Detecting Change-Points using a Bayesian Approach for Climate Data

Andrew Bartlett

Southern Illinois University Edwardsville, Edwardsville, IL 62025

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. Thus, acting as a catalyst for concern about whether or not the climate is actually changing. One challenge when scientifically trying to determine whether or not the climate is actually changing is a change-point. A change-point is defined as any abrupt change or shift in the distribution and is the single most important contributing factor for inaccurate or accurate results. Traditional change-point methods focus exclusively on detecting an alteration or a shift in the arithmetic mean.

In this paper we present a Bayesian change-point detection algorithm for detecting change-points in climate data. We first develop the theory for a Bayesian approach using a hierarchical model to estimate the location and number of change-points within a climatic time series. We then discuss the implementation of our Gibbs Sampler algorithm to obtain posterior probabilities of the location of multiple change-points. We finally investigate the performance of our Bayesian change-point approach through comparison with a standard frequentist method. Both methods are applied to simulated and real temperature data collected from Chula Vista, California.

Key words: Change-Point, Change-Point Analysis, Bayesian framework, Climatic Changes, Bayesian Analysis, Gibbs Sampling

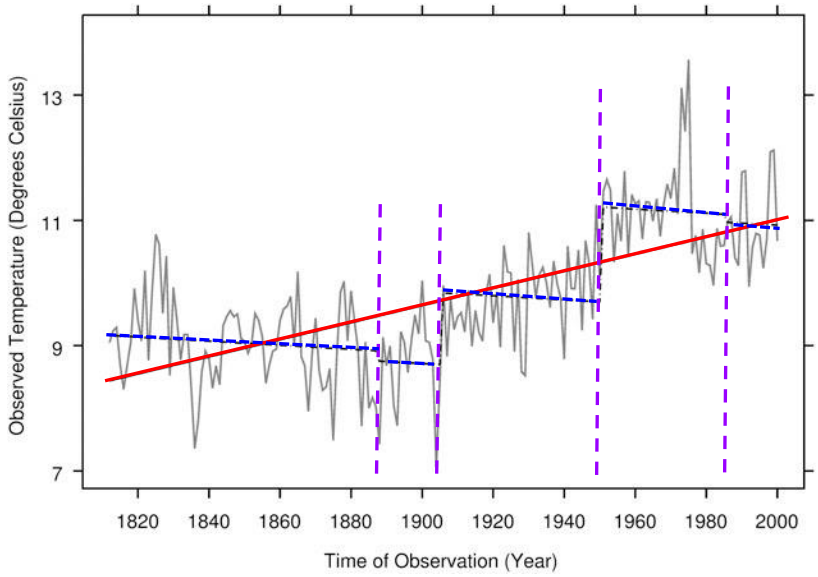
1 Introduction and Background

Recently, countries worldwide have been experiencing unusually hot days and nights and fewer unusually cold days and nights. Heavy downpours have become more frequent and intense. Droughts are becoming more serve. Despite the catastrophic nature of recent weather and climate events, society remains unconvinced that the climate is actually changing and as a result, fails to understand the affect this change will have on the economy, environment, and human well-being in the future.

In this paper, our primary focus is on providing an 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 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



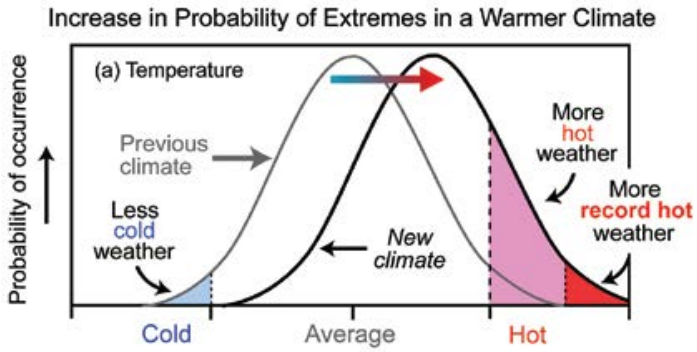
The example presented above is typical of climate time series in that they have a time trend and multiple change-points induced by both climatic and nonclimatic changes such as instrumentation changes, station changes, observer changes, etc., that act to influence overall trend inferences. The change-point 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; Page (1955), Hawkins (1977), Vincent (1998), Ducre-Robitaille et al. (2003), and Beaulieu et al. (2012).

For reasons described above, our main objective is to implement various Bayesian methods 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. For example, imagine a heart rate monitor which displays a line that is flat for the first 3 minutes and then begins to change to a sinusoidal pattern for the next 3 minutes and then all of a sudden

the monitor begins to display an intense up and down motion. The moment when we observed a sudden change in the heart rate would be described as a change-point in which the heart rate distribution changed. 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. For example, climate-related changes, non-climatic 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 Kuglitsch et al, (2009).

Up until this point, the method primarily used for detecting change-points within a climate time series is through an abrupt shift in the mean of the series. Recent work includes: Beaulieu et al (2012), Li and Lund (2012), Lu and Lund (2010), Toreti et al, (2012). However, extreme climatic events are relatively more sensitive to the variability of the climate than to its average and the sensitivity is relatively greater the more extreme the event. This becomes even more transparent in the discussion on climate change as performed by Crisci et al. (2002) and Katz and Brown (1992). The authors conclude that any climate change will first lead to changes in the frequency and intensity of extreme events. Figure 2 illustrates this point graphically. Observe how a change in temperature or climate will first be statistical and visually observed in the tails of the distribution. Statistically speaking, when an observation falls into the tail of a distribution, we refer to this observation as an extreme value. From Figure 2 we can visually see how the chances (probability) of an extreme event occurring, increases when there is a shift in the temperature or climate. This increase in probability is statistical understood by a shift of area from the center of the distribution to the tails. This in turns creates a distribution with fatter tails.

