*The Annals of Applied Statistics*
2010, Vol. 4, No. 1, 299–319
DOI: 10.1214/09-AOAS289


# AN MDL APPROACH TO THE CLIMATE SEGMENTATION PROBLEM[1]

BY QIQI LU, ROBERT LUND AND THOMAS C. M. LEE

*Mississippi State University, Clemson University, and Colorado State University and Chinese University of Hong Kong*

This paper proposes an information theory approach to estimate the number of changepoints and their locations in a climatic time series. A model is introduced that has an unknown number of changepoints and allows for series autocorrelations, periodic dynamics, and a mean shift at each changepoint time. An objective function gauging the number of changepoints and their locations, based on a minimum description length (MDL) information criterion, is derived. A genetic algorithm is then developed to optimize the objective function. The methods are applied in the analysis of a century of monthly temperatures from Tuscaloosa, Alabama.

**1. Introduction.** Changes in station instrumentation, location, or observer can often induce artificial discontinuities into climatic time series. For example, United States temperature recording stations average about six station relocation and instrumentation changes over a century of operation [Mitchell (1953)]. Many of these changepoint times are documented in station histories; however, other changepoint times are unknown for a variety of reasons. Even when a changepoint time is known, one may still question whether the change instills a mean shift in series observations. This paper proposes an information based approach to the multiple changepoint identification (segmentation) problem.

Our methods are specifically tailored to climatic time series in that they allow for periodicities and autocorrelations. Multiple changepoint detection procedures have been studied under the assumption that the series is driven by independent and identically distributed errors [Braun and Müller (1998), Caussinus and Mestre (2004), Menne and Williams (2005)]. This is unrealistic in climate settings where observations display moderate to strong serial autocorrelation. Ignoring autocorrelations can drastically alter changepoint inferences, as positive autocorrelation can be easily mistaken for mean shifts [see Berkes et al. (2006) and Lund et al. (2007)]. Multiple changepoint methods for time series data represent a very active area of current research [Davis, Lee, and Rodriguez-Yam (2006), Fearnhead (2006)]. Series recorded daily or monthly also display periodic dynamics.

Received December 2008; revised September 2009.

[1]Supported by NSF Grants DMS-07-07037 and DMS-09-05570 and Hong Kong Research Grant Council Grant CERG 401507.

*Key words and phrases.* Changepoints, genetic algorithm, level shifts, minimum description length, periodic autoregression, time series.

Our methods allow for seasonality by employing a time series regression model with periodic features. In short, this paper develops a multiple changepoint segmenter that applies to a variety of realistic climate series.

The rest of this paper is organized as follows. Section 2 introduces the time series regression model that underlies our work. Section 3 develops an objective function for the model. The objective function is a penalized likelihood whose penalty is based on the minimum description length (MDL) principle. This modifies Caussinus' and Mestre's (2004) model to allow for autocorrelation, seasonal effects, and also changes their likelihood penalty to an MDL-based penalty. Each segment of our model is allowed to have a distinct mean, but the autocovariance structure of each segment is constrained to be the same. Section 4 presents a genetic-type algorithm capable of optimizing the objective function to obtain estimates of the changepoint numbers, locations, and the time series regression parameters. Section 5 presents a short simulation study for feel. Section 6 applies the methods to a century of monthly temperatures from Tuscaloosa, Alabama and Section 7 concludes with comments.

**2. Model description.** The object under study is a time series $\{X_t\}$ governed by periodic errors and multiple level shifts. The period of the series is $T$ and is assumed known. The series observation during season $\nu$, $1 \le \nu \le T$, of the $(n+1)$st cycle is denoted by $X_{nT+\nu}$. The time-homogeneous and periodic notation $\{X_t\}$ and $\{X_{nT+\nu}\}$ are used interchangeably, the latter to emphasize seasonality. We index the first data cycle with $n = 0$ so that the first observation is indexed by unity. For simplicity, we take $d$ complete cycles of observations; specifically, the observed data are ordered as $X_1, \ldots, X_N$ and $d = N/T$ is assumed a natural number.

The model driving our work is a simple linear regression in a periodic environment:

$$X_{nT+\nu} = \mu_\nu + \alpha(nT + \nu) + \delta_{nT+\nu} + \varepsilon_{nT+\nu}. \tag{2.1}$$

In (2.1) $\alpha$ is a linear trend parameter that is assumed time homogeneous for simplicity; $\mu_\nu$ is the season $\nu$ location parameter (a detrended mean in the absence of changepoints). The errors $\{\varepsilon_t\}$ have zero mean and are a periodically stationary series with period $T$ in that

$$\mathrm{Cov}(\varepsilon_t, \varepsilon_s) = \mathrm{Cov}(\varepsilon_{t+T}, \varepsilon_{s+T}) \tag{2.2}$$

for all integers $t$ and $s$. Many climatic series have periodic second moments in the sense of (2.2). For a sample of size $N$ with $m < N$ changepoints, the ordered times of the changepoints are denoted by $1 < \tau_1 < \tau_2 < \cdots < \tau_m \le N$. The number of changepoints and the changepoint times are considered unknown. There are $m + 1$ different segments (regimes) during the observation record. At each changepoint

time, our model allows for a mean shift in the observations. Such a structure is described by

$$\delta_t = \begin{cases} \Delta_1, & 1 \le t < \tau_1, \\ \Delta_2, & \tau_1 \le t < \tau_2, \\ \vdots & \vdots \\ \Delta_{m+1}, & \tau_m \le t < N+1. \end{cases}$$

For parameter identifiability, we take $\Delta_1 = 0$; otherwise, the $\Delta_i$'s and $\mu_\nu$'s would become confounded. For a fixed $N$, the mean component $E[X_{nT+\nu}]$ in (2.1) depends on the $T + 1 + m$ parameters $\mu_1, \ldots, \mu_T$, $\alpha$, and $\Delta_2, \ldots, \Delta_{m+1}$. Generalizations of (2.1) are mentioned in Section 3 when we derive MDL codelengths.

To describe the time series component $\{\varepsilon_{nT+\nu}\}$, we use a causal periodic autoregression of order $p$ [PAR($p$)]. Such errors are the unique (in mean square) solution to the periodic linear difference equation

$$\varepsilon_{nT+\nu} = \sum_{k=1}^{p} \phi_k(\nu)\varepsilon_{nT+\nu-k} + Z_{nT+\nu}. \tag{2.3}$$

Here, $\{Z_t\}$ is zero mean periodic white noise with variance $\sigma^2(\nu)$ during season $\nu$. Solutions to (2.3) are indeed periodic with period $T$ in the sense of (2.2). PAR models are dense in the set of short memory periodic time series and parsimoniously describe many such series; explicit expressions for many time series quantities are available for PARs.

In many applications, reference series are available. A reference series is a series of the same genre as the series to be studied (the target series) that serves to aid changepoint identification. For example, with the Tuscaloosa temperatures examined later, series from nearby Greensboro AL, Selma AL, and Aberdeen MS are available over the same period of record. By constructing a target minus reference difference series, mean shifts induced by changepoints are sometimes illuminated. When the reference series is highly positively correlated with the target series, the target minus reference series will have smaller autocorrelations than the target series at all lags (this happens when the target and reference series have the same periodic autocovariance structure and the correlation between these two series exceeds $1/2$ at all times). Also, the linear trend assumption is typically more plausible for target minus reference differences than the target series, as long-memory and other nonlinear features can be eliminated in the subtraction. Moreover, the seasonal mean cycle is frequently reduced or altogether eliminated in target minus reference series. Drawbacks with reference series lie with additional undocumented changepoints that the reference series may introduce. Algorithms aimed at resolving which series among target and multiple references is responsible for any found changepoints are now available [see Menne and Williams (2005, 2009)], but these works do not consider seasonal features or autocorrelated errors.

Note that the difference of two series governed by (2.1) again lies in (2.1). Hence, in the next three sections we simply consider a single series satisfying (2.1). Reference series will return in Section 6.

