Statistics and Applications {ISSN 2454-7395 (online)}
Special Issue in Memory of Prof. C R Rao
Volume 22, No. 3, 2024 (New Series), pp 639–666
http://www.ssca.org.in/journal

Identification of Changes in Temperature and Precipitation in Cities Across the Contiguous United States through High Dimensional Change Point Analysis

Abhishek Kaul[1], Alexandros Paparas[2], Venkata K. Jandhyala[1] and Stergios B. Fotopoulos[3]

[1]*Department of Mathematics and Statistics, Washington State University, USA*
[2]*Department of Information Systems and Business Analytics, Eastern Washington University, USA*
[3]*Department of Finance and Management Science, Washington State University, USA*

Received: 06 June 2024; Revised: 16 October 2024; Accepted: 19 October 2024

Abstract

In this article, we carry out a simultaneous study of changes in temperature and precipitation variables in several cities across the contiguous United States for the time period 1948-2023. Included among the seven climatic variables that we consider are both extremes as well as averages. The data on all the climatic variables in this article is sourced from Global Summary of the Year (GSOY), provided by the National Center for Environmental Information (NCEI) under the National Oceanic and Atmospheric Administration (NOAA). The main goal of the article is to simultaneously detect abrupt changes in the averages of the seven climatic variables by implementing recently developed high dimensional change point methodology. The methodology identifies three years, namely, 1957, 1989, and 2010 as the years of changes taking place in the climatic variables. Extensive follow up analysis is carried out to determine clusters of cities such that the nature of changes are similar within each cluster and differ significantly between the clusters. The clustering is done based upon the magnitudes of change computed for each change year at each city and for each climatic variable. The clusters enabled us to identify regions within US in which increases/decreases have occurred at any given change year. For example, temperatures were found to decrease in the change year 1957 and this decrease occurred predominantly in the northeastern, southeastern and southern regions of the United States. More comprehensive summary of our findings can be found in the discussion section of the article. Some plausible reasons for changes such as solar dimming and solar brightening were discussed in the concluding remarks section.

Key words: Temperature and precipitation; High-dimensional; Change points; Climatic extremes.

AMS Subject Classifications: 62F12; 62P12

Corresponding Author: Venkata K Jandhyala
Email: jandhyala@wsu.edu

1. Introduction

The study of climatic changes through important climatic variables is fundamental to a proper understanding of the prevailing climatic conditions as well as the conditions we can expect in the years ahead. Such an understanding will enable humans to adapt to changing conditions and plan for taking timely actions to prevent the onslaught of extreme climatic conditions. Among all climatic variables, temperature and precipitation are by far the most important variables for purposes of observation, analysis and understanding.

In this paper, our main goal is to study changes in temperature and precipitation at 91 stations spread throughout the contiguous United States. Changes in temperature and precipitation can be studied separately for the two variables or could also be studied together in a combined way. Also, changes could be studied in averages, extremes, thresholds and each of such studies brings its own understanding of the variables under study. In this article we consider seven variables, all representing one of precipitation or temperature from each of the cities considered in this analysis. Specifically, we have considered two precipitation and five temperature variables:

PRCP1 Frequency of days with precipitation exceeding one inch

PRCP Total annual precipitation

TMAX32 Instances of maximum temperature dropping below 32°F

TMAX90 Occurrences of maximum temperature surpassing 90°F

TAVG Average annual temperature

TMAX Average annual maximum temperature

TMIN Average annual minimum temperature

There is a large body of literature on climatic studies including those on temperature and precipitation. Among them, most of the existing studies on temperature focus on extremes only. For example, heat waves in 1995 and 1999 resulted in 739 and 110 excess deaths, respectively, in the city of Chicago alone. Based on regional climate model simulations (RCMs), Kunkel *et al.* (2010) predicted that there is a high probability of heat waves of unprecedented severity by the end of twenty first century if the high emissions path is followed. Oswald (2018) studied spatially continuous homogenized climate data to examine changes in regularity of heat waves including nighttime and daytime temperatures across the continental United States. The analysis showed prevalence of heatwaves between mid-70s and 2015. This was preceded by a decrease since 1948, the beginning of the dataset. Earlier Oswald and Rood (2014) studied extreme heat event days (EHEs) in the continental US based on daily maximum, minimum temperatures. The study period was 1930-2010 and results showed negative trends in the interior while positive trends showed in coastal and southern areas. While decreases occurred between 1930-1970, these decreases were offset by increases between 1970-2010. Gaffen and Ross (1998) examined trends in the frequency of days with anomalously high apparent temperatures (ATs) across the United States from 1949 to 1995 and observed that the annual frequency of extreme minimum ATs increased at the greatest number of stations, particularly in the eastern and western United States. Extending the data for the years 1949-2010, Grundstein and Dowd (2011) found that an increase in occurrence of 1-day extreme minimum ATs was particularly notable, especially in the eastern and western United States. Lee *et al.* (2014) examined monthly maximum and minimum temperatures from 932 stations located across the contigouous US for the years

1897-2010 and found estimated trend for monthly maximum had a mean of 0.47°C/Century while the estimated trend for monthly minimum had a mean of 1.65° C/Century.

Studies that focus on precipitation changes are equally important. Events of extreme precipitation are among the costliest of natural disasters. They are associated with flooding, damage to infrastructure, and loss of life. In the United States alone, extreme precipitation events have caused more than $200 billion damages during 1988-2017, with an increasing trend in costs as these events have become more frequent. In a recent study, Martinez-Villalobos and Neelin (2023) used a probability distribution for precipitation and predicted that about 13% of the globe and 25% of the tropics have displayed increases in extreme precipitation. While studying changes in extreme precipitation in the northeastern united states, Nazarian *et al.* (2022) found that extreme precipitation increased throughout the region with the largest changes seen in the summer. Implementing dynamically downscaled simulation, most recently Nazarian *et al.* (2024) predicted that both mean and extreme precipitation will increase to the east of the Sierra Madre highland and that extreme precipitation events can be expected to double throughout the region. Earlier, under a predicted 2°C of global warming, Rupp *et al.* (2022) found large variability in the magnitude of extreme precipitation across the western United States. Specifically, they found that majority of the region showed heavier tails for extreme precipitation under warming, while plateaus of eastern Oregon and Washington, and the crest of the Sierra Nevada, showed a lightening of tails. Armal *et al.* (2018) developed a Bayesian multilevel model using data from 1244 rainfall stations throughout the contiguous united states and found statistically significant trends in extreme rainfall frequency in 742 of the 1244 stations. These stations were predominantly in US Southeast and Northeast regions. Also, the trends in 274 out of the 742 stations can be attributed to El Nino Southern Oscillation, the North Atlantic Oscillation, the Pacific decadal oscillation, and the Atlantic multidecadal oscillation along with changes in global surface temperature anomalies. These 274 stations are mainly found in the U.S. northwest, west, and southwest climate regions.

