Received: 25 October 2018 | Revised: 17 February 2019 | Accepted: 27 May 2019
DOI: 10.1002/env.2593

SPECIAL ISSUE PAPER

WILEY

Multiple change-point models for time series

I.B. MacNeill[1] | V.K. Jandhyala[2] | A. Kaul[2] | S.B. Fotopoulos[3]

[1]Department of Statistical and Actuarial Sciences, Western University, London, Canada

[2]Department of Mathematics and Statistics, Washington State University, Pullman, Washington

[3]Department of Finance and Management Science, Washington State University, Pullman, Washington

Correspondence
V. K. Jandhyala, Department of Mathematics and Statistics, Washington State University, Pullman, WA 99164-3113.
Email: jandhyala@wsu.edu

Abstract

The "Bayes-type" method of deriving change-point test statistics was introduced by Chernoff and Zacks (1964). Other authors subsequently adapted this approach and derived Bayes-type statistics for at most one change (AMOC), and for multiple change points, under a variety of model formulations. Asymptotic distribution theory has always been limited to the AMOC statistics because of the perceived complexity of multiple change-point statistics. In this article, it is shown that the Bayes-type statistic derived to test for multiple change points is directly proportional to the AMOC statistic. This result immediately provides distributional results for Bayes-type multiple change-point statistics. In addition, it fundamentally alters the current understanding of the AMOC statistic. It follows from this result that the Bayes-type statistic derived under AMOC conditions in fact tests for at least one change (ALOC), even though the statistic is derived under AMOC formulation. Under asymptotic consideration, the result also extends to the case of model errors following a stationary process. As an example, the classical Nile River data are revisited and analyzed for the presence of multiple change points.

KEYWORDS

at least one change, at most one change, Bayes-type tests, multiple change-points, Nile River series

1 | INTRODUCTION

The change-point problem is concerned with the time homogeneity of the parameters of a statistical model. Testing for the presence of change points and estimation of the locations of the change points are the two main inferential issues associated with the change-point problem. Statistical tests based on likelihood, Bayes, and nonparametric methods have been presented in the literature. The "Bayes-type" method for deriving change-point test statistics was first introduced by Chernoff and Zacks (1964) while deriving a one-sided statistic to test for a change point in the mean of a sequence of independent normal random variables. According to the method, one first assumes prior distributions for the unknown parameters associated with both null and alternative hypotheses. The nuisance parameters are then eliminated by deriving the unconditional likelihoods, and the Bayes-type statistic is obtained by considering the likelihood ratio.

The problem as formulated by Chernoff and Zacks (1964) considers a model with no change in the mean under the null hypothesis against a model with a single change point under the alternative hypothesis. This testing problem is referred to in the literature as the at most one change (AMOC) case. On the other hand, one might also be interested in testing for multiple change points under the alternative hypothesis.

The Bayes-type method of testing for change points has been considered by several authors, including Kander and Zacks (1966), Gardner (1969), Sen and Srivastava (1973, 1975), MacNeill (1974, 1978), Jandhyala and MacNeill (1989, 1991, 1992), Jandhyala (1993), Jandhyala and Minogue (1993), MacNeill and Jandhyala (1993), Tang and MacNeill (1993), Jandhyala and MacNeill (1997), Xie and MacNeill (2006), and Jandhyala, Fotopoulos, and You (2010). Recently, Jandhyala, Fotopoulos, MacNeill, and Liu (2013) have carried out a short review of Bayes-type methods for detection of change points in regression models.

Kander and Zacks (1966) extended the one-sided AMOC statistic to the one parameter exponential family. Gardner (1969) derived the two-sided statistic for sequences of normal random variables. His formulation allows for the detection of multiple change points; the derived statistic is a quadratic form in the residuals. While deriving the null asymptotic distribution theory, however, Gardner restricted his large-sample theory to the AMOC case, citing the more general case of multiple change points to be too complicated. MacNeill (1974) extended the large-sample distribution theory of the two-sided AMOC statistic to the one parameter exponential family. The case of testing for change points in the mean vector of a sequence of multivariate normal random variables was considered by Sen and Srivastava (1973, 1975). Jandhyala and MacNeill (1989, 1991, 1992) extended the Bayes-type methodology to the case of general linear regression models. As in the case of Gardner (1969), while the derivations of their statistics allow for arbitrarily many change points, their large-sample theory was again restricted to the AMOC case. Jandhyala and Minogue (1993) developed a general method for numerically solving Fredholm integral equations and applied the methodology to derive the large-sample distribution theory of two-sided AMOC statistics in the case of polynomial regression. Jandhyala and MacNeill (1997) derived a Bayes-type statistic to test for change points with continuity constraints. The statistic involves iterated partial sums of regression residuals whose finite and large-sample properties have been studied in this paper. Tang and MacNeill (1993) derived the large-sample distribution theory of change-detection statistics in the case of serially correlated noise. Considering the Tucumán annual mean rainfall data, Jandhyala et al. (2010) implemented the Bayes-type change detection methodology for serially correlated data, as developed by Tang and MacNeill (1993).

The spatial analogue of the change-point problem is one of detection and estimation of boundaries in spatial data. This problem has important applications in many areas including environmental monitoring, medical imaging, remote sensing, and forest management. Bayes-type statistics for the detection of boundaries in spatial data were first considered by MacNeill and Jandhyala (1993). These statistics have been extended by Xie and MacNeill (2006) where they studied the large-sample distribution theory of these boundary detection statistics.

In the regression setup, when a change point occurs, it may be a consequence of a change occurring in a single parameter, in a subset of the parameters, or in all the parameters at that single change point. Thus, this situation is still limited to the AMOC case. It is thus clear that asymptotic theory of Bayes-type change detection statistics has to this point been limited to the AMOC case.

In this article, we consider the case of testing for multiple change points under the “Bayes-type” approach. In the main result of this paper, we first consider the case where the errors are independent and then show that the statistic that tests for multiple change points (say, k change points) is proportional to the statistic derived under the AMOC case. The constant of proportionality is precisely k, the number of change points under the multiple-change alternative. This result has substantial implications for the change-point problem. It shows that both finite and large-sample distributions of multiple change statistics are completely determined by the respective distributions of the statistics derived under the AMOC case. More importantly, this requires a rethinking of the current understanding of the AMOC statistic. Because the k-change statistic is exactly equal to k times the AMOC statistic, it follows that a value of the AMOC statistic large enough to reject the null hypothesis of no change signifies a single change point or multiple change points. Thus, the AMOC statistic in fact shows evidence of at least one change (ALOC), and not just a single change as has been understood thus far. We discuss the asymptotic form of this main result under independence in the next section and follow it up in the subsequent section by considering stationary error structure.

From a practical point of view, the result establishes strong support in favor of the Bayes-type change-point statistics. While testing for the possible presence of change points in a time series, one would much prefer to test the hypothesis of no change against ALOC rather than against AMOC; that is, is a change-point model required? This is analogous to ANOVA-based tests, which test for the presence of one or more treatment effects. There is no known parallel result for testing no change against ALOC based on other procedures including generalized likelihood ratio criteria. Until now, only heuristic statistics with unspecified properties address the problem of testing the hypothesis of no change against an unspecified alternative. Because Bayes-type tests are known to satisfy local optimality power properties for each specified alternative, good power properties are assured under the ALOC case. The problem then of testing for exactly one change or exactly k changes under this setup now becomes an open question.

Other approaches to testing for single and multiple change points in time series have been considered in both statistical and econometric literature. Asymptotic theory for likelihood ratio statistics may be found in the monograph by Csörgő and Horváth (1997). Likelihood ratio (LR)- and Wald-type statistics and their variants for detection of multiple change points in time series have been considered by many authors in the econometrics literature; we only include here Andrews (1993), Vogelsang (1997), Bai and Perron (1998), and Qu and Perron (2007). For example, Andrews (1993) considers Wald-, Lagrange multiplier (LM)-, and LR-type test statistics and derives their asymptotic distributions. In particular, Andrews (1993) has shown that all three statistics have the same limiting distribution, namely, the supremum of the square of a Bessel process multiplied by a constant that becomes zero when the change occurs at either of the endpoints of the time series. Thus, these statistics go to infinity unless the change point is bounded away from both the endpoints. In contrast, instead of being a supremum, the Bayes-type statistic is a weighted average of partial sums of squared residuals, and there is no multiplier associated with the asymptotic distribution of the Bayes-type statistic that vanishes. As shown in Sections 3 and 4, asymptotically, the Bayes-type statistic is a stochastic integral (not a supremum) defined on the square of a generalized Brownian bridge process. Thus, unlike Wald-, LM-, and LR-type statistics, asymptotic distribution theory of Bayes-type statistics do not require the assumption that the change point be bounded away from both ends of the time series. In trying to avoid the assumption of the change point being bounded away from the end points, Csörgő and Horváth (1997) apply strong approximation theory to show that LR-type statistics converge in distribution to the double exponential distribution. However, being more conservative, double exponential distribution possesses lower power properties. As shown in this article, the ALOC interpretation of the AMOC statistic displayed by the Bayes-type statistic does not seem to apply to these other statistics such as LR-, LM-, or Wald-type statistic.

