AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Atmospheres

RESEARCH ARTICLE

10.1002/2013JD021446

Key Points:
- Estimating the probability of upper tercile events in the U.S. temperature record
- The probability of the event depends on the statistical model fitted
- We consider statistical models with trend, seasonality, and serial dependence

Supporting Information:
- Readme
- Figures S1–S7, Tables S1–S8, and Text S1
- U.S. CONUS temperatures
- Code S1
- Code S2
- Code S3
- Code S4
- Code S5
- Code S6
- Code S7
- Code S8
- Code S9
- Code S10
- Code S11
- Data S1

Correspondence to:
P. F. Craigmile,
pfc@stat.osu.edu

Citation:
Craigmile, P. F., P. Guttorp, R. Lund, R. L. Smith, P. W. Thorne, and D. Arndt (2014), Warm streaks in the U.S. temperature record: What are the chances?, *J. Geophys. Res. Atmos.*, *119*, 5757–5766, doi:10.1002/2013JD021446.

Received 2 JAN 2014
Accepted 24 APR 2014
Accepted article online 29 APR 2014
Published online 19 MAY 2014

Warm streaks in the U.S. temperature record: What are the chances?

Peter F. Craigmile[1,2], Peter Guttorp[3,4], Robert Lund[5], Richard L. Smith[6,7], Peter W. Thorne[8,9,10], and Derek Arndt[8]

[1]Department of Statistics, Ohio State University, Columbus, Ohio, USA, [2]School of Mathematics and Statistics, University of Glasgow, Scotland, UK, [3]Department of Statistics, University of Washington, Seattle, Washington, USA, [4]Norwegian Computing Center, Oslo, Norway, [5]Department of Mathematical Sciences, Clemson University, Clemson, South Carolina, USA, [6]Department of Statistics and Operations Research, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA, [7]Statistical and Applied Mathematical Sciences Institute, Research Triangle Park, North Carolina, USA, [8]National Climatic Data Center, Asheville, North Carolina, USA, [9]Department of Marine, Earth and Atmospheric Sciences, North Carolina State University at Raleigh, Raleigh, North Carolina, USA, [10]Nansen Environment and Remote Sensing Center, Bergen, Norway

Abstract A recent observation in NOAA's National Climatic Data Center's monthly assessment of the state of the climate was that contiguous U.S. average monthly temperatures were in the top third of monthly ranked historical temperatures for 13 straight months from June 2011 to June 2012. The chance of such a streak occurring randomly was quoted as $(1/3)^{13}$, or about one in 1.6 million. The streak continued for three more months before the October 2012 value dropped below the upper tercile. The climate system displays a degree of persistence that increases this probability relative to the assumption of independence. This paper puts forth different statistical techniques that more accurately quantify the probability of this and other such streaks. We consider how much more likely streaks are when an underlying warming trend is accounted for in the record, the chance of streaks occurring anywhere in the record, and the distribution of the record's longest streak.

1. Introduction

The U.S. National Oceanographic and Atmospheric Administration (NOAA) monitoring service provides real-time assessments of the climate in an accessible way to the public and other stakeholders. This type of reporting is challenging because it necessitates fast data processing, analysis, and reporting of results. In their monthly assessment of the state of the climate, NOAA made the following statement in July 2012:

"During the June 2011–June 2012 period, each of the 13 consecutive months ranked among the warmest third of their historical distribution for the first time in the 1895–present record. The odds of this occurring randomly is 1 in 1,594,323" [*NOAA*, 2012].

The statement, and in particular its subsequent reporting in traditional and nontraditional media, engendered much lively discussion in multiple fora. The streak continued for an additional 3 months after the statement and was broken when October 2012's temperature checked in slightly under the October tercile (upper third).

The above probability quote, which is 1 in 3^{13} (or 1 in 3^{16} if the longer streak is considered), assumes that the temperatures for each month are statistically independent from all other months and that only the most recent 13 (16) months were considered. We show that such month-to-month independence is invalid, as is well recognized within the climate community, and put forth several ways of more accurately estimating the chances of such streaks anywhere in the record. Specifically, we devise realistic estimates of this streak probability when the record is stationary and when a trend is allowed; the latter quantifies how streak probabilities change in a warming climate. Unlike the scientists involved in the original statement we can perform a less time-bound statistical analysis of the problem, which includes fewer simplifying assumptions. We hope that some of our statistical models and insights can be used to improve real-time reporting of such streaks in the future.

Before proceeding, we clarify what we mean by an upper tercile event. The practice of NOAA's National Climatic Data Center (NCDC) is to declare whether a given month's temperature is an upper tercile event

©2014. American Geophysical Union. All Rights Reserved.

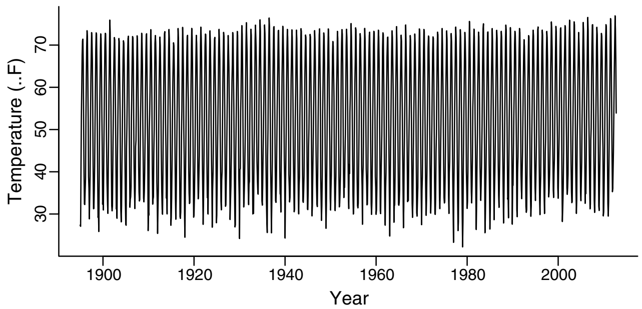

Figure 1. Time series of CONUS average monthly temperatures used in undertaking regular monthly NCDC monitoring reports. Version used: November 2012 report.

relative to all current and previous months in the record. This choice is driven by the stated mission of monitoring activities at NOAA to place the current period of record in historical context. Phrased mathematically, suppose there are N past and current temperatures for a given month. Let n be the nearest integer to $N/3$. The observation is declared an upper tercile event if and only if it is one of the largest n values among the N past and current observations for that month. Thus, a streak of 16 upper tercile events entails 16 consecutive monthly temperatures in the upper third of their historical record.

