WATER RESOURCES RESEARCH, VOL. 39, NO. 9, 1252, doi:10.1029/2003WR002081, 2003

A distribution-free test to detect gradual changes in watershed behavior

Vazken Andréassian

Cemagref, Water Quality and Hydrology Research Unit, Antony, France

Eric Parent

ENGREF, Laboratoire de Gestion des Risques en Sciences de l'Eau, Paris, France

Claude Michel

Cemagref, Water Quality and Hydrology Research Unit, Antony, France

Received 21 February 2003; revised 23 May 2003; accepted 24 June 2003; published 17 September 2003.

[1] This paper presents a distribution-free statistical test aimed at detecting gradual changes in the hydrological behavior of watersheds. The proposed test uses a rainfall-runoff model to identify watershed behavior over successive time periods and a resampling approach to quantify the significance of trends. The method can be applied with any model deemed suitable for the studied watershed. To assess test efficiency, we used three different case studies: An afforested agricultural watershed, a burnt-over forested watershed, and a watershed covered by old-growth forest. All three watersheds had a long period of rainfall and runoff records (60, 35, and 40 years, respectively), on which stationarity could be tested. The test was shown to adequately detect gradual changes, and it can therefore be useful to identify hydrological trends, wherever rainfall and streamflow time series are available.

INDEX TERMS: 1803 Hydrology: Anthropogenic effects; 1860 Hydrology: Runoff and streamflow; 1821 Hydrology: Floods; 1812 Hydrology: Drought; *KEYWORDS:* rainfall-runoff modeling, trend analysis, reforestation, distribution-free test, Coshocton, Réal Collobrier, Andrews LTER

Citation: Andréassian, V., E. Parent, and C. Michel, A distribution-free test to detect gradual changes in watershed behavior, *Water Resour. Res.*, *39*(9), 1252, doi:10.1029/2003WR002081, 2003.

1. Introduction

1.1. Reasons to Look for Changes in Watershed Behavior

[2] Humans need a reliable supply of water, and today the stability of water resources is more than ever a matter of concern, especially in dry areas. The sustainability of water supplies depends both on climate stability and the stationarity of watershed hydrological behavior. Here we focus on the second phenomenon: The change in watershed hydrological behavior, which corresponds to a change in the rainfall-runoff relationship at the watershed scale.

[3] Suspicion of nonstationarity in the hydrological behavior of watersheds has caused growing concern among the public as well as among watershed managers in the past few decades. During this time, floods have occurred repeatedly in Europe: On the Meuse River in 1993, 1994, and 1995; on the Oder River in 1997 and 2002; and on the Vilaine River in 1995 and 2001. Each one of these flood events invariably raised the same question in the press and elsewhere: Can recent land-use changes be held responsible for the apparent increase in disastrous events?

1.2. Possible Approaches to Assess Changes in Watershed Behavior

[4] Forecasting the consequences of land-use changes is one of the important tasks of modern hydrology scientists. Unfortunately, it must be recognized that we are still unable to provide a satisfactory answer to this question. Many hydrologists even consider that "there are still no credible models to predict the effect on hydrological response of land-use change in gauged catchments" [*Kokkonen and Jakeman*, 2002, p. 303].

[5] Therefore we must, for the time being, settle for an intermediate objective, i.e., the detection of watershed behavior changes once they have occurred. Throughout the twentieth century, experimental watersheds and the paired watershed principle have allowed hydrologists in general, and forest hydrologists in particular, to detect and study the impact of different types of treatments on small watersheds: deforestation, partial cutting, thinning, etc. However, there is now a need to assess possible changes both at a larger spatial scale, and at a much more gradual pace, as they occur in the real world. (Most watershed experiments have been concerned with detecting the impact of sudden changes such as clear-cutting a forest.)

[6] A recent report published by the World Meteorological Organization [*Kundzewicz and Robson*, 2000] presents the state of the art in trend detection in hydrological time series. Interestingly, all the methods presented in this report deal either with rainfall or runoff: There is no mention of any kind of methods for detecting trends in watershed behavior. We take this as clear evidence of a lack of research on this issue, which illustrates the need for methods to detect such trends.

0043-1397/03/2003WR002081$09.00

[7] This article will explore the ability of a statistical method to detect changes in the rainfall-runoff relationship at the watershed scale. In section 2, we review existing approaches to trend detection. After presenting a new statistical method in section 3, evaluation will be discussed in section 4, with section 5 illustrating its application using data from three experimental watersheds.

2. Relevant Literature

[8] For the study of watershed behavior changes, the methods available today are based either on paired watershed approaches (section 2.1) or on an adaptation of this concept involving a rainfall-runoff model (section 2.2).

2.1. The Paired Watershed Approach: A Reference Method for Detecting Changes in Hydrological Behavior

[9] The paired watershed approach involves selecting two similar watersheds, which are monitored simultaneously for a given time period to establish a univocal relationship between their hydrological behaviors [*Hewlett*, 1982]. Then, a land-cover treatment is applied to one of them, while the second remains unchanged. After the treatment, the initially derived relationship is used to reconstitute the behavior of the treated watershed (Figure 1). By comparing the actual (measured) flows and the reconstituted flows, we can assess the hydrological impact of the watershed treatment.

[10] Experiments using the paired watershed approach started at the beginning of the twentieth century, with the first one conducted in 1910 at Wagon Wheel Gap, Colorado [*Bates and Henry*, 1928]. Several reviews on this subject are available [*Hibbert*, 1967; *Rodda*, 1976; *Bosch and Hewlett*, 1982; *Hornbeck et al.*, 1993; *Stednick*, 1996]. Paired watersheds have been used to detect the impact of deforestation, reforestation, construction of forest roads, urbanization, etc. This method is particularly valuable in that the extent of change can be quantified in an easily interpretable way, i.e., directly in terms of an increase or decrease in water amounts, and it is efficient for detecting both sudden and gradual changes.

2.2. Rainfall-Runoff Models: An Alternative to Paired Watershed Schemes?

[11] Unfortunately, leaving the realm of experimental watersheds to investigate a trend in hydrological behavior on a real-world watershed, it is often impossible to identify a control basin, or to find enough time and money to design a controlled experiment. Usually, only rainfall and runoff records are available for the treated watershed. In these conditions, it is difficult to assess the effects of watershed behavior change, as prechange and postchange periods may differ in terms of climate [*Hewlett*, 1982; *Cosandey and Robinson*, 2000], thus blurring the signal to be analyzed. *Box et al.* [1975] describe in detail how fallacious statistical results can be produced by working with happenstance data without the framework of experimental design.

[12] To try to return to a paired watershed design, simulating a virtual control watershed using a rainfall-runoff model is possible. A common practice consists in calibrating a before-treatment model and using it as a virtual control watershed along with observed rainfall to reconstitute runoff after treatment, as if no change in the watershed behavior had occurred between the first and the second period. Treatment effects are then deducted by comparing the simulated and observed flows, with the hypothesis that the model fully captures the basin behavior under the conditions of the first time period. A review of such methods is given by *Refsgaard et al.* [1989] and *Lørup et al.* [1998]. There are several examples of studies using rainfall-runoff models to detect hydrological changes:

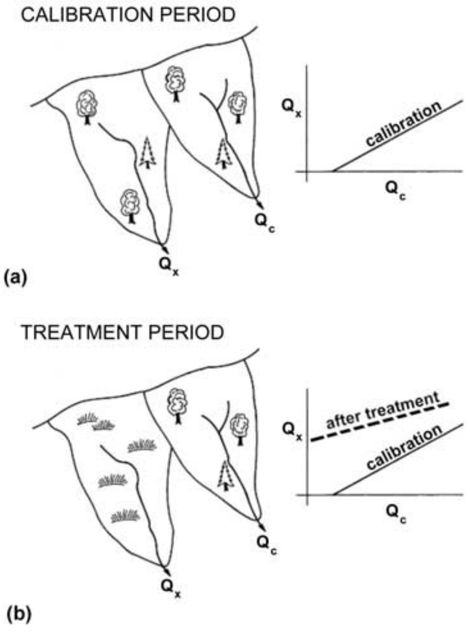


Figure 1. Sketch of a paired watershed experiment [from *Hewlett*, 1982].

