Detecting Change Points in Climatic Data Using Extreme Values

Andrew Bartlett * Cezareo Rodriguez †

September 28, 2016

Abstract

Extreme weather and climate events such as hot spells, snow storms, and floods have recently had a major impact on the economy, environment, and human well-being. However, when trying to develop a statistical model for climate data, a change-point is the single most important contributing factor for inaccurate or accurate results. Traditional change-point detection methods focus exclusively on detecting an alteration or a shift in the arithmetic mean. However, a change in the climate will first be recognized through changes in the frequency or intensity of extreme values. In this talk, we will first discuss our statistical method that uses extreme values to estimate the number and location of change-points within a climatic time series. Implementation of our Extreme value to various time series are discussed. We then compare the performance of our Extreme value method with a standard Bayesian method. Both methods are applied to simulated and real climate data.

Keywords: Climate change, Change-point, Change-point detection, Extreme values, Time Series, Bayesian framework

*University of Arkansas, 1 University of Arkansas, Fayetteville , AR 72701
†Southern Illinois University-Edwardsville, 1 Hairpin Dr., Edwardsville, IL 62026

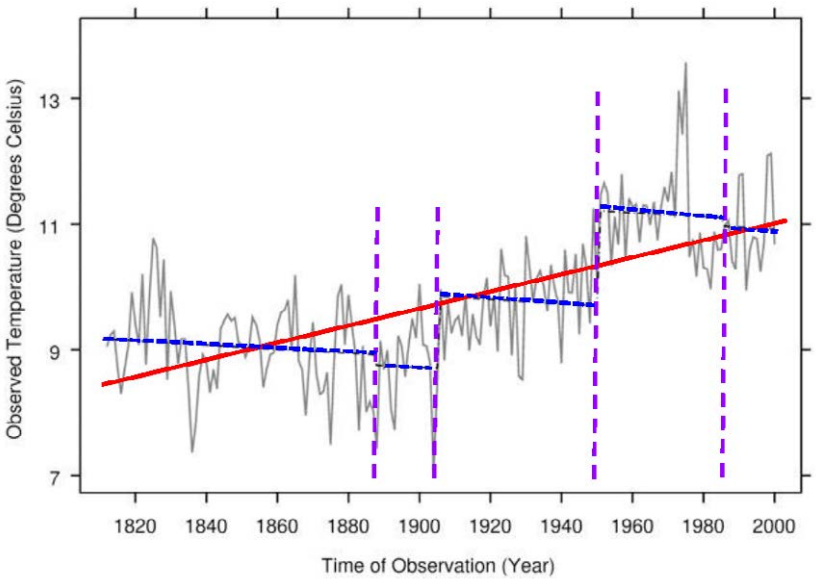
1 Background

With NASA's recent publications stating that 2015 broke 2014's record for warmest global temperatures, the climate change debate is hotter than ever. Extreme heat events are becoming more frequent, and rainfall (or lack thereof) is becoming increasingly volatile in recent years. Although the effects of climate change are likely to have catastrophic effects for life on Earth in the future, there is still a large portion of society who remains unconvinced on the existence of climate change. If this pattern continues, all aspects of human life will be adversely affected. In the second half of the twentieth century, for example, there have been 71 "billion dollar events" resulting from earthquakes that took over 670,000 lives and cost $\$345$ billion to fix [1]. In comparison, the same time period had over 170 extreme climatic events such as windstorms, floods, droughts, and heatwaves that took over 725,000 lives and generated economic losses over $\$700$ billion [2]. Furthermore, earlier snowmelt in western mountains means a longer dry season with widespread effects on the ecologies of plant and animal communities, fire threat, and human water resources. Melting of glaciers is also leading to rising sea levels - a phenomenon which threatens coastal cities around the world. With so much at stake, there is an obvious motivation to research these extreme climatic events to predict and take preventative measures against these events.

While there is no single definition of what constitutes an extreme event, typically extremes can be quantified on the basis of: how rare they are, which involved the notions of frequency of occurrence; or how intense they are, which involves notions of threshold exceedance. Thus, the term "climate extreme" is used to signify individual climate events that are unusual in their occurrence (minimally, the event must lie in the upper or lower tenth percentile of the distribution).

In this paper, our primary focus will be providing a more unified and mathematically justified procedure for detecting change-points within a climatic time series, as they can drastically alter estimates or predictions made from a statistical model. For example, Figure 1 below shows how making predictions based on the overall trend dramatically changes when change-point information is incorporated or neglected. This example considers annually averaged temperatures recorded at New Bedford, Masschusetts, from 1812 to 1999. The figure reports two statistical regression models: 1) a linear line (red) which has a positive rate of change for the overall trend from 1812 to 1999 and 2) 4 local linear lines (blue) from (i) 1812 to 1890, (ii) 1890 to 1902, (iii) 1902 to 1950, and (iv) 1950 to 2000 respectively to account for the four known change-points (1888, 1906, 1951, and 1985) represented by four vertical dashed purple lines. Observe that within each local segment that the rate of change is negative. Therefore, if the local lines (blue) are used we would conclude that 2 for every year that passes the temperature decreases at some constant rate, but if the overall trend line (red) is used we would conclude that the temperature increases, a major contradiction caused without a shadow of doubt by the station relocations during years 1888, 1906, 1950, and 1985 (i.e the change-points in the time series).

Figure 1: Impact of Change-points for a Climate Time Series



The example presented above is typical of climate time series in that they have a time trend and multiple change-points induced by both climatic and nonclimatic changes such as instrumentation changes, station changes, observer changes, etc., that act to influence overall trend inferences. The change-point problem is well known in the climate literature; numerous authors have presented change-point tests for the case of a single mean shift when the series has no trends. A partial list of references for this task includes; [3], [4], [5], and [6].

For reasons described above, our main objective in this study is to develop a scientific method that can accurately estimate the number and the location of change-points within a climate time series. A climate time series is a sequence of climate points (observations) measured at successive points in time (say annual) spaced at uniform time intervals which is described by an underlying distribution. Intuitively, we can think of a change-point as a point in time at which unknown quantities (parameters) of the distribution or model abruptly change. Finding change-points can also equivalently be seen as the subdivision of a series into segments characterized by homogeneous statistical features (e.g. mean and standard deviation). Establishing the existence, and ultimately the number and locations, of such change-points in climatic time series can be a extremely difficult task. Despite the difficultly, the need is essential and therefore has received much attention over the past 40 years as researchers seek irrefutable evidence of climatic change and its link to anthropogenic activities. Besides climate-related changes, other nonclimatic factors such as relocation of weather stations and changes of instrumentation are apt to cause sudden changes and these must be identified to properly analyze climatic time series; see [7].

The majority of existing algorithms for detecting the change-points of a time series are based on detecting changes in the mean or standard deviation of the time series. For shorter time series, the problem may be simplified into detecting a single change point (if it exists at all). The methods for solving these problems can often be generalized to account for cases with multiple change points or an unknown number of change points. [8] specifies a maximum number of discontinuities allowed, then uses both Haar (square) wavelets and a brute force minimization of the residual squared error to determine the placement of piecewise

continuous line segments. [9] automate the creation of a matrix of over-determined linear equations and consecutively solve this system for every possible combination of change points that satisfies their constraints. While techniques like these prove to be accurate in situations where normality assumptions are satisfied, they become very computationally intensive very quickly. [10] tried a dynamic programming method to make this computational burden more manageable, while [11] used Branch and Bound techniques to eliminate sub-optimal segmentations and reduce the number of computations needed.

Another approach is to model the time series as a discrete-time discrete-state Markov process [12]. In this method, we use Markov Chain Monte Carlo methods to sample from the posterior distribution of the change point locations. Our Bayesian approach in this paper falls into this category of strategies. Though all of these techniques are great for detecting changes in mean and variance, they have a tendency to fall short while detecting changes in non-normal distributions. We compare our Bayesian method to an Extreme Value approach in normal and non-normal situations.