Figure 2: Affect a Change in Extreme Values can have on the Distribution



The rest of the paper is organized as follows: the development of a simple frequentist approach that uses the Goodness of Fit test to determine possible change-points is presented in Section 2.1. Section 2.2 contains a Bayesian approach that uses prior information to estimate the number and location of unknown change-points. Section 2.3 describes our Gibbs sampler algorithm to obtain probabilities from our posterior distribution. Finally, in Section 3 we apply our theory developed in Section 2 to simulated and real temperature data. Both data sets are known to have three change-points. The performance of methods proposed in this paper are discussed and future work is introduced.

2 Methodology

In this section we introduce the theory necessary to implement our proposed Frequentist and Bayesian method to detect change-points in climate data sets.

2.1 Frequentist Approach

When dealing with a climatic or temperature data set, a typical approach to detecting change-points is linear functions. We begin by proposing a simple linear model. The advantage of implementing such a simple model is it does not involve advance statistical techniques so it is easy to understand and interpret the results. In addition, it can provide important insight into necessary modifications to existing methods of detecting change-points. Since this approach is quite standard in the change-point literature it serves as a base line for comparison with our alternative Bayesian method. We now begin by defining an indicator function

$$I_t(\theta) = \begin{cases} 1 & \text{if } t \geq \theta \\ 0 & \text{if } t < \theta, \end{cases} \tag{2.1}$$

where θ is the unknown change-point for $t = 1, \dots n$. The purpose of this indicator function is to partition the time series into two segments and compare the two distributions. This will allow us to determine statistical which time-point is then most likely to be a change-point. As described above, typically the mean of the two distributions are then compared and if there is a significant difference between the two distributions then the time-point at which the two distributions were partition is the change-point. We now consider the following simple regression model:

$$y_t = \beta_0 + t\beta_1 + \beta_2 I_t(\theta) + t\beta_3 I_t(\theta) + \epsilon_t, \tag{2.2}$$

where we assume that $\epsilon_t \sim N(0, \sigma^2)$. Now, in order to determine possible change-points under this parametric model, we must first fix our unknown change-point θ and fit n regression models using the ordinary least-square (LSE) method for parameter estimation. In doing so, we will define the value of θ which maximizes the coefficient of determination R^2 as the change-point in the data set. That is, we will obtain model estimates for $\beta = (\beta_0, \beta_1, \beta_2, \beta_3)$ and then compute the Goodness of Fit (GOF) statistic R^2 at each time point t. The GOF essentially determines how well the data fits the model by measuring the discrepancy between observed and expected values.

Remark: It turns out that this ad-hoc method of searching for the best-fitting change-point has some nice properties. Due to the fact that we are assuming that $\epsilon_t \sim N(0, \sigma^2)$, the time series of the R^2 statistic is proportional to the profile log likelihood for θ. Hence the time point of maximum R^2 is equivalent to the Maximum likelihood estimate (MLE) for θ.

A few limitations with this approach is the fact that it can only detect a single change-point and with all parametric models we have to assume the location of a change-point prior to our analysis. In doing so, this method should not be used to estimate regression coefficients due to the fact that we have ignored the uncertainty about the location of the change-point. We will now introduce our Bayesian approach.

2.2 Bayesian Approach

In the previous section we used a parametric model that assumed the location of a change-point, in this section we will use another parametric model under a Bayesian framework. That is, we will treat the location and the change-point as an unknown parameters to be estimated. This will allow us to obtain its probability distribution which in turn can then be used to determine the probability of a change-point at each time point. To obtain such probabilities we will build an Markov Chain Monte Carlo (MCMC) algorithm which allows us to sample from the conditional distribution of our unknown parameter θ to determine the location of one single change-point. Our first proposed Bayesian method is merely an extension of our frequentist approach presented in Section 2.1. In this section we will use the same simple linear regression model but use a prior distribution for our change-point θ. Thus, we begin with our simple regression model:

$$y_t = \beta_0 + t\beta_1 + \beta_2 I_t(\theta) + t\beta_3 I_t(\theta) + \epsilon_t, \text{ where } \epsilon_t \sim N(0, \sigma^2), \tag{2.3}$$

and

$$I_t(\theta) = \begin{cases} 1 & \text{if } t \geq \theta \\ 0 & \text{if } t < \theta, \end{cases} \tag{2.4}$$

as θ remains our unknown change-point. Using the LSE method we define $\hat{y}_t = X\hat{\boldsymbol{\beta}}$, where $\hat{\boldsymbol{\beta}} = (X'X)^{-1}X'y_t$. However, this time we assume that our time series $\{y_t|\theta\} \sim N(X\boldsymbol{\beta}, \sigma^2)$, where θ follows some prior distribution $p(\theta)$ which will be developed and discussed in section 2.2.1. For interpretation purposes we will work with the precision parameter $\tau = \frac{1}{\sigma^2}$ instead of σ^2. With the following setup we can now construct our Gibbs sampler. We begin by deriving the joint posterior distribution for $(\theta, \boldsymbol{\beta}, \tau)$ and then derive the marginal posterior distributions by conditioning on all other parameters. Using the fact that $\epsilon_t = (y_t - \hat{y}_t) \sim N(0, \sigma^2)$, the likelihood function is

$$p(y_t|\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_t - X\beta)^2}{2\sigma^2}}. \tag{2.5}$$