[13] 1. *Kuczera* [1987] used this approach to analyze water consumption of forest regrowth in Australia. He used an annual water balance model and studied residuals corresponding to the difference between observed data and predicted annual flows.

[14] 2. *Brandt et al.* [1988] studied the impact of clear-cutting on the hydrological response of small forested watersheds: They calibrated a rainfall-runoff model before clear-cutting and used calibrated parameters and observed rainfall after clear-cutting to simulate flows, which were then compared to observed flows.

[15] 3. *Cornish* [1993] studied changes in the annual water yield following forest exploitation and compared the paired watershed approach and the modeling approach (using an annual water balance model). He concluded that both approaches give similar results, though the use of a model results in greater associated uncertainty.

[16] 4. *Lavabre et al.* [1993] used a rainfall-runoff model at both monthly and daily time steps to assess hydrological changes following a wildfire, comparing observed flows and flows simulated with model parameters obtained before the fire.

[17] 5. *Nascimento* [1995] tested several ways of detecting changes by means of a rainfall-runoff model: He used a method based on the study of model parameter changes, as well as a method where he compared the outcome of model simulations using the same input series, and gave his preference to the latter.

[18] 6. *Schreider et al.* [2002] used an innovative approach: They calibrated a rainfall-runoff model at the beginning of their study period and then used the precipitation data to simulate flow over the entire study period. They computed simulation residuals (simulated minus measured flows), and then applied classic trend analysis techniques to the residual time series.

[19] Although the classic approach (involving real control watersheds) is efficient, it is seldom applicable in the real world, where watershed managers deal with problems at much larger scales: Hence hydrologists have to rely on a rainfall-runoff model to replace the control watershed. In this context, the method proposed by *Schreider et al.* [2002] is a step forward, as it improves the assessment of gradual changes (trends), although it does not provide a statistical test to evaluate the credibility of a trend.

3. A New Approach Involving Resampling of Model Simulations to Assess Trends in Watershed Behavior

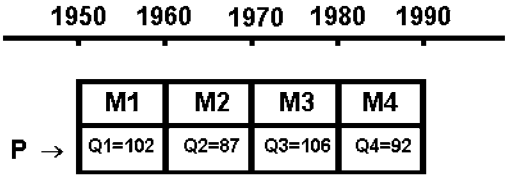
[20] In this section, we will present a method using a rainfall-runoff model to detect a gradual change in the hydrological behavior of a watershed.

3.1. Cross Simulations as a Basis for Statistical Testing

[21] In the following discussion, we consider a rainfall-runoff model, which can be calibrated against a time series of observed runoff and areal precipitation, to yield a set of parameters representative of watershed behavior during the calibration period. Note that this method is not specific to a given model, nor to a given time step (annual, monthly, or daily data could be used). Once model parameters are defined, the rainfall-runoff model can be used to simulate flows corresponding to a given rainfall time series, and a hydrological target variable can be computed. For the present discussion, let us consider that the target variable is total streamflow. The proposed method works in three steps:

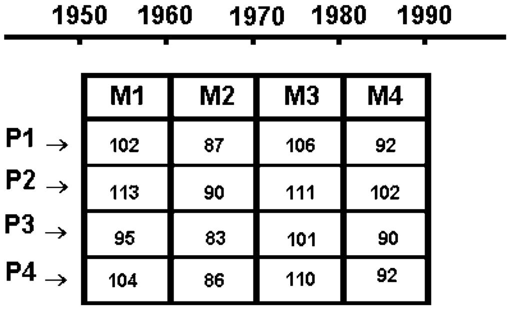
[22] 1. The period of study is divided into *n* successive periods of equal length, which must allow for a proper calibration of the rainfall-runoff model (2–3 years at least). For this example, 1950–1989 is the total period, with four subperiods of 10 years: 1950–1959, 1960–1969, 1970–1979, and 1980–1989. Calibration of the rainfall-runoff model yields four successive models of watershed behavior (defined by four different parameter sets): M1, M2, M3, and M4.

[23] 2. We can now use an observed rainfall time series in conjunction with the four optimized parameter sets and obtain the corresponding total streamflow simulated by the four models of watershed behavior (Figure 2). Note that these model outputs can be compared, since they have been simulated with the same precipitation time series as input.



	M1	M2	M3	M4
P →	Q1=102	Q2=87	Q3=106	Q4=92

Figure 2. Simulation of watershed flow (in mm) by the four models of watershed behavior, using the same precipitation time series P.



	M1	M2	M3	M4
P1 →	102	87	106	92
P2 →	113	90	111	102
P3 →	95	83	101	90
P4 →	104	86	110	92

Figure 3. Cross-simulation matrix gathering the results of flow simulations for four models (M1–M4) successively using four precipitation series (P1–P4).

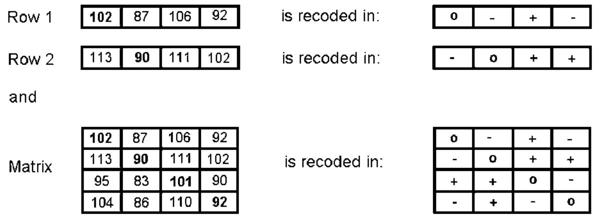
[24] 3. This operation can be repeated as long as independent precipitation time series are available, and the results can be displayed in matrix form. The more rows in the matrix, the more exhaustive the description of watershed behavior. At a minimum, we will obtain a square matrix (Figure 3), since the time series that were available for calibration can be used as precipitation input. In this "cross-simulation matrix" we find, in each cell (i, j), the total streamflow simulated with the model Mj (representative of watershed behavior during period j) using precipitation Pi.

3.2. Visual Analysis of Cross-Simulation Matrices

[25] Before performing any statistical test, a great deal of information on possible changes can be gained visually, by simply transforming the cross-simulation matrix figures into signs (plus or minus). Remember that for each row of the matrix, flows can be compared since they were simulated with the same precipitation input. In each row, we propose to take as a reference the value found on the matrix diagonal, i.e., the value simulated by model M_i using precipitation P_i (since P_i is used to calibrate M_i, the value $Q_{i,i}$ seems the most logical reference for the comparison). Thus, in case of a seemingly positive trend, we replace $Q_{i,j}$ with a plus sign. This happens for $j < i$, if $Q_{i,j} < Q_{i,i}$; and for $j > i$, if $Q_{i,j} > Q_{i,i}$. In case of a seemingly negative trend, we replace $Q_{i,j}$ with a minus sign. This happens for $j < i$, if $Q_{i,j} > Q_{i,i}$; and for $j > i$, if $Q_{i,j} < Q_{i,i}$. If $Q_{i,j} = Q_{i,i}$, we replace $Q_{i,j}$ with a letter O. An example of recoding is shown in Figure 4.

[26] Recoding simulated flow into signs provides a good basis for fast visual interpretation: A matrix with a majority of pluses means a gradual increase in the ability to produce the target flow, while a matrix with comparable numbers of pluses and minuses means overall stability. Moreover, recoded matrices can help identify periods of homogeneous behavior, to which the stationarity test can be applied (for example, to distinguish a period with a trend from a period of stability).

[27] However, this recoded matrix is less informative for a statistical test, since it only takes into account the sign of



Row 1

102	87	106	92

is recoded in:

o	-	+	-

Row 2

113	**90**	111	102

is recoded in:

-	o	+	+

and

Matrix

102	87	106	92
113	**90**	111	102
95	83	**101**	90
104	86	110	**92**

is recoded in:

o	-	+	-
-	o	+	+
+	+	o	-
-	+	-	o

Figure 4. Recoding of cross-simulation matrices for visual interpretation.

change and not its intensity. Therefore, in the following section, we use the original cross-correlation matrix to develop our test.