2 The Bayesian Approach

2.1 Prior Distributions

We start off with a Bayesian model of detecting change-points. The main idea of this approach is to determine a model that would lead to prior probabilities for the number and location of these change-points, then condition each interval on the previous change-point to derive a posterior distribution via Markov Chain Monte Carlo methods. In this approach, we assume a linear regression model: $y_i = \mu + \beta x_i + \epsilon, \epsilon \sim N(0, \sigma^2), i = 1, 2, ..., n$. We can then define a change point as a point where the parameters change: let $y_{s:t} = y_s, y_{s+1}, ..., y_t$ and $s \leq t^* \leq t$. We say t^* is a change point if $Y_{s:t^*} \sim N(\mu_1, \sigma_1^2)$ and $Y_{t^*:t} \sim N(\mu_2, \sigma_2^2)$, where $\mu_1 \neq \mu_2$ or $\sigma_1 \neq \sigma_2$. Among the x_i of the time series are K change-points $t_1, ..., t_K$. So, our problem boils down to finding the set $\{t_1, ..., t_K\}$ and unknown K. We then define:

$$P(\text{next change point at location t}|\text{change point at location s}) = g(|t - s|)$$

where $l = |t - s|$ corresponds again to the length of the interval between successive change points. This g assumes that this conditional probability depends only on the distance between 2 change points. This seems to be a reasonable assumption, as changes in distributions are certainly easier to detect when the distance between them is larger. If two change points are too close together, then it becomes difficult to determine how many change points are occurring in a small interval.

This g is also a probability mass function for the distance between successive change points s and t. We can also define the cumulative distribution function for distance as

$$G(l) = \sum_{i=1}^{l} g(i)$$

We can also interpret g as the pmf for the position of the first change-point, since $g(t) = g(|t - 0|)$, where s=0 marks the position of the first point of the time series. These functions g and G also imply a prior distributions on the number and location of change points [13].

A reasonable choice of g for climatic time series would be the geometric distribution with probability λ corresponding to the average length of time between change points. We

chose the geometric distribution for a multitude of reasons: each point of our time series is a Bernoulli trial (it either is a change point or it isn't), it assumes independence between each individual point, and it assumes that each point is equally likely to be a change point (an assumption we will have to test, but is reasonable for a starting prior). This geometric distribution implies that the number of change points is distributed BIN(n, λ) while the location of the change points is distributed UNIF$(0, n)$. We also put a prior on λ, as the Beta(a, b) distribution is the conjugate prior of parameter λ.

We next construct the pdf for the number of change-points K. Since $K \sim$ BIN(n, λ), we have the discrete PDF

$$f(K|\lambda) = \left(\begin{array}{c} n \\ K \end{array} \right) \lambda^K (1-\lambda)^{n-K}$$

After K is observed, $L(\lambda) = f(K|\lambda)$ defines a likelihood function $(0 < \lambda < 1)$. We also have the prior distribution on λ: $\lambda \sim$ Beta(a, b). This yields the pdf

$$B(\lambda) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \lambda^{a-1}(1-\lambda)^{b-1}, 0 < \lambda < 1$$

A beautiful property of the Beta distribution is that we have $E[\lambda] = \dfrac{a}{a+b}$ and Var$(\lambda) = \dfrac{E[\lambda](1-E[\lambda])}{a+b+1}$. We can use prior estimates for $E[\lambda]$ and Var(λ) to obtain values for a and b. With the likelihood function and prior distribution, we can finally construct the posterior distribution:

$$p(\lambda|K) = \frac{p(K|\lambda)p(\lambda)}{\int p(K|\lambda)p(\lambda)d\lambda}$$

which is

$$p(\lambda|K) = \frac{\dfrac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \lambda^{a-1}(1-\lambda)^{b-1} \left(\begin{array}{c} n \\ K \end{array} \right) \lambda^K (1-\lambda)^{n-K}}{\int_0^1 \dfrac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \lambda^{a-1}(1-\lambda)^{b-1} \left(\begin{array}{c} n \\ K \end{array} \right) \lambda^K (1-\lambda)^{N-K} d\lambda}$$

$$\implies p(\lambda|K) = \frac{\lambda^{a+K-1}(1-\lambda)^{b+(n-K)-1}}{I}$$

where

$$I = \int_0^1 \lambda^{a+K-1}(1-\lambda)^{b+(n-K)-1} d\lambda = \text{Beta}(a+K, b+n-K) = \frac{\Gamma(a+K)\Gamma(b+n-K)}{\Gamma(a+b+n)}$$

So,

$$p(\lambda|K) = \frac{\Gamma(a+b+n)}{\Gamma(a+K)\Gamma(b+n-K)} \lambda^{a+K-1}(1-\lambda)^{b+n-K-1}$$

In other words, $p(\lambda|K) \propto \lambda^{a+K-1}(1-\lambda)^{b+n-K-1}$. Notice that this is the kernel of Beta$(a+K, b+n-K)$. From this, we can also obtain the posterior mean for λ:

$$E[\lambda|K] = \frac{a+K}{a+b+n} = \frac{n}{a+b+n} \cdot \frac{K}{n} + \frac{a+b}{a+b+n} \cdot \frac{a}{a+b}$$

This has an easy interpretation, as the first product is the data weight times the data estimate, while the second product is the prior weight times the prior estimate.

When we look at the data itself, we are going to assume a linear regression model. I.e. we assume that on some segment $\{s+1, s+2, ..., t\}$ that $Y_{s+1:t} = X\beta + \varepsilon$, where X is a $(t-s) \times p$ matrix of basis functions, β is a $p \times 1$ vector of regression parameters, and ε is a $(t-s) \times 1$ vector of iid normal noise: $\varepsilon \sim N_{(t-s)}(0, \sigma^2 I)$. We assume conjugate priors in this regression framework, so σ^2 follows an Inverse Gamma distribution with parameters $\nu/2$ and $\gamma/2$. The hyperparameter ν represents the strength of the prior: $\nu = 2$ indicates a weak prior and can be increased depending on strength of the prior. Hyperparameter γ represents the expected variance in each segment. We can use the mode $\frac{\gamma}{\nu+2} = s^2$, where s is the observed standard deviation of the segment to find an appropriate choice for γ. The j^{th} component of β, β_j, is distributed Normal with mean 0 and variance $\sigma^2\delta_j^2$ with $\delta_j^2 = (X'_{s:t}X_{s:t})^{-1}_{jj}$. For each iteration, we simplify notation by denoting $Y_{s+1:t}$ as Y and $(t-s)$ as m. This gives us:

$$P(Y|\beta, \sigma^2) = \frac{1}{(2\pi)^{m/2}\sigma^m |I_m|^{1/2}} exp\bigg(-\frac{1}{2\sigma^2}(Y - X\beta)^T I_m^{-1}(Y - X\beta)\bigg) \tag{1}$$

$$P(\beta|D, \sigma^2) = \frac{1}{(2\pi)^{p/2}\sigma^p |D|^{1/2}} exp\bigg(-\frac{1}{2\sigma^2}\beta^T D^{-1}\beta\bigg) \tag{2}$$

$$P(\sigma^2|\nu, \gamma) = \frac{(\gamma/2)^{\nu/2}}{\Gamma(\nu/2)}(\sigma^2)^{-\nu/2-1} exp\bigg(-\frac{\gamma}{2\sigma^2}\bigg) \tag{3}$$

where $D = diag(\delta_1^2, ..., \delta_p^2)$ and I_m is an $m \times m$ identity matrix. From Baye's Theorem, we have

$$\begin{aligned} P(Y) &= P(Y|D, \nu, \gamma) \\ &= \iint P(Y, \beta, \sigma^2|D, \nu, \gamma) d\beta d\sigma^2 \\ &= \iint P(Y|\beta, \sigma^2) P(\beta|D, \sigma^2) P(\sigma^2|\nu, \gamma) d\beta d\sigma^2 \end{aligned} \tag{4}$$

