent, quantifiable, and predictable relationship between watershed response variables, whereas Loftis et al. (2001) illustrate that moderate correlation coefficients ($r \geq 0.6$) are adequate to detect treatment effects for paired watershed studies. For this study, all developed coefficients of determination (R^2) are greater than 0.8 ($r > 0.89$). The choice of data under a stable regression period meets the requirement for the relationship to be consistent by eliminating data that shifts this relationship, whereas the high R^2 meet the requirements for quantifiable and predictable relationships.

4.3 | Comparison of historical and 2012 LAI data

The similarity of the D1 and D2 flow calibration relationships for the historical (1988–1990) and the MOSUM-derived data (2010–2012) can be explained in part by the similarity of the LAI under the two corresponding periods. Figure 9 compares the LAI measured using LiCOR-PCA 2000 after the area between the pine rows on D1 was cleared to

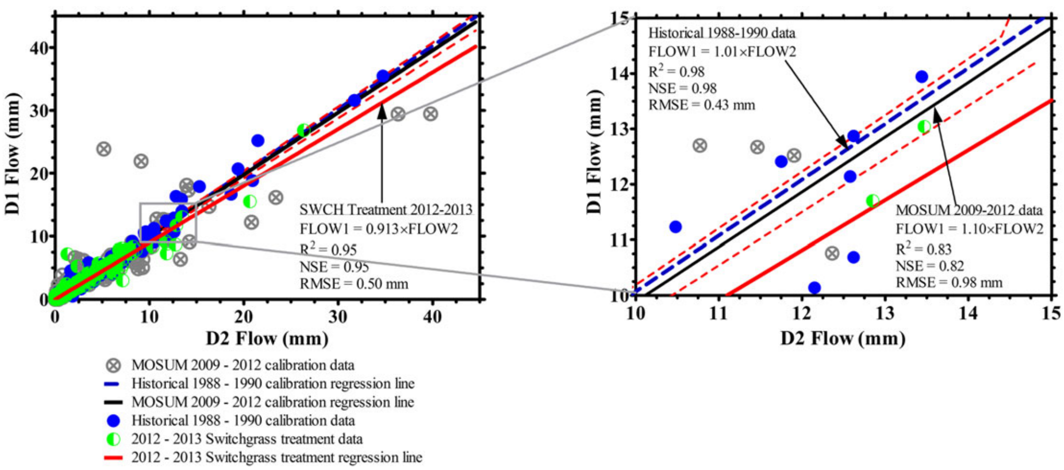

FIGURE 8 Comparison of flow calibration relationships of control and treatment watersheds (D2 and D1, respectively) during two different time periods. For the historical data (1988–1990), both watersheds were under mature pine whereas for the 2009–2012 data, D2 was under 14- to 15-year-old loblolly pine, and D1 was under 3- to 4-year-old loblolly pine. The moving sums (MOSUMs) of recursive residuals data were derived from the 2009–2012 period using only data between March 1, 2010, and March 31, 2012, because it generated the longest structurally stable calibration relationship. Structural stability or temporal stability of regression coefficients refers to non-significant change in the regression coefficients (intercept and slope) based on the analysis of MOSUM of recursive residuals. *NSE* = Nash-Sutcliffe efficiency; RMSE = root mean squared error