The parameters in the model will become important later. The PAR($p$) model, including the $T$ white noise variance parameters, has $(p+1)T$ autocovariance parameters. For a fixed $m$, there are also the changepoint times $\tau_1, \ldots, \tau_m$ and the mean shifts $\Delta_2, \ldots, \Delta_{m+1}$. Finally, a trend component $\alpha$ and the seasonal means $\mu_1, \ldots, \mu_T$ are present. Hence, given $p$ and $m$, there are $2m+1+(p+2)T$ model parameters. Given $p$ and $m$, we will need to estimate $\tau_1, \ldots, \tau_m$, $\Delta_2, \ldots, \Delta_{m+1}$, $\mu_1, \ldots, \mu_T$, $\alpha$, and all PAR($p$) parameters. Developing and optimizing an objective function for this purpose will be the subject of our next two sections.

Before leaving the model description, we make a comment. The model studied here allows for process changes at the changepoint times in the form of level shifts. This is reasonable in climate cases [Vincent (1998), Menne and Williams (2005), Lund et al. (2007)]. In other applications such as speech recognition and finance, it may be more realistic to keep mean process levels fixed and allow the time series parameters to change at each changepoint time [see Inclan and Tiao (1994), Chen and Gupta (1997), Davis, Lee, and Rodriguez-Yam (2006)].

**3. An MDL objective function.** To fit the above model, estimates of the changepoint numbers and locations, as well as the model parameters, are needed. Since different changepoint numbers refer to models with a different number of parameters, the model dimension will also need to be estimated. This is a model selection problem. Popular approaches to model selection problems include AIC (Akaike Information Criterion), BIC (Bayesian Information Criterion), cross-validation type methods, and MDL methods. For problems that involve the detection of regime changes, MDL methods often provide superior empirical results [e.g., Lee (2000, 2002), Davis, Lee, and Rodriguez-Yam (2006)]. This superiority is likely due to the fact that both AIC and BIC place the same penalty on all parameters, regardless of the nature of the parameter (e.g., mean shift magnitudes and changepoint times receive the same penalty). On the other hand, MDL methods can situationally tailor penalties for parameters of different natures, thereby accounting for whether the parameter is real or integer-valued.

The MDL principle was developed by Rissanen (1989, 2007) as a general method for solving model selection problems. It has roots in coding and information theories. In brief, MDL defines the best fitting model as the one that enables the best compression of the data; for the current problem, the data are the observed $\{X_t\}$. There exist several versions of MDL; the so-called two-part MDL is used here. For introductory MDL material, see Hansen and Yu (2001) and Lee (2001).

The rest of this section develops a two-part MDL objective function for fitting a good model. The main idea behind the two-part MDL is described as follows. First, the data $\{X_t\}$ is decomposed into two parts, the fitted candidate model and

its corresponding residuals. MDL methods then calculate the total codelength (i.e., the amount of computer memory) required for storing both parts as a sum of the codelength of the two parts. Finally, MDL methods define the best fitting model as one that produces a minimal codelength. Intuition behind MDL methods lies with why minimum codelength models are also good statistical models. Essentially, it is that both good compression and good statistical models are capable of capturing regularities in the data.

To proceed, let $\mathrm{CL}(z)$ denote the codelength of the object $z$. Also write a candidate fitted model as $\hat{\mathcal{M}}$ and its residuals as $\{\hat{\varepsilon}_t\}$. The codelength is additive in that

$$\mathrm{CL}(\{X_t\}) = \mathrm{CL}(\hat{\mathcal{M}}) + \mathrm{CL}(\{\hat{\varepsilon}_t\}). \tag{3.1}$$

The term $\mathrm{CL}(\hat{\mathcal{M}})$ in (3.1) can be viewed as a model complexity term, while $\mathrm{CL}(\{\hat{\varepsilon}_t\})$ can be viewed as a data fidelity term. Our next task is to obtain a computable expression for $\mathrm{CL}(\{X_t\})$ that can be minimized. We begin with the calculation of $\mathrm{CL}(\hat{\mathcal{M}})$.

An important result of Rissanen (1989) is that the maximum likelihood estimate of a real-valued parameter computed from a series of $N$ observations ($N$ is large) can be effectively encoded with $\log_2(N)/2$ bits. The trend parameter $\alpha$ hence requires $\log_2(N)/2$ bits to encode. The seasonal mean parameters $\mu_\nu$ are effectively estimated via seasonal sample means, each of which contributes $\log_2(d)/2$ bits to the codelength. Given values of the changepoint times $\tau_1, \ldots, \tau_m$, the mean shift parameter $\Delta_j$ can be estimated with data from the $j$th segment only. Hence, $\Delta_j$ requires $\log_2(\tau_j - \tau_{j-1})/2$ bits to encode for $2 \le j \le m+1$ ($\tau_{m+1} = N+1$ is taken as a convention). Hence, the portion of the codelength from mean parameters in the time series regression (i.e., $\alpha$, $\{\mu_\nu\}_{\nu=1}^{T}$, and $\{\Delta_j\}_{j=2}^{m+1}$) is

$$\frac{\log_2(N)}{2} + \frac{T\log_2(d)}{2} + \frac{1}{2}\sum_{j=2}^{m+1}\log_2(\tau_j - \tau_{j-1}). \tag{3.2}$$

The PAR($p$) time series parameters [$\phi_k(\nu)$ for $1 \le k \le p$; $1 \le \nu \le T$ and $\sigma^2(\nu)$ for $1 \le \nu \le T$] are also real valued. Because $\{\varepsilon_t\}$ is a zero mean process, we need only consider the zero mean version of this model. In this case, the PAR parameters can be estimated in an efficient manner via seasonal versions of the Yule–Walker equations [see Pagano (1978)]. The necessary equations for this task are presented in Shao and Lund (2004). Yule–Walker PAR parameter estimators are asymptotically most efficient [Pagano (1978)]; in fact, these estimators are the likelihood estimators except for the edge-effects (i.e., the likelihood is conditional on the first $p$ observations). The Yule–Walker estimators can be computed from the sample autocovariances $\gamma_\nu(h)$ over the lags $h = 0, \ldots, p$. The lag $h$ sample autocovariance at season $\nu$ is defined as $\hat{\gamma}_\nu(h) = d^{-1}\sum_{n=0}^{d-1}\varepsilon_{nT+\nu}\varepsilon_{nT+\nu-h}$, where $\varepsilon_t$ is taken as zero should a $t \le 0$ be encountered in the summation. Observe that $\hat{\gamma}_\nu(0)$ is

a function of $d$ series observations for each fixed $\nu$. Moreover, $\hat{\gamma}_\nu(h)$ is essentially computed from $2d$ observations. Hence, the total codelength from PAR($p$) parameters is

$$\frac{T\log_2(d)}{2} + \frac{pT\log_2(2d)}{2}. \tag{3.3}$$

The parameters $\tau_1, \ldots, \tau_m$ are integers and must be treated as such. Arguing as in Davis, Lee, and Rodriguez-Yam (2006), an integer parameter bounded by $Q$ takes $\log_2(Q)$ bits to encode. Since the $\tau_j$'s are ordered, we have $\tau_j < \tau_{j+1}$. This differs from Davis, Lee, and Rodriguez-Yam (2006) in that we do not loosely bound $\tau_j - \tau_{j-1}$ by $N$ for each $j$. In short, the codelength induced by the changepoint times that we use is

$$\sum_{j=2}^{m} \log_2(\tau_j) + \log_2(N). \tag{3.4}$$

Finally, the model orders $p$ and $m$ contribute

$$\log_2(p) + \log_2(m) \tag{3.5}$$

bits to the codelength. While $m$ is bounded by $N$, typical values of $m$ are significantly smaller than $N$ and a penalty of $\log_2(N)$ would be too much for $m$ changepoints.

