Vol.7.Issue.4.2019 (Oct-Dec)

http://www.bomsr.com
Email:editorbomsr@gmail.com

RESEARCH ARTICLE

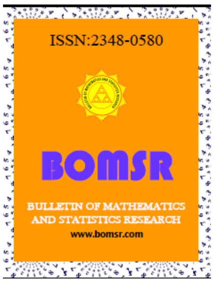
BULLETIN OF MATHEMATICS AND STATISTICS RESEARCH

A Peer Reviewed International Research Journal

INTERNATIONAL STANDARD SERIAL NUMBER 2348-0580

CHANGEPOINT ANALYSIS: A PRACTICAL TOOL FOR DETECTING ABRUPT CHANGES IN RAINFALL AND IDENTIFYING PERIODS OF HISTORICAL DROUGHTS: A CASE STUDY OF BOTSWANA

WILSON MOSEKI THUPENG*, THATO BASETSANA THEKISO
Department of Statistics, University of Botswana*,
Gaborone, Botswana, Private Bag UB00705, Gaborone, Botswana
*Corresponding author: thupengwm@yahoo.co.uk
DOI: 10.33329/bomsr.74.33



ABSTRACT

Rainfall is a very important meteorological variable for indicating climate change. Over the years there has been concerns of variability and changes in the climate around the world, including in Botswana, which have increasingly led to undesirable impacts like droughts, floods, destruction of the environment and property as well as loss of human life. The use of changepoint techniques to detect past abrupt shifts in climate time series has gained currency across the different regions of the world. The objective of this paper is to determine trends in the annual maximum average rainfall for Botswana during the years 1901-2012 and detect multiple change points in the series. Trend detection is carried out using the Mann-Kendall trend test. To test for stationarity, the Kwiatkowski, Phillips, Schmidt and Shin (KPSS) test is used. Multiple change points are detected using three widely used search methods, namely, the Segment Neighbourhood, Binary Segmentation and Pruned Exact Linear Time (PELT). The results show that the data are heavy-tailed, show no monotonic trend and are stationary. Both the Segment neighbourhood and Binary Segmentation detected multiple change points that correspond to documented periods of drought in the country. The Pruned Exact Linear Time (PELT) fails to detect a reasonable number of changepoints as it identifies almost all the years as changepoint locations.

Keywords: Rainfall, Trend, Changepoint detection, Segment Neighbourhood, Binary Segmentation, Pruned Exact Linear Time
AMS(2010) Subject Classification: 62G10, 62P12

1. Introduction

Precipitation and temperature extremes are considered to be the most important climate events and have been extensively explored over the past several decades, according Mothupi et al. (2016) [7]. Rainfall or precipitation is one of the prime precursors of climate change. Rainfall is one of the most important factors of hydrometeorology severely affected by climate change impact, according to Mazvimavi (2010) [6]. There is an increasing concern in Southern Africa about decline in rainfall and, Botswana like many countries has been experiencing low levels of rainfall over the years. The country is characterized by a semi-arid climate with an erratic, unpredictable and highly regional rainfall. It experiences very low rainfall, down pours can occur from November to April but, the wettest months in Botswana in which rainfall is usually higher are January and February (https://www.safaribookings.com/botswana/climate).

Rainfall is one of the climatic conditions or variables used for detecting changes in the climate.. . A change point is defined as that point in time at which the properties of the underlying distribution or the parameters of the model used to describe a time series abruptly change. Such abrupt changes include change in the mean, variance or in both and variance and change in the trend. Thus, cchangepoint detection is the problem of determining abrupt changes in the distribution of a stochastic process or a time series Changepoint detection methods are divided into two main branches: online methods that aim to detect changes as soon as they occur in a real-time setting, and offline methods that retrospectively detect changes when all samples are received Truonga *et al*. [13].

The use of changepoint techniques to detect past abrupt shifts in climate time series has gained currency across the different regions of the world. A number of studies have been carried out on trend analysis and offline changepoint detection of rainfall. For example, [6] investigates the validity of the assumption that rainfall is declining in Zimbabwe by using the Mann Kendall test to analyze the trend in the annual rainfall and total rainfall for the early parts of the rainfall season as well as the late parts rainfall. In order to examine the possible existence of a change or break points in the rainfall time series, [6] uses the Pettit test. However, this test is used to test for a single change point. Khapalova *et al*. (2013) [3] study global precipitation for the Northern, Tropical and Southern Latitude of the globe by using a change point methodology for Gaussian sequences in a multivariate framework. Their study seeks to give a better understanding of time periods in which changes in the annual mean precipitation have occurred during the period of study. They point out that precipitation should be monitored over time as significant changes my cause flooding or drought.

