Homogenization of Daily Temperature Data

Anuradha P. Hewaarachchi

Department of Statistics and Computer Science, University of Kelaniya, Kelaniya, Sri Lanka

Yingbo Li

*Department of Mathematical Sciences, Clemson University, Clemson, South Carolina, and
Department of Statistical Science, Southern Methodist University, Dallas, Texas*

Robert Lund

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

Jared Rennie

*Cooperative Institute for Climate and Satellites—North Carolina, North Carolina State University, Raleigh, and
NOAA/National Center for Environmental Information, Asheville, North Carolina*

(Manuscript received 14 February 2016, in final form 20 October 2016)

ABSTRACT

This paper develops a method for homogenizing daily temperature series. While daily temperatures are statistically more complex than annual or monthly temperatures, techniques and computational methods have been accumulating that can now model and analyze all salient statistical characteristics of daily temperature series. The goal here is to combine these techniques in an efficient manner for multiple changepoint identification in daily series; computational speed is critical as a century of daily data has over 36 500 data points. The method developed here takes into account 1) metadata, 2) reference series, 3) seasonal cycles, and 4) autocorrelation. Autocorrelation is especially important: ignoring it can degrade changepoint techniques, and sample autocorrelations of day-to-day temperature anomalies are often as large as 0.7. While daily homogenization is not conducted as commonly as monthly or annual homogenization, daily analyses provide greater detection precision as they are roughly 30 times as long as monthly records. For example, it is relatively easy to detect two changepoints less than two years apart with daily data, but virtually impossible to flag these in corresponding annually averaged data. The developed methods are shown to work in simulation studies and applied in the analysis of 46 years of daily temperatures from South Haven, Michigan.

1. Introduction

Climate time series often exhibit artificial discontinuities induced by station relocations, gauge changes, observer changes, and so on. Such changes may impart statistical discontinuities in associated data and are called changepoints (or breakpoints, or mean shifts). Mitchell (1953) estimates that U.S. temperature series experience about six breakpoints per century on average. Some, but not necessarily all, of these times induce mean shifts in the series. While the times of some gauge changes, station relocations, and other events are documented in station history logs (called metadata), these records are notoriously incomplete, and many breakpoint times are undocumented.

This paper seeks to identify all changepoint times in a daily temperature record while accounting for four critical aspects: metadata, a reference series, a seasonal cycle, and autocorrelation. While Li and Lund (2015) and Li et al. (2016) consider these features in annual and monthly series, this paper modifies the methods to accommodate the more complex features seen in daily data. Analyses of a single daily series by some existing methods may take days of computation time as a century of daily data has over 36 500 entries. Our methods are illustrated on single series only; homogenization of a temperature series network or comparison to other homogenization

Corresponding author e-mail: Robert Lund, lund@clemson.edu

methods is a worthy endeavor, but beyond our intended scope.

Undocumented changepoint identification is crucial in climate analysis (Potter 1981; Vincent 1998; Caussinus and Mestre 2004; Menne and Williams 2005, 2009; Lu and Lund 2007). The changepoint locations and mean shift sizes need to be estimated to make accurate inferences from the data; in fact, Lund et al. (2001) show that changepoint information is the single most important data feature to account for when reliably estimating a temperature trend at a fixed U.S. station. Once the changepoint times are identified, most other statistical inference procedures are relatively straightforward.

A common method used to identify multiple changepoints is any binary segmentation procedure and an at most one changepoint (AMOC) test. Workhorse AMOC procedures include the standard normal homogeneity (SNH) test, the nonparametric SNH test, and the two-phase regression of Lund and Reeves (2002) and Wang et al. (2014). These and other methods are reviewed in Reeves et al. (2007) and typically assume that the underlying regression model for the series is known and that the error terms in the regression model are independent and identically distributed. Such assumptions, especially independence, are violated with monthly or daily temperatures, which are highly correlated.

Binary segmentation techniques can turn any AMOC method into a multiple changepoint estimation scheme. In segmentation schemes, the time series is first classified as changepoint free or having a single changepoint. If one changepoint is declared, then the series is split into two segments about the changepoint time. AMOC methods are then applied to the two shorter segments to test for further changepoints. This procedure is repeated until all subsegments are declared changepoint free. Segmentation techniques have difficulty detecting two or more changepoints located closely in time (Li and Lund 2012). Moreover, when multiple changepoints shift the series mean higher at some changepoints and lower at others, an AMOC technique may fail to declare any changepoints whatsoever. For these reasons, multiple changepoint techniques are needed.

Efficient multiple changepoint algorithms that identify the number of changepoints and their locations are presented in Caussinus and Mestre (2004) and Davis et al. (2006). Caussinus and Mestre (2004) use a penalized log-likelihood criterion to estimate the number of changepoints, their locations, and any outliers. Davis et al. (2006) propose an automatic procedure to segment nonstationary time series into blocks of different autoregressive (AR) processes. The number of changepoints, their locations, and the orders of the AR models are estimated by optimizing a minimum description length (MDL) objective function via a genetic algorithm. Menne and Williams (2005) introduce semihierarchical splitting algorithms to multiple changepoint problems. There, a series is subdivided and several hypothesis tests are conducted to compare candidate changepoint configurations.

Li and Lund (2012) develop a multiple changepoint technique for annual climatic databased on an MDL penalized likelihood. There, the penalized likelihood is optimized by a genetic algorithm; however, their techniques apply to annual (nonperiodic) series and ignore trend features. Toreti et al. (2012) present a general segmentation method based on hidden Markov chains. They analyze annual winter precipitation, which does not exhibit high autocorrelation. Li and Lund (2015) develop Bayesian statistical methods to incorporate metadata in multiple changepoint detection and apply them to annual precipitation data. Prior distributions for the number of changepoints and their locations are constructed to reflect climatologists' belief that the metadata times are more likely to be changepoints. The prior distributions and the likelihood of the observed data are combined to form a posterior distribution of the changepoint configuration. The number of changepoints and their locations are estimated as those that maximize the posterior probability. We will borrow some of these techniques to handle metadata and correlation aspects in daily series.

The above literature studies monthly and annual series. Changepoint literature for daily data is scarcer. Homogenized daily data are useful in trend, extreme, and variability studies. Since a daily series contains many more observations than monthly or annual series, daily analyses will have a greater precision. On the other hand, analysis of daily data is more challenging due to the longer series lengths and the number of time series model parameters needed. In fact, a simple model for daily temperatures contains more than 1095 (365×3) parameters (see the next section).

Vincent and Zhang (2002) present a method to homogenize daily maximum and minimum temperatures over Canada. Their method homogenizes daily data based on the changepoints found and the subsequent adjustments made in corresponding monthly data. Daily temperature adjustments are conducted by linear interpolation, which preserves the long-term trend and variations in the monthly series. Della-Marta and Wanner (2006) propose a method to homogenize daily data that is capable of adjusting the series' mean and higher-order moments. Their method uses a nonlinear model to estimate the relationship between a target and reference series. Kuglitsch et al. (2009) present a quality control and homogenization method based on a penalized log-likelihood for a nonlinear model. The break detection

and correction methods there require a highly correlated reference series. The breakpoints are identified by applying the methods in Caussinus and Mestre (2004) to an annually differenced series. More recently, Trewin (2013) develops a percentile-matching algorithm to homogenize daily temperature data in Australia, which permits different adjustments based on where a temperature lies in its frequency distribution. Wang et al. (2014) and Xu et al. (2013) also use changepoints identified in the monthly averages to homogenize corresponding daily maximum and minimum temperatures. All of the above-mentioned daily homogenization methods are based on the changepoints identified in corresponding annual or monthly series. Often, correlation aspects are eschewed in these methods.

For daily precipitation, Wang et al. (2010) develop an AMOC method based on a two-phase regression model and a data-adaptive Box–Cox transformation for nonzero daily precipitation amounts, noting that it is wrong to change a dry day to a nondry day. Gallagher et al. (2012) also develop an AMOC technique for daily precipitation data via Markov chain and prediction methods. Their methods employ a background Markov chain to describe adjacent rainy and dry runs of days. While this model allows for correlation in the day-to-day precipitation amounts, the analyses becomes more mathematically complicated.

In this paper, a Bayesian MDL (BMDL) method is devised to estimate multiple changepoints in daily temperature data. Our method estimates the number of changepoints and their locations in data with autocorrelation, seasonality, and/or a linear trend. A genetic algorithm is devised to optimize the BMDL objective function, which is developed from a time series model for daily temperatures that allows for seasonality and autocorrelation. The model incorporates prior beliefs based on metadata records.

The rest of the paper is organized as follows. The next section introduces a model for daily temperature data. Section 3 develops the BMDL objective function for the problem. Section 4 deals with genetic algorithm aspects. Section 5 presents simulation studies showing that the methods can effectively and efficiently detect changepoints and accurately estimate their mean shift sizes. Section 6 presents a changepoint analysis of daily temperatures recorded at South Haven, Michigan. Section 7 concludes with comments.