Adding (3.2)–(3.5) gives

$$\begin{aligned} \mathrm{CL}(\hat{\mathcal{M}}) = {} & \frac{3}{2}\log_2(N) + T\log_2(d) + \frac{1}{2}\sum_{j=2}^{m+1}\log_2(\tau_j - \tau_{j-1}) \\ & + \frac{pT\log_2(2d)}{2} + \sum_{j=2}^{m}\log_2(\tau_j) + \log_2(m) + \log_2(p). \end{aligned} \tag{3.6}$$

Moving to $\mathrm{CL}(\{\hat{\varepsilon}_t\})$, a fundamental result of Rissanen (1989) is that this quantity equals the negative logarithm (base 2) of the likelihood of the fitted model $\hat{\mathcal{M}}$. For the present problem, this conditional likelihood can be calculated as follows. A Gaussian joint density of observations from the model, denoted by $L$, takes the classical innovations form modified to allow for series periodicities and level shifts at the changepoint times:

$$L = (2\pi)^{-N/2}\left(\prod_{t=1}^{N} v_t\right)^{-1/2} \exp\left[-\frac{1}{2}\sum_{t=1}^{N}\frac{(X_t - \hat{X}_t)^2}{v_t}\right]. \tag{3.7}$$

Here, $\hat{X}_t = P(X_t | X_1, \ldots, X_{t-1}, 1)$ is the best one-step-ahead predictor of $X_t$ from linear combinations of a constant and $X_1, \ldots, X_{t-1}$. Also, $v_t = E[(X_t - \hat{X}_t)^2]$ is the mean squared error (unconditional) of the one-step-ahead predictor.

The one-step-ahead prediction equations and mean squared errors for the PAR($p$) setup are easily expressed:

$$\hat{X}_{nT+\nu} = E[X_{nT+\nu}] + \sum_{k=1}^{p} \phi_k(\nu)(X_{nT+\nu-k} - E[X_{nT+\nu-k}]), \qquad nT + \nu > p,$$

where $E[X_{nT+\nu}] = \mu_\nu + \alpha(nT + \nu) + \delta_{nT+\nu}$ is the mean function. Computing $\hat{X}_t$ and $v_t$ for $t \le p$ is done as in Shao and Lund (2004). Taking a negative logarithm in (3.7) gives

$$\text{CL}(\{\hat{\varepsilon}_t\}) = \frac{N}{2}\log_2(2\pi) + \frac{1}{2}\sum_{t=1}^{N}\log_2(v_t) + \frac{1}{2}\log_2(e)\sum_{t=1}^{N}\frac{(X_t - \hat{X}_t)^2}{v_t}. \tag{3.8}$$

Substituting (3.6) and (3.8) into (3.1), we arrive at the following approximation:

$$\begin{aligned}\text{CL}(\{X_t\}) = \log_2(e)\Bigg[&\frac{3}{2}\ln(N) + T\ln(d) + \frac{1}{2}\sum_{j=2}^{m+1}\ln(\tau_j - \tau_{j-1}) + \frac{pT\ln(2d)}{2} \\ &+ \sum_{j=2}^{m}\ln(\tau_j) + \ln(m) + \ln(p) + \frac{N}{2}\ln(2\pi) \\ &+ \frac{1}{2}\sum_{t=1}^{N}\ln(v_t) + \frac{1}{2}\sum_{t=1}^{N}\frac{(X_t - \hat{X}_t)^2}{v_t}\Bigg].\end{aligned}$$

Because $N$, $d$, and $T$ are constant, our objective function for the model $\mathcal{M}$, denoted by MDL($\mathcal{M}$), can be taken as

$$\begin{aligned}\text{MDL}(\mathcal{M}) = &\frac{1}{2}\sum_{j=2}^{m+1}\ln(\tau_j - \tau_{j-1}) + \frac{pT\ln(2d)}{2} + \sum_{j=2}^{m}\ln(\tau_j) \\ &+ \ln(m) + \ln(p) + \frac{1}{2}\sum_{t=1}^{N}\ln(v_t) + \frac{1}{2}\sum_{t=1}^{N}\frac{(X_t - \hat{X}_t)^2}{v_t}.\end{aligned} \tag{3.9}$$

MDLs for variants of the model in (2.1) are worth mentioning. Should one also allow the trend to change with each regime, the codelength becomes, after appropriate modification of (3.2),

$$\begin{aligned}\text{MDL}(\mathcal{M}) = &\sum_{j=1}^{m+1}\ln(\tau_j - \tau_{j-1}) - \ln(\tau_1 - 1)/2 + \frac{pT\ln(2d)}{2} + \sum_{j=2}^{m}\ln(\tau_j) \\ &+ \ln(m) + \ln(p) + \frac{1}{2}\sum_{t=1}^{N}\ln(v_t) + \frac{1}{2}\sum_{t=1}^{N}\frac{(X_t - \hat{X}_t)^2}{v_t},\end{aligned}$$

where $\tau_0 = 1$ is taken as a convention. If the seasonal location parameters $\mu_\nu$ are consolidated to a single $\mu$, then an appropriate MDL (this assumes a single trend parameter) is

$$\begin{aligned}\mathrm{MDL}(\mathcal{M}) = {} & \frac{1}{2}\sum_{j=1}^{m+1} \ln(\tau_j - \tau_{j-1}) + \frac{pT\ln(2d)}{2} + \sum_{j=2}^{m} \ln(\tau_j) \\ & + \ln(m) + \ln(p) + \frac{1}{2}\sum_{t=1}^{N} \ln(v_t) + \frac{1}{2}\sum_{t=1}^{N} \frac{(X_t - \hat{X}_t)^2}{v_t},\end{aligned} \tag{3.10}$$

which is (3.9) expect for the $2^{-1}\ln(\tau_1 - \tau_0)$ term added in the first summation. MDLs for models where the structural form of the regression changes segment by segment are harder to quantify, but also have climate ramifications and are currently being investigated.

By an MDL model, we refer to a model $\hat{\mathcal{M}}$ that minimizes a MDL score over the class of models being considered. Practical minimization of $\mathrm{MDL}(\mathcal{M})$ over all admissible models is not a trivial task, which brings us to our next section.

**4. Optimizing the objective function.** First, suppose that we know $p$ and $m$ and the changepoint times $\tau_1, \ldots, \tau_m$. Then computation of $\mathrm{MDL}(\hat{\mathcal{M}})$ proceeds as follows. Computation of the model codelength given the parameters is straightforward. For computation of the likelihood contribution to the codelength, write (2.1) in the general linear models form

$$\vec{X} = D\vec{\beta} + \vec{\varepsilon}. \tag{4.1}$$

In (4.1), $\vec{\beta} = (\mu_1, \ldots, \mu_T, \alpha, \Delta_2, \ldots, \Delta_{m+1})'$, $\vec{X} = (X_1, \ldots, X_N)'$, $\vec{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_N)'$, and $D$ is the $N \times (T + 1 + m)$ design matrix

$$D = [S|C|R],$$

where $S$ is an $N \times T$ dimensional seasonal indicator matrix (all entries are zero except $S_{t,\nu} = 1$ if $t = \ell T + \nu$ for some $\ell \in \{0, \ldots, d-1\}$), $C$ is an $N \times 1$ vector with $C_t = t$, and $R$ is an $N \times m$ dimensional matrix with all zero entries except $R_{t,j} = 1$ when time $t$, $1 \le t \le N$, is observed during regime $j$ for $2 \le j \le m+1$.

We first estimate $\vec{\beta}$ with ordinary least squares methods. From the estimated $\vec{\beta}$, residuals of this model fit are next computed. From these residuals and a PAR order parameter $p$, estimates of $\phi_k(\nu)$ for $1 \le \nu \le T$ and $1 \le k \le p$ and $\sigma^2(\nu)$ for $1 \le \nu \le T$ are constructed via seasonal Yule–Walker moment estimation methods. With estimates of the $\phi_k(\nu)$'s and $\sigma^2(\nu)$'s, one can return to (4.1) and compute generalized weighted least squares estimators of $\vec{\beta}$. New residuals are computed and the process is iterated in a Cochrane–Orcutt fashion [see Cochrane and Orcutt (1949)] until convergence is achieved. The process gives jointly optimal estimators of $\vec{\beta}$ and the $\phi_k(\nu)$'s and $\sigma^2(\nu)$'s. Typically, only several iterations are needed.