3.3. A Resampling Scheme Using Permutation

[28] Among the many existing statistical methods, the nonparametric techniques [*Conover*, 1980] do not require the specification of a parametric structure for the null and the alternative hypotheses. In the nonparametric family, we found the methods based on resampling approaches particularly attractive, as they are typically distribution-free, i.e., they do not require the assumption of a particular distribution of the data when the test is made [*Robson et al.*, 2000]. They are thus well suited to hydrological applications. The null hypothesis (H_0) for which the test is designed is the absence of trend. If H_0 holds, the chronological order of the periods of observation is not important, and data can be shuffled many times; this is how a permutation test works. After each shuffle, a test statistic is calculated, so that at the end of the permutation round, a distribution of possible values of the test statistic under H_0 has been generated [*Robson*, 2000].

[29] The test is then based on the analysis of the relative position of the observed test statistic (i.e., its value for the original data) within the permutation distribution: If the observed statistic is somewhere in the middle of the distribution, we conclude that there is no evidence of a trend, i.e., there is no reason to reject H_0 (Figure 5a). If the observed statistic is larger or smaller than almost all the values of the distribution, we reject H_0 and we conclude that a trend exists, given that it seems unlikely that such a value is the result of chance (Figure 5b). Note that, as the test is based on a comparison of model simulations with different parameter sets (each representing a different subperiod), it is essential that model calibration is implemented in a comparable way. If manual calibration is used, we would therefore advise focusing on a single objective function, to avoid the subjectivity that some manual methods might introduce in the choice of the optimal parameter set.

3.4. Choice of a Representative Statistic

[30] Choosing a suitable test statistic to interpret the results of a permutation test is an important step in the definition of the test. The statistic S chosen to characterize the cross-simulation matrices is defined by

$$S = \sum_{i=1}^{n} \left[\sum_{j=1}^{i-1} \left(q_{ii} - q_{ij} \right) + \sum_{j=i+1}^{n} \left(q_{ij} - q_{ii} \right) \right] \tag{1}$$

where q_{ij} is the element found in the ith row and in the jth column.

[31] S can be interpreted as follows: A matrix showing a decreasing trend will have, in each row i, all q_{ij} mainly ranked in decreasing order. Thus S will be negative and of large absolute value. Symmetrically, for a matrix showing an increasing trend, S will be large and positive. In the absence of a trend, the absolute value of S will be low.

[32] In the proposed test, the distribution illustrated in Figure 5 represents the distribution of the S values obtained by shuffling the original matrix. The p-value synthesizes the results of the test: it corresponds to the probability of obtaining, within the null hypothesis, a value of the statistic smaller (or greater) than the statistic computed on the original matrix (depending on its value relative to the median of the distribution).

4. Methodology for Test Validation

[33] To assess the robustness of the proposed methodology, we will present three examples where the test is applied to well-documented case studies.

4.1. North Appalachian Experimental Watershed (NAEW)

[34] The North Appalachian Experimental Watershed (NAEW) is located in Coshocton, Ohio, USA. This exper-

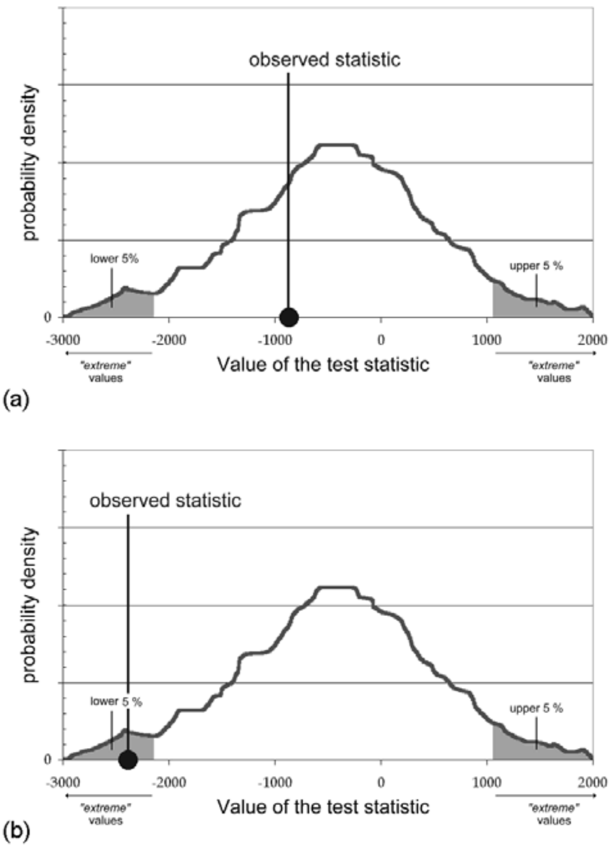
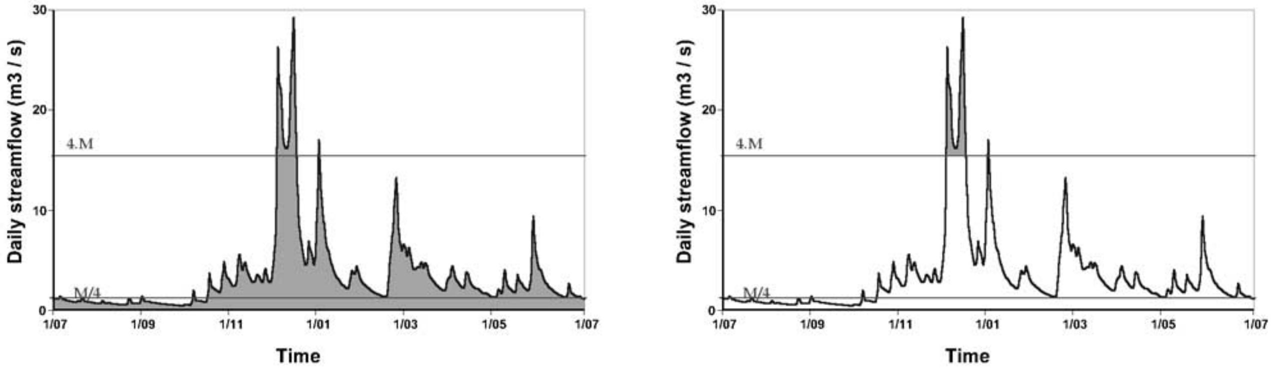
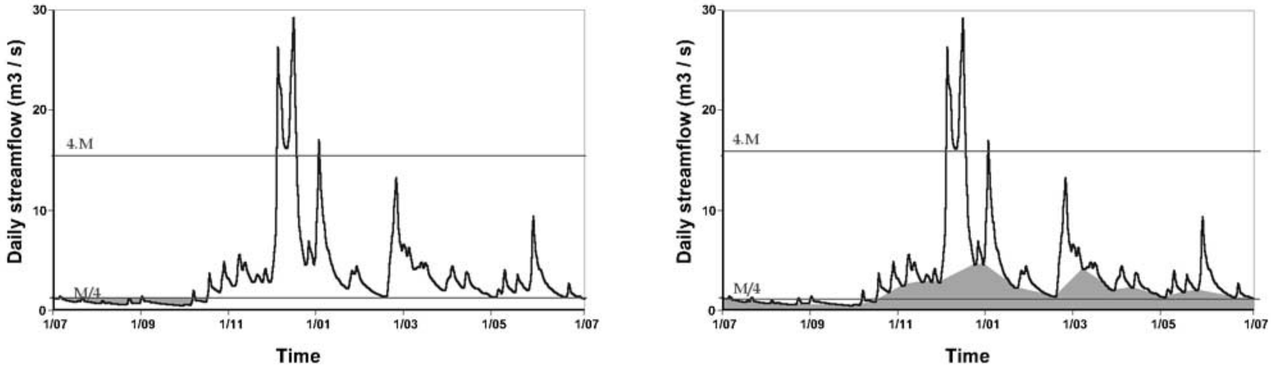


Figure 5. Two possibilities for the relative position of the observed (original) statistic within the distribution obtained by permutation. (a) The observed statistic is somewhere in the middle of the distribution; we conclude that there is no evidence of a trend. (b) The observed statistic is larger or smaller than almost all the values of the distribution; we reject H_0 and we conclude that a trend is present. See color version of this figure in the HTML.