With this setup, one key advantage lies in the fact that we are evaluating the likelihood function $p(y_t|\theta)$ at discrete points $t = 1, \ldots, n$. Under the frequentist approach, one must condition on θ to get accurate estimates for $\boldsymbol{\beta}$ whereas, a Bayesian framework allows us to account for this uncertainty by simulating draws from a discrete uniform or geometric distribution and then integrate over θ instead of conditioning on it.

2.2.1 Prior Distribution

Another obvious advantage of a Bayesian model is that it allows us to use prior information about certain parameters, which in turn will produce more accurate estimates. To this point we have introduced three parameters that we need to estimate $(\theta, \boldsymbol{\beta}, \tau)$; i.e. the unknown change-point, regression coefficients, and precision respectively. Within a Bayesian framework a very practical prior distribution to use is called the conjugate distribution. This type of distribution was chosen for computational reasons as it guarantees that the conditional posterior distribution will

have a closed form. With this being said, we begin by specifying prior distributions. Observe that the random unknown change-point θ is a discrete time point and thus we decided to assign a discrete uniform distribution

$$p(\theta) = \frac{1}{T-1}, \text{ where } \theta = 1, \dots T. \tag{2.6}$$

Remark: Another widely used prior distribution that we could consider for θ is the geometric distribution.

A common distribution used for the constant variance σ^2 is the inverse gamma distribution. Under a simple transformation $Y = \frac{1}{X}$, where $X \sim IVG(\alpha, \beta)$ it can be shown that $Y \sim gamma(\alpha, \beta^{-1})$. Hence, allowing $p(\sigma^2) \sim IVG\left(c_0, \frac{1}{s_0}\right)$ implies that

$$p(\tau) \sim Gamma(c_0, s_0). \tag{2.7}$$

We then used a non-informative hyper-prior distribution with initial values of $c_0 = .007, s_0 = .007$. Based upon our analysis, (c_0, s_0) are robust as results did not very greatly when values for c_0 and s_0 varied. We now turn our attention to the the prior distribution for the regression parameters $\boldsymbol{\beta}$. Using the fact that the conjugate prior distribution for the normal distribution is the normal distribution, we have

$$p(\boldsymbol{\beta}) \sim N(\boldsymbol{\beta}_0, \boldsymbol{\zeta}_0), \tag{2.8}$$

where we assigned another non-informative distribution with $\boldsymbol{\beta}_0 = 0$ and $\boldsymbol{\zeta}_0 = \left(\begin{smallmatrix} 100 & 0 \\ 0 & 100 \end{smallmatrix}\right)$.

Remark: While not proven in this section, it can be shown that values for $\boldsymbol{\zeta}_0$ larger than 100 are asymptotically the same.

Now that we have specified all prior distribution for our three parameters we are ready to present the joint and marginal posterior distributions.

2.2.2 Joint and Marginal Posterior Distributions

Applying Bayes Theorem for $\boldsymbol{\Theta} = (\theta, \boldsymbol{\beta}, \tau)$ we have

$$\begin{aligned} p(\boldsymbol{\Theta}|y_t) &= \frac{p(\boldsymbol{\Theta}, y_t)}{p(y_t)} \\ &\propto p(y_t|\boldsymbol{\Theta})p(\boldsymbol{\Theta}) \\ &= p(y_t|\boldsymbol{\Theta})p(\theta|\boldsymbol{\beta}, \tau)p(\boldsymbol{\beta}, \tau) \\ &= p(y_t|\theta, \boldsymbol{\beta}, \tau)p(\theta|\boldsymbol{\beta}, \tau)p(\boldsymbol{\beta}|\tau)p(\tau). \end{aligned}$$

Thus, the Joint Posterior Distribution $p(\boldsymbol{\Theta}|y_t) \propto$ is proportional to

$$exp\left(-\frac{\tau \sum_{t=1}^{T}(y_t - \hat{y}_t)^2 - 2\tau s_0}{2}\right) \cdot \left(\zeta_0^{-1}\right)^{T/2} exp\left(-\frac{\sum_{t=1}^{T}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^2}{2\zeta_0}\right) \tau^{(T/2+c_0-1)} \cdot \frac{1}{T-1}.$$

We now derive the marginal distributions for each of the three parameters $(\theta, \boldsymbol{\beta}, \tau)$. We naturally begin with the precision parameter τ. Defining $SSE = \sum_{t=1}^{T} \epsilon_t^2$, where $\epsilon_t = y_t - \hat{y}_t$ and $\hat{y}_t = X\hat{\boldsymbol{\beta}}$, we have

$$\begin{aligned} p(\tau|\theta, \boldsymbol{\beta}, y_t) &= \frac{p(\theta, \boldsymbol{\beta}, \tau, y_t)}{p(\theta, \boldsymbol{\beta}, y_t)} \\ &\propto \tau^{T/2} exp\left(-\frac{\tau SSE}{2}\right) \tau^{c_0 - 1} exp(-\tau/s_0) \\ &= \tau^{T/2 + c_0 - 1} e^{-\tau\left(\frac{SSE}{2} + s_0\right)}. \end{aligned}$$

