Statistics and Its Interface Volume 8 (2015) 19–31

Detecting change-points in extremes*

D. J. Dupuis†, Y. Sun, and Huixia Judy Wang

Even though most work on change-point estimation focuses on changes in the mean, changes in the variance or in the tail distribution can lead to more extreme events. In this paper, we develop a new method of detecting and estimating the change-points in the tail of multiple time series data. In addition, we adapt existing tail change-point detection methods to our specific problem and conduct a thorough comparison of different methods in terms of performance on the estimation of change-points and computational time. We also examine three locations on the U.S. northeast coast and demonstrate that the methods are useful for identifying changes in seasonally extreme warm temperatures.

AMS 2000 subject classifications: Primary 62G32; secondary 62P12.
Keywords and phrases: Tail behavior, Quantile methods.

1. INTRODUCTION

We consider time series and are interested in estimating the number and the location of change-points in such series. The time series is described by an underlying distribution or model, and a change-point is a point in time at which the parameters of the distribution or the model abruptly change. Finding change-points can also equivalently be seen as the subdivision of a series into segments characterized by homogeneous statistical features (e.g., mean and higher-order moments) or identical underlying distributions.

Establishing the existence, and ultimately the number and locations, of such change-points in climatic time series, for example, has received much attention over the last 40 years as researchers seek irrefutable evidence of climatic change and its link to anthropogenic activities. Besides climate-related changes, other nonclimatic factors such as relocation of weather stations and changes of instrumentation are apt to cause sudden changes and these must be identified to properly analyze climatic time series; see e.g. [17] and references therein. The detection of changes in variables other than temperature, on regional scales and in climate extremes, is also important for evaluating model simulations; see [7] for a list of these other variables and a discussion on their relevance.

Recent work on detecting mean shift in climatic series includes [1, 18, 19, 28]. [28] also provided a good literature review on the matter, while [25] provided a review and comparison of less recent techniques.

While most of the literature focuses on changes in the mean that lead to changes in extreme events, changes in the variance or in the tails can lead to more extreme events. For example, if we consider temperature distributions, Figure 1, similar to those in climate reports like [11], shows how changes in the probability of extreme events can also be the result of a change in variability (center plot) or a change in the tail (right plot), while keeping a fixed mean. Changes in tail behavior are the focus of this paper.

It was shown as early as in [15] that prediction of such extremities as hot spells, droughts and deep freezes lies largely in identifying changes in the variability (scale and/or shape) of the climate conditions and to a lesser extent on scenarios of changes in the average. However, identification of changes in scale and/or shape has proven to be a difficult task for environmental data sets. [13] outlined some of the problems with applying change-point detection methods to environmental data.

We need to distinguish between two different ideas: (1) identifying change-points in the mean of extreme data, e.g. in the location parameter of the generalized extreme value (GEV) distribution fitted to annual maxima (or minima), and (2) identifying change-points in the shape and/or scale of the GEV or of the tail distribution of extreme data. The former was first considered by [14], while the latter was first considered by [12], and more recently by [2, 4, 5]. [14] proposed test statistics for detecting a change in a location parameter of annual maxima and annual minima series. [12] tried to identify changes in the scale and shape parameters of the Weibull distribution used to model monthly minimum temperatures. [2] presented an empirical analysis focusing on the counts in extreme ranges of temperature, pressure and precipitation variables. [4] developed a likelihood approach for testing for changes in the scale or shape parameter of the generalized Pareto (GP) distribution used to model the upper tail of temperature and precipitation variables. [5] presented a change-point analysis to detect the time at which the crossing rate of a high threshold changes.

*This material was based upon work partially supported by the National Science Foundation (NSF) under Grant DMS-0635449 to the Statistical and Applied Mathematical Sciences Institute. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. We would also like to thank the support from Natural Sciences and Engineering Research Council of Canada for Dupuis, and NSF Award DMS-1007420 and NSF CAREER Award DMS-1149355 for Wang.
†Corresponding author.

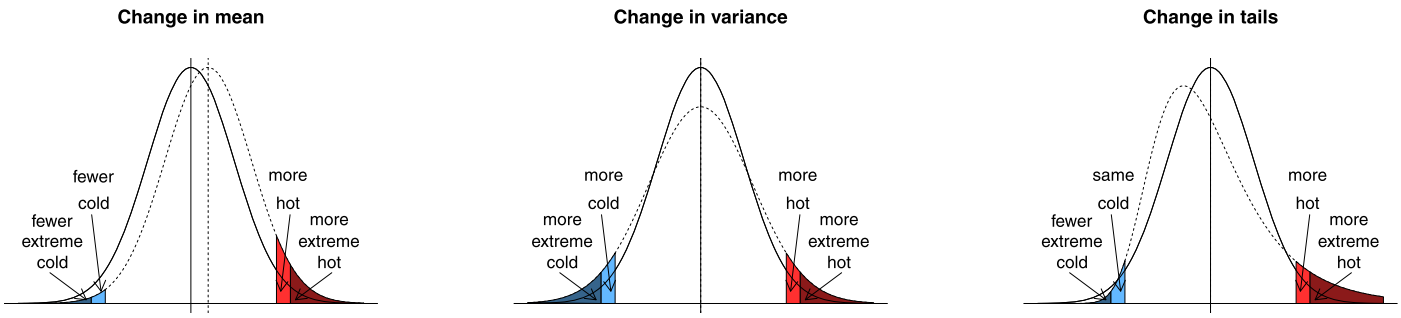

Figure 1. Change in the probability of exceeding an extreme value as the mean (left), the variance (center), and tail (right), changes from the solid density to the dashed density.

Beyond climate studies, there is also interest in economics and finance for detecting changes in the tail behavior of time series. [24] and [16] developed methods for detecting change-points in the tail index parameter of a distribution. Both methods are based on Hill's tail index estimator and thus are designed for heavy-tailed distributions.

In this paper, we focus on the detection of change-points in the upper-tail of the distribution of the variable of interest based on multiple cross sectional time series. For instance, consider daily time series data from multiple years. Year after year, the upper-tail of the distribution would change at some point(s), for example due to a *hurricane season* or a change in prevailing winds in the case of weather variables or an *earnings season* in the case of stock market returns. Even though there exists some work on tail change-point detection, none of the available methodology can be applied directly to multiple cross sectional time series, e.g., multiple-year daily data where multi-year seasonal change-points are sought.

Our contributions are three-fold. Firstly, we present a new multinomial-based method of estimating the change-points in the tail. Secondly, we adapt existing methods to our specific problem of identifying change-points using multiple cross sectional time series and compare the latter approaches to our new method, both in terms of performance and computational time. Computational time is an important issue if the change-point methods are to be applied to the massive simulations from climate models, for example. Thirdly, we discuss the connections and differences among all available methods.

The rest of the paper is organized as follows. In Section 2, we present the methods for identification of change-points in the tail of multiple time series data. In Section 3, we present the results of simulation studies that target the performance of these methods. We present an application to estimating seasonal change-points in noise about means for daily maximum temperatures in Section 4.

2. IDENTIFICATION OF CHANGE-POINTS

We adopt a change-point model and aim to identify change-points from multiple cross sectional time series. Suppose the data contain observations $\{Z_{it}, i = 1, \ldots, n, t = 1, \ldots, T\}$, where i is the index noting the individual (e.g. year) and t is the index for time (e.g. day). Let m be the unknown number of change-points resulting in $m + 1$ segments, and $T_1, \ldots, T_m$ be the corresponding unknown change-points. For every i, assume that $Z_{it} \sim f_j$ for $t \in [T_{j-1}, T_j)$, $j = 1, \ldots, m + 1$, where $T_0 = 1$, $T_{m+1} = T + 1$, and f_j are distributions such that f_j differ from both f_{j-1} and f_{j+1} in the right tail.

We propose several methods for identifying change-points across multiple time series, for instance, change-points that mark the change of seasons during the year based on daily data. The likelihood ratio, single and multiple quantiles, and Kim and Lee approaches are adapted from the existing work for single time series [4, 21, 16]. The multinomial approach in Section 2.2 is a new development. We provide a brief description of each method in Sections 2.1-2.5, and discuss the connections and distinctions of different methods in Section 2.6.