(1) total flow

(2) flood flow



(3) low-flow deficit

(4) base flow

Figure 6. Variables considered for trend analysis representing the four facets of a hydrograph. Horizontal lines identify thresholds defining floods (4 times mean discharge M) and low flows (one fourth of mean discharge M). See color version of this figure in the HTML.

imental watershed has been managed by the U.S. Department of Agriculture (USDA) Agricultural Research Service since 1935. One of the subwatersheds of the NAEW, watershed 172 (0.18 km^2), was reforested in 1938. The forest plantation was first thinned 30 years later, in 1967–1970 [*McGuinness and Harrold*, 1971]. The changes in hydrological behavior over the period 1938–1967 were documented by *Langford and McGuinness* [1976]; rainfall and runoff records are available for the period 1939–1999, with a short interruption between 1972 and 1975.

4.2. Réal Collobrier Research Watershed (RCRW)

[35] The Réal Collobrier research watershed is located in the South of France, close to the Mediterranean Sea. It has been managed by Cemagref since 1967. One of its subwatersheds, the Rimbaud watershed (1.5 km^2), was covered by *maquis* shrubland, which was entirely burnt over by a wildfire in August 1990 [*Lavabre et al.*, 1993]. The vegetation recovered in a couple of years. Daily rainfall and runoff records are available for the period 1967–2000.

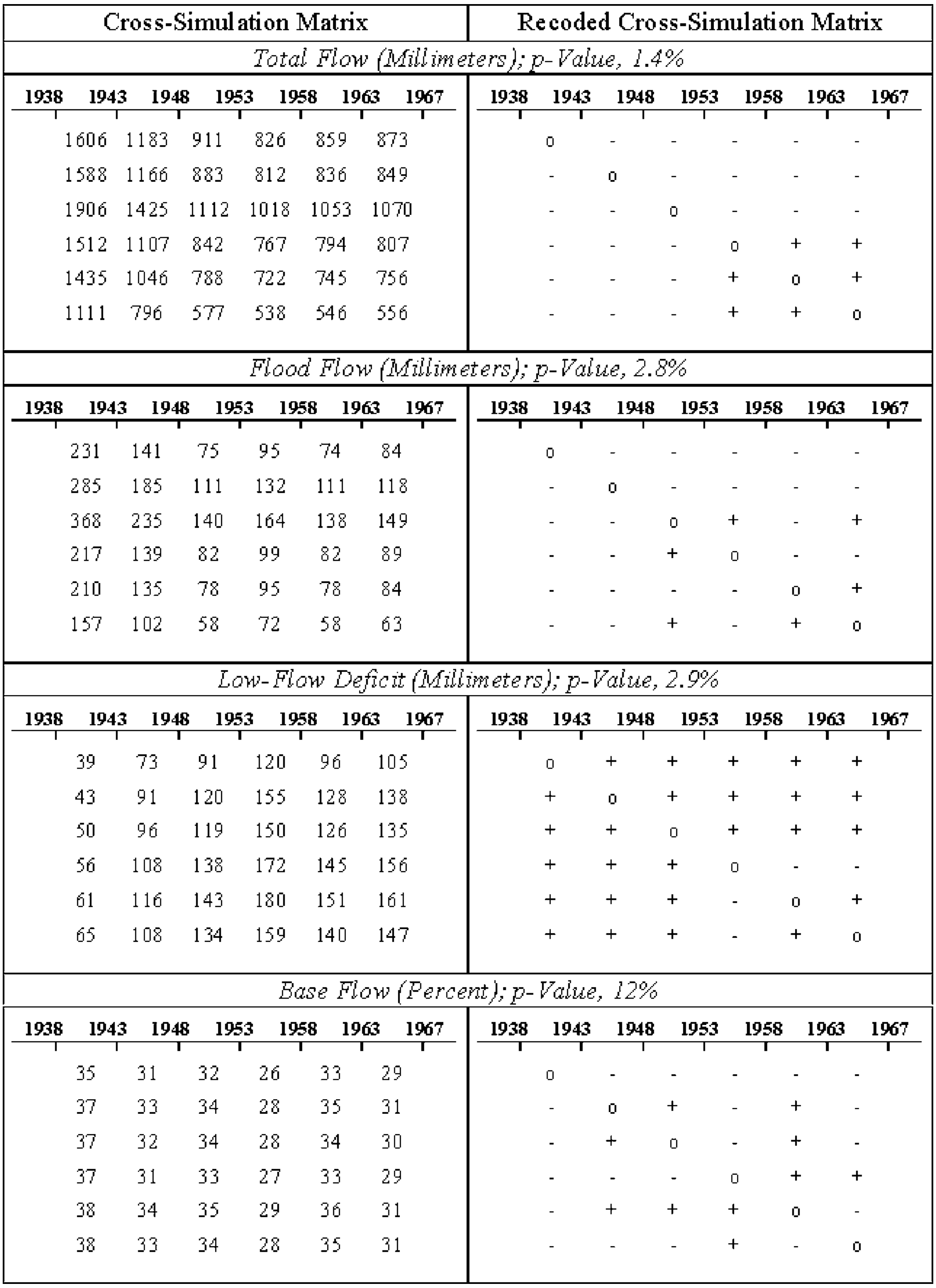
4.3. Andrews Experimental Forest (AEF)

[36] The H. J. Andrews Experimental Forest (now LTER) is located in Oregon, USA. This experimental forest has been managed by the USDA Forest Service since 1952. Watershed 2 (0.6 km^2) is a control watershed, covered by a forest of old-growth Douglas fir, which has been left untouched since the beginning of the experiment [*Post and Jones*, 2001]. Daily rainfall and runoff data were obtained for a period of 42 years (1958–1999).

4.4. The GR4J Daily Rainfall-Runoff Model

[37] As already mentioned, the test proposed in this article can be implemented with any rainfall-runoff model requiring calibration. Note, however, that the capacity of the test to detect the more subtle changes will be quite heavily dependent on the efficiency and robustness of the model (i.e., its capacity to represent the rainfall-runoff relationship equally well over all calibration subperiods, in calibration and in validation). Therefore it seems quite natural to recommend avoiding overparametrized rainfall-runoff modeling structures, as they tend to lack robustness.

[38] For the application to the case studies, we used GR4J, a reliable, continuous lumped rainfall-runoff model functioning at a daily time step [*Perrin et al.*, 2003]. A detailed discussion of the model is not within the scope of this paper. We will only mention that the GR4J model showed satisfactory versatility and robustness in the comparative study proposed by *Perrin et al.* [2001]. The model depends on four calibrated parameters, which account for water balance (X1, capacity of production store; X2, water exchange coefficient) and water transfer (X3, capacity of the nonlinear routing store; X4, unit hydrograph time base). The small number of parameters prevents problems of overparametrization.



Cross-Simulation Matrix | Recoded Cross-Simulation Matrix

Total Flow (Millimeters); p-Value, 1.4%

1938 1943 1948 1953 1958 1963 1967

1606	1183	911	826	859	873	o	-	-	-	-	-
1588	1166	883	812	836	849	-	o	-	-	-	-
1906	1425	1112	1018	1053	1070	-	-	o	-	-	-
1512	1107	842	767	794	807	-	-	-	o	+	+
1435	1046	788	722	745	756	-	-	-	+	o	+
1111	796	577	538	546	556	-	-	-	+	+	o

Flood Flow (Millimeters); p-Value, 2.8%

1938 1943 1948 1953 1958 1963 1967

231	141	75	95	74	84	o	-	-	-	-	-
285	185	111	132	111	118	-	o	-	-	-	-
368	235	140	164	138	149	-	-	o	+	-	+
217	139	82	99	82	89	-	-	+	o	-	-
210	135	78	95	78	84	-	-	-	-	o	+
157	102	58	72	58	63	-	-	+	-	+	o

Low-Flow Deficit (Millimeters); p-Value, 2.9%

1938 1943 1948 1953 1958 1963 1967

39	73	91	120	96	105	o	+	+	+	+	+
43	91	120	155	128	138	+	o	+	+	+	+
50	96	119	150	126	135	+	+	o	+	+	+
56	108	138	172	145	156	+	+	+	o	-	-
61	116	143	180	151	161	+	+	+	-	o	+
65	108	134	159	140	147	+	+	+	-	+	o

Base Flow (Percent); p-Value, 12%

1938 1943 1948 1953 1958 1963 1967

35	31	32	26	33	29	o	-	-	-	-	-
37	33	34	28	35	31	-	o	+	-	+	-
37	32	34	28	34	30	-	+	o	-	+	-
37	31	33	27	33	29	-	-	-	o	+	+
38	34	35	29	36	31	-	+	+	+	o	-
38	33	34	28	35	31	-	-	-	+	-	o

Figure 7. Results obtained for Coshocton WS172 over the period 1938–1967.