The remainder of this paper is arranged as follows. In section 2, brief characteristics of the monthly temperature data used in NCDC's monitoring products are established. Section 3 supplies probability streak calculations for independent data. Here we show that the monthly temperatures in question are dependent. Section 4 describes the individual statistical models and methods of estimation used to derive our various streak probability estimates. Section 5 presents and compares the estimated upper tercile streak probabilities under different model assumptions, including the absence and presence of a warming trend. Section 6 provides several comments, and section 7 summarizes our findings. The supporting information provides details of the modeling and computer code used in four different statistical analyses of the temperature data. The results we provide in this article are drawn from these analyses.

2. The Data

The data used in this analysis are derived from the U.S. Historical Climatology Network, Version 2.5 [*Menne et al.*, 2012] for the conterminous United States (the contiguous 48 states, abbreviated to CONUS). These data are a subset of the broader U.S. Cooperative Observer network (COOP) sites supported by NOAA's National Weather Service. This subset was assigned in the 1980s based upon geographical coverage and station completeness. In the Version 2.5 product, the stations have been homogenized using the NCDC's Pairwise Homogenization Algorithm [*Menne et al.*, 2009; *Williams et al.*, 2012a, 2012b], which makes pairwise comparisons between multiple neighboring stations to identify and adjust for nonclimatic artifacts. The adjustments use the full COOP network, but only long-term USHCN designated stations (about 1/6 of the total stations) are retained and analyzed after adjustment. The homogenized station data were then spatially averaged by NCDC's monitoring branch to create a national time series, which is studied here. The data contain monthly temperatures for the 118 year period from 1895 to 2012. We analyze the data through October 2012, giving 1414 monthly values.

Figure 1 shows a time series plot of the CONUS temperature series. There is a clear seasonal cycle on the order of 40°F that is much larger than any long-term warming trend. Consequently, any long-term trend is hard to discern without first deseasonalizing the series. Winter months are cooler and more variable than summer months. This is expected, given the partitioning of moist and dry energy terms with absolute temperature [*Peterson et al.*, 2011]. Fitting linear trends to the series for each month by ordinary least squares, the impact of seasonality on the trend becomes apparent (Table 1), with February showing the most warming (2.78°F per century) and October the least (0.57°F per century). Figure 2 shows the monthly distributions of temperatures throughout the record. On average, the warmest month is July and the coldest is January.

Table 1. Linear Trend Slope by Month in °F Per Century

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Slope	0.86	2.78	1.99	1.08	1.27	1.28	1.49	1.35	0.80	0.57	1.02	1.04

 ©2014. American Geophysical Union. All Rights Reserved.

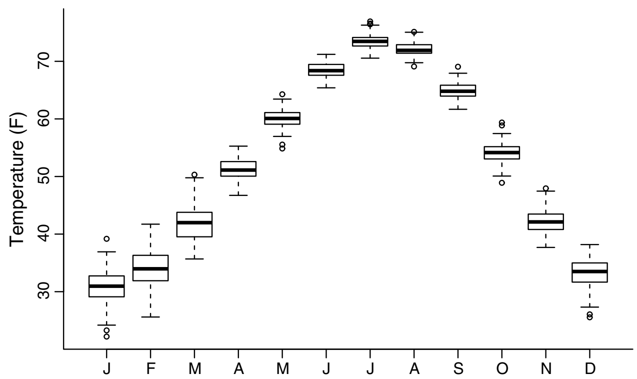

Figure 2. Side-by-side box plots [*Tukey*, 1977] of CONUS average monthly temperatures in °F. The thick horizontal line is the median, the box indicates the first and third quartiles (Q_1 and Q_3), and the whisker extends to the most extreme data point within 1.5 box heights (1.5 times the interquartile range, Q_3–Q_1). Remaining (even more extreme) data are plotted as circles.

Figure 3 shows a time series plot of the lower, middle, and upper tercile streaks for the CONUS series as sequentially calculated from the start of the record. Up to the year 2000, the longest upper tercile streak is 8 months. A single 12 month upper tercile streak ends in October 2000. The longest streak in the lower tercile is 6 months and occurs early in the record. The middle tercile has a number of streaks exceeding 5 months, with the longest streak of 13 months occurring early in the record. Long upper tercile streaks seem to have become more common since the late 1970s.

The 16 month streak being studied is the longest in any tercile in the entire record. While the choice of the tercile as our threshold of analysis is somewhat arbitrary, any choice between the top 42% and the top 26% has the same historical 16 consecutive month streak exceeding that percentile. Hence, the streak is relatively robust to the exact choice of threshold.