Thus, $p(\tau|\theta, \boldsymbol{\beta}, y_t) \sim Gamma(T/2 + c_0, SSE/2 + s_0)$. Now using the result proven in (Gelman 2004) for a normal conjugate prior we have

$$p(\boldsymbol{\beta}|\theta, \tau, y_t) = \frac{p(\theta, \boldsymbol{\beta}, \tau, y_t)}{p(\theta, \tau, y_t)} \propto exp\left(-\zeta_1 \frac{(\boldsymbol{\beta} - \zeta_1(\boldsymbol{\beta}_0 \zeta_0 + \tau X'\boldsymbol{y}))^2}{2}\right),$$

Thus, $p(\boldsymbol{\beta}|\theta, \tau, y_t) \sim MVN(\boldsymbol{\beta}_1, \zeta_1)$, where $\boldsymbol{\beta}_1 = \zeta_1(\boldsymbol{\beta}_0 \zeta_0 + \tau X'\boldsymbol{y})$ and $\zeta_1 = (\zeta_0^{-1} + \tau X'X)^{-1}$. Finally, we derive the marginal distribution for our change-point parameter θ. Observe that,

$$p(\theta|\boldsymbol{\beta}, \tau, y_t) \propto exp\left(-\frac{\tau SSE}{2}\right).$$

That is, we draw the change-point from the discrete Likelihood function

$$p(\theta = t|\boldsymbol{\beta}, \tau, \boldsymbol{y}) = \frac{L(\theta = t|\boldsymbol{y})}{\sum_{t=1}^{T} L(\theta = t|\boldsymbol{y})}, \text{ for } t = 1, \ldots, T-1. \tag{2.9}$$

2.3 Gibbs Sampler (MCMC)

In order to calculate probabilities from the posterior distribution of the change-point random variable θ, we construct a Markov Chain Monte Carlo algorithm. In order to implement our Gibbs sampler we used the statistical software R. We used 10,000 iterations for burn-in and 100,000 iterations for the chain in order to guarantee that the chain would converge to a stationary posterior distribution. The convergence was verified by running multiple chains with widely spaced starting values for the change-point θ. It should be noted that if the chain does not converge to a stationary distribution, then the change-point that Gibbs sampler produces will not be accurate. This is due to the fact that the mean of the chain shifts as it moves over the parameter space of our samples from the marginal posterior distributions. The following procedure gives a brief outline of our algorithm.

Step 1: Initial values

We begin by sampling each random variable $\theta, \boldsymbol{\beta}, \tau$ from its prior distribution. For θ we drew an integer between 1 and T from $p(\theta)$, where we set $T = 200$. We initialized τ by taking samples from Gamma(c_0, s_0), where $c_0 = .007$ and $s_0 = .007$. Finally we used the LSE estimates $\hat{\boldsymbol{\beta}}$ for $\boldsymbol{\beta}$ from our linear regression model in (2.2). Based upon the marginal distributions presented in Section 2.2.2 we perform the following updating process, where $*$ denotes an updated value.

Step 2: Sample for τ^*

$\tau^* \sim Gamma(T/2 + c_0, SSE^*/2 + s_0)$, where the initial values $\theta = \theta^*$ and $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}^*$.

Step 3: Sample for $\boldsymbol{\beta}^*$

$$\boldsymbol{\beta}^* \sim MVN(\beta_1, \zeta_1), \text{ where } \beta_1 = \zeta_1(\tau^* X'\boldsymbol{y}) \text{ and } \zeta_1 = (\zeta_0^{-1} + \tau^* X'X)^{-1}.$$

Step 4: Sample for θ^*

$$\theta^* \sim exp\left(-\frac{\tau^* SSE^*}{2}\right), \text{ with draws from } p(\theta|\boldsymbol{\beta}, \tau, \boldsymbol{y}^*) = \frac{L^*(\theta = t|\boldsymbol{y}^*)}{\sum_{t=1}^{T} L^*(\theta = t|\boldsymbol{y}^*)}.$$

In Sections 2.1 and 2.2 we used simplistic algorithms that could only find a single change-point. To find multiple change-points we modify our Gibbs sampler by conditioning on the previous change point. We begin by adding a prior onto the mean of the segment from time point $(i + 1)$ to time point (j). Then the prior distribution is

$$p(\mu_{ij}) \sim N\left(\mu_0, \frac{\sigma_0^2}{j - i}\right), \text{ where } \mu_0 = X\beta_0.$$

Now let p represent the probability of a change-point at time point i and $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_c)$, where c denotes the maximum number of possible change-points. We define a potential change-point as any point where the posterior probability is greater than .10. Within a certain interval, we consider all potential change-points and define a change-point as the time point with the maximum posterior probability. The transition probability in our Markov chain is

$$p(\theta_i) = \frac{p_i}{1 - p_i} = \frac{P(\theta_i = 1|\theta_{j\neq i}, y_t)}{P(\theta_i = 0|\theta_{j\neq i}, y_t)}.$$