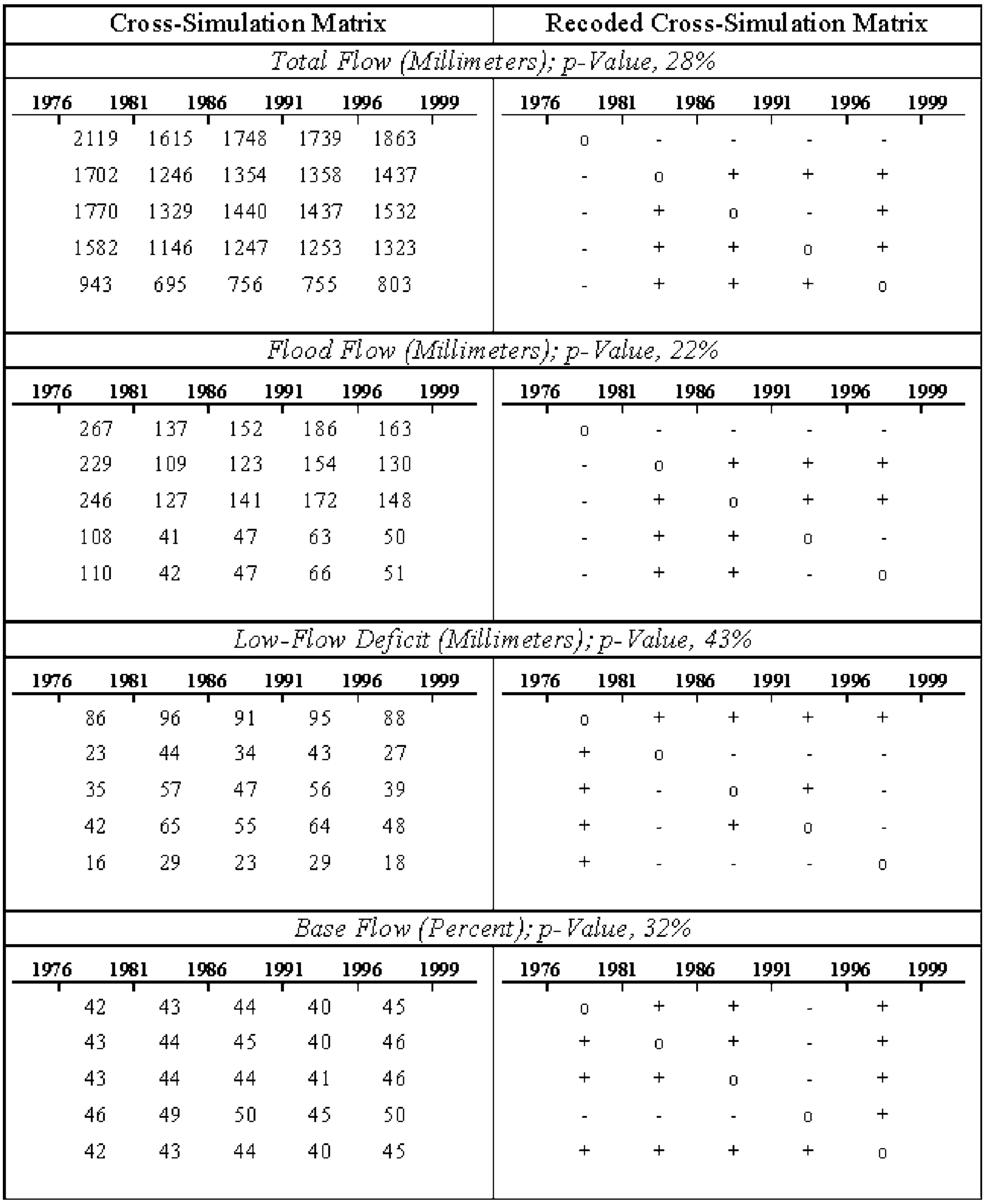
4.5. Four Target Variables to Analyze the Various Facets of Watershed Hydrology

[39] Four hydrological variables were used to characterize watershed behavior as completely as possible. These four variables make it possible to take into account the various facets of the hydrological regime (Figure 6): (1) total flow, (2) flood flows, (3) low flows, and (4) base flow. Flood flows are defined as the sum of daily flows above a threshold (arbitrarily taken equal to 4 times the mean annual discharge). Similarly, low flows are defined as the deficit of flow below a threshold (by convention, equal to one fourth of the mean annual discharge): This quantity is hereinafter called low-flow deficit. Base flow is computed according to the method proposed by *L'vovitch* [1979] and expressed as a percentage of total flow.

5. Discussion of Case Study Results

5.1. An Adequate Detection of the Impact of Forest Growth

[40] The North Appalachian experimental watershed 172 in Coshocton offers an excellent opportunity to



Total Flow (Millimeters); p-Value, 28%

Cross-Simulation Matrix					Recoded Cross-Simulation Matrix				
1976 1981 1986 1991 1996 1999					1976 1981 1986 1991 1996 1999				
2119	1615	1748	1739	1863	o	-	-	-	-
1702	1246	1354	1358	1437	-	o	+	+	+
1770	1329	1440	1437	1532	-	+	o	-	+
1582	1146	1247	1253	1323	-	+	+	o	+
943	695	756	755	803	-	+	+	+	o

Flood Flow (Millimeters); p-Value, 22%

Cross-Simulation Matrix					Recoded Cross-Simulation Matrix				
1976 1981 1986 1991 1996 1999					1976 1981 1986 1991 1996 1999				
267	137	152	186	163	o	-	-	-	-
229	109	123	154	130	-	o	+	+	+
246	127	141	172	148	-	+	o	+	+
108	41	47	63	50	-	+	+	o	-
110	42	47	66	51	-	+	+	-	o

Low-Flow Deficit (Millimeters); p-Value, 43%

Cross-Simulation Matrix					Recoded Cross-Simulation Matrix				
1976 1981 1986 1991 1996 1999					1976 1981 1986 1991 1996 1999				
86	96	91	95	88	o	+	+	+	+
23	44	34	43	27	+	o	-	-	-
35	57	47	56	39	+	-	o	+	-
42	65	55	64	48	+	-	+	o	-
16	29	23	29	18	+	-	-	-	o

Base Flow (Percent); p-Value, 32%

Cross-Simulation Matrix					Recoded Cross-Simulation Matrix				
1976 1981 1986 1991 1996 1999					1976 1981 1986 1991 1996 1999				
42	43	44	40	45	o	+	+	-	+
43	44	45	40	46	+	o	+	-	+
43	44	44	41	46	+	+	o	-	+
46	49	50	45	50	-	-	-	o	+
42	43	44	40	45	+	+	+	+	o

Figure 8. Results obtained for Coshocton WS172 over the period 1976–1999.

validate the proposed statistical test, as it has already been studied with classic tests by *McGuinness and Harrold* [1971] and found to show a clear trend over the 1938–1967 period.