To avoid estimated segments that are too short, for all the compared methods, we search for change-points over the following set of possible partitions: $\Lambda_\epsilon = \{(T_1, \ldots, T_m) : T_j - T_{j-1} \geq \epsilon T, T_1 \geq \epsilon T, T_m \leq (1 - \epsilon)T\}$, where $\epsilon > 0$ is a small constant. In our numerical studies, we focus on identifying seasonal change-points based on daily data from multiple years with $T = 365$, so we use $\epsilon = 0.082$, which ensures that the identified seasons are at least 30 days in length.

2.1 Likelihood ratio test

Consider a random process $\{X_t\}$. Suppose that the process is stationary and has marginal distribution F with upper end point x^F. [22] showed that if the distribution of excesses $X_t - u$ of a high threshold u, $u < x^F$, scaled as a function of u, converges to a non-degenerate limiting distribution as $u \to x^F$, that distribution must be the GP distribution. This derivation suggests that the GP distribution will be a practical family for statistical estimation when examining excesses over the threshold u, provided that u is taken sufficiently high.

We adapt the likelihood ratio test proposed in [4] for single time series to multiple time series. We consider the following tail model for Z_{it},

$$\Pr(Z_{it} > z + u | Z_{it} > u) = \left(1 + \frac{\xi_t z}{\beta_t}\right)_+^{-1/\xi_t}, \tag{1}$$

where $\beta_t > 0$ and ξ_t are the unknown scale and shape parameters, respectively, and $a_+ = \max(0, a)$. We define a time t^* to be a change-point if there is a change in either β_t or ξ_t before and after t^*.

[4] built a test statistic to test whether β_t and/or ξ_t change over time. Their construction is based on likelihood considerations and essentially implements the ideas of [3] to the extreme value context.

To test the existence of change-points, we consider the hypotheses $H_0 : (\beta_1, \xi_1) = \ldots = (\beta_T, \xi_T)$ versus $H_a : (\beta_1, \xi_1) = \ldots = (\beta_{t^*}, \xi_{t^*}) \neq (\beta_{t^*+1}, \xi_{t^*+1}) = \ldots = (\beta_T, \xi_T)$ for some t^*. The likelihood ratio test statistic is defined as

$$W_{nT} = \sqrt{\max_{t^* \in [\epsilon T, (1-\epsilon)T]} (-2 \log \Delta_{t^*})}, \tag{2}$$

where $\log \Delta_{t^*} = L(\widehat{\beta}, \widehat{\xi}) - L_1(\widehat{\beta}_{(1)}, \widehat{\xi}_{(1)}) - L_2(\widehat{\beta}_{(2)}, \widehat{\xi}_{(2)})$, $L(\widehat{\beta}, \widehat{\xi})$ is the log-likelihood function for (1) based on the positive excesses $\epsilon_{it} = Z_{it} - u$ for all $i = 1, \ldots, n$ and $t = 1, \ldots, T$, and $L_1(\widehat{\beta}_{(1)}, \widehat{\xi}_{(1)})$ and $L_2(\widehat{\beta}_{(2)}, \widehat{\xi}_{(2)})$ are the log-likelihood functions based on the samples up to and including the point at time t^* and the samples after time t^*, respectively. Similar to [4], we calculate the critical value by the $(1-\alpha)$th quantile of $\sqrt{\sup_{\epsilon \leq s < 1-\varepsilon} \{s(1-s)\}^{-1} B_2^2(s)}$, where $B_2(s)$ is the sum of two Brownian bridges and α is the significance level.

The method in [4] was designed for detecting a single change-point. We extend the method to identify multiple change-points using the following procedure. We first perform the above hypothesis testing using the entire sample. If W_{nT} does not exceed the critical value, we conclude there is no change-point and terminate the procedure. On the other hand, if the null hypothesis is rejected, we estimate the first change-point by $\widehat{T}_1 = \arg\max_{t^* \in [\epsilon T, (1-\epsilon)T]} (-2 \log \Delta_{t^*})$, and obtain two segments. Then we repeat the above testing procedure for each segment data separately, using a nominal level $\alpha/2$ for each test, and estimate the additional change-points if the hypothesis of no change-points in the segment is rejected. The process is continued in an analogous fashion until the test fails to reject the null in every identified segment or the shortest segment defined by adding another new change-point is shorter than ϵT.

2.2 Bernoulli and multinomial approaches

The method in Section 2.1 is based on approximating the distribution of the exceedances over a high threshold by the GP distribution. [6] developed an exact approach for finding MLE of the change-points $T_j, j = 1, \ldots, m$, and the within-segment parameters, when the functional form of the hypothesized within-segment distribution $f_i, i = 1, \ldots, m+1$ is within the general exponential family. Since the GP distribution is not within the general exponential family, the [6] approach cannot be applied directly.

[5] considered reducing the tail data to Bernoulli variates $B_{it} = I(Z_{it} > u)$: that is, the exceedance (success) or non exceedance (failure) of the observation over a high threshold u. Then the variables B_{it} with $T_{j-1} \leq t < T_j$ follow a Bernoulli model with parameter ζ_j. ML and the dynamic programming (DP) approach in [6] are used to identify $T_1, \ldots, T_m$.

As reducing the tail data to Bernoulli variates may be too crude an approximation, we propose to extend the idea by examining changes in the probability of exceeding multiple high thresholds. In our implementation, we consider three high thresholds $u_1 = q_{0.95}$, $u_2 = q_{0.98}$ and $u_3 = q_{0.99}$. For any given observation Z, there are four mutually exclusive outcomes

C_1: Z is below all three thresholds;
C_2: Z exceeds u_1 only;
C_3: Z exceeds u_1 and u_2 only;
C_4: Z exceeds all three thresholds,

and we consider the categorical variate X with sample space $\{C_1, \ldots, C_4\}$. Letting $\boldsymbol{X} = (X_1, \ldots, X_4)^T$ where X_k denotes the number of times that outcome C_k is observed, $\boldsymbol{X}$ has a multinomial distribution with parameters $n = 1$ and $\boldsymbol{p} = (p_1, \ldots, p_4)^T$ with $p_k \geq 0$ and $\sum_{k=1}^4 p_k = 1$.

Instead of choosing change-points based on MLE for a binomial likelihood as in [5], we use MLE for a multinomial likelihood. If the $\boldsymbol{X}_{h+1}, \ldots, \boldsymbol{X}_m$ are an independent sequence of multinomial variates with parameter $n = 1$ and $\boldsymbol{p} = (p_1, p_2, p_3, p_4)'$, the MLE of p_k is $\widehat{p}_k = (m-h)^{-1} \sum_{t=h+1}^m X_{tk}$. The maximized log-likelihood obtained by substituting $\widehat{p}_k$ for p_k, $k = 1, \ldots, 4$, can be easily calculated as

$$\begin{aligned} L(\hat{\boldsymbol{p}}; h, m) \quad \propto \quad & (m-h)\big\{\widehat{p}_2 \log(\widehat{p}_2) + \widehat{p}_3 \log(\widehat{p}_3) + \widehat{p}_4 \log(\widehat{p}_4) \\ & + (1 - \widehat{p}_2 - \widehat{p}_3 - \widehat{p}_4) \log(1 - \widehat{p}_2 - \widehat{p}_3 - \widehat{p}_4)\big\}. \end{aligned}$$

We reduce the tail data to multinomial variates $\boldsymbol{X}_{it}$. Observed $\boldsymbol{X}_{it}$ with $T_{j-1} \leq t < T_j$ follow a multinomial model with parameter $\boldsymbol{p}_j$. ML and the DP approach in [6] can then be used to identify $T_1, \ldots, T_m$.

[6] discussed testing for the number of change-points. The DP algorithm does not find the optimal number of change-points, but rather the optimal m change-points for a fixed m. We do not fix m *a priori* and initially allow for an arbitrary number of change-points, but we stop adding change-points when one of the segments that they define is shorter than ϵT.

2.3 Single quantile approach

Rather than focusing on the probability that Z_{it} exceeds a large value, we can examine the upper quantile of Z_{it} itself.

The single quantile method is a modification of the change-point estimation method proposed in [21]. The method focuses on identifying changes in the τth quantile of Z_{it}, where $\tau \in (0,1)$ is a prespecified high quantile level. [21] estimated change-points by fitting alternative models, which requires estimating the quantiles for $T(T+1)/2$ partitioned samples. To reduce the computational cost, we modify the procedure by estimating the change-points with the maximizers of the score-type statistic proposed in [23]. Suppose M is the prespecified maximum number of change-points. Our modified procedure requires estimating quantiles only $M(M+1)/2$ times, and thus greatly improves the computational efficiency.