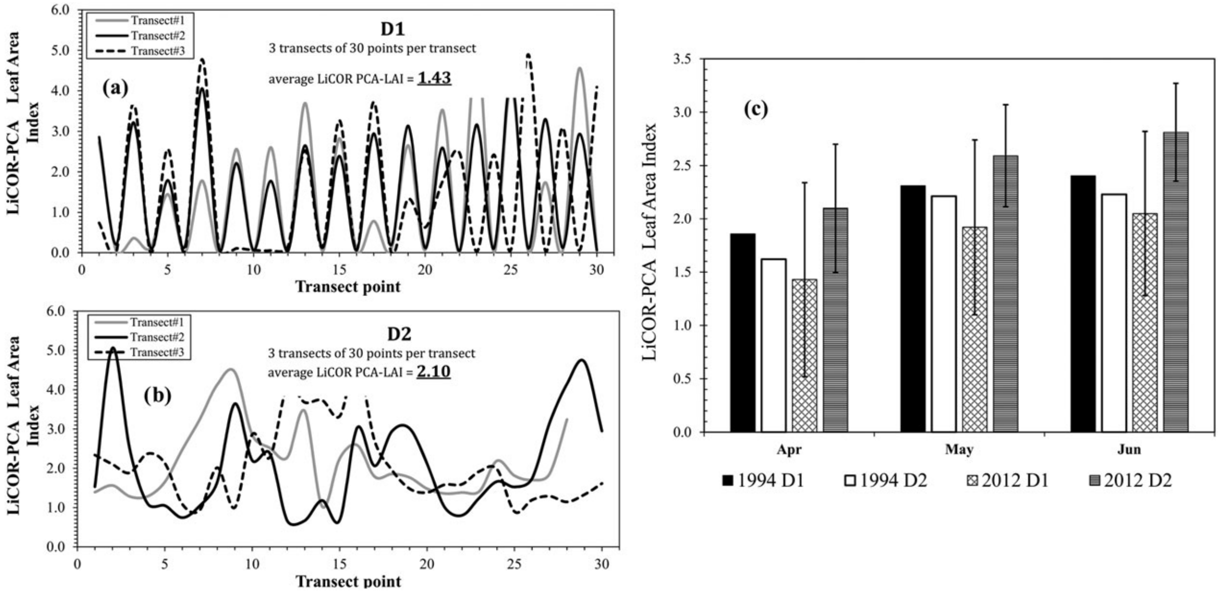

FIGURE 9 Comparison of leaf area index (LAI) of D1 (treatment) and D2 (control) watersheds during 1994 and 2012. Figure 9a,b shows LAI along three transects on watersheds D1 and D2, respectively. The LAI was taken in 2012 after the area between pine rows was cleared for switchgrass planting. Figure 9c compares the LAI of mature loblolly pine (20–21 years) in 1994 to 2012 where D2 was under 14- to 15-year-old loblolly pine and D1 was under 3- to 4-year-old loblolly pine. The error bars refer to associated standard deviations

illustrate that the LAI for watersheds D1 and D2 were still comparable to historical LAI data when both watersheds were under mature pine (about 20–21 years). Although LiCOR-PCA 2000 is known to underestimate actual LAI (Iiames, Congalton, Pilant, & Lewis, 2008), the values presented were measured using the same instrument to allow for data comparison between the two periods. The data choice was also based on LiCOR-PCA LAI data availability. The figure illustrates that even after clearing between pine rows in preparation for switchgrass planting on D1 (Figure 9a), the average LAI on D1 for April–June in 2012 was not significantly different from D2 (Figure 9b, 9c), and both values were comparable to historical data (figure 9c). Earlier work by Sampson et al. (2011) used data from the same study area and highlighted the role of emergent vegetation soon after harvest and early planting for regeneration. LAI may be assumed stable for D1 and D2 between 2010 and early 2012.

4.4 | Experimental design and study data

The experimental setup of a paired watershed approach requires at least one full annual cycle of pretreatment on control and treatment watersheds to capture seasonal variations. The full data used in this study covers the period from 2009 to 2012, during which some treatment watersheds underwent silvicultural management practices (e.g., D0 and D1 were harvested in the first 4 months of 2009). The WTE and flow calibration periods identified by the MOSUM approach as periods of structurally stable relationship between control (D2) and treatment (D1) watersheds began on March 1, 2010, and continued through March 31, 2012, just over 2 years. This period of minimal disturbance was between pine planting and just before the final site preparation for switchgrass broadcast on the intercropped site (D1). Although the MOSUM approach does not eliminate the need for the classic approach using data from a true calibration with both watersheds under same treatment, it minimizes the impacts of external disturbances other than the treatment of interest. Also, the usefulness and significance of this data for calibration is based on the generated quantifiable and predictable relationship between control and treatment watersheds due to the similarity of area, topography, drainage, and soil characteristics in addition to proximity of control and treatment watersheds when compared to other watershed studies. This is demonstrated by all coefficients of determination (R^2) greater than 0.8 (or correlation coefficients >0.89). We recommend further exploration of this approach's application for watersheds that are more seasonally variable or drier.