Pranuthi and Dubey (2014) [9] use change point methodology to detect abrupt changes in precipitation. The authors apply the Worsley likelihood technique and CUSUM method to detect changes. Trend is tested using the Mann-Kendall (MK) test for trend and the Spearman's Rho (SR) test, which are important and mostly used test when analysing rainfall as they are said to have capacity of testing for monotonic trend. In this study we apply only the MK test.

In the case of Botswana, Mphale *et al.* (2013) [8] study long term trends and abrupt changes of a sixty-year long rainfall series for the months of January, February and March for 4 stations of Maun, Ghanzi, Tsabong and Tshane. They apply the Mann Kendall trend test in their study. The test reveals that Maun, Tsabong and Tshane experienced a downward trend in rainfall series in those months. On the other hand, the test reveals a positive insignificant trend for Ghanzi. To detect abrupt changes in the monthly average rainfall series, [8] use the Pettit and Mann-Witney U tests. An abrupt change in rainfall regimes in these areas was investigated and was found to occur in the year 1982 for all stations [8]. This is certainly a very interesting result which, from a statistical perspective, can be explained by the fact that both the Pettit and Mann-Witney U tests are for detection of at most a single changepoint and that the largest change in the rainfall series might have happened in 1982 across all the 4 the stations for the reason posited by the authors. Applying multiple changepoint techniques to the same data would most probably have led to a different result.

Reeves *et al.* (2007) [10] provides a review and comparison of change point detection techniques for climatic data. The methods examined include the standard normal homogeneity (SNH) test, Wilcoxon's non parametric test, two-phase regression (TPR) procedures, inhomogeneity tests, information criteria procedures and various variants thereof. The work seeks to show the fundamental differences among the assumptions made by each procedure and provide guidelines for which procedures works best in different situations. The results indicates that the TPR and Sawa's Bayes criteria procedures seem optimal for most climate time series whereas the SNH procedure and its nonparametric variant are probably best when trend and periodic effects can be diminished by using homogenous reference series. However, as pointed by [10], the procedures presented in their work are heavily dependent on the "at most one changepoint" assumption. If the series is long and undocumented changepoints are plausible, then more than one undocumented changepoint may be present [10]. More recently, interest in changepoint detection problems has shifted to detection of multiple changes in a time series. As pointed out by Killick and Eckley (2014) [4], with increased collection of time series and signal stream there is growing need to be able to efficiently and accurately estimate the location of multiple change points. Many common statistical approaches to detecting changepoints can be formulated in terms of minimising a cost over segmentations. The present paper falls within the realm of offline changepoint detection of rainfall and focusses on trend analysis and multiple changepoint detection in the mean of the annual maximum average rainfall for Botswana during the period 1901-2012. An attempt to carryout testing for change in the variance and change in both the mean and variance did not yield any useful results, most probably due to the stationarity of the time series being analysed. Three commonly used search algorithms to identify multiple change points within time series, namely, Binary Segmentation, Segment Neighbourhood and Pruned Exact Linear Time (PELT) are employed.

The remainder of this article proceeds as follows. Section 2 reviews the methodology to be employed in the paper, which covers trend tests, stationary test, changepoint models and multiple changepoint search algorithms. Section 3 presents results of the study and a discussion comparing the results based on the search methods to documented changepoints of the average rainfall data for Botswana. Section 4 provides a conclusion of the investigation...

2. Materials and Methods

2.1 Data Description

The data used for this study are annual maximum average rainfall series for Botswana for the period of January 1901 to December 2012. These were extracted from the monthly average rainfall series for the country for the study period obtained from the World Bank website http://sdwebx.worldbank.org/climateportal/index.cfm?page=downscaled_data_downloadandmenu =historical. The data are regarded to be of very high quality with no missing values.

2.2 Descriptive Statistics

This section deals with basic summary statistics in order to get some sense of the basic distributional properties of the data such as measures of location and shape. More important are the coefficients of skewness and Kurtosis which, respectively, indicate whether or not the data is skewed or heavy-tailed. The population coefficient of skewness is defined by the formula $\gamma_1 = \frac{\mu_3^2}{\mu_2^3}$, where μ_2 and μ_3 are the second and third moments about the population mean μ.

Kurtosis is a measure of weather the data are heavily tailed or light tailed relative to the normal distribution. That is, Kurtosis describes the tail shape of a data distribution. The excess Kurtosis of a univariate population is defined by the following formula $\gamma_2 = \frac{\mu_4}{\mu_2^2} - 3$ where μ_2 and μ_2 are, respectively, the second and fourth central moments about the population mean μ. The normal distribution has zero excess kurtosis and, is thus used as the gold standard for checking the heavily-tailedness of a distribution of data.