3. Streak Probability Calculations

The calculation given by *NOAA* [2012] explicitly assumes that all months are independent, so that the probability that a given month's temperature exceeds its historical tercile is 1/3. One can view the month-to-month temperatures and the tercile as a sequence of biased coin tosses, where the probability

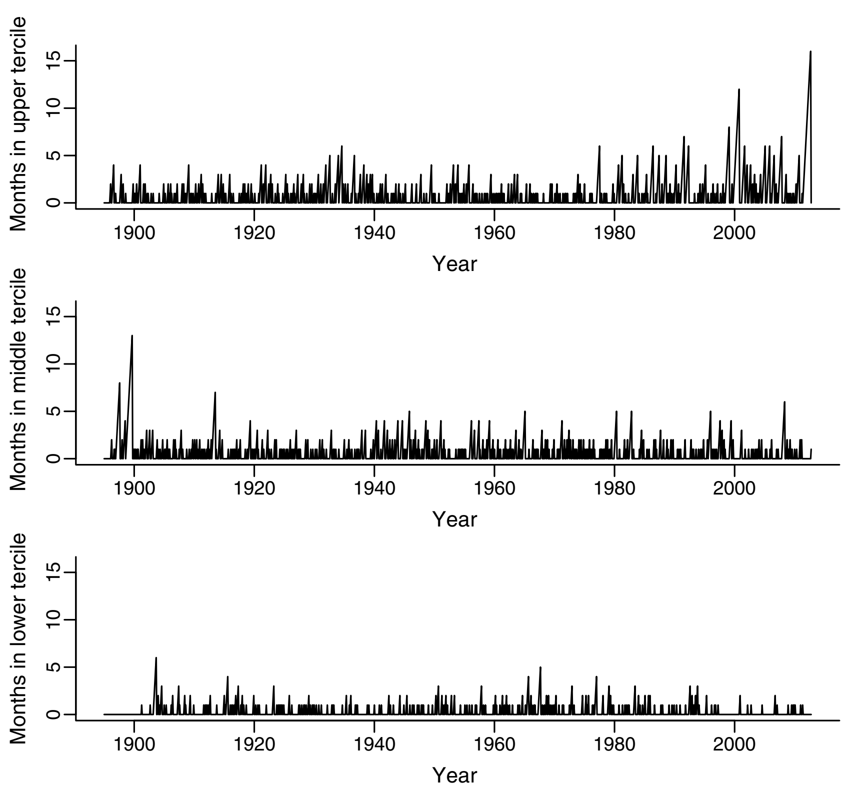

Figure 3. Time series plots of the number of consecutive months in the lower, middle, and upper terciles for the CONUS average monthly temperature record.

©2014. American Geophysical Union. All Rights Reserved.

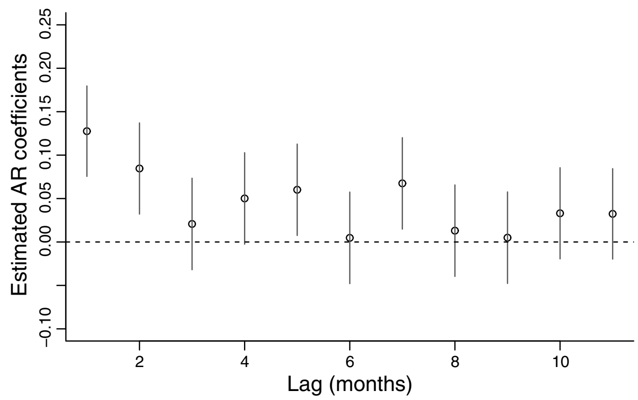

Figure 4. Estimated autoregressive (AR) coefficients for the monthly mean CONUS corrected series. The vertical lines are asymptotic pointwise 95% confidence intervals using the mle option in the ar function in R [*R Core Team*, 2012].

of heads (a top-tercile exceedence) in any toss is $p = 1/3$ (there is a 1/3 chance of an upper tercile event). When the tosses are independent, the chance of a particular $k = 16$ month sequence resulting in all heads is indeed $p^k = 3^{-16} = 2.323 \times 10^{-8}$, or about one in 43,046,721. This chance is not the probability that a 16 month streak of heads occurs *anywhere* within the first $N = 1414$ tosses. In tossing a coin independently with heads probability p ad infinitum, one expects to wait

$$\frac{1}{p} + \cdots + \frac{1}{p^k} = \frac{1 - p^{-k}}{p - 1}$$

tosses until encountering the first streak of k heads. For a run of 16 heads, one expects to wait 64,570,080 tosses. The probability of not encountering a streak of k or more heads in the N trials has a complicated expression [*Feller*, 1968, section XIII.7]. The chance of seeing at least one run of 16 heads in $N = 1414$ trials is approximately 2.166×10^{-5}, or about 1 in 46,171.

Independence between months is essential in the above calculations but seems questionable. An assumption of a stationary, but not independent, monthly temperature series may be more reasonable. When the coin flips are mathematically regarded as a stationary sequence (this assumes far less than month-to-month independence), streak probabilities can drastically change. Although, in principle, it is possible to construct stationary time series for which the probability of a 16 month upper tercile streak is anywhere between 0 and 1/3, in practice, we restrict our attention to time series models that are consistent with the observed data.