The above enables us to quickly compute a codelength for fixed values of $p$, $m$, and $\tau_1, \ldots, \tau_m$. However, not counting different values of $p$, there are $2^N$ different configurations of $m$ and $\tau_1, \ldots, \tau_m$ that must be considered. In other words, the parameter space has a huge cardinality. To optimize the codelength over this parameter space, we now introduce a 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), Reeves (1993)]. The basic principles of GAs were first developed by Holland (1975) and are designed to mimic the genetic process of natural selection and evolution. GAs start with an initial population of individuals, each representing a possible solution to the given problem. Each individual or *chromosome* in the population is evaluated to determine how well it scores with respect to the objective function. Highly fit individuals are more likely to be selected as parents for reproduction. In a *crossover* procedure, the *offspring* (*children*) share some characteristics of the parents. *Mutation* is often applied after crossover to introduce random changes to the current population with a small probability. Mutation increases population diversity. The offspring are used to construct a new generation by either a generational approach (replacing the whole population) or a steady-state approach (replacing a few of the less fit individuals). This process is repeated until an individual is found that roughly optimizes the objective function.

The GA used in this study is described as follows.

*Chromosome representation*: The first step in designing a GA is to create a suitable chromosome representation for the problem. Here, any individual (model) can be described as a set of parameters: the number of changepoints $m$, the order of the PAR model $p$, and the changepoint locations $\tau_1, \ldots, \tau_m$. Once these parameters are fixed, the regression parameters in the model (2.1) can be estimated using the methods described above. Hence, the chromosome, denoted by $u = (m, p, \tau_1, \ldots, \tau_m)$, is an integer vector of length $m+2$. The lengths of the chromosomes in the population depend on the number of changepoints. A minimum number of observations in each regime is set to $m_\ell T$ to ensure that reasonable mean shift estimates are obtained in all segments. Here, $m_\ell$ is the minimum number of cycles between adjacent changepoints. Our work will take $m_\ell = 1$ (no changepoints within a year for monthly data). Also, we impose the upper bound $p_{\max}$ for the order of the PAR($p$) model; $p_{\max} = 3$ is used in the forthcoming simulation study and examples.

*Initial population generation*: For each individual, the PAR order $p$ is first randomly selected with equal probabilities from the set $\{0, 1, 2, 3\}$. The changepoint numbers $m$ and locations are then independently simulated as follows. There is a probability $p_b$, essentially representing the probability that any admissible time is selected as a changepoint. Since there can be no changepoint before time $t = 1 + m_\ell T$, we first examine time $t = 1 + m_\ell T$, flipping a coin with heads probability $p_b$. If the flip is heads, $t = 1 + m_\ell T$ is declared to be the first changepoint ($\tau_1 = 1 + m_\ell T$) and attention shifts to the next possible changepoint time, which is time $1 + 2m_\ell T$. But if the flip is tails, $t = 1 + m_\ell T$ is not chosen as a changepoint

and we move to the next location at $t = 1 + m_\ell T + 1$, independently flipping the coin again. The process is continued in a similar manner until the last admissible changepoint time at $t = N - m_\ell T$ is exceeded. The population size $n_p = 30$ is used in this study and $p_b$ is set to be $0.06d$ (six changepoints over a century).

*Crossover*: Pairs of parent chromosomes, representing mother and father, are randomly selected from the initial population or current population by a linear ranking/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 0 and the most fit individual is assigned the rank $n_p - 1$. A crossover procedure, as explained in the paragraph below, is then applied to the parents to produce offspring for the next generation. The probability that any two parents have children, denoted by $p_c$, is set to $p_c = 1 - m_\ell/d$.

In our GA implementation, only one pair of parent chromosomes is chosen from the current generation and one child is produced by "mixing" two parent chromosomes with a uniform crossover. This works as follows. The child's PAR order $p$ is either the mother's or the father's PAR order, with both being equally likely. The child's changepoint locations are randomly selected using all admissible changepoint locations from *both* mother and father. For example, for $N = 1200$, $T = 12$, and $m_\ell = 1$, suppose the mother's chromosome has 3 changepoints at the times $t = 200, 320, 600$ and the father's chromosome has 4 changepoints at the times $t = 205, 300, 710, 850$. First, all changepoints from mother and father are mixed together and sorted from smallest to largest, yielding the string $(200, 205, 300, 320, 600, 710, 850)$. We select the first changepoint of the child at $t = 200$ with probability 0.5. If $t = 200$ is selected as a changepoint, then we discard the changepoint $t = 205$ (it would violate segmentation spacing requirements) and move to the next candidate changepoint at $t = 300$, again doing a fifty-fifty selection/inclusion randomization. If $t = 200$ is not chosen as one of the child's changepoints, we move to the next changepoint at $t = 205$ with the same fifty-fifty selection criterion. The child's $m$ is simply the number of retained changepoints.

*Mutation*: Mutation is applied to the child after crossover with a constant probability $p_m$. The probability $p_m$ is typically low; we use $p_m = 0.05$ in the following examples. The PAR order $p$ for the new chromosome produced by mutation is equal to the child's $p$ with a probability of 0.5. Then changepoint locations can either take on the corresponding changepoints from the child's or be a new set randomly selected from the parameter space. Mutation ensures that no solution in the admissible parameter space has a zero probability of being examined.

*New generation*: The steady-state replacement method with a duplication check as suggested by Davis (1991) is applied here to form a new generation. One advantage of the steady-state approach over the generational approach is that it typically finds better solutions faster. In our implementation of the steady-state approach, only one individual is replaced in the current generation by a child after crossover and/or mutation. This allows parents and offsprings to live concurrently, which is

true for long-lived species [Beasley, Bull, and Martin (1993)]. If the child is already present in the current generation, this child will be discarded and another child must be produced by the selection-crossover-mutation process. The duplication check is applied to all new children until a child is found that is not present in the current generation. In this way, duplicate solutions and premature convergence are significantly avoided.

*Migration*: Migrations act to speed up convergence of the GA and can be implemented via a parallel scheme [Davis (1991), Alba and Troya (1999)]. Migration also reduces the probability of premature convergence. The population is divided into several different sub-populations (islands). Highly fit individuals periodically migrate between the islands. The island model GA is controlled by several parameters, such as the number of islands $N_I$, the frequency of migration $M_i$, the number of migrants $M_n$, and the method used to select which individuals migrate. The migration policy used here is as follow. After every $M_i$ generations, the least fit individual on island $j$, $j = 1, \ldots, N_I$, is replaced by the best individual on island $i$, which is randomly selected among all other islands ($j \neq i$). Therefore, each island sends and receives individuals from different islands throughout the duration of the search process. Here, we set $N_I = 40$, $M_i = 5$, and $M_n = 1$.

*Convergence and stopping criteria*: We follow the criterion of Davis, Lee, and Rodriguez-Yam (2006) to declare convergence and terminate the GA. If the overall best individual at the end of each migration does not change for $M_c$ consecutive migrations, then the GA is deemed to have converged to this best individual. Additionally, if the total number of migrations exceeds a predetermined maximum number $M^*$, then the search process is terminated and the best individual in the $M^*$th migration is taken as the optimal solution to the given problem. The parameters $M_c$ and $M^*$ are taken as 10 and 25 in the study, respectively.

**5. A simulation study.** This section investigates the accuracy of the above methods via simulation. This study is designed to correspond to the simulation study in Caussinus and Mestre (2004). Elaborating, we will simulate a thousand series and apply our methods to each series. Each series contains a century ($d = 100$) of monthly data ($T = 12$) with six ($m = 6$) changepoints. This corresponds to the average number of changepoints over a century of operation reported in Mitchell (1953). The changepoint mean shifts in every series occur at the times $\tau_1 = 240$, $\tau_2 = 480$, $\tau_3 = 600$, $\tau_4 = 840$, $\tau_5 = 900$, and $\tau_6 = 1020$. The error terms $\{\varepsilon_t\}$ are simulated as a Gaussian first order periodic autoregression ($p = 1$) with parameters $\phi_1(v)$ and $\sigma^2(v)$ as specified in Table 1; the seasonal means $\mu_v$ are also listed in Table 1 and are in degrees Celsius. These values are those that were estimated for 50 years of monthly temperatures from Longmire, Washington, which was studied in Lund et al. (2007). The trend parameter $\alpha$ was set to zero in all simulations.