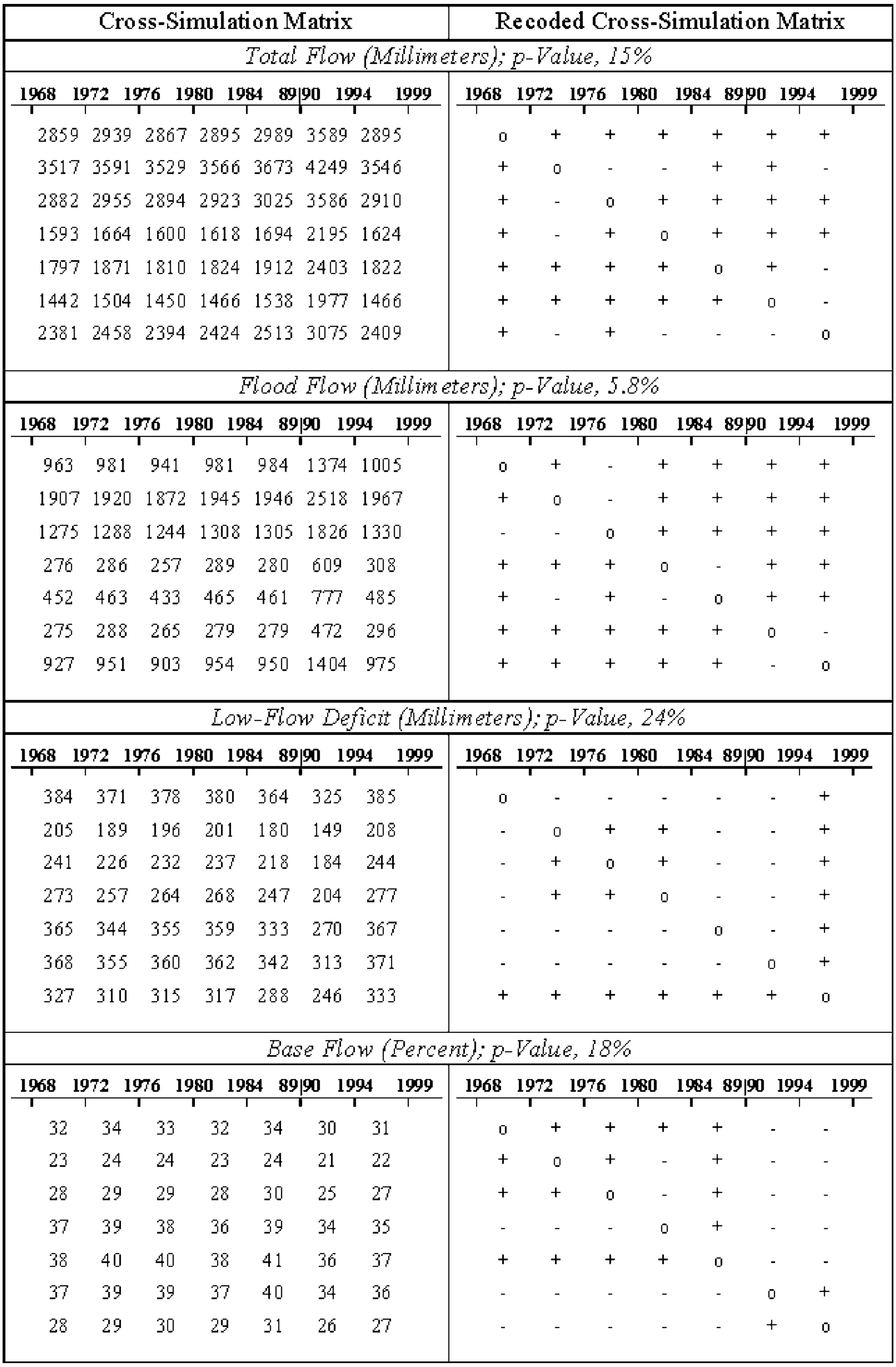
5.1.1. Application of the Test to the Forest Growth Period (1938–1967)

[41] Figure 7 shows the original and recoded cross-simulation matrices for the four hydrological variables defined in section 0. A visual analysis of the recoded matrices shows that there is a large majority of minuses for total and flood flows, which suggests a decreasing trend, while a large majority of pluses for the low-flow deficit suggests a positive trend (i.e., a decrease in the discharge in low-flow periods). The situation is much less contrasted for base flow.

[42] The *p*-values computed by the proposed resampling test lead to the same conclusion: A significant trend ($p < 5\%$) is found for the first three hydrological variables, confirming the results of *McGuinness and Harrold* [1971]: The afforestation in 1938 has led to a progressive change of watershed behavior in 30 years. Total flow has been reduced by half, flood flows have also been reduced, while the low-flow deficit has increased. There is no evidence of change in base flow on this watershed.

5.1.2. Application of the Test to a Steady Period (1976–1999)

[43] Figure 8 shows the matrices obtained over the 1976–1999 period, following forest thinning and a short interruption of gaging. No trend was detected either by visual analysis or resampling testing, and we can only conclude that after 38 years under forest cover, watershed behavior has now stabilized and reached a new equilibrium.



Cross-Simulation Matrix | Recoded Cross-Simulation Matrix

Total Flow (Millimeters); p-Value, 15%

1968 1972 1976 1980 1984 89|90 1994 1999

2859	2939	2867	2895	2989	3589	2895
3517	3591	3529	3566	3673	4249	3546
2882	2955	2894	2923	3025	3586	2910
1593	1664	1600	1618	1694	2195	1624
1797	1871	1810	1824	1912	2403	1822
1442	1504	1450	1466	1538	1977	1466
2381	2458	2394	2424	2513	3075	2409

1968 1972 1976 1980 1984 89|90 1994 1999

o	+	+	+	+	+	+
+	o	-	-	+	+	-
+	-	o	+	+	+	+
+	-	+	o	+	+	+
+	+	+	+	o	+	-
+	+	+	+	+	o	-
+	-	+	-	-	-	o

Flood Flow (Millimeters); p-Value, 5.8%

1968 1972 1976 1980 1984 89|90 1994 1999

963	981	941	981	984	1374	1005
1907	1920	1872	1945	1946	2518	1967
1275	1288	1244	1308	1305	1826	1330
276	286	257	289	280	609	308
452	463	433	465	461	777	485
275	288	265	279	279	472	296
927	951	903	954	950	1404	975

1968 1972 1976 1980 1984 89|90 1994 1999

o	+	-	+	+	+	+
+	o	-	+	+	+	+
-	-	o	+	+	+	+
+	+	+	o	-	+	+
+	-	+	-	o	+	+
+	+	+	+	+	o	-
+	+	+	+	+	-	o

Low-Flow Deficit (Millimeters); p-Value, 24%

1968 1972 1976 1980 1984 89|90 1994 1999

384	371	378	380	364	325	385
205	189	196	201	180	149	208
241	226	232	237	218	184	244
273	257	264	268	247	204	277
365	344	355	359	333	270	367
368	355	360	362	342	313	371
327	310	315	317	288	246	333

1968 1972 1976 1980 1984 89|90 1994 1999

o	-	-	-	-	-	+
-	o	+	+	-	-	+
-	+	o	+	-	-	+
-	+	+	o	-	-	+
-	-	-	-	o	-	+
-	-	-	-	-	o	+
+	+	+	+	+	+	o

Base Flow (Percent); p-Value, 18%

1968 1972 1976 1980 1984 89|90 1994 1999

32	34	33	32	34	30	31
23	24	24	23	24	21	22
28	29	29	28	30	25	27
37	39	38	36	39	34	35
38	40	40	38	41	36	37
37	39	39	37	40	34	36
28	29	30	29	31	26	27

1968 1972 1976 1980 1984 89|90 1994 1999

o	+	+	+	+	-	-
+	o	+	-	+	-	-
+	+	o	-	+	-	-
-	-	-	o	+	-	-
+	+	+	+	o	-	-
-	-	-	-	-	o	+
-	-	-	-	-	+	o

Figure 9. Results obtained for Réal Collobrier's Rimbaud subwatershed over the period 1967–2000.

5.2. Difficulties in the Detection of Sudden and Short-Lived Changes

[44] The Réal Collobrier research watershed presents a particular challenge to the proposed statistical test, as the hydrological changes induced by the wildfire were sudden but short-lived on this resilient watershed. However, although the test was designed to detect gradual changes, we wished to check it against a different kind of change.