For $j \geq 1$, let $G_{\alpha,j}$ be the $(1-\alpha)$th quantile of $G(x)^j$, where $G(\cdot)$ is the distribution function of $\sup_{\epsilon \leq s \leq 1-\epsilon} |B(s)|$, and B is a Brownian bridge. The detailed procedure is as follows.

Step 1. Test H_0: there is no change-point versus H_a: there is one change-point. Define $SQ_{\tau,1}$ as

(3)
$$\max_{\epsilon T \leq t^* \leq (1-\epsilon)T} \left| \{\tau(1-\tau)\}^{-1/2} \{S_{t^*}(\hat{q}_\tau) - t^*/T S_T(\hat{q}_\tau)\} \right|$$

where $S_{t^*}(\hat{q}_\tau) = (nT)^{-1/2} \sum_{i=1}^{n} \sum_{t=1}^{t^*} \{\tau - I(Z_{it} \leq \hat{q}_\tau)\}$ and $\hat{q}_\tau$ is the τth sample quantile of $\{Z_{it}, i = 1, \ldots, n, t = 1, \ldots, T\}$. If $SQ_{\tau,1} < G_{\alpha,1}$, we declare there is no change-point and terminate the procedure. Otherwise we reject H_0 and estimate the first change-point by $\hat{T}_1 = \arg\max_{\epsilon T \leq t^* \leq (1-\epsilon)T} |S_{t^*}(\hat{q}_\tau) - t^*/T S_T(\hat{q}_\tau)|$ and proceed.

Step 2. Start from $j = 1$. Suppose that in the previous stage, $j+1$ segments have been formed by change-points $\hat{T}_1, \ldots, \hat{T}_j$. We now test H_0: there are j change-points versus H_a: there are $j+1$ change-points. For each $k = 1, \ldots, j+1$, calculate the within-segment test statistic $SQ_{\tau,k}$ and its corresponding maximizer t_k as in Step 1 by using data from the kth segment. Let $SQ_\tau = \max_{1 \leq k \leq j+1} SQ_{\tau,k}$. If $SQ_\tau < G_{\alpha,j+1}$, we stop the procedure. Otherwise, we reject H_0 and estimate the new change-point as follows. For each $k = 1, \ldots, j+1$, the candidate change-point t_k and the existing change-points $\{\hat{T}_1, \ldots, \hat{T}_j\}$ form $j+2$ segments. Let $\hat{q}_{\tau,k}, k = 1, \ldots, j+2$ be the corresponding segment-wise τth sample quantile. The new change-point $\hat{T}_{j+1}$ is taken to be the t_k that gives the smallest quantile loss, defined as $\sum_{i=1}^{n} \sum_{t=1}^{T} \rho_\tau(Z_{it} - \hat{q}_{\tau,t})$, where $\rho_\tau(u) = \{\tau - I(u<0)\}u$ is the quantile loss function and $\hat{q}_{\tau,t} = \hat{q}_{\tau,k}$ if t falls into the kth segment. Finally increase the value of j.

Step 3. Repeat Step 2 until the test fails to reject the null or when j reaches M, the prespecified maximum number of change-points.

In our analysis, we consider three quantile levels $\tau = 0.95, 0.98$ and 0.99. Note that the single quantile approach implicitly sets the critical values to accommodate sequential testing and estimation so no further adjustments are required when trying to identify multiple change-points.

2.4 Multiple quantile approach

Suppose that quantiles in the interval $\mathcal{T} = [\tau_L, \tau_R]$ with $0 < \tau_L < \tau_R < 1$ are affected by structural changes at the same change-points but possibly change with different magnitudes. The multiple quantile method aims to detect the common change-points across quantiles. The procedure operates in the same way as the single quantile method except that the test statistic SQ is replaced by $MQ = \sup_{\tau \in \mathcal{T}} SQ_\tau$ and the critical value $G_{\alpha,j}$ is replaced by $\bar{G}_{\alpha,j}$, the $(1-\alpha)$th quantile of $\bar{Q}(x)^j$, where $\bar{Q}(\cdot)$ is the distribution of $\sup_{\tau \in \mathcal{T}} \sup_{\epsilon \leq s \leq 1-\epsilon} |B(s,\tau)|$, and $B(s,\tau)$ is the Brownian Pillow process; see [23]. In our analysis, we consider the upper quantile region $\mathcal{T} = [0.95, 0.99]$. As with the single quantile approach, sequential testing is implicit.

2.5 Kim and Lee approach

[16] proposed a cusum-based procedure for testing tail index changes in time series with Pareto-type tails. The tail index is a measurement of the fatness of the tail distribution. Here we adapt the idea in [16] to identify change points using multiple time series.

Consider testing the hypotheses H_0: there is no change in the tail index versus H_a: there exists at least one change-point. Let τ be a high quantile level close to one. Define the test statistic

$$T_n(\phi) = \frac{1}{\sqrt{nT\tau}} \max_{\epsilon T \leq t^* \leq (1-\epsilon)T} \left| M(t^*, \hat{q}_\tau) - \frac{t^*}{T} M(T, \hat{q}_\tau) \right|,$$

where $M(t^*, \hat{q}_\tau) = \sum_{i=1}^{n} \sum_{t=1}^{t^*} \phi\{\log(\frac{Z_{it}}{\hat{q}_\tau})\}$ and $\hat{q}_\tau$ is the τth sample quantile of $\{Z_{it}\}$. Two choices are considered for $\phi(\cdot)$, $\phi_1(x) = I(x > 0)$ and $\phi_2(x) = x_+$. For the score function $\phi_2(x) = x_+$, $M(t, \hat{q}_\tau)/(nT\tau)$ is the well known Hill's estimator ([8]) of the tail index for heavy-tailed distributions. The score function $\phi_2(x)$ incorporates exceedance values rather than the simple count of the number of exceedances measured by $\phi_1(x)$.

Similar to the likelihood-ratio method in Section 2.1, we identify the change-points sequentially using the following procedure. We first carry out the above testing, and determine there is no change-point and terminate the procedure if $T_n(\phi_1) < G_{\alpha,1}$ or $T_n(\phi_2) < \sqrt{2}\xi G_{\alpha,1}$, where $1/\xi$ is the tail index. Otherwise we reject H_0 and estimate the first change-point by $\hat{T}_1 = \arg\max_{\epsilon T \leq t^* \leq (1-\epsilon)T} |M(t^*, \hat{q}_\tau) - t^*/T M(T, \hat{q}_\tau)|$. We then repeat the hypothesis testing procedure for each identified segment separately, adjusting the nominal level as for the LRT in Section 2.1, to determine the other change-points until the test fails to reject the null hypothesis in all segments or until the shortest segment defined by adding another change-point is shorter than ϵT.

2.6 Connections and distinctions of different approaches

The test statistics of the Bernoulli, multinomial, single and multiple quantile approaches and the Kim and Lee approach with $\phi_1(\cdot)$ are all based on the counts of the number of exceedances. In contrast, the LRT and the Kim and Lee with $\phi_2(\cdot)$ incorporate the values of exceedances, and thus are more informative and are intended to characterize the tail distribution in a more comprehensive way. However, the paucity of data in the tail can make the estimation of the tail distribution more difficult, especially since the LRT approach relies on the approximating GP distribution of the exceedances.

All the approaches require the choice of a threshold (in one form or another) and the usual arguments of bias (too low a threshold) and variance (too high a threshold) prevail. To keep comparisons fair, in our numerical studies, we choose a reasonably high threshold for each method, and we apply the single quantile method at three different high quantiles to show the sensitivity.

2.6.1 Bernoulli and single quantile

Both Bernoulli and single quantile methods focus on the exceedance probability over a high threshold. The Bernoulli method reduces the tail data to Bernoulli variates $B_{it} = I(Z_{it} > u)$, indicating whether the observation Z_{it} on day t of the ith year exceeds the threshold u or not. Then the dynamic programming approach in [6] is applied to find the change-points that maximize the Bernoulli likelihood. In particular, the observed log-likelihood that needs to be maximized is of the form

$$\begin{aligned} &nt^*\{\hat{p}_a \log(\hat{p}_a) + (1-\hat{p}_a)\log(1-\hat{p}_a)\} \\ &+ n(T-t^*)\{\hat{p}_b \log(\hat{p}_b) + (1-\hat{p}_b)\log(1-\hat{p}_b)\} \end{aligned}$$

where $\hat{p}_a$ and $\hat{p}_b$ are the averages of B_{it} within the two segments formed by $[1, t^*]$ and $(t^*, T]$, respectively.