2. A multiple changepoint model for daily data

Our object of interest is a daily temperature series. Such series display autocorrelation, seasonal means and variances, a linear trend, and possible mean shifts at breakpoint times. A model that captures the above features will now be devised. We consider data $\mathbf{X} = (X_1, X_2, \ldots, X_N)'$ recorded daily. Here, $N = dT$, where $T = 365$ is the period of the series and d is the number of years of data. We assume data for d complete years to avoid trite work. The season (day of year) is indexed by $\nu \in \{1, 2, \ldots, T\}$. The notation $X_{nT+\nu}$ refers to the observation during the νth day of the nth year, for years $n \in \{0, 1, \ldots, d-1\}$. With daily data, leap year observations are omitted to enforce a period of $T = 365$.

Our fundamental model is a linear regression with seasonality, a linear trend, multiple possible mean shifts, and periodic random errors:

$$X_{nT+\nu} = \mu_\nu + \alpha(nT + \nu) + \delta_{nT+\nu} + \varepsilon_{nT+\nu}. \quad (1)$$

Here, μ_ν is the mean temperature on day ν (neglecting trend and mean shifts). We assume that the linear trend parameter, α, is time-homogeneous; other trend structures can be accommodated, but this is seldom necessary when examining target minus reference series as the subtraction greatly reduces any trends. The ordered changepoint times are denoted by $1 < \tau_1 < \tau_2 < \cdots < \tau_m \leq N$, where m is the unknown number of changepoints. Time 1 is not allowed to be a changepoint. The changepoint structure can be described by a binary indicator vector $\boldsymbol{\eta} = (\eta_2, \eta_3, \ldots, \eta_N)'$, with

$$\eta_t = \begin{cases} 1, & \text{if } \text{time } t \text{ is a changepoint}, \\ 0, & \text{otherwise}. \end{cases}$$

This model has 2^{N-1} distinct changepoint configurations. The m changepoints in $\boldsymbol{\eta}$ partition the series into $m+1$ distinct regimes. We take $\tau_0 = 1$ and $\tau_{m+1} = N+1$ for edge notations. The jth regime consists of the observations for times t with $t \in \{\tau_{j-1}, \ldots, \tau_j - 1\}$ for $j = 1, 2, \ldots, m+1$.

In regime $j \geq 2$, Δ_j denotes the magnitude that the mean has shifted relative to that in the first regime (neglecting the seasonal cycle and trend). For parameter identifiability, $\Delta_1 = 0$ is imposed. Define a vector $\boldsymbol{\Delta} = (\Delta_2, \Delta_3, \ldots, \Delta_{m+1})'$ whose entries are the mean shift sizes that need to be estimated. The component $\{\delta_{nT+\nu}\}$ in (1) has the shift structure

$$\delta_t = \begin{cases} \Delta_1 = 0, & \tau_0 \leq t < \tau_1, \\ \Delta_2, & \tau_1 \leq t < \tau_2, \\ \vdots \\ \Delta_{m+1}, & \tau_m \leq t < \tau_{m+1}. \end{cases}$$

The shift from regime j to regime $j+1$ changes mean temperatures (trend and seasonal cycle excluded) by $\Delta_{j+1} - \Delta_j$ degrees.

The model errors $\{\varepsilon_{nT+\nu}\}$ have zero mean and are autocorrelated. Daily temperatures are in fact heavily

correlated, with consecutive days often having correlation on the order of 0.7. As winter temperatures are about 2 to 5 times more variable than summer temperatures in the United States (as measured by standard deviation), a first-order periodic autoregressive time series [PAR(1)] (Lund et al. 1995) will be used to model the regression errors. A PAR(1) time series indeed has autocorrelation and periodic variances. A series $\{\varepsilon_t\}$ is said to be a PAR(1) series with zero mean if it satisfies the seasonal difference equation

$$\varepsilon_{nT+\nu} = \phi(\nu)\varepsilon_{nT+\nu-1} + Z_{nT+\nu}. \tag{2}$$

Here, $\phi(\nu)$ is the autoregressive parameter during day ν, and $\{Z_{nT+\nu}\}$ is zero mean periodic white noise with variance $\mathrm{Var}(Z_{nT+\nu}) = \sigma^2(\nu)$.

Our model has $3T + 2m + 2$ unknown parameters: the changepoint parameters m, $\tau_1, \tau_2, \ldots, \tau_m$, the seasonal means $\mu_1, \mu_2, \ldots, \mu_T$, the linear trend α, the mean shifts $\Delta_2, \Delta_3, \ldots, \Delta_{m+1}$, and the time series PAR(1) parameters $\phi(1), \phi(2), \ldots, \phi(T)$ and $\sigma^2(1), \sigma^2(2), \ldots, \sigma^2(T)$. In the next section, a BMDL objective function is introduced, which is subsequently minimized to estimate an optimal multiple changepoint configuration.

3. Bayesian minimum description lengths

This section develops an objective function that can be minimized to estimate the optimal changepoint configuration $\boldsymbol{\eta}$. The derivation is lengthy and similar to that in Li et al. (2016) for monthly data; hence, only an outline of the derivation steps and the end objective functions are listed here. This said, the derivation differs somewhat from Li et al.'s since $\{\varepsilon_t\}$ has periodic features here.

The MDL principle is used as our model selection criteria. An MDL objective function is a penalized likelihood with a smart penalty tailored to the changepoint problem. The MDL penalty, originally developed in Rissanen (1989) from information theory, has an analogous role to the Akaike information criterion (AIC) and Bayesian information criterion (BIC) penalties, but is more complicated than a simple multiple of the number of unknown parameters that characterize AIC and BIC penalties. In fact, the MDL penalty also depends on how far the changepoints lie from each other. Among a class of plausible models, the MDL principle seeks the model with the smallest (shortest) so-called description length. Better models should have shorter description lengths. For more background, see Hansen and Yu (2001) and Grünwald et al. (2005). The MDL principle has been utilized in climate changepoint detection problems (Davis et al. 2006; Lu et al. 2010; Li and Lund 2012), with good results. Recently,

Li et al. (2016) developed a new BMDL technique that uses metadata. Here, this method is tailored to accommodate daily data.

For a given candidate changepoint configuration $\boldsymbol{\eta}$, its BMDL has form

$$\mathrm{BMDL}(\boldsymbol{\eta}) = \mathcal{L}(\mathbf{X}\,|\,\boldsymbol{\eta}) + \mathcal{L}(\boldsymbol{\eta}), \tag{3}$$

where $\mathcal{L}(\mathbf{X}\,|\,\boldsymbol{\eta})$ is related to the data likelihood, and $\mathcal{L}(\boldsymbol{\eta})$ acts as a penalty for the changepoint number and locations. An outline of the steps to compute the BMDL is listed below:

1) Compute the PAR(1) likelihood function given the other model parameters,

$$f(\mathbf{X}\,|\,\boldsymbol{\Delta}, \boldsymbol{\mu}, \alpha, \boldsymbol{\phi}, \boldsymbol{\sigma}^2, \boldsymbol{\eta}),$$

where $\boldsymbol{\mu}$, $\boldsymbol{\phi}$, and $\boldsymbol{\sigma}^2$ are vectors containing all seasonal means, PAR(1) coefficients, and PAR(1) white noise variances, respectively.

2) Compute a marginal likelihood by integrating the regime mean shift sizes $\boldsymbol{\Delta}$ out under a Gaussian prior distribution, that is,

$$f(\mathbf{X}\,|\,\boldsymbol{\mu}, \alpha, \boldsymbol{\phi}, \boldsymbol{\sigma}^2, \boldsymbol{\eta}) = \int f(\mathbf{X}\,|\,\boldsymbol{\Delta}, \boldsymbol{\mu}, \alpha, \boldsymbol{\phi}, \boldsymbol{\sigma}^2, \boldsymbol{\eta})\pi(\boldsymbol{\Delta})\,d\boldsymbol{\Delta},$$

where the prior distribution $\pi(\boldsymbol{\Delta})$ is assumed to be composed of m independent normal distributions with zero mean and the same variance, that is, $\boldsymbol{\Delta} \sim N(0, \kappa g^2 \mathbf{I}_m)$, where $g^2 = \left[\prod_{\nu=1}^{T}\sigma^2(\nu)\right]^{1/T}$ is the geometric mean of $\sigma^2(1), \sigma^2(2), \ldots, \sigma^2(T)$. The parameter κ can be roughly viewed as the ratio of the variance of regime means relative to the variance of time series noises over a year. One does not need a precise value for κ. Here, κ is usually prespecified as some large value so that very little mean shift information is contained in the prior; we force mean shift sizes to be learned from the data. Our default takes $\kappa = 5$.