5 | CONCLUSIONS

Seasonal water table and flow response to rainfall were influenced by both silvicultural management and vegetation on four treatment watersheds in the coastal plain of North Carolina. This study laid out protocols to develop significant, efficient, stable, and predictable calibration relationships required to quantify hydrologic effects of intercropping switchgrass and pine and conversion of a pine forest to switchgrass on a watershed scale. Although the data spans the years from 2009 to 2012, a MOSUM of recursive residuals test was used to detect and omit the periods with instability in regression coefficients potentially due to silvicultural operations. The analysis using MOSUM demonstrated three important findings. First, use of all 2009 to 2012 data gave significantly different calibration relationships between control and treatment watersheds compared to use of only the data spanning the period when the linear regression coefficients of the WTE were stable. Second, the use of the MOSUM test to determine structurally stable relationships between control and treatment watersheds minimized uncertainty of regression relationships, and third, all calibration relationships were quantifiable, statistically significant, and consistent (R^2 and *NSE* greater than 0.85 for WTE and R^2 and

NSE greater than 0.80 for flow), a requirement for use of paired watershed approach. Therefore, though the MOSUM approach does not eliminate the need for true calibration data for a classic paired watershed approach, it does minimize the impacts of external disturbances on the data to develop alternative calibration equations for the paired watershed approach. The MOSUM method should be particularly helpful in studies where long-term monitoring data exists, even with the presence of true calibration data, to develop optimum and significant calibration relationship between the paired watersheds. It is also a useful stand-alone approach for detecting changes due to both natural and anthropogenic external disturbances.

ACKNOWLEDGMENTS

The authors thank Dr. M. Cristina Negri of Argonne National Lab for the physical space, technical, and financial support during preparation of this manuscript. The authors also would like to acknowledge Dr. Darren Miller of Weyerhaeuser Company for helpful review comments during manuscript preparation and Cliff Tyson of Weyerhaeuser Company for assisting with field data collection.

REFERENCES

Albaugh, J. M., Sucre, E. B., Leggett, Z. H., Domec, J.-C., & King, J. S. (2012). Evaluation of intercropped switchgrass establishment under a range of experimental site preparation treatments in a forested setting on the lower coastal plain of North Carolina, USA. *Biomass and Bioenergy*, *46*, 673–682.

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*.

Amatya, D. M., Gilliam, J. W., Skaggs, R. W., Lebo, M. E., & Campbell, R. G. (1998). Effects of controlled drainage on forest water quality. *Journal of Environmental Quality*, *27*, 923–935.

Amatya, D. M., Gregory, J. D., & Skaggs, R. W. (2000). Effects of controlled drainage on storm event hydrology in loblolly pine plantation. *JAWRA Journal of the American Water Resources Association*, *36*, 175–190.

Amatya, D. M., Skaggs, R. W., Gilliam, J. W., & Hughes, J. H. (2003). Effects of orifice-weir outlet on hydrology and water quality of a drained forested watershed. *Southern Journal of Applied Forestry*, *27*, 130–142.

Amatya, D. M., & Skaggs, R. W. (2011). Long-term hydrology and water quality of a drained pine plantation in North Carolina. *Transactions of the ASABE*, *54*, 2087–2098.

Amatya, D. M., Skaggs, R. W., Blanton, C. D., & Gilliam, J. W. (2006). Hydrologic and water quality effects of harvesting and regeneration of a drained pine forest. *Int'l Conference on Hydrology and Management of Forested Wetlands* (ed by. New Bern, NC.

