Non-stationarity of summer temperature extremes in Texas.

Meagan Carney* Robert Azencott† Matthew Nicol‡

September 21, 2018

Abstract

Modeling seasonal temperature extremes in weather patterns allows for better forecasting and prediction. Analysis of extreme values over a given time period is usually done by fitting a generalized extreme value (GEV) distribution to the maximum values in the data, however lack of sufficient weather recordings due to missing data or violation of independence assumptions (block maxima should be over a large time interval) may result in a poor fit for the GEV model. Modeling over larger, clustered regions may overcome some of these problems and can provide insight into macroscopic weather and climate changes.

In this paper we analyze temperature recordings in July and August taken from stations across Texas and New Orleans, Louisiana. We introduce clustering techniques which group stations by temperature trends and mutual information before performing extreme value analysis on the *clusters* of time series. This obviates some of the problems commonly encountered in analyzing single station weather data. Extreme analysis of the resulting clusters provides compelling evidence of non-stationarity of the distributional parameters in the GEV model and points to an increased likelihood from the period roughly 1980 to present of observing higher extreme temperatures for the months of July and August. We tabulate the probabilities of extreme temperatures in the clusters according to a non-stationary model. Our techniques can be easily adapted to a wide range of climatological problems.

1 Introduction

Extreme temperature events can make considerable impacts on society, most obviously on human health and power consumption. In addition there has been a growing interest in examining temperature extremes in relation to climate change [4][3][8][13][15]. In particular in the last decade the Texas Gulf Coast region has experienced high profile heat waves, such as the summer of 2011, and momentous summer rainfall and flooding.

*Department of Mathematics, University of Houston, Houston, USA. e-mail: meagan@math.uh.edu
Meagan Carney thanks the NSF for partial support on NSF-DMS Grant 1600780.
†Department of Mathematics, University of Houston, Houston, USA. e-mail: razencot@math.uh.edu
‡Department of Mathematics, University of Houston, Houston, USA. e-mail: nicol@math.uh.edu
Matthew Nicol thanks the NSF for partial support on NSF-DMS Grant 1600780.
Thanks due to the National Oceanic and Atmospheric Administration for providing the datasets from the isd lite data base.

In this paper we address the question of whether the probability of extreme summer temperatures in this region has increased by using clustering techniques and fitting generalized extreme value [2] (GEV) models to the data which allow for forecast modeling of the maxima by estimating the probability of high temperatures. Although the classical GEV model requires the mean and variance of the maxima to be stationary (not change in time) it is possible to adapt this probability model to allow for prediction in nonstationary scenarios[6][1][4], and that is what we do. We find compelling evidence that the probability of extreme temperatures during summer has increased. We do not consider rainfall patterns or the interaction between summer temperatures and rainfall in this paper.

It is necessary to have comprehensive information on extreme weather events to make reasonable conclusions from the data. Clustering techniques are very useful to allow enlargement of the time series data to permit better modeling. We apply clustering techniques to Texas wide weather stations to provide a larger pool of data for GEV fits of the distribution of maximum temperatures. Our approach is particularly effective when looking at large scale climate data and global environmental zones [17][16]. We use a combination of clustering methods, including K-means [7], with mutual information as a measure of similarity between weather stations. In this way we give a comprehensive extreme value analysis of summer extreme temperatures throughout Texas.

Recent extreme weather events in Houston, Texas, provide motivation for our station choice in this paper. Anecdotally, higher extreme temperature patterns have been recorded in Houston for the July-August months when compared to past records. The National Oceanic and Atmospheric Administration (NOAA) reports the longest stretch of record high temperatures August 1^{st} - August 24^{th} 2011 with the highest temperature ever recorded in Houston occurring on August 27, 2011. Preliminary results on the temperature vector from the Houston station suggest a significant change in the mean and variance of temperature maxima after year 1981. Other studies have also recorded more incidences of higher temperature outliers after 1980. [5][11][10][12] This provides motivation in the following analysis for breaking the time series into time windows 1941-1981, 1982-2017 and 1941-2017. We test these periods for stationarity and find across clusters the period 1941-1981 is stationary but this is not the case for the periods 1982-2017 and 1941-2017. Based on these results we fit nonstationary GEV models to estimate the probability of current temperature extremes in our clusters.

2 Data and Methods

2.1 Data

Each single time series in this analysis is defined as the NOAA hourly temperature recording vector for the July-August months every year from 1941-2017 for a single station. The 31^{st} of each month was not considered for 10 day block divisibility. See 2.2.2. Though many stations were listed by the NOAA across Texas the following stations were chosen for this analysis since they contain complete records aside from possible one

to two year gaps. New Orleans, Louisiana was included because of its known similarity to Houston weather. Missing data was excluded from this analysis.

Station	ID	Location	Type
1	690190	Abilene, TX	Airport
2	722310	New Orleans, LA	Airport
3	722436	Houston, TX	Airport
4	722505	Harlingen, TX	Airport
5	722517	Alice, TX	Airport
6	722530	San Antonio, TX	Airport
7	722533	Hondo, TX	Airport
8	722535	San Antonio, TX	Lackland AFB
9	722536	San Antonio, TX	Randolph AFB
10	722560	Waco, TX	Airport
11	722580	Dallas, TX	Airport
12	722595	Fort Worth, TX	Naval Air Station
13	722615	Del Rio, TX	Laughlin AFB
14	722640	Marfa, TX	Airport
15	722660	Abilene, TX	Airport
16	722700	El Paso, TX	Airport
17	723510	Wichita, Falls, TX	Sheppard AFB
18	723630	Amarillo, TX	Airport

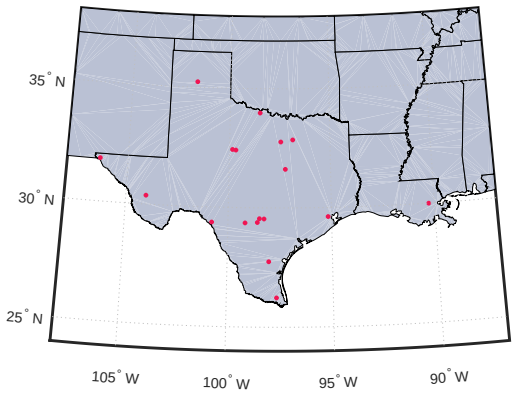


Figure 1: map of all listed stations.