In climate modeling, autoregressive (AR) processes, often of first order, are frequently used to capture dependence impacts in significance assessments [e.g., *Santer et al.*, 2008]. Figure 4 shows autoregressive coefficient estimates up to order 11 for our series (monthly sample means were first subtracted). The vertical lines are asymptotic 95% pointwise confidence intervals. These intervals show several significantly nonzero coefficients and imply that first-order autoregressive models are not appropriate here. Specifically, the AR coefficient estimates are significantly nonzero at lags 1, 2, 5, and 7. Nevertheless, AR time series processes are too restrictive for our present purposes and the bulk of the paper is concerned with two more general classes of models: those that combine autoregressive and moving-average components (abbreviated ARMA) and those based on fractional differencing (FD).

The next section describes the time series models we use in detail. These models are then used to simulate streak probabilities.

4. Methods of Analysis

Several time series models can reasonably describe the CONUS series. As the upper tercile streak probability will depend on the model choice, the use of different statistical models recognizes that there is no single correct way to estimate the streak's probability. In this paper, four statisticians analyzed the data independently, using different statistical models, though they shared intermediate results and made an effort to ensure that the approaches span a range of plausible approaches to the problem. In this section, we summarize the statistical models and methods of estimation. Section 5 provides the estimated streak probabilities from these models. The detailed analyses (including code) are described by each of the four statisticians in the supporting information.

Some general remarks are in order before proceeding. First, diagnostic tests indicate that the CONUS average monthly temperature series has approximately Gaussian marginal distributions. Second, all analyses take the periodicity in the mean and standard deviation (see Figure 2) into account in some manner.

©2014. American Geophysical Union. All Rights Reserved.

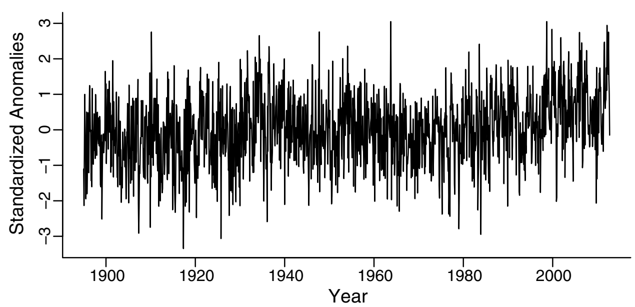

Figure 5. Time series plot of the standardized anomalies of CONUS temperatures.

4.1. Stationary Models With Seasonal Cycles

Let $\{X_t : t = 1, \dots, N\}$ denote the $N = 1414$ CONUS monthly temperatures. Let $m(t)$ denote the month of time t (January is 1, February is 2, …, and December is 12). Perhaps the simplest model for the data has the periodic stationary form

$$X_t = \alpha_{m(t)} + \sigma_{m(t)} Z_t, \qquad t = 1, \dots, N. \tag{1}$$

Here α_ν is the mean temperature of month ν and σ_ν is the standard deviation of the month ν temperatures, $\nu = 1, \dots, 12$. The model errors $\{Z_t\}$ are assumed to be a zero-mean unit-variance stationary time series. There is no trend in this model; section 4.2 modifies the above scenario to allow for trends.

Estimates of α_ν and σ_ν, denoted by $\hat{\alpha}_\nu$ and $\hat{\sigma}_\nu$, are simply the monthly sample averages and standard deviations, respectively. The seasonally standardized anomalies $\{\hat{Z}_t\}$ are calculated from these estimates via

$$\hat{Z}_t = \frac{X_t - \hat{\alpha}_{m(t)}}{\hat{\sigma}_{m(t)}}, \qquad t = 1, \dots, N.$$

Figure 5 shows a time series plot of $\{\hat{Z}_t\}$ and reveals a trend or long-range dependence. As seen in section 2, this series is not white noise. Thus, various stationary time series models will be considered for $\{\hat{Z}_t\}$ and compared for quality of fit.

One type of stationary model for $\{\hat{Z}_t\}$ employs an autoregressive moving-average ARMA(p, q) process [see *Brockwell and Davis*, 2002] of orders p and q. ARMA series parsimoniously describe many stationary time series. To select the autoregressive order p and the moving-average order q, we use the Akaike Information Criterion (AIC) [*Akaike*, 1974] or its asymptotically corrected version AICC [*Hurvich and Tsai*, 1988; *Brockwell and Davis*, 2002]. ARMA models are short-range dependent processes with autocorrelations that decay geometrically to 0 with increasing lag.

As an alternative, long-range dependent time series models for $\{\hat{Z}_t\}$ are also considered. Long-range dependent (also called long memory) processes are series in which the autocorrelations decay polynomially to 0 with increasing lag [see, e.g., *Beran*, 1994; *Palma*, 2007]. These models are popular because they can capture long-range variations in the atmosphere and are commonly used as temperature models [e.g., *Caballero et al.*, 2002; *Király and Jánosi*, 2005; *Craigmile and Guttorp*, 2011]. The simplest example of a long memory process is the fractionally differenced (FD) process [*Granger and Joyeux*, 1980; *Hosking*, 1981], although more complicated models such as autoregressive fractionally integrated moving averages and the fractional exponential models were also considered.

The simplest way to fit the model in (1) first estimates the α_ν and σ_ν parameters as the monthly sample means and standard deviations and then estimates the time series model from $\{\hat{Z}_t\}$. A more advanced paradigm estimates the α_ν and σ_ν parameters jointly with the parameters governing $\{Z_t\}$. In many cases, this makes little difference to the quality of fit compared with separately estimating the $\{\alpha_\nu, \sigma_\nu : \nu = 1, \dots, 12\}$ and time series parameters, but for trend models based on (2) below, the two parts of the estimation problem cannot be cleanly separated so we have to consider some form of joint estimation. The estimation techniques that were used include generalized least squares, maximum likelihood, and Bayesian estimation.

