Multiple Changepoint Detection via Genetic Algorithms

SHANGHONG LI AND ROBERT LUND

Department of Mathematical Sciences, Clemson University, Clemson, South Carolina

(Manuscript received 7 September 2010, in final form 26 June 2011)

ABSTRACT

This paper studies genetic algorithms as a means of estimating the number of changepoints and their locations in a climatic time series. Such methods bypass classic 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. The general objective functions allow for correlated data, reference station aspects, and/or nonnormal 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 randomly walks through a subset of good models in an intelligent manner. The methods are applied in the analysis of 173 yr of annual precipitation measurements from New Bedford, Massachusetts, and the North Atlantic basin tropical cyclone record.

1. 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 nonclimatic 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 6 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 owing to the huge number of admissible multiple changepoint configurations. This paper proposes an alternative to subsegmentation via genetic algorithms (GAs).

Genetic algorithms, 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. Genetic algorithms allow us to estimate the number of changepoints and their locations with minimal computational demands and without subsegmenting. This paper has a tutorial aspect in that

Corresponding author address: Robert Lund, Department of Mathematical Sciences, Clemson University, Clemson, SC 29634.
E-mail: lund@clemson.edu

DOI: 10.1175/2011JCLI4055.1

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 study annual data only; daily and/or monthly methods are also worth pursuing. 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 (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.

The rest of this paper proceeds as follows. The next section introduces a MDL penalized likelihood objective function; outside of Lu et al. (2010), MDL criterion have not been used in climatic changepoint research to date. Section 3 develops the likelihood and penalty terms in the objective function. Section 4 then devises a GA that is capable of finding the best changepoint configuration (model) among all possible models. Section 5 presents a short simulation study on series whose statistical properties are known. The paper closes with application to two datasets. First, section 6 examines a 173-yr series of annual precipitation from New Bedford, Massachusetts. This series is examined with and without a reference series and autocorrelation aspects. Second, section 7 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. The MDL setup

Our methods here seek a model (changepoint numbers and times) that minimizes an objective function. While the objective function needs to be tailored to individual situations, all objective functions here will minimize minimum description length scores of the form:

$$\text{MDL} = -\log_2(L_{\text{opt}}) + P. \tag{2.1}$$

In (2.1), 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 $-\log_2(L_{\text{opt}})$ becomes smaller; however, if adding additional parameters does not decrease $-\log_2(L_{\text{opt}})$ 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 the Akaike 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).

3. Objective function development

a. Annual precipitation series

As a first task, we develop the likelihood part of (2.1) 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:

$$f(x) = \frac{\exp\{-[\ln(x) - \mu]^2/2\sigma^2\}}{x\sigma\sqrt{2\pi}}, \quad x \geq 0,$$

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:

$$L(\mu, \sigma^2) = \prod_{t=1}^{N} f(X_t) = \frac{\exp\left\{-\frac{1}{2\sigma^2}\sum_{t=1}^{N}[\ln(X_t) - \mu]^2\right\}}{(\sigma\sqrt{2\pi})^N\left(\prod_{t=1}^{N} X_t\right)}. \tag{3.1}$$

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 (3.1) and setting the resulting expressions to zero, we obtain likelihood estimates of μ and σ^2:

$$\hat{\mu} = \frac{1}{N}\sum_{t=1}^{N}\ln(X_t), \quad \hat{\sigma}^2 = \frac{1}{N}\sum_{t=1}^{N}[\ln(X_t) - \hat{\mu}]^2. \tag{3.2}$$

The calculus computations are tedious but straightforward; the reader is referred to Casella and Berger (2002) for likelihood basics. Substituting the optimizing values of $\hat{\mu}$ and $\hat{\sigma}^2$ into (3.1) gives the optimal likelihood score:

$$\begin{aligned} -\ln(L_{\text{opt}}) &= -\ln[L(\hat{\mu}, \hat{\sigma}^2)] \\ &= \frac{N}{2}[1 + \ln(2\pi) + \ln(\hat{\sigma}^2)] + \sum_{t=1}^{N}\ln(X_t). \end{aligned} \tag{3.3}$$