In the single quantile method, the quantity $I(Z_{it} \le u)$ in the test statistic SQ_τ can be written as Bernoulli variates as well since $I(Z_{it} \le u) = 1 - B_{it}$. Then it is not difficult to show that maximizing the single quantile test statistic in (3) is equivalent to maximizing $\{t^*(T-t^*)\}|\hat{p}_a - \hat{p}_b|$ over t^*.

Both methods are therefore based on the MLE of the exceedance probability. The Bernoulli method estimates change-points by finding the MLE of the exceedance probability and maximizing a Bernoulli likelihood, while the single quantile approach estimates change-points by detecting the maximum fluctuation of the MLE between two segments.

2.6.2 Single quantile and Kim and Lee

The Kim and Lee method is closely related to the single quantile method. In fact, when the indicator function $\phi_1(x) = I(x > 0)$ is used, the test statistic of Kim and Lee is identical to that in the single quantile method. The procedure of Kim and Lee for detecting the existence of change-point is the same as Step 1 of the single quantile procedure. When $\phi_2(x) = x_+$ is used in Kim and Lee, the test statistic incorporates the relative excess and the method tends to be more informative but less robust compared to that based on $\phi_1(x)$.

2.6.3 Likelihood ratio test and Kim and Lee

The Kim and Lee method based on $\phi_2(\cdot)$ is also related to the likelihood ratio approach [4] for the GP distributions in that both utilize Hill's estimator to characterize the distribution's tail behavior. Instead of only considering the exceedance probability as in the Bernoulli and single quantile methods, both methods incorporate exceedance values through Hill's estimator. The connection between them is similar to that between the Bernoulli and the single quantile methods. The likelihood ratio method estimates change-points through the LRT where the shape parameter is estimated by Hill's estimator, while the Kim and Lee approach estimates change-points by detecting the maximum fluctuation of Hill's estimator.

3. SIMULATION STUDY

3.1 One change-point

To assess the implications of using a parametric or nonparametric approach, we compare the methods presented in Section 2 not only when the tail GP model is true, but also for cases with other possible (comparable) tail behaviors. We consider three different types of data: GP-, Beta- and t-distributed. We generate 50 years $\times$ 365 days of independent data and set one change-point at $T_1 = 182$ with distributions f_0 and f_1 as indicated in Table 1. The change in the tail, before and after the change-point, in each case is comparable as measured by the tail index $\tau(F)$ of [9] defined by

$$\tau(F) = \frac{F^{-1}(0.99) - F^{-1}(0.5)}{F^{-1}(0.75) - F^{-1}(0.5)} \Big/ \frac{\Phi^{-1}(0.99) - \Phi^{-1}(0.5)}{\Phi^{-1}(0.75) - \Phi^{-1}(0.5)},$$

where Φ is the standard normal distribution. The values of $\tau(F)$ are included in Table 1. The true parameter values were chosen based on the fact that, for the GP, going from $\xi = -0.05$ to $\xi = 0.05$ represented a reasonable challenge for the methods.

As noted in Section 2.1, Pickands' result says that if there is a non-degenerate asymptotic distribution for the excesses over a high threshold, then that distribution must be the GP. So, considering a distribution other than the GP reflects the idea that, in practice, we might not have reached the asymptotic behavior and the data are more related to the parent distribution. In this respect, we have chosen distributions that are likely. The Beta is in the domain of attraction of the GP($\xi < 0$). It is a good choice because it has finite upper support (which one would like to think is the case for temperature data for example). Note that $\beta = 1$ for both segments and thus the distribution of exceedances over an asymptotically high threshold is a GP with $\xi = -1$.

Table 1. Simulation details: unprocessed data, one change-point

	f_0	$\tau(F)$	f_1	$\tau(F)$
GP	GP($\xi = -0.05$, $\beta = 1$)	1.51	GP($\xi = 0.05$, $\beta = 1$)	1.78
Beta	Beta($\alpha = 0.25$,$\beta = 1$)	1.03	Beta($\alpha = 0.2$,$\beta = 1$)	1.29
t	Normal($\mu = 0, \sigma = 1$)	1.00	t($\nu = 6$)	1.26

Table 2. Comparison of TP, FP, FWER and computation time for GP-, Beta- and t- distributed tail data. Simulated data have one true change-point at $T_1 = 182$. Results for each distribution-method combination is based on 500 simulations. Fifty years of daily data are generated in each of the simulations. Methods are described in the following sections: LRT - §2.1, Be - §2.2, Mu - §2.2, SQ_τ - §2.3, MQ - §2.4, KL (using ϕ_2) - §2.5

Distribution	Method	TP	TPR (%)	FP1	FP2	FP	FWER (%)	Time (s)
GP	LRT	257	51.4	238	3	241	48.2	10.6
	Be	307	61.4	213	81	294	58.8	11.8
	Mu	353	70.6	196	58	254	50.8	17.9
	$\mathrm{SQ}_{0.95}$	300	60.0	206	5	211	42.2	0.01
	$\mathrm{SQ}_{0.98}$	312	62.4	199	4	203	40.6	0.01
	$\mathrm{SQ}_{0.99}$	304	60.8	204	8	212	42.4	0.01
	MQ	313	62.6	194	7	201	40.2	0.18
	KL	371	74.2	133	4	137	27.4	0.02
Beta	LRT	0	0	30	1	31	6.2	18.0
	Be	123	24.6	327	61	388	77.6	10.6
	Mu	112	22.4	290	57	347	69.4	15.7
	$\mathrm{SQ}_{0.95}$	128	25.6	259	9	304	60.8	0.01
	$\mathrm{SQ}_{0.98}$	52	10.4	185	3	188	37.6	0.01
	$\mathrm{SQ}_{0.99}$	25	5.0	103	1	104	20.8	0.01
	MQ	136	27.2	278	7	285	57.0	0.17
	KL	78	15.6	182	1	183	36.6	0.02
t	LRT	405	81.0	96	3	99	19.8	10.4
	Be	329	65.8	221	73	294	58.8	11.9
	Mu	417	83.4	192	54	246	49.2	18.9
	$\mathrm{SQ}_{0.95}$	347	69.4	155	8	163	32.6	0.01
	$\mathrm{SQ}_{0.98}$	395	79.0	110	8	118	23.6	0.01
	$\mathrm{SQ}_{0.99}$	374	74.8	135	13	148	29.6	0.01
	MQ	361	72.2	145	1	146	29.2	0.20
	KL	413	82.6	89	2	91	18.2	0.02

The maximum likelihood estimators do not obey the regularity conditions when $\xi < -0.5$ ([27]). This design could thus be quite problematic for the LRT. The Normal is in the domain of attraction of the GP($\xi = 0$), while the $t(\nu = 6)$ is in the domain of attraction of the GP($\xi > 0$). They are good choices based on empirical evidence in many data sets. We carry out 500 simulations for each distribution-method combination, and Table 2 shows:

TP: Total number of correctly identified change-points: true positives. For each simulation, if there is any change-point detected, we choose the one closest to the true change-point, and record a true positive only if the detected change-point is within five days of the true value;

TPR: Proportion of cases where the true change-point is identified, TPR = TP/500;

FP1: Number of cases where one false change-point is detected;

FP2: Number of cases where $\geq$ two false change-points are detected;

FP: Total number of cases with false positives, FP = FP1 + FP2; and

FWER: Proportion of cases where at least one false positive is obtained, i.e. the familywise error rate, FWER = FP/500. Note that FWER + TPR could be greater than 100% as a method could have identified two change-points for example, the true one and an additional (unnecessary) one.