2.2 Methods

2.2.1 Clustering Methods

2.2.1.1 Mutual Information and Data Compression

Mutual information between two time series is given as a function of the entropy and joint entropy and is used in analysis as a measurement of the similarity between two

time series. Suppose Z^1 and Z^2 are two time series then the mutual information is given by,

$$I(Z^1, Z^2) = H(Z^1) + H(Z^2) - H(Z^1, Z^2)$$

where $H(Z^1)$ is the calculated entropy given by,

$$H(Z^1) = -\sum_{i=1}^{m} p(z_i^1) \log p(z_i^1)$$

and the joint entropy is given by,

$$H(Z^1, Z^2) = -\sum_{i=1}^{m} \sum_{j=1}^{n} p(z_i^1, z_j^2) \log p(z_i^1, z_j^2)$$

where $p(z_i^1) = P(Z^1 = z_i^1)$ is the probability of the time series equal to $z_i^1 \in \{v_1, v_2 ... v_m\}$ and $p(z_i^1, z_j^2) = P(Z^1 = z_i^1, Z^2 = z_j^2)$ is the joint probability of the time series equal to $z_i^1 \in \{v_1, v_2 ... v_m\}$ and $z_j^2 \in \{u_1, u_2 ... u_n\}$. In this analysis mutual information between two time series was only calculated over non-missing years for *both* time series.

Given the data processing inequality,

$$I(T(X), R(Y)) \leq I(T(X), Y) \leq I(X, Y)$$

with equality only when $T(\cdot)$ and $R(\cdot)$ are invertible our goal is to find transformations (though not invertible by definition) which compress the continuous time series while preserving the maximum possible mutual information between each pair of time series.

We seek values $a_{\max}^{1,2,3,4}$ such that the mutual information between two centered time series $X^1 = (x_1^1, x_2^1 ... x_l^1)$ and $X^2 = (x_1^2, x_2^2 ... x_p^2)$ is maxmized where compression is given by,

If $x_k^1 > a^2$ then $z_k^1 = 1$

If $a^1 < x_k^1 < a^2$ then $z_k^1 = 0$

If $x_k^1 < a^1$ then $z_k^1 = -1$

and,

If $x_k^2 > a^4$ then $z_k^2 = 1$

If $a^3 < x_2^1 < a^4$ then $z_k^2 = 0$

If $x_k^2 < a^3$ then $z_k^2 = -1$

Since *entropy* is maximized when the time series is uniformly distributed, it is reasonble to choose starting points of gradient ascent a_0^1, a_0^3 at the 1/3 quantile of their respective time series (similarly for a_0^2, a_0^4 at the 2/3 quantile.) The gradient value of mutual information (now as a function of the interval endpoint) was estimated as,

$$\Delta I_t = \frac{I(a_t) - I(a_{t-1})}{a_t - a_{t-1}}$$

where a_t is the lefthand (righthand) value of the interval with fixed righthand (lefthand) at the current time step and our updated a_{t+1} is given by

$$a_{t+1} = a_t + \gamma \Delta I_t$$

where $\gamma = 0.1$ is the specified step size multiplier and $a_1 = a_0 + 1$. Gradient ascent was performed individually for each endpoint[1] starting at a^1 and finishing at a^4. Fixed values for all other endpoints corresponded to the initial value $a_0^{1,2,3,4}$ or, if previously calculated, the output value from maximization $a_{\max}^{1,2,3,4}$ of the mutual information. To deal with local maximum, if the mutual information reached a steady state a perturbation of 0.1 was added to a_t. The algorithm was stopped when the steady state with perturbation on $a_t^{1,2,3,4}$ remained for 100 time steps.

The number of values z_k in the kth state of our newly compressed time series can be seen as binomally distribution with $E(z_k) = Np_k$ and variance $V(z_k) = Np_k(1 - p_k)$ where the true probability of being in the kth state is given by p_k and N is the length of the time series. Then the error $\tilde{\epsilon}_k$ on the estimated proportion $\hat{p}_k = z_k/N$ has normal distribution with $E(\tilde{\epsilon}_k) = 0$ and $V(\tilde{\epsilon}_k) = \hat{p}(1 - \hat{p})/N$ for large enough N. The relative error size ϵ_k on $\hat{p}_k \log \hat{p}_k$ is taken as

$$\epsilon_k = \sum_k \frac{\delta(\hat{p}_k \log \hat{p}_k)}{\delta \hat{p}_k} \tilde{\epsilon}_k$$

but $E(\epsilon_k) = 0$ so,

$$V(\epsilon_k) = \sum_k (1 + \log \hat{p}_k)^2 \frac{\hat{p}_k(1 - \hat{p}_k)}{N}$$

Note that by choosing threshold values resulting in a uniform distribution for z_k, we have that the $V(\epsilon_k) \approx 0$ (entropy is maximized) for each time series so that we take the relative error to be the $V(\epsilon_k)$ taken over the approximate joint entropy. The relative error on the mutual information caluculated for each pair of time series is of order 10^{-5} so that the total error over all pairs of time series is of order 10^{-3}. See figure A.1 for histogram of maximized mutual information and A.2 for the error on the maximized mutual information located in the appendix.

2.2.1.2 Dimensionality Reduction

An undirected graph was created where nodes correspond to stations and an edge represents a strictly positive mutual information between stations. If the mutual information between two stations was less than 0.1 it was assumed to be zero. The goal of this section is to remove a set of edges from the chosen connected component which carry the lowest amount of information.

We begin by calculating the normalized laplacian for the connected component given by $L = I - D^{-1/2} S D^{-1/2}$ where D is the diagonal matrix with entries $D_{i,i} = \sum_{j=1}^{h} S_{i,j}$

[1] Our decision to fix all other end points and performing gradient ascent on the one-dimensional mutual information comes from the degree of uncertainty found in the data.

and $S = S_{i,j}$ is the symmetric matrix of mutual information between time series. Eigenvalues $\lambda_1, ..., \lambda_h$ and eigenvectors $V_1, ..., V_h$ were calculated for L then sorted in ascending order of eigenvalues. The sum of the eigenvalues $s(k) = \sum_{i=1}^{k} \lambda_k$ and the ratio $R(k) = \frac{s(k)}{s(h)}$ were calculated and plotted for each similarity matrix and each interval. A cut off point $k = J$ was chosen such that $R(J-1) < 0.15 \leq R(J)$. The number of nodes before this point are approximated $0.15 * h$ where h is the total number of nodes in the connected component.Vector projections for each node in the chosen connected component $W_1, ..., W_h$ onto the J dimensional subspace spanned by $V_1, ..., V_J$ were calculated by noting that $W_1, ..., W_h$ are the corresponding row vectors in the $h \times J$ matrix $[V_1 ... V_J]$ (a special property of the normalized laplacian). These J-dimensional vector projections $W_1, ..., W_h$ served as the inputs to the k-means clustering algorithm. See figure A.3 in appendix for eigenvalues and ratios.