4.2. Time Series Models With Trends

Figure 5 suggests that the standardized anomalies are increasing over the period of study. Thus, we consider a variety of models with a trend. As before, suppose that $\{Z_t\}$ is a zero-mean Gaussian series, $\{\alpha_\nu\}_{\nu=1}^{12}$ are monthly location parameters, and $\{\sigma_\nu\}_{\nu=1}^{12}$ are seasonal scale parameters. Additionally, let y_t denote a time index at time t. Two possible ways of defining this index are $y_t = t$ or y_t is the year corresponding to month t. Either way, a model for the temperatures $\{X_t\}$ that allows for a linear trend in the temperatures is

$$X_t = \alpha_{m(t)} + \beta y_t + \sigma_{m(t)} Z_t, \qquad t = 1, \dots, N. \tag{2}$$

©2014. American Geophysical Union. All Rights Reserved.

This model is a version of the periodic regression model adopted in *Lund et al.* [1995] and *Craigmile and Guttorp* [2011].

A more general nonlinear temperature trend model is

$$X_t = \alpha_{m(t)} + \mu_t + \sigma_{m(t)} Z_t, \tag{3}$$

where $\{\mu_t\}$ is a nonlinear trend component which smoothly varies over time, perhaps based on splines or wavelets.

An alternative parameterization of the trend supposes linearity after a standardization:

$$\frac{X_t - \alpha_{m(t)}}{\sigma_{m(t)}} = \beta y_t + Z_t, \qquad t = 1, \ldots, N,$$

which is equivalent to

$$X_t = \alpha_{m(t)} + \beta \sigma_{m(t)} y_t + \sigma_{m(t)} Z_t, \qquad t = 1, \ldots, N. \tag{4}$$

A generalization of this model allows the linear trend to vary by month:

$$X_t = \alpha_{m(t)} + \beta_{m(t)} \sigma_{m(t)} y_t + \sigma_{m(t)} Z_t, \qquad t = 1, \ldots, N. \tag{5}$$

As in the previous subsection, we consider a number of different stationary models for $\{Z_t\}$ in our above regression equations.

Comparing different ARMA model fits, we often found that a high-order model (such as ARMA(7,5)) had the smallest AIC or AICC. However, the AIC criterion tends to overparametrize [e.g., *Hurvich and Tsai*, 1988]. Specifically, the constraints on the ARMA model parameters tend to make the optimization procedure unstable. Lower order models often had nearly equivalent AICC scores without incurring the same instability issues. We therefore concentrate on two relatively low-order models: ARMA(3,1) and ARMA(4,2). Residuals from the ARMA(3,1) and ARMA(4,2) passed several white noise tests, including spectral analysis, further ARMA-fitting, and autocorrelation analysis, indicating a satisfactory fit. The higher-order ARMA models yield similar results in terms of simulated streak probabilities (see the supporting information for details).

As mentioned above, an alternative to ARMA modeling involves long-range dependent models. Simple fractionally differenced (FD) processes provide good fits to the data. Our analyses include both a simple FD model estimated by a joint maximum likelihood (FD-ML) approach and a Bayesian approach (FD-Bayes).

5. Estimating the Upper Tercile Streak Probability

We estimate the distribution of upper tercile streaks by simulation. Simulating many CONUS series from each model allows us to estimate the probability of obtaining an upper tercile streak length of at least 16 months in 1414 months, along with a measure of the uncertainty of that estimate. The simulation scheme is simple: for $k = 1, \ldots, K$, independent time series from a given statistical model are simulated. In simulation k, the longest upper tercile streak R_k is computed. Our estimate of obtaining an upper tercile streak of at least L months is

$$\hat{p} = \frac{\sum_{k=1}^{K} I(R_k \geq L)}{K},$$

where $I(A)$ is an indicator function defined to have value 1 if the statement A is true and 0 otherwise. This estimated probability has an estimated standard error (ese) of $\text{ese}\left(\hat{p}\right) = \sqrt{\hat{p}\left(1 - \hat{p}\right)/K}$. We take $L = 16$ to focus on the streak length of interest. In most cases, $K = 100{,}000$ simulations were conducted. For nearly all cases, $\hat{p} < 0.1$ and the standard error is hence less than 0.001. Some of the simulations also allow for variability of the parameter estimates by using the normality of estimates from Gaussian series or by sampling from the posterior distribution in the case of a Bayesian analysis.

Our first set of simulations assumes that the seasonally standardized anomalies are stationary as in (1) and considers different Gaussian time series models for $\{Z_t\}$. One should consider these as streak probability

©2014. American Geophysical Union. All Rights Reserved.