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 $\tau_1 < \tau_2 < \cdots < \tau_m$, 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 $l = 1, 2, \ldots, m + 1$. 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

$$f(x) = \frac{\exp\{-[\ln(x) - \mu_{r(t)}]^2/2\sigma^2\}}{x\sigma\sqrt{2\pi}}, \quad x > 0,$$

and

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

A derivation similar to that producing (3.2) gives

$$\hat{\mu}_l = \frac{1}{\tau_i - \tau_{i-1}}\sum_{t\in R_l}\ln(X_t), \quad \hat{\sigma}^2 = \frac{1}{N}\sum_{t=1}^{N}[\ln(X_t) - \hat{\mu}_{r(t)}]^2.$$

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

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

The optimal likelihood achieved is exactly as listed in (3.3), 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 minimum description length 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 $-\log_2(L_{\text{opt}})$. 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 $\log_2(\tau_l - \tau_{l-1})/2$ as it is estimated from the $\tau_l - \tau_{l-1}$ 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,

$$P = \frac{3\log_2(N)}{2} + \sum_{i=1}^{m+1} \frac{\log_2(\tau_i - \tau_{i-1})}{2} + \log_2(m) + \sum_{i=2}^{m} \log_2(\tau_i), \quad (3.4)$$

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 (2.1) as

$$\text{MDL} = \frac{N}{2}\ln(\hat{\sigma}^2) + \sum_{i=1}^{m+1} \frac{\ln(\tau_i - \tau_{i-1})}{2} + \ln(m) + \sum_{i=2}^{m} \ln(\tau_i). \quad (3.5)$$

We have made some simplifications in obtaining (3.5) that make it differ from the direct sum of (3.3) and (3.4). 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 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 lognormal 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

$$\text{L}(\mu_1, \mu_2, \ldots, \mu_{m+1}, \phi, \sigma^2) = (2\pi)^{-N/2}\left(\prod_{t=1}^{N} v_t\right)^{-1/2} \exp\left[-\frac{1}{2}\sum_{t=1}^{N}\frac{(S_t - \hat{S}_t)^2}{v_t}\right]. \quad (3.6)$$

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; Lu et al. 2010) and are similar to (3.6). Likelihood estimators of the mean for the lth segment are

$$\hat{\mu}_l = \frac{1}{\tau_l - \tau_{l-1}}\sum_{t \in R_l}^{N} S_t.$$

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

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

Substituting these values into (3.6) gives

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

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 minimum description length:

$$\text{MDL} = \frac{N}{2}\ln(\hat{\sigma}^2) + \sum_{i=1}^{m+1} \frac{\ln(\tau_i - \tau_{i-1})}{2} + \ln(m) + \sum_{i=2}^{m} \ln(\tau_i). \tag{3.7}$$

b. 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:

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

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

$$\hat{\lambda}_l = \frac{1}{\tau_l - \tau_{l-1}} \sum_{t\in R_l} X_t.$$

Substituting these values back into the likelihood gives

$$-\ln(L_{\text{opt}}) = -\ln[L(\hat{\lambda})] = -\sum_{l=1}^{m+1}\left[\ln(\hat{\lambda}_l)\sum_{t\in R_l} X_t\right] + \sum_{t=1}^{N} X_t + \sum_{t=1}^{N} \ln(X_t!).$$

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:

$$\text{MDL}(m, \tau_1, \ldots, \tau_m) = -\sum_{l=1}^{m+1}\left[\ln(\hat{\lambda}_l)\sum_{t\in R_l} X_t\right] + \sum_{i=1}^{m+1} \frac{\ln(\tau_i - \tau_{i-1})}{2} + \sum_{i=2}^{m} \ln(\tau_i) + \ln(m). \tag{3.8}$$

4. 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 6, this amounts to evaluating 1.2×10^{52} different MDL scores, a daunting task on even the fastest computer. This is where genetic algorithms will prove useful.