The computation time is also provided. All the computations are done by a single threaded application in R on a dual socket quad core 2.66 Ghz Intel Xeon Clovertown with 16 GB of RAM. For all methods, the tail data are those above the 95% quantile. For the case of GP-distributed data, the Kim and Lee approach outperforms all other methods, yielding both the largest TPR and the lowest FWER. The multinomial approach has the second largest TPR, but its

Table 3. Comparison of TP, FP, FWER for GP-, Beta- and t- distributed tail data. Simulated data have two true change-points at $T_1 = 122$ and $T_2 = 243$. Results for each distribution-method combination is based on 500 simulations. Fifty years of daily data are generated in each of the simulations. Methods are described in the following sections: LRT - §2.1, Be - §2.2, Mu - §2.2, SQ_τ - §2.3, MQ - §2.4, KL (using ϕ_2) - §2.5

Distribution	Method	TP1	TP2	TPR (%)	FP1	FP2	FP	FWER (%)
GP	LRT	180	88	17.6	187	71	258	51.6
	Be	239	175	35.0	216	139	355	71.0
	Mu	197	253	50.6	191	108	299	59.8
	$SQ_{0.95}$	219	221	44.2	218	54	272	54.4
	$SQ_{0.98}$	216	221	44.2	216	59	275	55.0
	$SQ_{0.99}$	236	191	38.2	236	68	304	60.8
	MQ	213	214	42.8	214	63	277	55.4
	KL	173	306	61.2	182	22	204	40.8
Beta	LRT	1	0	0	26	4	30	6.0
	Be	120	18	3.6	197	139	336	67.2
	Mu	77	8	1.6	217	90	307	61.4
	$SQ_{0.95}$	77	10	2.0	81	44	125	25.0
	$SQ_{0.98}$	18	5	1.0	45	7	52	10.4
	$SQ_{0.99}$	11	0	0	36	3	39	7.8
	MQ	62	11	2.2	80	35	115	23.0
	KL	15	2	0.4	16	3	19	3.8
t	LRT	162	309	61.8	165	23	188	37.6
	Be	226	206	41.2	220	128	348	69.6
	Mu	133	351	70.2	202	61	263	52.6
	$SQ_{0.95}$	212	236	47.2	212	49	261	52.2
	$SQ_{0.98}$	152	322	64.4	152	26	178	35.6
	$SQ_{0.99}$	170	316	63.2	170	14	184	36.8
	MQ	196	260	52.0	196	44	240	48.0
	KL	129	364	72.8	128	10	138	27.6

FWER is also the second largest. The parametric approach, i.e. LRT, falls apart in the case of the Beta-distributed data. The Bernoulli, multinomial, single quantile with $\tau = 0.95$ and multiple quantile manage to maintain a TPR of 25%, but do so at the expense of large FWER. The Kim and Lee approach, as one might expect, has very little power in the Beta-distributed case. For the t-distributed data, the Kim and Lee approach is once again the top performer, but single quantile with $\tau = 0.95$ and LRT are close behind. The multinomial approach yields the largest TPR, but has a FWER about twice that of the previous three methods. Overall, the multinomial works better than Bernoulli in terms of TPR and FWER for both the GP and t distributions, while the two are similar for the Beta distribution. The quantile methods, SQ (at the three quantiles) and MQ, perform similarly.

3.2 Multiple change-points

In this simulation study, we examine the performance of each of the methods for multiple change-points identification. A similar simulation design is used as in Section 3.1. The three types of distributions, GP, Beta and t, are considered, but with two change-points at $T_1 = 122$ and $T_2 = 243$. We generate 50 years $\times$ 365 days of independent data with distributions f_0, f_1 and $f_2 = f_0$ as indicated in Table 1. Then, change-points are identified by each of the methods following the multiple change-points identification procedures described in Section 2 and summarized in Table 4. With 500 replications, the histograms of the number of identified change-points for each of the methods are shown in Figure 2 under each of the three distributions. Table 3 shows similar measurements as in Table 2 except the following:

TP1: Number of cases where one true change-point is detected;

TP2: Number of cases where two true change-points are detected;

TPR: Proportion of cases where two true change-points are identified, TPR = TP2/500.

Results are similar to those for the one change-point case, except that the performance under the Beta-distributed tail is further reduced. A summary of the properties of our different change-point estimation methods appears in Table 4.

Other simulation studies (not shown) reveal that it becomes increasingly difficult to detect all the change-points when they increase in number, however the proportion of cases where at least one false positive is obtained only increases in the case of t-distributed data. As the number

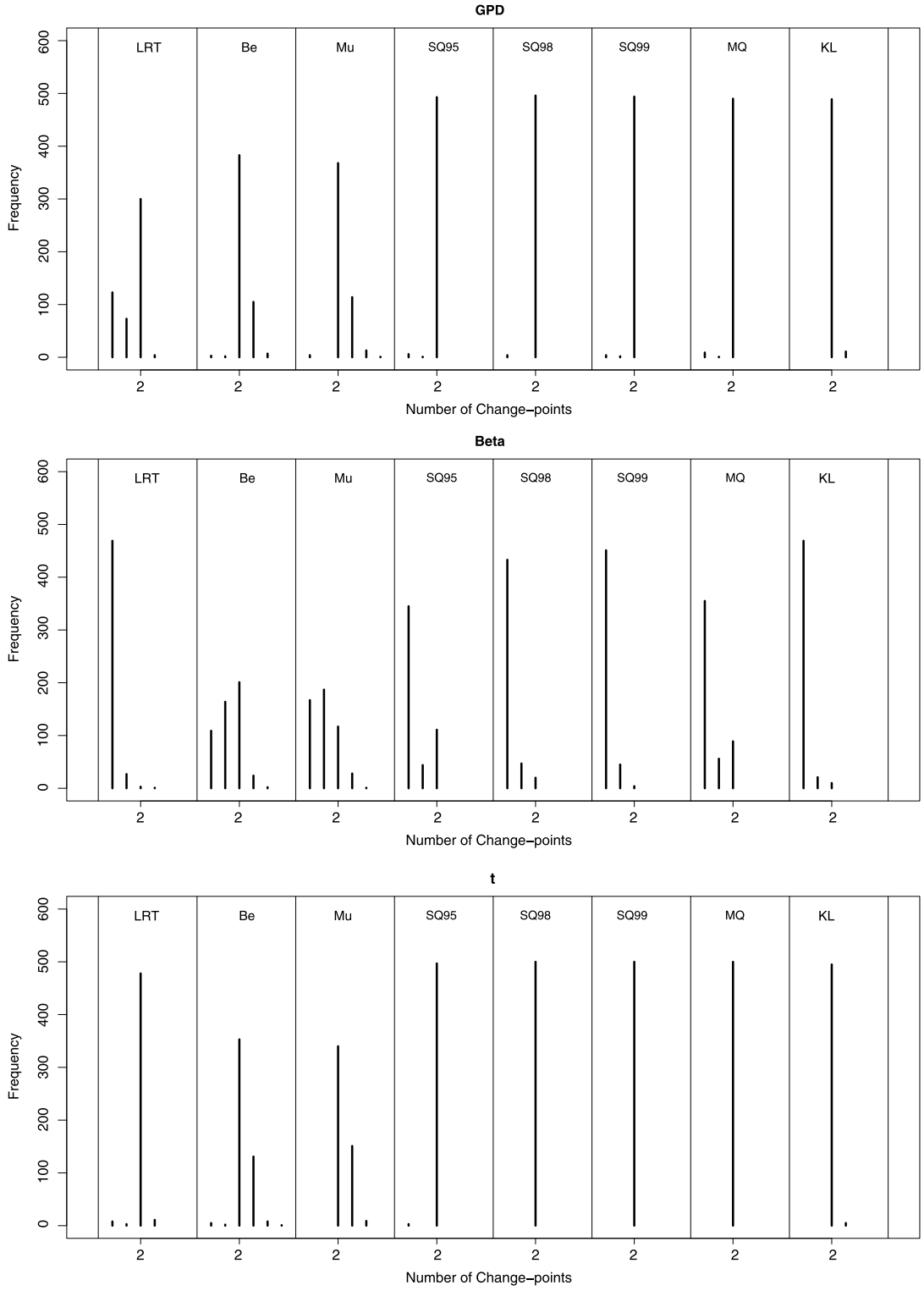

Figure 2. The histograms of the number of identified change-points in the multiple change-points simulation study for each of the methods for the GP, Beta and t distributions.

