12-2013

Genetic Algorithm Techniques in Climate Changepoint Problems

Shanghong Li
Clemson University, sli@g.clemson.edu

Genetic Algorithm Techniques in Climate Changepoint Problems

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Mathematical Sciences

---

by
Shanghong Li
December 2013

Accepted by:
Dr. Robert Lund, Committee Chair
Dr. Colin Gallagher
Dr. Peter Kiessler
Dr. Xiaoqian Sun

Abstract

The first part of this dissertation studies genetic algorithms as a means of estimating the number of changepoints and their locations in a climatic time series. Such methods bypass classical subsegmentation algorithms, which sometimes yield suboptimal conclusions. Minimum description length techniques are introduced. These techniques require optimizing an objective function over all possible changepoint numbers and location times. Our general objective functions allow for correlated data, reference station aspects, and/or non-normal marginal distributions, all common features of climate time series. As an exhaustive evaluation of all changepoint configurations is not possible, the optimization is accomplished via a genetic algorithm that random walks through a subset of good models in an intelligent manner. The methods are applied in the analysis of 173 years of annual precipitation measurements from New Bedford, Massachusetts and the North Atlantic Basin's tropical cyclone record.

In the second part, trend estimation techniques are developed for monthly maximum and minimum temperatures observed in the conterminous 48 United States over the last century. While most scientists concur that this region has warmed in aggregate, there is no a priori reason to believe that temporal trends in extremes will have same patterns as trends in average temperatures. Indeed, under minor regularity conditions, the sample partial sum and maximum of stationary time series are asymptotically independent. Climatologists have found that minimum temperatures are warming most rapidly; such an aspect can be investigated via our methods. Here, models with extreme value and changepoint features are used to estimate trend margins and their standard errors. A spatial smoothing is then done to extract general structure. The results show that monthly maximum temperatures are not significantly changing — perhaps surprisingly, in more cases than not, they are cooling. In contrast, the minimum temperatures show significant warming. Overall, the Southeastern United States shows the least warming (even some cooling) and the Western United

States, Northern Midwest, and New England have experienced the most warming.

Dedication

I dedicate this work to my loving family, especially to my upcoming baby. It is their love and support that made this work a complete one.

Acknowledgments

I would like to express my gratitude to many individuals who helped me in many ways during this work.

First and my foremost, I am indebted to my advisor Dr. Robert Lund for his guidance and support. His mathematical insights inspired me and helped me make this a success.

I would also like to thank Drs. X. Sun, and C. Gallagher and P. Kiessler for their insightful suggestions.

Last but not least, I would like to thank the Department of Mathematical Sciences for providing me financial support during my Ph.D studies.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 General Introduction

A changepoint is a time where the structural pattern of a time series first shifts. While we primarily study changepoints that induce mean shifts under a constant variance, changepoints in variances or quantiles can also be of interest. Mean shift changepoints are extremely important features to consider when analyzing climate time series. The shifts identified here can be used to adjust series for non-climatic factors (homogenization) or natural climate fluctuations (see Rodionov 2004). Our focus here is on detecting how many shifts and where they occur rather than their causes. The methods here can incorporate a reference series should one be available.

United States temperature stations, for example, move locations or change gauges or observing techniques an average of six times per century (Mitchell 1953). While it is recognized that changepoint issues are frequently paramount in climate change studies, many multiple changepoint analyses are based on subsegmentation techniques and at most one changepoint (AMOC) methods. By subsegmentation, we mean that the entire series is first analyzed and is judged to be changepoint free or have a single changepoint. If one changepoint is deemed to have occurred, then the series is partitioned into two shorter series about the flagged changepoint time; these two segments are then analyzed for additional changepoints. The process is repeated until no segment is judged to contain additional changepoints.

While there are many variants of the general subsegmentation algorithm (Hawkins 1976 discusses an attractive one), it is usually easy to construct multiple changepoint configurations that

evade detection by any specific subsegmenting algorithm. In particular, subsegmentation algorithms have difficulty identifying two changepoint times that occur close together, especially when the mean shifts induced by the two changepoints take opposite signs as this mimics a "run of outliers". Also, as the subsegmented series length becomes small, the detecting performance of many of the asymptotically tailored AMOC statistical tests degrades. On the other hand, subsegmentation algorithms require only minimal computing resources. An exhaustive multiple changepoint search is often not possible due to the huge number of admissible multiple changepoint configurations. This dissertation proposes an alternative to subsegmentation via genetic algorithms (GAs).

GAs, which are essentially intelligently designed random walk searches, use principles of genetic selection and mutation to drive the search for the best multiple changepoint configuration. GAs allow us to estimate the number of changepoints and their locations with minimal computational demands and without subsegmenting. This dissertation has a tutorial aspect in that GAs have not been widely used in climate research to date (Jann 2006 is an exception), but show potential in many climate optimization problems. We do not seek to overthrow subsegmenting techniques, but rather propose a competing method that gives realistic answers for many climate series. When coupled with minimum description length (MDL) objective function criterion, GAs show vast potential in changepoint research (Davis et al. 2006; Lu et al. 2010).

Our applications here first study annual data; While a detailed simulation study comparing subsegmenting techniques to genetic MDL methods is not presented, we will later compare GA and subsegmenting results in precipitation and tropical cyclone count series. Autocorrelation and reference stations aspects are developed here as these aspects are deemed crucial in making realistic changepoint conclusions.

For other relevant changepoint references, we cite Caussinus and Mestre (2004), Davis et al. (2006), and Lu et al. (2010) for non-Bayesian multiple changepoint techniques. Lund et al. (2007) consider AMOC tests with correlated data; Menne and Williams Jr. (2005, 2009) are good references to learn about reference station aspects. The standard normal homogeneity test used later to subsegment is reviewed in Reeves et al. (2007) among other classical references. Robbins et al. (2011) looks at the North Atlantic tropical cyclone record via a categorical data and subsegmenting approach. Hawkins (1977) considers Gaussian likelihood tests for a single mean shift while Potter (1981) and Buishand (1982) study this and other AMOC tests for precipitation series. Alexandersson and Moberg (1997), Ducré-Robitaille et al. (2003), and Reeves et al. (2007) study and review

AMOC changepoint tests for temperature series. Easterling and Peterson (1995) and Vincent (1998), study the AMOC problem when a trend component is involved; Lund and Reeves (2002) issue a correction to Easterling and Peterson (1995). Chen and Gupta (2000) is a comprehensive statistical changepoint reference.

Daily and/or monthly methods are also worth pursuing. Here, we extend results to monthly extreme data. Extreme temperatures have profound societal, ecological, and economic impacts. It is known that average temperatures in the contiguous United States since 1900 have warmed in aggregate, with the West, Northern Midwest, and New England showing the most warming and the Southeast showing little change (Lund et al. 2001). In fact, a linear trend estimate for the continental United States series of Menne et al. (2010), which aggregates over a thousand stations in the region on a day-by-day basis since 1895, is about 1.26°F per century. This trend applies to mean temperatures.

It is less clear whether minimum and/or maximum temperatures have changed during this period. In fact, maxima and averages are statistically independent in large samples. Specifically, if $\{X_t\}$ is a stationary time series, then $\sum_{t=1}^{n} X_t$ and $\max\{X_1, \ldots, X_n\}$, with n denoting the sample size, are asymptotically independent under minor regularity conditions (McCormick and Qi, 2000). The implication is that inferences involving first moment properties (such as a trend) and those from higher order statistics (such as extremes) need not necessarily exhibit the same patterns. Katz and Brown (1992) effectively argue that extremes are better attributed to variances than means. Mathematically, the limit theory of extremes is described solely by tail properties of the cumulative distribution function (Leadbetter et al. 1983; Coles 2001).

This dissertation seeks to quantify trend estimation procedures in monthly extreme temperature series and apply them to the United States' record. Specifically, monthly maximum series from 923 stations and monthly minimum series from 932 stations located in the conterminous 48 United States are examined. A monthly extreme high temperature for June at a station, for example, is the largest daily high temperature recorded from June 1 through June 30.

Other authors have looked at extreme temperature changes. For example, DeGaetano and Allen (2002) and DeGaetano et al. (2002) fix temperature thresholds and examine trends in the frequency of exceedances of these thresholds. This gives good rudimentary guidance, but does not incorporate the magnitudes of the extreme exceedances. Van de Vyver (2012), perhaps the most methodologically related study to ours, quantifies changes via extreme value distribution peaks over

threshold methods; however, gauge and station relocation effects are not considered there and this study's scope is for Belgium only.

Two prominent issues that we tackle below involve periodicities and changepoints. Periodic features naturally arise with monthly data; winter extremes are cooler and more variable than summer extremes. Changepoints here refer to mean shift structures that are induced by station location moves and temperature gauge changes. Moving a station can shift temperatures by several degrees. United States stations average six changepoints a century (Mitchell, 1953). About 60% of these changepoints are undocumented in station meta-data records and roughly half of the documented changepoint times do not impart mean shifts. As Lu and Lund (2007) show, the changepoint configuration of a station is the single most important piece of information in constructing an accurate trend estimate for that station. DeGaetano et al. (2002) recognize the importance of homogenizing extreme data for changepoint effects. Homogenized data is also useful in other climate studies.

Another issue common to all temperature trend analysis is worth addressing up front. First, this study examines linear trends only. While true temperature changes are surely non-linear in time, linear trends describe average changes over the period of record and provide good rudimentary guidance. As will be seen below, once multiple changepoint features are taken into account, the situation becomes complicated, linearly-based or not.

1.2 MDL Methods

The minimum description length (MDL) principle was developed by Rissanen (1989, 2007) for model selection problems. Deriving from the coding and information theories, the intuition behind MDL methods is that minimum codelength models are also good statistical models. MDL methods defines the best fitting model as the one with minimal codelength. Efficient compression of the data and good statistical models for the data are basically equivalently tasks.

When implementing MDL methods, an objective function or a penalized likelihood score will be optimized. While the objective function needs to be tailored to individual situations, all objective functions here will minimize MDL scores of the form

$$\text{MDL} = -\log_2(L_{opt}) + P. \tag{1.1}$$

L_{opt} is an optimized model likelihood, P is a penalty term that accounts for the number and

type of model parameters, and $\log_2$ indicates logarithm base 2. The more parameters the model has, the higher P becomes. As one adds more parameters to the model, the fit becomes better and $-\log_2(L_{opt})$ becomes smaller; however, if adding additional parameters does not decrease $-\log_2(L_{opt})$ more than the increased penalty for these extra parameters, the simpler model is preferred. Penalized likelihood methods are ubiquitous in modern statistical model selection problems (Rissanen 1989; Lee 2001; Davis et al. 2006) and include Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC).

MDL methods are penalized likelihood methods where the penalty is based on minimum description length information principles. Description lengths quantify how much computer storage space a model requires. Good models are stored with minimal space.

While formal MDL theory is technical and is rooted in a field called information theory and stochastic complexity (see Rissanen 1989, 2007; Hansen and Yu 2001), MDL methods essentially penalize integer-valued parameters, such as the number of changepoints or a changepoint location time, more heavily than a real-valued parameter such as a series mean or variance. This differs from AIC and BIC penalties, which are based on the total number of model parameters. Recently, MDL methods have proven useful in multiple changepoint detection (Davis et al. 2006; Lu et al. 2010).

The penalty term P is where the nuances of MDL methods are important. There are several principles needed to devise an appropriate penalty. First, if a real-valued parameter is estimated from k data points (values of the series), the penalty for it is $\log_2(k)/2$ (Davis et al. 2006).The second MDL penalty principle involves how much integer-valued parameters should be charged. From Davis et al. (2006), the penalty for an unbounded integer I is $\log_2(I)$. The penalty for an integer parameter I that is known to be bounded by an integer B is $\log_2(B)$ (the bound B should be taken as small as possible). The final principle of an MDL penalty is that of additivity: an end penalty is obtained by adding penalties for all model parameters.

1.3 Genetic Algorithm

A genetic algorithm (GA) is a stochastic search that can be applied to a variety of combinatorial optimization problems [Goldberg (1989), Davis (1991)]. The basic principles of GAs were first developed rigourously by Holland (1975) and are analogies of natural behavior, mimicking the genetic process of natural selection/evolution.

In nature, individuals in a population compete with each other for resources to survive. Those individuals which are most successful in surviving and fitting the environment will have relatively more of offspring. On contrast, poorly performing individuals will produce less or even no offspring at all and will eventually die out. This process implies that the genes from the highly "fit" individuals will continue to survive and evolve after successive generation. In this way, species evolve to become more and more adaptable to the environment.

Initial Population Generation: GAs start with an initial population of individuals. Each individual has a parameter configuration (chromosome) that is evaluated to determine the fitness score with respect to the objective function. Individuals with higher scores (highly "fit") are more likely to be selected as parents to produce (crossover) offsprings (children). Normally the initial population is generated randomly with a relative large sample size.

Crossover: Pairs of parent chromosomes, representing mother and father, are probabilistically combined to form a child chromosome. Members that are more fit are more likely to have children, thus mimicking natural selection principles.

Mutation: Mutation aspects are applied to each child individual after crossover. They randomly alter each gene with a small probability. As time increases, the GA evolves to contain "highly fit" individuals; mutations help ensure that the algorithm is not falsely deceived into premature convergence at a "local minimum".

New Generation: The steady-state replacement method with a duplication check as suggested by Davis (1991) is applied here to form new generations. Should a currently simulated child duplicate a previously simulated child in the current generation, the simulated child is discarded. The overall fitness of the population tends to increase with increasing generation since the fittest members of the current generation are more likely to breed.

Termination and Convergence: Generations are successively simulated until a termination condition has been reached. The solution to the optimization is deemed to be the fittest member of any simulated generation. Common terminating conditions are that 1) a solution is found that satisfies minimum criteria, 2) a fixed number of generations is reached, and/or 3) the generation's fittest ranking member is peaking (successive iterations no longer produce better results).

If the GAs has been correctly implemented and the parameters have been properly set, the population will evolve over successive generations so that the fitness of the best individual in each generation increases towards the global optimum. However, we have to admit that GAs cannot

identify a global optimum in all cases. The traditional view is that crossover is the most important aspect for rapidly exploring a parameter space. Mutation helps ensure that no point in the parameter space has a zero probability of being examined. Without mutation, the GA could evolve toward a "suboptimal colony". Such a colony might represent a local (rather than global) optimum of the objective function. Mutation effectively induces a random walk through the parameter space, while the other aspects serve to tune solutions in the "current vicinity of the algorithm".

There are many types/variants of GAs that may have better convergence features under different circumstance. Some, for example, involve multiple islands and immigration from island to island, where each island is itself a separate GA simulation aimed at optimizing the objective function. Holland (1975), Goldberg (1989), Davis (1991), Beasley et al. (1993), and Alba and Troya (1999) are computer science references that discuss standard GAs and their variants.

Chapter 2

Multiple Changepoint Detection via Genetic Algorithms

This chapter proceeds as follows. Section 2.1 develops the likelihood and penalty terms in the objective function. Section 2.2 then devises a GA that is capable of finding the best changepoint configuration (model) among all possible models. Section 2.3 presents a short simulation study on series whose statistical properties are known. The chapter closes with application to two data sets. First, Section 2.4 examines a 173 year series of annual precipitation from New Bedford, Massachusetts. This series is examined with and without a reference series and autocorrelation aspects. Second, Section 2.5 turns to a more controversial issue: the North Atlantic Basin tropical cyclone counts. There, we find a changepoint circa 1995 that is not easily explained by changes in observing techniques.

2.1 Objective Function Development

2.1.1 Annual Precipitation Series

As a first task, we develop the likelihood part of (1.1) for an annual precipitation series. In modeling annual precipitation data, lognormal distributions are worthy of consideration (Wilks 2006). The lognormal distribution has the probability density function

$$f(x) = \frac{\exp\left\{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}\right\}}{x\sigma\sqrt{2\pi}}; \qquad x \geq 0,$$

where μ and σ are location and scale parameters, respectively. If the data $X_1, \ldots, X_N$ are independent in time, the likelihood is simply the product of densities:

$$L(\mu, \sigma^2) = \prod_{t=1}^{N} f(X_t) = \frac{\exp\{-\frac{1}{2\sigma^2}\sum_{t=1}^{N}(\ln(X_t)-\mu)^2\}}{(\sigma\sqrt{2\pi})^N \left(\prod_{t=1}^{N} X_t\right)}. \tag{2.1}$$

noindent To compute L_{opt}, we must find the values of μ and σ^2, in terms of the observed data $X_1, \ldots, X_N$, that maximize $L(\mu, \sigma^2)$. Taking partial derivatives in (2.1) and setting the resulting expressions to zero, we obtain likelihood estimates of μ and σ^2:

$$\hat{\mu} = \frac{1}{N}\sum_{t=1}^{N}\ln(X_t), \qquad \hat{\sigma}^2 = \frac{1}{N}\sum_{t=1}^{N}(\ln(X_t)-\hat{\mu})^2. \tag{2.2}$$

The calculus computations are tedious but straightforward; the reader is referred to Casella and Berger (2002) for likelihood basics. Plugging the optimizing values of $\hat{\mu}$ and $\hat{\sigma}^2$ into (2.1) gives the optimal likelihood score

$$-\ln(L_{opt}) = -\ln(L(\hat{\mu}, \hat{\sigma}^2)) = \frac{N}{2}\left[1 + \ln(2\pi) + \ln(\hat{\sigma}^2)\right] + \sum_{t=1}^{N}\ln(X_t). \tag{2.3}$$

We now modify the above scenario for changepoints. A reasonable model might allow the location parameter μ to shift at each changepoint time. The scale parameter σ is held constant across different regimes. Such a scheme is equivalent to allowing a mean shift at each changepoint time. For a fixed number of changepoints, say m, occurring at the times $\tau_1 < \tau_2 < \ldots < \tau_m$, let $r(t)$ denote the regime number at which the time t data point is sampled from and let R_ℓ denote the set of all times in which regime ℓ held for $\ell = 1, 2, \ldots, m+1$. For example, if $N = 100, m = 1$, and $\tau_1 = 73$, then there are two regimes and $r(t) = 1$ when $t \in \{1, \ldots, 72\} = R_1$ and $r(t) = 2$ when $t \in \{73, \ldots, 100\} = R_2$. Then the marginal density of X_t is

$$f(x) = \frac{\exp\left\{-\frac{(\ln(x)-\mu_{r(t)})^2}{2\sigma^2}\right\}}{x\sigma\sqrt{2\pi}}, \qquad x > 0,$$

and

$$L(\mu_1, \ldots, \mu_m, \sigma^2) = \prod_{t=1}^{N} f(X_t) = \frac{\exp\{-\frac{1}{2\sigma^2}\sum_{t=1}^{N}(\ln(X_t) - \mu_{r(t)})^2\}}{(\sigma\sqrt{2\pi})^N \left(\prod_{t=1}^{N} X_t\right)}.$$

A derivation similar to that producing (2.2) gives

$$\hat{\mu}_\ell = \frac{1}{\tau_i - \tau_{i-1}} \sum_{t \in R_\ell} \ln(X_t), \qquad \hat{\sigma}^2 = \frac{1}{N}\sum_{t=1}^{N}(\ln(X_t) - \hat{\mu}_{r(t)})^2.$$

In piecewise notation, one can write the location parameters of the series as

$$\mu_{r(t)} = \begin{cases} \mu_1, & 1 \le t < \tau_1 \\ \mu_2, & \tau_1 \le t < \tau_2 \\ \quad \vdots & \quad \vdots \\ \mu_{m+1}, & \tau_m \le t < N+1 \end{cases}.$$

The optimal likelihood achieved is exactly as listed in (2.3) but the value of $\hat{\sigma}^2$ changes in form as it now involves an average of squared deviations about a piecewise mean.