In Section 3, we show the asymptotic analogue of the main result derived under independent errors. Then, we move on to consider in Section 4 the case where the error structure is a general stationary process and show that the asymptotic analogue of the main result under stationary errors can be obtained from the result derived under independent errors through a simple multiplying constant. This extension of the main result under independent errors to the case of stationary errors is facilitated by recalling the results in Tang and MacNeill (1993). In the following section, we review the simulation-based power comparison study carried out by Jandhyala and MacNeill (1991).

As an application, we consider the classical annual Nile River flows data at the Aswan Dam. These data have been analyzed for the presence of change points by many authors including Cobb (1978); Todini and O'Connell (1979); Hosking (1984); MacNeill, Tang, and Jandhyala (1991); Wu and Zhao (2007); Shao (2011); and Kim and Hart (2011). There is considerable discussion regarding the suitability of the assumption of independence for these data. In this article, we particularly note that the dependence structure is a result of a couple of extreme observations. Once these extreme observations are deleted, the data are well behaved and adheres to the assumption of independence. On this basis, we first test for the presence of ALOC point applying the Bayes-type methodology and subsequently identify the number and location of change points by minimizing the squared residuals and by applying the more recent pruned exact linear time (PELT) method developed by Killick, Fearnhead, and Eckley (2012). Both methods identify the years 1899, 1953, and 1965 as change points.

This paper is organized as follows. In the next section, we state the model, the main result under independent errors, and its implications. The asymptotic analogue of our main result under independent errors is considered in Section 3. Then, in Section 4, we consider the asymptotic analogue of our main result when the error structure is a general stationary process. Section 5 reviews a simulation study carried out by Jandhyala and MacNeill (1991) to compare powers of Bayes-type statistic against likelihood ratio statistics. The application to Nile River annual flows data at the Aswan Dam is carried out in Section 6.

2 | MODELS AND RESULTS

Let $Y_1, Y_2, \ldots, Y_n$ be a time series of observations satisfying the linear regression model

$$\mathbf{Y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{1}$$

where $\mathbf{Y}' = (Y_1, Y_2, \ldots, Y_n)$ is the vector of observations, $\boldsymbol{\beta}' = (\beta_0, \beta_1, \ldots, \beta_{p-1})$ is the vector of regression parameters, $X_{n\times p} = ((x_{ij}))$ is the design matrix, and $\boldsymbol{\varepsilon}' = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)$ is the vector of unobservable error random variables that are independent. Here, we assume that $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 I)$.

The change-point problem is concerned with determining whether the regression parameters are time homogeneous over the duration of the sampling period. Under the change-point formulation, the vector of regression parameters may change, possibly more than once during the sampling period. Accordingly, let $W = ((\omega_{ij}))$, $i = 1, 2, \ldots, n-1$; $j = 0, 1, \ldots, p-1$ be the change matrix. The (i,j)th element ω_{ij} is 1 or 0 according to whether or not a change occurs in β_j between the ith and $(i+1)$th time points. Furthermore, let $\Delta = ((\delta_{ij}))$, $i = 1, 2, \ldots, n-1$; $j = 0, 1, \ldots, p-1$ be the matrix representing the amounts of change in the parameters. The general change-point regression model that accommodates multiple change points may then be written as

$$Y_t = \mathbf{X}_t \boldsymbol{\beta}(t) + \varepsilon_t, \qquad t = 1, 2, \ldots, n, \tag{2}$$

where $\mathbf{X}_t$ is the tth row of the design matrix X, and $\boldsymbol{\beta}(t)$, which denotes the regression parameter vector associated with the tth observation Y_t, is given by

$$\boldsymbol{\beta}(t) = \begin{cases} \boldsymbol{\beta}, & t = 1 \\ \boldsymbol{\beta} + \left\{ \sum_{k=1}^{t-1} \omega_{k0}\delta_{k0}, \ldots, \sum_{k=1}^{t-1} \omega_{k0}\delta_{k,p-1} \right.', & t = 2, \ldots, n. \end{cases}$$

In the above, $\boldsymbol{\beta}(1)$ denotes the initial parameter vector $\boldsymbol{\beta}$.

An important inferential problem of the change-point formulation is that of testing for no change against the alternative of one or more change points in the parameters. Thus, the hypotheses associated with this testing problem are

$$H_0 : \Delta = 0 \qquad \text{versus} \qquad H_a : \Delta \neq 0. \tag{3}$$

The alternative hypothesis in (3) above is quite general, and specific alternatives such as for the AMOC case may be obtained from H_a in (3) by restricting the change matrix W accordingly. This will be elaborated upon below. In order to derive the corresponding Bayes-type statistic, Jandhyala and MacNeill (1991) assume the following prior distributions on the unknown parameters:

$$\boldsymbol{\beta} \sim N\left(\mathbf{0}, \tau^2 I\right) \quad \text{and} \quad \boldsymbol{\delta}_t \sim N\left(\mathbf{0}, \theta^2 I\right), \quad t = 1, 2, \ldots, n-1,$$

where $\boldsymbol{\delta}_t' = (\delta_{t0}, \ldots, \delta_{tp-1})'$, $t = 1, 2, \ldots, n-1$. Furthermore, it is assumed that $\boldsymbol{\beta}$, $\boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_{n-1}$, and $\boldsymbol{\varepsilon}$ are all distributed independently. The statistic derived by Jandhyala and MacNeill (1991) to test for H_0 against H_a in (3) is given by the following theorem. The theorem is proved based on the assumption that the errors are independent and that the error variance σ^2 is known. Without loss of generality, we let $\sigma^2 = 1$.

Theorem 1. *Suppose the errors ε_t, $t = 1,2, \ldots, n$ are independent. Let the corresponding two-sided Bayes-type likelihood ratio statistic for testing H_0 against H_a in (3) be denoted by V_n^I. Then, V_n^I is given by*

$$V_n^I = \sum_{\{W\}} p(W) \left\{ \mathbf{Y}'R \left(\sum_{i=0}^{p-1} C_i C_i' \right) R\mathbf{Y} \right\}, \tag{4}$$

where $p(W)$ represents a discrete prior probability function on the collection of change matrices $\{W\}$, $R = I - X(X'X)^{-1}X'$, and the matrices C_i, $i = 0,1, \ldots, p-1$ are given by

$$C_i = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & x_{2i}\omega_{1i} & 0 & \cdots & 0 \\ 0 & x_{3i}\omega_{1i} & x_{3i}\omega_{2i} & & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & x_{ni}\omega_{1i} & x_{ni}\omega_{2i} & \cdots & x_{ni}\omega_{(n-1)i} \end{bmatrix}. \tag{5}$$

The change detection statistics for several important special cases considered in the sequel may be derived from the statistic V_n^I in Theorem 1.

(i) AMOC case. First, we consider the AMOC case where a single change point occurring between m and $(m + 1)$, $m = 1, 2, \ldots, n-1$, is allowed with m being unknown. The change matrix W then consists of rows all zero except for the mth row, which is given by $(\omega_{m0}, \ldots, \omega_{m,p-1})$. Let $p(m)$ denote a prior on the change point m. Then, as a special case of Theorem 1, it follows that the corresponding two-sided Bayes-type statistic is given by

$$V_{1,n}^I = \sum_{m=1}^{n-1} p(m)\ \mathbf{Y}'RH_{(m)}H_{(m)}'R\mathbf{Y}, \tag{6}$$

where $H_{(m)}$ is a matrix with the first m rows being all identically zero and the subsequent rows given by $H_{(m)j} = (x_{j0}\omega_{m0}, \dots, x_{j,p-1}\omega_{m,p-1}), j = m+1, \dots, n$. In particular, if one is interested in testing for AMOC in the parameter β_i alone, then the statistic (6) becomes

$$U_{1,n}^{(i),I} = \sum_{m=1}^{n-1} p(m)\ \mathbf{Y}' R \mathbf{X}_{mi} \mathbf{X}_{mi}' R \mathbf{Y}, \tag{7}$$

where $\mathbf{X}_{mi}$ is the ith column vector of X with first m rows replaced by 0 s.