There are several articles in the literature that study changes in both temperature and precipitation. It is important to review some of such studies as well to obtain a better understanding of changes in these two climatic factors that are interdependent. Robinson (2021) reviews the observational evidence for climate-driven increases in extremes most relevant to the continental United States. Wang *et al.* (2015) applied dynamical-statistical downscaling approach for studying climate change impacts at local scales. They applied the methodology for projecting future climate over the province of Ontario, Canada and found that there would be a significant warming trend throughout this century for the entire province while less precipitation is projected for most of the selected weather stations. Later, Zhou *et al.* (2018) predicted that there will be an increasing pattern of temperature and precipitation extremes over Canada over two time- slices (*i.e.*, 2046-2065 and 2076-2095). The effects of climate change and global warming on Alaska are unequivocal. From 1949 to 2012, the annual mean temperature increased 1.78°C and annual precipitation increased 3.1mm; winter changes were most dramatic, with temperatures climbing 3.78°C and precipitation increasing by 7.2mm (Bieniek *et al.* (2014)). Isaac and Van Wijngaarden (2012) analyzed hourly values of temperatures and relative humidity observed at 309 stations located across North America for the period 1948-2010. Trends were determined based on straight line fits and results showed significant warming trends in the mid western US, Canadian prairies, and western arctic. Lai and Dzombak (2019) analyzed time series of historical annual average temperature, total precipitation, and extreme weather indices for 103 (for temperature indices) and 115 (for precipitation indices) U.S. cities with climate records starting from as early as 1870. Applying linear regression modelling, Lai and Dzombak (2019) constructed 95% confidence intervals for the mean rate of change. The results showed increases in annual average temperature and precipitation although there were spatial and temporal variations. Cities in the Northeast and Midwest showed significant increases in precipitation while no increases in temperature in Southeast regions were found. Earlier, Griffiths and Bradley (2007) examined changes in five temperature and five precipitation extreme indicators from the northeast US for the period 1926-2000. Their empirical orthogonal function (EOF) analysis showed increases in both temperature and precipitation extremes. High correlation was found between number of frost days and warm nights and Atlantic Oscillation (AO).

The above short review of studies regarding temperature and precipitation changes within the contiguous United States makes it clear that such climatic studies must continue. Then only the scientific community will be able to have a proper understanding of the dynamic behavior of various climatic variables including temperature and precipitation. It is also clear from the above review that change point methodology ((Zhao and Chu, 2010; Wang *et al.*, 2010; Villarini *et al.*, 2013; Lee *et al.*, 2014)) is a powerful way for modeling changes in climatic variables. In this article, we shall adapt this frequently implemented method for capturing changes in temperature and precipitation variables within the contiguous United States over the period 1948-2023.

The change point methodology has long been a tool for climatologists for estimation of unknown time points at which abrupt changes might have occurred in one or more climatic variables. See, *e.g.*, Jandhyala *et al.* (2013); Beaulieu *et al.* (2012); Reeves *et al.* (2007), and the many references therein. Change point models for climatic data can be implemented individually in a univariate way for each city, or can also be implemented simultaneously for all cities in a multivariate way. Clearly, simultaneous modelling accounts for dependencies among cities that would otherwise have been ignored. Moreover, changes detected from a univariate analysis are with respect to the corresponding variance in the one the dimensional series. Whereas, a change recovered from a multivariate series measures the total change (in ℓ_2 magnitude) across all components with respect to the total variance across all components. This distinction highlights the main advantage of multivariate change point estimation, *i.e.*, it brings out systemic macro-level (in this case country-wide) temporal changes providing a more robust perspective on large scale climatic changes. In contrast changes that are recovered componentwise may be localized at a city or other regional level that may instead be indicative of localized weather variations instead of the large scale climate.

In recent years, change point methods have been developed for modeling and analyzing high dimensional data where parameter size is much larger than the sample size. These high dimensional change point methods enable the implementation of models that were previously considered intractable. While high dimensional change point methods have been applied for the analysis of financial data (Cai and Wang (2023)), socio-economic data (Kaul *et al.* (2019)), and mortality data (Chen *et al.* (2023)), there has not yet been application of this methodology for modeling and analyzing climatological data. In this paper, our goal is to carry out a comprehensive high dimensional modeling and analysis of temperature and precipitation data from cities across the contiguous United States. We shall first present a brief review of recent advances in high dimensional change point methods.