The magnitude of the mean shifts $\Delta_2, \ldots, \Delta_7$ are critical. Big mean shifts make changepoints easier to detect. To facilitate interpretability, we use a common mean

TABLE 1
*Simulation parameters*

| $\nu$ | $\mu_\nu$ | $\phi_1(\nu)$ | $\sigma^2(\nu)$ |
|---|---|---|---|
| 1 | −0.61 | 0.272 | 2.713 |
| 2 | 0.99 | 0.284 | 2.748 |
| 3 | 2.35 | 0.478 | 1.871 |
| 4 | 4.91 | 0.286 | 1.717 |
| 5 | 8.74 | 0.335 | 2.474 |
| 6 | 12.15 | 0.279 | 2.403 |
| 7 | 15.51 | 0.245 | 2.569 |
| 8 | 15.47 | 0.137 | 1.910 |
| 9 | 12.79 | −0.127 | 2.826 |
| 10 | 7.82 | 0.082 | 2.488 |
| 11 | 2.32 | 0.196 | 2.394 |
| 12 | −0.25 | 0.214 | 2.256 |

shift magnitude $\Delta > 0$ at all changepoint times. For instance, if the current regime has mean level $c$ (trend and seasonal effects are assumed zero here), the next regime will have mean $c + \Delta$ or $c - \Delta$, with a fifty-fifty chance of shifting up or down at each changepoint time. It follows that $\Delta = |\Delta_j - \Delta_{j-1}|$ for $j = 2, \ldots, 7$.

The ability of our model to detect mean shifts can be roughly quantified by the mean shift magnitude relative to the process standard deviation (the latter averaged over a complete seasonal cycle). A parameter quantifying such aspects, denoted by $\kappa$, is

$$\kappa = \frac{\Delta}{\sqrt{T^{-1} \sum_{\nu=1}^{T} \mathrm{Var}(\varepsilon_{nT+\nu})}}.$$

Better quantifiers of changepoint detection power may well exist, but derivation of such quantities would be difficult and is tangential to our points. Below, we consider three different $\kappa$ values: 1.0, 1.5, and 2.0. The larger $\kappa$ is, the easier it is to detect changepoints. A realization of a temperature series with $\kappa = 1.5$ is plotted in Figure 1 for feel.

Table 2 and Figure 2 summarize the results of the simulations. Table 2 reports empirical frequency distributions of the number of estimated changepoints. Observe that the true value of six changepoints is obtained more frequently as $\kappa$ increases. When $\kappa = 2.0$, the percentage of simulations where the correct number of changepoints is estimated is 58.9%, which is better than the corresponding 43.4% reported in Caussinus and Mestre (2004) that applies to uncorrelated and time-homogeneous settings (i.e., 100 years of annual data). In fairness, we note that the equivalent sample size of our simulated series (the number of independent data points with the same periodic variances) translates to more than the 100 independent data points of Caussinus and Mestre (2004) (we will not quantify equiv-

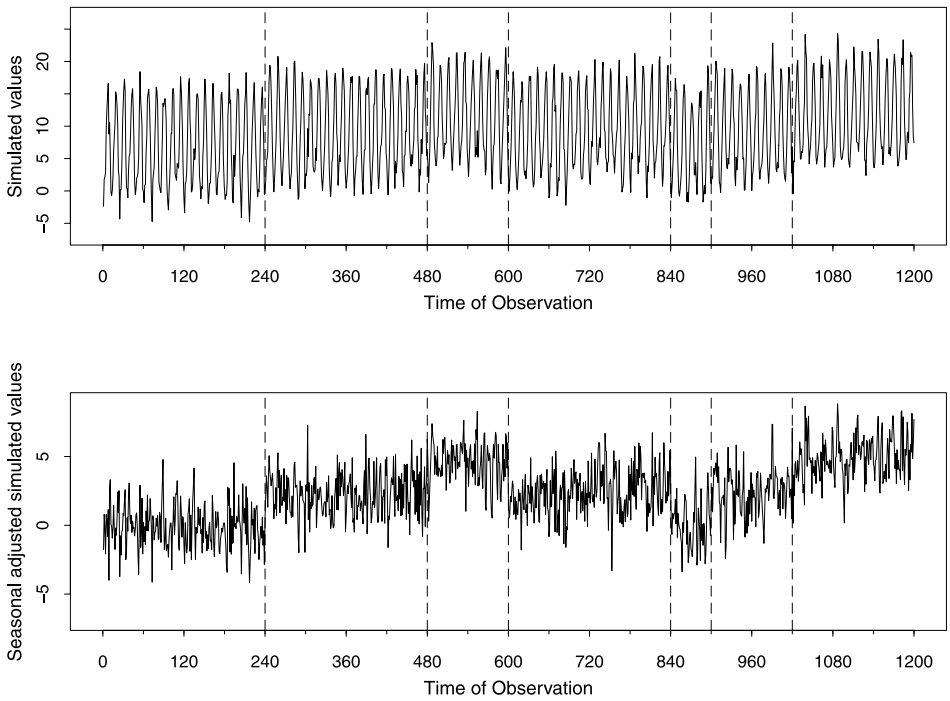


FIG. 1. *A simulated series with six changepoints.*

alent sample sizes further here). The correct number of changepoints is identified only 0.9% when $\kappa = 1.0$ [this, however, is also slightly better than the corresponding result in Caussinus and Mestre (2004)]. It is clear that changepoint numbers

TABLE 2
*Estimated changepoint numbers and PAR*(1) *order when* $m = 6$

| $m$ | $\kappa = 1.0$ | $\kappa = 1.5$ | $\kappa = 2.0$ |
|---|---|---|---|
| 0 | 0.1% | 0.0% | 0.0% |
| 1 | 12.2% | 2.6% | 0.0% |
| 2 | 30.1% | 10.9% | 3.7% |
| 3 | 34.5% | 17.5% | 5.4% |
| 4 | 18.3% | 23.6% | 11.8% |
| 5 | 3.9% | 12.9% | 12.5% |
| 6 | 0.9% | 31.9% | 58.9% |
| 7 | 0.0% | 0.6% | 7.1% |
| $>7$ | 0.0% | 0.0% | 0.6% |
| $p = 1$ | 99.9% | 100% | 100% |
| $p = 0$ | 0.1% | 0.0% | 0.0% |