The connection between MDL methods and likelihoods is the following result from information theory: the amount of information (also called code-length in the statistics literature) it takes to store the fitted model, given the model form, is $-\log_2(L_{opt})$. This gives us L_{opt}.

We now develop the penalty term P. This is where the nuances of MDL methods are important. There are several principles needed to devise an appropriate penalty. First, if a real-valued parameter is estimated from k data points (values of the series), the penalty for it is $\log_2(k)/2$ (Davis et al. 2006). For example, μ_ℓ, the location parameter for the ℓth regime, should be charged the penalty $\log_2(\tau_\ell - \tau_{\ell-1})/2$ as it is estimated from the $\tau_\ell - \tau_{\ell-1}$ data points in the ℓth regime. The boundary conventions $\tau_0 = 1$ and $\tau_{m+1} = N + 1$ are made for the first and last regimes. The parameter σ^2, which is estimated from all N data points, incurs a penalty of $\log_2(N)/2$.

The second MDL penalty principle involves how much integer-valued parameters such as m and $\tau_1, \ldots, \tau_m$ should be charged. From Davis et al. (2006), the penalty for an unbounded integer I is $\log_2(I)$. The changepoint count parameter m is only bounded by N (essentially unbounded); hence, we charge it a $\log_2(m)$ penalty. The penalty for an integer parameter I that is known to be bounded by an integer B is $\log_2(B)$ (the bound B should be taken as small as possible). The changepoint times $\tau_1, \ldots, \tau_m$ are parameters of this genre. Since $\tau_i < \tau_{i+1}$, τ_i is charged a $\log_2(\tau_{i+1})$

penalty.

The final principle of an MDL penalty is that of additivity: an end penalty is obtained by adding penalties for all model parameters. In the above scheme,

$$P = \frac{3\log_2(N)}{2} + \sum_{i=1}^{m+1} \frac{\log_2(\tau_i - \tau_{i-1})}{2} + \log_2(m) + \sum_{i=2}^{m} \log_2(\tau_i), \tag{2.4}$$

where the terms between the plus signs, from left to right, correspond to σ^2, the regime location parameters $\mu_1, \ldots, \mu_{m+1}$, the number of changepoints, and the changepoint time parameters $\tau_1, \ldots, \tau_m$. An objective function is now obtained from (3.8) as

$$\text{MDL} = \frac{N}{2}\ln(\hat{\sigma}^2) + \sum_{i=1}^{m+1} \frac{\ln(\tau_i - \tau_{i-1})}{2} + \ln(m) + \sum_{i=2}^{m} \ln(\tau_i). \tag{2.5}$$

We have made some simplifications in obtaining (2.5) that make it differ from the direct sum of (2.3) and (2.4). First, we have changed all base 2 logarithms to natural logarithms; this does not change where the minimum occurs since conversion of logarithm bases simply entails multiplying by a positive constant. Second, quantities that are constant in N or the data $X_1, \ldots, X_N$ will not effect where the minimum occurs and are discarded.

We now modify the above analysis to accommodate reference series and autocorrelation aspects. Changepoint detection in temperature series is greatly aided by the use of reference series (Mitchell 1953, Vincent 1998, Caussinus and Mestre 2004, Menne and Williams Jr. 2005, 2009). Lund et al. (2007) show that it is important to account for autocorrelations in changepoint detection techniques. In fact, the positive autocorrelations found in some climate series can induce features that resemble mean shifts. It is easy to erroneously conclude that a changepoint exists in positively correlated series.

Suppose that a reference series $Y_1, \ldots, Y_N$ is available to help identify changepoints in the "target series" $X_1, \ldots, X_N$. The log-normal distribution model simply asserts that the logarithm of each annual precipitation is normally distributed. In fact, if X_t and Y_t are independent and lognormally distributed, then $\ln(X_t) - \ln(Y_t) = \ln(X_t/Y_t)$ is normally distributed. Hence, it seems reasonable to model the logarithm of the precipitation ratios as a Gaussian series, allowing for mean shifts in the log-ratio at each changepoint time.

A model for an annual precipitation series $\{X_t\}$ that allows for a reference series $\{Y_t\}$ and autocorrelation can be devised as follows. Let $S_t = \ln(X_t/Y_t)$. Modeling $\{S_t\}$ as a correlated

Gaussian series requires that we quantify its autocovariance structure. For flexibility and computational simplicity, we will work with a simple first order autoregression (AR(1)) with first-order autocorrelation ϕ and white noise variance σ^2. Such a model satisfies

$$S_t = \mu_{r(t)} + \epsilon_t, \quad \epsilon_t = \phi\epsilon_{t-1} + Z_t,$$

where $\{Z_t\}$ is zero-mean white noise with variance σ^2.

The likelihood of this model, allowing for a mean shift at each changepoint time, is

$$\mathrm{L}(\mu_1, \mu_2, \ldots, \mu_{m+1}, \phi, \sigma^2) = (2\pi)^{-N/2} \left(\prod_{t=1}^{N} v_t \right)^{-1/2} \exp\left[-\frac{1}{2} \sum_{t=1}^{N} \frac{(S_t - \hat{S}_t)^2}{v_t} \right]. \tag{2.6}$$

Here, $\hat{S}_t$ is the best linear prediction of S_t from an intercept and the history $S_1, \ldots, S_{t-1}$, and $v_t = E[(S_t - \hat{S}_t)^2]$ is its mean squared prediction error. The AR(1) dynamics give

$$\hat{S}_t = \mu_{r(t)} + \phi[S_{t-1} - \mu_{r(t-1)}]$$

for $t \geq 2$ with the start-up condition $\hat{S}_1 = \mu_1$. The prediction errors are $v_t = \sigma^2$ for $t \geq 2$ with the start up condition $v_1 = \sigma^2/(1-\phi^2)$. While optimizing this likelihood is more complex with AR(1) autocorrelation than without, it is still not overly difficult. Methods with general pth order autoregressive correlation are possible (see Brockwell and Davis 1991 and Lu et al. 2010) and are similar to (2.6). Likelihood estimators of the mean for the ℓth segment are

$$\hat{\mu}_\ell = \frac{1}{\tau_\ell - \tau_{\ell-1}} \sum_{t \in R_\ell}^{N} S_t.$$

This estimator is asymptotically adjusted for edge-effects. An exact likelihood would need to be computed numerically for each and every changepoint configuration — an arduous task. The variance parameter σ^2 and autocorrelation parameter ϕ are estimated from all data points via

$$\hat{\sigma}^2 = \frac{\sum_{t=1}^{N}(S_t - \hat{S}_t)^2}{N}, \quad \hat{\phi} = \frac{\sum_{t=2}^{N}(S_t - \hat{\mu}_{r(t)})(S_{t-1} - \hat{\mu}_{r(t-1)})}{\sum_{t=2}^{N}(S_{t-1} - \hat{\mu}_{r(t-1)})^2}.$$

Plugging these values into (2.6) gives

$$-\ln(L_{opt}) = -\ln(L(\hat{\mu}_1, \ldots, \hat{\mu}_{m+1}, \hat{\sigma}^2, \hat{\phi})) = \frac{N}{2}\left[1 + \ln(2\pi) + \ln(\hat{\sigma}^2)\right].$$

The penalty for the model parameters is formulated via the same reasoning as before:

$$P = \log_2(m) + \sum_{i=2}^{m} \log_2(\tau_i) + 2\log_2(N) + \sum_{i=1}^{m+1} \frac{\log_2(\tau_i - \tau_{i-1})}{2}.$$

Changing all base 2 logarithms to natural logarithms and ignoring terms that are constants in N or the observations gives our MDL:

$$\text{MDL} = \frac{N}{2}\ln(\hat{\sigma}^2) + \sum_{i=1}^{m+1} \frac{\ln(\tau_i - \tau_{i-1})}{2} + \ln(m) + \sum_{i=2}^{m} \ln(\tau_i). \tag{2.7}$$

2.1.2 Tropical Cyclone Counts

As another example, we develop an objective function for annual tropical cyclone counts. Many authors use Poisson marginal distributions to describe cyclone counts (Mooley 1981, Thompson and Guttorp 1986, Solow 1989, Robbins et al. 2011). The Poisson probability function with parameter $\lambda > 0$ is $f(k) = e^{-\lambda}\lambda^k/k!$ at the integer $k \geq 0$. The mean of this distribution is λ. Allowing the mean parameter to shift at each of the m changepoint times $\tau_1 < \cdots < \tau_m$ produces the likelihood

$$L(\lambda_1, \ldots, \lambda_{m+1}) = \prod_{t=1}^{N} f(X_t) = \prod_{t=1}^{N} \frac{e^{-\lambda_{r(t)}}\lambda_{r(t)}^{X_t}}{X_t!} = \frac{\prod_{\ell=1}^{m+1} e^{-\lambda_\ell(\tau_\ell - \tau_{\ell-1})}\lambda_\ell^{\sum_{t\in R_\ell} X_t}}{\prod_{t=1}^{N} X_t!}.$$

The parameter estimates optimizing this likelihood are simply the segment means:

$$\hat{\lambda}_\ell = \frac{1}{\tau_\ell - \tau_{\ell-1}} \sum_{t\in R_\ell} X_t.$$

Plugging these values back into the likelihood gives

$$-\ln(L_{opt}) = -\ln(L(\hat{\lambda})) = -\sum_{\ell=1}^{m+1} \left(\ln(\hat{\lambda}_\ell) \sum_{t\in R_\ell} X_t \right) + \sum_{t=1}^{N} X_t + \sum_{t=1}^{N} \ln(X_t!).$$

It is easy to construct an MDL penalty here: the changepoint count parameter m is charged

a $\log_2(m)$ penalty and each changepoint time τ_i is charged a $\log_2(\tau_{i+1})$ penalty. Hence,

$$P = \log_2(m) + \sum_{i=2}^{m} \log_2(\tau_i) + \log_2(N) + \sum_{i=1}^{m+1} \frac{\log_2(\tau_i - \tau_{i-1})}{2}.$$

Converting to natural logarithms and ignoring terms that are constant in the sample size N, we arrive at an MDL of form

$$\text{MDL}(m, \tau_1, \ldots, \tau_m) = -\sum_{\ell=1}^{m+1} \left(\ln(\hat{\lambda}_\ell) \sum_{t \in R_\ell} X_t \right) + \sum_{i=1}^{m+1} \frac{\ln(\tau_i - \tau_{i-1})}{2} + \sum_{i=2}^{m} \ln(\tau_i) + \ln(m). \quad (2.8)$$

2.2 Genetic Algorithm Development

Our next task is to determine the optimal model; that is, the one that minimizes the MDL score. In statistical settings, this is termed a model selection problem. Our goal is to find the value of m and the changepoint locations $\tau_1, \ldots, \tau_m$ that minimize the MDL score. A naive approach would exhaustively evaluate the MDL at all possible values for m and $\tau_1, \ldots, \tau_m$. In a series of length N, there are $\binom{N}{m}$ distinct changepoint configurations with m changepoints. Summing this over $m = 0, 1, \ldots, N$ shows that an exhaustive search requires evaluation of 2^N different MDL scores. When $N = 173$, as in our application in Section 2.4, this amounts to evaluating 1.2×10^{52} different MDL scores, a daunting task on even the fastest computer. This is where GAs will prove useful.

GAs search for the optimizing values of m and $\tau_1, \ldots, \tau_m$ without evaluating the MDL score at every possible parameter configuration. They do this by taking an intelligent random walk through the space of admissible models that avoids evaluating MDL scores at models that are unlikely to be optimal. GAs are so-named because they contain aspects of genetic selection/evolution. In particular, each possible parameter configuration will be encoded as a "chromosome". GAs also allow for notions of generations. Two members in a generation are allowed to produce children. Specifically, the chromosome sets of mother and father are probabilistically combined to form a child chromosome. Members that are more fit (that is, they better optimize the objective function) are more likely to have children, thus mimicing natural selection principles. Mutations — cases where the children do not resemble either parent — are occasionally allowed. As time increases, the GA evolves to contain "highly fit" individuals; mutations help ensure that the algorithm is not falsely deceived into premature convergence at a "local minimum".

A GA to optimize (2.5) can be devised as follows. Each parameter configuration is expressed as a chromosome of the form $(m, \tau_1, \ldots, \tau_m)$. Chromosomes for 200 individuals (this generation size parameter can also be varied) in an initial generation were first simulated at random: each year is allowed to be a changepoint time, independent of all other changepoint times, with probability 0.06. The colony size of the initial generation is not overly important and can be varied if desired. This means that the number of changepoints in each initial generation chromosome has a binomial distribution with $N - 1$ trials. It is not necessary to get the changepoint probability accurate here; we use 0.06 to roughly correspond to average changepoint numbers quoted in Mitchell (1953) for US temperature stations.

Children of the first generation are made by combining the fitter individuals of the initial generation. Specifically, two parents (mother and father) are selected by sampling pairs of chromosomes in the initial generation via a linear ranking and selection method. That is, a selection probability is assigned to an individual that is proportional to the individual's rank in optimizing the objective function. The least fit individual is assigned the rank 1 and the most fit individual is assigned the rank N. Suppose that S_i is the rank of the ith individual in the initial population. First, a mother is selected from individuals 1 through 200, the ith chromosome being selected with probability

$$\frac{S_i}{\sum_{j=1}^{200} S_j}. \tag{2.9}$$

Once a mother is selected, the father is probabilistically selected by ranking the remaining 199 individuals akin to (2.9). Note that a mother and father are not allowed to be identical (the exact same chromosome).

Suppose that $(m, \tau_1, \ldots, \tau_m)$ and $(j, \eta_1, \ldots, \eta_j)$ represent the mother's and father's chromosomes, respectively. A child's chromosome is first set to $(m + j, \delta_1, \ldots, \delta_{m+j})$, where the $m + j$ changepoint times $\delta_1, \ldots, \delta_{m+j}$ contain all changepoints of *both* mother and father. The length of the child's chromosome may be shorter than $m + j$ by the number of changepoint times common to both mother and father. Next, we thin the changepoint times of the child, retaining each with an independent coin flip that has heads probability of 1/2. In this manner, the child keeps traits of both parents, but may not exactly duplicate either. For example, suppose that $N = 7$ and the two parent chromosomes are $(1, 6)$ and $(2, 3, 5)$. Then the child chromosome is first set to $(3, 3, 5, 6)$. A fair coin

is then flipped three times. If this sequence had resulted in tails, heads, and heads, the second and third changepoint times are retained and the child chromosome is set to $(2, 5, 6)$. We then allow some random changing of the location of the chromosomes. Specifically, for each changepoint in the child, we roll a three-sided die — say with outcomes -1, 0, and 1 and with respective probabilities $0.3, 0.4$ and, 0.3. If the coin flip is -1, we move the location of the changepoint downward by one unit; if it is $+1$, we move the changepoint location up one unit; if it is 0, we keep the changepoint position as is. Duplicate changepoint times are discarded as are changepoint moves to times 0 or $N + 1$. The above methods produce one child that we call child 1.

Children 2 through 200 are generated in the same manner. Should a currently simulated child duplicate a previously simulated child in this generation, the current child is discarded and we begin anew with the selection of "fresh parents". The 200 simulated children are viewed as the first generation. This process is repeated to obtain future generations. The overall fitness of the population tends to increase with increasing generation since the fittest members of the current generation are more likely to breed. However, without mutation, the GA could evolve toward a "suboptimal colony". Such a colony might represent a local (rather than global) optimum of the objective function.

Mutation ensures that the GA will sometimes explore chromosomes unlike those in the current generation and acts to inhibit premature convergence. Mutations keep the diversity of the population large and prevent convergence to suboptimal colonies. Our mutation mechanism allows a small portion of generated children to have extra changepoints. Specifically, after each child is formed from its parents, each and every non-changepoint time is independently allowed to become a changepoint time with probability p_m. Typically, p_m is small. Mutation effectively induces a random walk through the parameter space, while the other aspects serve to tune solutions in the "current vicinity of the algorithm".

Generations are successively simulated until a termination condition has been reached. The solution to the optimization is deemed to be the fittest member of any simulated generation. Common terminating conditions are that 1) a solution is found that satisfies minimum criteria, 2) a fixed number of generations is reached, and/or 3) the generation's fittest ranking member is peaking (successive iterations no longer produce better results).

There are many types/variants of GAs. Some, for example, involve multiple islands and immigration from island to island, where each island is itself a separate GA simulation aimed at op-

Table 2.1: Empirical proportions of estimated changepoint numbers I.

m	Percent
0	99.0 %
1	0.4 %
2	0.5 %
3+	0.1 %

timizing the objective function. Holland (1975), Goldberg (1989), Davis (1991), Beasley et al. (1993), and Alba and Troya (1999) are computer science references that discuss standard GAs and their variants.

2.3 A Simulation Study

To study the effectiveness of the proposed methods, we offer a short simulation study.