(ii) Multiple change-points case.

More generally, the statistic that tests for k change points, say, $m_1 < m_2 < \cdots < m_k$ in the parameter $\beta_i, i = 0,1, \dots, p-1$, is given by

$$U_{k,n}^{(i),I} = \sum_{m_1} \cdots \sum_{m_k} p(m_1, \dots, m_k) \left\{ \mathbf{Y}' R C_{m_1,\dots,m_k i} C'_{m_1,\dots,m_k i} R \mathbf{Y} \right\}, i = 0, 1, \dots, p-1, \tag{8}$$

where $p(m_1, \dots, m_k)$ is a joint prior on the unknown change points $m_1 < m_2 < \cdots < m_k$ and

$$C_{m_1,\dots,m_k i} = \begin{bmatrix}
0 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & \cdots & 0 \\
\vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots & & \vdots \\
\vdots & & \vdots & 0 & \vdots & & \vdots & \vdots & & \vdots & & \vdots \\
\vdots & & \vdots & x_{m_1+1,i} & \vdots & & \vdots & \vdots & & \vdots & & \vdots \\
\vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots & & \vdots \\
\vdots & & \vdots & \vdots & \vdots & & 0 & \vdots & & \vdots & & \vdots \\
\vdots & & \vdots & \vdots & \vdots & & x_{m_2+1,i} & \vdots & & \vdots & & \vdots \\
\vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots & & \vdots \\
\vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & & 0 & & \vdots \\
\vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & & x_{m_k+1,i} & & \vdots \\
\vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots & & \vdots \\
0 & \cdots & 0 & x_{ni} & 0 & \cdots & x_{ni} & 0 & & x_{ni} & \cdots & 0
\end{bmatrix}. \tag{9}$$

All of the above two-sided change detection statistics are quadratic forms in regression residuals, and their distribution theory has been regarded as complicated.

As discussed in Section 1, although some statistics have been derived using a multiple change-point formulation, there has been no previous attempt in the literature to deal with the distribution theory of the test statistic for multiple change points. Here, we show that the Bayes-type statistic that tests for multiple change points and the AMOC statistic are directly related. First, we state and prove the main result in the following theorem. Its implications are discussed subsequently.

Theorem 2. *Let $p(m_1, \dots, m_k)$ be the joint prior for the unknown change points $m_1, m_2, \dots, m_k$ such that $1 \le m_1 < m_2 < \cdots < m_k \le n-1$. In addition, let $p_j(\cdot)$ be the marginal probability of m_j such that $p_j(m_j) = \sum_{m_1} \cdots \sum_{m_j-1} \sum_{m_j+1} \cdots \sum_{m_k} p(m_1, \dots, m_k), j = 1, \dots, k$. Let $U_{k,n}^{(i),I}$ given by (8) be the two-sided Bayes-type statistic that tests for k-multiple change points in the regression parameter β_i, $i = 0, \dots, p-1$ when errors are independent. Furthermore, let $U_{1,n}^{(i),I}$ given by (7) be the AMOC statistic that tests for AMOC in the parameter β_i with prior probability function given by*

$$p(m) = \frac{1}{k} \left\{ p_1(m) + \dots + p_k(m) \right\}. \tag{10}$$

Then, the statistics $U_{k,n}^{(i),I}$ and $U_{1,n}^{(i),I}$ obey the relation given by

$$U_{k,n}^{(i),I} = k U_{1,n}^{(i),I}. \tag{11}$$

Proof. For $m = 1, \ldots, n-1$, define the matrix B_m by

$$B_m = \begin{array}{c} 1 \\ \vdots \\ m \\ m+1 \\ \vdots \\ \vdots \\ n \end{array} \begin{bmatrix} 0 & \cdots & 0 & \cdots & \cdots & \cdots & 0 \\ \vdots & & \vdots & & & & \\ 0 & \cdots & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \cdots & 0 & 1 & 1 & \cdots & 1 \\ 0 & \cdots & 0 & 1 & 1 & \cdots & 1 \\ \vdots & & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 1 & \cdots & 1 \end{bmatrix}.$$

Then, the statistic $U_{1,n}^{(i),I}$ in (7) may be written as

$$U_{1,n}^{(i),I} = \mathbf{Y}' R \mathrm{Diag}\left(\mathbf{X}_i\right) \sum\nolimits_{m=1}^{n-1} p\left(m\right) B_m \mathrm{Diag}\left(\mathbf{X}_i\right) R\mathbf{Y}, \tag{12}$$

where $\mathrm{Diag}(\mathbf{X}_i) = \mathrm{Diag}(x_{1i}, x_{2i}, \ldots, x_{ni})$ is a representation of the vector $\mathbf{X}_i$ in a diagonal matrix form with all off-diagonal elements being 0 s. Now, let

$$B_{m_1,\ldots m_k} = B_{m_1} + \cdots + B_{m_k}. \tag{13}$$

The statistic $U_{k,n}^{(i),I}$ in (8) may then be written as

$$\begin{aligned} U_{k,n}^{(i),I} &= \mathbf{Y}' R \mathrm{Diag}\left(\mathbf{X}_i\right) \left\{ \sum\nolimits_{m_1} \cdots \sum\nolimits_{m_k} p\left(m_1, \ldots, m_k\right)\ B_{m_1,\ldots m_k} \right\} \mathrm{Diag}\left(\mathbf{X}_i\right) R\mathbf{Y} \\ &= \mathbf{Y}' R \mathrm{Diag}\left(\mathbf{X}_i\right) \left\{ \sum\nolimits_{m_1=1}^{n-1} p_1\left(m_1\right) B_{m_1} + \cdots \sum\nolimits_{m_k=1}^{n-1} p_k\left(m_k\right) B_{m_k} \right\} \mathrm{Diag}\left(\mathbf{X}_i\right) R\mathbf{Y} \\ &= k\mathbf{Y}' R \mathrm{Diag}\left(\mathbf{X}_i\right) \left\{ \sum\nolimits_{m=1}^{n-1} p\left(m\right) B_m \right\} \mathrm{Diag}\left(\mathbf{X}_i\right) R\mathbf{Y}, \end{aligned}$$

where $p(m)$ satisfies (10). This completes the proof.

Theorem 2 reveals a new dimension to the current understanding of Bayes-type change detection statistics. This may be best understood in the uniform scenario. As a special case, let the joint prior $p(m_{1,\ldots,}m_k)$, $1 \leq m_1 < m_2 < \cdots < m_k \leq n-1$, be discrete uniform such that

$$p\left(m_1, \ldots, m_k\right) = \frac{1}{\binom{n-1}{k}}, \quad 1 \leq m_1 < m_2 < \cdots < m_k \leq n-1. \tag{14}$$

Then, from (10), the corresponding prior $p(m)$ in the AMOC case is also uniform and is given by

$$p\left(m\right) = \frac{1}{n-1}, \quad 1 \leq m \leq n-1. \tag{15}$$

Thus, when the prior for the AMOC case is uniform, we have

$$U_{k,n}^{(i),I} = kU_{1,n}^{(i),I}, \quad k = 1, \ldots, n-1, \tag{16}$$

with the prior for the k-multiple changes also being uniform. Now, suppose x_α is an α-significance value for $U_{1,n}^{(i),I}$ such that

$$P\left(U_{1,n}^{(i),I} > x_\alpha\right) = \alpha. \tag{17}$$

Then, from (15), it follows that

$$P\left(U_{1,n}^{(i),I} > x_\alpha\right) = P\left(U_{k,n}^{(i),I} > kx_\alpha\right) = \alpha, \ \ k = 1, \ldots, n-1. \tag{18}$$

