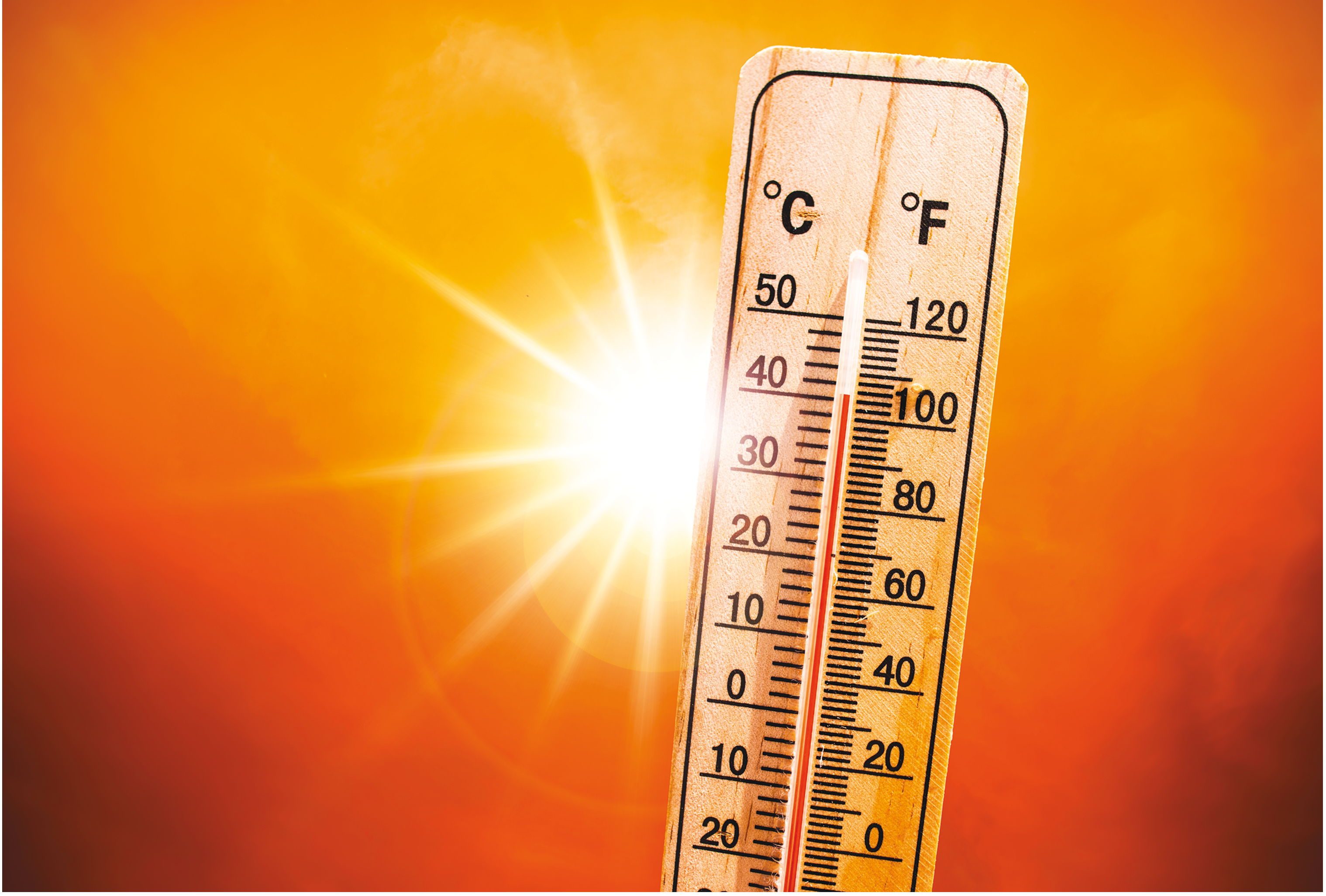
CHANCE

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/ucha20

Changepoint Methods in Climatology

Robert Lund & Xueheng Shi



Changepoint Methods in Climatology

Robert Lund and Xueheng Shi

Changepoints are discontinuity times (abrupt changes) in a time-ordered sequence of data. In climate settings, changepoints often occur when measuring stations are relocated or gauges are changed. Moving a climate station even 100 yards, for example, can shift temperatures by several degrees, especially if the solar exposure of the new location differs. In the United States, first-order climate stations average roughly six relocations or gauge changes per century. In many instances, these change times are not known to users analyzing the records. Changepoint times can also be triggered by natural causes when weather patterns shift.