Table 4. Summary of properties of our change-point estimation methods

Method	Robustness/Flexibility	Multiple change-point detection	Speed
LRT	Relies on the GP distribution assumption/approximation for exceedance over threshold.	Applies single-change-point-detection method within each subsegment.	Slow
Be/Mu	Nonparametric	Uses multiple change-points estimation method based on the dynamic programming approach	Slow
SQ/MQ	Nonparametric	Uses a sequential test to determine number of change-points and estimate multiple change-points	Fast
KL	Nonparametric	Same as LRT	Fast

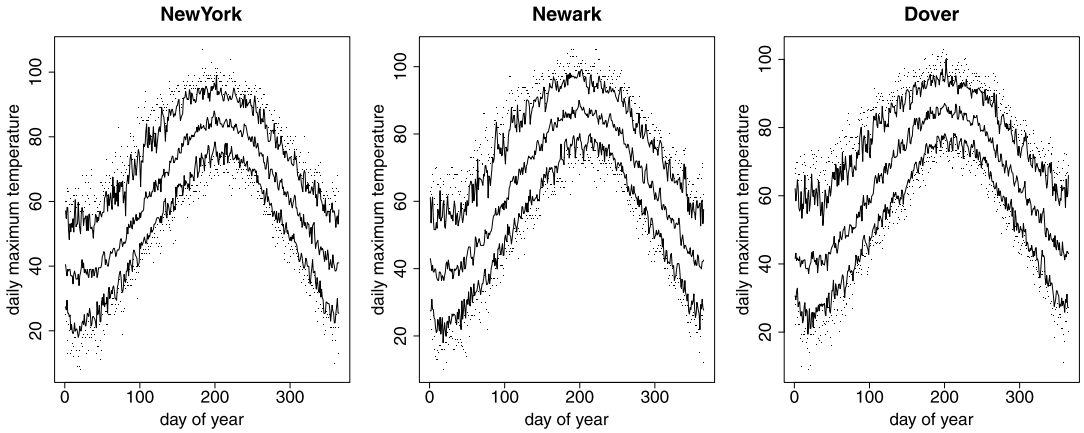

Figure 3. Minimum, 5% quantile, median, 95% quantile and maximum observed daily maximum temperatures (in degrees Fahrenheit) for New York, Newark and Dover, 1956-2005. WBAN station numbers are 14732, 14734 and 13707, respectively.

of days in a year is fixed to 365, an increased number of change-points means fewer observations between these change-points so these results are as expected.

4. CHANGE-POINTS IN TEMPERATURE DATA

We reconsider the historical daily maximum temperatures from 1956 to 2005, recorded to the nearest degree Fahrenheit, for New York City, analyzed in [5], and add the historical daily maximum temperatures at Newark, New Jersey and Dover, Delaware, over the same 1956 to 2005 period. The latter daily maximum temperatures are recorded to the nearest tenth of a degree Fahrenheit and are obtained from the National Oceanic and Atmospheric Administration (NOAA) at www.nesdis.noaa.gov. There were only one and two missing observations, respectively, and we simply filled in missing points by linear interpolation. Quantile curves for the observed daily maximum temperatures are plotted in Figure 3.

As daily maximum temperature data are non-stationary, exhibiting at least seasonality in the mean and in the variability, and autocorrelated, we first consider the time series model proposed in [5] to remove these complicating factors. Scaled residuals Z_{it} are then calculated and inspected for serial correlation. Estimated lag coefficients are very small (not shown) and there are no significant lags. The non-extreme part of the distribution of the Z_{it} is modeled by its empirical distribution $\tilde{F}_{Z,u}$, but we need to check for, and model if necessary, any non-stationarity in the extremes of the $\{Z_{it}\}$ process. The fundamental differences in heat-generating mechanisms at different times of the year, e.g. in winter, daily maximum temperature is driven by advection, while in the summer, clear skies and light winds can be the driving forces, encourage non-stationary extremes of Z_{it}. Figure 4 shows the scaled residuals, by season, for each of the three cities. Under the null hypothesis that the Z_{it} are *i.i.d.*, the expected number of the exceedances of any T of them over a high threshold u_p such that $\Pr(Z_{it} > u_p) = p$ is Tnp, and the total number of exceedances, $\sum_{i=1}^{n}\sum_{t=1}^{T} I(Z_t > u_p)$, follows a binomial$(nT, p)$ distribution. Table 5 shows the results of the tests by season. We clearly see that the Z_{it} are not *i.i.d.* over the year.

It is inadequate to simply adopt a model using the calendar-based seasons and we use our estimation methods to estimate more appropriate change-points. Consider a number of change-points, $T_1, T_2, \ldots, T_m$ such that the observations Z_{it} with $T_{j-1} \le t < T_j$, $j = 1, \ldots, m+1$ are con-

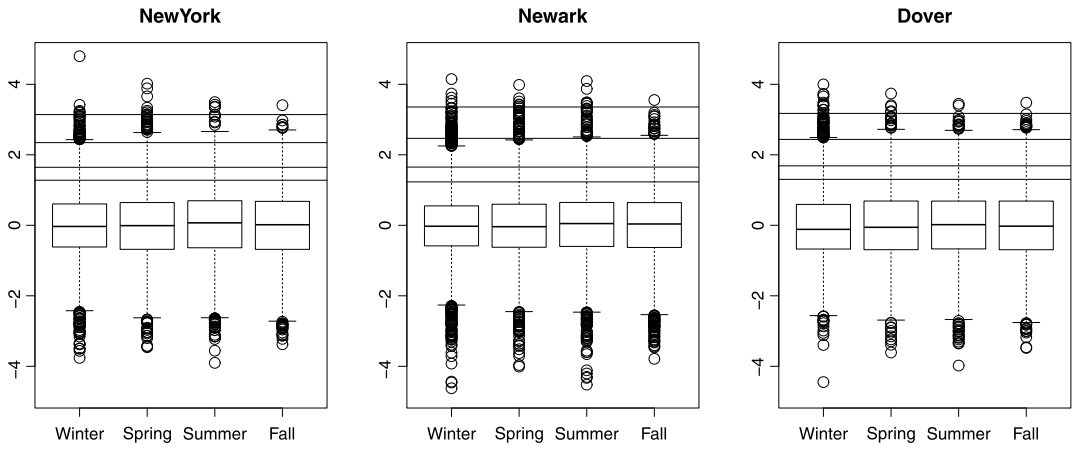

Figure 4. Residuals by season. Horizontal lines indicate 90%, 95%, 99% and 99.9% quantiles, respectively.

Table 5. Number of exceedances of residuals Z_{it} over high thresholds $q_{0.90}$ [1], $q_{0.95}$ [2], $q_{0.99}$ [3] and $q_{0.999}$ [4]. Expected number of exceedances over the four thresholds are approximately 455, 227.5, 45.5, and 4.55, respectively. Results of binomial test for too few or too many exceedances over the threshold: $*$ - significant at 0.05, $**$ - significant at 0.01

	New York				Newark				Dover			
	[1]	[2]	[3]	[4]	[1]	[2]	[3]	[4]	[1]	[2]	[3]	[4]
Winter	482	250	61*	5	475	244	59*	5	489	288**	69**	11**
Spring	467	241	61*	8	478	254	54	6	476	232	44	5
Summer	422	193*	37	5	427	193*	38	7	428	189**	30*	2
Fall	454	229	24**	1	445	222	32	1	432	204	40	1

sidered in the same season. Estimated seasonal breaks are shown in Figure 5. We see that while the optimal seasons are city-dependent and differ somewhat across the eight estimation methods, there is some general agreement. However, when examining the data, we note that for a given season, the GP parameters β and ξ could also be time-varying because of changes in weather patterns over the years due to climate change, industrialization, etc. For a season defined by $t \in [T_{j-1}, T_j)$, we allow one *change* over the 50-year period and refer to the year η_j in which it occurred as the change-year. We fit a GP with various time-covariate models: (i) allowing for one change-year in the value of ξ, (ii) allowing for one change-year in the value of β, (iii) allowing for one (common) change-year for the values of both ξ and β. More precisely, for Z_{it} with $t \in [T_{j-1}, T_j)$, $j = 1, \ldots, m+1$, we model $(Z_{it} - u_{z,j}) | Z_{it} > u_{z,j} \sim \text{GP}(\beta_{ij}, \xi_{ij})$ where model (i) $\xi_{ij} = \xi_{j0} + \xi_{j1} * I(i > \eta_j)$ and $\beta_{ij} = \beta_j$, model (ii) $\beta_{ij} = \beta_{j0} + \beta_{j1} * I(i > \eta_j)$ and $\xi_{ij} = \xi_j$, model (iii) $\xi_{ij} = \xi_{j0} + \xi_{j1} * I(i > \eta_j)$ and $\beta_{ij} = \beta_{j0} + \beta_{j1} * I(i > \eta_j)$, and $u_{z,j}$ is a high threshold. Models (i) and (ii) are nested in model (iii) and we carry out model selection based on likelihood considerations, performing likelihood ratio tests to assess evidence in favor of model (iii). Note that the change-point methods are not applied in this case. We simply consider integer-valued η_j such that $1966 \leq \eta_j < 1996$, and estimate η_j as the year which maximizes the likelihood for the 50-year period. The constraints on η_j are necessary to ensure that there is a sufficient number of points before, and after, the change-year so that the GP can actually be fitted. Results are detailed in Figure 5. All data and most methods are pointing in the direction of a change-point in the April to May time period and, for the January to April season, a change-year in the shape (shape and scale for Newark) of the upper tail during the mid-70s.