2.2.1.3 k-Means Clustering

The k-means clustering algorithm seeks to minimize cost by performing the following until a steady state is reached:

$$\min \sum_{k=1}^{N} \sum_{l=1}^{M} ||n(l) - C(k)||_{\mathbb{R}^J}$$

where M is the number of nodes n in the cluster k, $C(k)$ is the centroid of cluster k and N is the number of clusters. The built-in MATLAB k-means algorithm was run 1,000 times on our J-dimensional vectors $W_1, ..., W_h$ and $k = 4$ clusters. Determining the "accuracy" of resulting clusters is a wide topic of dicussion in unsupervised learning. In our analysis we consider a good cluster one which has (1) a low dispersion of points and (2) a large separation between centroids.
The davies-bouldin index, given by

$$\frac{1}{N} \sum_{1}^{N} \max \frac{MSE_j + MSE_k}{||C(j) - C(k)||}$$

where $j, k = 1, ..., N$, N is the number of clusters and $MSE_j = \frac{1}{M} \sum_{n(l) \in C(j)} d(n(l), C(j))^2$ is often used to determine how separated a cluster is. We note that small values of the davies-bouldin index indicate larger separation of clusters. This value was calculated for each run of the k-means algorithm. The set of clusters associated to the minimum davies-bouldin index was used in the following extreme value analysis.

2.2.2 Extreme Value Methods

In this paper we define an extreme value or maxima by the maximum temperature over 10 day time blocks where each block is a disjoint series of hourly temperature records $\sim$ 240 data points long.

2.2.2.1 Extreme Value Distribution

We seek to fit the of from each cluster of time series (that is, ~ 456 block maxima taken from each station so that our distribution represents the *combined* block maxima of all the time series in the cluster ~ 456 times the number of time series in the cluster) to a generalized extreme value distribution given by,

$$f(x|k,\mu,\sigma) = \frac{1}{\sigma}\exp(-(1+k\frac{x-\mu}{\sigma})^{\frac{1}{k}})(1+k\frac{x-\mu}{\sigma})^{-1-\frac{1}{k}}$$

for,

$$1+k\frac{x-\mu}{\sigma} > 0$$

with location parameter μ, scale parameter σ and shape parameter $k \neq 0$. $k > 0$ corresponds to the Type II case, while $k < 0$ corresponds to the Type III case. When $k = 0$ (Type I), the density is,

$$f(x|0,\mu,\sigma) = \frac{1}{\sigma}\exp(-\exp(-\frac{x-\mu}{\sigma}) - \frac{x-\mu}{\sigma})$$

We estimate the parameter values k, μ and σ for the GEV fit through maximum likelihood estimation (MLE). Independence assumptions are met by performing a chi-square test of independence over blocks of 10 days where every 10 day block reflects independence at the $\alpha = 0.05$ confidence level. The goal of MLE is the maximize the log-likelihood function for the GEV of $M_1, ..., M_n$ (maxima over n blocks) given by,

$$\log L_{k,\mu,\sigma} = \log\prod_{i=1}^{n} f_{k,\mu,\sigma}(M_i) = \sum_{i=1}^{n} l_{k,\mu,\sigma}(M_i)$$

$$l_{k,\mu,\sigma}(x) = -(1+1/k)\log(1+k\frac{x-\mu}{\sigma}) - (1+k\frac{x-\mu}{\sigma})^{-1/k} - \log\sigma$$

or equivalently in element form,

$$l_{k,\mu,\sigma}(x) = -(1+1/k)\sum_{i=1}^{n}\log(1+k\frac{x_i-\mu}{\sigma}) - \sum_{i=1}^{n}(1+k\frac{x_i-\mu}{\sigma})^{-1/k} - n\log\sigma$$

After maximum likelihood estimation of the distributional parameters were calculated the Anderson-Darling (A-D) goodness of fit test was performed on the binned data and resulting distributions. Within the 95% confidence intervals associated to each MLE parameter 10 points were chosen so that a total of 1,000 permutations of 3 parameters were tested for goodness of fit on the binned data. The histogram of 1,000 p values associated to each run of the A-D test for 1941-1981, 1982-2017 and 1941-2017 was plotted to determine whether a difference in GEV fit exists between each time interval.

Stationarity of the time series is a necessary condition when fitting a GEV distribution to the maxima. A random process $x_1, x_2, ...$ is said to be stationary if, given any set of integers $\{i_1, ..., i_k\}$ and any integer m, the joint distributions of $(x_{i_1}, ..., x_{i_k})$ and

$(x_{i_1+m}, ..., x_{i_k+m})$ are identical. Such a requirement excludes time series with trends, seasonality and other deterministic cycles. When a poor GEV fit is returned a common step is to check whether the parameters are time-dependent. In this analysis we perform non-parametric Mann-Kendall and Theil-Sen and parametric linear regression tests on the yearly mean and variance of the *combined* cluster of time series.[2]

2.2.2.2 Nonstationary Extreme Value Distribution

The goal of this section is to find a model which more accurately represents the *combined* generalized extreme value distribution for the block maxima of each cluster of time series. Trend and regression tests suggest a time dependent quadratic and linear model for the mean and variance respectively. See 3.2 for results. We will discuss maximum likelihood estimates with non-stationary parameters for the individual time series and their relationship to the combined non-stationary GEV distribution for each cluster.

Recall the log-likelihood function for the stationary GEV,

$$l_{k,\mu,\sigma}(x) = -(1+1/k)\sum_{i=1}^{n}\log(1+k\frac{x_i-\mu}{\sigma}) - \sum_{i=1}^{n}(1+k\frac{x_i-\mu}{\sigma})^{-1/k} - n\log\sigma$$

where n is the number of block maxima. Then the log-likelihood function for GEV with non-stationary, quadratic mean $\mu(t) = \beta_0+\beta_1 t+\beta_2 t^2$ and linear variance $\sigma(t) = \alpha_0+\alpha_1 t$ is given by,

$$l_{k,\beta_0,\beta_1,\beta_2,\alpha_0,\alpha_1}(x) = -(1+1/k)\sum_{i=1}^{n}\log(1+k\frac{x_i-(\beta_0+\beta_1 t+\beta_2 t^2)}{(\alpha_0+\alpha_1 t)}) - \sum_{i=1}^{n}(1+k\frac{x_i-(\beta_0+\beta_1 t+\beta_2 t^2)}{(\alpha_0+\alpha_1 t)})^{-1/k}$$
$$-n\log(\alpha_0+\alpha_1 t)$$

Parameters were estimated by maximizing the negative log-likelihood function with time dependent models of mean and variance for the individual time series within each cluster where t is given as the block maxima index vector.