That is, at each step in the Markov chain we draw $\theta_{i+1} \sim P(\theta_{i+1} = 1|\theta_i, y_t)$. Therefore, we can compute the odds of a change-point at time point $i + 1$ using the conditional probability of a change-point at time point j. The addition to our joint posterior distribution is

$$\frac{P(\theta_i = 1|\theta_{j\neq i}, y_t)}{P(\theta_i = 0|\theta_{j\neq i}, y_t)} = \frac{\int_0^a p^{(b+2)/2}(1-p)^{(T-\gamma-3)/2}dp}{\int_0^b p^{(\gamma+1)/2}(1-p)^{(T-\gamma-2)/2}dp} \cdot \frac{\int_0^\xi p^{\gamma}(1-p)^{(T-\gamma-1)/2}dp}{\int_0^\xi p^{(\gamma-1)}(1-p)^{(T-\gamma)/2}dp}.$$

3 Results

In this section we apply our Frequentist and Bayesian method for detecting change-points to both simulated and climatic data sets.

3.1 Simulations

To assist the quality of each method we generate data with known change-points positions, with various signal to noise ratio. For each simulation we take 10,000 simulations from the posterior distribution of change-points. In both cases we used a constant function. For case 1: Let $Z_i \sim N(0, .3)$ for $i = 1, \ldots, 100$ such that

$$y_t = \begin{cases} -1 + Z_i & 1 \leq i \leq 25, \\ 1 + Z_i & 26 \leq i \leq 50, \\ 3 + Z_i & 51 \leq i \leq 75, \\ 10 + Z_i & 76 \leq i \leq 100. \end{cases}$$

This gives us quite a strong signal to noise ratio with three obvious change-points at $t = 25, 50$, and 75.

Figure 3: Three known Change-points with strong to noise ratio

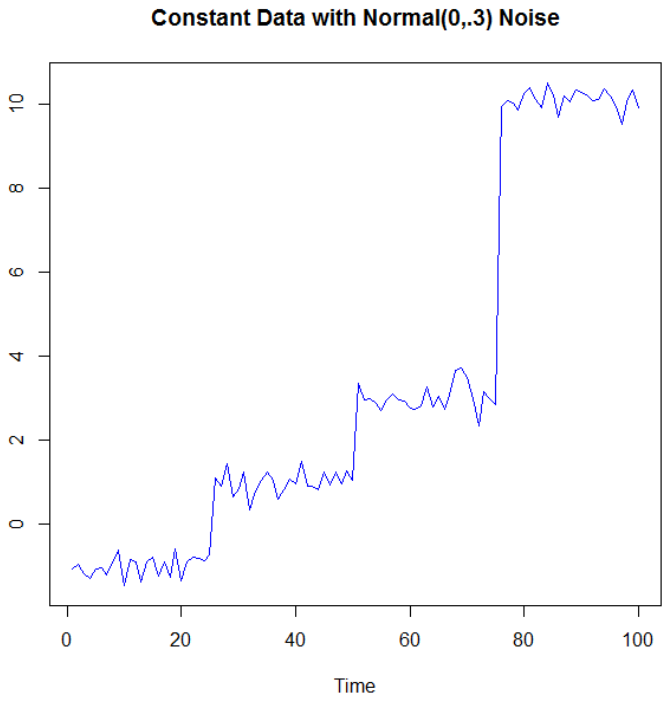
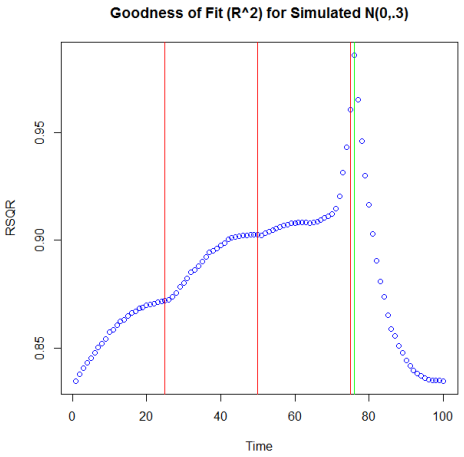
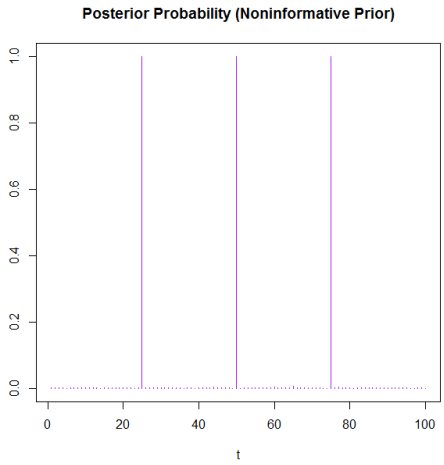


Figure 4: Case 1: $Z_i \sim N(0, .3)$

(a) Frequentist



(b) Bayesian



As expected both methods had no problems of detecting these change-points. In our final case we increased the variance to .8, reducing the signal to noise ratio. Case 2: Let $Z_i \sim N(0, .8)$ for $i = 1, \ldots, 100$ such that

$$y_t = \begin{cases} -1 + Z_i & 1 \leq i \leq 25, \\ 1 + Z_i & 26 \leq i \leq 50, \\ 3 + Z_i & 51 \leq i \leq 75, \\ 10 + Z_i & 76 \leq i \leq 100. \end{cases}$$