The implication of (18) is extremely important. For a given data set, if $U_{1,n}^{(i),I} > x_\alpha$ holds, thus showing evidence of a single change at level α, then (18) implies that $U_{k,n}^{(i),I} > kx_\alpha$ holds automatically at the same level α, for any $k = 1, \dots, n-1$. This shows that evidence for a single change point at a given level in a datum implies there is evidence for k-multiple change points at the same level for any arbitrary k. Thus, the AMOC statistic not only tests for the occurrence of a single change point but also shows evidence for arbitrarily many multiple change points. It thus follows that the AMOC statistic should be viewed as an ALOC statistic implying the AMOC statistic in fact tests for ALOC point. The result also holds more generally when one tests for change points in a subset or all of the parameters. For example, if one were to test for two-sided changes in the first two parameters β_0 and β_1, then the statistic for the AMOC case is

$$U_{1,n}^{(0,1),I} = \sum_{m=1}^{n-1} p(m)\ \mathbf{Y}'R\mathbf{X}_{m0}\mathbf{X}_{m0}'R\mathbf{Y} + \sum_{m=1}^{n-1} p(m)\ \mathbf{Y}'R\mathbf{X}_{m1}\mathbf{X}_{m1}'R\mathbf{Y}. \tag{19}$$

The corresponding statistic that tests for k-multiple changes is

$$\begin{aligned} U_{k,n}^{(0,1),I} = &\sum_{m_1} \cdots \sum_{m_k} p(m_1, \dots, m_k)\ \mathbf{Y}'R\left(\sum_{i=1}^{k} \mathbf{X}_{m_i 0}\mathbf{X}_{m_i 0}'\right)R\mathbf{Y} \\ &+ \sum_{m_1} \cdots \sum_{m_k} p(m_1, \dots, m_k)\ \mathbf{Y}'R\left(\sum_{i=1}^{k} \mathbf{X}_{m_i 1}\mathbf{X}_{m_i 1}'\right)R\mathbf{Y}. \end{aligned} \tag{20}$$

Again, it can be shown that

$$U_{k,n}^{(0,1),I} = kU_{1,n}^{(0,1),I}, \quad k = 1, \dots, n-1, \tag{21}$$

The main result and the interpretation also hold for one-sided changes in the parameters. As derived by Jandhyala and MacNeill (1991), the AMOC statistic that tests for one-sided change in the parameter β_i under independent errors is

$$T_{1,n}^{(i),I} = \sum_{m=1}^{n-1} p(m)\ \mathbf{Y}'R\mathbf{X}_{mi}. \tag{22}$$

The corresponding statistic that tests for k-multiple change points is

$$T_{k,n}^{(i),I} = \sum_{m_1} \cdots \sum_{m_k} p(m_1, \dots, m_k)\ \mathbf{Y}'R\sum_{j=1}^{k} \mathbf{X}_{m_j i}. \tag{23}$$

Here again, it can be shown that

$$T_{k,n}^{(i),I} = kT_{1,n}^{(i),I},\ k = 1, \dots, n-1. \tag{24}$$

□

3 | ASYMPTOTIC DISTRIBUTIONS OF BAYES-TYPE STATISTICS

This section presents a brief outline of the asymptotic distributional result corresponding to the main result in Theorem 2 derived under the assumption that the errors ε_t, $t = 1,2, \dots ,n$ are independent. Earlier, Jandhyala and MacNeill (1989, 1991) derived the asymptotic distribution theory of the two-sided statistic for the AMOC case. While their results hold more generally for testing changes in a subset or all the parameters, here, we restrict the attention to the case of testing for changes in a single parameter β_i only. Asymptotic theory for change-point statistics requires that the formulation of the regression model be based on regressor functions.

Thus, for $t \in [0,1]$, let $f_i(t)$ be the ith regressor function and let the regression model be given by

$$Y_t = \sum_{j=0}^{p-1} \beta_i f_i\left(\frac{t}{n}\right) + \varepsilon_t, \quad t = 1, 2, \dots, n, \tag{25}$$

where ε_t, $t = 1,2, \dots , n$ are independent. When the regressor functions are continuously differentiable, Jandhyala and MacNeill (1989, 1991) showed that the asymptotic distribution of the AMOC statistic $U_{1,n}^{(i),I}$ in (7) is given by

$$n^{-1}U_{1,n}^{(i),I} \xrightarrow{D} \int_0^1 \psi(t) \left\{B_p^{(f_i)}(t)\right\}^2 dt, \quad i = 0, 1, \dots, p-1, \tag{26}$$

where $\psi(t)$, $t \in [0,1]$ is a weight function reflecting the prior on the unknown change point and $\{B_p^{(f_i)}(t)\}$, $t \in [0,1]$ is a generalized Brownian bridge process. The exact form of $\{B_p^{(f_i)}(t)\}$ may be found in Jandhyala and MacNeill (1989). Their proof may be applied to show that the asymptotic distribution of the k-multiple change detection statistic $U_{k,n}^{(i),I}$ in (8) is given by

$$n^{-1}U_{k,n}^{(i),I}\xrightarrow{D}\int_{t_1=0}^{1}\int_{t_2=t_1}^{1}\cdots\int_{t_k=t_k-1}^{1}\psi\left(t_1,t_2,\ldots,t_k\right)\left\{B_p^{(f_i)}\left(t_1,t_2,\ldots,t_k\right)\right\}^2dt_k\cdots dt_1, \tag{27}$$

where $\psi(t_1, t_2, \ldots, t_k)$ is a weight function reflecting the joint prior on the k change points and $\{B_p^{(f_i)}(t_1, t_2, \ldots, t_k)\}, 0 \le t_1 < t_2 < \ldots < t_k \le 1$ is an appropriately defined generalized Brownian bridge. Then, the following result holds as a consequence of Theorem 2:

$$\int_{t_1=0}^{1}\int_{t_2=t_1}^{1}\cdots\int_{t_k=t_k-1}^{1}\psi\left(t_1,t_2,\ldots,t_k\right)\left\{B_p^{(f_i)}\left(t_1,t_2,\ldots,t_k\right)\right\}^2dt_k\cdots dt_1=_Dk\int_0^1\psi\left(t\right)\left\{B_p^{(f_i)}\left(t\right)\right\}^2dt, \tag{28}$$

where $\psi(t)$ satisfies the relation

$$\psi\left(t\right)=\frac{1}{k}\left\{\psi_1\left(t\right)+\cdots+\psi_k\left(t\right)\right\}. \tag{29}$$

In the above, $\psi_i(t)$, $i = 1,2, \ldots, k$ is the ith marginal function derived from the prior $\psi(t_1, t_2, \ldots, t_k)$. From (27) and (28), one may conclude that

$$n^{-1}U_{k,n}^{(i),I}\xrightarrow{D}k\int_0^1\psi\left(t\right)\left\{B_p^{(f_i)}\left(t\right)\right\}^2dt. \tag{30}$$

4 | DETECTION OF CHANGE POINTS UNDER DEPENDENCE STRUCTURE

The goal of this section is to consider the problem of detecting multiple change points in regression models via a Bayes-type method when the error process obeys a dependence structure characterized by a stationary process. Accordingly, we begin by considering the regression model given by

$$Y_t=\sum\nolimits_{j=0}^{p-1}\beta_i f_i\left(\frac{t}{n}\right)+\varepsilon_t,\quad t=1,2,\ldots,n, \tag{31}$$

where ε_t, $t = 1,2, \ldots, n$ is a zero-mean, discrete-time stationary time series with covariance function $R(\nu) = E[\varepsilon_t\varepsilon_{t+\nu}]$. The problem of studying the effect of stationary errors on the Bayes-type change detection statistic derived under independence has been previously tackled by Tang and MacNeill (1993). We find that the main result in their paper is directly applicable to the problem we consider in this section. In order to state their result, we let $f_\varepsilon(\lambda)$ denote the spectral density function of the error process ε_t such that $f_\varepsilon(\lambda) = (1/2\pi)\sum_{|\nu|<\infty}e^{-i\lambda\nu}R(\nu)$, $\lambda \in [-\pi, \pi]$. Note that the spectral density function exists when $\sum_{\nu=-\infty}^{\infty}|R(\nu)| < \infty$. Here, we pursue the effect of stationary error structure upon the AMOC change detection static $U_{1,n}^{(i),I}$ in (7), which is derived when the errors are independent. Thus, let the same AMOC test statistic in (7), when applied to the case where the error process ε_t is stationary, be denoted by $U_{1,n}^{(i),S}$. Then, assuming (i) $f_\varepsilon(\lambda) \ge a > 0$, $\lambda \in [-\pi, \pi]$, and (ii) $|C_{k+1}(\nu_1, \ldots, \nu_k)| < L_k \Big/ \left(\prod_{j=1}^{k}\left(1+\nu_j^2\right)\right)$, Tang and MacNeill (1993) have shown that the AMOC statistic $U_{1,n}^{(i),S}$ in (7) will converge in distribution as follows:

$$n^{-1}U_{1,n}^{(i),S}\xrightarrow{D}\frac{2\pi f_\varepsilon\left(0\right)}{\int_{-\pi}^{\pi}f_\varepsilon\left(\lambda\right)d\lambda}\int_0^1\psi\left(t\right)\left\{B_p^{(f_i)}\left(t\right)\right\}^2dt,\qquad i=0,1,\ldots,p-1, \tag{32}$$

where $\psi(t)$ and $B_p^{(f_i)}(t)$ are as described in Section 3. Subsequently, while studying distributions of F-type change detection statistics, Bai and Perron (1998) have also noted that a similar multiplicative constant applied to the statistic under independence accounts for serial correlations in time series.