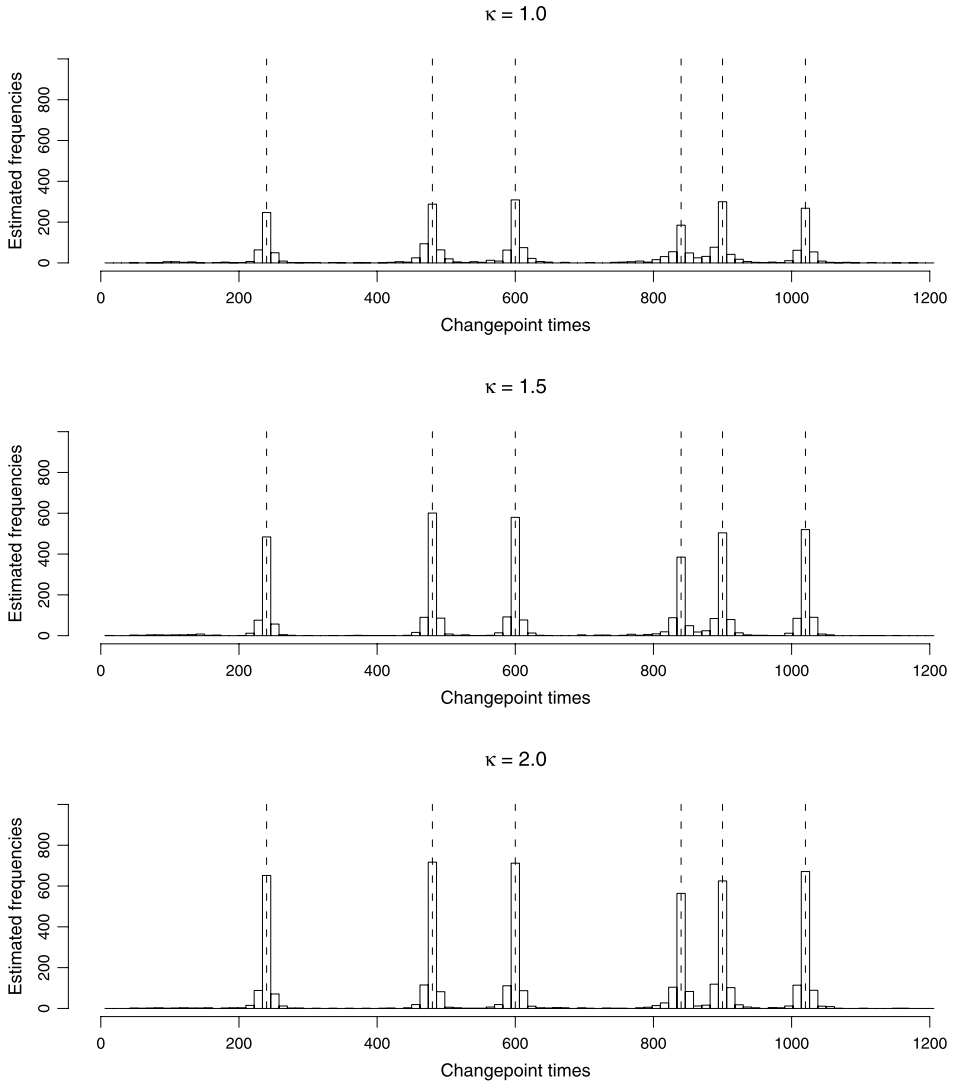


FIG. 2. *Histograms of estimated changepoint times.*

are underestimated in settings with relatively small $\kappa$. In fact, the empirical mean (standard deviation in parentheses) of the distributions in Table 2 are 2.74 (1.07) for $\kappa = 1.0$, 4.34 (1.48) for $\kappa = 1.5$, and 5.413 (1.20) for $\kappa = 2.0$. Overall, one sees that changepoint shift sizes are critical in changepoint detection, that the detection situation is difficult when $\kappa$ is small, but that methods work reasonably well when $\kappa$ is relatively large. Using monthly data (as opposed to annual averages) also seems to improve changepoint detection power.

As for where the changepoints are estimated to occur, Figure 2 shows histograms of the estimated changepoint locations, reporting the total number of times a changepoint is signaled at time $t$ for $1 \leq t \leq N$ in the 1000 simulations. Observe that the histograms have modes around the actual changepoint times. It is also evident that the changepoints at times 840 and 900 were the most difficult to detect, a feature attributed to the close proximity of the times of these two changepoints (with the fifty-fifty up/down mean shift randomization employed, the sign of these two mean shifts differ with probability $1/2$, in which case their detection is relatively more difficult).

Note that the correct autoregressive order $p = 1$ was obtained virtually all of the time. Hence, the time series model selection component seems to be working well. As changing the trend parameter did not appreciably affect results, we will not report separate tables with nonzero trends.

We now compare the MDL penalty more closely with the Caussinus–Lyazrhi penalty used in Caussinus and Mestre (2004). The Caussinus–Lyazrhi penalty is larger than AIC or BIC penalties, but does not penalize parameters in the mean function or consider autocorrelation aspects. To make this comparison, 1000 series of length 100 were simulated with six changepoints always occurring at the times 20, 40, 50, 70, 75, and 85. The mean shift size parameter $\kappa$ was changed to the parameter $a$ in Caussinus and Mestre (2004) to mimic their simulations. The errors in the model were assumed to be Gaussian and independent. Note that the level of changepoint activity relative to the sample size has increased 12-fold from the previous simulations. Table 3 below lists estimates of the relative frequencies of changepoints found by the genetic algorithm with an MDL penalty when $\mu_\nu$ is held constant with $\nu$. No trends were considered in this setup nor was tuning of the genetic algorithm (varying its mutation probabilities, etc.) considered in detail. The frequency distributions in Table 3 are approximately the same as those in Caussinus and Mestre (2004), perhaps slightly worse, in all cases. For this sample size and level of changepoint activity, an MDL penalty seems to perform about the

TABLE 3
*Estimated number of changepoints for $n = 100$*

| $m$ | $a = 1.0$ | $a = 2.0$ | $a = 3.0$ |
|---|---|---|---|
| 0 | 6.2% | 0.1% | 0.0% |
| 1 | 29.5% | 0.3% | 0.0% |
| 2 | 32.5% | 4.2% | 0.0% |
| 3 | 22.8% | 5.4% | 0.0% |
| 4 | 7.4% | 35.3% | 7.7% |
| 5 | 1.5% | 33.0% | 4.4% |
| 6 | 0.1% | 20.9% | 81.7% |
| 7 | 0.0% | 0.8% | 6.1% |
| $>7$ | 0.0% | 0.0% | 0.1% |

same as the Caussinus–Lyazrhi penalty. Of course, we reiterate that some gains are made by considering monthly data in lieu of annual averages.

**6. The Tuscaloosa data.** Figure 3 plots a century of monthly data from Tuscaloosa, Alabama recorded from January, 1901–December, 2000. A seasonal mean cycle is visually evident in the data, but trends and mean shifts are not readily apparent. Comparing the year-to-year jaggedness of the seasonal throughs (the winter minimums) against the year-to-year seasonal peaks (July maximums), it is discerned that this series has a periodic variance with winter temperatures being much more variable than summer temperatures. In fact, as we will see, the entire autocorrelation structure of the series is periodic.

The Tuscaloosa series is one in which the station history is reasonably documented. In particular, a catalog (called meta-data) exists that notes the circumstances under which the data were recorded, including the times of station relocations and instrumentation changes. This said, meta-data files are notoriously incomplete [Menne and Williams (2005)] and "undocumented" changepoints may lurk. The Tuscaloosa series also has a moderately clean changepoint record with only four major documented changes over a century of operation; as noted before,

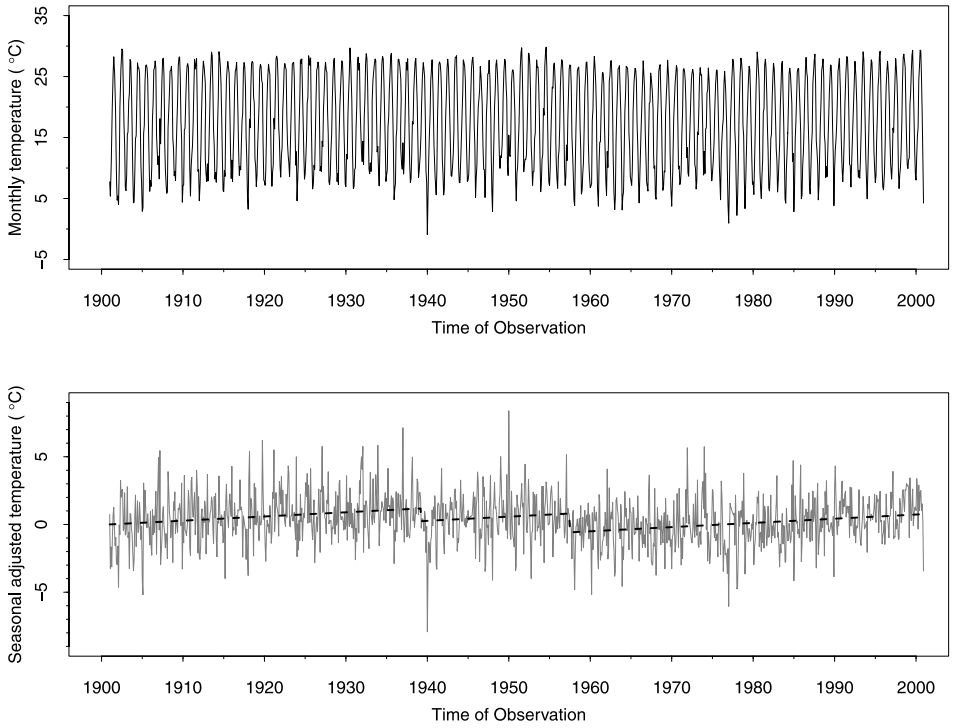


FIG. 3. *The Tuscaloosa data with changepoint structure imposed.*

the average United States temperature station experiences about six changepoints per century [Mitchell (1953)]. In short, the Tuscaloosa is a good "proving ground" series for changepoint methods.