3) Maximize the marginal likelihood function over the model parameters $\boldsymbol{\mu}$, α, $\boldsymbol{\phi}$, $\boldsymbol{\sigma}^2$, and obtain the description length of the observed data $\mathcal{L}(\mathbf{X}\,|\,\boldsymbol{\eta})$, that is,

$$f(\mathbf{X}\,|\,\boldsymbol{\eta}) = f(\mathbf{X}\,|\,\hat{\boldsymbol{\mu}}, \hat{\alpha}, \hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\sigma}}^2, \boldsymbol{\eta}),$$
$$\mathcal{L}(\mathbf{X}\,|\,\boldsymbol{\eta}) = -\log[f(\mathbf{X}\,|\,\boldsymbol{\eta})],$$

where $\hat{\boldsymbol{\mu}}$, $\hat{\alpha}$ are the ordinary least squares estimators, and $\hat{\boldsymbol{\phi}}$, $\hat{\boldsymbol{\sigma}}^2$ are the Yule–Walker moment estimators for the PAR(1) model, computed from standard time series methods (Lund et al. 1995).

4) Compute the MDL of the description length of the changepoint configuration $\boldsymbol{\eta}$ via $\mathcal{L}(\boldsymbol{\eta}) = -\log[\pi(\boldsymbol{\eta})]$,

where $\pi(\boldsymbol{\eta})$ is the prior discrete probability mass function of $\boldsymbol{\eta}$. Metadata is incorporated in this prior distribution. Elaborating, a beta-binomial prior is put on $\boldsymbol{\eta}$. This prior assumes that 1) each undocumented time is a changepoint with probability ρ_1, 2) each documented time is a changepoint with probability ρ_2, and 3) documented times are more likely than undocumented times to be changepoints: $\rho_2 > \rho_1$. In the absence of information beyond the metadata record, changepoints declarations at all distinct time points are assumed to be statistically independent. Since we do not know ρ_1 and ρ_2, we model these in a Bayesian hierarchical fashion. Elaborating, ρ_1 is modeled as a Beta$(1, \beta_1)$ random variable; ρ_2 is modeled as a Beta$(1, \beta_2)$ variable. Our default values take $\beta_1 = 365/.06$ and $\beta_2 = 4$, which reflects our prior belief that $E[\rho_1] = 1/(1 + 365/.06) \approx .06/365$ (approximately six changepoints per century) and $E[\rho_2] = 1/(1 + 4) = 0.2$ (one out of every five metadata times induces a true mean shift). The parameters β_1 and β_2 can be changed by users should changepoints be believed to occur at different rates. Detection results are relatively stable under a wide range of parameter choices (Li and Lund 2015).

For a given candidate changepoint configuration $\boldsymbol{\eta}$ with at least one changepoint, the BMDL objective function is

$$\text{BMDL}(\boldsymbol{\eta}) = \frac{m}{2}\log(\kappa g^2) + \frac{1}{2}\sum_{t=1}^{N}\log[\sigma^2(t)] + \frac{1}{2}\log(|\mathbf{B}|) + \frac{1}{2}\sum_{t=1}^{N}\frac{Y_t^2}{\sigma^2(t)} - \frac{1}{2}\mathbf{b}'\mathbf{B}^{-1}\mathbf{b}$$
$$- \log[\Gamma(1 + m_1)\Gamma(\beta_1 + n_1 - m_1)\Gamma(1 + m_2)\Gamma(\beta_2 + n_2 - m_2)]. \tag{4}$$

In the above, $\Gamma(x)$ is the gamma function at the argument x, all logarithms are natural-based, and $|\mathbf{B}|$ is the determinant of the matrix $\mathbf{B}$, whose formula is given below. The optimal changepoint configuration is the $\boldsymbol{\eta}$ that minimizes BMDL$(\boldsymbol{\eta})$. For each candidate changepoint configuration $\boldsymbol{\eta}$, the mean shift, trend, and time series parameters are optimally estimated (see the simulation example in section 5 for a detailed illustration). Our next objective is to explain the quantities arising in (4). For this, the prediction residuals $\{Y_t\}_{t=1}^{N}$ are computed from

$$Y_t = (X_t - \mu_t - \alpha t) - \phi(t)[X_{t-1} - \mu_{t-1} - \alpha(t-1)],$$

with the convention that $Y_1 = X_1 - \mu_1 - \alpha$. During the jth regime, we define

$$a_j = \sum_{t=\tau_{j-1}}^{\tau_j - 1}\frac{1}{\sigma^2(t)} + \sum_{t=\tau_{j-1}}^{\tau_j - 1}\frac{\phi^2(t+1)}{\sigma^2(t+1)} - 2\sum_{t=\tau_{j-1}+1}^{\tau_j - 1}\frac{\phi(t)}{\sigma^2(t)},$$

$$b_j = \sum_{t=\tau_{j-1}}^{\tau_j - 1}\frac{Y_t}{\sigma^2(t)} - \sum_{t=\tau_{j-1}}^{\tau_j - 1}\frac{Y_{t+1}\phi(t+1)}{\sigma^2(t+1)},$$

and $c_j = \phi(\tau_{j-1})/\sigma^2(\tau_{j-1})$. Also, $\mathbf{b} = (b_2, \ldots, b_{m+1})'$, and $\mathbf{B}$ is an $m \times m$ symmetric matrix with form

$$\mathbf{B} = \begin{pmatrix} a_2 + \dfrac{1}{\kappa g^2} & -c_3 & 0 & 0 & 0 & \cdots & 0 \\ -c_3 & a_3 + \dfrac{1}{\kappa g^2} & -c_4 & 0 & 0 & \cdots & 0 \\ 0 & -c_4 & a_4 + \dfrac{1}{\kappa g^2} & -c_5 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -c_m & a_m + \dfrac{1}{\kappa g^2} & -c_{m+1} \\ 0 & 0 & 0 & 0 & \cdots & -c_{m+1} & a_{m+1} + \dfrac{1}{\kappa g^2} \end{pmatrix}.$$

Finally, m_1 and m_2 are the number of undocumented and documented changepoints in $\boldsymbol{\eta}$, respectively. The total number of changepoints in $\boldsymbol{\eta}$ is $m = m_1 + m_2$; moreover, n_2 is the number of metadata points and $n_1 = N - 1 - n_2$.

The BMDL for the changepoint configuration with no changepoints, denoted by $\boldsymbol{\eta}_0$, is

$$
\begin{aligned}
\text{BMDL}(\boldsymbol{\eta}_0) = {} & \frac{1}{2} \sum_{t=1}^{N} \log[\sigma^2(t)] + \frac{1}{2} \sum_{t=1}^{N} \frac{Y_t^2}{\sigma^2(t)} \\
& - \log[\Gamma(\beta_1 + n_1)\Gamma(\beta_2 + n_2)].
\end{aligned} \quad (5)
$$

This allows one to compare models with changepoints to the model with no changepoints and fixes an issue with the methods in Davis et al. (2006) and Li and Lund (2012), where the term $\log(m)$ arises (this is undefined for $m = 0$, the no-changepoint configuration).

4. BMDL minimization

The best changepoint configuration is the one (or more) that minimizes the BMDL score. A naive approach to find such a configuration is to perform an exhaustive search. Such an approach requires 2^{N-1} distinct BMDL evaluations. Thus, for a century of daily data, 2^{36500} BMDL evaluations need to be conducted, an infeasible task for even the world's fastest computers. Hence, an efficient optimization algorithm is needed to find the best model. For this, a genetic algorithm (GA), which is an intelligent stochastic search that quickly visits only good changepoint configurations, is devised to perform the minimization.

GAs are popular optimization tools (Goldberg and Holland 1988) that are inspired by natural selection and genetics. Like Darwin's theory of evolution, GAs have aspects of genetic evolution that allow the fittest models to survive in a random walk stochastic search. GAs usually converge to global optimums. Beasley et al. (1993) and the references therein compare GAs to other optimization methods.

GAs encode each model as a chromosome. Here, a chromosome is represented by a binary indicator vector $\boldsymbol{\eta} = (\eta_2, \eta_3, \ldots, \eta_N)'$ as in section 2. The number of changepoints in $\boldsymbol{\eta}$ is $m = \sum_{t=1}^{N} \eta_t$ and the changepoint locations $\tau_1, \tau_2, \ldots, \tau_m$ are the nonzero positions in $\boldsymbol{\eta}$. The GA begins with a randomly generated initial population of chromosomes and evaluates the BMDL score at each generated chromosome. The GA then simulates successive generations of chromosomes via a series of operations: parent selection, breeding (crossover), and mutation. Chromosomes with smaller BMDLs are viewed as fitter and are more likely to be selected to bear children. In each generation, two chromosomes (parent chromosomes) are selected to breed. These parent chromosomes are combined cross-overed to form a new chromosome called a child. The child's chromosome is allowed to mutate before joining the next generation.