Our first simulation set is designed as a control run. Here, one thousand series of length $N = 200$ were simulated with no changepoints. The simulation parameters were selected to mimic the New Bedford data in Section 2.4. In particular, a log-normal setup was considered with the μ parameter set to 6.8 at all times. The autocorrelation parameter chosen was $\phi = 0.2$ (this represents slightly less correlation than the New Bedford series displays when gauged against a reference series; Section 2.4 elaborates) and the AR(1) white noise variance selected was $\sigma^2 = 0.025$. As there are no changepoints, the true value of m is zero.

Table 2.1 displays the proportion of simulations which yielded various estimated values of m. The genetic algorithm has correctly estimated the series to have no changepoints in 990 of the 1000 runs (99.0%). In four of the simulations, one changepoint was estimated. Overall, the algorithm seems to have a very low false alarm rate. As only ten of the runs estimated changepoints, the location of the estimated changepoint times is of little concern.

Our second simulation set retains the above parameter choices except that three mean shifts are added to every simulated series. We place the mean shifts uniformly in time. Specifically, μ_t is set to 6.8 from times 1 to 49, rises to 7.0 from times 50 through 99, increases to 7.2 at times 100 through 149, and increases again to 7.4 for times 150 through 200. This configuration represents mean shifts in one direction (increasing), all having the same shift magnitudes (on the log scale). Table 2.2 shows empirical proportions of estimates of m. The methods estimate the correct

Table 2.2: Empirical proportions of estimated changepoint numbers II.

m	Percent
0	0.0 %
1	3.6 %
2	28.8 %
3	63.1 %
4	4.3 %
5+	0.2 %

Table 2.3: Empirical proportions of estimated changepoint numbers III.

m	Percent
0	0.0 %
1	6.0 %
2	19.5 %
3	69.2 %
4	5.1 %
5+	0.2 %

changepoint order 63.1% of the time, which is quite admirable. The methods favor underestimation of the changepoint numbers as 28.8% of runs estimate two changepoints while only 4.3% of runs estimate four changepoints. As for the times at which the changepoints are estimated, Figure 2.1 shows a count histogram for the 1000 runs. Elaborating, if a changepoint is estimated at time t in any simulation, the count scale is increased at time t by unity. For example, approximately 300 of the simulations flag the time 50 changepoint exactly at time 50. Fig. 2.1 reveals little bias: the times of the detected changepoints cluster about their true values in a symmetric fashion. Also, the three mean shifts appear "equally easy" to detect.

Simulation set III is akin to Simulation set II except that the changepoint times have been moved and the shift magnitudes are altered. We start with a series whose μ_t is 6.8 from times 1

Table 2.4: Empirical proportions of estimated changepoint numbers IV.

m	Percent
0	0.0 %
1	7.5 %
2	90.7 %
3+	1.8 %

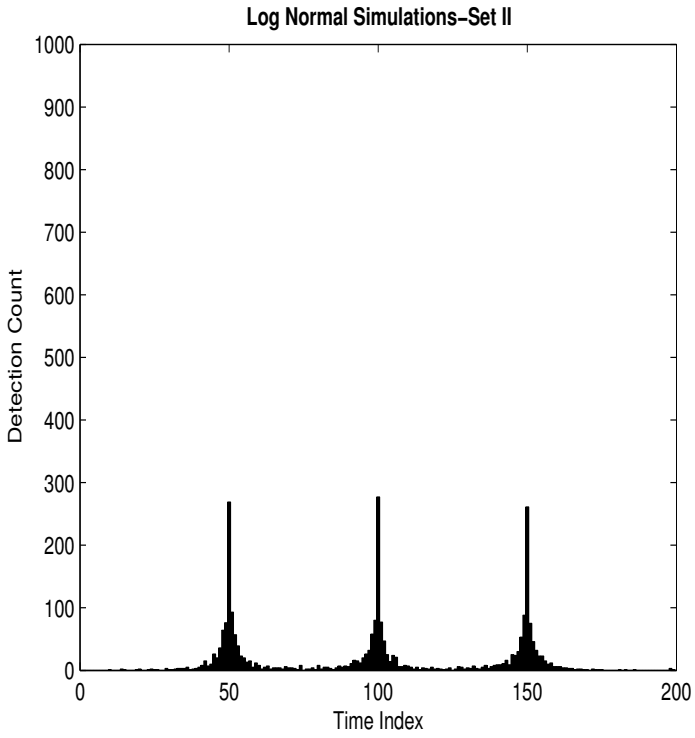

Figure 2.1: Count detection histogram I.

through 24, shifts up to 7.0 for times 25 through 74, moves downward to 6.6 for times 75 through 99, and then shifts upwards to the initial level of 6.8 from times 100 through 200. Table 2.3 shows the empirical proportions of estimated changepoint numbers and has a similar structure to the numbers reported in Table 2.2. Fig. 2.2 displays a count histogram akin to Fig. 2.1. While the estimated changepoint times still cluster symmetrically about their true values, the changepoint at time 75 was the easiest to detect. This is because the time 75 mean shift is twice the magnitude of the other mean shifts. It is interesting to note that the changepoints at times 25 and 75 were approximately equally difficult to detect (there are 99 shift-free data points after the time 100 changepoint, but only 24 shift-free points before the time 25 changepoint).

We also ran a Poisson simulation designed to mimic the annual tropical cyclone count data in Section 2.5. Here, we take $n = 160$ and superimpose two changepoints. Specifically, we start with a mean of 7.0 from times 1 through 79, shift upwards to 10 from times 80 to 145, and then move to 15 at times 145 through 160. Table 2.4 reports estimated values of m. The correct order $m = 2$ was estimated in 90.7% of the 1000 simulations. Fig. 2.3 shows a count histogram and reveals good

location performance. The changepoint at time 145 was slightly easier to detect, presumably due to its bigger mean shift.

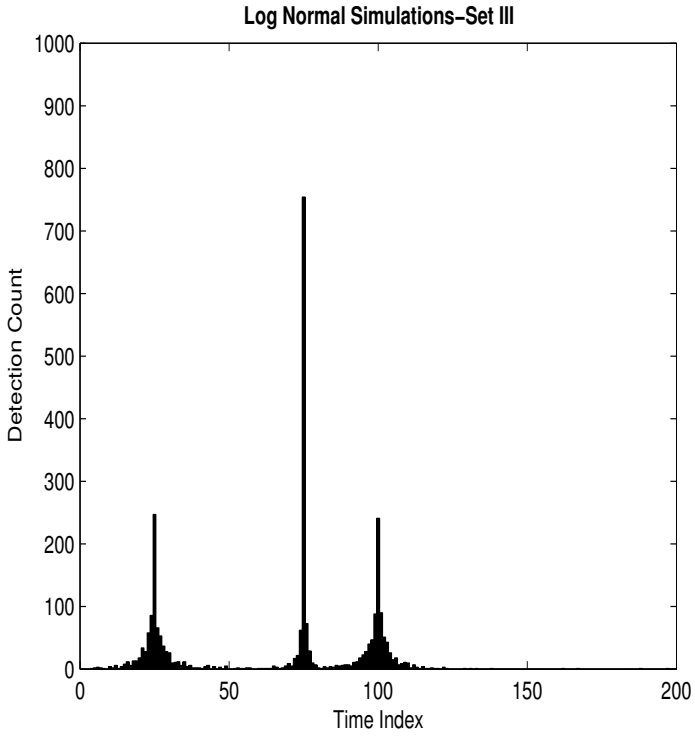

Figure 2.2: Count detection histogram II.

These and other simulations reveal the following themes. The closer the changepoints are in time, the more difficult they are to detect. Mean shifts in a monotone direction (all up or down) are easiest to detect. Also, as the autocorrelation in the series increases, detection power decreases.

2.4 The New Bedford Series

Fig. 2.4 plots a $N = 173$-year annual precipitation series from New Bedford, MA during 1818 — 1990. For this data, we first ran a GA with initial generation changepoint probability $p_i = 0.06$, generation size 200, and mutation probability $p_m = 0.003$. The model did not include autocorrelation; that is, we take $\phi = 0$ and optimize the MDL score in (2.5). The algorithm converged to a model with four changepoints at times $1867, 1910, 1965$, and 1967. The minimum MDL score achieved was -309.8570. This segmentation is graphed in Fig. 2.5 against the data and

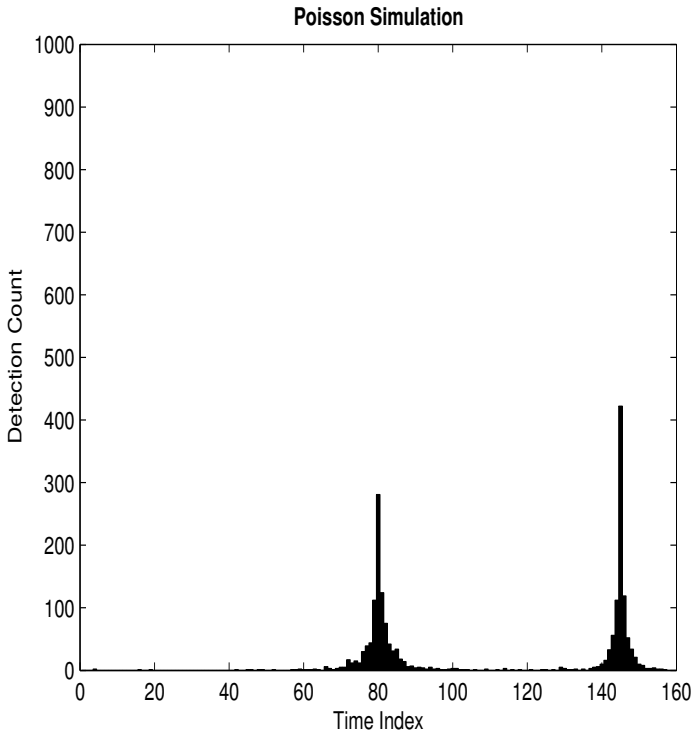

Figure 2.3: Poisson count detection histogram.

appears visually reasonable. The optimal segmentation has a short segment containing only 1965 and 1966, suggesting perhaps an outlier. The legitimacy of the small precipitation in 1966 may be questioned; however, the station's meta-data record, discussed further below, does not suggest the point is in error.

Application of a GA requires specification of the generation size, mutation probability p_m, and initial generation changepoint probability p_i. We have found that the GA will converge for a wide variety of choices of these parameters. Table 2.5 shows results for nine other GA runs with varying parameter settings. Except for the two runs with a generation size of 50, the GA has converged to the same four changepoint configuration with an MDL score of -309.8570.

As a check of this result, Table 2.6 shows optimum MDL scores for various numbers of model segments (the number of segments is one more than the number of changepoints). These values were obtained by exhaustive search of all candidate models. For instance, the minimal MDL score with three changepoints is -309.2878 and places the changepoints at times 1867, 1910, and 1967. The three changepoint optimal MDL score is slightly worse than the globally optimal model found by

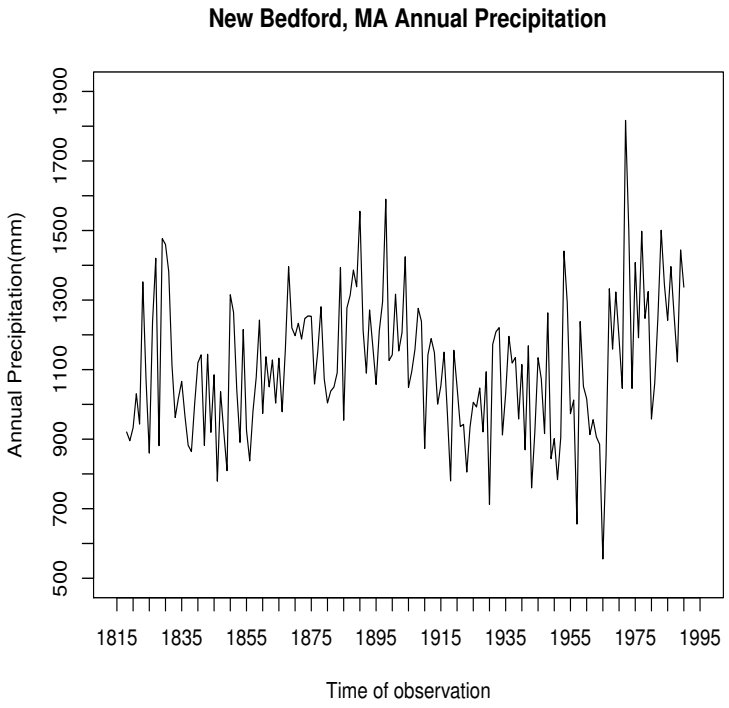

Figure 2.4: Annual precipitation at New Bedford, MA

the GA (which has four changepoints). In fact, the GA has selected (exactly) the best five segment model identified in Table 2.6; the times of all four changepoints in this model are identical. It should be emphasized that the GA implicitly estimates how many changepoints are present in the data, a seminal problem in itself. The exhaustive check of all six segment models alone took a week on a personal computer while the GA ran in several seconds.

To assess the effects of autocorrelation on the conclusions, the above analysis was rerun allowing the AR(1) parameter ϕ to be non-zero. A GA was run to minimize (2.7) and converges to the same four changepoint configuration with changepoints at times 1867, 1910, 1965, and 1967. The minimum MDL score was -309.9003 and the parameters of the GA are the same as those in the above analysis. The estimated autocorrelation coefficient was $\hat{\phi} = 0.021$, which is very close to zero (no correlation). The estimated white noise variance is $\hat{\sigma}^2 = 0.022$.

A reference series from Boston, MA is available to help make conclusions. The New Bedford to Boston ratios were computed and are displayed in Figure 6. A genetic algorithm was then constructed to minimize the MDL score in (2.7). This model allows for AR(1) autocorrelation. The

Table 2.5: GA convergence results with varying parameters for the New Bedford series. Most runs converge to a four changepoint model with an MDL of -309.8570.

Run #	p_m	p_i	Generation Size	MDL Score	Changepoint #
1	0.003	0.06	200	-309.8570	4
2	0.003	0.06	200	-309.8570	4
3	0.005	0.10	150	-309.8570	4
4	0.005	0.10	150	-309.8570	4
5	0.010	0.04	50	-307.6775	3
6	0.010	0.04	50	-308.4426	3
7	0.002	0.10	300	-309.8570	4
8	0.002	0.10	300	-309.8570	4
9	0.007	0.04	200	-309.8570	4
10	0.007	0.04	200	-309.8570	4

Table 2.6: Optimum MDL scores for various numbers of segments

# Segments	Changepoint Times	MDL Score
1	—	-296.7328
2	1967	-303.8382
3	1917, 1967	-306.6359
4	1867, 1910, 1967	-309.2878
5	1867, 1910, 1965, 1967	-309.8570
6	1829, 1832, 1867, 1910, 1967	-308.2182

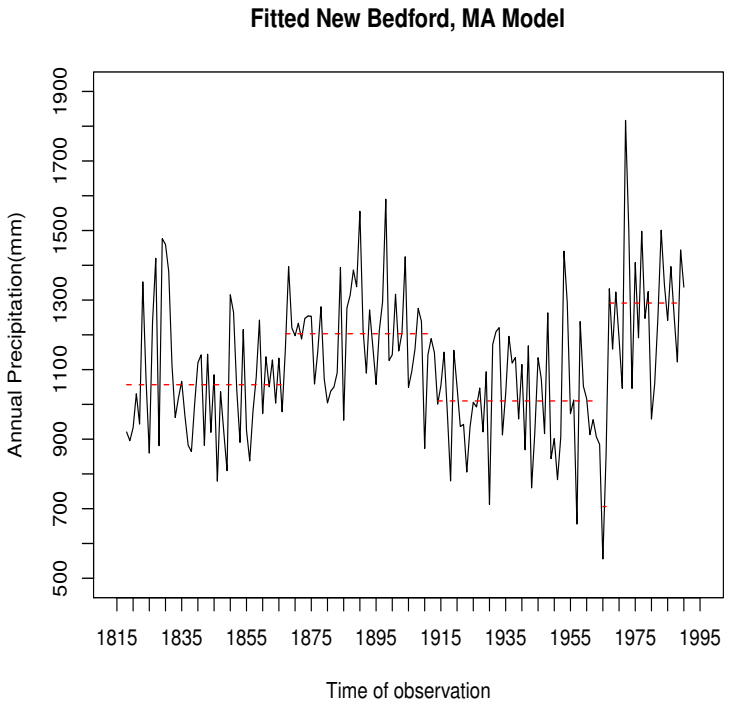

Figure 2.5: Optimal model with data superimposed. The optimal model has four changepoints. The fitted mean configuration (dashed) follows the data fluctuations well.

genetic parameters used are $p_m = 0.03$, $p_i = 0.06$, and a generation size of 200. Table 2.7 shows nine other GA runs with different parameter selections. All runs converge to the same MDL score of -327.1603 except the two runs with a generation size of 50. With this series, many competing models existed that had varying numbers of changepoints with slightly worse MDL scores than the -327.1603 optimum found.

The best fitting model now has four segments with changepoint times at 1886, 1917, and 1967. The data averages of the segments are plotted against the data in Fig. 2.3 and appear to move with the fluctuations of the target to reference ratios. The estimated AR(1) parameters are $\hat{\phi} = 0.31$ and $\hat{\sigma}^2 = 0.02$ and the optimal MDL achieved was -327.1603. This model fit contains considerably more autocorrelation than the reference-neglected fit.