Level shifts in temperature series are arguably the most important factor in assessing temperature trends [see Lu and Lund (2007)]. It has been argued in climate settings that the manner in which changepoints are handled may be the most critical factor in the global warming debate. Supporting this, temperature insurance treaties on Wall Street are based solely on station location and gauge properties, while ignoring long-term trends altogether.

Our methods were applied to the Tuscaloosa data. A reference series was constructed by averaging three neighboring series located at Selma, AL, Greensboro, AL, and Aberdeen, MS over the century of record. We work with one reference series that averages three neighboring series to expedite the discourse; methods that analyze all $\binom{4}{2}$ pairs of stations are discussed in Menne and Williams (2005, 2009).

First, we examine the Tuscaloosa series without a reference. The fitted MDL model has two changepoints at times 460 (April, 1939) and 679 (July, 1957). The mean function induced by these two changepoints, less the seasonal cycle but including the trend, is plotted against the data in Figure 3. The mean shift magnitudes of the 1939 and 1957 changepoints, in degrees Celsius, are both negative: $\hat{\Delta}_2 = -0.94 \pm 0.20$ and $\hat{\Delta}_3 = -2.33 \pm 0.30$. The estimated trend parameter is $\hat{\alpha} = 0.00258 \pm 0.00039$. The standard errors were estimated from the fitted time series regression model with generalized weighted least squares techniques. The estimated order of the PAR model is $p = 1$; a consequence of this is that the autocovariance structure in the errors of the fitted models is indeed periodic.

Second, we examine the Tuscaloosa minus the reference series. This seasonally adjusted difference series is plotted in Figure 4. In this target minus reference, four changepoints are flagged: March 1909, December 1919, July 1933, and August 1990. The estimates of the $\Delta_i$'s are $\hat{\Delta}_2 = 0.76 \pm 0.11$, $\hat{\Delta}_3 = 0.26 \pm 0.11$,

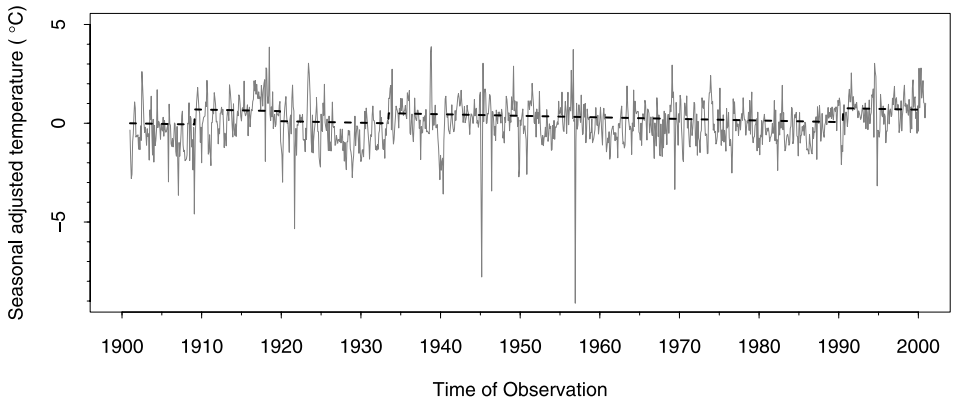


FIG. 4. *The Tuscaloosa minus the reference data with changepoint structure imposed.*

$\hat{\Delta}_4 = 0.77 \pm 0.13$, and $\hat{\Delta}_5 = 1.47 \pm 0.19$. Note that a mean shift with the small magnitude of 0.26 has been flagged. The trend estimate is $\hat{\alpha} = -0.00066 \pm 0.00015$ and the selected order of the autoregression is $p = 1$. Observe that the fitted order of the autoregression did not reduce from that for the raw series; that is, periodic autocorrelation still exists in the target minus reference series. Also, the trend for the target minus reference series appears to be significantly negative. We comment that the two large negative values occurring in the 1940s and the 1950s appear to be decimal typos in the raw data; that is, the monthly average temperature for Tuscaloosa was entered as ten degrees too small. We make this claim after examining additional reference series from various cities close to Tuscaloosa. Whereas the series in this database have been quality checked to some degree, errors like this may still exist. We reran the analysis above after replacing these two values by (1) their estimated seasonal means $\hat{\mu}_\nu$ and (2) what we believe are the correct values, that is, adding 10 degrees to both outliers. In both cases, four changepoints with similar magnitudes and times to the ones above are found. The 1957 changepoint flagged in the target series has not been flagged in any version of the target minus difference series. The 1909 changepoint is possibly attributed to a changepoint in the reference series: Greensboro reports a time of observation change in 1906 and Aberdeen reports a station relocation in 1915.

The meta-data show four changepoints in this series: the first was a station relocation in November of 1921, the second was a station relocation in March of 1939, the third was a station relocation during June of 1956 and an accompanying instrumentation change in November of 1956 (we regard this as one changepoint), and the fourth is a station relocation and instrumentation change in May of 1987. The reference series analysis seems to have correctly identified three of these four changepoints (we are liberally including the 1933 flagged changepoint time as correctly identifying the 1939 changepoint), missing the 1956 changepoint and adding a 1909 changepoint. The raw target series analysis misses the 1921 and 1987 changepoints, but finds the 1956 changepoint; also, the estimated time of the 1939 changepoint is much closer to its true value than that for the reference analysis. Overall, it seems that the reference analysis is superior to simple target series analysis, but that one can learn something with both analyses.

We caution the reader that trends in some monthly temperature series, especially when the series is aggregated over a large geographic region, may not be well described by a linear regression component. As noted by Handcock and Wallis (1994), trends at localized series are more likely to be adequately described with a simple linear structure. As a final diagnostic check, residuals from the model fits were computed. Figure 5 shows the sample autocorrelation of the residuals for the target series over the first 60 lags. The dashed lines are 95% pointwise confidence bounds for white noise. As only three of the sample autocorrelations lie outside the bounds (and then only slightly so), the model appears to have fitted the data well. Figure 6 shows the periodogram of the residuals from the target minus reference series. A long memory structure is not readily evident in this plot.