This process is repeated until a certain number of children are produced for the generation. The resulting population is referred to as the next generation. A prespecified number of generations are often simulated. If implemented correctly, the overall fitness of the generation, which is the BMDL score of the fittest individual in the generation, converges to the global minimum (Cerf 1998). Details on how to implement our GA are now given.

a. Initial generation

An initial population often simply simulates a set of chromosomes at random. Here, each position in a chromosome is allowed to be a changepoint with some preset probability. For daily data, this probability is set to 6/36500 [following Mitchell (1953), this corresponds to an average of six changepoints per century]. While small generation sizes might not explore enough different chromosomes, larger generation sizes slow the algorithm down. A generation size of 150 will be taken here for illustration purposes. One need not consider metadata aspects in the initial generation; this is accounted for in the BMDL score.

b. Parent selection

Once the initial generation is simulated, parents (mother and father chromosomes) are selected to breed. To generate fitter offspring, a parent selection technique is needed. This technique should be more likely to choose fitter individuals to bear children. Several selection mechanisms are listed in Beasley et al. (1993). Here, a linear ranking is used to select the parents from the 150 chromosomes. First, the 150 chromosomes' BMDL scores are ranked in a descending order; the chromosome with the highest BMDL (the least fit) has rank 1 and the chromosome with the smallest BMDL (the most fit) has rank 150. Parents are chosen with probabilities proportional to their ranks: if the rank of the ith chromosome is R_i, it is selected as a father with probability $R_i / \sum_{j=1}^{150} R_j = R_i/11{,}325$. The most fit chromosome has a 0.01324 chance of being selected as a father for any child; the least fit chromosome has a 0.00008809 chance of fatherhood. Mothers are then selected in the same way from all nonfather chromosomes.

c. Crossover

Crossover mechanisms combine mother and father chromosomes in a random manner to generate a child chromosome. The child chromosome ideally contains changepoint characteristics of both parents. Our crossover mechanism allows changepoints in either parent to be changepoints of the child. The general idea is best illustrated with an example: suppose, with $N = 6$, that

the mother chromosome is $\boldsymbol{\eta}_1 = (0,1,0,1,0,0)'$ and the father chromosome is $\boldsymbol{\eta}_2 = (0,0,0,1,0,1)'$ (the time slot 1 restriction $\eta_1 = 0$ is appended in the $\boldsymbol{\eta}$ vectors for clarity). Here, the mother has changepoints at times 2 and 4 and the father has changepoints at times 4 and 6. The child chromosome is first set to have changepoints of either mother or father: $(0,1,0,1,0,1)'$. At this point the child likely has more changepoints than the mother or father; hence, some child chromosomes are randomly discarded. With the aforementioned child chromosome, a fair coin is flipped three times, one at each of the three changepoint times, and all changepoints with tails are discarded. If the resulting coin-flip sequence is heads, tails, heads, then the second changepoint at time 4 is discarded and the resulting chromosome becomes $(0,1,0,0,0,1)'$.

Since the number of distinct changepoint configurations is enormous, changepoint locations are perturbed to speed algorithm convergence. Next, the location of each changepoint is shifted via an integer-valued random variable with zero mean. To execute this, two independent Poisson random numbers D_1 and D_2 are generated at each changepoint time and the changepoint's location is then shifted by $D_1 - D_2$ time units. For example, a chromosome containing three changepoints might see Poisson differences of -1, 0, and 3, respectively. Then the first changepoint is shifted downward one day, the second changepoint time is not shifted, and the third changepoint time is shifted upward three days. Should any of the shifted times be less than day 2 or more than day N, the changepoint is altogether eliminated. Choosing the best Poisson parameter λ can be tricky, but it is important for computational speed. In early generations, a larger λ is needed to explore new changepoint locations; in later generations, a smaller value of λ is preferred to slightly tune the likely good changepoint configurations in the current models being explored. Selection of λ is described further below.

d. Mutation

Each child is allowed to mutate after crossover. Mutation changes randomly selected bits of each chromosome. If mutation is not allowed, the GA can hone in to a local minimum; with mutation, radically different chromosomes are continually explored. Mutation essentially ensures the exploration of whole changepoint configuration space, maintaining a diversity of the chromosome population and preventing premature GA convergence. Our mutation mechanism selects a random number of locations in a child and flips the changepoint at each of these selected locations. For example, if position 100 is chosen for mutation and is not a changepoint in the child, it is flipped to a changepoint; should time 100

already be a changepoint, it is flipped to a non-changepoint. In our algorithm, each time is allowed to mutate independently with a very small probability (described below). In many chromosomes, no mutation occurs.

e. Islands and migration

There can be a huge number of distinct changepoint configurations in a daily series. In such settings, researchers often suggest island versions of the GA approach. In an island GA, populations are divided into several subpopulations, called islands. GAs are run simultaneously on each island. The islands are largely isolated, but migrations are allowed to occur between islands every now and again. This allows very fit chromosomes to change islands. Migration increases chromosome diversity and prevents the algorithm from converging to a local BMDL minimum. A migration policy specifies the number of islands, the migration rate (number of individuals to migrate), and the migration interval (the frequency of migrations). Our migration policy replaces the least-fit individual on each island by the best-fit individual of a randomly selected different island, once every five generations.

f. Stopping rule and parameter choices

The GA is terminated when a prescribed stopping criterion is reached. Frequently used stopping criteria are that a prespecified maximum number of generations are reached, or that there is no improvement in the most-fit member in many successive generations. The most-fit chromosome of the last generation (among all islands) is taken as the estimated changepoint configuration.

GA convergence depends on parameters such as the number of islands, the population size of each island, the mutation probability, and the Poisson parameter λ. Our experience suggests that the GA will converge under a range of parameter choices, which suggests that one does not have to tune these parameters optimally to get good results; however, an efficient algorithm is usually appreciated. In our subsequent work, the following parameter settings are used: 1) with 46 years of daily data, two islands of size 75 were used, the mutation probability was set to 0.0001, and $\lambda = 50$. For 10 years of daily data, three islands of size 50 were used, the mutation probability was set to 0.0001, and $\lambda = 1$. In the next section, we also explore how long a GA takes to find the best changepoint configuration with various parameter settings. Users can experiment with other parameter settings that may work faster for their particular series, but GA algorithm convergence is usually not a parameter selection issue.
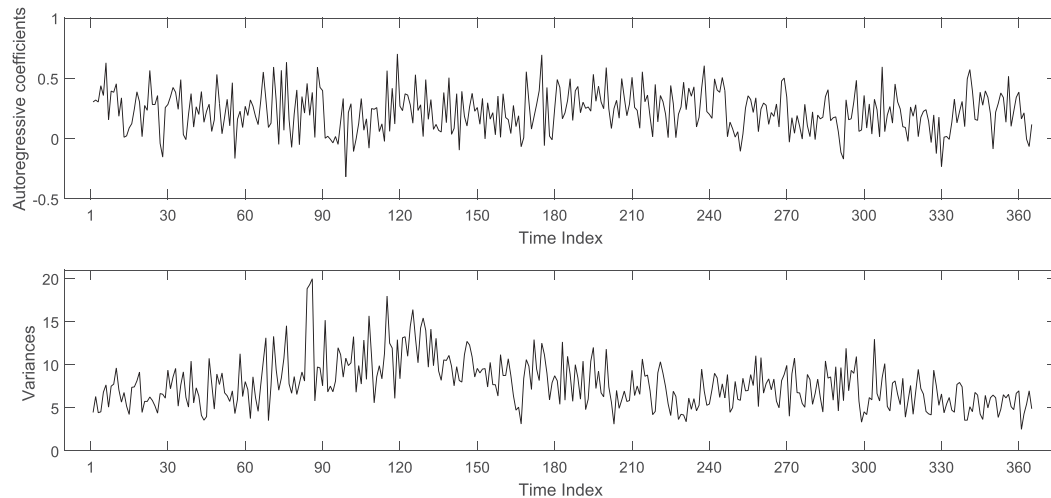
FIG. 1. Periodic autoregressive coefficients $\phi(\nu)$ and variances $\sigma^2(\nu)$ of the target minus reference series.