For comparison's sake, we ran a simple segmentation algorithm on the log-ratios. A modified standard normal homogeneity test (SNHT) as discussed in Reeves et al. (2007) was used at level 95% to make AMOC conclusions. Subjecting the whole series to the SNHT reveals a changepoint at

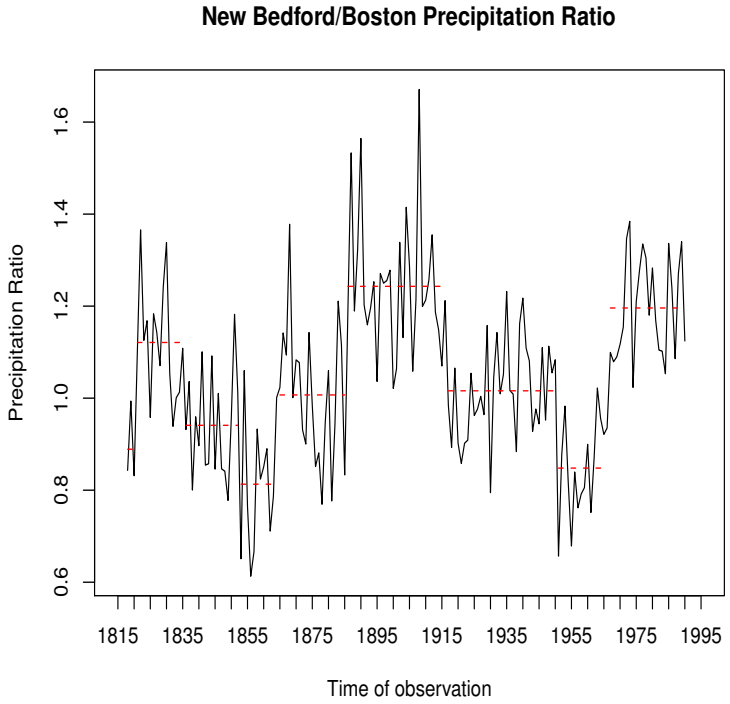

Figure 2.6: Optimal MDL model for precipitation ratios with data superimposed. Three estimated changepoint times are estimated. The fitted mean configuration (dashed) follows the data fluctuations well.

1883, which is close to the 1886 changepoint time flagged by the MDL segmenter. A SNHT analysis of data points 1818-1882 concludes another changepoint at time 1836. The 1818-1835 segment tests positively for another changepoint at time 1821, while the 1836-1882 segment tests as homogeneous. Turning to the 1883-1990 segment, a changepoint at 1917 is seen to be highly significant, which duplicates one of the MDL changepoint times verbatim. No further changepoints are found in the 1883-1916 segment, but the 1917-1990 segment is found to have a changepoint at time 1967, which is (exactly) a time the MDL segmentation flagged. The 1967-1990 segment tests as homogeneous, but the 1917-1966 segment is found to have changepoints at times 1951 and 1963. The SNHT segmentation has created a very short segment containing only 1963, 1964, and 1965 that the MDL method does not believe is distinct. In summary, a simple segmentation algorithm locates seven changepoints (four more than the MDL configuration) at times 1821, 1836, 1883, 1917, 1951, 1963, and 1967, and two times were flagged by both methods. The "mean shift configuration" of this

Table 2.7: GA convergence results with varying parameters for the New Bedford to Boston precipitation ratio series. Most runs converge to a three changepoint model with an MDL of -327.1603.

Run #	p_m	p_i	Generation Size	MDL Score	Changepoint #
1	0.003	0.06	200	-327.1603	3
2	0.003	0.06	200	-327.1603	3
3	0.005	0.10	150	-327.1603	3
4	0.005	0.10	150	-327.1603	3
5	0.010	0.04	50	-321.2423	2
6	0.010	0.04	50	-323.7953	2
7	0.002	0.10	300	-327.1603	3
8	0.002	0.10	300	-327.1603	3
9	0.007	0.04	200	-327.1603	3
10	0.007	0.04	200	-327.1603	3

segmentation is plotted against the data in Fig. 2.7. We reiterate that segmentation algorithms can be made smarter by reconsidering past conclusions once new subsegments are found (Hawkins 1976). In this case, the estimated MDL configuration has fewer changepoints than the estimated SNHT configuration. Part of this aspect is likely explained by the significant non-zero autocorrelation in the target to reference ratios. The reader is referred to Lund et al. (2007) for the influence of autocorrelation on changepoint detection.

The MDL results for the raw series also differ from the MDL results where the Boston reference was used. In particular, only the circa 1910 and circa 1965 changepoints were flagged in both analyses. Of course, one should trust the target to reference analysis more as this comparison reduces variability by removing some of the natural fluctuations common to both series.

The meta-data for the New Bedford station indicates station relocations in 1906 and 1974, changes in observation recording frequencies in 1854 and 1861, a change in the daily time that observations are recorded in 1951, and a change in the height of the precipitation gauge in 1985. The Boston reference series (NOAA 9699) is currently located at Logan Airport. We have been unable to obtain reliable meta-data for this station that spans its entire record (or even since Logan Airport's birth in 1923). Hence, it is difficult to attribute any of the changepoint times to specific station changes; however, the 1974 change is reasonably close to the 1967 breakpoint time flagged by both MDL and SNHT segmentations. We again refer the reader to Menne and Williams Jr. (2005, 2009) for an algorithm that discerns which station is responsible for the changepoint when many reference series are compared to the target in a pairwise fashion.

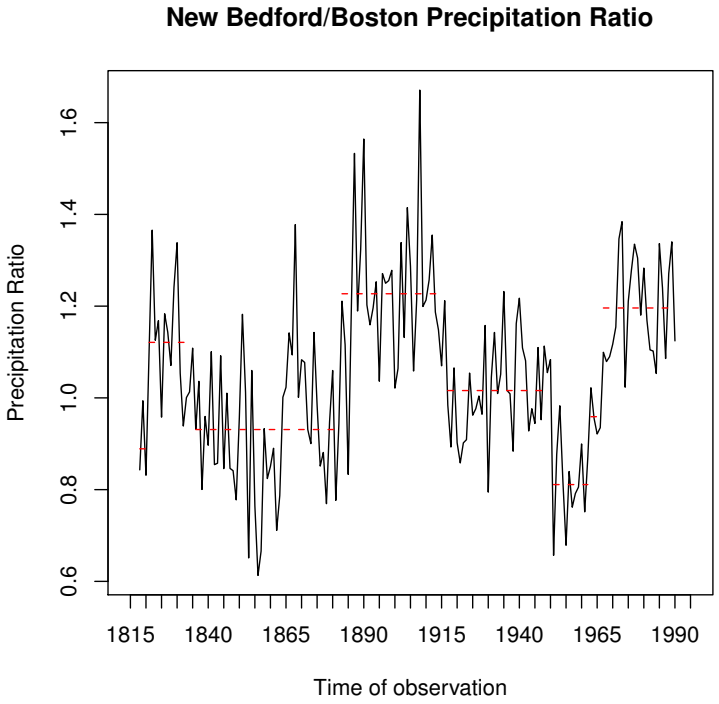

Figure 2.7: Model for precipitation ratios estimated via SNHT segmentation. This segmentation estimates seven changepoints.

2.5 The North Atlantic Tropical Cyclone Record

Our second application examines the North Atlantic Basin's annual tropical cyclone counts from 1851-2009. Here, we seek to identify times of statistical discontinuities in the record. The counts are plotted in Fig. 2.8 and include all storms that made at least tropical storm strength at any time during the storm's life, and were taken from the HURDAT data set, which is available on the National Oceanic Atmospheric Administration's website. In total, there are 1410 storms in the record. The record is thought to contain inconsistencies due to advances in measurement techniques. For instance, counts of landfalling cyclones before 1900 are considered unreliable (Landsea et al. 1999) due to sparse population along coastlines. Also, as Landsea et al. (1999) and Neumann et al. (1999) observe, aircraft reconnaissance towards the end of World War II (around 1944) improved detection of non-landfalling storms. Robbins et al. (2011) examines this record from a segmentation approach and finds two prominent changepoints at times 1931 and 1995. It would seem interesting to see how

a multiple changepoint segmenter compares to this result.

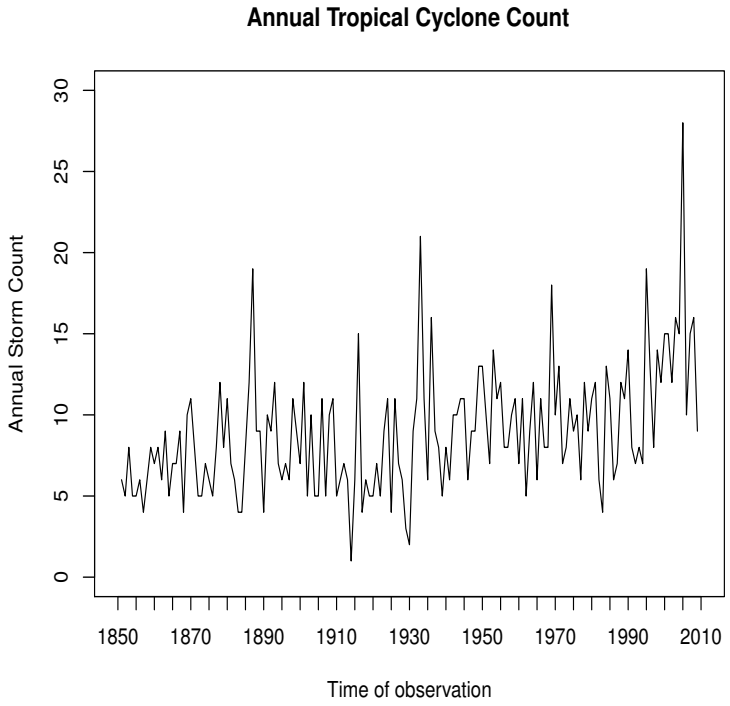

Figure 2.8: Annual Atlantic Basin tropical cyclone counts

Obviously, we do not have a reference series for this data. Also no definitive meta-data record exists. For a single site analysis, our model uses Poisson marginal distributions for the counts. Poisson distributions are natural count models and are known to describe tropical cyclone counts reasonably well (Mooley 1981; Thompson and Guttorp 1986; Solow 1989; Robbins et al. 2011). In truth, there is some overdispersion in the annual cyclone counts. This means that the variance of the annual counts is slightly higher than the mean — recall that a Poisson distribution has equal mean and variances — but this overdispersion is slight. Moreover, it does not appear that autocorrelation is present in the annual cyclone counts. The lack of correlation is confirmed in the empirical calculations in Robbins et al. (2011). Of course, if significant correlation in the annual counts did exist, it would be easier to forecast future year's counts one or more years in advance (one can have some forecasting power with shorter lead times). In short, we will base our model on the MDL developed in (2.8).

Fig. 2.9 graphically displays the mean structure of the optimal segmentation found by the

Poisson MDL segmenter. The parameters in the GA were taken as $p_m = 0.03$, $p_i = 0.06$, and generation size of 200. Here, the MDL judges three segments as optimal: one from 1851-1930, one from 1931-1994, and one for 1995-2009. The optimal MDL was -3131.40. This segmentation agrees exactly with that in Robbins et al. (2011). Table 2.8 displays convergence results for nine other GA runs. All runs opt for a two changepoint model, but the two runs with a generation size of 50 place the changepoint times slightly differently to 1931 and 1995.

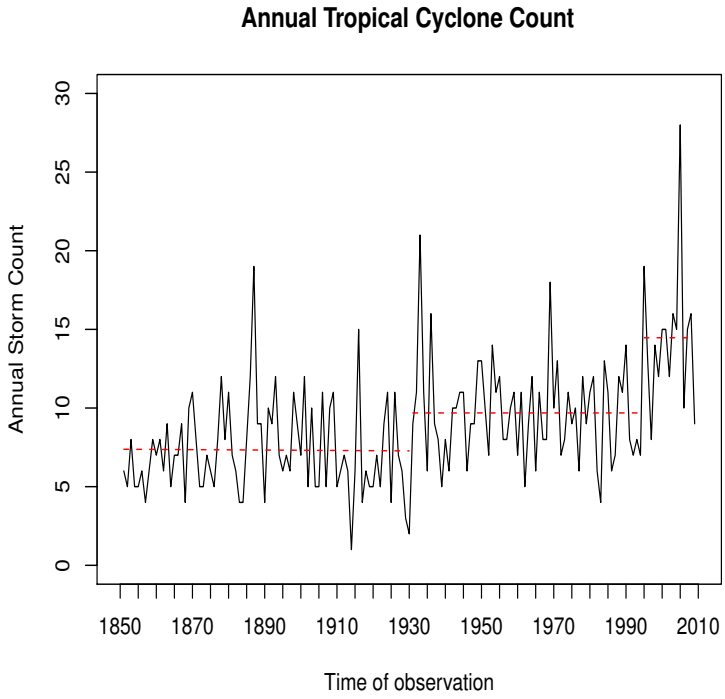

Figure 2.9: Optimal MDL model for cyclone count data with data superimposed. There are two estimated changepoint times and the fitted mean shift configuration (dashed) follows the data fluctuations well.

Overall, the cyclone counts appear to be increasing. The authors are unaware of any data collection changes that explain the 1995 changepoint. The 1931 changepoint is perhaps explained by the onset of aircraft surveillance, although it seems to occur about 10 years too early.

For end conclusions, it appears that tropical cyclone counts have increased recently (circa 1995). This contradicts July 28, 2009 Senate testimonial to the United States Senate by Dr. Kelvin Drogemeier that says the counts have remained stable. A deeper probabilistic assessment of the circa

Table 2.8: GA convergence results with varying parameters for the Atlantic tropical cyclone data. Most runs converge to a two changepoint model with an MDL of -3130.40.

Run #	p_m	p_i	Generation Size	MDL Score	Changepoint #
1	0.003	0.06	200	-3130.40	2
2	0.003	0.06	200	-3130.40	2
3	0.005	0.10	150	-3130.40	2
4	0.005	0.10	150	-3130.40	2
5	0.010	0.04	50	-3129.30	2
6	0.010	0.04	50	-3129.30	2
7	0.002	0.10	300	-3130.40	2
8	0.002	0.10	300	-3130.40	2
9	0.007	0.04	200	-3130.40	2
10	0.007	0.04	200	-3130.40	2

1995 changepoint is presented in Robbins et al. (2011) and examines the storm counts restricted to the post satellite era 1965-2008. For this segment, Robbins et al. (2011) again find a changepoint at 1995 with a p-value of 0.0234. Hence, it does appear that North Atlantic Basin tropical cyclone counts have recently increased.

2.6 Comments

This chapter presented a technique to estimate the number of changepoints and their locations in a climatic time series of annual values. The statistical rudiments of the methods were taken from information theory and are known as Minimum Description Length (MDL) techniques. MDL methods are penalized likelihood techniques, but differ from classic penalties like AIC by penalizing integer-valued parameters such as the changepoint numbers and locations more heavily than real-valued parameters such as a Poisson mean. Determining the number of changepoints and their locations is hence reduced to a statistical model selection problem. Because the model selection optimization entails searching a huge number of admissible changepoint configurations, a genetic algorithm (GA) was introduced that intelligently walks through the model space, discarding models that have little chance of being good. It was shown how to incorporate reference station aspects and autocorrelation features into the methods. The procedure estimated plausible changepoint numbers and configurations in the New Bedford, MA annual precipitation series and the annual North Atlantic Basin tropical cyclone counts.

Modifications of the methods here are worth pursuing. In particular, this study examined

annual data. Techniques for monthly and daily data with periodic features are worth developing should homogenization need to be done on such time scales. Also, our discourse here centered on mean shifts. It would be worthwhile to consider other regression structures. For example, a linear trend is plausible with temperature data. This is a simple matter of adding a linear trend into the regression setup and modifying the results. We caution that one should not apply our setup to data where there are clearly seasonal components, trends, etc. and expect good answers.

Finally, it would be seem useful to construct versions of MDL methods where the metadata is used to form a prior distribution of the changepoint configuration for a Bayesian analysis. Bayesian techniques have recently been used in climate changepoint research (Beaulieu et al. 2010) and seem promising.

Chapter 3

Trends in Extreme United States Temperatures

This chapter proceeds as follows. Section 3.1 describes our data while Section 3.2 presents our analysis methods. A series of monthly maximum temperatures from Jacksonville, Illinois is introduced in Section 3.1 and is analyzed in detail in Section 3.3. Section 3.4 reports results for all stations, and Section 3.5 concludes with comments.

3.1 The Data

Our monthly extremes are taken from the National Climatic Data Center's (NCDC's) United States Historical Climatology Network (USHCN) data. The USHCN data contains daily maximum and minimum temperatures for 1218 stations located throughout the 48 contiguous United States through December of 2010 (at the time of our analysis). The USHCN data are located at `http://cdiac.ornl.gov/epubs/ndp/ushcn/ushcn.html`.

Erroneous data entries do exist, but are not overly prevalent. Sometimes erroneous negative signs or extra digits were keyed in with the data. For example, some observations exceed the US record high temperature (134°F, Death Valley, California, July 10, 1913) or are lower than the US record low (−70°F, Rogers Pass, Montana, January 20, 1954). The NCDC has flagged inconsistent entries with quality control checks. All flagged temperatures are regarded as missing. Burt (2004)

Figure 3.1: Station locations.

contains a good compilation of United States temperature records.

Almost all stations in the USHCN daily data set have some missing data. One missing daily observation could change a monthly extreme greatly if the extreme, in truth, occurred on that missing day. Because of this, a monthly extreme is flagged as missing if one or more of the days within that month are missing. About 75% of the months in the maxima data are non-missing.

Stations where data begins or ends in the interior of a calendar year are cropped to full calendar years: each station's record begins with a January observation and ends with a December observation. This simplifies our notation and analysis. After this cropping, a station is required to have at least 75 years of observations with a missing rate of at most 33.3%, or have at least 50 years of record with a missing rate of at most 5% to make this study. These requirements leave 923 stations for maximum series. Fig. 3.1 graphically depicts the spatial location of these stations. The spatial coverage over the 48 contiguous United States is reasonable. When the above requirements are applied to each station's monthly minimum temperatures, 932 stations remain. The spatial coverage of the minimum stations is similar to that of the maximum stations.