Amatya, D. M., Skaggs, R. W., & Gregory, J. D. (1996). Effects of controlled drainage on the hydrology of drained pine plantations in the North Carolina coastal plain. *Journal of Hydrology*, *181*, 211–232.

Arthur, M. A., Coltharp, G. B., & Brown, D. L. (1998). Effects of best management practices on forest streamwater quality in eastern Kentucky. *JAWRA: Journal of the American Water Resources Association*, *34*, 481–495.

Bai, J. (1994). Least squares estimation of a shift in linear processes. *Journal of Time Series Analysis*, *15*, 453–472.

Bai, J. (1997). Estimating multiple breaks one at a time. *Econometric Theory*, *13*, 315–352.

Bai, J., & Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, 47–78.

Bai, J., & Perron, P. (2003a). Critical values for multiple structural change tests. *The Econometrics Journal*, *6*, 72–78.

Bai, J., & Perron, P. (2003b). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, *18*, 1–22.

Bauer, P., & Hackl, P. (1978). The use of MOSUMS for quality control. *Technometrics*, *20*, 431–436.

Birol, F. (2014). World energy outlook 2013. *International Energy Agency*.

Blanco-Canqui, H., Gantzer, C. J., Anderson, S. H., Alberts, E. E., & Thompson, A. L. (2004). Grass barrier and vegetative filter strip effectiveness in reducing runoff, sediment, nitrogen, and phosphorus loss. *Soil Science Society of America Journal*, *68*, 1670–1678.

Blanton, C., Skaggs, R. W., Amatya, D. M., & Chescheir, G. (1998). Soil hydraulic property variations during harvest and regeneration of drained, coastal pine plantations. *Proc. ASAE Annual Int. Meeting, Orlando, FL* (ed by, pp. 12–16.

Bliss, C. M., & Comerford, N. B. (2002). Forest harvesting influence on water table dynamics in a Florida flatwoods landscape. *Soil Science Society of America Journal*, *66*, 1344–1349.

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.

Brater, E. F., & King, H. W. (1976). *Handbook of hydraulics for the solution of hydraulic engineering problems*. Handbook of hydraulics for the solution of hydraulic engineering problemsMcGraw-Hill.

Bren, L. J., & Lane, P. N. (2014). Optimal development of calibration equations for paired catchment projects. *Journal of Hydrology*, *519*, 720–731.

Brown, R. L., Durbin, J., & Evans, J. M. (1975). Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society. Series B (Methodological)*, 149–192.

Chu, C.-S. J., Hornik, K., & Kaun, C.-M. (1995). MOSUM tests for parameter constancy. *Biometrika*, *82*, 603–617.

Clausen, J., & Spooner, J. (1993). Paired watershed study design. In. Environmental Protection Agency, Washington, DC (United States). Office of Wetlands, Oceans and Watersheds

Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap*. CRC press.

Elkinton, J. S., Healy, W. M., Buonaccorsi, J. P., Boettner, G. H., Hazzard, A. M., & Smith, H. R. (1996). Interactions among gypsy moths, white-footed mice, and acorns. *Ecology*, *77*, 2332–2342.

Fipps, G., & Skaggs, R. (1986). Effect of canal steepage on drainage to parallel drains. *Transactions of the ASAE American Society of Agricultural Engineers*, *29*, 1278–1283.

Friedman, J., Bohonak, A. J., & Levine, R. A. (2013). When are two pieces better than one: Fitting and testing OLS and RMA regressions. *Environmetrics*, *24*, 306–316.

Georgescu, M., & Lobell, D. B. (2010). Perennial questions of hydrology and climate. *Science (Washington)*, *330*, 33.

Ghosh, S. (2009). Import demand of crude oil and economic growth: Evidence from India. *Energy Policy*, *37*, 699–702.