If we multiply (1) and (2), we obtain

$$\begin{aligned} P(Y|\beta, \sigma^2)P(\beta|D, \sigma^2) &\propto exp\bigg(\frac{-1}{2\sigma^2}((Y - X\beta)^T(Y - X\beta) + \beta^T D^{-1}\beta)\bigg) \\ &\propto exp\bigg(\frac{-1}{2\sigma^2}(Y^T Y - 2Y^T X\beta + \beta^T X^T X\beta + \beta^T D^{-1}\beta)\bigg) \end{aligned}$$

From here, we let

$$\begin{aligned} M &= (X^T X + D^{-1})^{-1} \\ P &= (I - XMX^T) \\ ||Y||_p^2 &= Y^T PY \end{aligned}$$

Then the above expression reduces to

$$P(Y|\beta, \sigma^2)P(\beta|D, \sigma^2) \propto exp\bigg(\frac{-1}{2\sigma^2}((\beta - MX^T Y)^T M^{-1}(\beta - MX^T Y) + ||Y||_p^2\bigg)$$

So, the posterior distribution for β is still normal with mean $MH^T Y$ and variance $\sigma^2 M$:

$$P(\beta|D, \sigma^2) = \frac{1}{(2\pi)^{q/2}\sigma^q |M|^{1/2}} exp\bigg(\frac{-1}{2\sigma^2}(\beta - MX^T Y)^T M^{-1}(\beta - MX^T Y)\bigg) \tag{5}$$

Integrating out β, we then obtain

$$\begin{aligned} P(Y|D,\sigma^2) &= \int P(Y|\beta,\sigma^2)P(\beta|D,\sigma^2)d\beta \\ &= \frac{1}{(2\pi)^{m/2}\sigma^m}\frac{(2\pi)^{q/2}\sigma^q|M|^{1/2}}{(2\pi)^{q/2}\sigma^q|D|^{1/2}}exp\bigg(\frac{-1}{2\sigma^2}||Y||_p^2\bigg) \\ &= \frac{1}{(2\pi)^{m/2}\sigma^m}\bigg(\frac{|M|}{|D|}\bigg)^{1/2}exp\bigg(\frac{-1}{2\sigma^2}||Y||_p^2\bigg) \end{aligned}$$

We can next multiply (3) and (6) to obtain

$$P(Y|D,\sigma^2)P(\sigma^2|\nu,\gamma) \propto (\sigma^2)^{-m/2-\nu/2-1}exp\bigg(-\frac{\gamma+||Y||_P^2}{2\sigma^2}\bigg)$$

So, the posterior distribution on σ^2 is still Inverse Gamma with scale parameter $(m+\nu)/2$ and shape parameter $(\gamma+||Y||_P^2)$.

$$P(\sigma^2|\nu,\gamma) = \frac{((\gamma+||Y||_P^2)/2)^{(m+\nu)/2}}{\Gamma((m+\nu)/2)}(\sigma^2)^{-(m+\nu)/2-1}exp\bigg(-\frac{\gamma+||Y||_P^2}{2\sigma^2}\bigg)$$

Finally, we integrate out σ^2 to obtain

$$\begin{aligned} P(Y_{s+1:t}) &= P(Y|D,\nu,\gamma) \\ &= \int P(Y|D,\sigma^2)P(\sigma^2|\nu,\gamma)d\sigma^2 \\ &= \bigg[\frac{1}{(2\pi)^{m/2}\sigma^m}\bigg(\frac{|M|}{|D|}\bigg)^{1/2}\bigg]\bigg[\frac{(\gamma/2)^{\nu/2}}{\Gamma(\nu/2)}\bigg]\bigg[\frac{\Gamma((m+\nu)/2)}{((\gamma+||Y||_P^2)/2)^{(m+\nu)/2}}\bigg] \\ &= \pi^{-m/2}\bigg(\frac{|M|}{|D|}\bigg)^{1/2}\frac{(\gamma)^{\nu/2}}{(\gamma+||Y||_P^2)^{(m+\nu)/2}}\frac{\Gamma((m+\nu)/2)}{\Gamma(\nu/2)} \end{aligned} \tag{6}$$

To make this easier to implement, we perform our calculation in log space:

$$\begin{aligned} log(P(Y_{s+1:t})) = &-\frac{m}{2}log(\pi)+\frac{1}{2}(log|M|-log|D|)+\frac{\nu}{2}log(\gamma) \\ &-\frac{m+\nu}{2}log(\gamma+||Y||_P^2)+log(\Gamma((m+\nu)/2))-log(\Gamma(\nu/2)) \end{aligned} \tag{7}$$

2.2 Developing Estimates

Since our number of change-points K is unknown, we must develop a way of estimating it to make use of our posterior distribution. In this model we define a change-point as a point where the time series changes in mean or variance, so it is natural to try a regression approach to this problem. Between any two change-points we can expect some regression function to hold, i.e. between change points $T_i = C_j$ and $T_{i+1} = C_k$,

$$E[Y_{i:j}] = \beta_0 + \sum_{l=1}^{m}\beta_l X_l$$

So, we can find the probability of the observed data given these models. I.e. we calculate $f(Y_{i:j}) = f(Y_{i:j}|X)$ for each substring in our data.

In coded form, this would be the following nested for loop:

for $i = 1$ to N

for $j = i + d_{min}$ to N

$Y_{i:j}|X \sim \mathrm{N}(X\beta, \sigma^2 I)$

where σ^2 is the residual variance and d_{min} is the minimum distance between two change-points. Using these small regression models, we can create a method for estimating K. First, define $P_k(Y_{1:j}) = P_k(Y_{1:j}|X)$ as the probability of having k change points in the first j data points. So, we obtain

$$P_1(Y_{1:j}) = \sum_{v=1}^{j-1} P(Y_{1:v}) \cdot P(Y_{v+1:j})$$

In other words, we start by finding the probability of one change point in the first j observations by multiplying probabilities of two non-overlapping strings and summing over all possible placements of the change-point.

From here, we can find $P_2(Y_{1:j})$ by multiplying the probability of the first segment having a change-point, $P_1(Y_{1:v}$, by $f(Y_{v+1:j})$ to fill out the rest of the segment. Again, we sum over all possible placements of the second change-point to obtain the total probability of having two change-points in the first j data points.

$$P_2(Y_{1:j}) = \sum_{v=1}^{j-1} P_1(Y_{1:v}) \cdot P(Y_{v+1:j})$$

Continuing this pattern, we obtain

$$P_k(Y_{1:j}) = \sum_{v=1}^{j-1} P_{k-1}(Y_{1:v}) \cdot P(Y_{v+1:j})$$

This pattern can be continued until $k = k_{max}$, the maximum number of change-points allowed. This would be determined by d_{min} described above. Once this process is finished, we can choose the k with the highest probability to be our estimate for K.

To estimate the vector of change points $\mathbf{t}$, we use the following algorithm:

1. For $s = 1, ..., N-1$ and $t = 2, ..., N$, calculate the log-likelihood $log(P(Y_{s:t})$.
2. Create process $Q(s) = P(Y_{s:N}|s-1$ is a change-point). We know that $Q(N) = P(Y_{N:N})$, and for $s < N$,

$$Q(s) = \sum_{t=s}^{N-1} P(Y_{s:t})Q(t+1)g(t-s+1) + P(Y_{s:N})(1 - G(N-s))$$

for g and G the prior distributions on change-points as described above.