5. Simulation studies

Using simulation examples, this section first assesses the performance of our daily homogenization methods, illuminates its advantages over monthly homogenization techniques, and explores different GA parameter choices and their runtimes. One thousand series, each containing 10 years of daily data ($N = 3650$) were simulated under various scenarios. For application realism, the daily means and linear trend were set to those estimated in the South Haven, Michigan, daily temperature series after adjusting for a reference series. The South Haven series is studied in detail in the next section. The parameters of the PAR(1) model were set to those estimated in the South Haven minus reference series of the next section. Figure 1 graphically displays these parameters. In each simulated series, a metadata record was posited to contain five points at the times $450, 900, 1500, 3100,$ and 3350.

a. No changepoints

As a control run, 1000 Gaussian series were simulated under the above specifications without changepoints and our methods were applied. A GA with two islands was used to optimize the BMDL; the other GA settings are as specified in the last section. Two hundred generations were simulated in the analysis of each series. The results estimated 962 series with no changepoints, 33 series with one changepoint, and five series with two changepoints. The false-positive rate (3.8%) is reasonably low. The average runtime of the GA for each series in this section was about nine minutes on a Dell OptiPlex 9020 computer. MATLAB R2015 software was used to run the genetic algorithm and the code is available from the authors upon request.

b. Three changepoints: One documented and two undocumented

Next, 1000 Gaussian series were simulated with three mean shift changepoints at the times $\tau_1 = 900$ (19 June, year 3), $\tau_2 = 1800$ (6 December, year 5), and $\tau_3 = 2700$ (25 May, year 8). Here, the first changepoint is also a metadata time. The series mean shifts upward by 2°F at each changepoint time. The average variance (over all days of year) of any simulated series is roughly 7.8°F. Hence, the signal to noise ratio is $2/\sqrt{7.8} = 0.716$. The top panel of Fig. 2 shows a simulated series. The metadata times are marked with crosses on the x axis. Detection percentages for this case (at the exact changepoint time) are displayed in the bottom panel of Fig. 2. Since τ_1 is a metadata point, τ_1 should be easier to flag as a changepoint, all other things being equal. As expected, τ_1 has the highest detection rate. Although the mean shift sizes at times τ_2 and τ_3 are identical, the detection rate of τ_2 is higher. This is because τ_2 occurs on 6 December, which is a season of less variability than τ_3, which occurs on 25 May. From the bottom panel of Fig. 1, $\sigma(\nu)$ at τ_2 (6 December) and τ_3 (25 May) are 1.89° and 3.34°F, respectively. Higher variability makes changepoint detection harder (this seasonality issue is explored further below). Among the 1000 simulated series, the GA estimated the true number of changepoints ($m = 3$) correctly in 975 of the series. In the remaining 25 cases, 10 of the series were estimated to have two changepoints, 12 series to have four changepoints, two series to have one changepoint, and one series to be changepoint free.

To evaluate the detection performance under different shift sizes, three changepoints are placed at the times $\tau_1 = 900$, $\tau_2 = 1825$ (31 December, year 5), and
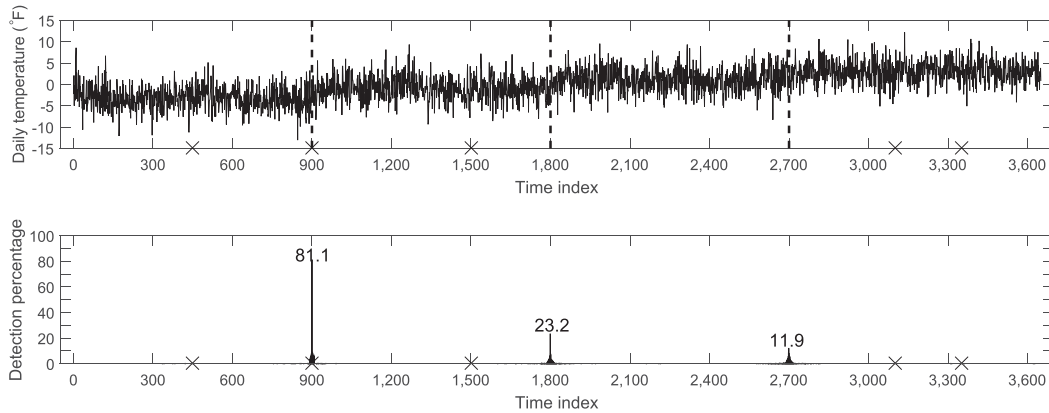
FIG. 2. A simulated daily temperature series with three changepoints. Vertical dashed lines demarcate the three mean shift at times 900, 1800, and 2700. Crosses on the axes mark metadata times. The bottom plot shows detection percentages.

$\tau_3 = 2700$, shifting the mean upward by $\Delta°\text{F}$, then downward by $2\Delta°\text{F}$, and then upward by $\Delta°\text{F}$, respectively; the last shift returns the series to its starting mean level. The shift size Δ (in °F) is varied in $\{0.5, 1.0, 1.5, 2.0\}$. All other simulation parameters, including metadata, remain as in the above paragraph. Table 1 shows, for all Δ considered, the percentages of exact changepoint hits (estimating the changepoint on its exact day of occurrence), the number of hits that flag the changepoint within ± 15 days of its true occurrence (say a monthly hit), and the estimated total number of changepoints. Since τ_1 is a metadata time, it has the highest exact detection percentage under most Δs. Because the mean shift size at the second changepoint is twice as large as the other two mean shifts, the detection rates of τ_2 within ± 15 days are higher than those of τ_1 and τ_3. Obviously, performance worsens as the shift size Δ become smaller.

c. Two changepoints in different seasons

Next, 1000 Gaussian series with two changepoints at the times $\tau_1 = 821$ (1 April, year 3) and $\tau_2 = 2758$ (22 July, year 8) were simulated. The variability of the series at the April changepoint is about 3.87°F and the variability at the July changepoint is about 1.76°F. Both changepoints are posited to be undocumented and shift the series mean 2°F upward. Figure 3 displays a simulated series and an associated histogram of detection percentages. The detection rate (exact hit) of the April changepoint is indeed 13% lower than the detection rate of the July changepoint, confirming that changepoints occurring in high variability seasons (winter or spring) are more difficult to detect than changepoints occurring in low variability seasons (summer). The true number of changepoints ($m = 2$) was correctly estimated in 951 of the 1000 runs; 14 series were estimated to have one changepoint, and 35 series to have three changepoints.

d. Estimation of shift sizes

To investigate the estimation accuracy of mean shift sizes, 1000 Gaussian series with one changepoint in the middle of the record, $\tau_1 = 1825$, were generated, with mean shift sizes (in °F) varied in $\{\pm 0.5, \pm 1.0, \pm 1.5, \pm 2.0\}$. For simplicity, no trend or seasonality is considered. In this setting, the errors were assumed to follow a first-order autoregressive [AR(1)] model with $\phi = 0.7$ and white noise variance $1 - \phi^2 = 0.51$. These settings induce a unit variance in all simulated temperatures. Table 2 displays the means of the estimated shift sizes, aggregated only from the runs where a single changepoint was detected. As long as the signal-to-noise ratio (mean shift size to standard deviation, which is the mean shift size in this case) is not too small, our method accurately estimates the mean shift size.

e. Daily versus monthly changepoint detection

Changepoints located close to each other can be hard to detect, in which case the increased number of observations

TABLE 1. (top) Detection percentages for the three changepoint simulated example. (bottom) Estimated number of changepoints, out of 1000 independent realizations for each shift size Δ.

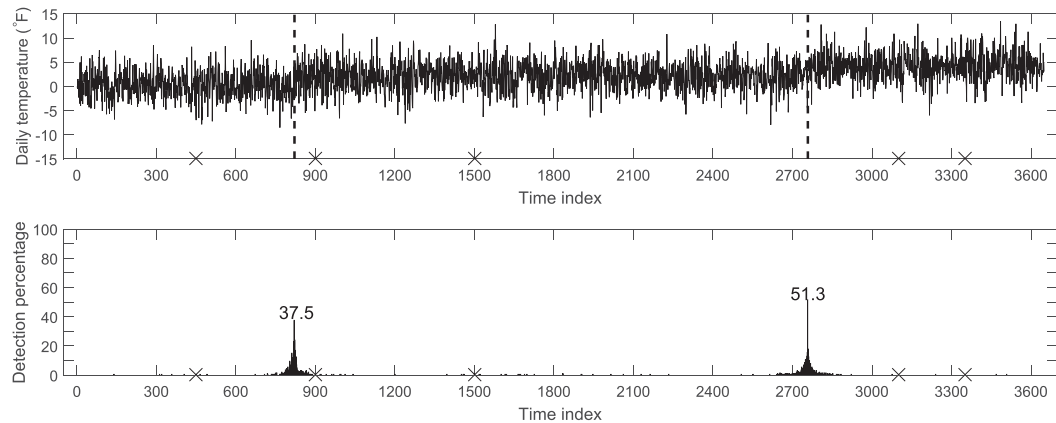| Shift size Δ (°F)	Within ± 15 days			On the exact day		
	τ_1	τ_2	τ_3	τ_1	τ_2	τ_3
2.0	57.4	96.1	43.0	46.4	34.8	15.8
1.5	35.0	89.2	22.0	27.3	22.6	8.8
1.0	9.4	74.6	4.5	8.3	10.9	1.0
0.5	3.3	36.0	1.2	3.1	3.0	0.8
Shift size Δ (°F)	0	1		2	3	4
2.0	0	77		13	888	22
1.5	10	378		87	521	4
1.0	20	771		103	105	1
0.5	129	801		67	3	0

FIG. 3. A simulated daily temperature series with two changepoints. Vertical dashed lines demarcate the two mean shifts at times 821 and 2757. Crosses on the axes mark the metadata times. The bottom plot shows detection percentages.

in daily data can be helpful. Here, for each of the 1000 Gaussian series [no trend, no seasonality, AR(1) with $\phi = 0.7$ and a series variance of unity], 10 years of daily data were generated, with three nonmonotonic mean shifts placed at days 1735, 1825, and 1915 that shift the series mean by $+2°$, $-4°$, and $+2°F$, respectively. Each series was analyzed twice, once using our daily homogenization techniques, and once using a monthly version of BMDL after monthly averaging the daily data. No metadata are assumed to be available here. For the monthly averaged data, the true changepoints occur at months 57, 60, and 63.