Iiames, J. S., Congalton, R., Pilant, A., & Lewis, T. (2008). Validation of an integrated estimation of loblolly pine (*Pinus taeda* L.) leaf area index (LAI) using two indirect optical methods in the southeastern United States. *Southern Journal of Applied Forestry*, *32*, 101–110.

Jaynes, D. B., Dinnes, D. L., Meek, D. W., Karlen, D. L., Cambardella, C. A., & Colvin, T. S. (2004). Using the late spring nitrate test to reduce nitrate loss within a watershed. *Journal of Environmental Quality*, *33*, 669–677.

Jokela, W. E., & Casler, M. D. (2011). Transport of phosphorus and nitrogen in surface runoff in a corn silage system: Paired watershed methodology and calibration period results. *Canadian Journal of Soil Science*, *91*, 479–491.

de Jong, R., Verbesselt, J., Zeileis, A., & Schaepman, M. E. (2013). Shifts in global vegetation activity trends. *Remote Sensing*, *5*, 1117–1133.

Khaliq, M. N., Ouarda, T. B., Gachon, P., Sushama, L., & St-Hilaire, A. (2009). Identification of hydrological trends in the presence of serial and cross correlations: A review of selected methods and their application to annual flow regimes of Canadian rivers. *Journal of Hydrology*, *368*, 117–130.

King, K. W., Smiley, P. C., Baker, B. J., & Fausey, N. R. (2008). Validation of paired watersheds for assessing conservation practices in the Upper Big Walnut Creek watershed, Ohio. *Journal of Soil and Water Conservation*, *63*, 380–395.

Kundzewicz, Z. W., & Robson, A. J. (2004). Change detection in hydrological records—A review of the methodology. *Hydrological Sciences Journal*, *49*, 7–19.

Laurén, A., Heinonen, J., Koivusalo, H., Sarkkola, S., Tattari, S., Mattsson, T., … Finér, L. (2009). Implications of uncertainty in a pre-treatment dataset when estimating treatment effects in paired catchment studies: Phosphorus loads from forest clear-cuts. *Water, Air, and Soil Pollution*, *196*, 251–261.

Lemke, A. M., Kirkham, K. G., Lindenbaum, T. T., Herbert, M. E., Tear, T. H., Perry, W. L., & Herkert, J. R. (2011). Evaluating agricultural best management practices in tile-drained subwatersheds of the Mackinaw River, Illinois. *Journal of Environmental Quality*, *40*, 1215–1228.

Loftis, J. C., MacDonald, L. H., Streett, S., Iyer, H. K., & Bunte, K. (2001). Detecting cumulative watershed effects: The statistical power of pairing. *Journal of Hydrology*, *251*, 49–64.

McCarthy, E. J., Flewelling, J. W., & Skaggs, R. W. (1992). Hydrologic model for drained forest watershed. *Journal of Irrigation and Drainage Engineering*, *118*, 242–255.

McCarthy, E. J., Skaggs, R. W., & Farnum, P. (1991). Experimental determination of the hydrologic components of a drained forest watershed. *Transactions of the ASAE*, *34*, 2031–2039.

Muwamba, A., Amatya, D. M., Ssegane, H., Chescheir, G. M., Appelboom, T., Tollner, E., Nettles, J. E., Youssef, M. A., Birgand, F., Skaggs, R. W., & Tian, S. (2015). Effects of site preparation for switchgrass-pine forest intercropping on drainage water quality. *[Accepted] Journal of Environmental Quality*, *44*, 1263–1272.

Olmo, J., Pilbeam, K., & Pouliot, W. (2011). Detecting the presence of insider trading via structural break tests. *Journal of Banking & Finance*, *35*, 2820–2828.

Plotnick, R. E. (1989). Application of bootstrap methods to reduced major axis line fitting. *Systematic Biology*, *38*, 144–153.

Politis, D. N. (2003). The impact of bootstrap methods on time series analysis. *Statistical Science*, *18*, 219–230.