2.3 Testing for Trend and Stationarity

2.3.1 Mann Kendall Trend Test

Mann Kendall test (MK) is one of the non-parametric statistical tests widely used for detecting climatic trends. The purpose of the test is to statistically assess if there is a monotonic upward or downward trend in rainfall over a stated period of time.

Let $\boldsymbol{x}$ be a vector of *n* observation (annual maximum average rainfall), i.e. $\boldsymbol{x} = (x_1, x_2, .., x_n)$, where x_j represents a data point at time j.

Let $x_j - x_k$ be the possible difference between the observations corresponding to the *i*[th] and *j*[th] years and $sign(x_j - x_k)$ be an indicator function for the difference, where x_j and x_k are the sequential annual maximum average rainfall observations in years j and k, (j>k), respectively. See Equation (1).

$$i.e.\ sign(x_j - x_k) = \begin{cases} 1 & if \quad x_j - x_k > 0 \\ 0 & if \quad x_j - x_k = 0 \\ -1 & if \quad x_j - x_k < 0 \end{cases} \tag{1}$$

The Mann Kendall statistics (S) is defined as the sum of the number of positive differences minus negative differences and is defined in Equation (2).

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} sign(x_j - x_k) \tag{2}$$

with variance given by Equation (3).

$$var(S) = \frac{1}{18}\{n(n-1)(2n+5) - \sum_{p=1}^{g} t_p\,(t_p - 1)(2t_p + 5)\} \tag{3}$$

where *g* is the number of tied groups and t_p is the number of observations in the *p*[th] group.

Two-sided Mann Kendall test

H_0 : No monotonic trend over time

H_1: For one or more seasons, there is an upward or downward monotonic trend over time

Test Statistic is given by

$$Z = \begin{cases} (S-1)\,/\,\sqrt{(var(S))} & for \quad S > 0 \\ 0 & for \quad S = 0 \\ (S+1)\,/\,\sqrt{(var(S))} & for \quad S < 0 \end{cases}$$

Decision Rule: Reject H_0 if $|Z| > |Z_{1-\alpha/2}|$ and accept H_1 otherwise at a given level of significance.

2.3.2 Kwiatkowski, Phillips, Schmidt and Shin test (KPSS)

This is test for stationarity was introduced by Kwiatkowski, Phillips, Schmidt and Shin (1992). Its null hypothesis is that there is stationarity in the series around either the mean or a linear trend; the alternative hypothesis is that the series is non stationary due to the presence of a unit root. In the KPSS test, a series of observations is represented by a sum of three components: a deterministic trend, random walk and a stationary error term. The model has the form of Equation (4).

$$x_t = \xi_t + r_t + \varepsilon_t \tag{4}$$

with $r_t = r_{t-1} + \mu_t$ and $t = 1, 2, 3 \ldots, T$,

where x_t is the value of the series of observations of rainfall at time t, ξ_t is the deterministic trend, r_t is random walk process, ε_t is the error term for x_t and is assumed to be stationary, μ_t is the error term for the random walk $r_t$and is assumed to be independently and identically distributed with mean 0 and variance σ^2. The preceding assumption implies that the initial value of $r_t = r_{t-1} + \mu_t$ is a constant. Hence, the KPPS test tests for a null hypothesis of stationary around a linear trend versus the alternative of the presence of a unit root. This null hypothesis is equivalent to the assumption that the variance of the random walk (σ_r^2) equals zero. For the alternative hypothesis, when σ_r^2 is greater than zero the series will be non-stationary (as a sum of trend and random walk) due to the presence of a unit root.

Test the hypotheses

$H_0 : \sigma_r^2 = 0$

$H_1 : \sigma_r^2 > 0$

Test statistic is given as

$$K = \frac{\frac{1}{T^2}\sum_{t=1}^{T} S_t^2}{\hat{\sigma}_t^2}$$

where t is an index for the series of rainfall observations over the T years, i.e., $t = 1, 2, 3 \ldots, T$, $\hat{\sigma}_t^2$ is the estimate of variance and $S_t^2 = \sum_{i=1}^{T} e_i$.

Decision Rule: Reject H_0 if K is greater than 100(1- α) % quantile from the appropriate asymptotic distribution or if the p value is improbably small.

2.4 Change Point Detection

Change point analysis is a statistical tool able to detect single or multiple changes in a time series. Change points techniques have been developed to detect abrupt shifts in the parameters of a distribution or in the coefficients of the regression model.

A change point is said to occur within a set of data when there exists a time after which the statistical properties of the distribution change in some way. It is a point in time at which the parameters of the underlying distribution or the parameters of the model used to describe the time series abruptly change.