The longest maximum temperature record comes from Atlantic City, New Jersey (137 years), and the shortest maximum record occurs at three stations, Bedford and Reading, Massachusetts, and Las Cruces, New Mexico (51 years). Fig. 3.2 presents a time series plot of the monthly

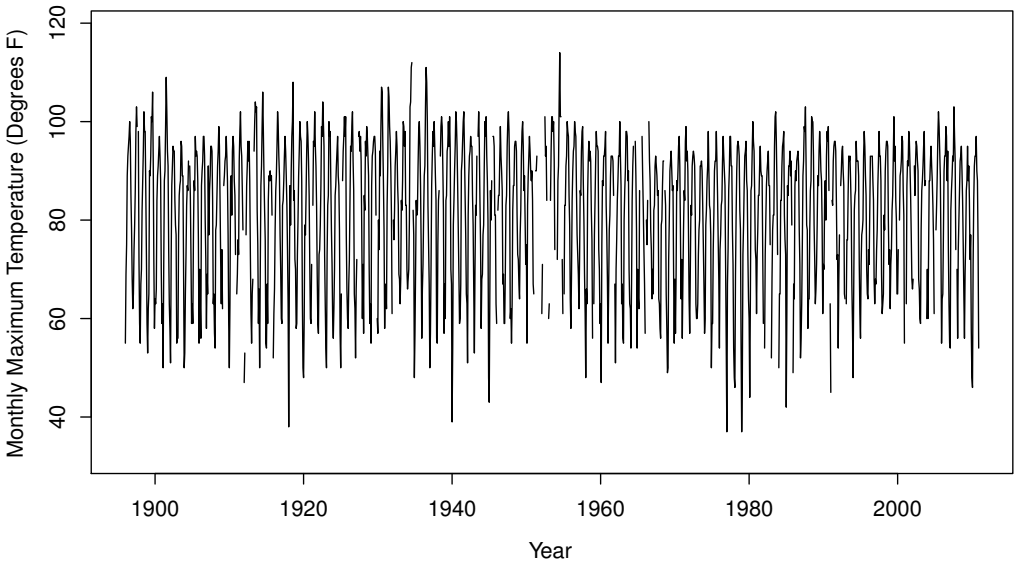

Figure 3.2: Monthly maxima at Jacksonville, Illinois from January 1896 — December 2010.

maxima observed at Jacksonville, Illinois. This station will be analyzed in detail in Section 3.3. The Jacksonville maximum series begins in January of 1896, ends in December of 2010, and has 115 years of monthly data with a missing rate of 4.42%. The Jacksonville maxima exhibit periodicity: winter temperatures are cooler and a little more variable than summer temperatures. A seasonal variability cycle is also evident: compare the year-to-year jaggedness of the summer peaks (smaller) in Fig. 3.2 to their winter counterparts (larger).

3.2 Methods

3.2.1 GEV Models

Our mathematical model for the monthly extremes $\{X_t\}$ at a fixed station is as follows. We assume that $\{X_t\}$ is independent in time t and marginally follows the generalized extreme value (GEV) distribution with location parameter μ_t, scale parameter $\sigma_t > 0$, and shape parameter ξ at

time t. The cumulative distribution function of X_t is, for $t = 1, \ldots, N$,

$$P[X_t \leq x] = \exp\left[-\left\{1 + \xi\left(\frac{x - \mu_t}{\sigma_t}\right)\right\}_+^{-1/\xi}\right], \tag{3.1}$$

where the subscript + indicates that the support set of the distribution in (3.1) is all x with $1 + \xi(x - \mu_t)/\sigma_t > 0$. In the case where $\xi = 0$, the distribution is taken as Gumbel (take limits as $\xi \to 0$). When $\xi < 1$,

$$E[X_t] = \mu_t + \frac{\sigma_t}{\xi}\left[\Gamma(1 - \xi) - 1\right], \tag{3.2}$$

and for $\xi < 1/2$,

$$Var(X_t) = \frac{\sigma_t^2}{\xi^2}\left[\Gamma(1 - 2\xi) - \Gamma^2(1 - \xi)\right], \tag{3.3}$$

where $\Gamma(\cdot)$ denotes the usual Gamma function.

To allow for time changes, the location parameter μ_t is parameterized by a linear trend with shifts at all changepoint times:

$$\mu_t = m_t + \alpha\left(\frac{t}{100T}\right) + \delta_t.$$

Here, m_t is the location parameter for month t, assumed periodic with period $T = 12$ ($m_{t+T} = m_t$), α is a linear trend parameter (our focus), and δ_t is a location shift changepoint factor obeying the shift structure

$$\delta_t = \begin{cases} \Delta_1, & \text{if } t = 1, \ldots, \tau_1 - 1; \\ \Delta_2, & \text{if } t = \tau_1, \ldots, \tau_2 - 1; \\ \vdots & \vdots \\ \Delta_{k+1}, & \text{if } t = \tau_k, \ldots, N. \end{cases}$$

In this setting, k is the number of changepoints and $\{\tau_1, \ldots, \tau_k\}$ are the ordered changepoint times. The number of changepoints k, their locations τ_i, $i \in \{1, \ldots, k\}$, and their associated location shifts Δ_j, $j \in \{2, \ldots, k+1\}$, are all unknown. To keep model parameters statistically identifiable, no shift parameter is allowed in the first regime; that is, $\Delta_1 = 0$.

To allow for periodic structures in $\{X_t\}$, the first-order Fourier representation

$$\sigma_t = c_0 + c_1 \cos\left(\frac{2\pi t}{T}\right) + c_2 \sin\left(\frac{2\pi t}{T}\right) \tag{3.4}$$

is used, where c_0, c_1, and c_2 are free parameters. In pilot computations, the seasonal location cycle $\{m_t\}_{t=1}^{T}$ is often inadequately described by a short Fourier series, but the first-order parametrization in (3.4) seems to work well for the parameters $\{\sigma_t\}_{t=1}^{T}$. One could also allow ξ to depend on time in a periodic way, but Coles (2001) advises (at least initially) to keep this parameter time-constant.

Our primary inferential objective involves the trend parameter α. Positive values of α indicate warming extremes; a negative α represents cooling extremes. The expected change in extremes over a century is obtained from (3.2) and is uniform in the season ν:

$$E[X_{(n+100)T+\nu}] - E[X_{nT+\nu}] = \mu_{(n+100)T+\nu} - \mu_{nT+\nu} = \alpha, \tag{3.5}$$

assuming that no changepoints occur between times $nT + \nu$ and $(n + 100)T + \nu$. This relation remains valid under any periodic form for σ_t or even if ξ were allowed to periodically vary.

Because the data are extremes, autocorrelation in $\{X_t\}$ is not allowed in our analysis. This is not to say that correlation is totally absent, but month-to-month temperature extremes typically exhibit weaker dependence than month-to-month temperature averages (this is intuitive as any 'freak' observation can serve to set a monthly extreme while monthly sample means are pulled toward a central tendency via the averaging). While correlation often does not change the limiting GEV distribution of the scaled process extremes (see Leadbetter et al. 1983), it would admittedly be better to block threshold each and every series, the typical way that dependent extremes are handled. Unfortunately, this is not feasible given the changepoint and periodicity features considered and the large number of stations in our study. Residual autocorrelation plots will be analyzed later to scrutinize this issue in finite samples. The issue is not found to be overly problematic.

Likelihood methods will be used to fit the extreme models. Given the number of changepoints k and their location times $\tau_1, \ldots, \tau_k$, a GEV likelihood is

$$L = L(k; \tau_1, \ldots, \tau_k) = \prod_{t=1}^{N} f_{X_t}(X_t), \tag{3.6}$$

where $f_{X_t}(x) = \frac{d}{dx} P[X_t \leq x]$ is the extreme value density of X_t. The optimal likelihood L_{opt} is

the likelihood L optimized over the parameters $m_1, \ldots, m_T$, α, $\Delta_2, \ldots, \Delta_{k+1}$, c_0, c_1, c_2, and ξ. This optimum needs to be found numerically.

A standard error $\widehat{\mathrm{Var}}(\hat{\alpha})^{1/2}$ for the expected extreme change over a century estimated via (3.5) is calculated by the usual information matrix associated with the likelihood fit. Later, these standard errors will be used in a spatial smoothing procedure.

The likelihood in assumes that the changepoint numbers and times are known. Unfortunately, this is not true in practice. Whereas files exist showing some of the station relocation and instrumentation change histories (the so-called meta-data), these files are notoriously incomplete. For a very rough flavor, it is estimated that only 40% of occurring changepoints made the meta-data logs; of the changepoint times that were documented, only about half of these induce shifts in the series.

Trends for individual stations are usually distrusted if homogenization has not been first attempted. The case study in the next section will reinforce this point. Our homogenization methods take the classic reference series approach. A reference series is a series from a location near the series being studied; the series being studied is called the target series. A good reference series is relatively changepoint-free and experiences similar weather to the target. The target minus reference subtraction serves to reduce variabilities and illuminate the locations of any shifts. In good target minus reference comparisons, the seasonal mean cycle and series variances are "reduced" compared to those in the target series.

The reference methods in this section allow us to construct a reasonable reference series for each target series, hence yielding estimates of the changepoint times and locations in the target series. Once the changepoint count and location times are known, it is easy to fit the GEV model to $\{X_t\}$ and obtain an estimate of the trend.

Multiple reference series for a given target series are often helpful (Menne and Williams 2005, 2009). Current NCDC methods compare over 40 distinct references to a given target (Menne and Williams 2009) before making changepoint conclusions. Issues arise in multiple reference station comparisons. Foremost, any changepoint in the reference will likely impart a changepoint in the target minus reference — one adds to the changepoint numbers by making reference comparisons. Menne and Williams (2009) devise the so-called pairwise algorithm to address this issue. The pairwise procedure is complicated, especially when assigning where the changepoints occur. To keep changepoint issues manageable but realistic, our approach will construct a composite reference series

by averaging many individual reference series. Strength is gained by considering multiple references, but issues of additional changepoints induced by the reference series are minimized in the averaging.

For each station, the 100 nearest ("as the crow-flys") neighboring stations are first selected. Since good reference stations are heavily correlated with the target, the correlation between the target and these 100 nearest neighbors is next computed. As suggested by Peterson et al. (1998), this correlation is computed after differencing at lag $T = 12$. Differencing at lag $T = 12$ eliminates the seasonal mean cycle and most of the changepoint shifts. In fact, $\delta_t - \delta_{t-T}$ is nonzero only when one or more changepoints occur between times $t - T$ and t. Let $\{X_t\}$ denote the target and $\{Y_t\}$ a candidate reference. With $U_t = \nabla_T X_t = X_t - X_{t-T}$ and $V_t = \nabla_T Y_t = Y_t - Y_{t-T}$, a good reference series maximizes the correlation

$$Corr(\{U_t\}, \{V_t\}) = \frac{\sum_{t=T+1}^{N}(U_t - \bar{U})(V_t - \bar{V})}{\left[\sum_{t=T+1}^{N}(U_t - \bar{U})^2\right]^{1/2}\left[\sum_{t=T+1}^{N}(V_t - \bar{V})^2\right]^{1/2}}. \tag{3.7}$$

Here, $\bar{U} = \sum_{t=T+1}^{N} U_t/(N-T)$ and $\bar{V} = \sum_{t=T+1}^{N} V_t/(N-T)$. The correlation in (3.7) is computed over the 100 nearest neighboring candidate references; time t data is not included should any missing quantities be encountered.

Our reference series will average the 40 neighboring series that have the largest correlation, as computed in (3.7), to the target. One caveat is made in selecting these 40 stations: only stations whose correlation to the target, as in (3.7), exceeds 0.5 are used. Subtracting a reference whose correlation does not exceed 0.5 can actually increase data variability. In our analysis of the 923 maximum stations, only 76 stations had less than 40 candidate reference series with the required 0.5+ correlation. Should there be less than 40 such reference stations, our composite reference simply averages over the number of stations that have the required correlation. It is noteworthy that four stations had no references (and these are only for the maxima): Eureka, CA; Fort Lauderdale, FL; Tarpon Springs, FL; and Brookings, OR. Interestingly, these stations are all coastal and are known for micro-climates, especially Eureka. These four stations are analyzed without a reference.

A more subtle issue involves the starting date for some of the longer series. Specifically, no reference station exists for the January 1874 data point at Atlantic City, New Jersey, the longest record in the study. To accommodate, the starting year of the Atlantic City series was advanced to 1901, which is the median starting year of the 40 reference stations with the highest correlation over times that are common to both records. By doing this, there are at least 20 reference stations at

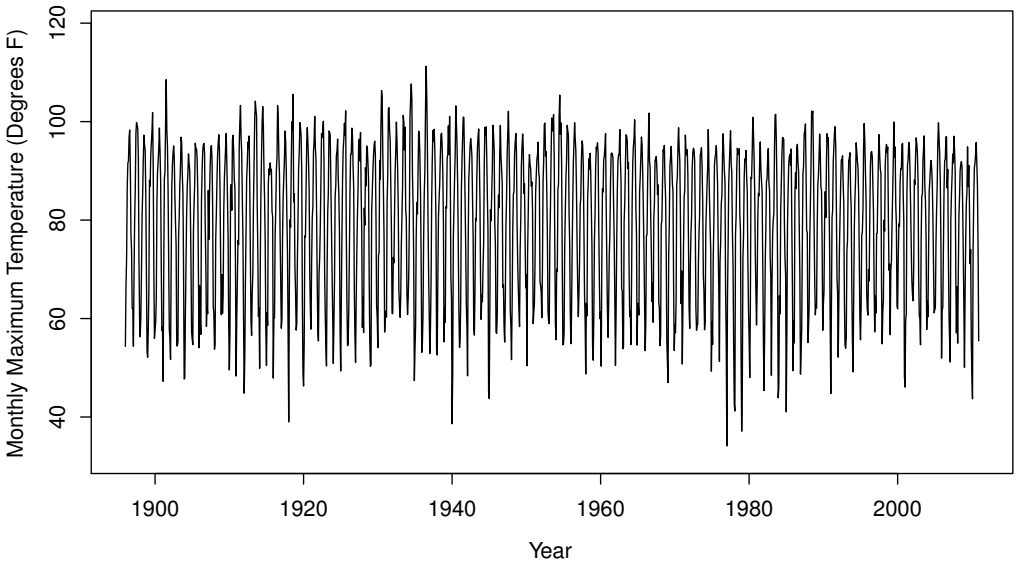

Figure 3.3: Jacksonville composite maxima reference series.

all times past 1901 for the Atlantic City series. A similar rubric is used for ending years, although this issue arises less frequently. If data is missing in one or more of the references, the denominator of the composite reference average is simply set to the number of references with non-missing data. Because of this, composite reference series do not usually have any missing data.

Fig. 3.3 shows our composite reference series for the maximum temperatures at Jacksonville, Illinois. Fig. 3.4 displays a histogram of the target minus reference differences. While not exactly Gaussian (formal normality tests are not passed at level 95%), it may be surprising that the target minus reference series' marginal distribution is not radically non-Gaussian. Pilot computations with the target minus reference series reveal seasonal means and variances, but no other periodic structure. Elaborating, the coherence tests of [**?**] were applied to assess whether or not the differenced series is stationary after subtraction of a linear trend and monthly sample means and division by a monthly sample standard deviation. Fig. 3.5 shows a coherence plot with a 99% pointwise confidence threshold for these seasonally adjusted differences. As there are no large exceedances of the 99% threshold, one concludes that the target minus reference series, beyond monthly means and variances, has no additional periodic structure. This simplifies our model in the next subsection.

Other stations were also scrutinized; in all cases, the conclusions are the same as those for

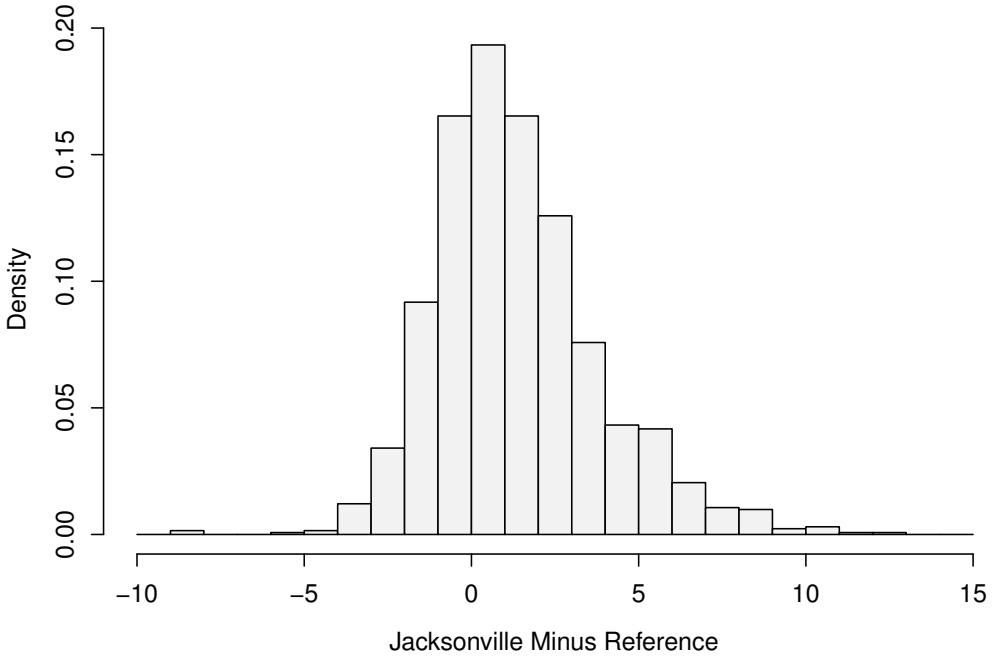

Figure 3.4: Histogram of the Jacksonville maximum target minus reference differences.

the Jacksonville series. Hence, we move to our next task — finding the changepoint locations in any target minus composite reference series. Suppose that a target minus composite reference difference series $\{D_t\}$, where $D_t = X_t - \tilde{Y}_t$ with $\{\tilde{Y}_t\}$ as the composite reference series described in the previous subsection, has been computed at the times $t = 1, \ldots, N$. We assume that $N = dT$ for some whole number d so that there are d complete cycles of data available (that is, d is a whole number).