Table 2. Estimate of the Probability of Obtaining an Upper Tercile Streak of at Least 16 Months, Assuming Different Statistical Models for the Temperature[a]

	Assumption for $\{Z_t\}$			
Model	ARMA(3,1)	ARMA(4,2)	FD-ML	FD-Bayes
Stationary model (1)	0.031	0.034	0.019	0.035
Trend model (2)	0.065	0.069	0.116	0.145
Model (2), zero slope	0.007	0.008	0.008	0.013
Nonlinear trend model (3)	0.135	0.141	0.269	0.163
% Increase	830	790	1260	1030

[a]The last line shows the percentage increase in the probability as we go from model (2) with a zero slope to model (2) with the actual slope observed for the temperature series. The first three columns are taken from Table S8 of the supporting information and agree (subject to the margin of error) with results presented in Table S3 and on p. 12 of Text S1. The last column for the fractionally differenced Bayesian model is taken from the $p = 0$ case of Figure S5.

estimates under a stationary climate (seasonal periodicities are allowed). To generate a realization of X_t, a simulated Z_t is multiplied by the estimated σ_t and then added to the estimated α_t. The estimated probabilities in the stationary case are summarized in first line of Table 2 and are clearly much larger than $(1/3)^{16}$.

The second set of simulations move to cases with a linear trend as in (2). The only difference to the previous set is that a linear trend is added to the simulated values to allow for climate change (in fashions depending on the model). These models are listed in the second line of Table 2. The streak probabilities are much higher.

The third set of simulations views, in essence, the residuals from the linear trend fit as a model of the natural variability of the CONUS climate. In this case, we use model (2) but set the slope parameter, β, equal to 0. This demonstrates how the streak probabilities would behave under the assumption of stationary climate, observed without the trend component. The results from simulations based on these models are given in the third line of Table 2. The fourth set of simulations is based on nonlinear trends, as in model (3), represented by either splines or wavelets and are given in the fourth line of Table 2.

A number of comparisons are possible based on these simulations, but we focus on the effects of including a linear trend. Specifically, the fifth line of the table shows how much more likely the observed streak is in a changing climate than in a stationary one, based on the second and third lines of Table 2.

Figure 6 shows a histogram of the maximum streak lengths when $\{Z_t\}$ is an ARMA(3,1) process under three different trend assumptions. All histograms are right skewed. The centers and spread of the distributions vary by model; for example, the streak distribution clearly has smallest spread for the linear trend model with a zero slope.

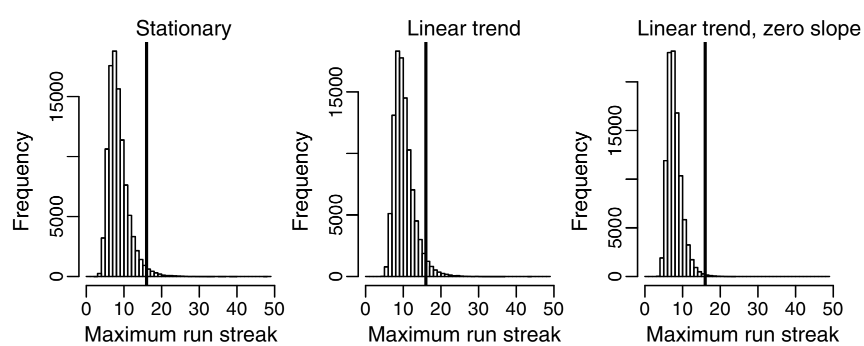

Figure 6. Histograms of the maximum run of upper tercile streak when $\{Z_t\}$ is an ARMA(3,1) process for different assumptions made for the trend.

©2014. American Geophysical Union. All Rights Reserved.

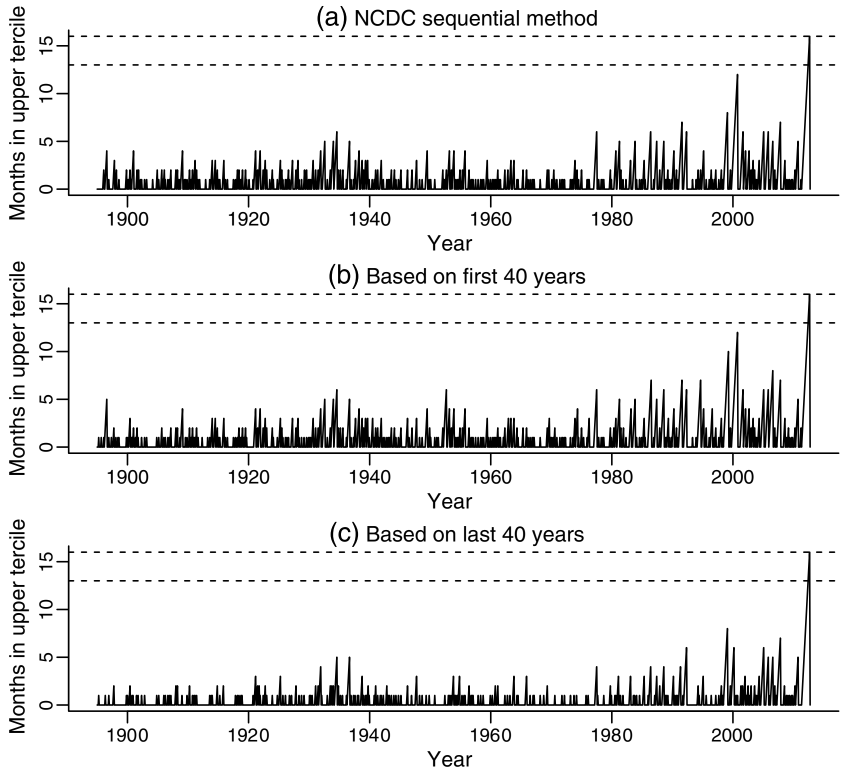

Figure 7. Time series plots of the number of months in the upper terciles for the CONUS average monthly temperature record, according to calculating the upper tercile (a) using the NCDC sequential method, (b) based on the first 40 years, and (c) based on the last 40 years. The horizontal dashed lines in each panel denote 13 and 16 months, respectively.