3. Compute posterior probability of $t_{k+1}|t_k$ for $k = 1, ..., K$. In this step, we start by computing

$$P(t_{k+1}|t_k, Y_{1:N}) = \frac{P(Y_{t_k+1:t_{k+1}})Q(t_{k+1}+1)g(t_{k+1}-t_k)}{Q(t_k+1)}$$

From here, we simulate $t_{k+1}$from $P(t_{k+1}|t_k, Y_{1:N})$, then increment $k := k + 1$. If $t_k < N$, then we start the algorithm over with the next point in the series. Otherwise, we output the simulated change points $t_1, ..., t_K$.

3 The Extreme Value Approach

Our primary goal here is to develop the statistical theory needed to use extreme values to estimate the number and location of change-points in a climatic time series. Instead of simply observing changes in the mean as in traditional methods, we observe tail values (the extreme temperature events in each year) to find estimates of where scale and shape parameters may change. For example, a year with several record high temperatures and less extreme cold temperatures may indicate a shift of our entire distribution towards higher temperatures. Since extreme temperatures are often easier to observe than changes in mean, we hypothesize that an extreme value method will outperform traditional methods in accuracy.

To accomplish this, we employ a likelihood ratio test used by [14] in which we assume the climatic time series distribution is non-Gaussian with β_t and ξ_t as unknown scale and shape parameters (respectively). From this test, we can test whether β_t and ξ_t change over the entire segment or at each individual point. We will start off using the extreme values as estimates for where change points occur, then can refine this estimate to find spikes in likelihood where the change-points are most likely to be.

We start off with a time series $\{Z_{it} : i = 1, 2, ..., n; t = 1, 2, ..., m\}$, where i is the index for each individual year and t is the index for time (in this case, day of the year). We label our K change-points $t_1, t_2, ..., t_K$, which cuts our time series into $K + 1$ segments. In each of these segments, we assume that $Z_{it} \sim f_j, j = 1, 2, ..., K + 1$. In other words, each segment has its own distribution function that differs from those of adjacent segments (f_{j-1} and f_{j+1}) in right tail. We also define $t_0 = 1$ and $t_{K+1} = n + 1$

3.1 Likelihood Ratio Test

Before going into the likelihood ratio test, we will need a few definitions.

Definition 3.1. *A random process $\{X_t\}$ is **stationary** if it is a stochastic process whose joint probability distribution does not change when shifted in time, i.e. its parameters do not change over time.*

Definition 3.2. *A random process $\{X_t\}$ is **degenerate** if it converges to a single value. In other words, there exists some number t^* and some constant c such that for $i > t^*$, $P(X_i = c) = 1$. A **non-degenerate** process is one that is not degenerate.*

Consider a random process $\{X_t\}$. Assume that the process is stationary and has marginal distribution F with upper end point x^F. [15] showed that if the distribution of excesses $X_t - u$ of a high threshold u ($u < x^F$), scaled as a function of u, converges to a non-degenerate limiting distribution as $u \to x^F$, that distribution must be the General Pareto distribution (see (9) below).

As a result, for sufficiently high thresholds u, the General Pareto distribution will be the most practical family of distributions for statistical estimations of excesses over u.

To observe these excesses, we create the sequence $Z_t = \max(Z_{it}), t = 1, ..., T$. Going along with this, we use the aforementioned likelihood ratio test proposed by [14]. Consider the following tail model:

$$P(Z_{it} > z + u | Z_{it} > u) = \Big(1 + \frac{\xi_t z}{\beta_t}\Big)_+^{-1/\xi_t} \tag{8}$$

where $\beta_t > 0$ and ξ_t are unknown scale and shape parameters (respectively), and $a_+ = max(0, a)$. We can then say that t^* is a change-point if β_t or ξ_t significantly changes before and after t^*. This comes down to the
hypothesis test

$$H_0 : (\beta_1, \xi_1) = ... = (\beta_T, \xi_T)$$

vs

$$H_A : (\beta_1, \xi_1) = ... = (\beta_{t^*}, \xi_{t^*}) \neq (\beta_{t^*+1}, \xi_{t^*+1}) = ... = (\beta_T, \xi_T)$$

for some point t^*. To test this, [16] use the test statistic $Z_n = \sqrt{\max_{1 \leq m \leq n}(-2log \wedge_m)}$, where

$$\wedge_m = \frac{\sup_{\theta,\eta} \prod_{i=1}^{n} f(X_i; \theta, \eta)}{\sup_{\theta,\tau,\eta} \prod_{i=1}^{m} f(X_i; \theta, \eta) \prod_{i=m+1}^{n} f(X_i; \tau, \eta)}$$

In the context of our problem, $\log \wedge_{t^*} = \mathrm{L}(\hat{\beta}, \hat{\xi}) - \mathrm{L}_1(\hat{\beta}_1, \hat{\xi}_1) - \mathrm{L}_2(\hat{\beta}_2, \hat{\xi}_2)$ and our test statistic becomes $Z_n = \sqrt{\max_{t^* \in [\epsilon T, (1-\epsilon)T]}(-2log \wedge_{t^*})}$, where $\mathrm{L}(\hat{\beta}, \hat{\xi})$ is the log-likelihood for (9) based on the positive excesses $\epsilon_{it} = Z_{it} - u$ for all $i = 1, ..., n$ and $t = 1, ..., T$. $\mathrm{L}_1(\hat{\beta}_1, \hat{\xi_1})$ is the log-likelihood function based on the positive excesses of samples up to and including t^*, while $\mathrm{L}_2(\hat{\beta}_2, \hat{\xi}_2)$ is based on the positive excesses of samples after t^*.

From here, we calculate a critical value V from the $(1 - \alpha)^{th}$ quantile of $\sqrt{\max_{\epsilon \leq s \leq 1-\epsilon} \Big[\frac{B_2^2(s)}{s(1-s)}\Big]}$ where $B_2(s)$ is the sum of two Brownian bridges. If our $Z_n < V$, then we fail to reject our null hypothesis - there is not sufficient evidence of a change-point in our segment. In this case, our algorithm is finished. Otherwise we have $Z_n > V$, and we reject H_0 at significance level α. From here, we can estimate our first change-point with $\hat{T}_1 = \mathrm{argmax}_{t \in [\epsilon T, (1-\epsilon)T]}(-2log \wedge_{t^*})$ and obtain two segments. We can then repeat this process on these two segments and subsequently created segments until we fail to reject H_0 for all segments. Note that we do not carry out our hypothesis tests on segments with length less than or equal to ϵT (we can consider these segments to have automatically failed to reject H_0) , and we must be sure to adjust our α by using $\frac{\alpha}{2}$ each time we divide a segment.

4 Simulation Study

In our simulation study, we perform each of the previously described methods on a normally distributed data set, a t-distributed data set, a Beta-distributed data set, and a Pareto-distributed data set. For each of these data sets, we also try cases where the number of change-points in the data ranges from 0 to 3. We will also vary whether these change points result in changes in mean, standard deviation, or other parameters. In the Bayesian case, our data is Y_t where $t = 1, 2, ..., 748$. In the Extreme value case, our data is Z_{it}, where $i = 1, 2, ..., 365$ and $t = 1, 2, ..., 748$. The red vertical lines in each graph correspond to the change points detected by each algorithm.