A single changepoint for a random sequence $\{X_t\}_{t=1}^{N}$ of length N is an unknown time $\tau \in \{2,3,\ldots,N\}$ (time 1 is not allowed to be a changepoint) where X_t for $t \leq \tau$ has a particular marginal distribution, and this distribution changes after $t > \tau$. There are many ways in which the marginal distribution can change. Perhaps the simplest way allows series means to shift at the changepoint time, but keeps the same distribution type before and after the changepoint. While the mean shift case is the focus here, researchers have also studied shifts in process variabilities (volatilities) in finance and in autocovariances in speech recognition.

In the mean shift case, a regression model for the series is

$$X_t = E[X_t] + \epsilon_t,$$

where $E[X_t]$ is the mean of the series at time t and ϵ_t is random error with a zero mean and finite variance σ^2. The single mean shift case has the structure

$$E[X_t] = \begin{cases} \Delta_1, & 1 \leq t \leq \tau \\ \Delta_2, & \tau < t \leq N \end{cases}$$

for an unknown changepoint time $\tau \in \{2,3,\ldots,N\}$. The quantity $\Delta_2 - \Delta_1$ is the size of the mean shift. In the

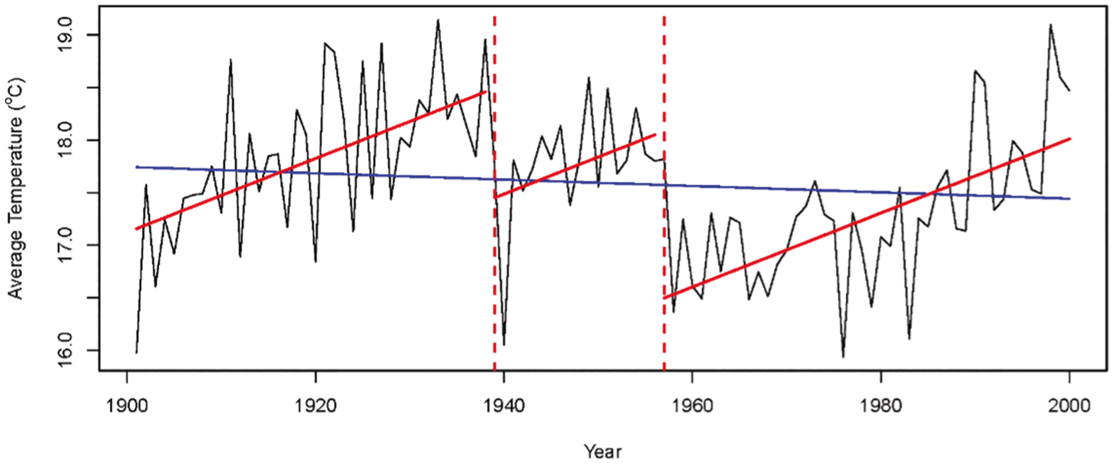


Figure 1. The Tuscaloosa temperature series in degrees Celsius. A fitted simple linear regression has a slightly negative trend slope. This slope becomes significantly positive when two mean shifts, estimated to occur in 1939 and 1957, are incorporated into the model fit.

multiple case where more than one changepoint might exist, m denotes the unknown number of changepoints. The piece-wise mean structure is

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

and allows each of the (m+1) "regimes" of the series to have a distinct mean.

Deleterious Effects of Changepoints

The havoc that changepoints wreak on trend estimation can be illustrated with an analysis of a century of annual temperatures in Tuscaloosa, Alabama, observed from 1901–2000. This series is plotted in Figure 1, along with two regression fits that are explained further below. The analysis examines the linear change rate of the series to assess climate change.

Suppose that we were naïve and neglected all changepoint issues. A ground-zero approach to the problem fits the simple linear regression $E[X_t] = \alpha + \beta t$ to the series. Here, α is a location parameter and β is the linear trend slope. Fitting this model provides the estimates

$$\hat{\alpha} = 17.74^{\circ}\text{C}$$
$$\hat{\beta} = -0.302^{\circ}\text{C/Century}$$

These estimates are not concerned with standard errors; rather, the trend slope estimate $\hat{\beta}$ is negative, suggesting that Tuscaloosa has cooled (albeit slightly) during the century. This fit is depicted by the unbroken line in Figure 1. Cooling is not necessarily absurd for this location in a global warming era: The Earth has not warmed uniformly by spatial location and the southeastern United States is known to have enjoyed somewhat of a "warming hole" during this era.

In contrast, consider a model fitted with both a linear trend and an unknown number of mean shifts. This is parametrized by $E[X_t] = \Delta_i + \beta t$ for times t in the i^{th} regime. The trend slope is the same in all regimes. The model is fitted by a penalized Gaussian likelihood with first-order autoregressive errors and a Bayesian Information Criterion (BIC) penalty. This fit estimates the number of changepoints m and their locations $\tau_1, \cdots, \tau_m$ and is plotted in Figure 1.