Figure 4 shows detection percentages at exact times. The extra precision in the daily record substantially improved detection accuracy over monthly data, while not increasing false detections. The analysis with monthly series typically misses all three changepoints. For a fairer comparison between daily and monthly analyses, an exact hit with daily data is better viewed as a hit if a changepoint is flagged ± 15 days from the true mean shift, which is a "monthly window." With this definition of a hit time, the daily detection rates of the three Fig. 4 changepoints increase to 53.6%, 80.4%, and 52.0%, respectively. Before leaving this issue, we comment that some series have erroneous observations (outliers) that do not appear blatantly wrong at a first glance. Often, MDL adjacent routines will flag a pair of adjacent times, indicating that such a point could be an outlier, perhaps in need of further examination.

f. GA parameters and runtimes

Finally, runtimes (minutes) are explored for a 10-yr daily series with three changepoints at days 900, 1800, and 2700 (Fig. 2 graphs an example of such a series) for different GA parameter settings. The optimum changepoint configuration was determined by running a genetic

algorithm many times and recording the absolute best BMDL. Then, a GA was run under various different parameter settings until it found this optimal changepoint configuration, and then terminated. For each different parameter configuration, a GA was run 25 times and average runtimes were computed.

The top portion of Table 3 fixes the mutation probability as 0.0001. GA convergence slows as the population size (the number of islands times the island size) grows. With the same population size of 100, a GA with two islands slightly outperforms that with a single island. The bottom three rows in Table 3 fix the parameters at their best values in top nine rows of this table: $\lambda = 1$, two islands, and an island size of 50. The mutation probability of 0.0001 was found as optimal among the ones explored.

6. Analysis of daily data from South Haven, Michigan

Figure 5 (left panels) displays average daily temperatures at South Haven, Michigan, from 1 January 1953

TABLE 2. Mean shift size estimation.

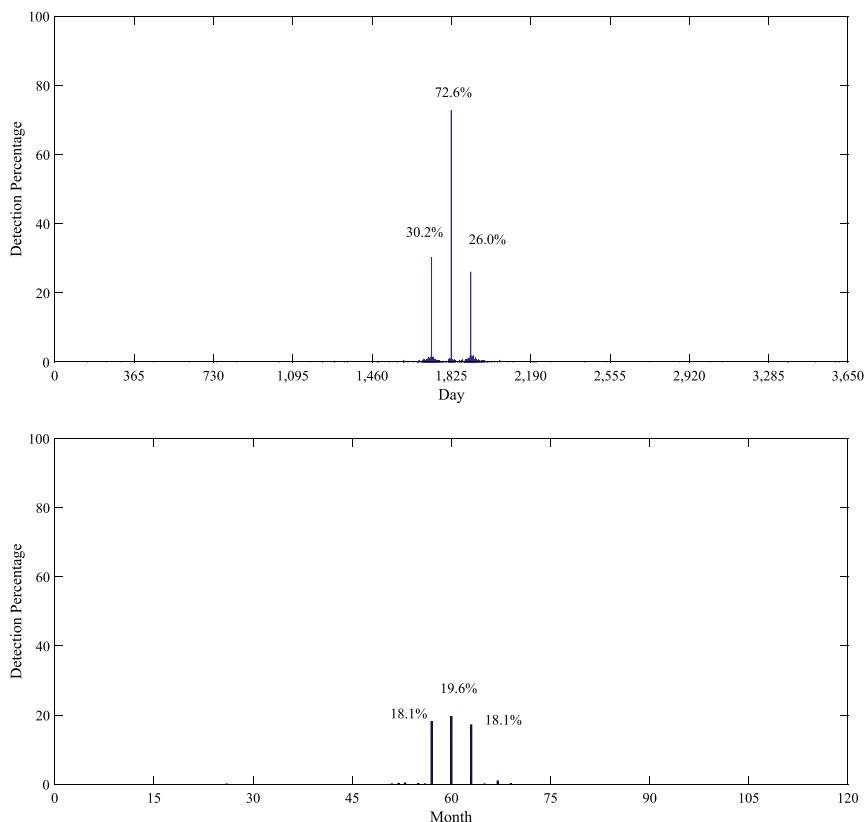| Shift size Δ (°F)	A single changepoint detected (%)	Mean of estimated shifts
0.5	9.4	0.765
−0.5	11.3	−0.797
1.0	95.6	1.029
−1.0	95.8	−1.028
1.5	99.9	1.508
−1.5	100.0	−1.504
2.0	100.0	1.997
−2.0	100.0	−1.993

FIG. 4. (top) Daily vs (bottom) monthly detection, aggregated from 1000 independent datasets.

to 31 December 1998 (46 yr). The bottom plot shows seasonally adjusted temperature anomalies, where a daily sample mean has been subtracted. Leap year data were omitted; hence, there are $365 \times 46 = 16790$ data points. The periodic mean cycle of the daily temperatures in Fig. 5 is evident; however, it is difficult to visually see any changepoints in these plots. To illuminate mean shifts and to lessen trends and seasonal cycles, reference series are often used (Menne and Williams 2005, 2009). For the South Haven series, reference series are available from the nearby Michigan stations at Shelby, Benton Harbor, and Pellston Regional Airport. We use Benton Harbor as our reference series since it is located on the eastern shore of Lake Michigan, like South Haven. The right panels in Fig. 5 show average daily and daily adjusted temperature anomalies at Benton Harbor. Even though subtraction of a reference series often lessens trends and seasonal mean cycles, these components are still retained in our model. Indeed, as Liu et al. (2016) shows, target minus reference subtractions often do not completely remove the seasonal cycle.

The records at South Haven (the target series) and Benton Harbor are mostly complete, with only a few sporadic missing data points (less than 1.3% of the record). For simplicity, missing data were infilled in our four series (maximums and minimums at the target and reference stations). To do this, a first-order vector autoregressive model was fitted to the four series in tandem.

TABLE 3. GA runtimes.

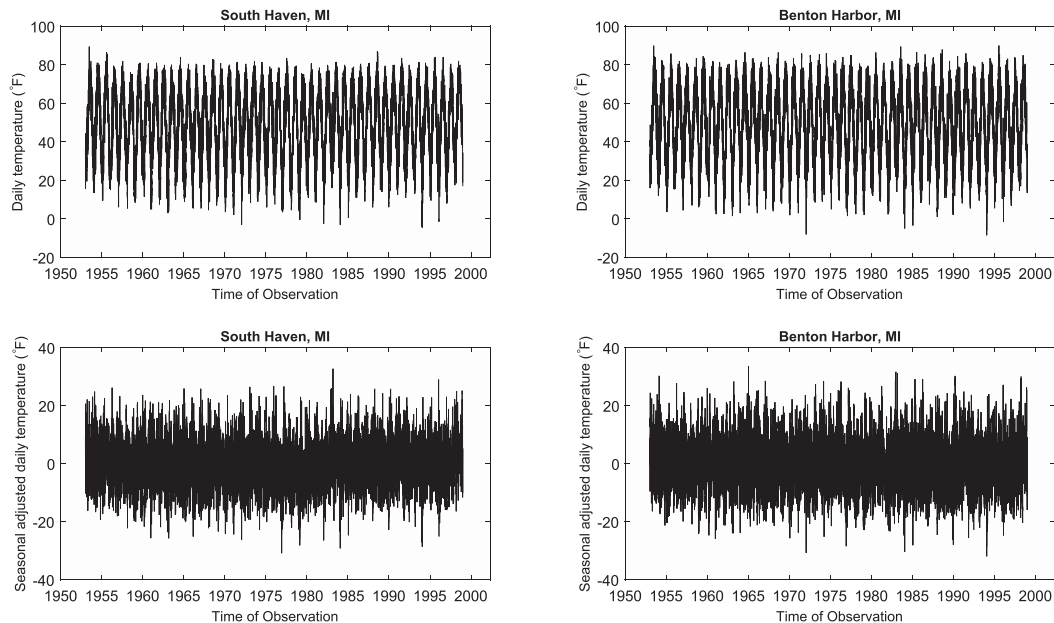| λ	Mutation prob.	Number of islands	Island size	Average runtime
1	0.0001	1	100	28.73
5	0.0001	1	100	38.39
10	0.0001	1	100	41.25
1	0.0001	2	50	27.02
5	0.0001	2	50	37.06
10	0.0001	2	50	36.37
1	0.0001	4	50	47.94
5	0.0001	4	50	44.48
10	0.0001	4	50	73.04
λ	Mutation prob.	Number of islands	Island size	Average runtime
1	0.00001	2	50	33.60
1	0.0001	2	50	27.86
1	0.001	2	50	43.10

FIG. 5. (left) South Haven daily average temperatures (top) before and (bottom) after subtracting a daily sample mean. (right) Analogous plots for the Benton Harbor station.