In this study, changepoint analysis will be used to detect how many change points have occurred over the years in Botswana's annual maximum average rainfall over the study period and estimate the locations of the different changes.

2.4.1 Multiple Change point Detection Model

The general formulation of multiple change point analysis is as follows.

Given a set of observed time series of annual maximum average rainfall $x_1, x_2, .., x_n$ with probability density functions $p(x_1), p(x_2), \dots, p(x_n)$, the multiple change point model can be formulated as in Equation (5).

$$x_t = \begin{cases} p_1(x_1|\theta_1) & 1 \le t \le t_1 \\ p_2(x_2|\theta_2) & t_1 \le t \le t_2 \\ \vdots & \\ p_n(x_n|\theta_n) & t_{n-1} \le t \le t_n \end{cases} \tag{5}$$

where θ_t denotes the unknown parameter (e.g., mean or variance) of the probability model for segment t. The parameters $\theta_1, \theta_2, .., \theta_n$ and the changepoints $t_1, t_2, .., t_{n-1}$ are unknown.

The entire series can be divided into n homogenous groups called segments if locations of change points is detected. If there is only one change point Equation (5) is called a single change point model, as discussed in subsection 2.4.2

2.4.2 Single Changepoint Detection Model

Changes in the mean and variance are important when performing a changepoint analysis or detecting a changepoint in a time series. There can be a change in the mean, variance or a change in both the mean and variance. To detect a changepoint in all the three possible shifts, search algorithms or methods can be used. For example, to detect a change in the mean, the model with a shift is compared with a model without a shift after the most likely time for a shift is identified. To formalise things, consider the model

$$x_t = \mu + \varepsilon_t \ , \varepsilon_t \sim N(0, \sigma^2)$$

where x_t is value of the response variable at time t, $t = 1, 2, 3 \dots, T$, μ is the overall mean and ε_t is the random error term with mean 0 and variance σ^2.

Model 1: Change in the mean

To detect a change point in a series we test the null hypothesis

H_0: There is no change in the mean

against the alternative

H_1: There is a change in the mean

These hypotheses can be formally represented by the model represented by Equation (6).

$$x_t = \begin{cases} \mu_1 + \varepsilon_t, & \varepsilon_t \sim N(0, \sigma^2), \quad t = 1, 2, \dots, k \\ \mu_2 + \varepsilon_t, & \varepsilon_t \sim N(0, \sigma^2), \ t = k+1, \dots, T \end{cases} \tag{6}$$

where μ_1 and μ_2 are the means before and after the unknown change at time *k* occurred?

Sometimes interest may be in detecting change in the variance. The appropriate model is then stated formally as Model 2 represented by Equation (7).

Model 2: Change in the variance

Similarly, to test for change in variance, we have the models

$$x_t = \begin{cases} \mu + \varepsilon_t, & \varepsilon_t \sim N(0, \sigma_1^2), \quad t = 1, 2, \dots, k \\ \mu + \varepsilon_t, & \varepsilon_t \sim N(0, \sigma_2^2), \ t = k+1, \dots, T \end{cases} \tag{7}$$

where σ_1^2 and σ_2^2 are. respectively, the variances before and after the unknown change point at time k.

Model 3: Change in both mean and variance

Model 3, expressed in Equation (8), is used for testing of changes in both the mean and variance we have the following models.

$$x_t = \begin{cases} \mu_1 + \varepsilon_t, & \varepsilon_t \sim N(0, \sigma_1^2), \quad t = 1, 2, \dots, k \\ \mu_2 + \varepsilon_t, & \varepsilon_t \sim N(0, \sigma_2^2), \ t = k+1, \dots, T \end{cases} \tag{8}$$

2.5 Search methods for identifying multiple change points

In this study we consider three search algorithms/methods for multiple changepoints discussed in Killick and Eckley (2014) [4]: Binary Segmentation, Segment Neighbourhood and Pruned Exact Liner Time (PELT). According [4], given m segments of the data, all the three methods minimise the function contained in Equation (9).

$$\sum_{i=1}^{m+1} \left[c\left(x_{(t_{i-1}+1):t_i} \right) \right] + \beta f(m) \tag{9}$$

where c is a cost function (for example, negative log-likelihood) for a segment and $\beta f(m)$ is a penalty to guard against overfitting, a multiple changepoint version of the threshold c.

2.5.1 Binary Segmentation

The Binary Segmentation proposed by Scott and Knott (1974) [11] is arguably the most widely used changepoint search method (Killick *et al.*, 2012) [5]. This method originates from the work of Edwards & Cavalli-Sforza (1965) [2]. It first detects a single changepoint of the entire data. If a change point is identified the data is split into two segments at the change point location and the single change point procedure is repeated on the two data sets, before and after the change point. If further change points are identified the data are split again. The procedure continues until no change points are found in any parts of the data segments. However, according to the authors, the procedure is an approximate maximization of the function given by Equation (9).