The fit estimates two changepoints at the years 1939 and 1957. While the fit improves on the simple linear fit, something sinister has happened: The trend slope sign has reversed, becoming a whopping 3.52°C per century and suggesting significant warming. With the Tuscaloosa series, the changepoint times are realistic—the extended station logs list a station relocation in 1939 and two gauge changes during 1956 (station logs, called metadata, are often unavailable).

The Tuscaloosa series is not pathological; indeed, a typical century-long temperature series has multiple changepoints during its record, with some shifts being a degree or two in magnitude. These shifts often move the series in opposite directions and make it a daunting task to estimate accurately a long-term trend, which is often less than the magnitude of one of the mean shifts over the record period. Due to the large number of recording stations measuring temperatures, trend estimation is not entirely hopeless; however, climatologists now realize that trends computed for a single station are untrustable when changepoint effects are neglected. In most changepoint problems, the fundamental issue lies with estimating the number of changepoints and where they occur; once this is done, most statistical procedures move in a straightforward manner that allows for a breakpoint at the identified changepoint times.

Changepoint Uses

Homogeneity Adjustments

Perhaps the chief use of changepoint methods in climatology is to homogenize series. Climate homogenizers seek to adjust series for any human-made effects by subtracting/adding estimates of the Δ_i's to the series. This rids the series of the human effects induced by gauge changes, station relocations, or other measurement change methods, leaving a record that can be attributed only to natural forcings. Climatologists prefer to leave changepoint effects from natural causes in the series, deeming them part of the natural record. Adjusting a series for an excessive number of changepoints risks removing an excessive amount of variation in the record, making other fluctuations seem more significant than they truly are.

To homogenize series, climatologists often compare a series under study, called the target series, to a series of the same quantity collected at a nearby recording station that experiences similar weather, called a reference series. Often, many reference series are used to make conclusions, sometimes as many as 40 in practice.

If $\{X_t\}$ denotes the target series and $\{Y_t\}$ the reference series, climatologists often analyze the difference $\{X_t - Y_t\}$ for changepoints. If the reference series matches the target series well, both series should experience similar fluctuations (weather) and the subtraction should eliminate most naturally occurring fluctuations. However, if the target series has a shift induced by a human-made change (such as a station relocation or a gauge change), then this time will also be a changepoint in the target minus reference series.

Unfortunately, by making target minus reference comparisons, changepoints in either the target or reference series become changepoints in the target minus reference series. This, in essence, doubles the number of changepoints in the series, complicating matters for the homogenizer. Modern homogeneity methods combat this issue by examining multiple target minus reference series: If the target minus reference series flags the same changepoint time for 10 distinct station references, then evidence suggests that the changepoint time is attributable to the target series.

There are stations where multiple good reference series do not exist; the North Atlantic Basin tropical cyclone analysis in the next subsection, for example, has no reference series whatever.

Stationarity Assessments

Another use of changepoint methods is with stationarity assessments. A common null hypothesis for climatologists is that some aspect of climate is static (stationary). Changepoint methods can be used to assess this hypothesis: If a multiple changepoint procedure flags one or more changepoints, then the record is deemed non-stationary. The form of $E[X_t]$ is not important here; linear trends, sinusoidal components, and even hinge-shaped structures have arisen in practice.

In general, a multiple mean shift changepoint procedure will flag changepoints in a pattern that attempts to follow the mean of the series. For example, if the series has an increasing linear trend, then a mean shift changepoint procedure will attempt to approximate this by flagging multiple changepoints that produce an "increasing staircase" for the series mean.

As an example of a stationarity check, consider the North Atlantic Ocean Basin's tropical cyclone counts for the 53-year period 1970–2022, plotted in Figure 2 along with a mean shift changepoint fit that is described below.

A Poisson marginal distribution provides a good model of annual tropical cyclone counts in the North Atlantic Basin. Hence, a reasonable model posits that the storm count for year t, denoted by X_t, , has a Poisson distribution with mean $E[X_t]$. Here, we conduct a multiple changepoint analysis with

$$E[X_t] = \begin{cases} \lambda_1, & 1 \le t \le \tau_1, \\ \lambda_2, & \tau_1 < t \le \tau_2, \\ \vdots & \vdots \\ \lambda_{m+1}, & \tau_m \le t \le N \end{cases}$$

The year-to-year storm counts are taken to be statistically independent; indeed, there is very little predictive power in a tropical cyclone count forecast even 12 months in advance (annual conferences are held each year in the spring, just before the start of hurricane season, where various modelers release their annual count forecasts). This fit optimizes a BIC-penalized likelihood based on independent Poisson counts from year to year.