Figure 5: Three known Change-points with increased noise

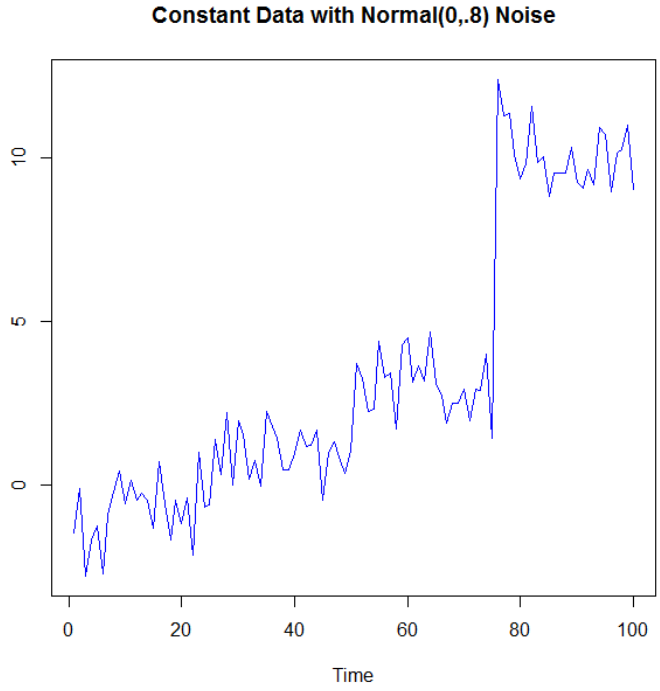
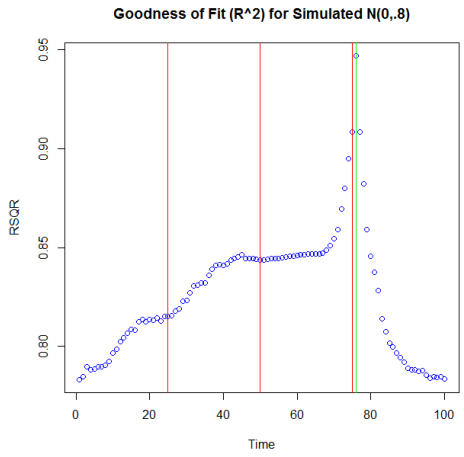


Figure 6: Case 1: $Z_i \sim N(0, .8)$

(a) Frequentist



(b) Bayesian

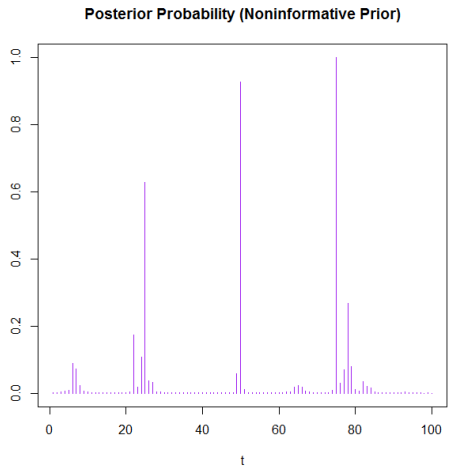
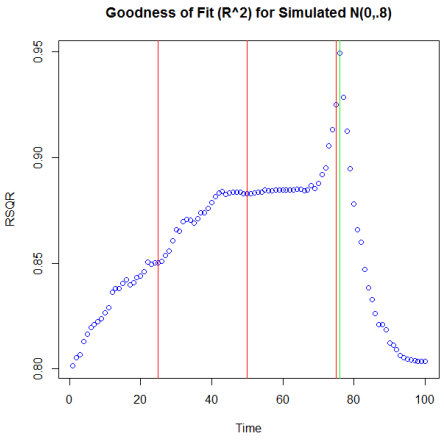


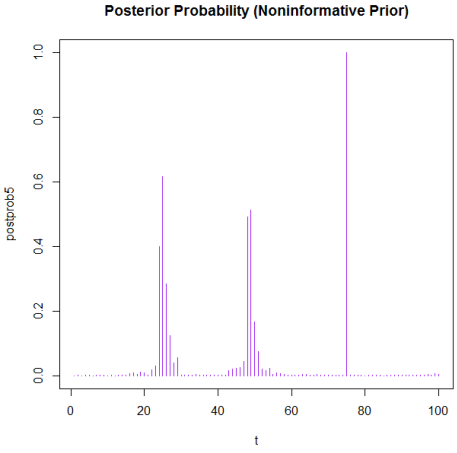
Figure 7 shows that the last change-point ($t = 75$) is found with probability one, as expected. However, our Bayesian method was not as confident with the first change-points, which we expect with the small signal to noise ratio. Interestingly, the second change-point was detected with certainty, despite the step size the same as the second. This is the result of random variability as verified in Figure 8.