Genetic algorithms (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. Genetic algorithms are so-named because they contain aspects of genetic selection/evolution. In particular, each possible parameter configuration will be encoded as a "chromosome." Genetic algorithms 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 (i.e., they better optimize the objective function) are more likely to have children, thus mimicking 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 (3.5) 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 U.S. 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:

$$S_i \Big/ \sum_{j=1}^{200} S_j. \tag{4.1}$$

Once a mother is selected, the father is probabilistically selected by ranking the remaining 199 individuals akin to (4.1). 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 ½. 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 time 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 nonchangepoint 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

TABLE 1. Empirical proportions of estimated changepoint numbers. The correct value of m is zero.

m	Percent
0	99.0
1	0.4
2	0.5
3+	0.1

TABLE 2. Empirical proportions of estimated changepoint numbers. The correct value of m is 3.

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

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 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.

5. 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 6. In particular, a lognormal 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 6 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 1 displays the proportion of simulations that 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 4 of the simulations, one changepoint was estimated. Overall, the algorithm seems to have a very low false alarm rate. As only 10 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 for times 1–49, rises to 7.0 for times 50–99, increases to 7.2 at times 100–149, and increases again to 7.4 for times 150–200. This configuration represents mean shifts in one direction (increasing), all having the same shift magnitudes (on the log scale). Table 2 shows empirical proportions of estimates of m. The methods estimate the correct 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, Fig. 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. Figure 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

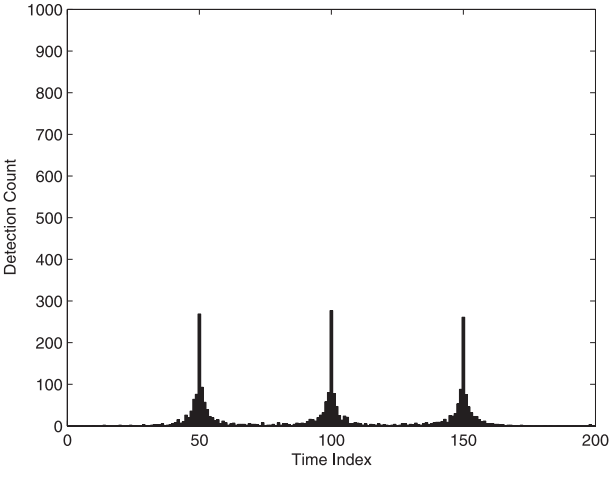

FIG. 1. Count detection histogram. The detected changepoint times cluster around their true values of 50, 100, and 150.

TABLE 3. Empirical proportions of estimated changepoint numbers. The correct value of m is 3.

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

TABLE 4. Empirical proportions of estimated changepoint numbers. The correct value of m is 2.

m	Percent
0	0.0
1	7.5
2	90.7
3+	1.8

shift magnitudes are altered. We start with a series whose μ_t is 6.8 for times 1–24, shifts up to 7.0 for times 25–74, moves downward to 6.6 for times 75–99, and then shifts upward to the initial level of 6.8 for times 100–200. Table 3 shows the empirical proportions of estimated changepoint numbers and has a similar structure to the numbers reported in Table 2. Figure 2 displays a count histogram akin to Fig. 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 7. Here, we take $n = 160$ and superimpose two changepoints. Specifically, we start with a mean of 7.0 for times 1–79, shift upward to 10 for times 80–145, and then move to 15 at times 145–160. Table 4 reports estimated values of m. The correct order $m = 2$ was estimated in 90.7% of the 1000 simulations. Figure 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.

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.

6. The New Bedford series

Figure 4 plots a $N = 173$-yr annual precipitation series from New Bedford, Massachusetts, during 1818–1990. For this data, we first ran a genetic algorithm 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 (3.5). 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. 5 against the data and appears visually reasonable. The optimal segmentation has a short segment containing only 1965 and 1966, suggesting perhaps an

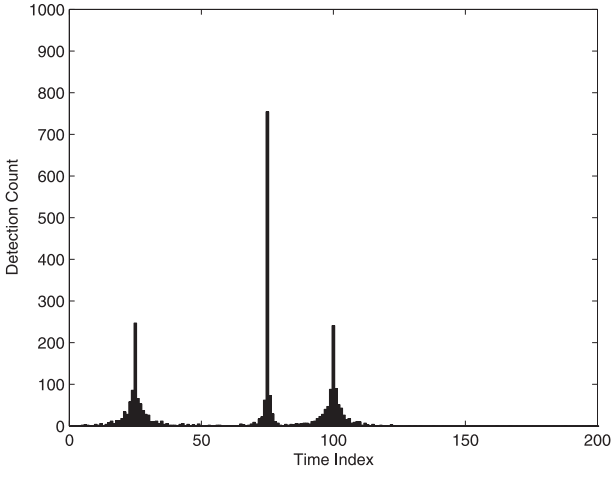

FIG. 2. Count detection histogram. The detected changepoint times cluster around their true values of 25, 75, and 100.