The simplest and most natural case to consider first is when all of the data is normally distributed. We first begin by setting a change point at $t^* = 400$.

$$Y_t, Z_{it} \sim \begin{cases} N(15, 2) & \text{if t} \leq 400 \\ N(10, 4) & \text{if t} > 400 \end{cases}$$

In the two change point case, we have

$$Y_t, Z_{it} \sim \begin{cases} N(10, 4) & \text{if } t \leq 270 \\ N(15, 4) & \text{if } 270 < t \leq 570 \\ N(15, 2) & \text{if t} > 570 \end{cases}$$

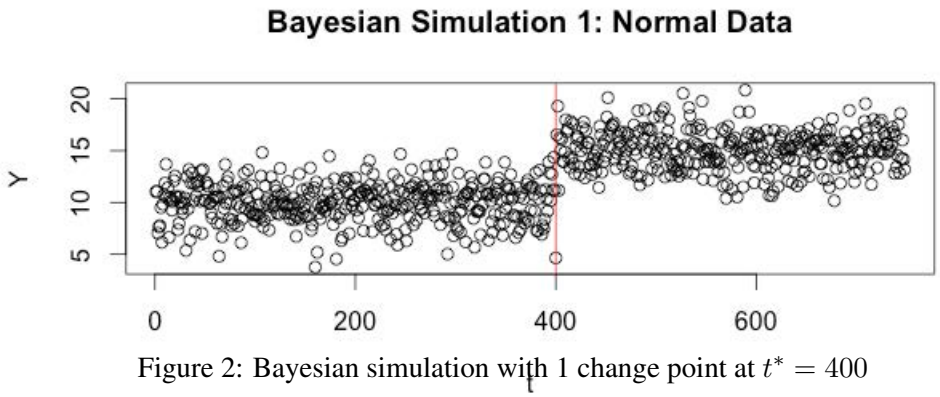


Figure 2: Bayesian simulation with 1 change point at $t^* = 400$

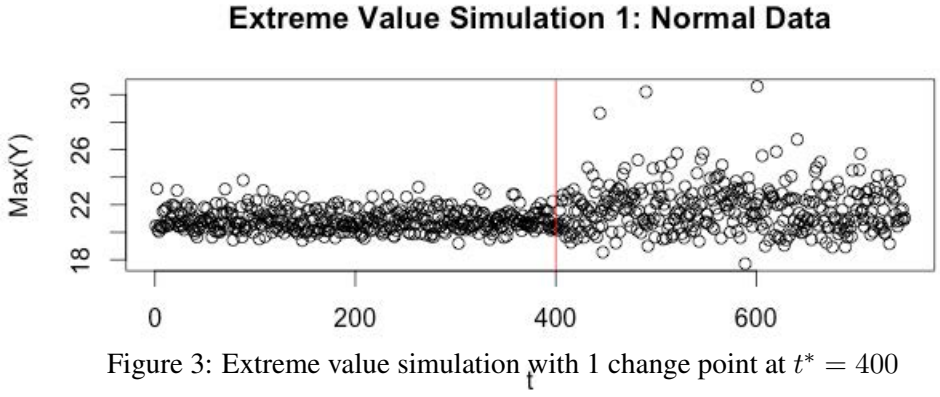


Figure 3: Extreme value simulation with 1 change point at $t^* = 400$

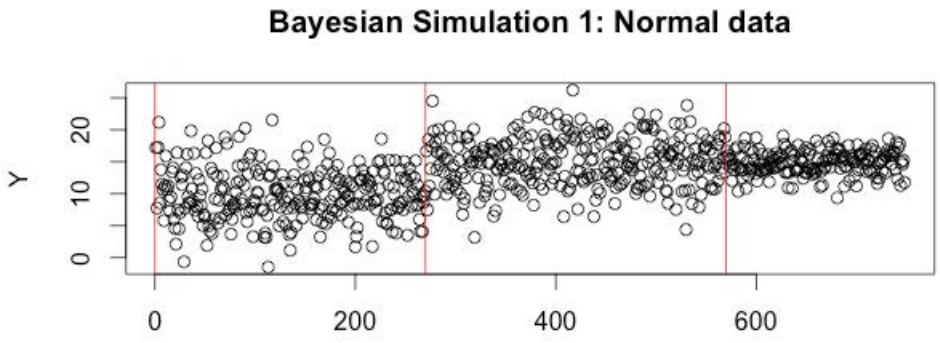


Figure 4: Bayesian simulation with 2 change points at $t^* = 270$ and $t^* = 570$. Posterior probabilities = 1.00

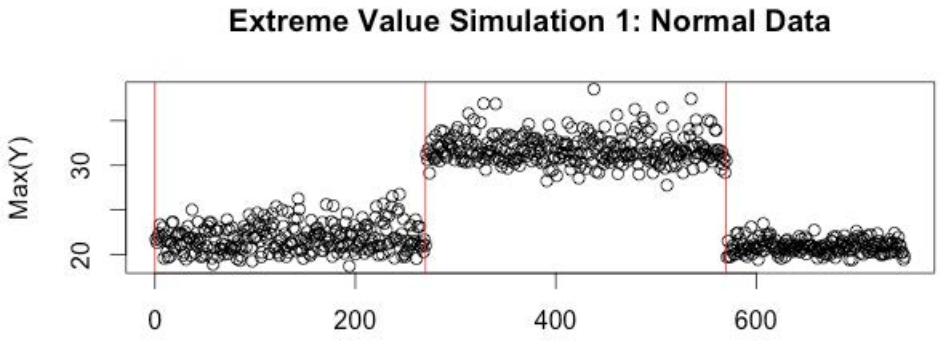


Figure 5: Extreme value simulation with 2 change points at $t^* = 270$ and $t^* = 570$

With this simple case working well for both methods, we next try our hand at some non-normal data. Simulation 2 is run with t-distributed data. Again in our first case we have a single change-point at $t* = 400$.

$$Y_t, Z_{it} \sim \begin{cases} t(30) + 40 & \text{if t} \leq 400 \\ t(40) + 10 & \text{if t} > 400 \end{cases}$$

In the multiple change point case we have

$$Y_t, Z_{it} \sim \begin{cases} t(30) + 40 & \text{if } t \leq 270 \\ t(10) + 20 & \text{if } 270 < t \leq 570 \end{cases}$$

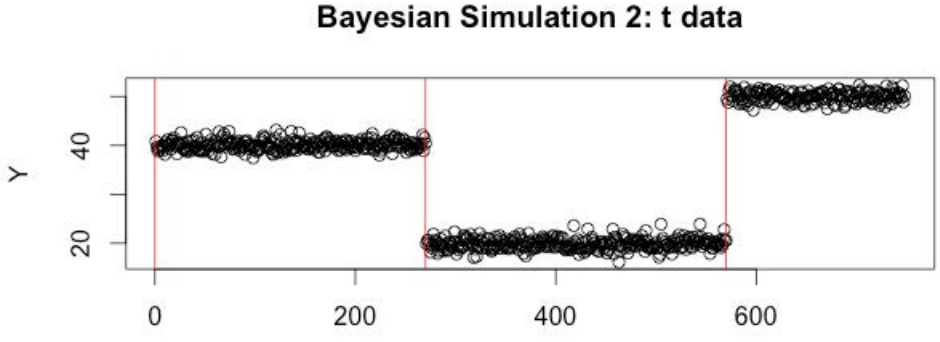


Figure 6: Bayesian simulation with 2 change points at $t^* = 270$ and $t^* = 570$. Posterior probabilities = 1.00

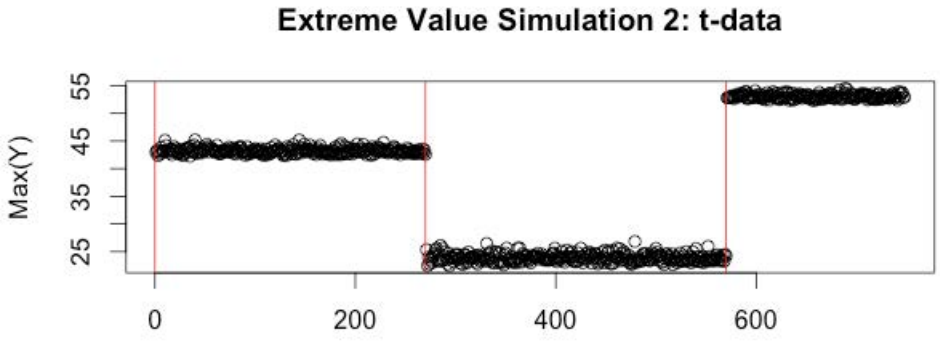


Figure 7: Extreme value simulation with 2 change points at $t^* = 270$ and $t^* = 570$