Fixed dimensional mean shift models and other variants have existed for several decades with well-known monographs being available, *e.g.*,Csorgo and Horváth (1997). The multivariate and high dimensional versions of these non-stationary models have seen significant recent research with an overwhelming proportion devoted to estimation methodologies for change points. A common thread of available methods is the use of general purpose algorithm's such as *binary segmentation* Venkatraman (1992), *wild binary segmentation* Fryzlewicz (2014) and *minimal partitioning via dynamic programming* Jackson *et al.* (2005). The first two work as extensions of single change point methods to multiple changes. From a methods perspective, the literature on estimation of change points under high dimensions can be forked into two general approaches, **(a).** Regularized cumulative sum (CUSUM) based recovery that is typically built on ℓ_1, ℓ_2 or ℓ_∞ aggregations of a cumulative sum metric. **(b).** Regularized M-estimation type recovery that is typically built upon a squared loss or a likelihood function. The former considered in (Enikeeva and Harchaoui, 2019; Jirak, 2015) which are based on an ℓ_2, ℓ_∞ aggregation, respectively. Other CUSUM based estimators include (Cho and Fryzlewicz, 2015; Cho, 2016; Wang and Samworth, 2018) amongst others, with the last allowing for sparsity of parameters and thus allowing for high dimensional means. Approach (b) is taken in (Wang *et al.*, 2020; Kaul *et al.*, 2021). Algorithmic advancements pertaining to minimal partitioning that is particularly critical for M-estimation type change point recovery is developed in Killick *et al.* (2012). Several other types of high dimensional change point models have also been studied in the recent literature, *e.g.*, linear regression, Bernoulli networks, graphical models, see, *e.g.*, (Kaul *et al.*, 2019, 2023; Lee *et al.*, 2016; Bhattacharjee *et al.*, 2020; Wang *et al.*, 2021) and several others. The problem of post-estimation inference on change points is a much lesser studied aspect in comparison to estimation alone, however some recent works have developed significant results under large data designs. Fundamental results under univariate $p = 1$ designs are available in *e.g.*, (Bai, 1994; Eichinger and Kirch, 2018; Cho and Kirch, 2022; Fotopoulos *et al.*, 2010). The case of diverging p is considered in Bhattacharjee *et al.* (2017). The article Kaul *et al.* (2021) which considers the high dimensional case, in a single change point setting ($N = 1$).

The article is organized as follows. Section 2 describes data analyzed in the article. Section 3 discusses published results on high dimensional change point methods that are utilized for the analysis in the paper, and Section 4 presents the implementation of high dimensional change point methods and their results. Section 5 is dedicated to a comprehensive discussion of the results and Section 6 ends the paper with some concluding remarks.

2. Temperature and precipitation data from contiguous United States

The data on temperature and precipitation variables is spread across the contiguous United States. It originates from the Global Summary of the Year (GSOY), provided by the National Center for Environmental Information (NCEI) under the National Oceanic and Atmospheric Administration (NOAA). It is available publicly and can be accessed from the NOAA GSOY Database. While the complete NCEI dataset is more comprehensive, we have meticulously collected data only on temperature and precipitation variables from 91 cities for the period 1948-2023 spread across the 48 contiguous states of the US. The dataset collected and analyzed in this article includes two precipitation and five temperature variables. Amongst these variables, three are discrete and the remaining are continuous. These include: PRCP1: # of days in a year with precipitation exceeding one inch, and PRCP: total annual precipitation measured in mm; and five temperature variables - TMAX32: # of instances in days of maximum temperature dropping below 32°F, TMAX90: # of occurrences in days of maximum temperature surpassing 90°F, TAVG: average annual temperature in °C computed by adding the unrounded monthly average temperatures and dividing by 12, TMAX: average annual maximum temperature in °C obtained as average of the mean monthly maximum temperatures, and TMIN: average annual minimum temperature in °C obtained as average of the mean monthly minimum temperatures. It may be noted that among the seven climatic variables, the variables PRCP1, TMAX32, TMAX90, TMAX, and TMIN represent extremes with PRCP1 being the only extreme variable for precipitation.

The collected data spanning years 1948-2023 demonstrates substantial diversity in temporal scope and geography. Thus, the selected cities ensure comprehensive coverage of various geographical regions and climatic conditions of the US. The dataset includes not only big metropolitan cities, but also rural areas surrounding these urban centers, offering a comprehensive representation of climatic conditions beyond the city limits. Along the East Coast, cities such as New York, Boston, and Philadelphia offer insights into the climatic nuances of the Northeast. Moving southward, vibrant urban centers like Atlanta, Miami, and New Orleans provide a glimpse into the subtropical climates of the Southeast. Across the Midwest, cities like Chicago, Minneapolis, and Kansas City showcase the variability of continental climates. In the Great Lakes region, cities such as Buffalo, Cleveland, and Milwaukee experience the moderating effects of the large bodies of water, influencing their climate patterns. On the West Coast, cities such as Los Angeles, San Francisco, and Seattle offer perspectives on the mild coastal climates of the Pacific. In the Southwest, cities like Phoenix, Las Vegas, and Albuquerque experience arid desert climates, while Denver and Salt Lake City experience the high-altitude conditions of the Rocky Mountains. Our dataset also includes cities in the Mountain West, Great Plains, and Pacific Northwest, providing a comprehensive understanding of climatic variations across the United States.

3. High dimensional methods for identifying change points in time series

We adopt a high dimensional multiple mean shift framework to model the considered climate data, specifically,

$$y_t = \sum_{j=1}^{N+1} \theta^0_{(j)} \mathbf{1}[\tau^0_{j-1} < t \le \tau^0_j] + \varepsilon_t, \qquad \text{for } t = 1, ..., T, \tag{1}$$

wherein $y_t = (y_{t1}, y_{t2}, ..., y_{tp})^T \in \mathbb{R}^p$ denotes the underlying temperature (5 variables for each city) and precipitation (2 variables for each city) variable across all considered cities.

There are 91 cities in the data set, resulting in $p = 628$ variables. The Model 1 assumes there are an unknown number $N \in \mathbb{N}^+ = \{1, 2, ...\}$ of change points in the underlying mean vectors $\theta^0_{(j)} \in \mathbb{R}^p$, $j = 1, ..., (N+1)$. , where their locations in the sampling period are denoted by $\tau^0 = (\tau^0_1, \tau^0_2, ..., \tau^0_N)^T \subseteq \{1, ..., T\}^N$. Our analysis to follow allows for spatial dependence across variables y_{tj}, $j = 1, ..., p$, *i.e.*, it allows for a dependence between temperature and precipitation variables as well as across cities. However, we assume temporal independence.

Remark 1: While the modelling structure adopted induces a large number of parameters, our chosen methodology is capable of allowing such high dimensionality as explained below.