A minimum description length (MDL) criterion for estimating the number and location of the changepoint times minimizes a penalized likelihood score of form

$$\text{MDL}(k, \tau_1, \ldots, \tau_k) = -\log_2(L_{opt}) + P. \tag{3.8}$$

In (3.8), L_{opt} is an optimized model likelihood *given the number of changepoints and where they occur*, P is a penalty term that accounts for the number and type of model parameters, and $\log_2$ indicates logarithm base 2. MDL methods have yielded promising results in recent changepoint studies (Davis et al. 2006; Lu et al. 2010; Li and Lund 2012). The MDL penalty is based on minimum description length information theoretic principles. While the reader is referred to the above

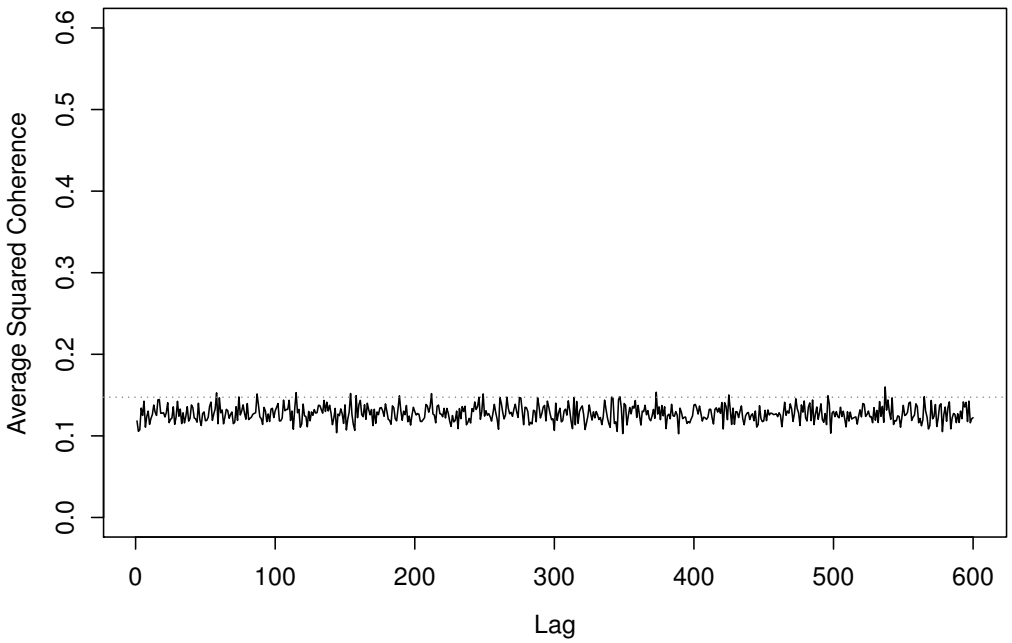

Figure 3.5: Average squared coherences for the seasonally adjusted Jacksonville, IL (maxima) target minus reference data. The absence of values exceeding the pointwise 99% confidence threshold suggests stationarity.

references for technicalities, the key point distinguishing MDL penalties from classical statistical penalties such as AIC and BIC is that MDL penalties are not solely based on the total number of model parameters, but also account for the parameter type and changepoint numbers and locations. Elaborating, MDL penalties penalize integer-valued parameters, such as the changepoint numbers and locations, more heavily than real-valued parameters such as the trend. MDL penalties also account for the changepoint configuration, penalizing configurations where the changepoint times occur close together more heavily than uniformly spaced configurations.

Our methods take $\{D_t\}$ as Gaussian, equipped with periodic means and variances, to estimate the changepoint count and location(s). Gaussianity is only used to estimate the changepoint number(s) and location(s); GEV models will be fitted after the changepoint configuration is estimated. This allows us to incorporate autocorrelation aspects into all changepoint inferences.

Mathematically, the model for $\{D_t\}$ has the periodic simple linear regression form

$$D_{nT+\nu} = m^*_\nu + \alpha^*(nT+\nu) + \delta_{nT+\nu} + \sigma^*_\nu \omega_{nT+\nu}. \tag{3.9}$$

The terms in (3.9) are described as follows. First, a periodic notation is used where $T = 12$ is the period and $\nu \in \{1, \ldots, 12\}$ signifies the month (season) corresponding to time $nT + \nu$. The m^*_ν terms allow for a periodic monthly mean cycle satisfying $m^*_{t+T} = m^*_t$ for all t. Observe that m^*_ν and m_ν may differ as may α and α^* and σ_ν and σ^*_ν. The term σ^*_ν is included to describe the periodic variances present in $\{D_t\}$. Again, σ^*_ν and σ_ν are different parameters. As our case study in the next section shows, constructing a target minus reference difference will not necessarily completely eliminate the seasonal mean and variance structures in $\{D_t\}$. The error terms $\{\omega_t\}$ are posited to be first-order autoregressive noise with lag-one autocorrelation parameter $\phi \in (-1, 1)$ and white noise variance σ^2. As it is not overly important to model the autocorrelation structure of $\{D_t\}$ to exactitudes — and the correlation structure of $\{D_t\}$ is often simple due to the differencing — a first-order autoregression is used. It is straightforward to extend methods to higher order autoregressions should this be desired. This said, one does not want to ignore correlation aspects completely as they can drastically influence changepoint conclusions (Lund et al. 2007). Elaborating, neglecting positive autocorrelations can induce the spurious conclusion of an excessive number of changepoints. We prefer to allow a linear trend parameter α^* in the target minus reference representation, which again need not be the same as the trend parameter α in the representation for $\{X_t\}$, for the following reason. If target series $\{X_t\}$ has a linear trend that is not the same as that in the reference, then a linear trend exists in the target minus reference. Such a situation could arise if, for example, the target is experiencing heating due to urban sprawl while its neighbors in the reference are not. When changepoint methods that assume no trend are applied to data with trends, they often spuriously flag many changepoints. This is a situation to avoid.

We now develop the penalty term in (3.8). In computing an MDL penalty, three principles are needed. First, the penalty for a real-valued parameter estimated from g data points is $\log_2(g)/2$. Second, the penalty for an integer-valued parameter I that is known to be bounded by the integer M is $\log_2(M)$. If no bound for I is known, the parameter is penalized $\log_2(I)$ units. Third, the model penalty P is obtained by adding the penalty for all model parameters.

To derive an MDL penalty, we assume first that there is no missing data. Then the three

parameters α^*, ϕ, and σ^2 are all real-valued and estimated from all N data points. Hence, they are charged a $\log_2(N)/2$ penalty each. The seasonal location and variance parameters m_ν^* and σ_ν^*, $\nu \in \{1, \ldots, T\}$, are real-valued and estimated via the data from season ν only; hence, they are each penalized $\log_2(d)/2$. The jth-regime location parameter Δ_j, $j \in \{2, \ldots, k+1\}$ (recall that $\Delta_1 = 0$ for model identifiability), is real-valued and estimated from data in the jth regime (the times from τ_{j-1} through $\tau_j - 1$). Thus, Δ_j is penalized $\log_2(\tau_j - \tau_{j-1})/2$. The boundary conventions $\tau_0 = 1$ and $\tau_{k+1} = N + 1$ are made for the first and last regimes. The number of regimes parameter is $k + 1$ and is charged $\log_2(k + 1)$ since this integer-valued parameter is unknown. Finally, since τ_i is integer-valued and $\tau_i < \tau_{i+1}$, τ_i is charged a $\log_2(\tau_{i+1})$ penalty. Adding the above together gives the penalty

$$\frac{3}{2}\log_2(N) + T\log_2(d) + \frac{1}{2}\sum_{j=2}^{k+1}\log_2(\tau_j - \tau_{j-1}) + \log_2(k+1) + \sum_{j=2}^{k}\log_2(\tau_j) + \log_2(N+1).$$

Notice that this penalty depends on the changepoint count k and the changepoint configuration $\{\tau_1, \ldots, \tau_k\}$. Since terms that are constant in N or d will not change where the minimal MDL is achieved, the above penalty is simplified to

$$P = \frac{1}{2}\sum_{j=2}^{k+1}\log_2(\tau_j - \tau_{j-1}) + \log_2(k+1) + \sum_{j=2}^{k}\log_2(\tau_j).$$

For cases with missing data, one simply changes $\tau_j - \tau_{j-1}$ to the number of data points in the jth regime, etc.

The likelihood used in (3.8) is developed in detail in Lu et al. (2010). It is Gaussian, conditional on the stipulation that k changepoints occur at the times $\{\tau_1, \ldots, \tau_k\}$, and can be written in the Innovations form (see Brockwell and Davis, 1991):

$$L = (2\pi)^{-N/2}\left(\prod_{t=1}^{N} v_t^{-1/2}\right)\exp\left[-\frac{1}{2}\sum_{t=1}^{N}\frac{(D_t - \hat{D}_t)^2}{v_t}\right]. \tag{3.10}$$

Here, $\hat{D}_t = P(D_t|1, D_1, \ldots, D_{t-1})$ is the best linear prediction of D_t from past observations and a constant, and $v_t = E[(D_t - \hat{D}_t)^2]$ is its unconditional mean squared error. For a given changepoint

configuration $\{\tau_1, \ldots, \tau_k\}$, we can further express $\hat{D}_t$ and v_t via the AR(1) prediction relationships:

$$\hat{D}_{nT+\nu} = E[D_{nT+\nu}] + \frac{\phi\sigma^*_\nu}{\sigma^*_{\nu-1}} \left(D_{nT+\nu-1} - E[D_{nT+\nu-1}]\right), \tag{3.11}$$

$$v_{nT+\nu} = \sigma^{*2}_\nu(1-\phi^2), \tag{3.12}$$

with the startup conditions $\hat{D}_1 = E[D_1]$ and $v_1 = \sigma^{*2}_1$. Here, all terms in (3.11) and (3.12), excluding ϕ, are treated as being periodic with period T, and the mean in (3.11) is

$$E[D_{nT+\nu}] = m^*_\nu + \alpha^*(nT+\nu) + \delta_{nT+\nu}.$$

The likelihood in (3.10) can then be computed. For each changepoint configuration, maximum likelihood estimators of $m^*_1, \ldots, m^*_T$, α^*, $\Delta_2, \ldots, \Delta_{k+1}$, $\sigma^*_1, \ldots, \sigma^*_T$, ϕ, and σ^2 are obtained. This computation is not overly difficult and is described in Li and Lund (2012).

A serious computational issue now arises. It is not feasible to compute the penalized likelihood in (3.8) over all possible changepoint numbers k and configurations $\{\tau_1, \ldots, \tau_k\}$ when N is large. Indeed, there are $\binom{N}{k}$ ways to arrange k changepoints in N places. Summing this count over all k from $0, 1, \ldots, N$ and applying the binomial theorem shows that there are 2^N distinct changepoint configurations. For $N = 1200$ (a century of monthly data), an exhaustive check of all changepoint configurations would require 2^{1200} different likelihood fits, which is not feasible. In the next subsection, a genetic algorithm is introduced that intelligently walks through this huge sample space and avoids evaluating the likelihood at configurations that are likely to be suboptimal.

3.2.2 The Genetic Algorithm

A genetic algorithm (GA), which is essentially a Markov stochastic search, will be used to estimate the number of changepoints and their times in the target minus reference difference series. The GA used here will be similar to that developed in Li and Lund (2012), but has seasonal aspects.

Genetic algorithms are described via chromosomes. Chromosomes here have the form $(k; \tau_1, \ldots, \tau_k)$ and contain all changepoint information. Each different chromosome is viewed as a different individual in a population. One can compute an MDL score for a fixed chromosome from the methods in the last subsection. Individuals in the population are termed fitter (relatively) when

they have a smaller (relatively) MDL score.

GAs need to breed two chromosome configurations, called the mother and father, in a probabilistic manner to form a child. The better fit individuals will be more likely to breed and pass on their chromosomes to the next generation, thus mimicking natural selection principles. Suppose a generation contains L individuals. A mother and father are selected from these L chromosomes as follows. The ith chromosome is selected as the father with probability $R_i / \sum_{j=1}^{L} R_j$, where R_i is the MDL rank of the ith chromosome (the best MDL score is given rank L). A mother is then chosen from all remaining chromosomes (excluding the father) after reranking all non-father chromosomes.

From a mother and father chromosome, a child chromosome is randomly generated as follows. Suppose $(i; \varsigma_1, \ldots, \varsigma_i)$ and $(j; \tau_1, \ldots, \tau_j)$ are the mother and father chromosomes, respectively. The child's chromosome is produced in three steps. First, the mother and father's chromosomes are combined by forming the chromosome $(i + j; \kappa_1, \ldots, \kappa_{i+j})$. Here, the κ_ℓs contain all changepoint times of *both* mother and father. The number of changepoints is strictly less than $i + j$ should the mother and father have some common changepoint times. Second, the κ_ℓ combined changepoints are then retained/discarded with independent coin flips with success probability 0.5. This acts to thin the number of changepoints. Finally, we allow the changepoint times that remain to move their locations slightly: each changepoint location stays the same with probability 0.4, moves to one time smaller with probability 0.3, or moves to one time larger with probability 0.3 (subject to the changepoint time being in $\{1, \ldots, N\}$). For example, with $N = 8$, suppose that a mother and father have the chromosome $(1; 6)$ and $(3; 3, 5, 6)$, respectively. Then the child chromosome is first set to $(3; 3, 5, 6)$. Three fair coins are then flipped independently. Should this have resulted in success, failure, and success, the chromosome is thinned to $(2; 3, 6)$. Two draws from the above location shift generation mechanism might then, for example, keep the time 3 changepoint where it is and shift the time 6 changepoint to 7. This yields the end chromosome $(2; 3, 7)$. Once one child is generated, the process is repeated until L new children are formed. These children represent the next generation. We do not allow different children to have the exact same chromosome; however, a mother and father could be the parents of more than one child.

Mutation is an aspect of GAs added to prevent premature convergence to poor solutions (local minima). Our mutation mechanism allows a small portion of children to have "extra changepoints". Specifically, after each child is formed from its parents, each and every non-changepoint time is independently allowed to become a changepoint time with probability p_{mut}. In the computations

below, $p_{mut} = 0.003$ is used.

In this manner, successive generations are simulated. The solution to the optimization problem is taken as the fittest chromosome in the terminating generation. One terminates the GA when there is little or no improvement to the fittest member of a few successive generations. The specifics of how this is done are usually of little consequence.

One must deal with missing data in the above setup. In the GAs, we simply do not allow a changepoint to occur at a time where the target series is missing (as noted above, the reference series is almost never missing). If a generated chromosome attempts to put the changepoint at a time where the target series is missing, we move the changepoint rightwards (higher) to the first time point with present data. The likelihood in (3.10) also needs to be modified to sum only over the present data. Should we wish to predict $D_{nT+\nu}$ and the most recent non-missing data point is $D_{nT+\nu-k}$, then the prediction now become k-step-ahead

$$\hat{D}_{nT+\nu} = E[D_{nT+\nu}] + \frac{\phi^k \sigma^*_\nu}{\sigma^*_{\nu-k}} \left(D_{nT+\nu-k} - E[D_{nT+\nu-k}] \right).$$

The mean square prediction error $v_{nT+\nu}$ is changed to $\sigma^{*2}_\nu(1-\phi^{2k})$ for times $nT+\nu = 2, \ldots, N$ and σ^{*2}_1 at time $nT+\nu = 1$.

3.2.3 Spatial Smoothing Methods

After the GEV likelihoods are fitted, each station will have an estimated trend for its minimum and maximum series. Also computed are standard errors for these trend margins. To help interpret the geographical pattern of the results, the estimated trends will be spatially smoothed. Specifically, the head-banging algorithm, discussed in Hansen (1991), will be applied to smooth the raw trends and their z-scores $Z = \hat{\alpha}/\mathrm{Var}(\hat{\alpha})^{1/2}$ by station longitude and latitude.

The head-banging algorithm is a robust median-polished smoother that extracts general structure well from noisy data. It is named from a child's game where a face is banged against a board of nails protruding at various lengths, leaving an impression of the face, but smoothing the residual nail lengths. Briefly, head-banging techniques are local median methods that class stations into many subsets of neighboring stations over which median trends are taken. Taking local medians accounts for spatial correlation in the trend estimates in a nonparametric manner. To run the head-banging algorithm, one only needs to set a parameter, called the number of triples. The number of

Table 3.1: Jacksonville GEV monthly location estimates in degrees F

Month	Changepoint ignored GEV fit	Two changepoint GEV fit
January	55.703 (0.705)	56.846 (0.948)
February	60.464 (0.686)	61.626 (0.941)
March	73.766 (0.638)	74.907 (0.894)
April	83.198 (0.564)	84.141 (0.857)
May	88.347 (0.493)	89.322 (0.808)
June	94.039 (0.436)	95.054 (0.777)
July	97.142 (0.424)	98.182 (0.764)
August	96.115 (0.425)	97.129 (0.775)
September	91.881 (0.474)	92.925 (0.808)
October	84.364 (0.541)	85.388 (0.848)
November	72.326 (0.623)	73.307 (0.901)
December	59.908 (0.679)	60.905 (0.936)

triples essentially represents the number of neighboring stations that will be used to compute the smoothed values.

3.3 A Case Station Study

This section examines the monthly maximum series from Jacksonville, IL introduced in Section 3.1. The GA applied to the target minus reference differences estimates two changepoints at the times 49 (January, 1900) and 730 (October, 1956). The estimated AR(1) coefficient in this fit is $\hat{\phi} = 0.160$.

When the two changepoints are ignored and the GEV model is fitted — this allows for general monthly means and a first order Fourier expansion for $\{\sigma_t\}_{t=1}^{T}$ σ_t — the trend estimate is $\hat{\alpha} = -1.667$ (0.399)°F per century (error margins are one standard error). The estimated GEV shape parameter is $\hat{\xi} = -0.182$ (0.012). Table 3.1 (second column) shows monthly GEV estimates of m_ν. The estimated coefficients in the first-order Fourier expansion of σ_t are $\hat{c}_0 = 5.077$ (0.101), $\hat{c}_1 = 1.286$ (0.147), and $\hat{c}_2 = 0.879$ (0.139). From these statistics, one might conclude that the Jacksonville monthly maxima are cooling. It is also worth noting that the estimated shape parameter ξ is negative, implying a finite upper limit for temperatures (ignoring trends).