Recall that for a distribution created from n samples of m distributions we have the mean of the distribution,

$$\mu = \frac{\sum_{i=1}^{m} n\bar{x}_i}{mn}$$

where $\bar{x}_i$ is the mean of the sample coming from one of the m distributions and the variance of the distribution,

$$\sigma^2 = \frac{\sum_{i=1}^{m} n(s_i^2+d_i^2)}{mn}$$

where s_i^2 is the variance of the sample and $d_i = \bar{x}_i-\mu$. Then for the time dependent mean and variance parameters of the combined GEV distribution created from the clustered time series we have,

$$\mu(t) = \frac{\sum_{i=1}^{m} n\bar{x}_i(t)}{mn} \tag{1}$$

[2]The mean and variance of the block maxima is calculated over all time series in the cluster for that year resulting in a single mean and variance per year.

and

$$\sigma^2(t) = \frac{\sum_{i=1}^{m} n(s_i^2(t) + d_i^2(t))}{mn} \tag{2}$$

where $\bar{x}_i(t)$ and $s_i^2(t)$ are the maximum likelihood estimate models from the individual time series. Under the assumption of a "perfect" model for $\mu(t)$ and $\sigma(t)$ we should have each individual time series of block maxima $X^i \sim GEV(k, \bar{x}_i(t), s_i(t))$ so that the normalized time series of block maxima

$$Z^i = \frac{x_j - \mu(t = j)}{\sigma(t = j)}$$

where j is the index of the block maxima and the *combined* normalized time series[3], $Z = \{Z^i\} \sim GEV(\hat{k}, 0, 1)$ as $j \to \infty$ which implies that the estimated combined $\hat{\mu}(t) \to 0$ and $\hat{\sigma}(t) \to 1$ as $t \to \infty$.

Maximum likelihood estimation was used to determine the parameters of the combined normalized GEV distribution. The Anderson-Darling goodness of fit test was performed on the data with estimated parameters.

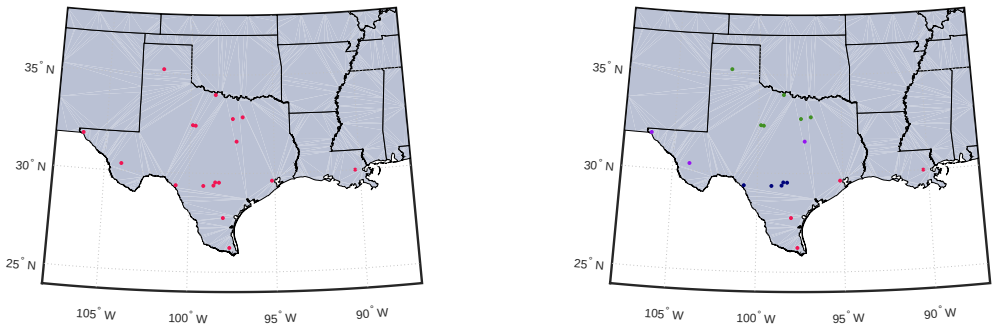
3 Results and Discussion

3.1 Clustering Results

The following table reflects the clustered stations associated to the minimum Davies-Bouldin index. Clusters are stable for each time window. Geographical locations of clusters are given by (1) Coastal (2) Southern Texas (3) Northern Texas (4) Along the Central Band.

time window	1941-1981	1982-2017	1941-2017
cluster (1)	2,3,4,5	”	”
cluster (2)	6,7,8,9,13	”	”
cluster (3)	1,11,12,15,17,18	”	”
cluster (4)	10,14,16	”	”
davies-bouldin index	7.75e-4	8.94e-4	6.34e-4

[3]Issues of modality are assumed to be negligable since the distribution is unimodal and each mean in the cluster is within 2 standard deviations of the other.



(a) BEFORE (b) AFTER

Figure 2: regional map with clustered stations. colors represent clusters.

3.2 Extreme Value Results

3.2.1 Stationary Extreme Value Distribution

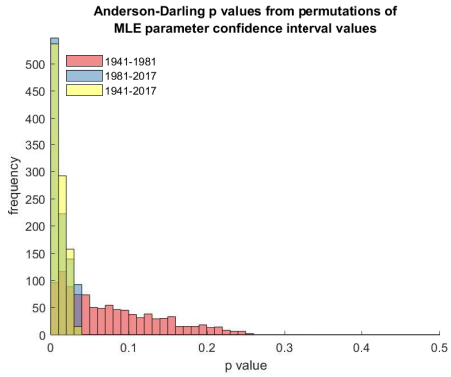
H_0: The data follows the theoretical GEV distribution with estimated parameters
H_A : The data does not follow the distribution

Anderson-Darling p-values for MLE parameters

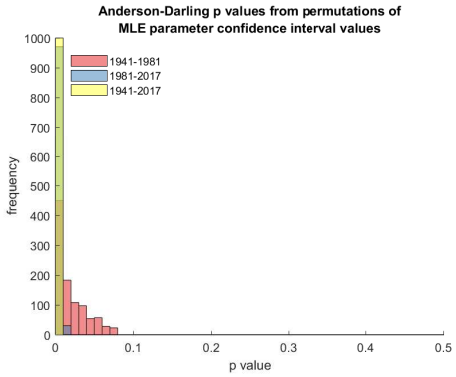
years	cluster(1)	cluster(2)	cluster(3)	cluster(4)
1941-1981	0.23	0.05	0.06	0.04
1982-2017	0.03	0.01	0.08	0.03
1941-2017	0.03	7.17e-4	0.01	6.21e-4

Results from the Anderson-Darling goodness of fit test suggest a better fit for 1941-1981 and 1982-2017 when compared to 1941-2017 for all clusters. The histogram of p-values (see figure 3 and section 2.2.2.1) for clusters (1) and (2) show higher likelihoods of being below the alpha = 0.05 confidence limit (e.g. conclude the data do not come from the specified distribution) for groups 1982-2017 and 1941-2017. Clusters (3) and (4) show higher likelihoods of being below alpha = 0.05 confidence limit for 1941-2017.

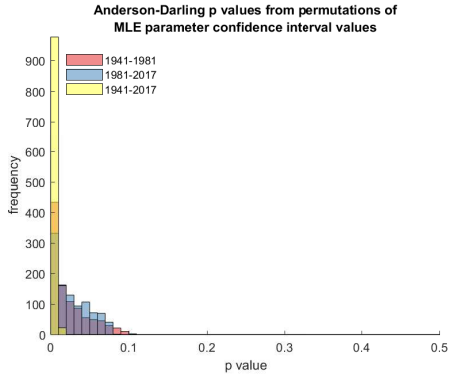
It is the case for all clusters that the GEV model with maximum likelihood estimates does not fit for the whole of 1941-2017 where it is reasonable to conclude from the results that a change in parameters happens between groups 1941-1981 and 1982-2017. Moreover, a change in parameters continues to be significant enough after 1981 for clusters (1) and (2) to result in a poor fit for 1982-2017.