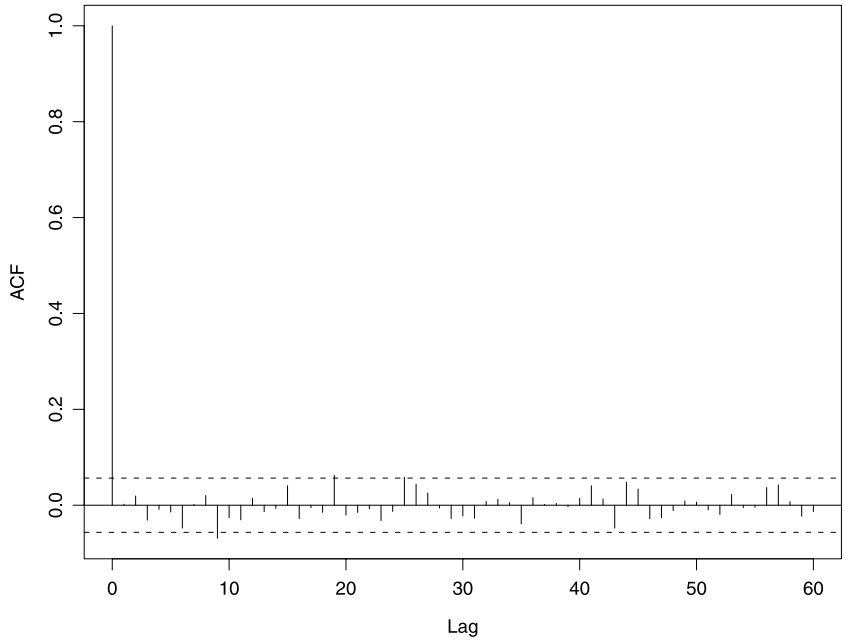


FIG. 5. *Sample autocorrelations of target series residuals.*

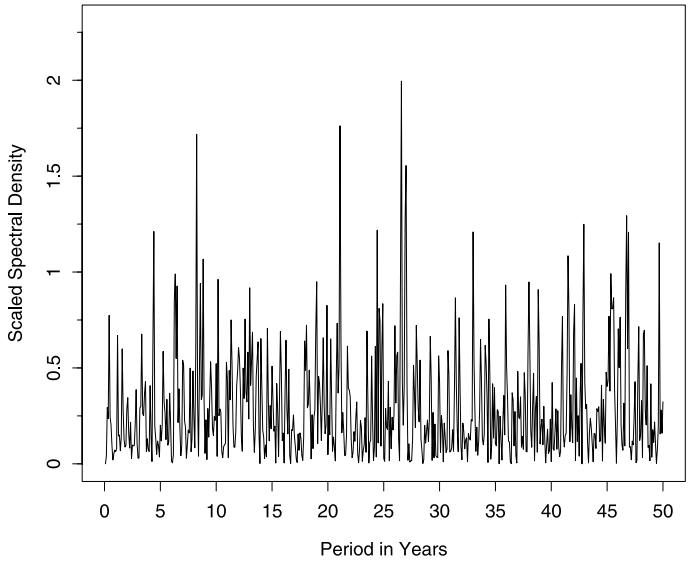


FIG. 6. *Periodogram of target minus reference series residuals.*

**7. Comments.** Time series parsimony may be an issue with periodic data. Specifically, the penalty in (3.3) essentially assumes that the PAR(1) model requires $(p+1)T$ distinct parameters. In practice, changes in climate processes from season to season are slow/smooth. Low order Fourier series expansions, such as those in Lund, Shao, and Basawa (2005), can statistically simplify the model and serve to lessen the penalty for time series components. This issue is likely to be paramount should daily data be considered.

**Acknowledgments.** Comments from two referees and the editor substantially improved this paper.

## REFERENCES

ALBA, E. and TROYA, J. M. (1999). A survey of parallel-distributed genetic algorithms. *Complexity* **4** 31–52. MR1688681

BEASLEY, D., BULL, D. R. and MARTIN, R. R. (1993). An overview of genetic algorithm: Part 1, fundamentals. *University Computing* **15** 58–69.

BERKES, I., HORVATH, L., KOKOSZKA, P. and SHAO, Q. M. (2006). On discriminating between long-range dependence and changes in mean. *Ann. Statist.* **34** 1140–1165. MR2278354

BRAUN, J. V. and MÜLLER, H. G. (1998). Statistical methods for DNA sequence segmentation. *Statist. Sci.* **13** 142–162.

CAUSSINUS, H. and MESTRE, O. (2004). Detection and correction of artificial shifts in climate series. *J. Roy. Statist. Soc. Ser. C* **53** 405–425. MR2075499

CHEN, J. and GUPTA, A. K. (1997). Testing and locating variance changepoints with application to stock prices. *J. Amer. Statist. Assoc.* **92** 739–747. MR1467863

COCHRANE, D. and ORCUTT, G. H. (1949). Application of least squares regression to relationships containing autocorrelated error terms. *J. Amer. Statist. Assoc.* **44** 32–61.

DAVIS, L. (1991). *Handbook of Genetic Algorithm*. Reinhold, New York.

DAVIS, R. A., LEE, T. C. M. and RODRIGUEZ-YAM, G. A. (2006). Structural break estimation for nonstationary time series models. *J. Amer. Statist. Assoc.* **101** 223–239. MR2268041

FEARNHEAD, P. (2006). Exact and efficient Bayesian inference for multiple changepoint problems. *Statist. Comput.* **16** 203–213. MR2227396

GOLDBERG, D. E. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley, Reading, MA.

HANDCOCK, M. S. and WALLIS, J. R. (1994). An approach to statistical spatial-temporal modeling of meteorological fields. *J. Amer. Statist. Assoc.* **89** 368–378. MR1294070

HANSEN, M. H. and YU, B. (2001). Model selection and the principle of minimum description length. *J. Amer. Statist. Assoc.* **96** 746–774. MR1939352

HOLLAND, J. H. (1975). *Adaptation in Natural and Artificial Systems*, MIT Press, Cambridge, MA. MR0441393

INCLAN, C. and TIAO, G. C. (1994). Use of cumulative sums of squares for retrospective detection of changes of variance. *J. Amer. Statist. Assoc.* **89** 913–923. MR1294735

LEE, T. C. M. (2000). A minimum description length based image segmentation procedure, and its comparison with a cross-validation based segmentation procedure. *J. Amer. Statist. Assoc.* **95** 259–270. MR1803154

LEE, T. C. M. (2001). An introduction to coding theory and the two-part minimum description length principle. *International Statistical Review* **69** 169–183.

LEE, T. C. M. (2002). Automatic smoothing for discontinuous regression functions. *Statist. Sinica* **12** 823–842. MR1929966

LU, Q. and LUND, R. B. (2007). Simple linear regression with multiple level shifts. *Canad. J. Statist.* **37** 447–458. MR2396030

LUND, R. B., SHAO, Q. and BASAWA, I. V. (2005). Parsimonious periodic time series modeling. *Aust. N. Z. J. Statist.* **48** 33–47. MR2234777

LUND, R. B., WANG, X. L., LU, Q., REEVES, J., GALLAGHER, C. and FENG, Y. (2007). Change-point detection in periodic and autocorrelated time series. *Journal of Climate* **20** 5178–5190.

MITCHELL, J. M., JR. (1953). On the causes of instrumentally observed secular temperature trends. *Journal of Applied Meteorology* **10** 244–261.

MENNE, J. M. and WILLIAMS, C. N., JR. (2005). Detection of undocumented changepoints using multiple test statistics and composite reference series. *Journal of Climate* **18** 4271–4286.

MENNE, J. M. and WILLIAMS, C. N., JR. (2009). Homogenization of temperature series via pairwise comparisons. *Journal of Climate* **22** 1700–1717.

PAGANO, M. (1978). On periodic and multiple autoregressions. *Ann. Statist.* **6** 1310–1317. MR0523765

REEVES, C. (1993). *Modern Heuristic Techniques for Combinatorial Problems*. Wiley, New York. MR1230642

RISSANEN, J. (1989). *Stochastic Complexity in Statistical Inquiry*. World Scientific, Singapore. MR1082556

RISSANEN, J. (2007). *Information and Complexity in Statistical Modeling*, Springer, New York. MR2287233

SHAO, Q. and LUND, R. B. (2004). Computation and characterization of autocorrelations and partial autocorrelations in periodic ARMA models. *J. Time Ser. Anal.* **25** 359–372. MR2062679

VINCENT, L. A. (1998). A technique for the identification of inhomogeneities in Canadian temperature series. *Journal of Climate* **11** 1094–1104.

Q. LU
DEPARTMENT OF MATHEMATICS AND STATISTICS
MISSISSIPPI STATE UNIVERSITY
MISSISSIPPI STATE, MISSISSIPPI 39762
USA
E-MAIL: qlu@math.msstate.edu

R. LUND
DEPARTMENT OF MATHEMATICAL SCIENCES
CLEMSON UNIVERSITY
CLEMSON, SOUTH CAROLINA 29634-0975
USA
E-MAIL: lund@clemson.edu

T. C. M. LEE
DEPARTMENT OF STATISTICS
COLORADO STATE UNIVERSITY
FORT COLLINS, COLORADO 80523
USA
AND
DEPARTMENT OF STATISTICS
CHINESE UNIVERSITY OF HONG KONG
SHATIN, HONG KONG
E-MAIL: tlee@sta.cuhk.edu.hk