Aspects change when the two changepoints are considered. While all changepoints are deemed to induce significant location shifts by the GA, they may not be significant as gaged by the GEV likelihood. The least "GEV significant" changepoint time (at the 5% significance level)

is eliminated, and the GEV likelihood is then refit sequentially until all changepoints are deemed significant at level 5%. Elaborating, the jth changepoint, where $j = 2, \ldots, k+1$, is GEV significant if $\Delta_j - \Delta_{j-1}$ is significantly non-zero. To gauge this, the z-score

$$\frac{\hat{\Delta}_j - \hat{\Delta}_{j-1}}{\mathrm{Var}(\hat{\Delta}_j - \hat{\Delta}_{j-1})^{1/2}}$$

is computed, and the jth changepoint is eliminated if its absolute z-score is smallest and less than 1.96. Recall that $\Delta_1 = 0$ was taken for parameter identifiability. The covariances $\mathrm{Cov}(\hat{\Delta}_j, \hat{\Delta}_{j-1})$ needed to estimate $\mathrm{Var}(\hat{\Delta}_j - \hat{\Delta}_{j-1})$ are extracted from the information matrix in the GEV fit. For the Jacksonville maximum series, $\hat{\Delta}_2 = -2.492$ (0.765) and $\hat{\Delta}_3 - \hat{\Delta}_2 = -4.869$ (0.541). As such, the two changepoints are GEV significant, and we do not eliminate either of them. The estimated shape parameter is $\hat{\xi} = -0.198$ (0.014), and the estimated trend changes to $\hat{\alpha} = 4.892$ (0.812)°F per century. Table 3.1 (third column) shows estimated monthly estimates of m_ν. They are all larger than the column 2 estimates, reflecting perhaps the extra uncertainty the two changepoints add. The estimated coefficients in the first-order Fourier expansion of σ_t become $\hat{c}_0 = 4.972$ (0.099), $\hat{c}_1 = 1.343$ (0.141), and $\hat{c}_2 = 0.887$ (0.134).

The crux here is that the estimated trend $\hat{\alpha}$ reverses sign from the no changepoint fit: *from* $-1.667(0.399)$°F per century *to* $4.892(0.812)$°F per century. A plot of the estimated mean function of the no- and two-changepoint models is superimposed upon the raw series after monthly subtraction of $\hat{m}_\nu$ in . The two-changepoint model seems to describe the series well. Obviously, trend inferences greatly change when changepoint features are incorporated into the analysis.

To assess the importance of autocorrelation in the month-to-month extremes, Fig. 3.7 plots the sample autocorrelations of the seasonally adjusted Jacksonville extreme residuals

$$\frac{X_{nT+\nu} - \hat{E}[X_{nT+\nu}]}{\widehat{Var}(X_{nT+\nu})^{1/2}},$$

where the mean and variance are computed from (3.2) and (3.3). Pointwise 95% bounds for white noise are included. It appears that the autocorrelations at lags one and two are nonzero (the lag-one sample autocorrelation is 0.212), but that higher order autocorrelations are essentially zero. While this moderate amount of autocorrelation is not completely ignorable, accounting for it in the GEV likelihood will not change our trend inferences appreciably.

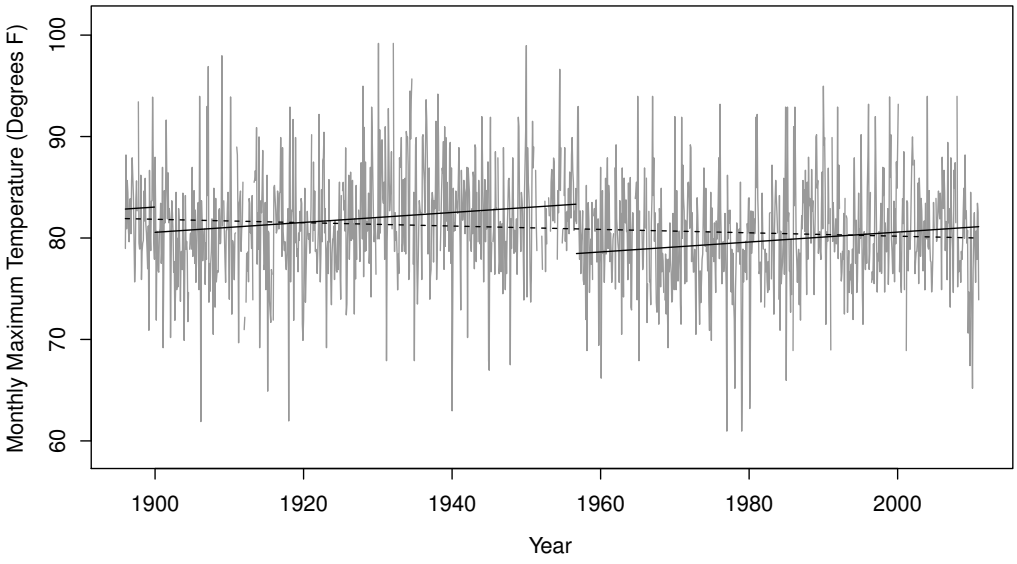

Figure 3.6: Fitted model structures for Jacksonville series.

3.4 Results for All Stations

Raw trends for the maximum temperatures for all 923 stations are displayed in Fig. 3.8. While it may be surprising that 583 of the 923 stations had negative trend estimates, a "warming hole" in the Eastern United States has been previously noted (Lund et al. 2001; Robinson et al. 2002; DeGaetano and Allen 2002; Lu et al. 2005; Kunkel et al. 2006; Meehl et al. 2012, among many others). A histogram of the 923 GEV trends is supplied in Fig. 3.9. The head-banging algorithm was applied to the raw trends with a smoothing parameter of 10 triples. The result is depicted in Fig. 3.10. Here, color shades run from deep red (the most warming) to deep blue (the most cooling). In aggregate, monthly maximum temperatures are decreasing in the Eastern United States, with the exception of New England. In contrast, the Western United States' maximum temperatures are slightly warming in aggregate. Head-banging smoothed Z-scores for the trends are displayed in Fig. 3.11. Our inferences are more confidently made in the Southeast (cooling) and the Northern Rockies (warming).

The trends for minimum temperatures show a different pattern. The raw trends for all 932 stations are displayed in Fig. 3.12. Here, the majority of the trends are increasing (728 of the 932 stations). A histogram of the 932 trends is shown in Fig. 3.13. The head-banging smoothed

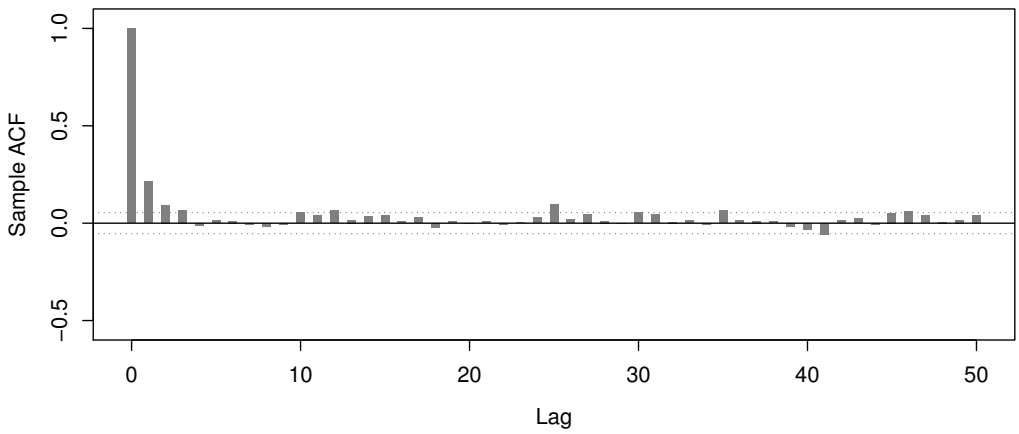

Figure 3.7: Sample autocorrelations of the seasonally scaled residuals.

minimum trends (10 triples again) are shown in Fig. 3.14. With the exception of two localized pockets in the Southeast and Colorado, cooling is sparse. Head-banging smoothed Z-scores for the trends (10 triples again) are displayed in Fig. 3.15.

The average trend in the maxima is -0.842°F per century with an average standard deviation of 3.686°F per century (over all 923 stations). The average trend in the minima is 2.962°F per century with an average standard deviation of 4.293°F per century (this is over 932 stations). It is interesting to assess the role of changepoints. The average numbers of "GEV significant" changepoints are 1.74 for the maximum series stations and 1.91 for the minimum series. When GEV likelihoods are fitted to the station maximum data without changepoint effects, the average trend is -0.756°F per century; the corresponding average trend in the minima when changepoints are ignored is 1.824°F per century. Including changepoints acts to give us more cooling in maximum temperatures (slightly) and more warming in minimum temperatures.

Comparing to the trends in mean United States temperatures reported in Lund et al. (2001) and Lu et al. (2005), similar spatial patterns emerge: a cooling Southeastern United States with warming elsewhere. One noticeable discrepancy from Lu et al. (2005) involves the cooling seen in the monthly maxima in the Southern Great Lakes and Ohio Valley. The overall results also support the statement that minimum temperatures are warming much more rapidly than maximum temperatures, a hypothesis generally believed to be consistent with warming induced by carbon dioxide (Karl et al. 1991; Jones et al. 1999).

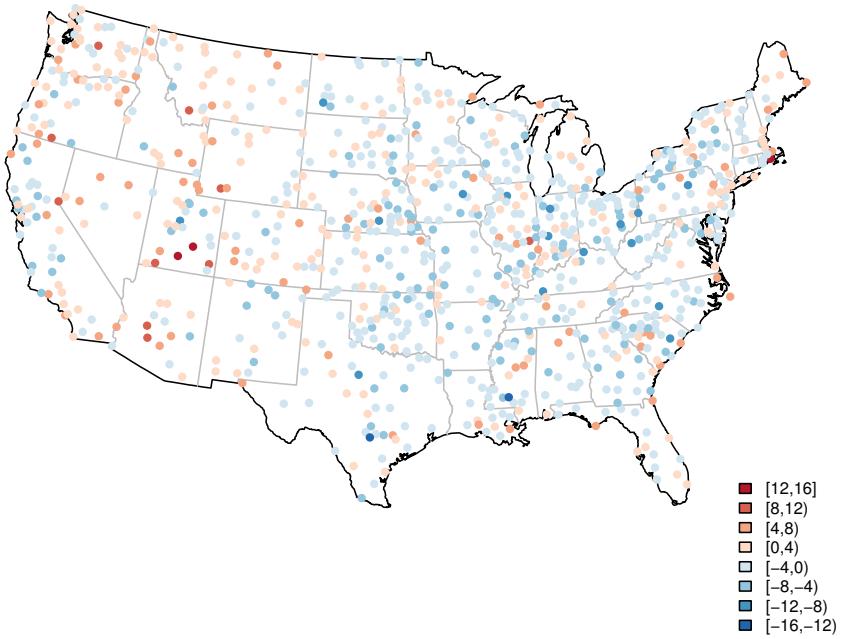

Figure 3.8: Trends of United States monthly maximum temperatures.

3.5 Comments

While one may want to select a common series starting date for all stations (say 1900) for interpretability, this issue does not greatly impact our results. In fact, most station starting dates are close to 1900. Mathematically, linear trend analyses are invariant of the starting time; thus, to minimize statistical variability, it makes sense to use the longest record possible. This issue would need to be further scrutinized should non-linear trends be considered.

An improvement to our methods would allow estimation of the changepoint times and locations in the GEV likelihood, accounting for autocorrelation and reference station aspects. This would eliminate the Gaussian analysis step to discover the station changepoint configurations. We have attempted to develop such methods but failed. Handling autocorrelation in extremes is very difficult.

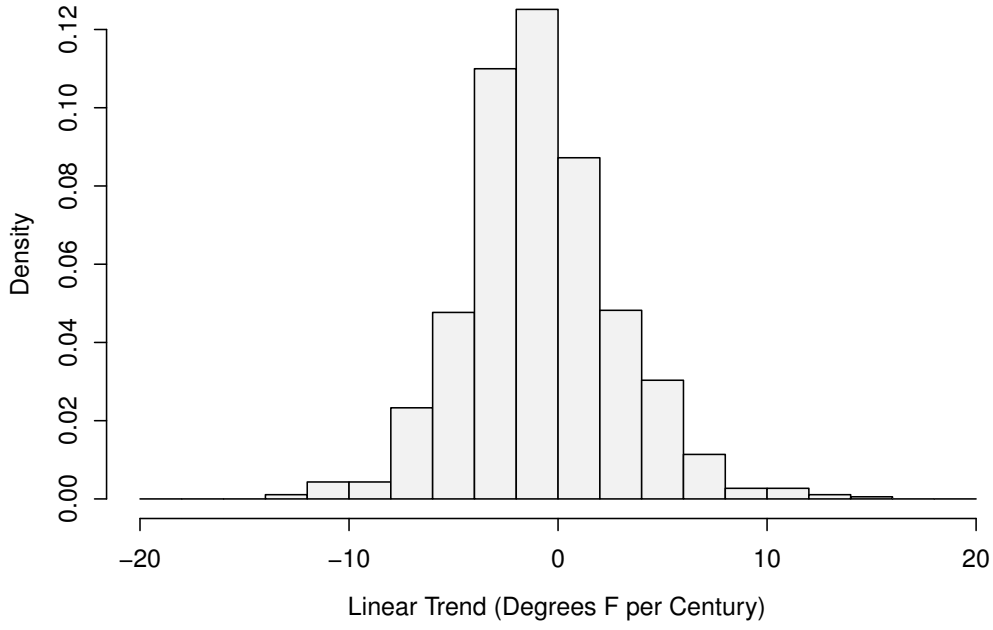

Figure 3.9: Histogram of maximum temperature trends.

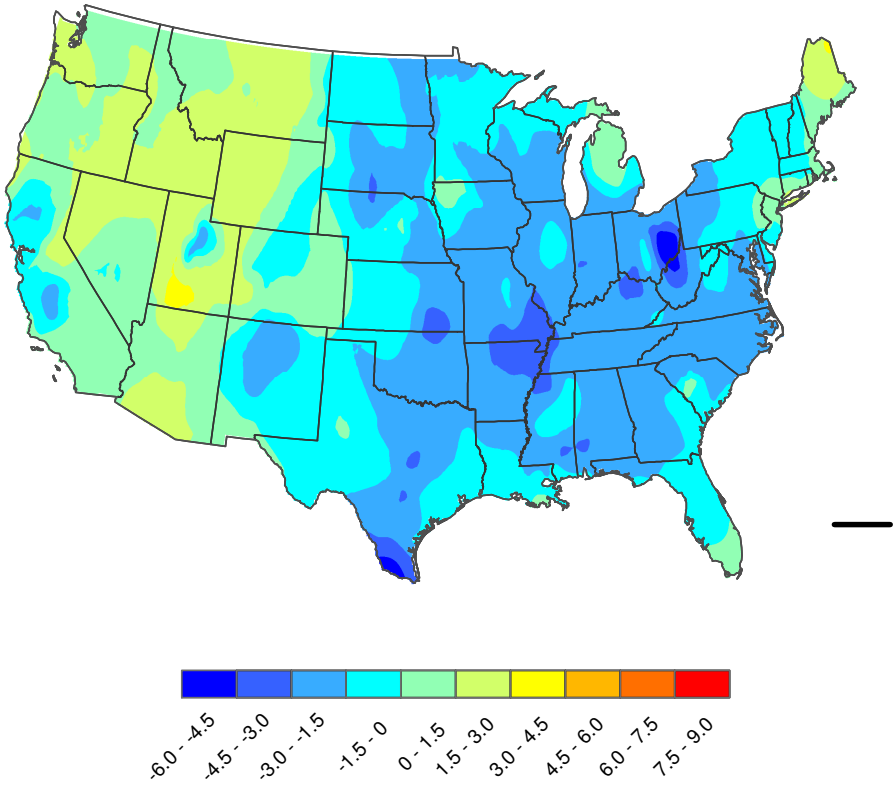

Figure 3.10: Head-banging smoothed trends of United States monthly maximum temperatures. The Eastern US shows cooling and the Western US warming.

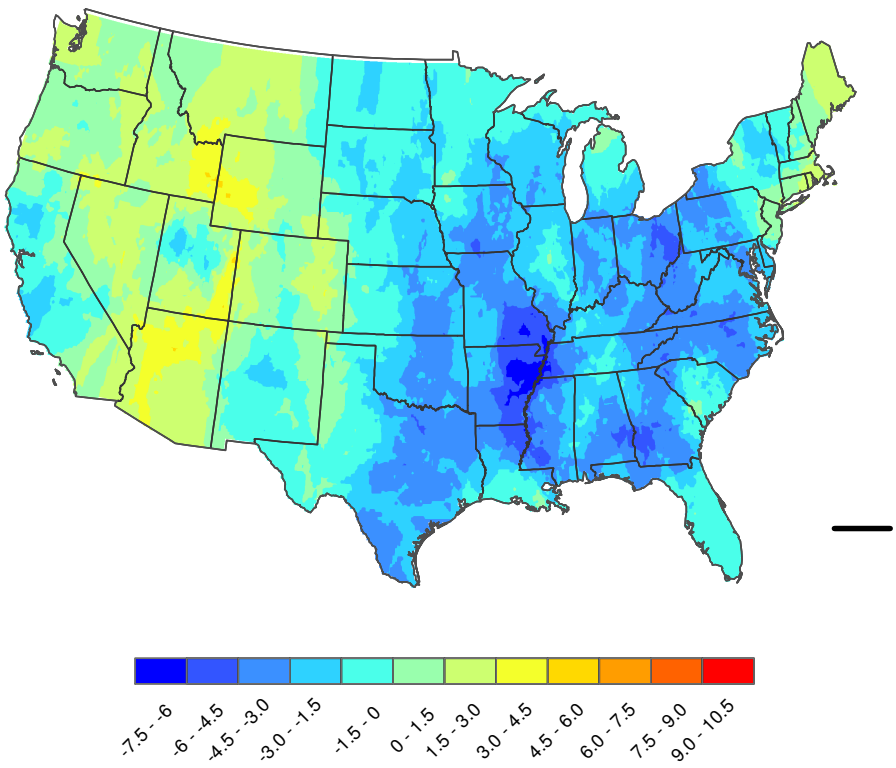

Figure 3.11: Z-scores for trends of United States maximum monthly temperatures.