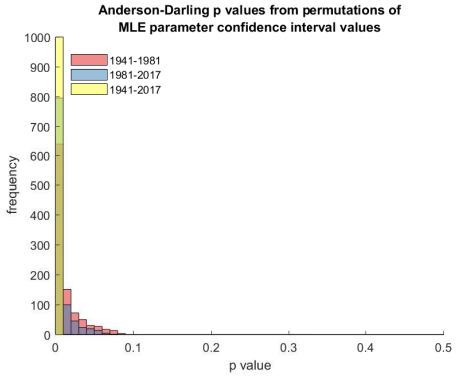
(a) cluster(1)



(b) cluster(2)



(c) cluster(3)



(d) cluster(4)

Figure 3: histogram of Anderson-Darling p-values obtained by varying the estimated parameters within the 95% confidence interval of the MLE parameters.

3.3 Mean and Variance Tests for Trend

3.3.0.1 Mean

Non-Parametric Mann-Kendall trend test
H_0: There is no monotonic trend in the data.
H_A: There exists a monotonic trend in the data.

years	cluster(1)	cluster(2)	cluster(3)	cluster(4)
1941-1981	2.20e-3	0.01	0.03	0.29
1982-2017	1.4e-3	0.13	0.32	0.92
1941-2017	6.82e-5	0.55	0.74	3.61e-4

Non-Parametric Theil-Sen estimator

cluster	(1)		(2)		(3)	
years	m	b	m	b	m	b
1941-1981	-0.08	96.73	-0.08	100.11	-0.09	100.78
1982-2017	0.05	96.03	0.05	98.15	0.04	98.71
1941-2017	0.04	94.80	0.01	98.66	-4.3e-3	99.48

(4)	
m	b
-0.03	99.29
3.3e-3	96.67
-0.04	99.38

Parametric Linear Regression

cluster	(1)			(2)			(3)		
years	m	b	r	m	b	r	m	b	r
1941-1981	-0.07	96.50	-0.46	-0.09	100.16	-0.48	-0.07	100.70	-0.35
1982-2017	0.06	96.11	0.44	0.04	98.34	0.24	0.05	98.55	0.26
1941-2017	0.04	94.71	0.45	0.01	98.35	0.10	3.82e-4	99.32	3.9e-3

(4)		
m	b	r
3.9e-3	98.17	0.02
0.01	96.33	0.09
-0.03	98.63	-0.32

Correlation coefficients for the mean for clusters (1) , (2) and (3) stay the same or decrease for 1941-2017. Moreover, coefficients switch signs from 1941-1981 and 1982-2017. Results from the non-parametric Mann-Kendall test suggest a monotonic trend ($\alpha = 0.05$ significance level) for the mean exists in 1941-1981 for clusters (1), (2) and (3) and 1941-2017 for clusters (1) and (4), however linear regression outputs suggest that the phenomenon of observed trend may be the result of cutting the mean vector mid cycle and motivate fitting the mean to a quadratic model.

3.3.0.2 Variance

Non-Parametric Mann-Kendall trend test
H_0: There is no monotonic trend in the data.
H_A: There exists a monotonic trend in the data.

years	cluster(1)	cluster(2)	cluster(3)	cluster(4)
1941-1981	0.66	0.21	0.55	0.14
1982-2017	0.54	0.15	0.71	0.75
1941-2017	0.02	1.74e-4	0.36	1.26e-6

Non-Parametric Theil-Sen estimator

cluster	(1)		(2)		(3)	
years	m	b	m	b	m	b
1941-1981	0.03	8.38	0.05	6.91	0.06	11.59
1982-2017	-0.04	11.96	0.13	8.78	0.03	12.90
1941-2017	0.06	7.98	0.09	5.81	0.03	11.89

(4)	
m	b
0.17	9.48
0.04	25.15
0.27	8.67

Parametric Linear Regression

cluster	(1)			(2)			(3)		
years	m	b	r	m	b	r	m	b	r
1941-1981	0.01	9.58	0.02	0.06	6.43	0.18	0.09	12.74	0.16
1982-2017	-0.03	13.58	-0.05	0.12	9.42	0.22	0.05	14.18	0.10
1941-2017	0.06	8.92	0.25	0.09	5.85	0.43	0.03	13.77	0.11

(4)		
m	b	r
0.27	10.03	0.32
-0.07	32.78	-0.04
0.34	10.38	0.41

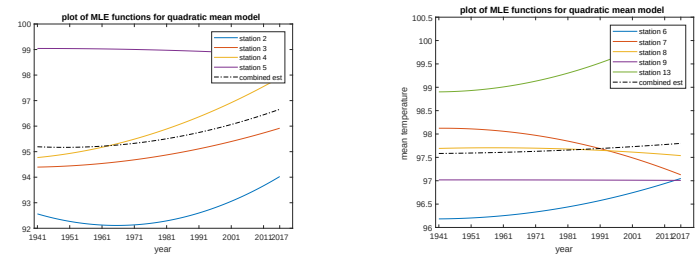
The non-parametric Mann-Kendall test returns no monotonic trend in the variance over 1941-1981 and 1982-2017 for all clusters and a monotonic trend for clusters (1) (2) and (4) over the time period 1941-2017 at the $\alpha = 0.05$ confidence level. Though cluster (3) does not report a monotonic trend for 1941-2017 it does return a lower p-value when compared to 1941-1981 and 1981-2017. Such a result suggests a significant difference in variance between 1941-1981 and 1982-2017. Higher correlation coefficients returned by linear regression for 1941-2017 support conclusions from the Mann-Kendall test. Theil-Sen and linear regression models have similar values for slopes and intercepts which imply that the monotonic trend which exists is not the result of extremal outliers. Results from all three tests provide motivation for fitting the variance parameters for 1941-2017 to a linear model.[4]

3.3.1 Nonstationary Extreme Value Distribution

The quadratic and linear models of the time dependent mean and variance parameters for the individual time series given by maximum likelihood estimation of the nonstationary GEV model (See 2.2.2.2) and the time-dependent model of the cluster are plotted in

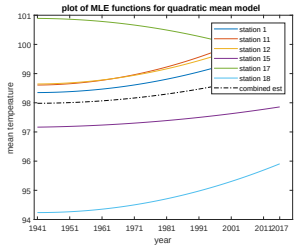
[4]Regression results for cluster (1) report sign changes between 1941-1981 and 1982-2017 which suggest a possible quadratic trend exists for the variance. Our justification for keeping a linear model for cluster (1) comes from the existance of a monotonic trend over 1941-2017 and all other clusters favoring a linear model, however the final variance model reflects a quadratic trend.

figures 4 and 5 on page 14. Maximum likelihood estimates, 95% confidence intervals and A-D goodness of fit p-values before and after nonstationary modeling of the combined GEV for each cluster are also given. See tables on page 14.