Finally, our third simulation applies our methods to Pareto-distributed data. With the Pareto distribution being a distribution often used to model extreme events, we predicted that the extreme value approach would be much more valuable in predicting these change points. As before, we used all of these same change-points as in previous simulations. As in the Beta case, the extreme value approach is far more accurate than the Bayesian MCMC method. Once again we attempt multiple cases:

$$Y_t, Z_{it} \sim \begin{cases} Pareto(2,2) & \text{if t} \leq 400 \\ Pareto(5,2) & \text{if t} > 400 \end{cases}$$

In the multiple change point case we have

$$Y_t, Z_{it} \sim \begin{cases} Pareto(2,2) & \text{if } t \leq 270 \\ Pareto(5,2) & \text{if } 270 < t \leq 570 \end{cases}$$

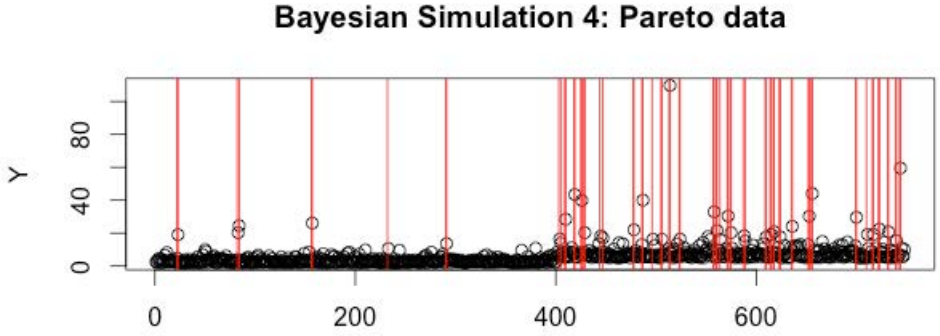


Figure 8: Bayesian simulation with 1 change point at $t^* = 400$. All posterior probabilities = 1.00

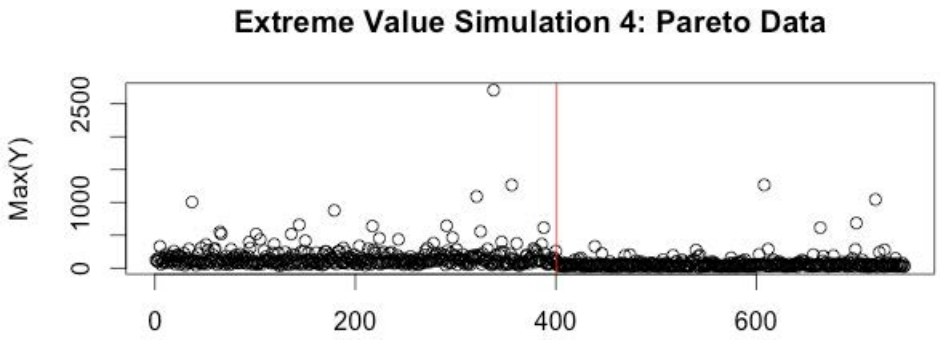


Figure 9: Extreme value simulation with 1 change point at $t^* = 400$

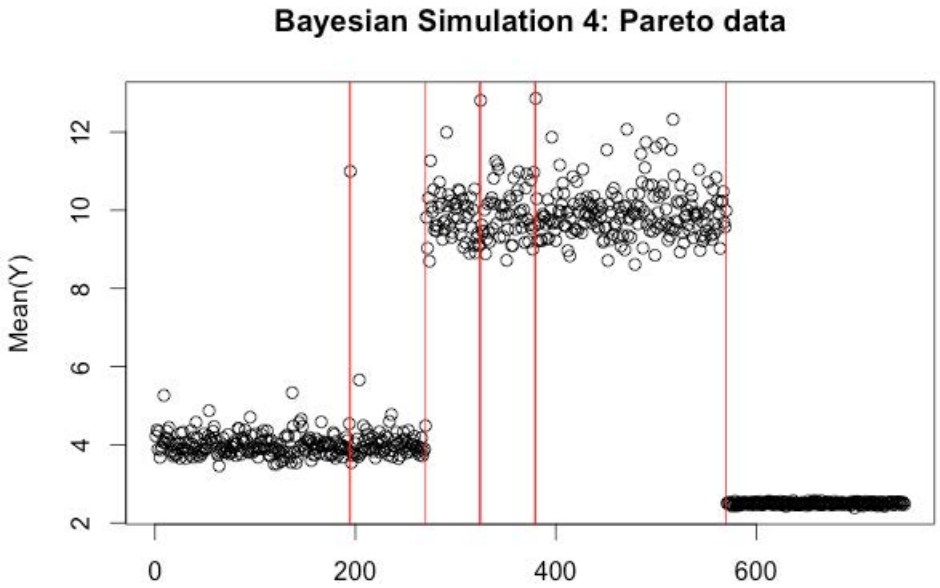


Figure 10: Bayesian simulation with 2 change points at $t^* = 270$ and $t^* = 570$. Posterior probabilities = 1.00

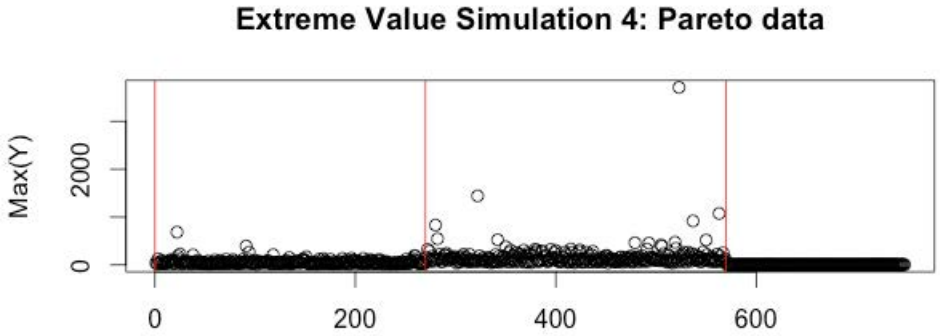


Figure 11: Extreme value simulation with 2 change points at $t^* = 270$ and $t^* = 570$

Our detected change points turn out being very similar in "approximately normal" data, but differed considerably when the data was clearly not normal. In such cases, the extreme value method clearly dominates the traditional Bayesian approach. The extreme value method also turned out to be considerably faster to run than the Bayesian approach.

5 Application to Real Data Set

After testing our methods on synthetic data sets, we turn our attention to a set of climate data from Chula Vista, California. This data set contains the mean temperatures, recorded monthly, of Chula Vista from years 1919 to 1996.