A potential alternative for reduced modelling dimensions is to perform a coarser aggregation of cities into regional blocks (*e.g.*, North, Northeast, East, Midwest, South, Southeast, Southwest, West, and Northwest), however such an approach may lead to compromise on the post-hoc identification of the natural homogeneity of climatic changes amongst considered cities and instead force these to be on the chosen coarser grid.

We utilize the method and results of Kaul *et al.* (2021) for inference on change points. They proposed an iterative estimation procedure between squared loss based change point recovery and a ℓ_1-regularized squared loss recovery of mean estimates. While this article is developed under the assumption of a single change point $N = 1$, we utilize its natural extensions to multiple change points via the extensively chosen principle of binary segmentation.

The following provides a brief description of the methods and main results of Kaul *et al.* (2021) utilized here. Let $\bar{y} = \Big(\sum_{t=1}^{T} y_t/T\Big)$ and $x_t = (y_t - \bar{y})$, be the globally centered observations. Then under a single change point ($N = 1$), define a squared loss,

$$Q\Big(\tau, \theta\Big) = \sum_{t=1}^{\tau} \|x_t - \theta_{(1)}\|_2^2 + \sum_{t=\tau+1}^{T} \|x_t - \theta_{(2)}\|_2^2. \tag{2}$$

Additionally define ℓ_1 regularized mean estimated at any given τ as,

$$\hat{\theta}_{(j)}(\tau) = k_{\lambda_j}\Big(x_{(j)}(\tau)\Big), \quad j = 1, 2 \tag{3}$$

with $k_\lambda(x) = \mathrm{sign}(x)(|x| - \lambda)_+$, $\lambda > 0$, $x \in \mathbb{R}^p$, is the *soft-thresholding* operator, where $\mathrm{sign}(\cdot)$, $|\cdot|$, and $(\cdot)_+$[1] are applied component-wise. Then Algorithm 1 provides a twice-iterative method for recovery of the change point, where $\gamma > 0$ is a tuning parameter.

Next we briefly discuss properties of the estimator $\hat{\tau}$ that are relevant for our analysis. These properties assume suitable regularity conditions. Among the two most relevant ones, first we allow for spatial dependence with the underlying distribution being of a sub-exponential type (see *e.g.*, Vershynin (2019)). Next, given the high dimensional nature of the considered problem, an underlying sparsity of the mean parameters is also assumed. Further details are omitted here in view of simplicity of exposition.

The change point estimate from Algorithm 1 possesses desirable statistical properties in context of both estimation and inference, despite the underlying high dimensionality of mean parameters. To characterize the limiting distribution of the estimate $\tilde{\tau}$, we require the following negative drift two-sided random walk initializing at the origin,

$$\mathcal{C}_\infty(\zeta) = \begin{cases} \sum_{t=1}^{\zeta} z_t, & \zeta \in \mathbb{N}^+ \\ 0, & \zeta = 0 \\ \sum_{t=1}^{-\zeta} z_t^*, & \zeta \in \mathbb{N}^-, \end{cases} \tag{4}$$

Here z_t, z_t^* are independent copies of a normal distribution $\mathcal{N}(-\xi_\infty^2, 4\xi_\infty^2\sigma_\infty^2)$, which are also independent over all t. Here the parameters $\xi_\infty = \lim_{T\to\infty} \xi > 0$ and $\sigma_\infty^2 = \lim_{T\to\infty} \sigma^2$, where both ξ and σ^2 are as defined as follows,

$$\eta^0 = \Big(\theta_{(1)}^0 - \theta_{(2)}^0\Big), \quad \xi = \|\eta^0\|_2, \quad \text{and} \quad \sigma^2 = \eta^{0T}\Sigma\eta^0/\xi^2$$

[1]For $x \in \mathbb{R}$, $(x)_+ = x$, if $x \geq 0$, and $x = 0$ if $x < 0$.

Algorithm 1 (KFJS 2021): Estimation of τ^0 with boundary selection (under $N = 1$)

(Initialize): Select a preliminary evenly spaced coarse grid $\mathcal{D} \subset \{1, ..., T\}$ of cardinality $\log T$. Select an initializer $\check{\tau} \in \mathcal{D}$ as the best fitting value to the data $\left\{x_t\right\}_{t=1}^{T}$.

Step 1: Obtain estimates $\check{\theta}_{(j)} = \hat{\theta}_{(j)}(\check{\tau})$, $j = 1, 2$, and update change point estimates as

$$\hat{\tau} = \underset{\tau \in \{1,...,(T-1)\}}{\arg\min} Q(\tau, \check{\theta}),$$

and perform an ℓ_0 regularization as

$$\hat{\tau}^* = \begin{cases} T(\text{no change}) & \text{if } \{Q(T, \check{\theta}) - Q(\hat{\tau}, \check{\theta})\} < \gamma \\ \hat{\tau} & \text{else.} \end{cases}$$

Step 2: If $\hat{\tau}^* = T$ the set $\tilde{\tau} = T$, else if $\hat{\tau}^* > 0$, obtain estimates $\hat{\theta}_{(j)} = \hat{\theta}_{(j)}(\hat{\tau})$, $j = 1, 2$, and refit change point as,

$$\tilde{\tau} = \underset{\tau \in \{1,...,(T-1)\}}{\arg\min} Q(\tau, \hat{\theta}),$$

(Output): $\tilde{\tau}$

where, $\Sigma = E(\varepsilon_t \varepsilon_t^T)$ is the underlying covariance structure of Model 1. Finally, normality of the increments z_t in (4) is also a consequence of the normality assumption on the distribution underlying Model 1. Then, we have,

$$(\tilde{\tau} - \tau^0) \Rightarrow \underset{\zeta \in \mathbb{Z}}{\arg\max}\, \mathcal{C}_\infty(\zeta), \tag{5}$$

We utilize (5) to construct asymptotically valid confidence intervals for the change point parameters. Specifically, these are obtained as $[\tilde{\tau} - q_{(1-\alpha/2)}, \tilde{\tau} + q_{(1-\alpha/2)}]$ where $q_{(1-\alpha/2)}$ is the $(1 - \alpha/2)^{th}$ quantile of the considered arg max of a two sided random walk with a negative drift. Since no analytical form of this distribution is available, we obtain these quantiles via a monte-carlo simulation, *i.e.*, simulating the two-sided random walk process and in turn obtaining realizations from the distribution under consideration.