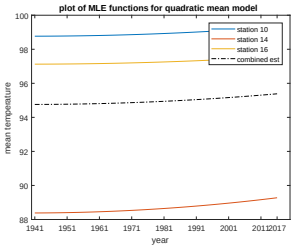


(a) cluster(1)

(b) cluster(2)



(c) cluster(3)



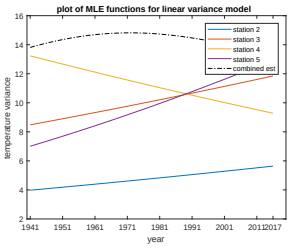
(d) cluster(4)

Figure 4: mean models of block maxima for individual and combined time series.

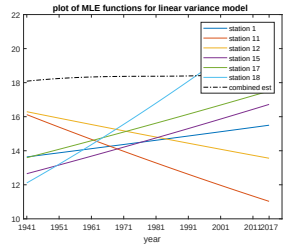
Confidence intervals from maximum likelihood estimates suggest asymptotic convergence of $\hat{\mu} \to 0$ and $\hat{\sigma} \to 1$ as the number of block maxima tend to infinity for clusters (1), (2) and (3) which provide motivation for retaining our choice of mean and variance model. Normalized GEV distributions with nonstationary mean (quadratic) and variance (linear) parameters report significantly better fits from the stationary GEV distribution for all clusters. Moreover, clusters (1) and (3) conclude the normalized block maxima come from the GEV distribution with MLE parameters at the $\alpha = 0.05$ significance level over all time intervals.

cluster(1): MLE of normalized GEV

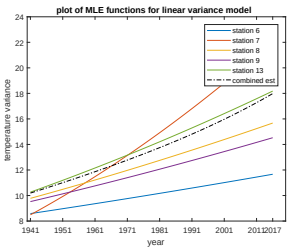
params	k			σ			μ		
years	value	CI		value	CI		value	CI	
1941-1981	-0.18	-0.23	-0.14	0.87	0.82	0.92	-0.32	-0.39	-0.24
1982-2017	-0.25	-0.28	-0.23	0.10	0.95	1.05	-0.08	-0.15	-5.1e-3
1941-2017	-0.23	-0.25	-0.21	0.95	0.92	0.99	-0.18	-0.23	-0.13



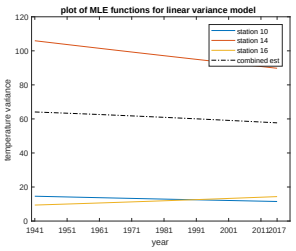
(a) cluster(1)



(b) cluster(2)



(c) cluster(3)



(d) cluster(4)

Figure 5: variance models of block maxima for individual and combined time series.

A-D	p-values	
years	before	after
1941-1981	0.23	0.52
1982-2017	0.03	0.19
1941-2017	0.03	0.26

cluster(2): MLE of normalized GEV

params	k			σ			μ		
years	value	CI		value	CI		value	CI	
1941-1981	-0.25	-0.28	-0.22	0.99	0.94	1.04	-0.09	-0.16	-0.02
1982-2017	-0.27	-0.29	-0.24	1.00	0.96	1.04	-0.02	-0.09	0.04
1941-2017	-0.26	-0.28	-0.24	1.00	0.97	1.03	-0.05	-0.10	-2.7e-3

A-D	p-values	
years	before	after
1941-1981	0.05	0.14
1982-2017	0.01	0.01
1941-2017	7.16e-4	4.60e-3

cluster(3): MLE of normalized GEV

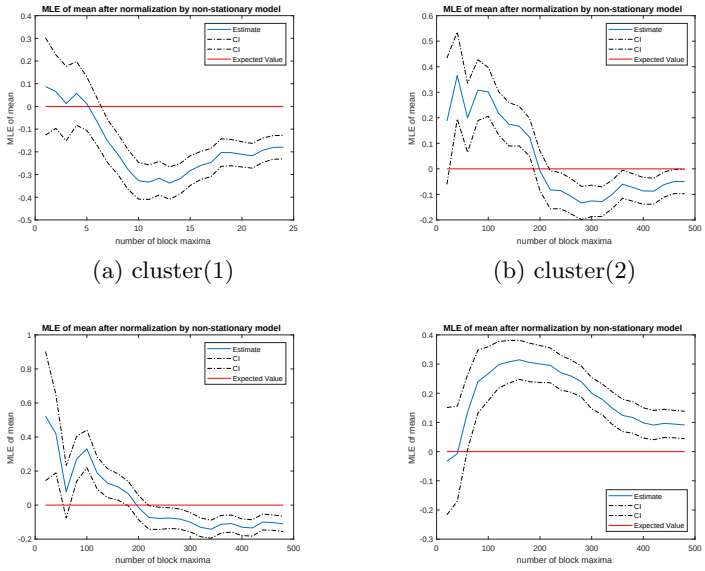
params	k			σ			μ		
years	value	CI		value	CI		value	CI	
1941-1981	-0.26	-0.28	-0.24	1.04	1.00	1.09	-0.08	-0.14	-0.01
1982-2017	-0.30	-0.33	-0.27	1.01	0.97	1.06	-0.12	-0.18	-0.06
1941-2017	-0.26	-0.27	-0.24	1.02	0.99	1.05	-0.11	-0.16	-0.07

A-D	p-values	
years	before	after
1941-1981	0.06	0.18
1982-2017	0.08	0.70
1941-2017	0.01	0.17

cluster(4): MLE of normalized GEV

params	k			σ			μ		
years	value	CI		value	CI		value	CI	
1941-1981	-0.32	-0.35	-0.28	0.59	0.55	0.63	0.27	0.21	0.33
1982-2017	-0.41	-0.43	-0.38	0.80	0.76	0.84	-0.02	-0.08	0.05
1941-2017	-0.37	-0.39	-0.36	0.74	0.71	0.77	0.09	0.04	0.14

A-D	p-values	
years	before	after
1941-1981	0.04	0.05
1982-2017	0.034	0.05
1941-2017	6.21e-4	8.68e-4

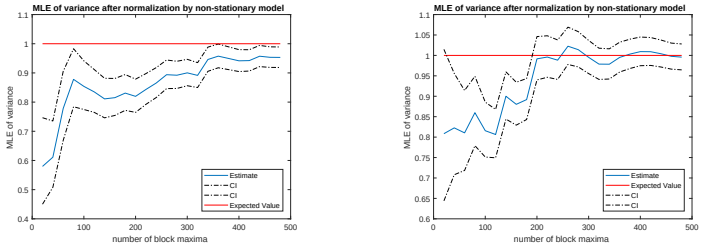


(a) cluster(1) (b) cluster(2)

(c) cluster(3) (d) cluster(4)

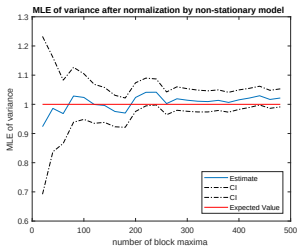
Figure 6: Mean maximum likelihood estimates for combined gev model with number of block maxima contribution per time series. Red line represents theoretical value under the assumption of a perfect time dependent model.