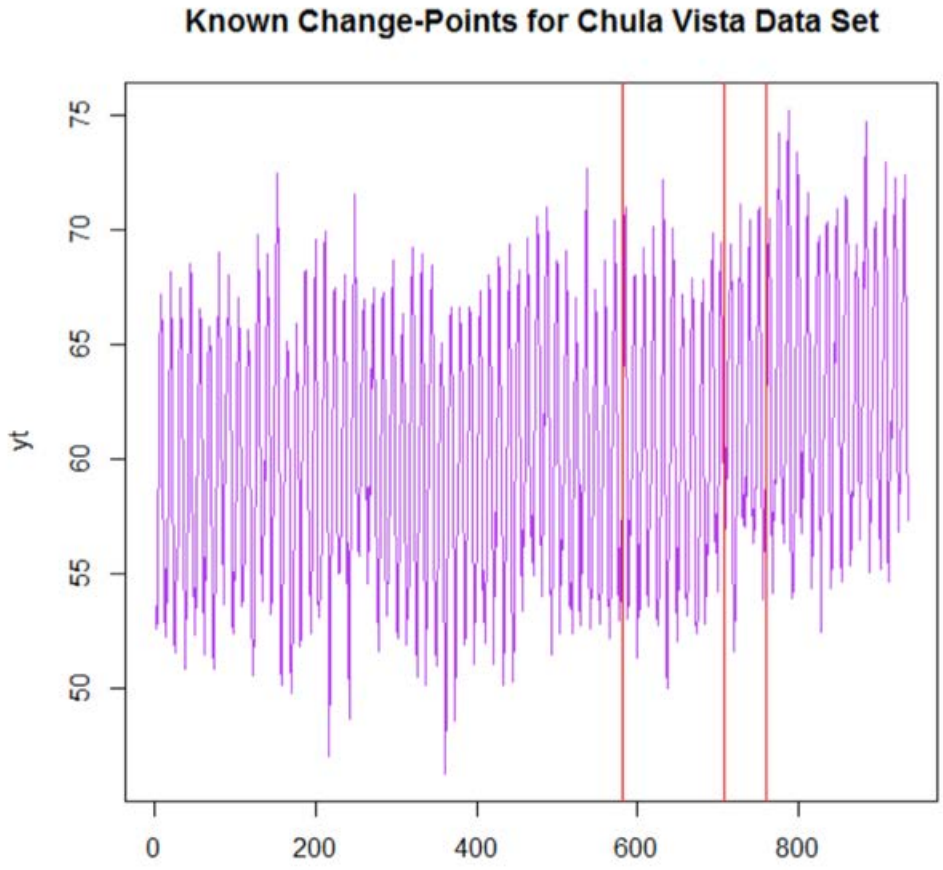


Figure 12: Known change-points lie at $t^* = 582, 708, 760$ which correspond to July 1966, December 1981, and April 1985 respectively.

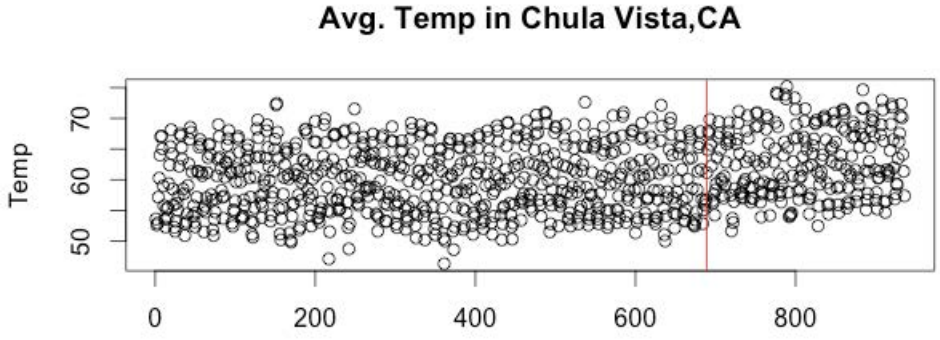


Figure 13: Bayesian simulation with 1 change point at detected at $t^* = 689$. Posterior probability = 0.60

Our Bayesian MCMC method finds one change-point at $t = 689$, which is fairly close to the true change-point at $t = 708$. Indeed, at first glance the graph seems to make its most drastic change in this area. After checking this result with multiple Bayesian change point packages in R, this result appears to be the best that we can do with this approach.

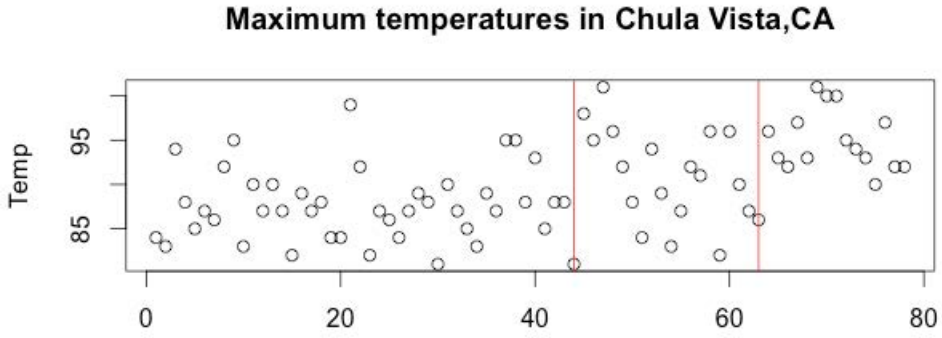


Figure 14: Extreme Value method with change points detected at $t^* = 44$ and $t^* = 63$.

In our extreme value case, we group the data into years and create a sequence out of each year's highest temperatures. Though we have fewer points in this sequence, we pick up an extra change point and some surprisingly nice results. The change-points in this graph lie at $t = 44$ and $t = 63$, which correspond to the years 1962 and 1981. These are close to the true change points, but we lose some accuracy in switching our t values from months to years.

6 Conclusions and Further Research

Overall, we found the extreme value approach to have several advantages over the traditional Bayesian approach. The algorithm produced faster and more accurate results with minimal knowledge of the data's distribution. Possible areas of further research include testing for robustness of this method and optimization of our R code. I am also interested in seeing if the two methods may be combined and experimenting with how various priors may further improve our estimates.

7 Appendix: R Functions

```
#prior distributions
g <- function(k,lambda)
{
  p = ((1-lambda)^(k-1))*lambda
  return(p)
}
G <- function(k,lambda)
{
  p = 1-(1-lambda)^k
  return(p)
}
#Function for computing log likelihoods
llikelihoods <- function(Y, nu = 2, gam = 2){

  for(i in 1:n){
    for(j in 1:i){
      m <- i-j+1
      Yst <- matrix(Y[i:j], nrow=m)
      Hst <- matrix(H[i:j,], nrow=m, ncol=2)
      gam = 4*var(Yst)
      if(is.na(gam)){
        gam=1
      }
      M <- solve(t(Hst)%*%Hst + solve(D))
      P <- diag(rep(1,m)) - Hst%*%M%*%t(Hst)
      Y2p <- t(Yst)%*%P%*%Yst
      L[i,j] <- (-m/2)*log(pi)-.5*(log(det(M))-log(det(D)))+(nu/2)*log(gam)
       - ((m+nu)/2)*log(gam+Y2p)+lgamma((n+nu)/2)-lgamma(nu/2)
    }
  }
  return(t(L))
}
#Backward simulation function
backward <- function(L, kguess=1){

  Q[n] <- min(exp(L[n,n]),1)
  for(s in (n-1):1){
    for(t in s:(n-1)){
      p = min(exp(L[s,t]),1)
      Q[s] <- Q[s]+p*Q[t+1]*g(t-s+1,kguess/n)
    }
    p2 <- min(1,exp(L[s,n]))
    Q[s] <- Q[s] + p2*(1-G(N-s,kguess/n))
  }
```

```
  return(Q)
}
#Forward simulation function
forward <- function(Q,L,kguess=1){
  n <- length(Q)

    #i corresponds with tao_k, j with tao_(k+1)
    #each row of tao corresponds to posterior P(tao_(k+1)=j|tao_k=i)
    for(i in 1:(n-1)){
      for(j in (i+1):n){
        if(j == n){
          p = min(exp(L[i+1,n]),1)
          tao[i,j] <- p*(1-G(n-i,kguess/n))/Q[i+1]
        } else{
          p = min(exp(L[i+1,j]),1)
          tao[i,j] <- p*Q[j]*g(j-i,kguess/n)/Q[i+1]
        }
      }
      t[i] <- which.max(tao[i,])
      if(t[i] >= n){
        break
      }
    }
  return(t)
}

#Extreme Value change point search function
cpsearch = function(Z, alpha){
  n = nrow(Z)
  if(n<3 || alpha < .00625){
    return(matrix(0,1,1))
  }

  #compute likelihood ratios
  for(i in 1:(n-1)){
    Zi = as.matrix(Z[1:i])
    Ziplus1 = as.matrix(Z[(i+1):n])
    L1 = f(Zi, beta1, xi1)
    L2 = f(Ziplus1, beta2, xi2)
    lambda[i] = L - L1 - L2
  }

  for(i in 2:n){
    delta0.hat = delta0.hat + (Z[i] - phi0.hat*Z[i-1])/(n-1)
  }

  for(i in 2:n){
    s2 = s2 + (1/(n-2))*(Z[i] - delta0.hat - phi0.hat*Z[i-1])^2
```

```
    Omega[i] = Z[i] - delta0.hat - phi0.hat*Z[i-1]
  }
  for(i in 2:n){
    OmegaSum[i] = OmegaSum[i-1] + (1/(s2*sqrt(n)))*Omega[i]
  }

  k.hat = which.max(abs(OmegaSum))

  #test statistic from Brownian bridge
  OmegaSumOrd = OmegaSum[order(OmegaSum[,1])]

  for(t in 1:10000){
    s = runif(1,0,1)
    ns = as.integer(n*s)
    Os = Omega[1:ns]
    Bnt = (1/sqrt(s2*n))*sum(Os)
    B[t] = sqrt(((2*Bnt)^2)/(s*(1-s)))
  }
  Bord = B[order(B)]
  V = Bord[critIndex]
  V
  cps = matrix(0,1,1)
  if(is.finite(V) && is.finite(Zn) && V < Zn){
    cp.hat = which.max(abs(lambda))
  } else {
    cp.hat = 0
  }
  cps[1] = cp.hat
  if(cp.hat > 0){
    Z1 = matrix(Z[1:cp.hat])
    Z2 = matrix(Z[(cp.hat+1):n])
    cp1 = cpsearch(Z1, alpha/2)
    cp2 = cpsearch(Z2, alpha/2)
    cps = as.matrix(c(cps, cp1, cp2))
    zeros = which(cps == 0)
    cpsord = cps[order(cps[-zeros])]
    if(length(cpsord) < 1){
      cps = matrix(0,1,1)
    } else{
      cps = cpsord
    }
  }
  return(cps)
}
```