The above results are under a single change point assumption, whereas the model and data under consideration have multiple change points. For this extension, we adopt the fairly standard practice of implementing binary segmentation, *i.e.*, recursively split data into binary partitions until no further change points are observed. This process utilizes Algorithm 1 in each recursive step, however, this algorithm is implemented only upto the ℓ_0 regularization of Step 1 (stated as Algorithm 2). The entire process of estimating multiple change points is then stated as Algorithm 3 (KFJS+BS) below.

As suggested in Kaul *et al.* (2021) a further local refitting is carried out of the change point estimates (analog of Step 2 of Algorithm 1). Specifically, Let $\hat{\tau}$ and $\hat{\mathbb{N}}$ represent the location and number of change point estimates obtained from Algorithm 2 and $\hat{\theta}(\hat{\tau})$ represent

Algorithm 2 (KFJS 2021): Estimation of τ^0 with boundary selection (under $N = 1$)

(Initialize): Select a preliminary evenly spaced coarse grid $\mathcal{D} \subset \{1, ..., T\}$ of cardinality $\log T$. Select an initializer $\check{\tau} \in \mathcal{D}$ as the best fitting value to the data $\left\{x_t\right\}_{t=1}^{T}$.

Step 1: Obtain estimates $\check{\theta}_{(j)} = \hat{\theta}_{(j)}(\check{\tau})$, $j = 1, 2$, and update change point estimates as

$$\hat{\tau} = \underset{\tau \in \{1,...,(T-1)\}}{\arg\min}\, Q(\tau, \check{\theta}),$$

and perform an ℓ_0 regularization as

$$\hat{\tau}^* = \begin{cases} T(\text{no change}) & \text{if } \{Q(T, \check{\theta}) - Q(\hat{\tau}, \check{\theta})\} < \gamma \\ \hat{\tau} & \text{else.} \end{cases}$$

(Output): $\hat{\tau}^*$

Algorithm 3 (KJFS+BS): Extension of KJFS to multiple changes via binary segmentation

(Initialize): $\hat{\tau}_{\text{st}} = \phi$ collecting all change points to be estimated.
Implement $\hat{\tau}$= Algorithm 2 $\left(\{1, ..., T\}\right)$.
  If $\hat{\tau} = T$ (no change) then **Stop**
  Else $\hat{\tau}_{\text{up}} = (\tau_{\text{st}}, \hat{\tau})$ (updated vector of estimated change points)
While length($\hat{\tau}_{up}$) > length($\hat{\tau}_{st}$) **do**
  $\hat{\tau}_{st} = \hat{\tau}_{up}$
  for $m \in 1 : (\text{length}(\tau_{st}) + 1)$ **do**
  $\text{partition}_m = \{\tau_{st(m-1)}, ..., \tau_{st(m)}\}$
  $\hat{\tau} = \text{Algorithm 2}(\text{partition}_\text{m})$
  If $\hat{\tau}$ is away from boundary of sampling period of partition **then**
  $\hat{\tau}_{up} = (\hat{\tau}_{st}, \hat{\tau})$
(Output): all estimated change points of vector $\hat{\tau}_{up}$ sorted in ascending order.

the mean estimates obtained from the associated partitioning via 3. Further, let,

$$Q_j\Big(\tau_j, \tau_{-j}, \theta\Big) = \sum_{t=\tau_{j-1}+1}^{\tau_j} \|x_t - \theta_{(j)}\|_2^2 + \sum_{t=\tau_j+1}^{\tau_{j+1}} \|x_t - \theta_{(j+1)}\|_2^2. \tag{6}$$

Then define the locally refitted estimates as,

$$\tilde{\tau}_j := \tilde{\tau}_j\Big(\hat{\tau}_{-j}, \hat{\theta}\Big) = \underset{\hat{\tau}_{j-1} < \tau_j < \hat{\tau}_{j+1}}{\arg\min}\, Q_j\Big(\tau_j, \hat{\tau}_{-j}, \hat{\theta}\Big), \quad j = 1, ..., \hat{N} \tag{7}$$

Then confidence intervals for the parameters τ_j^0's are obtained by utilizing the change point estimates $\tilde{\tau}_j$ and by a piecewise application of (5).

4. Results of the high dimensional change point analysis

As described in Section 2, the data consists of two precipitation and five temperature variables collected from stations at 91 cities spread across the contiguous United States over the period 1948-2023. We implement the analysis with all the seven climatic variables in the model ($p = 628$). Here, it should be noted that the value of p is lower than what one would expect. This happened because some of the variables had no variability in their values throughout the sampling period, and hence such variables were removed from the analysis. All computations are carried out in the statistical software R.

Before we begin the presentation of results, we would like to bring to the attention of the reader the inadequacy of identifying merely the point estimates of unknown change points. The first step of the high dimensional change point analysis is the implementation of Algorithm 3, mainly for point estimation of change points in the climatic data. For each change point identified in the mean vector, one may only conclude that there is a change in at least one of the component climatic variables.

This leaves still the question of identifying further the actual climatic variables in which changes have occurred in their means. For this purpose we perform component-wise t-tests for a comparison of pre and post means across estimated change points. A further Bonferroni correction is made based on the number of tests performed in order to control the family wise error rate of the procedure. Based on this inferential procedure, we draw conclusions about changes in the climatic variables comprehensively.