Now, let $U_{k,n}^{(i),S}$ be the statistic to detect k-multiple changes in β_i, $i = 0,1, \ldots, p-1$ when the error process is stationary. Noting that the asymptotic distribution of $U_{k,n}^{(i),I}$ under independence is given in (27), similar to (32), it can be shown that the asymptotic distribution of $U_{k,n}^{(i),S}$ under stationary errors is related to that of $U_{k,n}^{(i),I}$ as follows:

$$n^{-1}U_{k,n}^{(i),S}\xrightarrow{D}\frac{2\pi f_\varepsilon\left(0\right)}{\int_{-\pi}^{\pi}f_\varepsilon\left(\lambda\right)d\lambda}\int_{t_1=0}^{1}\int_{t_2=t_1}^{1}\cdots\int_{t_k=t_k-1}^{1}\psi\left(t_1,t_2,\ldots,t_k\right)\left\{B_p^{(f_i)}\left(t_1,t_2,\ldots,t_k\right)\right\}^2dt_k\cdots dt_1, \tag{33}$$

where $\psi(t_1, t_2, \dots, t_k)$ and $\left\{B_p^{(f_i)}(t_1, t_2, \dots, t_k)\right\}$ are as in Section 3. It follows from (33), (28), and (30) that

$$n^{-1} U_{k,n}^{(i),S} \xrightarrow{D} k \frac{2\pi f_\varepsilon(0)}{\int_{-\pi}^{\pi} f_\varepsilon(\lambda)\, d\lambda} \int_0^1 \psi(t) \left\{B_p^{(f_i)}(t)\right\}^2 dt. \tag{34}$$

We shall now consider the special case where the error process ε_t follows an AR(1) process—an autoregressive process of order one. Accordingly, let ε_t be such that

$$\varepsilon_t + b_1 \varepsilon_{t-1} = \eta_t, \qquad t = 1, 2, \dots, n, \tag{35}$$

where η_t is a white noise with error variance σ_η^2. Then, it can be shown that

$$\frac{2\pi f_\varepsilon(0)}{\int_{-\pi}^{\pi} f_\varepsilon(\lambda)\, d\lambda} = \frac{2\pi\left(\sigma_\eta^2 / \left(2\pi\left(1 + b_1^2\right)\right)\right)}{\sigma_\eta^2 / \left(1 - b_1^2\right)} = \frac{1 - b_1}{1 + b_1}. \tag{36}$$

Let $U_{k,n}^{(i),\mathrm{AR}(1)}$ be the statistic to detect k-multiple changes under the AR(1) model (35). In addition, asymptotically, let $n^{-1} U_{k,n}^{(i),\mathrm{AR}(1)} \xrightarrow{d} U_k^{(i),\mathrm{AR}(1)}$. Similarly, under independence, let $n^{-1} U_{k,n}^{(i),I} \xrightarrow{d} U_k^{(i),I}$. When $k = 1$, it is clear from (26) that $U_1^{(i),I} =_D \int_0^1 \psi(t) \left\{B_p^{(f_i)}(t)\right\}^2 dt, i = 0, 1, \dots, p-1$. Let $u_1^{(i),I}(\alpha)$ be the αth quantile for $U_1^{(i),I}$ such that $P\left(U_1^{(i),I} < u_1^{(i),I}(\alpha)\right) = \alpha$. Then, it follows from (34) and (36) that

$$P\left(U_k^{(i),\mathrm{AR}(1)} < k \frac{1 - b_1}{1 + b_1} u_1^{(i),I}(\alpha)\right) = \alpha, \quad i = 0, 1, \dots, p-1. \tag{37}$$

Thus, if $u_k^{(i),\mathrm{AR}(1)}(\alpha)$ denotes the αth quantile for $U_k^{(i),\mathrm{AR}(1)}$, then (37) implies that

$$u_k^{(i),\mathrm{AR}(1)}(\alpha) = k \frac{1 - b_1}{1 + b_1} u_1^{(i),I}(\alpha), \quad i = 0, 1, \dots, p-1. \tag{38}$$

Similar simple multiplying factors can also be derived for the quantiles of other examples of stationary processes.

5 | SIMULATIONS AND POWER COMPARISONS

It is of interest to evaluate the power properties of the Bayes-type statistic for detection of single and multiple change points. However, because the Bayes-type statistic that tests for AMOC also tests for ALOC, if we evaluate the power properties of the AMOC test, then the same power properties also extend to the case of testing for ALOC. Furthermore, as shown in Section 4, for large samples, the case of dependence structure among errors can be handled in a straight forward manner through a multiplicative factor of the statistic under independence. In this sense, it is quite sufficient to study the power properties of the AMOC Bayes-type statistic when it is derived under the independent structure. For purposes of comparison, while evaluating the power properties of AMOC Bayes-type statistic, it is relevant to compare its power against other statistics such as the likelihood ratio type. Clearly, such power comparisons can be carried out only via simulations. It turns out that Jandhyala and MacNeill (1991) carried out a simulation study to compare powers of AMOC Bayes-type statistic under independence against two other statistics: (i) likelihood ratio statistic of Quandt (1958), derived assuming variances before change and after change differ, and (ii) likelihood ratio statistic of Worsley (1983), which is derived assuming variances before and after change are the same. Because their simulation study is quite adequate and relevant, instead of performing an altogether new simulation study, we shall briefly summarize the study performed by Jandhyala and MacNeill (1991).

While carrying out the simulations, Jandhyala and MacNeill (1991) considered the following simple polynomial regression model under the null hypothesis of no change:

$$Y_t = \beta_0 + \beta_1 (t/n) + \varepsilon_t, \qquad t = 1, 2, \dots, n, \tag{39}$$

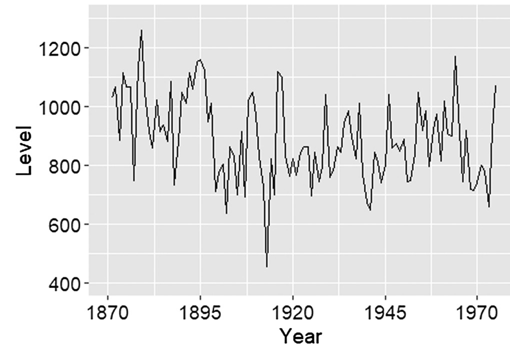

FIGURE 1 Annual flows of the Nile River for the years 1870–1975 in 10^{13} m^3 for months July–June

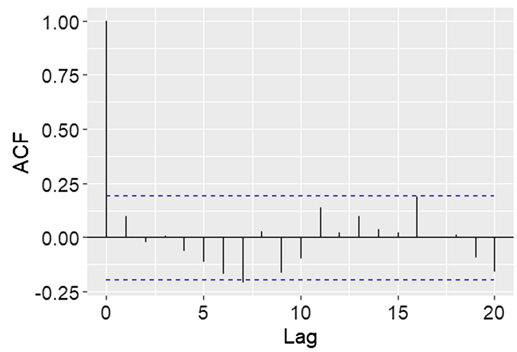

FIGURE 2 Autocorrelation correlation function (ACF) for residuals of pruned Nile River data where residuals are computed based on means within each of the four segments determined by the years of change 1899, 1953, and 1965

where ε_t~i.i.d. $N(0, \sigma^2)$. Under the alternative hypothesis of a change-point model, first they have considered change in the intercept β_0 alone such that the model subsequent to change is given by

$$Y_t = \beta_0 + \delta_0 + \beta_1 (t/n) + \varepsilon_t, \qquad t = m + 1, 2, \dots, n. \tag{40}$$

Next, they have considered change in slope β_1 alone such that the model after change is given by

$$Y_t = \beta_0 + (\beta_1 + \delta_1)(t/n) + \varepsilon_t, \qquad t = m + 1, 2, \dots, n. \tag{41}$$