Cluster (4) returns poor results throughout this analysis. Though the model for the mean provides similar convergence results to that of the previous clusters the normalized variance seems to converge to a value less than 1. Station 14 may be contributing to the deviation of the maximum likelihood estimates and goodness of fit results. The time

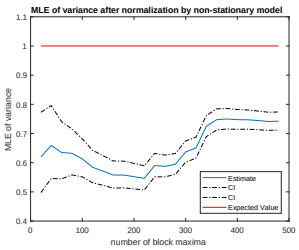


(a) cluster(1)

(b) cluster(2)



(c) cluster(3)



(d) cluster(4)

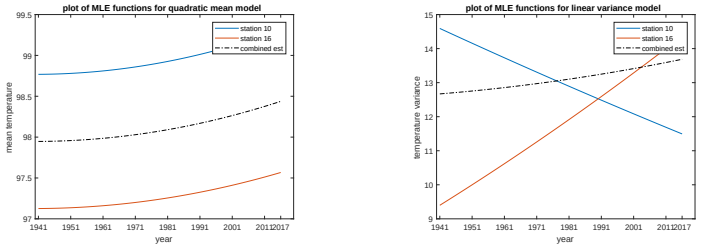
Figure 7: Variance maximum likelihood estimates for combined gev model with number of block maxima contribution per time series. Red line represents theoretical value under the assumption of a perfect time dependent model.

dependent variance models from the individual time series for station 14 suggest large differences from the other stations in the cluster which most likely affected the model for variance of the cluster. Station 14 was removed from cluster (4) and A-D goodness of fit and MLE parameter convergence was reevaluated. Convergence of the estimate $\hat{\sigma} \to 1$ as the number of block maxima tends to infinity. Goodness of fit results conclude the block maxima come from the GEV distribution given by the parameters for MLE at the $\alpha = 0.05$ significance level.

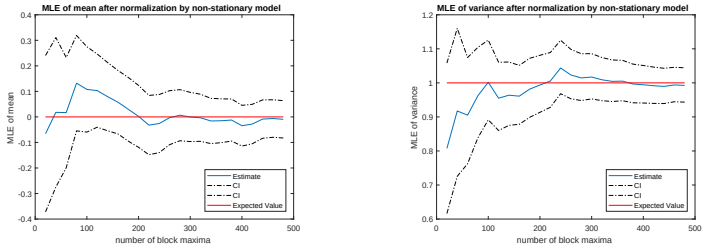
cluster(4) without station 14: MLE of normalized GEV

params	k			σ			μ		
years	value	CI		value	CI		value	CI	
1941-1981	-0.26	-0.31	-0.21	0.95	0.88	1.02	0.01	-0.09	0.11
1982-2017	-0.25	-0.30	-0.20	1.04	0.97	1.13	-0.03	-0.14	-0.09
1941-2017	-0.25	-0.28	-0.22	0.99	0.94	1.04	-0.01	-0.08	0.06

A-D	p-values	
years	before	after
1941-1981	0.04	0.34
1982-2017	0.03	0.50
1941-2017	6.21e-4	0.50



(a) mean models (b) variance models



(c) mean (d) variance

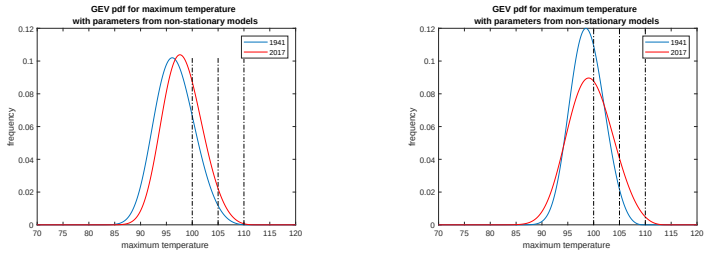
Figure 8: Cluster(4) without station 14: (a),(b) mean and variance models of block maxima for individual and combined time series. (c),(d) mean and variance maximum likelihood estimates for combined gev model with number of block maxima contribution per time series. Red line represents theoretical value under the assumption of a perfect time dependent model.

3.4 Discussion

Results throughout this analysis reflect a need for non-stationary GEV distribution modeling for the maxima of temperature values for every cluster of stations. Particularly interesting are the differences in GEV fit between 1941-1981 and 1982-2017 which suggest significant differences in the mean and variance parameters. We note that the most prominent example of monotonic trend in variance occurs in cluster (2) while the most prominent example of monotonic trend in mean occurs in cluster (1). Such monotonic trends suggest that the time dependence of the parameters is not the result of standard temperature cycles for all clusters.

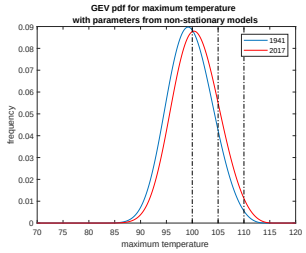
GEV distributions were generated for 1941 and 2017 based on parameter modeling [5]. In general, if variance parameters of the cluster change a significant amount only a small amount of change is seen in the mean (see clusters (2) and (4)). Conversely, if a small amount of change is seen in the variance a significant amount of change is seen in the mean (see clusters (1) and (3)). A comparison of GEV distributional fits for 1941 and 2017 reflect an increase in the probablity for right-hand (higher) temperature extremes for all clusters particularly for temperature values greater than 100 degrees farenheit. In

[5]Since the shape parameter of the distributions is assumed to be stationary the shape parameter estimated by MLE for 1941-2017 was used to generate the GEV pdfs for 1941 and 2017.

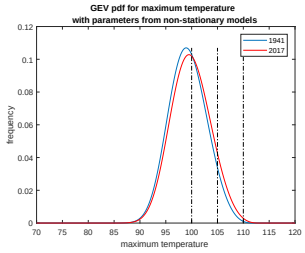


(a) cluster(1)

(b) cluster(2)



(c) cluster(3)



(d) cluster(4)

Figure 9: Non-stationary generalized extreme value pdfs. Cluster(4) distributions exclude station 14.

fact, we see that the probability of occurence in 2017 is double[6] that of 1941.