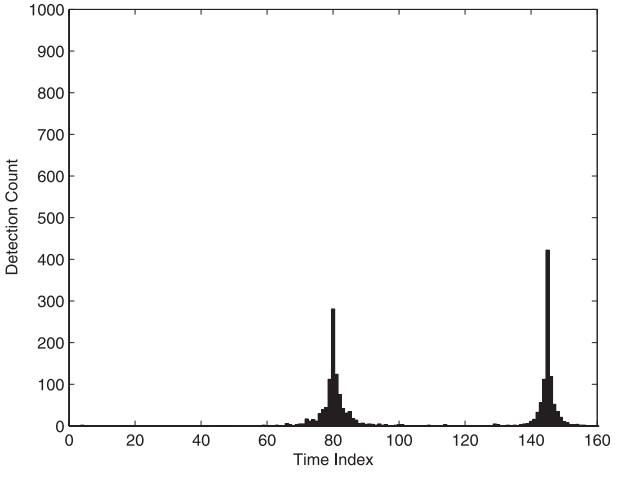

FIG. 3. Poisson count detection histogram. The detected changepoint times cluster around their true values of 80 and 145.

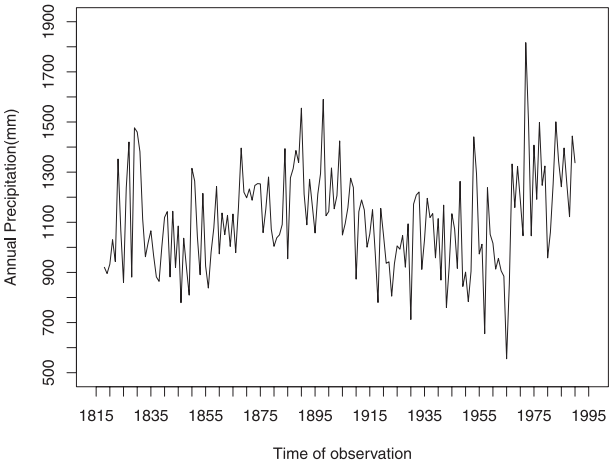

FIG. 4. Annual precipitation at New Bedford, Massachusetts.

outlier. The legitimacy of the small precipitation in 1966 may be questioned; however, the station's metadata 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 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 6 shows optimum MDL scores for various numbers of model segments

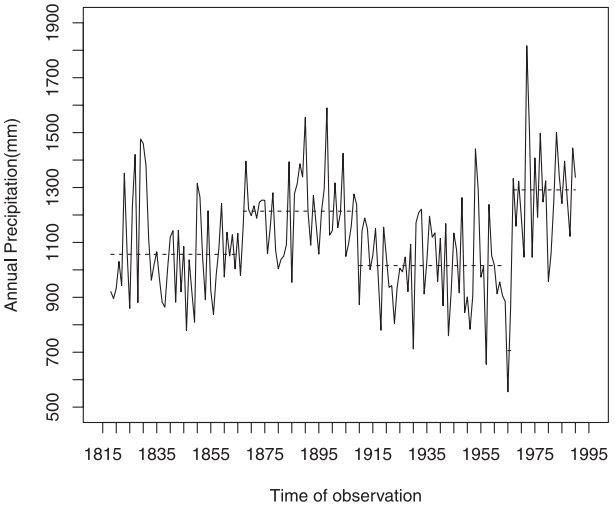

FIG. 5. Optimal model with data superimposed. The optimal model has four changepoints. The fitted mean configuration (dashed) follows the data fluctuations well.

TABLE 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

(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 the GA (which has four changepoints). In fact, the GA has selected (exactly) the best five-segment model identified in Table 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 nonzero. A GA was run to minimize (3.7) 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 $\hat{\sigma}^2 = 0.022$.