Jandhyala and MacNeill (1991) have also considered a third model under the alternative by imposing the continuity constraint when change in the slope β_1 occurred. However, we shall not pursue this third model in our discussion here. While the Bayes-type statistic for change in β_i alone (i = 1,2) is given by $U_{1,n}^{(i),I}$ in (7), expressions for the statistics of Quandt (1958) and Worsley (1983) may be found in Jandhyala and MacNeill (1991). In order to compare the effect of sample size n, Jandhyala and MacNeill (1991) considered n = 20,30,40,50,60,80,100. Next, to compare the effects of amounts of change, the corresponding parameter values considered were δ_0 = 0.5,1.0,2.0 under model (40), and δ_1 = 0.5,1.0,2.0 under model (41). Noting that the effect of sample size is minimal, Jandhyala and MacNeill (1991) depicted the power comparisons for the case of n = 100 in Figure 1a–c for change in β_0 alone and in Figure 2a–c for change in β_1 alone. In comparing Figure 1a–c (change in β_0 alone), they have found that powers of all statistics were symmetric around the middle of the data. However, the highest power was not centered in the middle, but was concentrated toward one quarter and three quarters of the length of the data. When change was small $\delta_0 = 0.5$, the Bayes-type statistics showed far greater power, whereas when $\delta_0 = 2.0$, the Bayes-type and Worsley's statistics showed relatively similar powers. In the case of change in β_1 alone, Figure 2a–c showed that the highest power occurred near three quarters of the length of the data. In this case, the power curves were not symmetric, and once again, Bayes-type statistics showed far greater power when the amount of change was small ($\delta_1 = 0.5$). A more extensive discussion of the power comparisons may be found in the article of Jandhyala and MacNeill (1991).

6 | MULTIPLE CHANGE–POINT ANALYSIS OF THE NILE RIVER FLOWS

As pointed out in Section 1, the annual Nile River discharges at Aswan have been the subject of a number of articles. The Nile River data have been reported for the years 1871–1975, a total of 105 data points. In reporting the data, the hydrological year has been taken to run from July of the calendar year to June of the next so as to include the annual Nile flood, which begins approximately in August, within one hydrological year (Hosking, 1984). In addition, as observed by Hosking (1984), the raw data might have been affected by nonhomogeneity or inconsistency caused by construction of the Aswan Dam in 1903 and the use of a rating curve calibrated from discharge during the period 1903–1939 to infer the pre-1903 discharges from the measured stages. Consequently, Todini and O'Connell (1979) investigated this effect and suggested an 8% reduction be applied to the pre-1903 flows. While some authors have worked with the corrected data (Hosking, 1984; MacNeill et al., 1991), many others have worked with the original raw data (Cobb, 1978; Shao, 2011; Wu & Zhao, 2007). In addition, it should be pointed out that those who worked with the raw data have considered data only for the years 1871–1970, consisting of 100 data points. In this article, adopting Hosking (1984) and MacNeill et al. (1991), we consider the corrected data for 105 years, 1871–1975. Even though the Nile River data are well studied, the issue of possible presence of multiple change points in the data has not been studied adequately. Here, our goal is to study the Nile River data for presence of multiple change points in the mean level of the annual flows.

Let the corrected annual Nile River flows data for the years 1871–1975 be denoted by Y_i, $i = 1,2, \dots , 105$; see Figure 1 for a plot of the data. Visually, it is not particularly easy to identify the presence of changes in the mean of the data series. If there are no change points in the mean, then the observations Y_i will admit the following common mean model:

$$Y_t = \beta_0 + \varepsilon_t, \qquad 1 \le t \le 105, \tag{42}$$

where the common mean β_0 is unknown. In the above, we assume that the error variables ε_t~i.i.d. $N(0, \sigma^2)$. In contrast, if multiple change points are present in the mean, then the following multiple change-point model may be relevant for the data series:

$$Y_i = \begin{cases} \beta_0 + \varepsilon_i & 1 \le i \le m_1 \\ \beta_0 + \delta_1 + \varepsilon_i & m_1 + 1 \le i \le m_2 \\ \quad\vdots & \quad\vdots \\ \beta_0 + \delta_1 + \cdots + \delta_k + \varepsilon_i & m_k + 1 \le i \le 105, \end{cases} \tag{43}$$

where all the parameters in the above model are unknown including β_0 being the initial mean, k being the number of change points, $m_1, \dots ,m_k$ being the locations of the change points, and $\delta_0, \dots ,\delta_{k-1}$ being the amounts of changes at each of the k change points. Once again, we let ε_t~i.i.d. $N(0, \sigma^2)$. Clearly, the error variance σ^2 remains unknown for the Nile River data series. Prior to estimation of the parameters under either model (42) or model (43), one should carry out a test for the validity of one model against the other. Thus, it is important to carry out a test of the following hypotheses:

$$H_0 : \text{No change model (42) is true} \tag{44}$$

$$\text{versus} \qquad H_a : \text{Multiple change-point model (43) is true for some } k, \text{where } 1 \le k \le 104. \tag{45}$$

Clearly, the above is a case of testing for no change in the mean under H_0 against ALOC under H_a. From Theorem 2 in Section 2, it follows that it is sufficient to apply the AMOC test statistic in order to test for ALOC in the mean of the data series. Throughout this section, we shall assume that the prior distribution on the locations of k-multiple change points is jointly uniform for any value of k, $1 \le k \le 104$. Then, as suggested by (14) and (15), the uniform prior assumption will ensure that condition (10) in Theorem 2 is automatically satisfied. Thus, in the presence of error variance σ^2, and when the prior on the unknown change point is uniform, the corresponding AMOC Bayes-type test statistic $U_{1,n}^{(0),I}$ in (7) can be written as

$$U_{1,n}^{(0),I} = \frac{1}{104 \times \sigma^2} \sum_{m=1}^{104} \mathbf{Y}' R \mathbf{X}_{m0} \mathbf{X}'_{m0} R \mathbf{Y}, \tag{46}$$

where $\mathbf{X}_{m0}$ is the vector of 1 s with first m elements replaced by 0 s. In its present form, the statistic $U_{1,n}^{(0),I}$ in (46) cannot be implemented because σ^2 is unknown. However, because the sample size is large, it is sufficient to replace σ^2 by a suitable consistent estimator. In order to retain good power properties of the statistic $U_{1,n}^{(0),I}$, it is preferable that the estimate of σ^2 be consistent not only under H_0 but also under H_a. Such an estimator can be obtained by de-trending the series through differencing. Here, we obtain $\hat{\sigma}_1^2$ from first differencing of $\mathbf{Y}$, and $\hat{\sigma}_2^2$ from the second differencing. Upon computing, we obtain $\hat{\sigma}_1^2 = 131.32$ and $\hat{\sigma}_2^2 = 121.55$. The corresponding values of the AMOC statistic are $U_{1,n}^{(0),I}\left(\hat{\sigma}_1^2\right) = 1.981$ and $U_{1,n}^{(0),I}\left(\hat{\sigma}_2^2\right) = 2.141$. From Jandhyala and MacNeill (1989, 1991), it follows that the limiting distribution of the AMOC statistic $U_{1,n}^{(0),I}$ under uniform prior on the change point ($\psi(t) = 1$) is given by the following form:

$$n^{-1}U_{1,n}^{(0),I} \xrightarrow{D} \int_0^1 \left\{B_1^{(f_0)}(t)\right\}^2 dt, \tag{47}$$

where $\left\{B_1^{(f_0)}(t)\right\}$, $t \in [0,1]$ is the Brownian bridge given by $B_1^{(f_0)}(t) = B(t) - tB(1)$. Here, $\{B(t)\}$, $t \in [0,1]$ represents the standard Brownian motion. Quantiles for the stochastic integral in (6) when the underlying process is the Brownian bridge $\{B_1^{(f_0)}(t)\}$, $t \in [0,1]$, have been computed by several authors including Jandhyala and MacNeill (1989). Suppose $u_{0.99}$ represents the 99th percentile for the Brownian bridge, then we have $u_{0.99} = 0.7435$. Clearly, p value for the AMOC statistic is much less than 0.01 (p value $< < 0.01$) under both $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$. Thus, we conclude that there is strong evidence for at most once change in the mean of the data series. However, because AMOC implies ALOC from Theorem 2, we also conclude that the multiple change-point model (43) applies for the Nile River data. Thus, we must estimate the parameters β_0; k; $m_1, \dots, m_k$; and $\delta_0, \dots, \delta_{k-1}$.