8 Bibliography

References

[1] M. Re, "Topics annual review: natural catastrophes 2002," *Munich. Available at www. munichre. com/pdf/topics*, 2002.

[2] C. Gallagher, R. Lund, and M. Robbins, "Changepoint detection in climate time series with long-term trends," *Journal of Climate*, vol. 26, no. 14, pp. 4994–5006, 2013.

[3] E. Page, "A test for a change in a parameter occurring at an unknown point," *Biometrika*, vol. 42, no. 3/4, pp. 523–527, 1955.

[4] D. M. Hawkins, "Testing a sequence of observations for a shift in location," *Journal of the American Statistical Association*, vol. 72, no. 357, pp. 180–186, 1977.

[5] J.-F. Ducré-Robitaille, L. A. Vincent, and G. Boulet, "Comparison of techniques for detection of discontinuities in temperature series," *International Journal of Climatology*, vol. 23, no. 9, pp. 1087–1101, 2003.

[6] C. Beaulieu, J. Chen, and J. L. Sarmiento, "Change-point analysis as a tool to detect abrupt climate variations," *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 370, no. 1962, pp. 1228–1249, 2012.

[7] F. G. Kuglitsch, A. Toreti, E. Xoplaki, P. M. Della-Marta, J. Luterbacher, and H. Wanner, "Homogenization of daily maximum temperature series in the mediterranean," *Journal of Geophysical Research: Atmospheres*, vol. 114, no. D15, 2009.

[8] D. R. Easterling, G. A. Meehl, C. Parmesan, S. A. Changnon, T. R. Karl, and L. O. Mearns, "Climate extremes: observations, modeling, and impacts," *science*, vol. 289, no. 5487, pp. 2068–2074, 2000.

[9] A. Tomé and P. Miranda, "Piecewise linear fitting and trend changing points of climate parameters," *Geophysical Research Letters*, vol. 31, no. 2, 2004.

[10] E. Ruggieri, T. Herbert, K. T. Lawrence, and C. E. Lawrence, "Change point method for detecting regime shifts in paleoclimatic time series: Application to δ18o time series of the plio-pleistocene," *Paleoceanography*, vol. 24, no. 1, 2009.

[11] H. Aksoy, A. Gedikli, N. E. Unal, and A. Kehagias, "Fast segmentation algorithms for long hydrometeorological time series," *Hydrological Processes*, vol. 22, no. 23, pp. 4600–4608, 2008.

[12] J. Albert and S. Chib, "Bayesian tests and model diagnostics in conditionally independent hierarchical models," *Journal of the American Statistical Association*, vol. 92, no. 439, pp. 916–925, 1997.

[13] X. Xuan, *Bayesian inference on change point problems*. PhD thesis, University of British Columbia, 2007.

[14] G. Dierckx and J. L. Teugels, "Change point analysis of extreme values," *Environmetrics*, vol. 21, no. 7-8, pp. 661–686, 2010.

[15] J. Pickands III, "Statistical inference using extreme order statistics," *the Annals of Statistics*, pp. 119–131, 1975.

[16] M. Csörgö and L. Horváth, *Limit theorems in change-point analysis*, vol. 18. John Wiley & Sons Inc, 1997.

[17] T. L. Lai and H. Xing, "A simple bayesian approach to multiple change-points," *Statistica Sinica*, pp. 539–569, 2011.

[18] A. Mohammad-Djafari and O. Féron, "Bayesian approach to change points detection in time series," *International Journal of Imaging Systems and Technology*, vol. 16, no. 5, pp. 215–221, 2006.

[19] X. Wang and J. W. Emerson, "Bayesian change point analysis of linear models on graphs," *arXiv preprint arXiv:1509.00817*, 2015.

[20] P. J. Brockwell and R. A. Davis, *Introduction to time series and forecasting*. Springer Science & Business Media, 2006.

[21] A. C. Bartlett, *Extreme Value Estimators for Various Non-Negative Time Series with Heavy-tail Innovations*. uga, 2013.

[22] B. Boukai, "An explicit expression for the distribution of the supremum of brownian motion with a change point," *Communications in Statistics-Theory and Methods*, vol. 19, no. 1, pp. 31–40, 1990.

[23] D. Dupuis, Y. Sun, and H. J. Wang, "Detecting change-points in extremes," *Statistics and Its Interface*, 2015.

[24] A. Bartlett and W. McCormick, "Estimation for nonnegative first-order autoregressive processes with an unknown location parameter," *Applied Mathematics*, vol. 3, no. 12, p. 2133, 2012.

[25] A. Bartlett and W. McCormick, "Estimation for non-negative time series with heavy-tail innovations," *Journal of Time Series Analysis*, vol. 34, no. 1, pp. 96–115, 2013.

[26] H. Drees, A. Ferreira, and L. De Haan, "On maximum likelihood estimation of the extreme value index," *Annals of Applied Probability*, pp. 1179–1201, 2004.

[27] G. Dierckx, "Trends and change points in the tail behaviour of a heavy tailed distribution," in *Proceedings of the 58th World Statistics Congress (ISI2011), Dublin*, 2011.

[28] D. J. MacKay, "Introduction to monte carlo methods," in *Learning in graphical models*, pp. 175–204, Springer, 1998.

[29] L. Aksoy, E. Da Costa, P. Flores, and J. Monteiro, "Exact and approximate algorithms for the optimization of area and delay in multiple constant multiplications," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 27, no. 6, pp. 1013–1026, 2008.