TABLE 6. Optimum MDL scores for various numbers of segments.

No. of 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

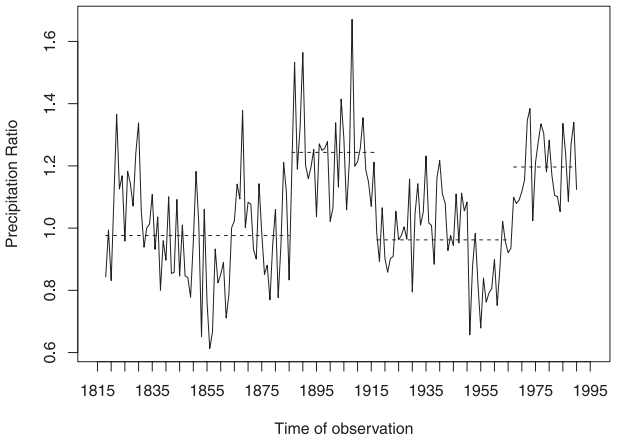

FIG. 6. Optimal MDL model for precipitation ratios with data superimposed. Three changepoint times are estimated. The fitted mean configuration (dashed) follows the data fluctuations well.

A reference series from Boston, Massachusetts, is available to help make conclusions. The New Bedford to Boston ratios were computed and are displayed in Fig. 6. A genetic algorithm was then constructed to minimize the MDL score in (3.7). This model allows for AR(1) autocorrelation. The genetic parameters used are $p_m = 0.03$, $p_i = 0.06$, and a generation size of 200. Table 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. 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 1883, which is close to the 1886 changepoint time flagged by the MDL segmenter. An SNHT analysis of data points 1818–82 concludes another changepoint at time 1836. The 1818–35 segment tests positively for another changepoint at time 1821, while the 1836–82 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 1917–90 segment is found to have a changepoint at time 1967, which is (exactly) a time the MDL segmentation flagged. The 1967–90 segment tests as homogeneous, but the 1917–66 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 segmentation is plotted against the data in Fig. 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

TABLE 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	Change point
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

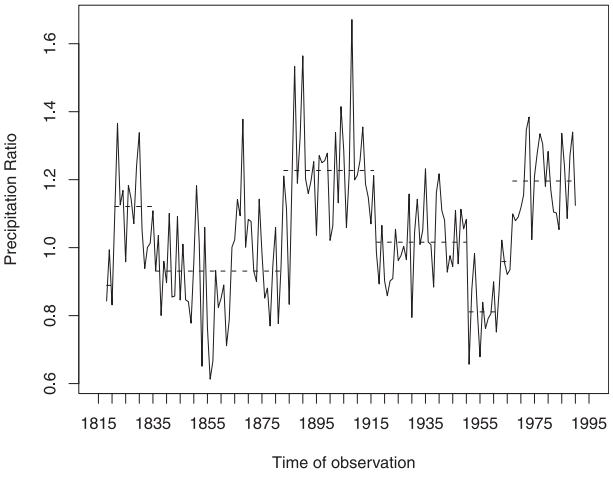

FIG. 7. Model for precipitation ratios estimated via SNHT segmentation. This segmentation estimates seven changepoints.

estimated MDL configuration has fewer changepoints than the estimated SNHT configuration. Part of this aspect is likely explained by the significant nonzero 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 metadata 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 metadata 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 (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.

7. The North Atlantic tropical cyclone record

Our second application examines the North Atlantic basin's annual tropical cyclone counts from 1851 to 2009. Here, we seek to identify times of statistical discontinuities in the record. The counts are plotted in Fig. 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 dataset, which is available on NOAA's Web site (see online at http://www.aoml.noaa.gov/hrd/data_sub/hurdat.html). 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) owing to sparse population along coastlines. Also, as Landsea et al. (1999) and Neumann et al. (1999) observe, aircraft reconnaissance toward the end of World War II (around 1944) improved detection of nonlandfalling 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.