In recent times, there has been growing literature for identifying multiple change points—both the number and their locations. Among the many that are available in the literature, we mention here few of them: binary segmentation (BS; Hawkins, 2001), optimal partitioning (OP; Jackson et al., 2005), PELT (Killick, Eckley, & Haynes, 2014; Killick et al., 2012), Bayesian method (Y. Li & Lund, 2015), and minimum description length (MDL) method (Davis, Lee, & Rodriguez-Yam, 2006; S. Li & Lund, 2012). Here, we implement the PELT method for estimating the number of change points and their locations in the Nile River flows data. The PELT is an exact method and under mild conditions has a computational cost that is linear in the number of data points. In addition, PELT is better than OP (Killick et al., 2012) and is faster than other exact search methods (Wambui, Waititu, & Wanjoya, 2015). The method can also be easily implemented in R by using the package "Changepoint" (Killick et al., 2014). While implementing PELT, we have set the minimum segment length to be 10 so as to avoid identifying anomalous data points as change points. The method estimated the number of change points in the data mean to be $\hat{k} = 3$. The corresponding locations were estimated as $\hat{m}_1 = 28$, $\hat{m}_2 = 83$, and $\hat{m}_3 = 95$. At this stage, while we accepted the PELT estimate $\hat{k} = 3$ for the number of change points, we decided to cross-validate the PELT estimates $\hat{m}_1$, $\hat{m}_2$, and $\hat{m}_3$ of their locations. For this purpose, we decided to obtain new estimates $\widetilde{m}_1$, $\widetilde{m}_2$, and $\widetilde{m}_3$ by minimizing the sum of squared residuals among all the $\binom{104}{3}$ 3-tuple change points. The best 3-tuple change points that minimized the sum of squared residuals yielded $\widetilde{m}_1 = 28$, $\widetilde{m}_2 = 45$, and $\widetilde{m}_3 = 47$. These new estimates differ from the PELT estimates, and this is clearly a matter of concern. However, on closer examination, it may be noted that the segment between $\widetilde{m}_2 = 45$ and $\widetilde{m}_3 = 47$ consists of only two observations, namely, Y_{46} and Y_{47}. The question arises whether these two observations are anomalous. In order to ascertain this, we compare the sum of squared residuals for the data $Y_{29}, \dots, Y_{105}$ against the data $Y_{29}, \dots, Y_{45}, Y_{48}, \dots, Y_{105}$ where the potential anomalous observations Y_{46} and Y_{47} have been removed. In computing the sum of squared residuals, we used $\hat{m}_2 = 83$ and $\hat{m}_3 = 95$ as change points for the data $Y_{29}, \dots, Y_{105}$, while these change points were shifted down by two to account for the deletion of two observations in the data set $Y_{29}, \dots, Y_{45}, Y_{48}, \dots, Y_{105}$. The corresponding sum of squared residuals were 14,867.6 and 13,340.17, respectively. Thus, there is a reduction of as much as 11% by the deletion of observations Y_{46} and Y_{47}. On this basis, we decided Y_{46} and Y_{47} were anomalous and removed them from the data, altogether. We then computed both PELT estimates and the best 3-tuple change points by minimizing the sum of squared residuals for the pruned data $Y_1, \dots, Y_{45}, Y_{48}, \dots, Y_{105}$, consisting of 103 data points. This time, both PELT and the best 3-tuple change-point estimates were identical; the estimated change points common to both methods were $\hat{m}_1 = 28$, $\hat{m}_2 = 81$, and $\hat{m}_3 = 93$.

It should be noted that the statistic $U_{1,n}^{(0),I}$ for detecting ALOC and the PELT method for estimating the number and locations of change points both assume independence of the data series within each segment. Thus, before validating the estimates $\hat{m}_1 = 28$, $\hat{m}_2 = 81$, and $\hat{m}_3 = 93$, we needed to cross-validate the assumption of independence. For this, we constructed plots of both autocorrelation function (ACF; see Figure 2) and partial ACF (PACF; see Figure 3) for residuals of pruned data $Y_1, \dots, Y_{45}, Y_{48}, \dots, Y_{105}$. Neither plot showed any clear evidence against the independence assumption. The assumption of normality is required only for deriving the Bayes-type statistic, and as such, it is not required for establishing its asymptotic distribution.

Thus, we concluded that the pruned Nile River data $Y_1, \dots, Y_{45}, Y_{48}, \dots, Y_{105}$ admit a multiple change-point model with three change points with the four segments being $Y_1, \dots, Y_{28}$; $Y_{29}, \dots, Y_{45}, Y_{47}, \dots, Y_{83}$; $Y_{84}, \dots, Y_{95}$; and $Y_{96}, \dots, Y_{105}$. In terms of calendar years, the estimated change points represent the years 1899, 1953, and 1965. Let μ_i, $i = 1,2,3,4$ represent the mean river flow within each segment. Then, the estimated mean flows for the four segments are $\hat{\mu}_1 = 1,009.93$,

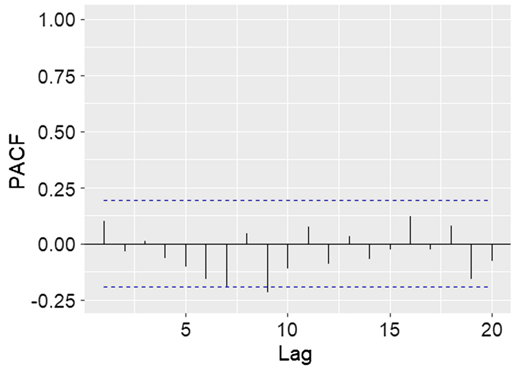

FIGURE 3 Partial autocorrelation correlation function (PACF) for residuals of pruned Nile River data where residuals are computed based on means within each of the four segments determined by the years of change 1899, 1953, and 1965

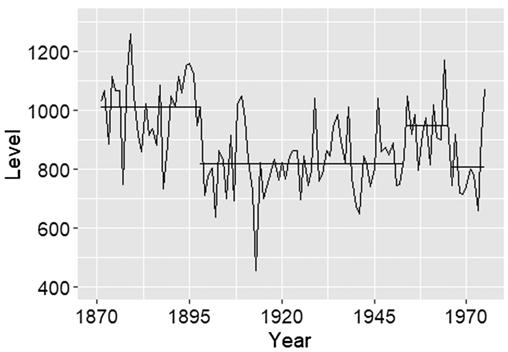

FIGURE 4 Pruned Nile River data together with means within each of the four segments determined by the years of change 1899, 1953, and 1965

$\hat{\mu}_2 = 819.59$, $\hat{\mu}_3 = 847.73$, and $\hat{\mu}_4 = 806.70$. A plot of the pruned data together with the estimated means for the four segments is presented in Figure 4.

Earlier, while searching for a single change point, many authors estimated the change point in the year 1899 (see Cobb, 1978; Kim & Hart, 2011; Shao, 2011; Wu & Zhao, 2007). In contrast, MacNeill et al. (1991) identified three change points in the data and estimated the change points to have occurred during the years 1899, 1954 and 1965. In identifying these change points, MacNeill et al. (1991) modeled the data series from year 1907 onwards as an ARMA (2, 2) time series, prior to identifying change points. It should be noted that while MacNeill et al. (1991) worked with full data, our modeling is based upon the pruned data. In any case, both approaches have yielded essentially the same change-point estimates. This illustrates the robustness of the identified change points for the Nile River data.

MacNeill et al. (1991) discussed at length the possible reasons for the changes identified in the years 1899, 1954, and 1965. Here, we only summarize their observations. The change in the year 1899 coincided with the beginning of the dam construction at Aswan. While the change in the year 1954 could be attributed to completion of Owen Falls Dam on Lake Victoria as well as other significant hydropolitical events, the change in the year 1965 was attributed to the beginning of filling the High Dam reservoir in that same year.