The multiple changepoint fit estimates one changepoint in 1995. The estimated mean of the series is plotted in Figure 2 against the storm counts. Clearly, we have entered into an era of enhanced tropical cyclone activity circa the mid-1990s, a concerning aspect for East Coast residents.

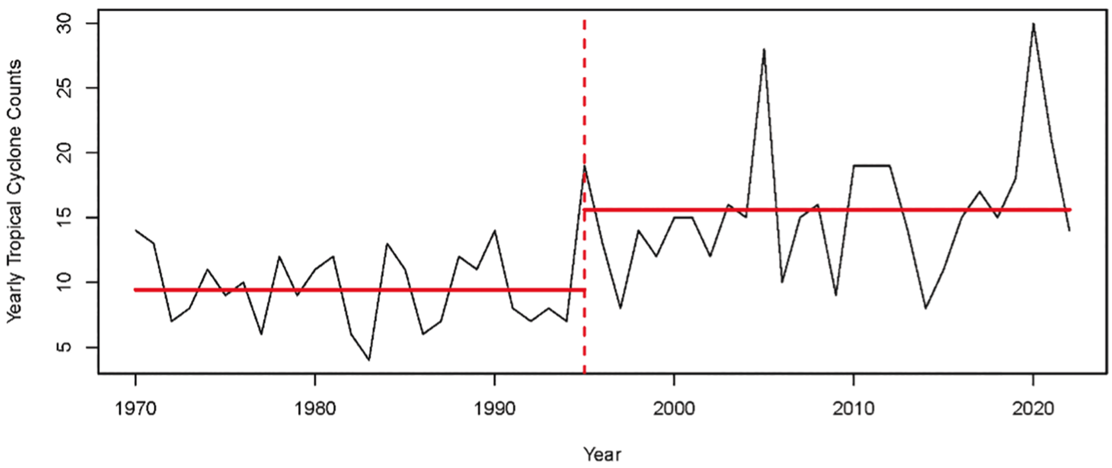


Figure 2. Tropical cyclone counts in the North Atlantic Ocean since 1970. The prominent changepoint in 1995 indicates that the record is non-stationary, with the region experiencing enhanced storm activity since that time.

Comments

Changepoint methods have multiple uses in climatology, including stationary checks and record homogenization. There are still many open problems in the area, especially in the multiple changepoint setting. Statisticians are needed to both help develop the methods and analyze the data.

These multiple changepoint configurations were estimated via a penalized likelihood with a BIC penalty, with a genetic algorithm to optimize the penalized likelihood. Penalized likelihoods select the changepoint configuration that gives the best model likelihood, subject to a cost placed on the changepoint configuration that prevents fitting too many changepoints. The penalty function used here was the classical BIC penalty $m \ln(N)$ for a model with m changepoints. The Tuscaloosa temperature series used Gaussian likelihoods with first-order autoregressive time series errors for the likelihood. For the North Atlantic tropical cyclone counts, the likelihood is based on independent Poisson counts. C

Further Reading

Davis, Richard A., Lee, Thomas, C. M., and Rodriguez-Yam, Gabriel A. 2006. Structural break estimation for nonstationary time series models. *Journal of the American Statistical Association* 223–239.

Lu, QiQi, and Lund, Robert B. 2007. Simple linear regression with multiple level shifts. *Canadian Journal of Statistics* 447–458.

Lund, Robert B., Beaulieu, Claudie, Killick, Rebecca, Lu, QiQi, and Shi, Xueheng. 2022. Good Practices and Common Pitfalls in Climate Time Series Changepoint Techniques: A Review. *arXiv preprint*.

Menne, Matthew J., and Williams, Jr., Claude N. 2009. Homogenization of temperature series via pairwise comparisons. *Journal of Climate* 1700–1717.

Robbins, Michael W., Gallagher, Colin M., and Lund, Robert B. 2016. A general regression changepoint test for time series data. *Journal of the American Statistical Association* 670–683.

Shi, Xueheng, Beaulieu, Claudie, Killick, Rebecca, and Lund, Robert B. 2022. Changepoint Detection: An Analysis of the Central England Temperature Series. *Journal of Climate* 2729–2742.

Shi, Xueheng, Gallagher, Colin, Lund, Robert B., and Killick, Rebecca. 2021. A Comparison of Single and Multiple Changepoint Techniques for Time Series Data. *Computational Statistics and Data Analysis*. *doi.org/10.1016/j.csda.2022.107433*.

About the Authors

Robert Lund is a professor and chair of the Statistics Department at the University of California - Santa Cruz. His research expertise lies in time series, statistical climatology, and stochastic processes.

Xueheng Shi is an assistant professor of statistics at the University of Nebraska – Lincoln. His research interests include time series, changepoint analysis, and statistical computing.