6. Discussion

Our analysis takes the CONUS record as correct. There are at least two issues with this. First, our analysis does not consider uncertainties involved with changepoint adjustments made on the data [*Williams et al.*, 2012c; *Muller et al.*, 2013; *Fall et al.*, 2011]. Changepoint adjustments are made on the individual series for station moves, instrument changes, etc. There is substantial evidence that the NCDC algorithm used to homogenize the U.S. temperature record is conservative [*Williams et al.*, 2012c, *Venema et al.*, 2012] in that the true long-term warming trend is underestimated. Second, the uncertainty in the CONUS series due to spatial averaging has not been taken into account. Different plausible ways to calculate a CONUS-type series for the U.S. from the spatially and temporally incomplete station network could produce different conclusions, permitting an assessment of uncertainty. Having such series would allow us to improve our trend estimates. *Shen et al.* [2012] calculated some uncertainties for a variant of the CONUS data; the results changed the significance of the trend estimate [*Guttorp and Kim*, 2013].

Our analysis broadly adopts the tercile definitions of NOAA's NCDC. Specifically, the terciles are updated with each new data point. For example, when a new January temperature is added, January terciles are recalculated from this and all previous January measurements. This is one of several approaches that could be employed in creating terciles. One could equally well decide that terciles should be based on a recent subperiod of the record to reflect the recent range of normal, which the general public may better understand. One could also use a fixed historical period to compute terciles to show how climate has changed relative to this past period. Approaches exist that are intermediate between these. Figure 7 shows how NCDC upper tercile streak lengths change when the terciles are computed from the first and last 40 years. Additional data will inevitably change some of the terciles. However, as noted in section 2, the particular streak of 16 months was very robust to the actual quantile used. Substantial changes in future data would be needed to invalidate the 16 month streak.

The comparison in the bottom line of Table 2 is only one of many possible comparisons. It is a rough attempt to compare the streak probability based on a linear trend to what it would be if there were no such trend.

©2014. American Geophysical Union. All Rights Reserved.

Apparently, the series in Figure 5 is not stationary; hence, fitting a stationary model to the data tends to push the parameter estimates toward a nonstationary condition. In effect, the fitted stationary model is doing its best to reproduce the observed nonstationarity. This explains why the probabilities estimated under the stationary model are closer to those estimated from the trend model than to those estimated when the trend is set to 0.

Finally, as discussed in section 3 there is a distinction to be made between the chances of a streak arising anywhere within the series and the chances that it arises in a given location within the series. The latter is a specific subset of the former, with a smaller probability as a result. Therefore, it is important to understand the hypothesis being tested. NOAA's monitoring activities are generally assessing the probability of the restricted test of the streak arising at a fixed location of "present" at any given time. Such location-specific probabilities will always be lower than the probability of an event arising anywhere within a series. In the presence of a trend they also will not be time-invariant.

7. Conclusions

The streak of 16 consecutive months of upper tercile mean temperatures in the contiguous United States that occurred between June 2011 and October 2012 was an unusual event in the context of the observational record that started in 1895. Indeed, it was the longest streak in any tercile anywhere in the record by a substantial margin. However, it is certainly far less rare than would be implied by naively assuming that each monthly temperature is independent of all preceding and following temperatures. There are two factors, which are both present in the record and need to be taken into account, in assessing this streak: temporal dependence and an underlying trend. Several approaches to describe these factors in a statistical manner were documented and explored. Our resulting calculations imply that in the absence of a trend, the probability of a 16 month or greater streak is likely in the range of 0.02 to 0.04, which is small but not so small as to make the event completely implausible. When a linear trend of the magnitude observed in the historical record is included, this probability increases to the range of 0.06 to 0.15. Even larger probabilities were obtained when nonlinear trend models were considered. However, the estimated streak probabilities were roughly 10 times smaller in cases where the time series model was first fitted with trend and then the trend set to 0 for the simulations. In most cases, estimated streak probabilities for FD models were larger than those for ARMA models, which is expected given the larger high-lag correlations of the FD models; however, the results were still qualitatively similar under the two sets of models.

Overall, the paper shows that in the absence of trend, the probability of a streak of 16 consecutive top-tercile events is low enough that one could legitimately query its plausibility. When either a linear or nonlinear trend is included, the probability increases to the point where such a result is not out of the ordinary.

Acknowledgments
Craigmile is supported in part by the U.S. National Science Foundation under grant SES-1024709. The work of Guttorp had partial support from the U.S. National Science Foundation under grant DMS-1106862, the STATMOS research network on Statistical Methods in Oceanic and Atmospheric Sciences. Smith was partially supported by the National Science Foundation under grant DMS-1127914 to the Statistical and Applied Mathematical Sciences Institute and by the Office of Science, Biological and Environmental Research, U.S. Department of Energy, Research grant DE-SC0004956. Jessica Matthews of CICS-NC provided help in bringing together the analysis group and scoping the remit of the subsequent analysis.