As a final thought, we wish to comment on the assumption of uniform prior on the change points while implementing the change detection methodology. The uniform prior essentially implies that we have no prior knowledge about the locations of change points in the data. Because we do have the knowledge that a Dam was constructed in the year 1903, could we have not imposed a more informative prior to reflect this knowledge? While this proposition has merit, in practice, it is perhaps preferable to impose a uniform prior while detecting change points. Because the uniform prior assumes no knowledge about the presence of change points, it can be seen as being more conservative. Thus, if a particular change point has been detected under a uniform prior, then one can conclude that a more informative prior would have identified the same change point with greater power, provided the prior information is accurate. However, if the prior information on the potential presence of a change point is erroneous, then imposing an informative prior on the basis of the erroneous information can lead to identification of a nonexisting change point. Thus, even if conservative, a uniform prior guards against unintentional errors that might occur while imposing more informative priors on the unknown change points. This is all the more important because the influences of known external factors on a time series are often uncertain in nature. However, in the event that the prior information is reliably correct, one can then impose an informative prior given by, say, $p(m_1, \dots, m_k)$, $1 \leq m_1 < m_2 < \cdots < m_k \leq n-1$. Then, for Theorem 2 to be applicable, the prior $p(m)$ on the single change point must satisfy (10), which may not be uniform any further. In this case, when one moves on to the asymptotic scenario, the function $\psi(t)$ in (29) will then be such that $\psi(t) \neq 1$, $t \in (0,1)$. Because quantiles for the stochastic integral (26) are currently available only for the case $\psi(t) = 1$, $t \in (0,1)$, one will not be able to apply the Bayes-type change detection methodology as readily as in the uniform case.

For example, in our analysis, despite imposing the uniform prior, the methodology identified a change point in the year 1899. As pointed earlier, this change coincides with the beginning of the Dam construction at Aswan. Furthermore, this change in the year 1899 has been detected despite the fact that we have used in our analysis the corrected data of Todini and O'Connell (1979) where observations prior to 1903 have been reduced by a factor of 8%. Earlier, while analyzing the original raw data, Cobb (1978), Wu and Zhao (2007), and Shao (2011) have also identified a change in the Nile River flow in the year 1899. Thus, the change in the year 1899 in the Nile River flow is robust enough that neither the conservative uniform prior nor the reduction in river flows (as implemented by Todini & O'Connell, 1979) have any impact on the observed change.

REFERENCES

Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change-point. *Econometrica*, *61*, 821–856.

Bai, J., & Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, *66*, 47–78.

Chernoff, H., & Zacks, S. (1964). Estimating the current mean of a normal distribution which is subject to changes in time. *Annals of Mathematical Statistics*, *35*, 999–1018.

Cobb, G. W. (1978). The problem of the Nile: Conditional solution to a changepoint problem. *Biometrika*, *65*, 243–251.

Csörgő, M., & Horváth, L. (1997). *Limit theorems in change-point analysis*. New York, NY: John Wiley & Sons.

Davis, R. A., Lee, T. C. M., & Rodriguez-Yam, G. A. (2006). Structural estimation for nonstationary time series models. *Journal of the American Statistical Association*, *101*, 223–239.

Gardner, L. A. (1969). On detecting changes in the mean of normal variates. *Annals of Mathematical Statistics*, *40*, 116–126.

Hawkins, D. M. (2001). Fitting multiple-changepoint models to data. *Computational Statistics and Data Analysis*, *37*, 323–341.

Hosking, J. R. M. (1984). Modelling persistence in hydrological time using fractional differencing. *Water Resources Research*, *20*, 1898–1908.

Jackson, B., Sargle, J. D., Barnes, D., Arabhi, S., Alt, A., Gioumousis, P., … Tsai, T. T. (2005). An algorithm for optimal partitioning of data on interval. *IEEE Signal Processing Letters*, *12*, 105–108.

Jandhyala, V. K. (1993). A property of partial sums of regression least squares residuals and its applications. *Journal of Statistical Planning and Inference*, *37*, 317–326.

Jandhyala, V. K., Fotopoulos, S. B., MacNeill, I. B., & Liu, P. (2013). Inference for single and multiple change-points in time series. *Journal of Time Series Analysis*, *34*, 423–446.

Jandhyala, V. K., Fotopoulos, S. B., & You, J. (2010). Change-point analysis of mean annual rainfall data from Tucumán, Argentina. *Environmetrics*, *21*, 687–697.

Jandhyala, V. K., & MacNeill, I. B. (1989). Residual partial sum limit processes for regression models with application to detecting parameter changes at unknown times. *Stochastic Processes and their Applications*, *33*, 309–323.

Jandhyala, V. K., & MacNeill, I. B. (1991). Tests for parameter changes at unknown times in linear regression models. *Journal of Statistical Planning and Inference*, *27*, 291–316.

Jandhyala, V. K., & MacNeill, I. B. (1992). On testing for the constancy of regression coefficients under random walk and change-point alternatives. *Econometric Theory*, *8*, 501–517.

Jandhyala, V. K., & MacNeill, I. B. (1997). Iterated partial sum sequences of regression residuals and tests for changepoints with continuity constraints. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *59*, 147–156.

Jandhyala, V. K., & Minogue, C. D. (1993). Distributions of Bayes-type change-point statistics under polynomial regression. *Journal of Statistical Planning and Inference*, *37*, 271–290.

Kander, Z., & Zacks, S. (1966). Test procedures for possible changes in parameters of statistical distributions occurring at unknown time points. *Annals of Mathematical Statistics*, *37*, 1196–1210.

Killick, R., Eckley, I., and Haynes, K. (2014). Changepoint: An R package for changepoint analysis. R Package Version 1.1.5. http://CRAN.R-project.org/package=changepoint

Killick, R., Fearnhead, P., & Eckley, I. A. (2012). Optimal detection of change-points with a linear computational cost. *Journal of the American Statistical Association*, *107*, 1590–1598.

Kim, J., & Hart, J. D. (2011). A change-point estimator using local Fourier series. *Journal of Nonparametric Statistics*, *23*, 83–98.

Li, S., & Lund, R. (2012). Multiple change-point detection via genetic algorithms. *Journal of Climate*, *25*, 674–686.

Li, Y., & Lund, R. (2015). Multiple changepoint detection using metadata. *Journal of Climate*, *28*, 4199–4216.

MacNeill, I. B. (1974). Tests for change of parameter at unknown times and distributions of some related functionals on Brownian motion. *Annals of Statistics*, *2*, 950–962.

MacNeill, I. B. (1978). Properties of sequences of partial sums of polynomial regression residuals with applications to tests for change of regression at unknown times. *Annals of Statistics*, *6*, 422–433.

MacNeill, I. B., & Jandhyala, V. K. (1993). Change-point methods for spatial data. In G. P. Patel & C. R. Rao (Eds.), *Multivariate environmental statistics* (pp. 289–306). Amsterdam, The Netherlands: Elsevier Science Publishers BV.

MacNeill, I. B., Tang, S. M., & Jandhyala, V. K. (1991). A search for the source of the Nile's change-points. *Environmetrics*, *2*, 341–375.

Qu, Z., & Perron, P. (2007). Estimating and testing structural changes in multivariate regressions. *Econometrica*, *75*, 459–502.

Quandt, R. E. (1958). The estimation of the parameters of a linear regression system obeying two separate regimes. *Journal of American Statistical Association*, *53*, 873–880.

Sen, A. K., & Srivastava, M. S. (1973). On multivariate tests for detecting change in mean. *Sankhyā: The Indian Journal of Statistics, Series A*, *35*, 173–186.

Sen, A. K., & Srivastava, M. S. (1975). On tests for detecting change in mean. *Annals of Statistics*, *3*, 98–108.

Shao, X. (2011). A simple test of changes in mean in the possible presence of long-range dependence. *Journal of Time Series Analysis*, *32*, 598–606.

Tang, S. M., & MacNeill, I. B. (1993). The effect of serial correlation on tests for parameter change at unknown time. *Annals of Statistics*, *21*, 552–575.

Todini, E., & O'Connell, P. E. (1979). *Hydrological simulation of Lake Nasser. Volume 1: Analysis and results*. Segrate, Italy: IBM Italia Scientific Centers.

Vogelsang, T. J. (1997). Wald-type tests for detecting breaks in the trend function of a dynamic time series. *Econometric Theory*, *13*, 818–848.

Wambui, G. D., Waititu, G. A., & Wanjoya, A. (2015). The power of the pruned exact linear time (PELT) test in multiple changepoint detection. *American Journal of Theoretical and Applied Statistics*, *4*, 581–586.

Worsley, K. J. (1983). Testing for a two phase multiple regression. *Technometrics*, *25*, 35–42.

Wu, W. B., & Zhao, Z. (2007). Inference of trends in time series. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, *69*, 391–410.

Xie, L., & MacNeill, I. B. (2006). Spatial residual processes and boundary detection. *South African Statistical Journal*, *40*, 33–53.

How to cite this article: MacNeill IB, Jandhyala VK, Kaul A, Fotopoulos SB. Multiple change-point models for time series. *Environmetrics*. 2020;31:e2593. https://doi.org/10.1002/env.2593