2.5.2 Segment Neighbourhood

The Segment Neighbourhood method was proposed by Auger and Lawrence (1989) [1]. This method minuses the function (9) exactly using a dynamic programming technique to obtain the optimal segmentation for m+1 changepoints reusing information that was calculated from change points. This method is considered to be exact.

2.5.3 Pruned Exact Liner Time (PELT)

This method was proposed by Killick *et al.* (2012) [5]. It is similar to the Segment Neighbourhood since it provides an exact segmentation. Due to its construction it can be computationally efficient, due to its use of dynamic programming.

The descriptive statistics, trend analysis and change point detection techniques discussed in this section are implemented using the R Software.

3. Results and Discussion

3.1 Descriptive Statistics

Figure 3.1 shows a time series plot of the maximum annual average rainfall for Botswana during the years 1901-2012. From the figure, it is clear that the series of maximum annual average rainfall is characterised by troughs and peaks. Based on a visual inspection of the data in Figure 3.1, it appears that the series has no monotonic trend, is stationary but that multiple changes in mean may exist.

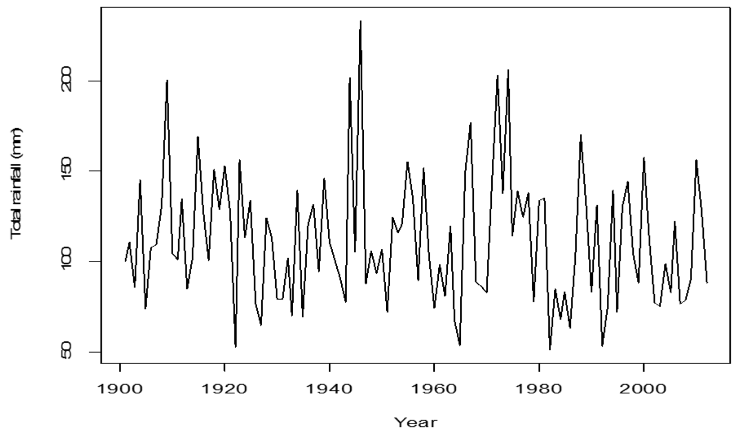


Figure 3.1. Time Series Plot of maximum annual average rainfall years 1901 to 2012

The results contained in Table 1 indicate that during the years 1901 to 2012, Botswana received the highest annual maximum average rainfall of 233.60 mm and the lowest of 51.02 mm. The kurtosis coefficient of $\gamma_2 = 3.610997$, which is greater than that of the normal distribution (which is 3), shows that the data are heavy-tailed relative to the normal distribution. Furthermore, the coefficient of skewness of $\gamma_1 = 0.77479$ indicates that the data are not symmetric but rather positively skewed.

Table 3.1: Summary statistics of annual average rainfall

Minimum value	Mean	Median	Maximum value	Skewness	Kurtosis
51.02	112.1	106.6	233.60	0.77479	3.610997

3.2 Trend and Stationarity Tests

3.2.1 Mann Kendall Trend Test

The following are the results from the Mann Kendall test:

Two-sided homogeneity test

H0: No monotonic trend over time

H1: There is monotonic trend (an upward or downward) overtime

Statistics for total series

$S=$-661 $Var(S)=$ 4195871 $Z=$ -0.3 tau =-0.012 $p-value=$ 0.7473

The p-value from the Mann Kendall test is significantly large. This means we fail to reject the null and conclude that the data do not show any monotonic trend, implying that the observations are identically and independently distributed.

3.2.2 Kwiatkowski, Philips, Schmidt and Shin (KPSS) test

For the KPSS test, we obtain the following results:

KPSS Level = 0.05193 Truncation lag parameter = 4 $p-value=$ = 0. 1

P-value greater than printed p-value

The p-value is significantly large, and we, therefore, fail to reject the null hypothesis of stationarity and conclude that the data is stationary.

3.3 Change Point Detection

The following results are for change points detection based on the three multiple change point search methods discussed earlier on in this report.

3.3.1 Binary Segmentation

Change point type: Change in mean

Method of analysis: Binary Segmentation (BinSeg)

Minimum Segment Length: 1

Maximum no. of change points: 5

Change point Locations: 58 70 74 81 86

Range of segmentations:

	[,1]	[,2]	[,3]	[,4]	[,5]
[1,]	81	NA	NA	NA	NA
[2,]	81	70	NA	NA	NA
[3,]	81	70	74	NA	NA
[4,]	81	70	74	86	NA
[5,]	81	70	74	86	58

Table 3.2: Change point locations detected using Binary Segmentation

Year	1958	1970	1974	1981	1986
Month	1	12	1	1	1
Rainfall (mm)	152.13	83.08	206.39	139.12	63.13

The R output above and Table 3.2 show the multiple change points in the mean detected in the annual maximum average rainfall for Botswana from January 1901 to December 2012 using the Binary Segmentation search. As seen from these results, the Binary Segmentation method detected

5 change points that occurred in January 1958, December 1970, January 1974, January 1981 and January 1986. Figure 3.2 shows a graphical representation of the change points plus one for the last segmentation.