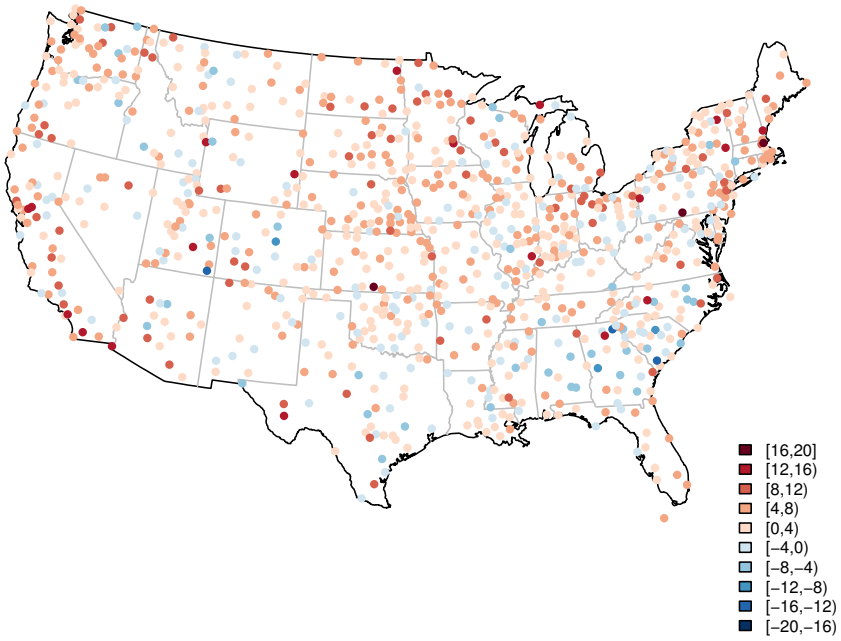

Figure 3.12: Trends of United States monthly minimum temperatures.

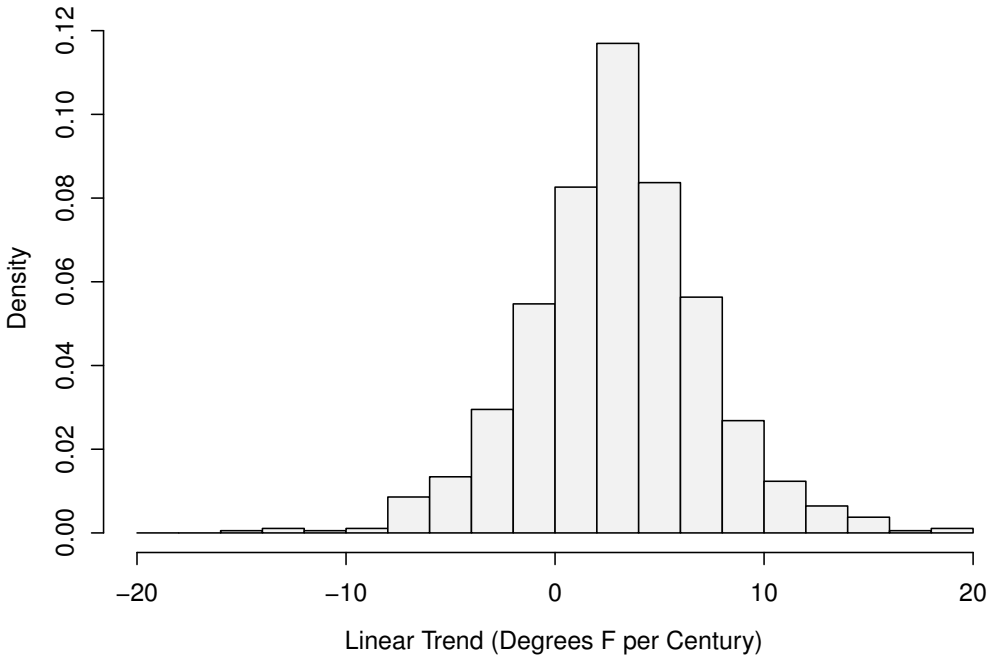

Figure 3.13: Histogram of minimum temperature trends.

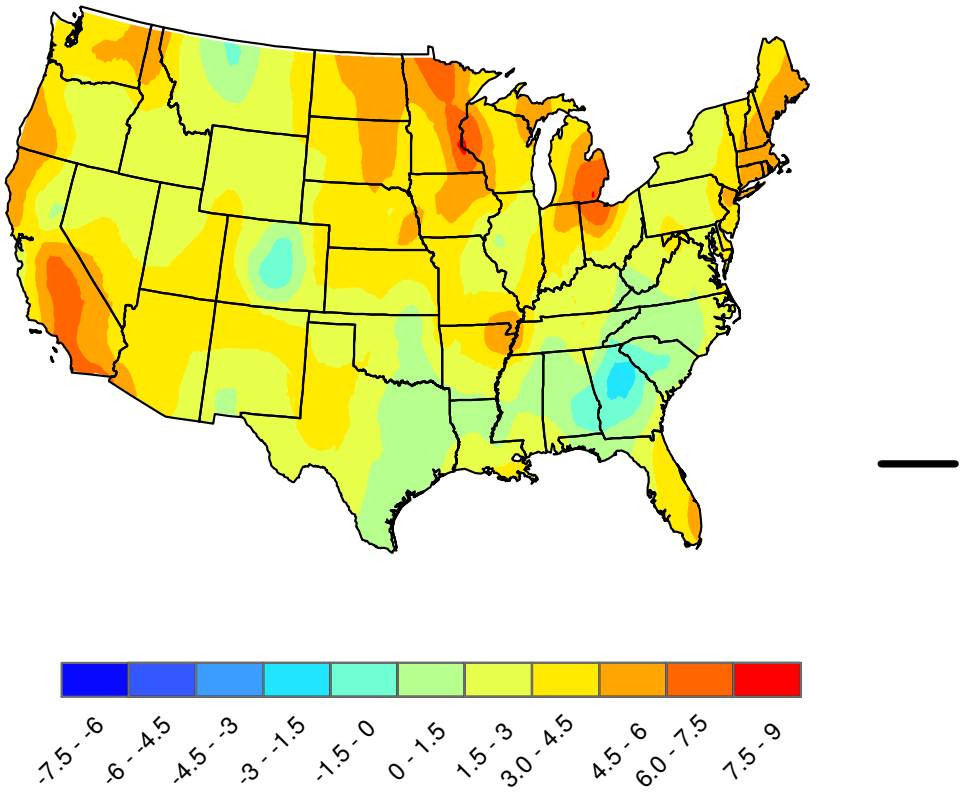

Figure 3.14: Head-banging smoothed trends of United States minimum monthly temperatures.

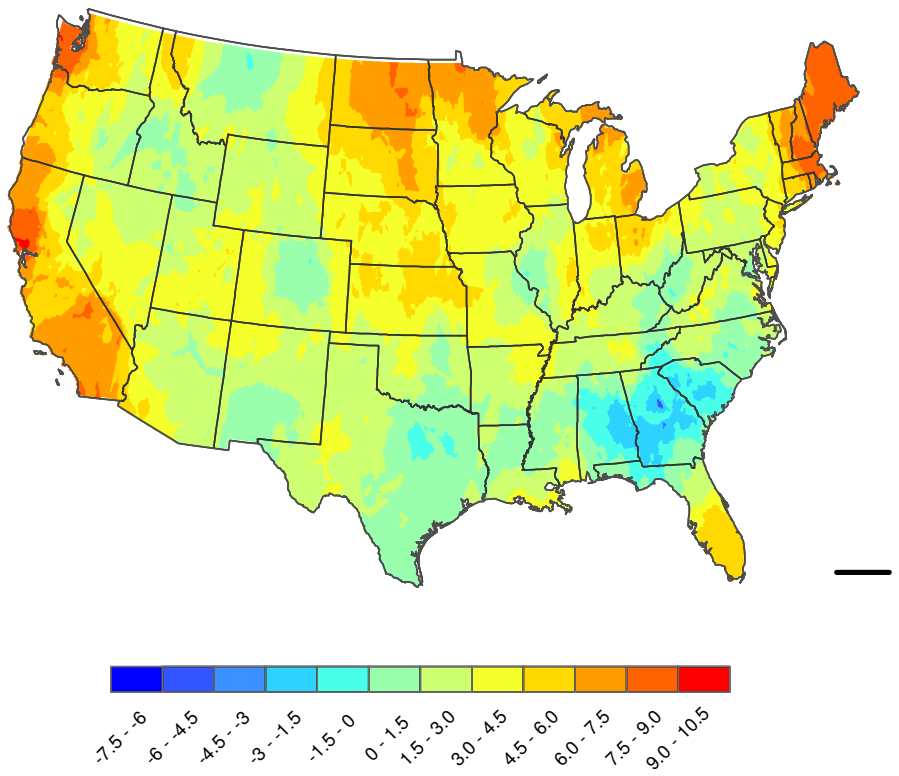

Figure 3.15: Z-scores for trends of United States minimum monthly temperatures.

Bibliography

[1] Alba, E., and J. M. Troya, 1999: A survey of parallel-distributed genetic algorithms. *Complexity*, **4**, 31-52.

[2] Alexandersson, H., and A. Moberg, 1997: Homogenization of Swedish temperature data. Part I: Homogeneity tests for linear trends. *International Journal of Climatology*, **17**, 25-34.

[3] Beasley, D., D. R. Bull, and R. R. Martin, 1993: An overview of genetic algorithm: Part 1, fundamentals. *University Computing*, **15**, 58-69.

[4] Beaulieu, C., T. B. M. J. Ourada, and O. Seidou, 2010: A Bayesian normal homogeneity test for the detection of artificial discontinuities in climatic series. *International Journal of Climatology*, doi:10.1002/joc2056.

[5] Brockwell, P. J., and R. A. Davis, 1991: *Time Series: Theory and Methods*, Second Edition, Springer: New York City.

[6] Buishand, T. A., 1982: Some methods for testing the homogeneity of rainfall records. *Journal of Hydrology*, **58**, 11-27.

[7] Burt, C. C., 2004: *Extreme Weather*. W. W. Norton & Company, 304 pp.

[8] Casella, G., and R. L. Berger, 2002: *Statistical Inference*, Second Edition, Duxbury: Australia.

[9] Caussinus, H., and O. Mestre, 2004: Detection and correction of artificial shifts in climate. *Journal of the Royal Statistical Society, Series C*, **53**, 405-425.

[10] Chen, J., and A. K. Gupta, 2000: *Parametric Statistical Change Point Analysis*. Birkhäuser: Boston.

[11] Coles, S., 2001: *An Introduction to Statistical Modelling of Extreme Values*. Springer, 224 pp.

[12] Davis, L. 1991: *Handbook of Genetic Algorithm*, Van Nostrand Reinhold: New York.

[13] Davis, R. A.,T. C. M. Lee, and G. A. Rodriguez-Yam, 2006: Structural break estimation for nonstationary time series models. *Journal of the American Statistical Association*, **101**, 223-239.

[14] DeGaetano, A. T., and R. J. Allen, 2002: Trends in Twentieth-Century temperature extremes across the United States. *Journal of Climate*, **15**, 3188-3205.

[15] DeGaetano, A. T., Allen, R. J., and K. P. Gallo, 2002: A homogenized historical extreme dataset for the United States. *Journal of Atmospheric and Oceanic Technology*, **19**, 1267-1284.

[16] Ducré-Robitaille, J-F., L. A. Vincent, and G. Boulet, 2003: Comparison of techniques for detection of discontinuities in temperatures. *International Journal of Climatology*, **23**, 1087-1101.

[17] Easterling, D. A., and T. C. Peterson, 1995: A new method for detecting undocumented discontinuities in climatological time series. *International Journal of Climatology*, **15**, 369-377.

[18] Goldberg, D.E. 1989: *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison Wesley: Reading, Massachusetts.

[19] Hansen, K. M., 1991: Head-banging: robust smoothing in the plane. *IEEE Transactions on Geoscience and Remote Sensing*, **29**, 369-378.

[20] Hansen, M. H. and B. Yu, 2001: Model selection and the principle of minimum description length. *Journal of the American Statistical Association*, **96** 746-774.

[21] Hawkins, D. M., 1976: Point estimation of the parameters of piecewise regression models. *Journal of the Royal Statistical Society, Series C*, **25**, 51-57.

[22] Hawkins, D. M., 1977: Testing a sequence of observations for a shift in location. *Journal of the American Statistical Association*, **72**, 180-186.

[23] Holland, J. H. 1975: *Adaptation in Natural and Artificial Systems*, MIT Press, Cambridge, Massachusetts.

[24] Jann, A., 2006: Genetic algorithms: Towards their use in the homogenization of climatological records. *Croatian Meteorological Journal*, **41**, 3-19.

[25] Jones, P. D., New, M., Parker, D. E., Martin, S., and I. G. Rigor, 1999: Surface air temperature and its changes over the past 150 years. *Reviews of Geophysics*, **37(2)**, 173-199.

[26] Karl, T. K., Kukla, G., Changery, M. J., Quayle, R. G., Heim Jr., R. R., Easterling, D. R., and C. B. Fu, 1991: Global warming: evidence for asymmetric diurnal temperature change. *Geophysical Research Letters*, **18(12)**, 2253-2256.

[27] Katz, R. W., and B. G. Brown, 1992: Extreme events in a changing climate: variability is more important than means. *Climatic Change*, **21**, 289-302.

[28] Kunkel, K. E., Liang, X-Z., Zhu, J., and Y. Lin, 2006: Can CGCMs simulate the twentieth-century warming hole in the central United States? *Journal of Climate*, **19**, 4137-4153.

[29] Landsea, C. W., Pielke, R. A., Mestas-Nuñez, A. M., and J. A. Knaff, 1999: Atlantic Basin hurricanes: indices of climatic changes. *Climatic Change*, **42**, 89-129.

[30] Leadbetter, M. R., Lindgren, G., and H. Rootzen, 1983: *Extremes and Related Properties of Random Sequences and Processes*. Springer-Verlag, 336 pp.

[31] Lee, T. C. M., 2001: An introduction to coding theory and the two-part minimum description length principle. *International Statistical Review*, **69**, 169-183.

[32] Li, S., and R. B. Lund, 2012: Multiple changepoint detection via genetic algorithms, *Journal of Climate*, **25**, 674-686.

[33] Lu, Q., and R. B. Lund, 2007: Simple linear regression with multiple level shifts. *Canadian Journal of Statistics*, **37**, 447-458.

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

[35] Lu, Q., Lund, R. B., and P. L. Seymour, 2005:, An update of United States temperature trends. *Journal of Climate*, **18**, 4906-4914.

[36] Lund, R. B., Hurd, H., Bloomfield, P., and R. L. Smith, 1995: Climatological time series with periodic correlation. *Journal of Climate*, **11**, 2787-2809.

[37] Lund, R. B., and J. Reeves, 2002: Detection of undocumented changepoints — a revision of the two-phase regression model. *Journal of Climate*, **15**, 2547-2554.

[38] Lund, R. B., Seymour, P. L., and K. Kafadar, 2001: Temperature trends in the United States. *Environmetrics*, **12**, 673-690.

[39] Lund, R. B., X. L. Wang, Q. Lu, J. Reeves, C. Gallagher, and Y. Feng, 2007: Changepoint detection in periodic and autocorrelated time series. *Journal of Climate*, **20**, 5178-5190.

[40] McCormick, W., and Y. Qi, 2000: Asymptotic distribution for the sum and maximum of Gaussian processes. *Journal of Applied Probability*, **8**, 958-971.

[41] Meehl, G. A., Arblaster, J. M., and G. Branstator 2012: Mechanisms contributing to the warming hole and the consequent U.S. Eastwest differential of heat extremes. *Journal of Climate*, **25**, 6394-6408.

[42] Menne, J. M. and C. N. Williams Jr., 2005: Detection of undocumented changepoints using multiple test statistics and composite reference series. *Journal of Climate*, **18**, 4271-4286.

[43] Menne, J. M. and C. N. Williams Jr., 2009: Homogenization of temperature series via pairwise comparisons. *Journal of Climate*, **22**, 1700-1717.

[44] Menne, J. M., Williams, C. N., Jr., and M. A. Palecki, 2010: On the reliability of the U.S. surface temperature record. *Journal of Geophysical Research*, **115**, D11108, doi:10.1029 /2009JD013094.

[45] Mitchell, J. M. Jr., 1953: On the causes of instrumentally observed secular temperature trends. *Journal of Applied Meteorology*, **10**, 244-261.

[46] Mooley, D. A., 1981: Applicability of the Poisson probability model to the severe cyclonic storms striking the coast around the Bay of Bengal. *Sankhyā, Series B*, **43**, 187-197.

[47] Neumann, C. J., Jarvinen, B. R., McAdie, C. J., and J. D. Elms, 1999: *Tropical Cyclones of the North Atlantic Ocean, 1871-1998.* National Climatic Data Center, Asheville, NC, pp. 206.

[48] Peterson, T. C., Karl, T. R., Jamason, P. F., Knight, R., and D. R. Easterling, 1998: First difference method: maximizing station density for the calculation of long-term global temperature change. *Journal of Geophysical Research*, **103, D20**, 25967-25974.

[49] Potter, K. W., 1981: Illustration of a new test for detecting a shift in mean in precipitation series. *Monthly Weather Review*, **109**, 2040-2045.

[50] Reeves, J., J. Chen, X. L. Wang, R. B. Lund, and Q. Lu, 2007: A review and comparison of changepoint detection techniques for climate data. *Journal of Applied Meteorology and Climatology*, **46**, 900-915.

[51] Rissanen, J., 1989: *Stochastic Complexity in Statistical Inquiry*, World Scientific, Singapore.

[52] Rissanen, J., 2007: *Information and Complexity in Statistical Modeling*, Springer, New York.

[53] Robbins, M., R. B. Lund, C. Gallagher, and Q. Lu 2011: Changepoints in the North Atlantic tropical cyclone record, *Journal of the American Statistical Association*, **106**, 89-99.

[54] Robinson, W. A., Reudy, R., and J. E. Hansen, 2002: On the recent cooling in the East-central United States. *Journal of Geophysical Research*, **107**, 4748, doi:10.1029/2001JD001577.

[55] Rodionov, S. N., 2004: A sequential algorithm for testing climate regime shifts. *Geophysical Research Letters*, **31**, L09204, doi:10.1029/2004GL019448.

[56] Solow, A. R., 1989: Reconstructing a partially observed record of tropical cyclone counts. *Journal of Climate*, **2**, 1253-1257.

[57] Thompson, M. L., and P. Guttorp, 1986: A probability model for the severe cyclonic storms striking the coast around the Bay of Bengal. *Monthly Weather Review*, **114**, 2267-2271.

[58] Van de Vyver, H., 2012: Evolution of extreme temperatures in Belgium since the 1950s. *Theoretical and Applied Climatology*, **107**, 113-129.

[59] Vincent, L. A., 1998: A technique for the identification of inhomogeneities in Canadian temperature series. *Journal of Climate*, **11** 1094-1104.

[60] Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences*, Second Edition, Academic Press: Amsterdam.