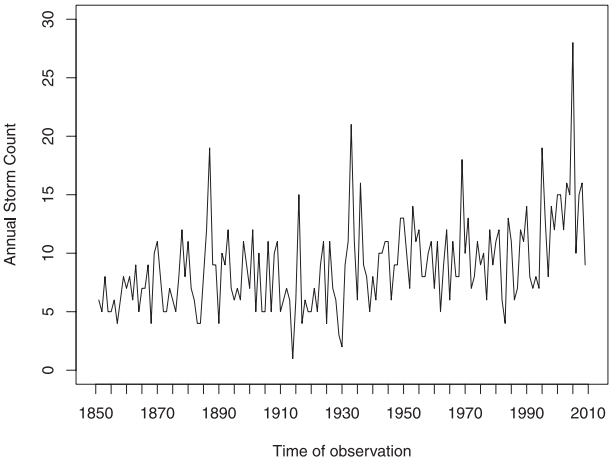

FIG. 8. Annual Atlantic basin tropical cyclone counts.

Obviously, we do not have a reference series for this data. Also no definitive metadata 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 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 (3.8).

Figure 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 200. Here, the MDL judges three segments as optimal: one for 1851–1930, one for 1931–94, and one for 1995–2009. The optimal MDL was -3131.40. This segmentation agrees exactly with that in Robbins et al. (2011). Table 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.

Overall, the cyclone counts appear to be increasing. The authors are unaware of any data collection changes

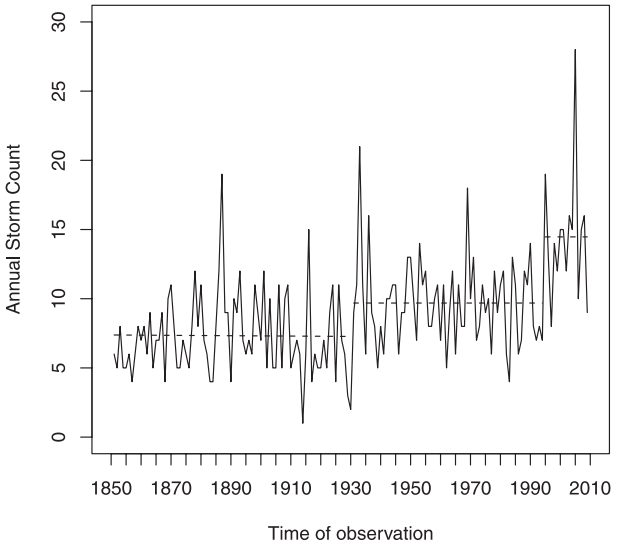

FIG. 9. Optimal MDL model for cyclone count data with data superimposed. There are two estimated changepoint times and the fitted mean shift configuration (dashed) follows the data fluctuations well.

that explain the 1995 changepoint. The 1931 changepoint is perhaps explained by the onset of aircraft surveillance, although it seems to occur about 10 yr too early.

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

8. Comments

This paper 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, Massachusetts, annual precipitation series and the annual North Atlantic basin tropical cyclone counts.

TABLE 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	Change point
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

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 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.

Acknowledgments. The authors acknowledge National Science Foundation support from Grant DMS 0905570. Comments from three referees substantially improved this paper.

REFERENCES

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

Alexandersson, H., and A. Moberg, 1997: Homogenization of Swedish temperature data. Part I: Homogeneity tests for linear trends. *Int. J. Climatol.,* **17,** 25–34.

Beasley, D., D. R. Bull, and R. R. Martin, 1993: An overview of genetic algorithm. Part 1: Fundamentals. *Univ. Comput.,* **15,** 58–69.

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. *Int. J. Climatol.,* **30,** 2342–2357, doi:10.1002/joc.2056.

Brockwell, P. J., and R. A. Davis, 1991: *Time Series: Theory and Methods*. 2nd ed. Springer, 584 pp.

Buishand, T. A., 1982: Some methods for testing the homogeneity of rainfall records. *J. Hydrol.,* **58,** 11–27.

Casella, G., and R. L. Berger, 2002: *Statistical Inference*. 2nd ed. Duxbury Press, 7000 pp.

Caussinus, H., and O. Mestre, 2004: Detection and correction of artificial shifts in climate. *J. Roy. Stat. Soc.,* **53C,** 405–425.