Table 1: Estimated change points in years via the implementation of algorithm-3

Climatic variables (#)	Number of parameters (p)	Estimated change points
Temperature and precipitation (7)	628	1957, 1989, 2010

Table 2: Confidence interval for each of the three change points detected via Algorithm 3 together with estimated jump sizes (ξ) and estimated variances(σ^2_∞)

Estimated change point	95% confidence interval	Estimated jump size (ξ)	Estimated variance (σ^2_∞)
1957	(1955, 1959)	16.02	36.82
1989	(1986, 1992)	14.2	44.59
2010	(2009, 2011)	18.09	43.34

As for presentation of results, we begin with presenting in Table 1 the change points identified by Algorithm 3. The fitted model with all the seven temperature and precipitation variables consisted of three change points estimated in years as 1957, 1989, 2010. Confidence intervals for the true change points along with estimated jump size, and estimated variance are presented in Table 2. As described above, confidence intervals for change point in each of the component climatic variables enabled us to determine whether a change has occurred in that component variable or not. Upon applying this method at each of the three change points, we were able to determine the number of changes identified at each city and for each climatic variable. Results from this analysis are presented in Tables 3-5, and Tables 6-7. Specifically, Table 3 consists of list of cities that have undergone a change in their mean at the change year 1957 and the listing is made for each of the seven climatic variables PRCP1, PRCP, TMAX32, TMAX90, TAVG, TMAX, TMIN. Table 4 and Table 5 consist of similar listings of cities for change years 1989 and 2010, respectively. Tables 6 and 7 consist of the listing of all 91 cities together with the number of change points in each climatic variable for any given city.

Upon identifying changes, it is important to compute the magnitudes of change in each climatic variable at each city. The magnitudes of change enable us to clearly understand the nature and severity of changes in the climatic variables under consideration. Moving further, based upon the magnitudes of change, we can also identify clusters of cities so that different clusters may identify groups of cities with different magnitudes while maintaining similarity in changes within each cluster. Often such clusters of cities can be associated with a particular region, and such information is extremely important for interpreting changes in climatic studies. Among the plethora of clustering methods, K-means clustering stands out as a widely adopted technique for segmenting datasets into a predefined number of groups, denoted as 'k clusters'. Its primary objective is to categorize objects into clusters, maximizing intra-class similarity while minimizing inter-class dissimilarity. In the K-means approach, each cluster is characterized by a centroid, computed as the mean of points within the cluster. The process begins with specifying the desired number of clusters (k), followed by the random selection of k objects from the dataset to serve as initial centroids. Subsequently, each remaining object is assigned to the nearest centroid based on Euclidean distance, a step known as the 'cluster assignment' step. The algorithm then updates the mean value of each cluster, termed the 'centroid update' step, iteratively repeating these steps until convergence is attained. Convergence indicates stability, signifying that cluster assignments remain unchanged between successive iterations.

In this study, the K-means clustering method was implemented using the 'kmeans' function from the 'cluster' (Maechler *et al.* (2013)) and 'factoextra' (Kassambara and Mundt (2021)) packages in R. The clusters resulting from the K-means cluster analysis for each of the three change points together with a comment on the nature of each cluster are presented in Tables 8-10 The actual magnitudes of change in each cluster for each climatic variable are presented in Table 11.

5. Discussion of results

We shall begin our discussion with Tables 1-2 that identify the change points in years through the application of Algorithm 3 to data on temperature and precipitation variables. The change years for the model with all seven temperature and precipitation variables are 1957, 1989, and 2010. The 95% confidence intervals presented in Table 2 for each of the three true change years are very tight (at most +/-3 years), thus indicating the high precision with which the change years have been estimated. Further, we look at Tables 3-5 lists cities that have undergone a change, respectively, in the years 1957, 1989, and 2010, for each of the seven climatic variables. Focusing on the two precipitation variables PRCP1 and PRCP we notice that changes in PRCP1 occurred at 9 cities in 1957, at 6 cities in 1989, and at only 3 cities in 2010, whereas similar numbers for PRCP are 5, 8, and 1, respectively. Similar city

Table 3: List of cities that have undergone a change in the year 1957 for each of the seven climatic variables, namely, PRCP1, PRCP, TMAX32, TMAX90, TAVG, TMAX, TMIN

Variable	Cities
PRCP1	Baton Rouge, Brownsville, Columbia, Eugene, Milwaukee, New York City, Oklahoma City, Tallahassee, Wichita
PRCP	Albuquerque, Baton Rouge, New York City, Tallahassee, Wichita
TMAX32	Atlanta, Birmingham, Buffalo, Charleston, Charlotte, Chattanooga, Cincinnati, Cleveland, Columbus, Knoxville, Lexington, Louisville, Nashville, New Orleans, Philadelphia, Pittsburgh, Richmond, Saint Louis, Wichita
TMAX90	Atlanta, Austin, Boise, Charlotte, Chattanooga, Cleveland, Columbia, Columbus, Jacksonville, Knoxville, Lexington, Macon, Montgomery, Portland OR, Raleigh, Sacramento, San Francisco, Seattle, Tucson, Wichita
TAVG	Albuquerque, Atlanta, Augusta, Bakersfield, Baton Rouge, Birmingham, Brownsville, Buffalo, Burlington, Charlotte, Chattanooga, Cincinnati, Cleveland, Columbia, Columbus, Detroit, El Paso, Fresno, Greensboro, Jacksonville, Knoxville, Lexington, Little Rock, Los Angeles, Louisville, Macon, Montgomery, Nashville, New Orleans, Philadelphia, Phoenix, Pittsburgh, Portland OR, Raleigh, Reno, Richmond, Sacramento, San Antonio, San Diego, San Francisco, Seattle, Tallahassee, Virginia Beach, Wichita
TMAX	Atlanta, Augusta, Austin, Baton Rouge, Birmingham, Buffalo, Burlington, Charlotte, Chattanooga, Cleveland, Columbia, Columbus, Detroit, Greensboro, Houston, Jacksonville, Knoxville, Lexington, Los Angeles, Louisville, Macon, Miami, Mobile, Montgomery, Nashville, New Orleans, Philadelphia, Pittsburgh, Portland OR, Raleigh, Sacramento, San Antonio, San Francisco, Seattle, Virginia Beach, Wichita
TMIN	Albuquerque, Atlanta, Augusta, Bakersfield, Baton Rouge, Birmingham, Brownsville, Burlington, Charlotte, Chattanooga, Cincinnati, Cleveland, Columbia, Dayton, Detroit, El Paso, Eugene, Fresno, Jacksonville, Knoxville, Las Vegas, Little Rock, Los Angeles, Macon, Madison, Miami, Nashville, New Orleans, Philadelphia, Phoenix, Pittsburgh, Portland, Raleigh, Reno, Richmond, Sacramento, San Diego, San Francisco, Seattle, Tallahassee, Virginia Beach, Wichita