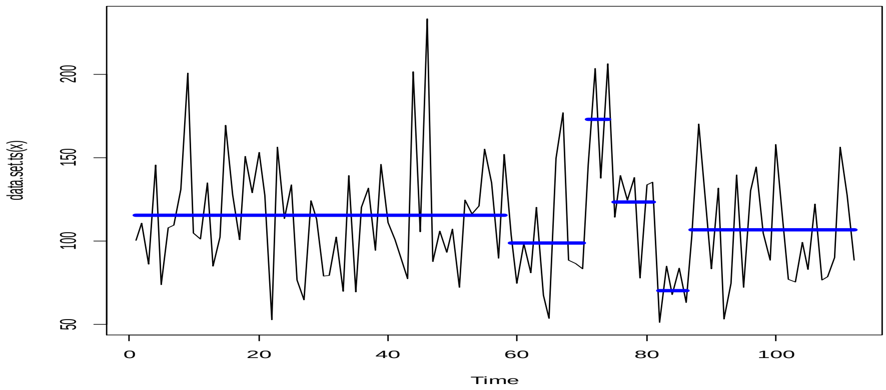


Figure 3.*2* Binary Segmentation changepoints with default penalty

3.3.2 Segment Neighbourhood

Change point type: Change in mean

Method of analysis: Segment Neighbourhood (SegNeigh)

Minimum Segment Length: 1

Maximum no. of Change points: 15

Change point Locations: 43 46 51 58 65 67 70 74 81 86 89 90 91 92

Table 3.3 Changepoint locations detected using Segment Neighbourhood

Year	Month	Rainfall(mm)
1943	1	77.43
1946	1	233.58
1951	1	72.07
1958	1	152.13
1965	1	53.38
1967	1	176.96
1970	12	83.08
1974	1	206.39
1981	1	135.12
1986	1	63.19
1989	2	131.47
1990	1	83.12
1991	1	131.6
1992	2	53.12

The results summarised in Table 3.3 and Figure 3.3 show that, in contrast to the Binary Segmentation method, which detected only 5 changepoint locations during the stated period of time, the Segment Neighbourhood search method has detected 14 changepoint locations in the rainfall series in the years and months indicated in the table. Figure 3.3 shows an additional changepoint for the last segmentation.

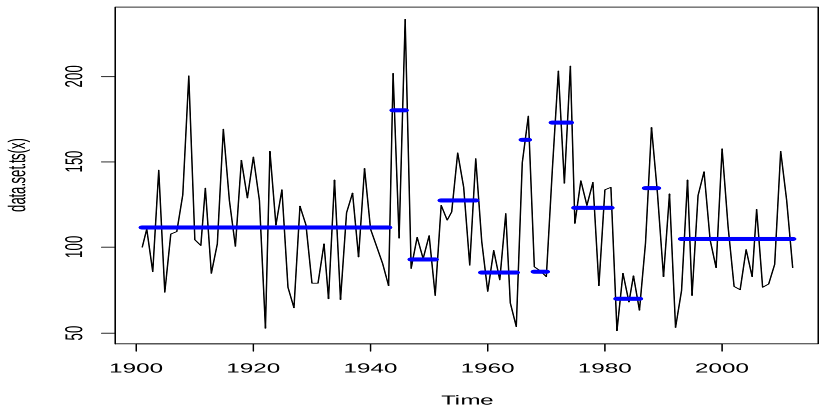


Figure 3.2 Segment Neighbourhood changepoints with manual penalty

3.3.3 Pruned Exact Liner Time (PELT)

Changepoint type: Change in mean

Method of analysis: PELT

Minimum Segment Length: 1

Maximum no. of cpts : Inf

Number of changepoints: 102

1 2 3 4 5 7 8 9 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 54 55 56 57 58 59 60 61 62 63 64 65 66 67 70 71 72 73 74 75 76 77 78 79 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 103 104 105 106 108 109 110 111