Figure 7: Case 2: $Z_i \sim N(0, 1)$

(a) Frequentist



(b) Bayesian

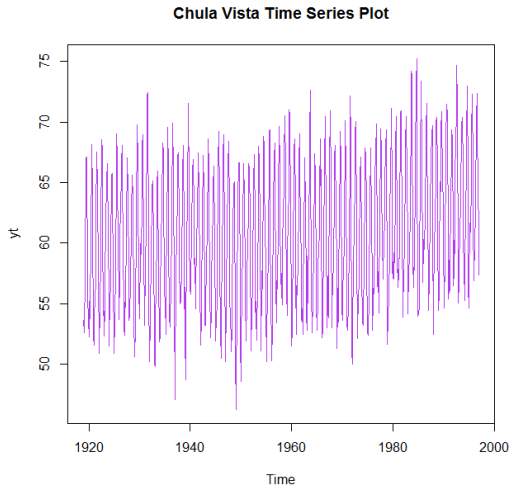


3.2 Chula Vista Data Set

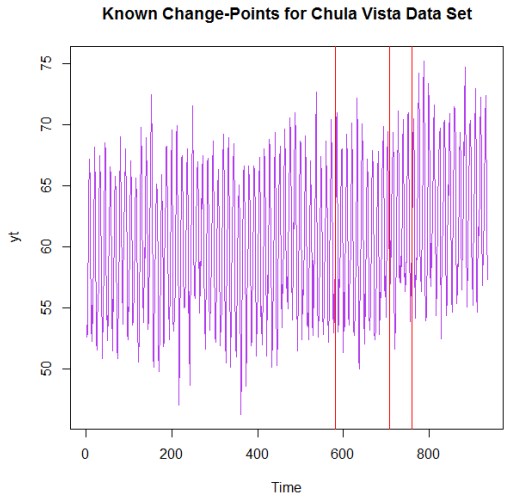
In this section we will apply the theory developed in section 2.1 and 2.2 to a real temperature data set collected from Chula Vista, California. The data set came from the United States Historical Climatology Network (USHCN). In this section we perform our change-point analysis to determine the number and location of possible site changes at the Chula Vista station. This data set contains 936 data points recorded monthly over the years 1919 - 1996. The first observation was in January 1919 and the last observation was December 1996. The upper subplot of Figure 9 displays a time series plot of the data set. The lower subplot of Figure 9 displays the location of the three known change-points represented in red lines provided by (USHCN) located at $t = 582$, (July 1966), $t = 708$, (Dec 1981), and $t = 760$, (April 1985).

Figure 8: Chula Vista Time Data Set

(a) Actual Data Set



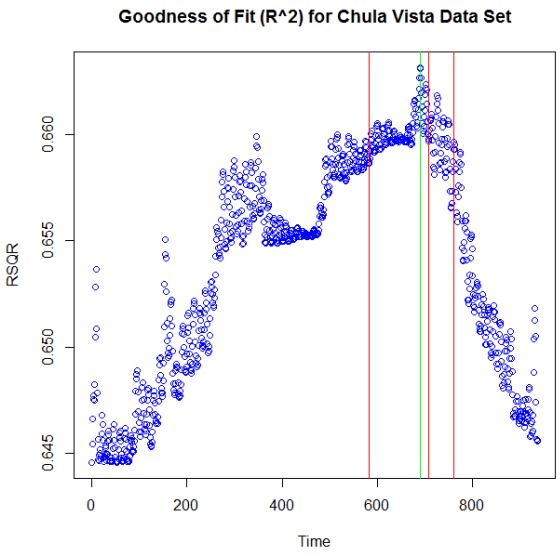
(b) Known Change Points



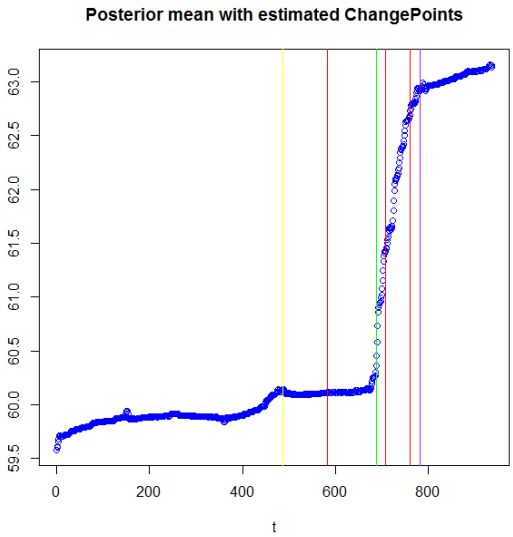
The green line located at $t = 701$ in the upper plot of Figure 10 shows our Frequentist change-point estimate. For such a simple method, it appears to find the most significant change-point with fairly high confidence ($R^2 = .68$). The lower plot of Figure 10 displays the posterior mean. Observe that our Gibbs sampler after running 100,000 iterations estimated three change-points at $t = 482$ (yellow line), $t = 701$ (green), $t = 765$ (purple), whereas the USHCN determined the true change-point to occur at $t = 582$, $t = 708$, and $t = 760$. Estimates for the second and third change-points appear to be very close to the truth.