References

Akaike, H. (1974), A new look at the statistical model identification, *IEEE Trans. Autom. Control*, *19*, 716–723.

Beran, J. (1994), *Statistics for Long Memory Processes, Monographs on Statistics and Applied Probability*, vol. 61, Chapman and Hall, New York.

Brockwell, P. J., and R. A. Davis (2002), *Introduction to Time Series and Forecasting,* 2nd ed., Springer, New York.

Caballero, R., S. Jewson, and A. Brix (2002), Long memory in surface air temperature: Detection, modeling, and application to weather derivative valuation, *Clim. Res.*, *21*, 127–140.

Craigmile, P. F., and P. Guttorp (2011), Space-time modeling of trends in temperature series, *J. Time Ser. Anal.*, *32*, 378–395.

Fall, S., A. Watts, J. Nielsen-Gammon, E. Jones, D. Niyogi, J. R. Christy, and R. A. Pielke (2011), Analysis of the impacts of station exposure on the U.S. Historical Climatology Network temperatures and temperature trends, *J. Geophys. Res.*, *116*, D14120, doi:10.1029/2010JD015146.

Feller, W. (1968), *An Introduction to Probability Theory and Its Applications,* 3rd ed., vol. I, Wiley, New York.

Granger, C. W. J., and R. Joyeux (1980), An introduction to long-memory time series models and fractional differencing, *J. Time Ser. Anal.*, *1*, 15–29.

Guttorp, P., and T. Y. Kim (2013), Uncertainty in ranking the hottest years of US surface temperatures, *J. Clim.*, *26*, 6323–6328, doi:10.1175/JCLI-D-12-00760.1.

Hosking, J. R. M. (1981), Fractional differencing, *Biometrika*, *68*, 165–176.

Hurvich, C. M., and C.-L. Tsai (1988), Regression and time series model selection in small samples, *Biometrika*, *76*, 297–307.

Király, A., and I. M. Jánosi (2005), Detrended fluctuation analysis of daily temperature records: Geographic dependence over Australia, *Meteorol. Atmos. Phys.*, *88*, 119–128.

Lund, R., H. Hurd, P. Bloomfield, and R. Smith (1995), Climatological time series with periodic correlation, *J. Clim.*, *11*, 2787–2809.

Menne, M. J., C. N. Williams Jr., and R. S. Vose (2009), The United States Historical Climatology Network Monthly Temperature Data–Version 2, *Bull. Am. Meteorol. Soc.*, *90*, 993–1007, doi:10.1175/2008BAMS2613.1.

Menne, M. J., I. Durre, R. S. Vose, B. E. Gleason, and T. G. Houston (2012), An overview of the Global Historical Climatology Network-Daily Database, *J. Atmos. Oceanic Technol.*, *29*, 897–910, doi:10.1175/JTECH-D-11-00103.1.

©2014. American Geophysical Union. All Rights Reserved.

Muller, R. A., J. Wurtele, R. Rhode, R. Jacobsen, S. Perlmutter, A. Rosenfeld, J. Curry, D. Groom, and C. Wickham (2013), Earth atmosphere land surface temperature and station quality in the contiguous United States, *Geoinfor. Geostat.: An Overview*, *1*, 1–6.

NOAA (2012), U. S. climate highlights 12-month period (July 2011-June 2012). [Available at http://www.ncdc.noaa.gov/sotc/summary-info/national/2012/6, Accessed Sep. 26, 2012.]

Palma, W. (2007), *Long-Memory Time Series: Theory and Methods*, Wiley-Interscience, Hoboken, New Jersey.

Peterson, T. C., K. M. Willett, and P. W. Thorne (2011), Observed changes in surface atmospheric energy over land, *Geophys. Res. Lett.*, *38*, L16707, doi:10.1029/2011GL048442.

R Core Team (2012), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Santer, B. D., et al. (2008), Consistency of modelled and observed temperature trends in the tropical troposphere, *Int. J. Climatol.*, *28*, 1703–1722.

Shen, S. S. P., S. K. Lee, and J. Lawrimore (2012), Uncertainties, trends, and hottest and coldest years of U.S. surface air temperature since 1895: An update based on the USHCN V2 TOB data, *J. Clim.*, *25*, 4185–4203.

Tukey, J. W. (1977), *Exploratory Data Analysis*, Addison-Wesley, Reading, Mass.

Venema, V., et al. (2012), Benchmarking homogenization algorithms for monthly data, *Clim. Past*, *8*, 89–115, doi:10.5194/cp-8-89-2012.

Williams, C. N., M. J. Menne, and J. H. Lawrimore (2012a), Modifications to Pairwise Homogeneity Adjustment software to improve run-time efficiency, *Tech. Rep. GHCNM-12-01R*, NOAA-NCDC.

Williams, C. N., M. J. Menne, and J. H. Lawrimore (2012b), Modifications to Pairwise Homogeneity Adjustment software to address coding errors and improve run-time efficiency, *Tech. Rep. GHCNM-12-02*, NOAA-NCDC.

Williams, C. N., M. J. Menne, and P. W. Thorne (2012c), Benchmarking the performance of pairwise homogenization of surface temperatures in the United States, *J. Geophys. Res.*, *117*, D05116, doi:10.1029/2011JD016761.

 ©2014. American Geophysical Union. All Rights Reserved.