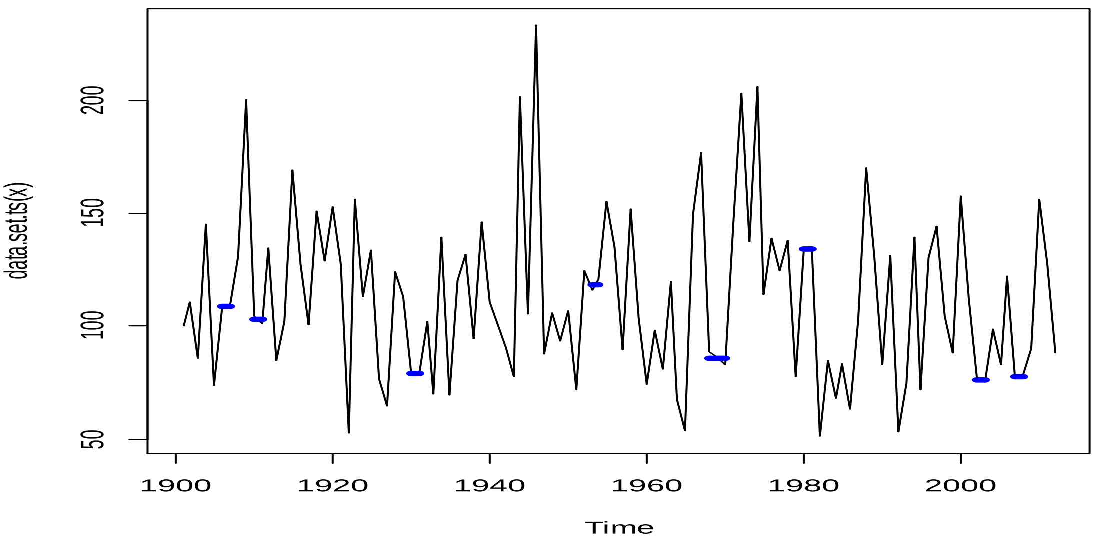


Figure 3.4 PELT changepoints with manual penalty

As indicated earlier on in this paper, the results show that both the Binary Segmentation and Segment neighbourhood techniques have detected the same multiple changepoints (January 1958, December 1970, January 1974, January 1981 and January 1986), with the Segment neighbourhood detecting extra multiple changepoints. The PELT, on the other hand fails to pinpoint an optimal solution. These results are surprising as the PELT algorithm proposed by [5] is similar to that of the Segment Neighbourhood algorithm since it provides an exact segmentation [4]. It is clear from the results in this subsection that the PELT method has failed to detect a reasonable number of changepoints in the mean of Botswana's annual maximum average rainfall series for the period 1901-2012. The technique identifies almost all the years of the 112-year series as changepoint locations (from 1901 to 2011). This behaviour of the algorithm is unusual as the PELT is most efficient in applications where the number of changepoints is large. Perhaps the method requires a longer time series than the 112-year series under study.

To assess the utility of multiple changepoint detection techniques for rainfall data considered in this investigation, we compare the results from the Binary Segmentation and Segment neighbourhood techniques with the documented 'Occurrences of droughts in Botswana' declared by the country's State Presidents during the period 1961-2015, Table 4.1a of STATISTICS BOTSWANA. BOTSWANA ENVIRONMENT STATISTICS: NATURAL DISASTERS DIGEST 2015. The table is reproduced her for ease of reference. We believe that, for practical purposes, this comparison is reasonable and practicable since Botswana, being an arid to a semi-arid country, is more prone to severe and prolonged droughts than any other factor, like floods or change of weather monitoring station, which might lead to an abrupt change in the rainfall series.

Table 4.1a: Occurrences of Drought in Botswana (Drought declarations)

Year	Impact area
1961-1965	Drought year (North East, Central)
1968-1971	Non-drought years
1979-1980	Drought year (Bobirwa)
1981-1987	**Drought years (Whole Country)**
1991-1999	**Drought years (Whole Country**
2001-2005	**Drought years (Whole Country)**
2007-2008	**Drought year (Whole Country)**
2001-2005	**Drought years (Whole Country)**
2007-2008	**Drought year (Whole Country)**
2008-2009	Non-drought year
2009-2010	**Whole Country**
2010-2011	**Whole Country**
2011-2012	**Whole Country**
2012-2013	**Drought year (Whole Country)**
2013-2014	Non-drought year
2014-2015	**Drought year (Whole Country)**