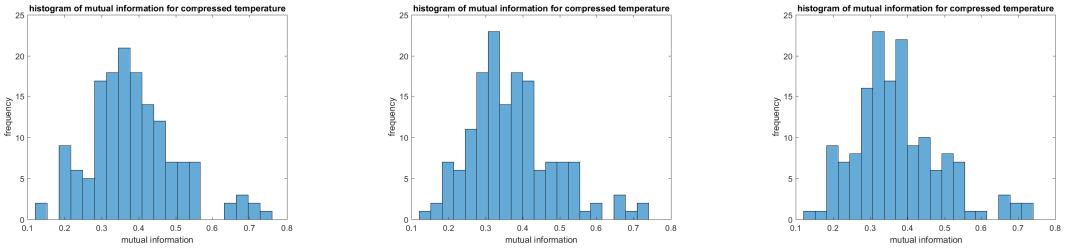
In the future it may be interesting to evaluate the data for long-term temperature extremes such as in [9][14] or look for early prediction modeling by examining the trajectories which result in a temperature extreme. Precipitation and how it relates to temperature extremes (such as in [12]) across Texas may also be of interest particularly in the case of unprecendented flooding as seen in Houston, Texas, in August 2017.

values	μ		σ		$P(X \geq 100)$		$P(X \geq 105)$		$P(X \geq 110)$	
cluster	1941	2017	1941	2017	1941	2017	1941	2017	1941	2017
(1)	95.18	96.70	13.76	13.28	0.07	0.09	0.01	0.02	1.00e-4	7.00e-4
(2)	97.58	97.81	10.13	18.13	0.11	0.09	0.02	0.04	0	4.80e-3
(3)	97.98	99.09	18.07	18.84	0.09	0.09	0.04	0.06	5.50e-3	1.13e-2
(4)	97.95	98.45	12.66	13.71	0.10	0.10	0.03	0.04	1.00e-3	2.90e-3

2017/1941	$P(X \geq 100)$	$P(X \geq 105)$	$P(X \geq 110)$
(1)	1.32	1.89	7.00
(2)	0.80	1.88	-
(3)	0.99	1.30	2.06
(4)	0.99	1.26	2.90

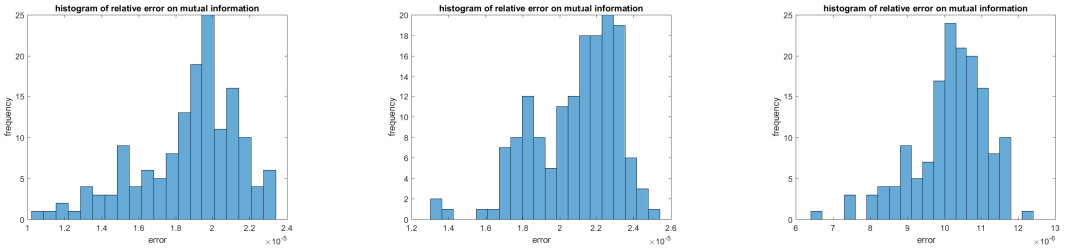
[6]Where this is seen in the tails depends on the location of the pdf.

A Figures



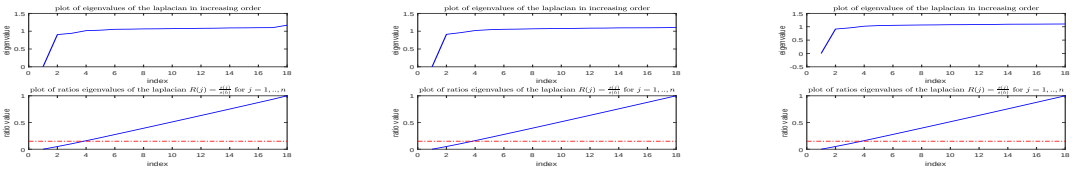
(a) 1941-1981 (b) 1982-2017 (c) 1982-2017

Figure A.1: Calculated maximized mutual information between stations.



(a) 1941-1981 (b) 1982-2017 (c) 1982-2017

Figure A.2: Calculated error on mutual information between stations after maximization.



(a) 1941-1981 (b) 1982-2017 (c) 1941-2017

Figure A.3: Eigenvalues and ratio cut-off for the normalized laplacian caluclated from the similarity matrix.

References

[1] L. Cheng, A. AghaKouchak, E. Gilleland, R.W., Katz. (2014). *Non-stationary Extreme Value Analysis in a Changing Climate*, Climatic Change. **2** 127.

[2] S. Coles (2001). *An Introduction to Statistical Modeling of Extreme Values*, Springer.

[3] J. Finkel, J. I. Katz (2017). *Changing US extreme temperature statistics*, Int. J. Climatol.,**37** 4749-4755.

[4] M. Gao, H. Zheng (2018). *Nonstationary extreme value analysis of temperature extremes in China*, Stochastic Environmental Research and Risk Assessment. **32** 5: 1299-1315.

[5] J. Hansen, M. Sato, R. Ruedy (2012). *Perception of climate change*, PNAS, **109** 37: E2415-E2423.

[6] H. Hasan, N. Radi, S. Kassim (2012). *Modeling of Extreme Temperature Using Generalized Extreme Value (GEV) Distribution: A Case Study of Penang*, Proceedings of the World Congress on Engineering. **1**

[7] T. Hastie, R. Tsibirani, J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference and Prediction* 2nd Edition, Springer.

[8] J. Kyselý (2002). *Probability Estimates of Extreme Temperature Events: Stochastic Modelling Approach vs. Extreme Value Distributions* Studia Geophysica et Geodaetica, **46** 1: 93-112.

[9] J. Kyselý (2002). *Temporal fluctuations in heat waves at Prague–Klementinum, the Czech Republic, from 1901–97, and their relationships to atmospheric circulation*, Int. J. Climatol., **22** 33-50.

[10] G. Meehl, C. Tebaldi, D. Adams-Smith (2016). *US daily temperature records past, present, and future*, PNAS, **113** 49: 13977-13982.

[11] NOAA National Centers for Environmental Information, *State of the Climate: Global Climate Report for July 2017 Mean Temp Anomalies*, published online August 2017, retrieved on September 19, 2018 from https://www.ncdc.noaa.gov/sotc/global/201703.

[12] R. Portmann, S. Solomon, G. Hegerl (2009). *Spatial and seasonal patterns in climate change, temperatures, and precipitation across the United States*, PNAS, **106** 18: 7324-7329.

[13] S. Rahmstorf, D. Coumou (2011). *Increase of extreme events in a warming world*, PNAS, **108** 44: 17905-17909.

[14] C. Schär, P. Vidale, D. Lüthi, C. Frei, C. Häberli, M. Liniger, C. Appenzeller (2004). *The role of increasing temperature variablility in European summer heatwaves*, Nature, **427** 332-336.

[15] G. Wergen, J. Krug (2010). *Record-breaking temperatures reveal a warming climate*, Europhysics Letters, **92** 3.

[16] X. Zhang, X. Yan (2014). *Spatiotemporal change in geographical distribution of global climate types in the context of climate warming*, Climate Dynamics. **43** 3-4: 595-605.

[17] J. Zscheischler, M. Mahecha, S. Harmeling (2012). *Climate Classifications: the Value of Unsupervised Clustering*, Procedia Computer Science, **9** 897-906.