Figure 9: Chula Vista Change-Point Estimates

(a) Frequentist



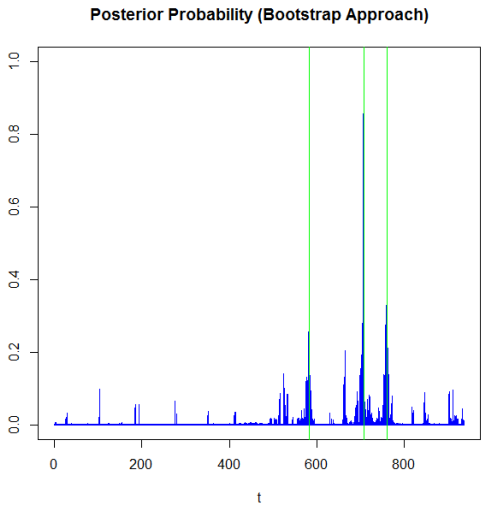
(b) Posterior Mean



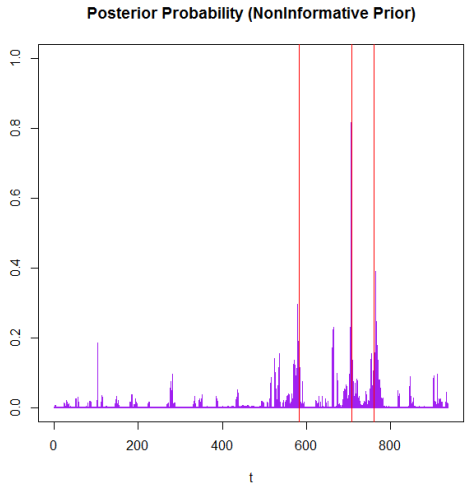
Using the statistical software R we were able to implement our modified Gibbs sampler to compute the posterior probabilities of change-points for the Chula Vista data set using both a non-informative prior and bootstrap approach. Figure 11 displays our results for our bootstrap and non-informative approach. In both cases, we were able to estimate with high accuracy the location of all three change-points. The green vertical lines display USHCN belief for the location of the actual three change-points. Comparing our results with the results in Figure 10 (b) it appears that our Bayesian method was significantly closer to the known change-points. However, a draw back with our method was that many neighboring time points near our estimated change-point had small probabilities, which in turn forced probabilities for our estimated change-points down. This is known as leakage and was corrected by implementing Richen's (2007) peak algorithm to converge neighboring time points to one single change-point.

Figure 10: Location of Change-points for Chula Vista Data Set

(a) Bootstrap



(b) Non-Informative Prior



4 Conclusion

In conclusion, change-points reflect occurrences of important events that can drastically alter estimates or predictions from a statistical model. In this paper, theory for a bayesian method was developed with the ability of detecting multiple change-points for climatic data. We considered both a bootstrap and non-informative prior approach to compute posterior probabilities of a change- point. Difficulties with estimating the location of a change-point with precision was handled by adding neighboring probabilities to the time point with the relative maximum within the interval. Both the simulated data and Chula Vista data set demonstrated that even when there is subtle changes within the distribution that we are able to estimate the location(s) of the change-point(s) with high accuracy. Compared with the standard frequentist approach, multiple advantages in terms of flexibility, sensitivity, accuracy, and certainly where observed in favor of our Bayesian method.

5 References

[1] Beaulieu, C., Chen, J., and Sarmiento, J. L. (2012). Changepoint analysis as a tool to detect abrupt climate variations. *Philosophical Transactions of the Royal Society A - Mathematical, Physical and Engineering Sciences.* **370**, 1228 - 1249.

[2] Crisci A, Gozinni B, Meneguzzo F, Pagliara S, Maracchi G. (2002). Extreme reinfall in a changing climate: regional analysis and hydrological implications in Tuscany. *Hydological Processes*, **16**, 1261 - 1274.

[3] Ducré-Robitaille, J.-F., L. A. Vincent, and G. Boulet, (2003). Com- parison of techniques for detection of discontinuities in temperature series. *Int. J. Climatology.*, **23**, 1087 - 1101.

[4] Gelman A, Carlin J, Stern H, Rubin D. (2004). Bayesian Data Analysis. 2nd Edition. Chapman and Hall/CRC: New York.

[5] Hawkins, D. M., (1977). Testing a sequence of observations for a shift in location. *J. Amer. Stat. Assoc.*, **72**, 180 - 186.

[6] Katz R., Brown B. (1992). Extreme events in a changing climate: Variability is more important than averages. Journal of Climatic Change. 1992. Vol: 21 Issue 3, 289 - 302.

[7] Kuglitsch, F. G., Toreti, A., Xoplaki, E., Della-Marta, P. M., Luterbacher, J. and Wanner, H. (2009). Homogenization of daily maximum temperature series in the Mediterranean. *J. Geophysical Research*, **114**, DOI: 10.1029/2008JD011606.

[8] Li, S. and Lund, R. (2012). Multiple Changepoint Detection via Genetic Algorithms. *J. Climate*, **25**, 674 - 686.

[9] Lund R, Seymour L, Kafadar K. (2001). Temperature trends in the US. Journal of Environmetrics 2001; 12: 673-690.

[10] Lu, Q. Q., Lund, R. and Lee, T. C. M. (2010). An MDL approach to the climate segmentation problem. *Annals of Applied Statistics*, **4**, 299 - 319.

[11] Page, E. S., (1955). A test for a change in a parameter occurring at an unknown point. *Biometrika*, **42**, 523 - 527.

[12] Richens, A (2007). Detecting change points in Time Series Using the Bayesian approach perfect simulation. Journal of American Statistical Association, 87, 109-209.

[13] Toreti, A., Kuglitsch, F. G., Xoplaki, E., and Luterbacher, J. (2012). A novel approach for the detection of inhomogeneities affecting climate time series. *Journal of Applied Meteorology and Climatology*, **51**, 317 - 326.

[14] Vincent, L. A., (1998). A technique for the identification of in- homogeneities in Canadian temperature series. *J. Climate*, **11**, 1094 - 1104.