Source: Environment Statistics 2000: Rural Development Council

From results contained in Tables 3.2 and 3.3 (and Figures 3.2 and 3,3, respectively) and Table 4.1a suggest that the changepoints in the mean of Botswana’s annual maximum average rainfall that occurred in the years 1965, 1981, 1986, 1991 and 1992 are attributable to drought. Drought can be defined as a deficiency in rainfall in terms of its timing, spatial-temporal distribution, and/or overall amounts received and whether they were severe enough to negatively affect plant growth, water supplies, wildlife condition and ultimately human livelihoods and food security in general [12]. The results in Table 3.3, third column (Rainfall (mm)) are consistent with this definition of drought. For example, the results show that annual maximum average rainfall in Botswana dropped from a high of 152.13 mm in 1958 to a low of 53.38 mm in 1965, representing an 8-year decline of 98.75 mm. Similarly, the results show a decline in the annual maximum average rainfall from 206.39 mm in 1974 to 135.12 mm in 1981, representing an 8-year decrease in magnitude of 71.27 mm. Continuing this way using the results in Table 3.3 reveals that, actually, most of the changepoints in the mean of Botswana’s annual maximum average rainfall are attributable to drought

4. Conclusion

The results of this investigation indicate that annual maximum average rainfall series in Botswana for the period 1901-2012 is not normally distributed. Also, the result of the Mann Kendall trend test show that the rainfall series has no monotonic trend whilst the KPSS stationary test shows that the series is stationary.

Furthermore, for this time series, of the three search techniques of detecting multiple changepoints, the Segment neighbourhood and the Binary Segmentation seem to perform far better than the Pruned Exact Linear Trend (PELT) in terms of finding reasonable optimal changepoints. Both the Segment neighbourhood and the Binary Segmentation have detected the same changepoints, with the Segment neighbourhood detecting extra multiple changepoints. Thus, compared to the Binary segmentation and the Segmentation Neighbourhood techniques, the PELT method has failed to detect a reasonable number of changepoints in the mean of Botswana’s annual maximum average rainfall series for 1901-2012. The results further suggest that most of the changepoints in the mean of Botswana’s annual maximum average rainfall are attributable to drought

5. References

[1]. Auger, I. E. & Lawrence, C. E. (1989). Agorithms for the optimal identification of segment neighborhoods. *Bulletin of Mathematical Bilogy* **1** (51), 39-54.

[2]. Edwards, A. W. & Cavalli-Sforza, L. L. (1965). A Method for Cluster Analysis. *Biometrics*(21), 362-375.

[3]. Khapalova, E. A., Jandhyala, V. K. & Fotopoulos , S. B. (2013, February). Change-Point Analysis of Annual Mean Precipitation for Northern, Tropical and Southern Latitued of the Globe in the Past Century. *Journal of Environmental Statistics, **4*** (3). http://www.jenvstat.org.

[4]. Killick, R. & Eckley, I. A. (2014). Changepoint: An R Package for Changepoint Analysis. Journal of Statistical Software **58** (3). https://www.jstatsoft.org/issue/view/v058.

[5]. Killick, R., Fearnhead, P. & Eckley, I. A. (2012). Optimal detection of change points with a linear computational cost. *JASA,* **100** (500), 1590-1598.

[6]. Mazvimavi, D. (2010). Investigating changes over time of annual rainfall in Zimbabwe. *Hydrology and Earth System Sciences* **5** (4), 2671-2679. DOI: 10.5194/hessd-5-1765-2008

[7]. Mothupi, T., Thupeng, W. M., Mashabe, B. & Mokoto, B. (2016). Estimating Extreme Quantiles of the Maximum Surface Air Temperatures for the Sir Seretse Khama International Airport uisng the Generalized Extreme Value Distribution. *American Journal of Theoretical and Applied Statistics* **5** (6), 365-375.

[8]. Mphale, K. M., Dash, S. K., Adedoyin, A. & Panda, S. K. (2013). Rainfall regime changes and trends in Botswana Kalahari Transect's late summer precipitation. **116,** 75–91, (2014) doi:10.1007/s00704-013-0907-z.

[9]. Pranuthi , G. & Dubey, S. K. (2014). Trend and Change Point Detectiono of Precipitation in Urbanizing Districts of Uttarakhand in India. *Indian Journal of Science and Technology,* **7** (10), 1573-1582.

[10]. Reeves Jaxk, Chen, J., Wang, X. L., Lund, R. & Lu, Q. (2006). A Review and Comparison of Changepoint Detection Techniques for Climate Data. *Jounal Of Applied Meteorology And Climatology* **46** (6), 900-915.

[11]. Scott, A. J. and Knott, M. (1974). A cluster analysis method for grouping means in the analysis of variance. Biometrics, 30(3):507–512.

[12] STATISTICS BOTSWANA. BOTSWANA ENVIRONMENT STATISTICS: NATURAL DISASTERS DIGEST 2015. statsbots.org.bw/sites/default/files/publications/Botswana Environment Natural...

[13]. Truonga, C., Oudreb, L. & Vayatisa, N. .Preprint submitted to Signal Processing January 17, 2019. arXiv:1801.00718v2 [cs.CE].