Estimated 100-year return levels using the [16] approach are shown for the three cities in Figure 6. We see that the ordering has been maintained over the 50 years and the estimates are most similar for the three cities during the problematic May period. Figure 7 shows increases in estimated 100-year return levels, from 1956 to 2005, for the three cities. We see that New York and Newark have seen comparable increases in the winter months, but the increase in Newark is about twice that in New York, 2 degrees Fahrenheit compared to 1, for the rest of the year. Dover seems to have been immune to these increases, indicating that perhaps increases in the other cities are the result of a heat-island effect. The heat-island effect increases with population, population density and urbanization. Dover has seen smaller increases in all three compared to New York and Newark.

5. DISCUSSION

There are three approaches to analyzing the extremes of non-stationary data: (i) including covariates in the parameters of the usual extreme value models for stationary series,

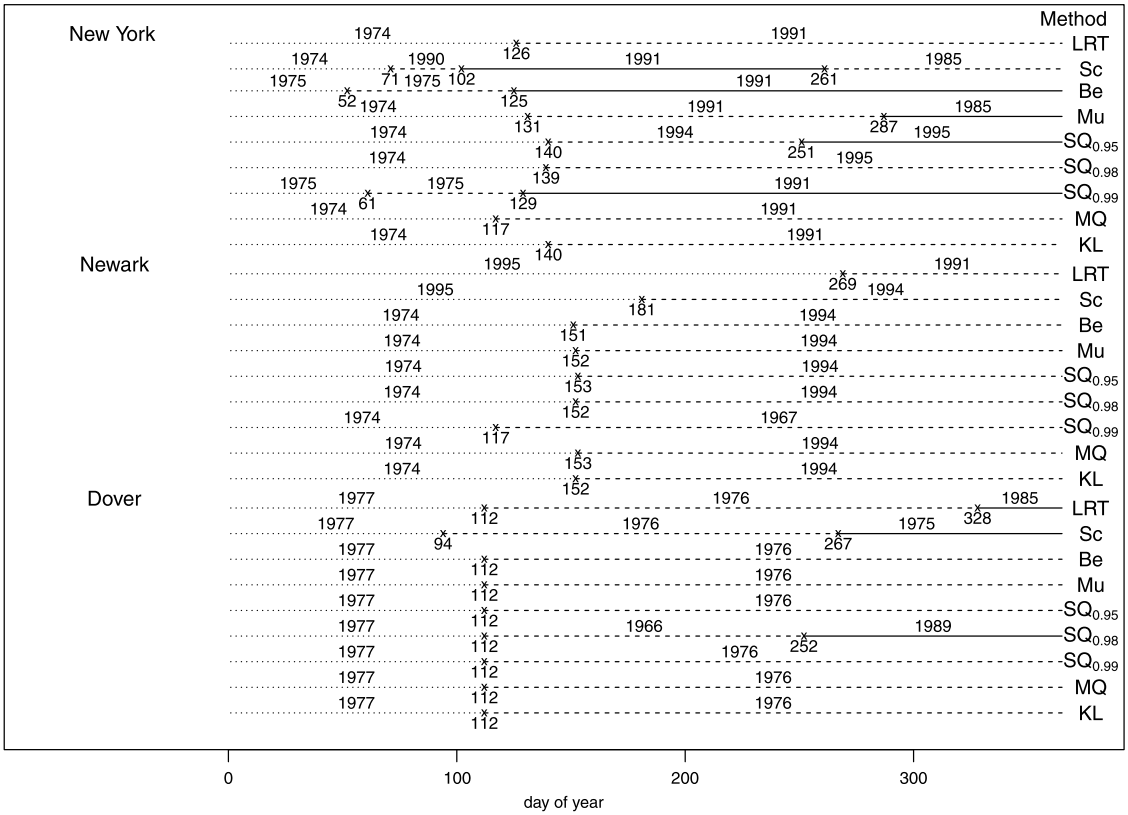

Figure 5. Estimated change-point day τ (under segment) marking the change of seasons during the year and estimated change-year η (over segment) at which there was a change in the shape (all three cities, all seasons) and scale (season 1, Newark only) parameter of the GP of scaled residuals, 1956-2005. E.g. The LRT approach estimates that there are two seasons in New York: Jan 1 to May 6 (Day 126) and May 7 to Dec 31. The GP has shape parameter ξ_{10} for Jan 1 to May 6 data up to 1974, and shape parameter $\xi_{10} + \xi_{11}$ for Jan 1 to May 6 data after 1974. Analogously, the GP has shape parameter ξ_{20} for May 7 to Dec 31 data up to 1991, and shape parameter $\xi_{20} + \xi_{21}$ for May 7 to Dec 31 data after 1991. Methods are labelled as in Table 2.

(ii) using a time-varying (or covariate-dependent) threshold, and (iii) preprocessing the original series before analyzing extremes in terms of residuals. Approach (i) above is not workable with the daily maximum temperature data. The complexity of the nonstationarity can simply not be captured by including covariates in the model parameters. One often encounters convergence problems with MLE when even only a few covariates are included, sometimes as few as one in the case of the shape parameter ξ. There is also much climatological evidence towards heavy day-to-day dependence in daily maximum temperatures (as there is empirical evidence in [5] where seasonal disturbances have large first-order regression coefficient) and approach (i) cannot handle the autoregressive behavior. A few have attempted approach (ii), see [5]. We took approach (iii) and preprocessed the maximum temperature series, accounting for location and scale of the daily maximum temperature distribution. When interest lies in changes in extreme tails, these are more affected by the shape of the error distribution. The methods reviewed and developed in this paper are used in Section 4 to detect changes in the latter.

Change-point analysis for extremes can focus on either the rate of exceeding of a high threshold, the excess over the high threshold, or both components. For the methods discussed in the paper, the Bernoulli and single quantile methods treat only the first component, while the likelihood ratio test and Kim and Lee method treat only the second component. The multiple quantile method aims to detect changes at any high quantiles $\tau \in \mathcal{T}$ and thus treats both components when $\mathcal{T}$ is an interval. When calculating the multiple quantile test statistic MQ, a grid of quantile levels from $\mathcal{T}$ are considered. Our numerical investigation shows that the multiple quantile method is insensitive to the choice of number of grid points used. In our applications, we considered only three thresholds for the multinomial method but more thresholds can be incorporated for the multinomial method to cover the far upper tail region. However, our additional simulations reveal that adding additional thresholds to the multinomial approach do not further improve its performance.

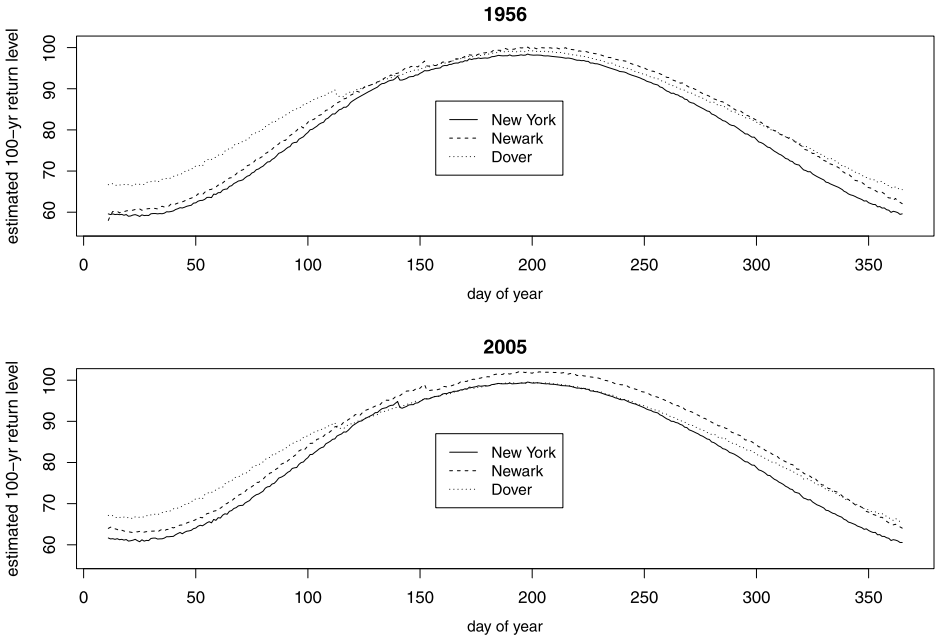

Figure 6. Estimated 100-year return levels for 1956 and 2005 for New York, Newark and Dover when using the Kim and Lee (2009) approach to find the change-points (i.e. determine seasons). Estimates are values that were only exceeded in 500 of 50,000 Monte Carlo simulated years.