Chen, J., and A. K. Gupta, 2000: *Parametric Statistical Change Point Analysis*. Birkhäuser, 192 pp.

Davis, L., 1991: *Handbook of Genetic Algorithm*. Van Nostrand Reinhold, 385 pp.

Davis, R. A., T. C. M. Lee, and G. A. Rodriguez-Yam, 2006: Structural break estimation for nonstationary time series models. *J. Amer. Stat. Assoc.,* **101,** 223–239.

Ducré-Robitaille, J.-F., L. A. Vincent, and G. Boulet, 2003: Comparison of techniques for detection of discontinuities in temperatures. *Int. J. Climatol.,* **23,** 1087–1101.

Easterling, D. A., and T. C. Peterson, 1995: A new method for detecting undocumented discontinuities in climatological time series. *Int. J. Climatol.,* **15,** 369–377.

Goldberg, D. E., 1989: *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley, 432 pp.

Hansen, M. H., and B. Yu, 2001: Model selection and the principle of minimum description length. *J. Amer. Stat. Assoc.,* **96,** 746–774.

Hawkins, D. M., 1976: Point estimation of the parameters of piecewise regression models. *J. Roy. Stat. Soc.,* **25C,** 51–57.

——, 1977: Testing a sequence of observations for a shift in location. *J. Amer. Stat. Assoc.,* **72,** 180–186.

Holland, J. H., 1975: *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. University of Michigan Press, 183 pp.

Jann, A., 2006: Genetic algorithms: Towards their use in the homogenization of climatological records. *Croatian Meteor. J.,* **41,** 3–19.

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

Lee, T. C. M., 2001: An introduction to coding theory and the two-part minimum description length principle. *Int. Stat. Rev.,* **69,** 169–183.

Lu, Q., R. B. Lund, and T. C. M. Lee, 2010: An MDL approach to the climate segmentation problem. *Ann. Appl. Stat.,* **4,** 299–319.

Lund, R. B., and J. Reeves, 2002: Detection of undocumented changepoints—A revision of the two-phase regression model. *J. Climate,* **15,** 2547–2554.

——, X. L. Wang, Q. Lu, J. Reeves, C. Gallagher, and Y. Feng, 2007: Changepoint detection in periodic and autocorrelated time series. *J. Climate,* **20,** 5178–5190.

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

——, and ——, 2009: Homogenization of temperature series via pairwise comparisons. *J. Climate,* **22,** 1700–1717.

Mitchell, J. M., Jr., 1953: On the causes of instrumentally observed secular temperature trends. *J. Meteor.,* **10,** 244–261.

Mooley, D. A., 1981: Applicability of the Poisson probability model to the severe cyclonic storms striking the coast around the Bay of Bengal. *Sankhya,* **43B,** 187–197.

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

Potter, K. W., 1981: Illustration of a new test for detecting a shift in mean in precipitation series. *Mon. Wea. Rev.,* **109,** 2040–2045.

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. *J. Appl. Meteor. Climatol.,* **46,** 900–915.

Rissanen, J., 1989: *Stochastic Complexity in Statistical Inquiry*. World Scientific, 250 pp.

——, 2007: *Information and Complexity in Statistical Modeling*. Springer, 152 pp.

Robbins, M., R. B. Lund, C. Gallagher, and Q. Lu, 2011: Changepoints in the North Atlantic tropical cyclone record. *J. Amer. Stat. Assoc.,* **106,** 89–99.

Rodionov, S. N., 2004: A sequential algorithm for testing climate regime shifts. *Geophys. Res. Lett.,* **31,** L09204, doi:10.1029/2004GL019448.

Solow, A. R., 1989: Reconstructing a partially observed record of tropical cyclone counts. *J. Climate,* **2,** 1253–1257.

Thompson, M. L., and P. Guttorp, 1986: A probability model for the severe cyclonic storms striking the coast around the Bay of Bengal. *Mon. Wea. Rev.,* **114,** 2267–2271.

Vincent, L. A., 1998: A technique for the identification of inhomogeneities in Canadian temperature series. *J. Climate,* **11,** 1094–1104.

Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences*. 2nd ed. Academic Press, 648 pp.