[45] Figure 9 presents the results obtained for the burnt-over watershed: The test did not detect the effects of the 1990

Total Flow (Millimeters); p-Value, 32%

1958 1964 1970 1976 1982 1988 1994 1999

1958 1964 1970 1976 1982 1988 1994 1999

Cross-Simulation Matrix							Recoded Cross-Simulation Matrix						
8276	7428	8227	8077	7537	7644	8120	o	-	-	-	-	-	-
7237	6469	7183	7039	6563	6655	7081	-	o	+	+	+	+	+
9072	8274	9029	8886	8284	8492	8864	-	+	o	-	-	-	-
7421	6603	7369	7221	6757	6801	7297	-	+	-	o	-	-	+
7632	6841	7580	7435	6927	7039	7476	-	+	-	-	o	+	+
6519	5760	6460	6311	5890	5931	6383	-	+	-	-	+	o	+
8617	7891	8590	8469	7899	8105	8448	-	+	-	-	+	+	o

Flood Flow (Millimeters); p-Value, 27%

1958 1964 1970 1976 1982 1988 1994 1999

1958 1964 1970 1976 1982 1988 1994 1999

Cross-Simulation Matrix							Recoded Cross-Simulation Matrix						
890	728	953	987	849	844	1019	o	-	+	+	-	-	+
1092	877	1151	1181	1002	1004	1212	-	o	+	+	+	+	+
1550	1279	1638	1694	1402	1471	1681	+	+	o	+	-	-	+
1051	860	1118	1157	975	998	1159	+	+	+	o	-	-	+
770	587	841	875	727	715	911	-	+	-	-	o	-	+
310	190	357	382	305	273	420	-	+	-	-	-	o	+
1464	1218	1562	1612	1370	1404	1628	+	+	+	+	+	+	o

Low-Flow Deficit (Millimeters); p-Value, 4.1%

1958 1964 1970 1976 1982 1988 1994 1999

1958 1964 1970 1976 1982 1988 1994 1999

Cross-Simulation Matrix							Recoded Cross-Simulation Matrix						
853	891	881	906	934	901	910	o	+	+	+	+	+	+
689	739	719	748	791	747	756	+	o	-	+	+	+	+
720	774	752	783	820	781	784	+	-	o	+	+	+	+
704	757	737	768	809	765	771	+	+	+	o	+	-	+
619	663	650	680	716	675	688	+	+	+	+	o	-	-
647	695	679	711	743	708	719	+	+	+	-	-	o	+
482	521	507	530	564	528	535	+	+	+	+	-	+	o

Base Flow (Percent); p-Value, 1.5%

1958 1964 1970 1976 1982 1988 1994 1999

1958 1964 1970 1976 1982 1988 1994 1999

Cross-Simulation Matrix							Recoded Cross-Simulation Matrix						
63	62	61	58	59	58	57	o	-	-	-	-	-	-
49	48	47	45	45	45	45	-	o	-	-	-	-	-
49	47	47	44	45	44	44	-	-	o	-	-	-	-
51	50	49	45	46	49	45	-	-	-	o	+	+	+
51	49	48	45	46	46	45	-	-	-	+	o	-	-
56	53	53	49	51	50	50	-	-	-	+	-	o	-
50	46	45	43	44	43	44	-	-	-	+	+	+	o

Figure 10. Results obtained for Andrews Experimental Forest WS2 over the period 1958–1999.

wildfire, which occurred after more than 20 years of stability and was followed by a rapid return to the previous state. However, this change is easy to identify visually from the cross-simulation matrix (the 1990–1994 column presents the highest total and flood flows in each row). Note that if the episode of sudden change is the last one of the total period, the resampling test can identify it: When the test was run over the period 1968–1994, where 1990–1994 is the last period, we were able to detect a significant trend for total, flood, and low flows (p-values were 1.0, 1.8, and 1.7%, respectively). As all statistical tests based on resampling, this test does have some sensitivity to edge effects.

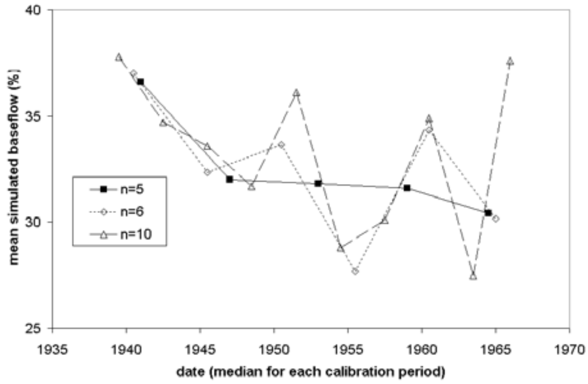


Figure 11. Evolution of the mean simulated base flow on Coshocton WS172 between 1938 and 1967 for three different number *n* of subperiods.

[46] Finally, the test could also be used to assess the progressive return to initial conditions after the wildfire. We ran the test over five successive periods of 1990–2000 and detected a decreasing trend both for total and flood flows (*p*-values were 0.8 and 3.3%, respectively).

5.3. Unexpected Trends for Low and Base Flow in an Apparently Steady Basin

[47] Figure 10 presents the results obtained for WS2 of the Andrews Experimental Forest, a forested watershed that can be considered to have reached a steady hydrological behavior under its present old-growth cover. Over the 40 years of monitoring, we identified no trends for the first two hydrological variables (total and flood flows): All *p*-values were greater than 5%; we had no reason to reject our null hypothesis.

[48] Surprisingly, the deficit in low flows shows a positive trend, and the base flow index a negative one. However, even though these trends are statistically significant, their actual variation range remains very limited (low-flow deficit increases from 74 mm to 84 mm, and the base flow percentage decreases from 54% to 49% on average) and small in comparison with the total streamflow volumes on this watershed. Further investigations are needed to determine whether these trends represent an intrinsic change in watershed behavior or if it is an artifact caused by unaccounted-for climate evolution or a change in measuring techniques over the 40 years of monitoring.

5.4. Sensitivity of the Test to the Choice of Subperiods

[49] A key factor for the detection of changes using the proposed test is obviously the selection of the number of subperiods to be used. Increasing the number of subperiods theoretically allows for a more detailed analysis of a gradual trend. However, longer periods have a higher information content, which provides better calibrations. A trade-off must thus be found between the number *n* of subperiods and their length. To illustrate the sensitivity of the proposed test, we applied it to the period of forest growth on WS172 in Coshocton, starting with a large number of short subperiods and moving toward a low number of long subperiods. Table 1 presents the results for the four hydrological target variables studied in this article. For total flow, flood flow, and low-flow deficit, the test gave consistent results across the range of subperiod numbers explored: *p*-values of the three variables identified as unsteady remain consistently under 5% when the number of subperiods is decreased from $n = 14$ to $n = 6$. There is, however, an understandable tendency toward a decreasing power of the test (for total flow and flood flow at least) when the number of subperiods falls to five: The reason is obvious and lies in the low number of possible resampling combinations. For base flow, the test also shows consistent conclusions when *n* drops from 14 to 6, but surprisingly, it seems to indicate a trend when *n* falls to 5. To understand what happened, in Figure 11 we plotted the progression of the mean simulated base flow (i.e., the average for all the rows of the matrix) for $n = 5$, 6, and 10. Figure 11 shows that the rather surprising *p*-value obtained with $n = 5$ is due to the same edge effect already documented in section 5.2: There is a net decrease in base flow between the first and the second period, which then remains stable from the second to the fifth period. For $n = 5$, the test is influenced by this sudden change, which had previously remained undetected with a larger number of subperiods.

[50] The detection capacity of the proposed test thus appeared relatively independent of the number of subperiods, as long as this number does not fall below $n = 6$. We would therefore advise against using the test blindly with *n* less than 6.

6. Conclusions

[51] The objective of this research was to set up a statistical test able to detect gradual changes in the hydrological behavior of watersheds, and we chose to investigate a method based on resampling of the simulations by a rainfall-runoff model. The proposed test was found to be capable of detecting gradual changes but, not surprisingly, poorly adapted to the detection of sudden changes.