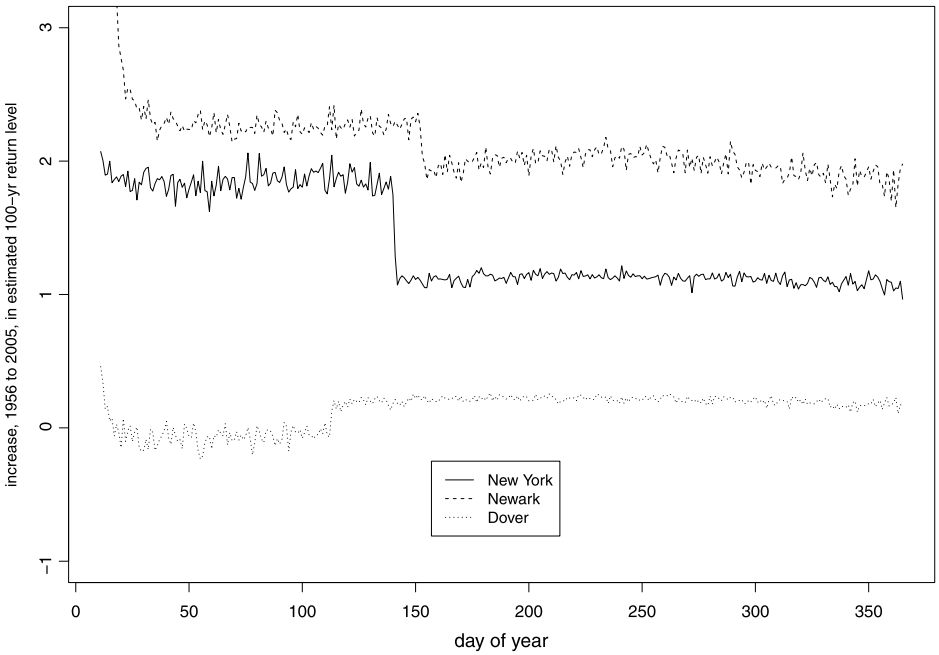

Figure 7. Increase in estimated 100-year return levels, from 1956 to 2005, for New York, Newark and Dover. Estimates are values that were only exceeded in 500 of 50,000 Monte Carlo simulated years.

Received 1 October 2013

REFERENCES

[1] Beaulieu, C., Chen, J., and Sarmiento, J. L. (2012). Change-point analysis as a tool to detect abrupt climate variations. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **370** 1228–1249.

[2] Beniston, M. (2009). Decadal-scale changes in the tails of probability distribution functions of climate variables in Switzerland. *International Journal of Climatology* **29** 1362–1368.

[3] Csörgő, M. and Horváth, L. (1997). *Limit Theorems in Change Point Analysis*, Wiley, Chichester. MR2743035

[4] Dierckx, G. and Teugels, J. L. (2010). Change point analysis of extreme values. *Environmetrics* **21** 661–686. MR2838438

[5] Dupuis, D. J. (2012). Modeling waves of extreme temperature: the changing tails of four cities. *Journal of the American Statistical Association* **107** 24–39. MR2949339

[6] Hawkins, D. M. (2001). Fitting multiple change-point models to data. *Computational Statistics & Data Analysis* **37** 323–341. MR1856677

[7] Hegerl G. C., Karl, T. R., Allen, M., Bindoff, N. L., Gillett, N., Karoly, D., Zhang, X. and Zwiers, F. (2006). Climate change detection and attribution: beyond mean temperature signals. *Journal of Climate* **19** 5058–5077.

[8] Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *Annals of Statistics* **3** 1163–1174. MR0378204

[9] Hoaglin, D. C., Mosteller, F. and Tukey, J. W. (Eds.) (1983). *Understanding Robust and Exploratory Data Analysis*, Wiley, New York. MR1800901

[10] Horvath, L. and Parzen, E. (1994). Limit theorems for Fisher-Score change processes. *IMS Lecture Notes – Monograph Series* **23**. MR1477922

[11] IPCC (2012). Summary for Policymakers. In: Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation [Field, C. B., V. Barros, T. F. Stocker, D. Qin, D. J. Dokken, K. L. Ebi, M. D. Mastrandrea, K. J. Mach, G.-K. Plattner, S. K. Allen, M. Tignor, and P. M. Midgley (eds.)]. *A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, UK, and New York, NY, USA, 1–19.

[12] Jandhyala, V. K., Fotopoulos, S. B. and Evaggelopoulos, N. (1999). Change-point methods for Weibull models with applications to detection of trends in extreme temperatures. *Environmetrics* **10** 547–564.

[13] Jaruskova, D. (1997). Some problems with application of change-point detection methods to environmental data. *Environmetrics* **8** 469–483.

[14] Jaruskova, D. and Rencova, M. (2008). Analysis of annual maximal and minimal temperatures for some European cities by change point methods. *Environmetrics* **19** 221–233. MR2420467

[15] Katz, R. W. and Brown, B. G. (1992). Extreme events in a changing climate: variability is more important than averages. *Climatic Change* **21** 289–302.

[16] Kim, M. and Lee, S. (2009). Test for tail index change in stationary time series with Pareto-type marginal distribution. *Bernoulli* **15** 325–356. MR2543865

[17] Kuglitsch, F. G., Toreti, A., Xoplaki, E., Della-Marta, P. M., Luterbacher, J. and Wanner, H. (2009). Homogenization of daily maximum temperature series in the Mediterranean. *Journal of Geophysical Research* **114** D15108, doi:10.1029/2008JD011606.

[18] Li, S. and Lund, R. (2012). Multiple Changepoint Detection via Genetic Algorithms. *Journal of Climate* **25** 674–686.

[19] Lu, Q. Q., Lund, R. and Lee, T. C. M. (2010). An MDL approach to the climate segmentation problem. *Annals of Applied Statistics* **4** 299–319. MR2758173

[20] Lund, R. B., Wang, X. L., Reeves, J., Lu, Q., Gallagher, C. and Feng, Y. (2007). Changepoint detection in periodic and autocorrelated time series. *Journal of Climate* **20** 5178–5190.

[21] Oka, T. and Qu, Z. (2011). Estimating structural changes in regression quantiles. *Journal of Econometrics* **162** 248–267. MR2795616

[22] Pickands, J. (1975). Statistical inference using extreme order statistics. *Annals of Statistics* **3** 119–131. MR0423667

[23] Qu, Z. (2008). Testing for structural change in regression quantiles. *Journal of Econometrics* **146** 170–184. MR2459652

[24] Quintos, C., Fan, Z. H. and Phillips, P. C. B. (2001). Structural change tests in tail behaviour and the Asian crisis. *Review of Economic Studies* **68** 633–663. MR1855475

[25] Reeves, J., Chen, J., Wang, X. L., Lund, R. and Lu, Q. (2007). A review and comparison of changepoint detection techniques for climate data. *Journal of Applied Meteorology and Climatology*, **46** 900–915.

[26] Robbins, M. W., Lund, R. B., Gallagher, C. M. and Lu, Q. Q. (2011). Changepoints in the North Atlantic tropical cyclone record. *Journal of the American Statistical Association* **106** 89–99. MR2816704

[27] Smith, R.L. (1985). Maximum likelihood estimation in a class of non-regular cases. *Biometrika* **72** 67–90. MR0790201

[28] Toreti, A., Kuglitsch, F. G., Xoplaki, E., and Luterbacher, J. (2012). A novel approach for the detection of inhomogeneities affecting climate time series. *Journal of Applied Meteorology and Climatology*, **51** 317–326.

D. J. Dupuis
Department of Management Sciences
HEC Montréal
3000, chemin de la Côte-Sainte-Catherine
Montréal (Québec)
Canada
H3T 2A7
E-mail address: debbie.dupuis@hec.ca

Y. Sun
Department of Statistics
Ohio State University
Columbus, OH 43210
USA
E-mail address: sun.1268@osu.edu

Huixia Judy Wang
Department of Statistics
George Washington University
Washington, DC 20052
USA
E-mail address: huixia@gmail.com