Missing data were infilled with best linear predictions. For example, if the maximum temperature of the reference series at time t was missing, this point was estimated by its best linear predictor from all nonmissing observations of the other three series at times t, $t - 1$, and $t + 1$. Runs of missing values were infilled one at a time.

Figure 6 plots the difference of daily average temperatures (daily average temperatures are the average of daily maximum and minimum temperatures) at South Haven and Benton Harbor. The graph appears to have some mean shifts, possibly attributable to either station. The metadata records for South Haven and Benton

Harbor list three changes from 1953 to 1998. According to South Haven's metadata, traditional liquid-in-glass maximum–minimum thermometers were replaced by electronic maximum–minimum temperature sensors on 22 August 1990. The station at Benton Harbor was relocated on 8 December 1993 and 19 June 1996. The 8 December 1993 relocation moved the station 600 ft south. Besides latitude and longitude details, the metadata do not provide a description of the second relocation. These three times were declared metadata times in the analysis. An island GA algorithm with two islands, a population size of 75 on each island, and 2000
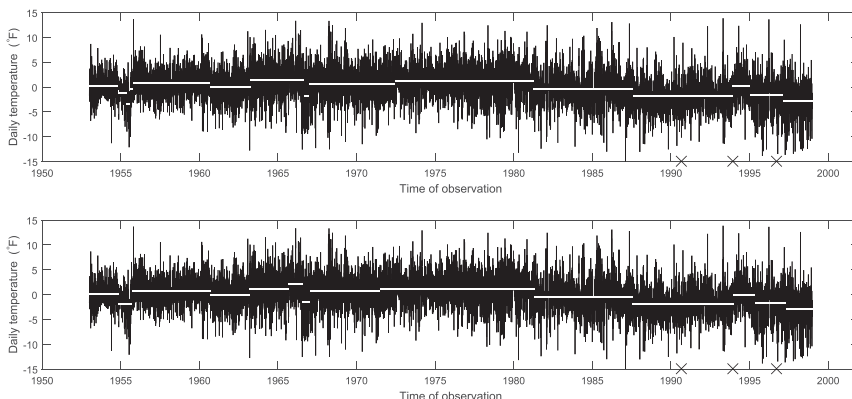

FIG. 6. The South Haven minus the Benton Harbor series, showing the changepoint structure (top) without and (bottom) with a linear trend. The estimated changepoint structure is superimposed on the graph and reveals 15 mean shifts without the linear trend and 13 mean shifts with a linear trend.

TABLE 4. Estimated changepoint times and corresponding mean shift sizes.

Changepoint	Time	Mean shift	Nearest metadata
666	27 Oct 1954	−1.3110°F	No metadata
673	3 Nov 1954	−1.3478°F	before 1990
868	17 May 1955	−3.5676°F	
955	12 Aug 1955	−0.6234°F	
1022	18 Oct 1955	0.6108°F	
2807	8 Sep 1960	−0.1867°F	
3750	9 Apr 1963	1.3158°F	
4985	27 Aug 1966	−2.0380°F	
5098	18 Dec 1966	0.6040°F	
7090	3 Jun 1972	1.1607°F	
10315	4 Apr 1981	−0.4865°F	
12632	9 Aug 1987	−1.9119°F	22 Aug 1990
14942	7 Dec 1993	0.2161°F	8 Dec 1993
15340	9 Jan 1995	−1.5768°F	8 Dec 1993
16104	12 Feb 1997	−2.7843°F	19 Jun 1996

generation iterations converged to a changepoint configuration with 13 changepoints (the bottom panel of Fig. 6). The runtime was about 19 h on a Dell optiPlex 9020 computer. Among the 13 flagged changepoints, only the 26 December 1993 changepoint is close to a metadata time (8 December 1993). This metadata time is the first station relocation of the Benton Harbor station. Neither the equipment change at South Haven nor the second relocation at Benton Harbor were judged to induce mean shifts. At the second relocation (19 June 1996), it is not clear if the station actually moved or if the latitude and longitude were updated to a higher precision.

The estimated PAR(1) autoregressive coefficients and their periodic variances are those displayed in Fig. 1. The estimated linear trend parameter is $-0.4440°$ F century^{-1} and has a standard error of $1.35°$ F century^{-1} (the standard error was computed with a time series regression model and allows for autocorrelation). Since the linear

trend is insignificant at the 95% significance level, and trend aspects are crucial in changepoint analyses (Gallagher et al. 2012), the target minus reference series was reanalyzed without a trend component. The resulting changepoint structure has 15 changepoints and is displayed in the top panel of Fig. 6. Table 4 displays the estimated changepoints, their occurrence times, and their corresponding mean shifts. Ten of the shifts move the series to colder regimes and five to warmer regimes.

To complement the daily analysis, annual and monthly target minus reference temperature series (Figs. 7 and 8) were also analyzed. The model in (1) with period 12 was fitted to the monthly averaged data. A GA was used to minimize the BMDL in (4) and revealed two changepoints at August 1980 and December 1987. For the annually averaged series, a multiple changepoint model with time-homogeneous AR(1) errors was fitted to the data. A GA analysis revealed six changepoints at the times 1955, 1956, 1992, 1994, 1995, and 1997. Figure 8 shows the changepoints of the annual target minus reference series. While 13 changepoints were found in the daily series, only five were found in the annual analysis, demonstrating the extra precision gained with daily series. In fact, dips in the series circa 1956 and 1967 are flagged as changepoints in the daily series, but not in the monthly series (even though the monthly graph is "suspect" at these two times).

7. Comments

This paper modified the BMDL techniques of Li et al. (2016) to accommodate daily temperature series. A BMDL objective function is minimized to estimate the best changepoint configuration. The BMDL here accounts for trends, metadata, seasonal means, autocorrelation, and seasonal variabilities. An island version of the GA was
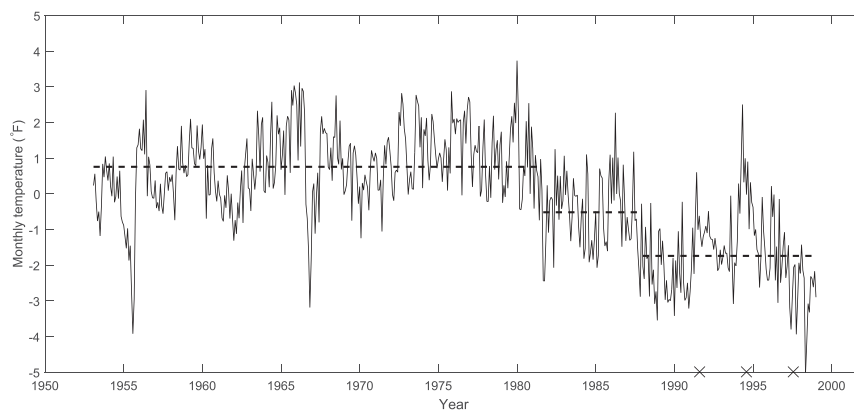

FIG. 7. Monthly South Haven minus Benton Harbor series with optimal changepoint configuration superimposed.
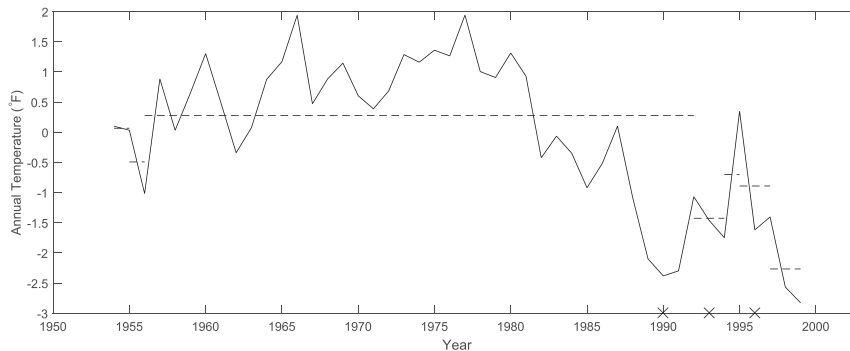
FIG. 8. Annual South Haven minus Benton Harbor series with optimal changepoint configuration superimposed.

implemented as a numerical optimization tool. Identifying changepoints in daily data is challenging due to long series lengths, large seasonal cycles, and the large number of model parameters.

The mean shift magnitudes in our model are non-seasonal; the mean shift changes temperatures on all days by the same amount. Should one expect a seasonal mean shift structure (say with winter shifts being larger than summer shifts), this could be allowed in the modeling procedure, although it would take work to accommodate such a structure. Future work might combine our techniques with the quantile matching methods of Trewin (2013) to investigate series changes that are not mean shifts.

The MDL methods here and elsewhere (Li and Lund 2012), which do not require data samples before and after a changepoint time to be large, may flag two changepoints at times close to each other. Often, this is suggestive of an outlying observation in need of confirmation or a run of outliers. While the time scale of homogenization is ultimately up to the homogenizer, MDL techniques also appear helpful in assessing data quality.