[52] An unexpected and interesting result came from the analysis of AEF watershed 2, where the old-growth forest cover should theoretically guarantee the stability of all the facets of hydrological behavior. Although the watershed

Table 1. Sensitivity of Test Results (Expressed as *p*-Values) to the Number of Subperiods[a]

Length of the Subperiods, years	Total Number of Subperiods (*n*)	Number of Possible Combinations for Resampling (*n*!)	*p*-Value of the Statistical Test, %			
			Applied to Total Flow	Applied to Flood Flow	Applied to Low-Flow Deficit	Applied to Base Flow
2	14	$14! = 8.7 \times 10^{10}$	0.03	0.70	0.60	11
3	10	$10! = 3.6 \times 10^{6}$	1.63	4.36	4.87	20
4	7	$7! = 5040$	0.44	4.82	1.03	10
5	6	$6! = 720$	1.39	2.78	2.92	12
6	5	$5! = 120$	5.00	10	1.67	2.50

[a] Results obtained for Coshocton WS 172, over the period 1938–1967.

behavior seems steady in terms of total and flood flow, we identified a significant trend in terms of base flow and low flow. Detailed analysis of simulation results showed a trend, albeit of limited extent. Further investigations are needed to understand the phenomenon.

[53] We feel that the proposed test can be used with confidence to detect gradual changes. It can be implemented in an automated fashion, but the expertise of a hydrologist will still be required to define the periods where the test can best be applied.

[54] **Acknowledgments.** The authors wish to thank Jacques Bernier for his kind hospitality and his expert assistance with the design of resampling tests, Charles Perrin for his support, and Gunilla de Marsily for her careful editing of the English manuscript. Data for Coshocton watershed 172 were kindly provided by the USDA-ARS, who have maintained the North Appalachian Experimental Watershed since its creation in 1935. Data for the Réal Collobrier research watershed were kindly provided by Jacques Lavabre of Cemagref Aix-en-Provence, which has maintained the watershed since its creation in 1967. Data for the Andrews experimental forest were obtained from the Forest Science Data Bank, a partnership between the Department of Forest Science, Oregon State University, and the USFS Pacific Northwest Research Station. Funding for these data was provided by the NSF LTER program (grant number DEB-9632921). Last, we wish here to thank Mark E. Harmon, lead PI at the Andrews LTER, as well as two anonymous reviewers for their encouraging suggestions.

References

Bates, C. G., and A. J. Henry, Forest and streamflow experiment at Wagon Wheel Gap, Colorado, *Mon. Weather Rev. Suppl.*, *30*, 1–79, 1928.

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

Box, G. E. P., W. G. Hunter, and J. S. Hunter, *Statistics for Experimenters: An Introduction to Design, Data Analysis and Model Fitting*, John Wiley, Hoboken, N. J., 1975.

Brandt, M., S. Bergström, and M. Gardelin, Modelling the effects of clearcutting on runoff—Examples from central Sweden, *Ambio*, *17*, 307–313, 1988.

Conover, J. W., *Practical Non Parametric Statistics*, John Wiley, Hoboken, N. J., 1980.

Cornish, P. M., The effects of logging and forest regeneration on water yields in a moist eucalypt forest in New South Wales, Australia, *J. Hydrol.*, *150*, 301–322, 1993.

Cosandey, C., and M. Robinson, *Hydrologie Continentale*, Armand Colin, Paris, 2000.

Hewlett, J. D., *Principles of Forest Hydrology*, Univ. of Ga. Press, Athens, 1982.

Hibbert, A. R., Forest treatment effects on water yield, in *Forest Hydrology: Proceedings of a National Science Foundation Advanced Science Seminar*, edited by W. E. Sopper and H. W. Lull, pp. 527–543, Pergamon, Tarrytown, N. Y., 1967.

Hornbeck, J. W., M. B. Adams, E. S. Corbett, E. S. Verry, and J. A. Lynch, Long-term impacts of forest treatments on water yield: A summary for northeastern USA, *J. Hydrol.*, *150*, 323–344, 1993.

Kokkonen, T. S., and A. J. Jakeman, Structural effects of landscape and land use on streamflow response, in *Environmental Foresight and Models: A Manifesto*, edited by M. B. Beck, pp. 303–321, Elsevier Sci., New York, 2002.

Kuczera, G., Prediction of water yield reductions following a bushfire in ash-mixed species eucalypt forest, *J. Hydrol.*, *94*, 215–236, 1987.

Kundzewicz, Z. W., and A. Robson, Detecting trend and other changes in hydrological data, *Tech. Doc. 1013*, 159 pp., World Meteorol. Organ., Geneva, 2000.

Langford, K. J., and J. L. McGuinness, A comparison of modeling and statistical evaluation of hydrologic change, *Water Resour. Res.*, *12*, 1322–1324, 1976.

Lavabre, J., D. Sempere Torres, and F. Cernesson, Changes in the hydrological response of a small Mediterranean basin a year after a wildfire, *J. Hydrol.*, *142*, 273–299, 1993.

Lørup, J. K., J. C. Refsgaard, and D. Mazvimavi, Assessing the effect of land use change on catchment runoff by combined use of statistical tests and hydrological modelling: Case studies from Zimbabwe, *J. Hydrol.*, *205*, 147–163, 1998.

L'vovitch, M. I., (Ed.), *World Water Resources and Their Future*, AGU, Washington, D. C., 1979.

McGuinness, J. L., and L. Harrold, Reforestation influences on small watershed streamflow, *Water Resour. Res.*, *7*, 845–852, 1971.

Nascimento, N., Appréciation à l'aide d'un modèle empirique des effets d'actions anthropiques sur la relation pluie-débit à l'échelle d'un bassin versant, Ph.D. thesis, Ecole Nationale des Ponts et Chaussées, Paris, 1995.

Perrin, C., C. Michel, and V. Andréassian, Does a large number of parameters enhance model performance? Comparative assessment of common catchment model structures on 429 catchments, *J. Hydrol.*, *242*, 275–301, 2001.

Perrin, C., C. Michel, and V. Andréassian, Improvement of a parsimonious model for streamflow simulation, *J. Hydrol.*, in press, 2003.

Post, D. A., and J. A. Jones, Hydrologic regimes of forested, mountainous, headwater basins in New Hampshire, North Carolina, Oregon, and Puerto Rico, *Adv. Water Resour.*, *24*, 1195–1210, 2001.

Refsgaard, J. C., W. M. Alley, and V. S. Vuglinsky, Methodology for distinguishing between man's influence and climatic effects on the hydrological cycle, IHP-III Project 6.3, UNESCO, Paris, 1989.

Robson, A., Analysis guidelines, in Detecting trend and other changes in hydrological data, edited by Z. W. Kundzewicz and A. Robson, pp. 11–14, *Tech. Doc. 1013*, World Meteorol. Organ., Geneva, 2000.

Robson, A., A. Bardossy, D. Jones, and Z. W. Kundzewicz, Statistical methods for testing for change, in Detecting trend and other changes in hydrological data, edited by Z. W. Kundzewicz and A. Robson, pp. 49–65, *Tech. Doc. 1013*, World Meteorol. Organ., Geneva, 2000.

Rodda, J. C., Basin studies, in *Facets of Hydrology*, edited by J. C. Rodda, pp. 257–297, John Wiley, Hoboken, N. J., 1976.

Schreider, S. Y., A. J. Jakeman, R. A. Letcher, R. J. Nathan, B. P. Neal, and S. G. Beavis, Detecting changes in streamflow response to changes in non-climatic catchment conditions: Farm dam development in the Murray-Darling basin, Australia, *J. Hydrol.*, *262*, 84–98, 2002.

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

V. Andréassian and C. Michel, Cemagref, Water Quality and Hydrology Research Unit, BP 44, 92163 Antony cedex, France. (vazken.andreassian@cemagref.fr)

E. Parent, ENGREF, Laboratoire de Gestion des Risques en Sciences de l'Eau, Paris, France.