Table 4: List of cities that have undergone a change in the year 1989 for each of the seven climatic variables, namely, PRCP1, PRCP, TMAX32, TMAX90, TAVG, TMAX, TMIN

Variable	Cities
PRCP1	Albany, Dayton, Fort Wayne, Knoxville, Macon, Tallahassee
PRCP	Albany, Concord, Dayton, Fargo, Fort Wayne, Madison, Rochester, Tallahassee
TMAX32	Albany, Albuquerque, Allentown, Amarillo, Atlanta, Augusta, Birmingham, Burlington, Charleston, Charlotte, Chattanooga, Cleveland, Colorado Springs, Columbia, Columbus, Dallas, Greensboro, Harrisburg, Indianapolis, Little Rock, Louisville, Macon, Memphis, Milwaukee, Mobile, Nashville, New York City, New Orleans, Oklahoma, Philadelphia, Pittsburgh, Providence, Raleigh, Richmond, Saint Louis, Tulsa, Virginia Beach, Wichita
TMAX90	Austin, Boise, Brownsville, Fargo, Miami, New Orleans, Raleigh, San Diego, Sioux Falls, Tallahassee, Tucson
TAVG	Albany, Albuquerque, Allentown, Amarillo, Atlanta, Augusta, Austin, Baton Rouge, Birmingham, Boise, Boston, Brownsville, Buffalo, Burlington, Charleston, Charlotte, Chattanooga, Cheyenne, Chicago, Cincinnati, Cleveland, Columbia, Columbus, Concord, Dallas, Denver, Des Moines, Detroit, El Paso, Fargo, Fort Wayne, Fresno, Greensboro, Harrisburg, Hartford, Houston, Indianapolis, Knoxville, Las Vegas, Lexington, Little Rock, Louisville, Madison, Memphis, Miami, Milwaukee, Minneapolis, Montgomery, Nashville, New York City, New Orleans, Oklahoma City, Orlando, Philadelphia, Phoenix, Pittsburgh, Portland OR, Portland ME, Providence, Raleigh, Reno, Richmond, Saint Louis, Salt Lake, San Antonio, Springfield, Tallahassee, Tampa, Tucson, Tulsa, Virginia Beach, Washington DC, Wichita
TMAX	Allentown, Amarillo, Atlanta, Augusta, Austin, Baton Rouge, Birmingham, Boise, Brownsville, Buffalo, Burlington, Charleston, Charlotte, Chattanooga, Cleveland, Columbus, Concord, Dallas, Denver, Detroit, El Paso, Fort Wayne, Greensboro, Harrisburg, Indianapolis, Little Rock, Louisville, Memphis, Miami, Milwaukee, Montgomery, New Orleans, Oklahoma, Orlando, Philadelphia, Phoenix, Pittsburgh, Portland ME, Providence, Raleigh, Saint Louis, San Antonio, San Diego, San Francisco, Tallahassee, Tucson, Virginia Beach
TMIN	Albany, Albuquerque, Atlanta, Austin, Billings, Birmingham, Boise, Boston, Brownsville, Buffalo, Burlington, Charleston, Charlotte, Chattanooga, Cheyenne, Chicago, Cincinnati, Cleveland, Columbia, Columbus, Concord, Dallas, Dayton, Des Moines, Detroit, El Paso, Fargo, Fort Wayne, Fresno, Greensboro, Harrisburg, Hartford, Houston, Indianapolis, Knoxville, Las Vegas, Lexington, Little Rock, Louisville, Madison, Memphis, Miami, Milwaukee, Minneapolis, Nashville, New York City, New Orleans, Oklahoma, Omaha, Philadelphia, Phoenix, Pittsburgh, Portland OR, Portland ME, Providence, Raleigh, Reno, Richmond, Rochester, Saint Louis, Salt Lake, San Antonio, Seattle, Sioux Falls, Springfield, Tallahassee, Tampa, Tucson, Tulsa, Virginia Beach, Washington DC, Wichita

Table 5: List of cities that have undergone a change in the year 2010 for each of the seven climatic variables, namely, PRCP1, PRCP, TMAX32, TMAX90, TAVG, TMAX, TMIN

Variable	Cities
PRCP1	Baton Rouge, Cincinnati, Raleigh
PRCP	Eugene
TMAX32	None
TMAX90	Albuquerque, Amarillo, Austin, Bakersfield, Boise, Brownsville, Colorado Springs, Dallas, Denver, Des Moines, El Paso, Eugene, Houston, Miami, Nashville, Orlando, Reno, Saint Louis, Seattle, Wichita
TAVG	Albany, Albuquerque, Allentown, Amarillo, Atlanta, Augusta, Austin, Bakersfield, Baton Rouge, Birmingham, Boise, Boston, Brownsville, Burlington, Charleston, Charlotte, Chattanooga, Cincinnati, Cleveland, Colorado Springs, Columbia, Concord, Dallas, Dayton, El Paso, Fresno, Greensboro, Harrisburg, Hartford, Houston, Jacksonville, Knoxville, Las Vegas, Lexington, Louisville, Miami, Montgomery, Nashville, New York City, New Orleans, Omaha, Orlando, Philadelphia, Phoenix, Portland ME, Providence, Raleigh, Reno, Richmond, Rochester, Sacramento, Salt Lake, San Antonio, San Diego, Seattle, Spokane, Tallahassee, Tampa, Tucson, Virginia Beach, Washington DC, Wichita
TMAX	Albany, Albuquerque, Allentown, Amarillo, Atlanta, Augusta, Austin, Bakersfield, Boston, Brownsville, Burlington, Charleston, Charlotte, Cheyenne, Cleveland, Colorado Springs, Columbia, Dallas, Dayton, El Paso, Eugene, Fresno, Hartford, Houston, Jacksonville, Las Vegas, Lexington, Louisville, Macon, Miami, Montgomery, Nashville, New Orleans, Orlando, Phoenix, Portland ME, Reno, Rochester, Sacramento, Saint Louis, Salt Lake, San Diego, Seattle, Tallahassee, Tampa, Tucson, Virginia Beach, Washington DC, Wichita
TMIN	Albany, Albuquerque, Allentown, Amarillo, Atlanta, Augusta, Austin, Bakersfield, Baton Rouge, Birmingham, Boise, Boston, Brownsville, Buffalo, Burlington, Charleston, Chattanooga, Cincinnati, Cleveland, Colorado Springs, Columbia, Columbus, Concord, Dallas, Dayton, El Paso, Fresno, Greensboro, Harrisburg, Hartford, Houston, Jacksonville, Knoxville, Las Vegas, Louisville, Miami, Montgomery, Nashville, New York City, New Orleans, Omaha, Orlando, Philadelphia, Phoenix, Pittsburgh, Portland OR, Portland ME, Raleigh, Reno, Richmond, Rochester, Salt Lake, San Antonio, Seattle, Spokane, Tallahassee, Tampa, Tucson, Virginia Beach, Washington DC