While our study examined temperature series, our methods can be applied to other climatic series with non-Gaussian dynamics. For example, Poisson-based likelihoods could be used for count series such as the monthly number of snow or thunderstorm days. While this research only considered univariate series, the methods could be modified to analyze multiple daily series.

Further improvements in computational speed of the algorithm are possible. The current GA runtimes make application of the methods to a large network of L temperature series infeasible, where all $\binom{L}{2}$ pairwise differences series need to be examined for changepoints. This said, multiple changepoint computing is an active area of current statistics research. Improvements in computer speed, GA parameter tuning, and methods such as wild binary segmentation (Fryzlewicz 2014) may render this drawback moot in the near future. Markov

chain Monte Carlo (MCMC) methods could also be used to help identify the optimal model; these techniques are developed in Li et al. (2016). Prescreening methods (Chan et al. 2014; Yau and Zhao 2016) that seed the genetic algorithm with initial chromosomes that are likely to be very good can further accelerate GA computational speed. For long series, it may be possible to analyze the series in smaller blocks.

Finally, it would be worthwhile to compare the detection methods here to some of the computer packages used in today's temperature homogenization problems; see Venema et al. (2012). Such a comparison, while beyond our scope here, should put all methods on the same footing. For example, with daily series that have high positive autocorrelation, one should penalize for false changepoint declarations, which would happen frequently if the method does not allow for autocorrelation.

REFERENCES

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

Caussinus, H., and O. Mestre, 2004: Detection and correction of artificial shifts in climate series. *J. Roy. Stat. Soc.*, **53C**, 405–425, doi:10.1111/j.1467-9876.2004.05155.x.

Cerf, R., 1998: Asymptotic convergence of genetic algorithms. *Adv. Appl. Probab.*, **30**, 521–550, doi:10.1017/S0001867800047418.

Chan, N. H., C. Y. Yau, and R.-M. Zhang, 2014: Group LASSO for structural break time series. *J. Amer. Stat. Assoc.*, **109**, 590–599, doi:10.1080/01621459.2013.866566.

Davis, R. A., T. C. M. Lee, and G. A. Rodrigues-Yam, 2006: Structural break estimation for nonstationary time series models. *J. Amer. Stat. Assoc.*, **101**, 223–239, doi:10.1198/016214505000000745.

Della-Marta, P. M., and H. Wanner, 2006: A method of homogenizing the extremes and mean of daily temperature measurements. *J. Climate*, **19**, 4179–4197, doi:10.1175/JCLI3855.1.

Fryzlewicz, P., 2014: Wild binary segmentation for multiple change-point detection. *Ann. Stat.*, **42**, 2243–2281, doi:10.1214/14-AOS1245.

Gallagher, C., R. Lund, and M. Robbins, 2012: Changepoint detection in daily precipitation series. *Environmetrics*, **23**, 407–419, doi:10.1002/env.2146.

Goldberg, D. E., and J. H. Holland, 1988: Genetic algorithms and machine learning. *Mach. Learn.*, **3**, 95–99, doi:10.1023/A:1022602019183.

Grünwald, P. D., I. J. Myung, and M. A. Pitt, 2005: *Advances in Minimum Description Length: Theory and Applications*. MIT Press, 444 pp.

Hansen, M. H., and B. Yu, 2001: Model selection and the principle of minimum description lengths. *J. Amer. Stat. Assoc.*, **96**, 746–774, doi:10.1198/016214501753168398.

Kuglitsch, F. G., A. Toreti, E. Xoplaki, P. M. Della-Marta, J. Luterbacher, and H. Wanner, 2009: Homogenization of daily maximum temperature series in the Mediterranean. *J. Geophys. Res.*, **114**, D15108, doi:10.1029/2008JD011606.

Li, S., and R. Lund, 2012: Multiple changepoint detection via genetic algorithms. *J. Climate*, **25**, 674–686, doi:10.1175/2011JCLI4055.1.

Li, Y., and R. Lund, 2015: Multiple changepoint detection using metadata. *J. Climate*, **28**, 4199–4216, doi:10.1175/JCLI-D-14-00442.1.

——, ——, and H. A. Priyadarshani, 2016: Bayesian minimal description lengths for multiple changepoint detection. [Available online at https://arxiv.org/abs/1511.07238.]

Liu, G., Q. Shao, R. Lund, and J. Woody, 2016: Testing for seasonal means in time series data. *Environmetrics*, **27**, 198–211, doi:10.1002/env.2383.

Lu, Q., and R. Lund, 2007: Simple linear regression with multiple level shifts. *Can. J. Stat.*, **35**, 447–458, doi:10.1002/cjs.5550350308.

——, ——, and T. Lee, 2010: An MDL approach to the climate segmentation problem. *Ann. Appl. Stat.*, **4**, 299–319, doi:10.1214/09-AOAS289.

Lund, R., and J. Reeves, 2002: Detection of undocumented changepoints: A revision of the two-phase regression model. *J. Climate*, **15**, 2547–2554, doi:10.1175/1520-0442(2002)015<2547:DOUCAR>2.0.CO;2.

——, H. Hurd, P. Bloomfield, and R. Smith, 1995: Climatological time series with periodic correlation. *J. Climate*, **8**, 2787–2809, doi:10.1175/1520-0442(1995)008<2787:CTSWPC>2.0.CO;2.

——, L. Seymour, and K. Kafadar, 2001: Temperature trends in the United States. *Environmetrics*, **12**, 673–690, doi:10.1002/env.468.

Menne, M. J., and C. N. Williams Jr., 2005: Detection of undocumented changepoints using multiple test statistics and composite reference series. *J. Climate*, **18**, 4271–4286, doi:10.1175/JCLI3524.1.

——, and ——, 2009: Homogenization of temperature series via pairwise comparisons. *J. Climate*, **22**, 1700–1717, doi:10.1175/2008JCLI2263.1.

Mitchell, J. M., 1953: On the causes of instrumentally observed secular temperature trends. *J. Meteor.*, **10**, 244–261, doi:10.1175/1520-0469(1953)010<0244:OTCOIO>2.0.CO;2.

Potter, K. W., 1981: Illustration of a new test for detecting a shift in mean in precipitation series. *Mon. Wea. Rev.*, **109**, 2040–2045, doi:10.1175/1520-0493(1981)109<2040:IOANTF>2.0.CO;2.

Reeves, J., J. Chen, X. Wang, R. Lund, and Q. Q. Lu, 2007: A review and comparison of changepoint detection techniques for climate data. *J. Appl. Meteor. Climatol.*, **46**, 900–915, doi:10.1175/JAM2493.1.

Rissanen, J., 1989: *Stochastic Complexity in Statistical Inquiry*. World Scientific Publishing, 188 pp.

Toreti, A., F. G. Kuglitsch, E. Xoplaki, and J. Luterbacher, 2012: A novel approach for the detection of inhomogeneities affecting climate time series. *J. Appl. Meteor. Climatol.*, **51**, 317–326, doi:10.1175/JAMC-D-10-05033.1.

Trewin, B., 2013: A daily homogenized temperature data set for Australia. *Int. J. Climatol.*, **33**, 1510–1529, doi:10.1002/joc.3530.

Venema, V., and Coauthors, 2012: Benchmarking homogenization algorithms for monthly data. *Climate Past*, **8**, 89–115, doi:10.5194/cp-8-89-2012.

Vincent, L. A., 1998: A technique for the identification of inhomogeneities in Canadian temperature series. *J. Climate*, **11**, 1094–1104, doi:10.1175/1520-0442(1998)011<1094:ATFTIO>2.0.CO;2.

——, and X. Zhang, 2002: Homogenization of daily temperatures over Canada. *J. Climate*, **15**, 1322–1334, doi:10.1175/1520-0442(2002)015<1322:HODTOC>2.0.CO;2.

Wang, X. L., H. Chen, Y. Wu, Y. Feng, and Q. Pu, 2010: New techniques for the detection and adjustment of shifts in daily precipitation data series. *J. Appl. Meteor. Climatol.*, **49**, 2416–2436, doi:10.1175/2010JAMC2376.1.

——, Y. Feng, and L. A. Vincent, 2014: Observed changes in one-in-20 year extremes of Canadian surface air temperatures. *Atmos.-Ocean*, **52**, 222–231, doi:10.1080/07055900.2013.818526.

Xu, W., Q. Li, X. L. Wang, S. Yang, L. Cao, and Y. Feng, 2013: Homogenization of Chinese daily surface air temperatures and analysis of trends in the extreme temperature indices. *J. Geophys. Res.*, **118**, 9708–9720, doi:10.1002/jgrd.50791.

Yau, C. Y., and Z. Zhao, 2016: Inference for multiple change points in time series via likelihood ratio scan statistics. *J. Roy. Stat. Soc.*, **78B**, 895–916, doi:10.1111/rssb.12139.