Sampson, D. A., Amatya, D. M., Lawson, C. D., & Skaggs, R. W. (2011). Leaf area index (LAI) of loblolly pine and emergent vegetation following a harvest. *Transactions of the ASABE*, *54*, 2057–2066.

Schmer, M. R., Liebig, M. A., Vogel, K. P., & Mitchell, R. B. (2011). Field-scale soil property changes under switchgrass managed for bioenergy. *GCB Bioenergy*, *3*, 439–448.

Skaggs, R. W., Amatya, D. M., Chescheir, G. M., & Blanton, C. D. (2006). Soil property changes during loblolly pine production. In: *2006 ASABE Annual International Meeting*. American Society of Agricultural and Biological Engineers, Portland, OR.

Skaggs, R. W., Breve, M. A., & Gilliam, J. W. (1994). Hydrologic and water quality impacts of agricultural drainage*. *Critical Reviews in Environmental Science and Technology*, *24*, 1–32.

Sørensen, R., Ring, E., Meili, M., Högbom, L., Seibert, J., Grabs, T., … Bishop, K. (2009). Forest harvest increases runoff most during low flows in two boreal streams. *Ambio: A Journal of the Human Environment*, *38*, 357–363.

Ssegane, H., Amatya, D. M., Chescheir, G. M., Skaggs, W. R., Tollner, E. W., & Nettles, J. E. (2013). Consistency of hydrologic relationships of a paired watershed approach. *American Journal of Climate Change*, *2*, 147.

Tiwari, A. K., Shahbaz, M., & Islam, F. (2012). Does financial development increase rural-urban income inequality? Cointegration analysis in the case of Indian economy. *International Journal of Social Economics*, *40*, 4–4.

Tomer, M. D., & Schilling, K. E. (2009). A simple approach to distinguish land-use and climate-change effects on watershed hydrology. *Journal of Hydrology*, *376*, 24–33.

USGS. (1997). National handbook of recommended methods for water data acquisition. Office of Water Data Coordination, Geological survey, US Department of the Interior, Reston, VA.

Verbesselt, J., Hyndman, R., Newnham, G., & Culvenor, D. (2010). Detecting trend and seasonal changes in satellite image time series. *Remote Sensing of Environment*, *114*, 106–115.

Verbesselt, J., Zeileis, A., & Herold, M. (2012). Near real-time disturbance detection using satellite image time series. *Remote Sensing of Environment*, *123*, 98–108.

Vogl, A. L., & Lopes, V. L. (2009). Evaluating watershed experiments through recursive residual analysis. *Journal of Irrigation and Drainage Engineering*, *136*, 348–353.

Webb, A. A., Kathuria, A., & Turner, L. (2012). Longer-term changes in streamflow following logging and mixed species eucalypt forest regeneration: The Karuah experiment. *Journal of Hydrology*, *464*, 412–422.

Xu, Y.-J., Burger, J. A., Aust, W. M., Patterson, S. C., Miwa, M., & Preston, D. P. (2002). Changes in surface water table depth and soil physical properties after harvest and establishment of loblolly pine (*Pinus taeda* L.) in Atlantic coastal plain wetlands of South Carolina. *Soil and Tillage Research*, *63*, 109–121.

Zegre, N. (2008). The history and advancement of change detection methods in Forest hydrology. Society of American Foresters Water Resources Working Group Newsletter, Bethesda, MD.

Zégre, N., Skaugset, A. E., Som, N. A., McDonnell, J. J., & Ganio, L. M. (2010). In lieu of the paired catchment approach: Hydrologic model change detection at the catchment scale. *Water Resources Research*, 46.

Zeileis, A., Leisch, F., Hornik, K., Kleiber, C., Hansen, B., & Zeileis, M. A. (2013) Package 'strucchange'.

How to cite this article: Ssegane H, Amatya DM, Muwamba A, et al. Calibration of paired watersheds: Utility of moving sums in presence of externalities. *Hydrological Processes*. 2017;1–14. https://doi.org/10.1002/hyp.11248