count for temperature variables are: TMAX 32 – 19, 38, 0; TMAX90 – 20, 11, 20; TAVG – 44, 79, 63; TMAX – 36, 48, 49; and TMIN – 42, 72, 60, respectively. It is also informative to see the same numbers for each of the three change years. Thus the number of cities in which changes have occurred at each of the change years in climatic variables PRCP1, PRCP, TMAX32, TMAX90, TAVG, TMAX, and TMIN, respectively are: in 1957 – 9, 5, 19, 20, 44, 36, 42; in 1989 – 6, 8, 38, 11, 79, 48, 72; and in 2010 – 3, 1, 0, 20, 63, 49, 60.

Clearly, changes in temperature variables dominate the changes in precipitation variables. Also, changes in continuous variables (PRCP, TAVG, TMAX, TMIN) are significantly higher compared to changes in the three discrete variables (PRCP1, TMAX32, TMAX90). Perhaps this can be anticipated ahead because the information content in continuous variables is much more than that available in discrete variables and hence changes in continuous variables can be detected with higher precision. Among continuous temperature variables, changes in TAVG (44, 79, 63) and TMIN (42, 72, 60) are significantly higher compared to changes in TMAX (36, 48, 49). While the three variables had similar number of changes in

Table 6: List of cities along with corresponding number of change points for data on PRCP1, PRCP, TMAX32, TMAX90, TAVG, TMAX, and TMIN during the years 1948-2023

City	State	PRCP1	PRCP	TMAX32	TMAX90	TAVG	TMAX	TMIN
Albany	New York	1	1	1	0	2	1	2
Albuquerque	New Mexico	0	1	1	1	3	1	3
Allentown	Pennsylvania	0	0	1	0	2	2	1
Amarillo	Texas	0	0	1	1	2	2	1
Atlanta	Georgia	0	0	2	1	3	3	3
Augusta	Georgia	0	0	1	0	3	3	2
Austin	Texas	0	0	0	3	2	3	2
Bakersfield	California	0	0	0	1	2	1	2
Baton Rouge	Louisiana	2	1	0	0	3	2	2
Billings	Montana	0	0	0	0	0	0	1
Birmingham	Alabama	0	0	2	0	3	2	3
Boise	Idaho	0	0	0	3	2	1	2
Boston	Massachusetts	0	0	0	0	2	1	2
Brownsville	Texas	1	0	0	2	3	2	3
Buffalo	New York	0	0	1	0	2	2	2
Burlington	Vermont	0	0	1	0	3	3	3
Charleston	South Carolina	0	0	2	0	2	2	2
Charlotte	North Carolina	0	0	2	1	3	3	2
Chattanooga	Tennessee	0	0	2	1	3	2	3
Cheyenne	Wyoming	0	0	0	0	1	1	1
Chicago	Illinois	0	0	0	0	1	0	1
Cincinnati	Ohio	1	0	1	0	3	0	3
Cleveland	Ohio	0	0	2	1	3	3	3
Color. Spring	Colorado	0	0	1	1	1	1	1
Columbia	South Carolina	1	0	1	1	3	2	3
Columbus	Ohio	0	0	2	1	2	2	2
Concord	New Hampshire	0	1	0	0	2	1	2
Dallas	Texas	0	0	1	1	2	2	2
Dayton	Ohio	1	1	0	0	1	1	3
Denver	Colorado	0	0	0	1	1	1	0
Des Moines	Iowa	0	0	0	1	1	0	1
Detroit	Michigan	0	0	0	0	2	2	2
El Paso	Texas	0	0	0	1	3	2	3
Eugene	Oregon	1	1	0	1	0	1	1
Fargo	North Dakota	0	1	0	1	1	0	1
Fort Wayne	Indiana	1	1	0	0	1	1	1
Fresno	California	0	0	0	0	3	1	3
Greensboro	North Carolina	0	0	2	0	3	2	2
Harrisburg	Pennsylvania	0	0	2	0	2	1	2
Hartford	Connecticut	0	0	0	0	2	1	2
Houston	Texas	0	0	0	1	2	2	2
Indianapolis	Indiana	0	0	1	0	1	1	1
Jacksonville	Florida	0	0	0	1	2	2	2
Kansas City	Kansas	0	0	0	0	0	0	0
Knoxville	Tennessee	1	0	1	1	3	1	3

Table 7: List of cities along with corresponding number of change points for data on PRCP1, PRCP, TMAX32, TMAX90, TAVG, TMAX, and TMIN during the years 1948-2023. (Continued).

City	State	PRCP1	PRCP	TMAX32	TMAX90	TAVG	TMAX	TMIN
Las Vegas	Nevada	0	0	0	0	2	1	3
Lexington	Kentucky	0	0	1	1	3	2	1
Little Rock	Arkansas	0	0	1	0	2	1	2
Los Angeles	California	0	0	0	0	1	1	1
Louisville	Kentucky	0	0	2	0	3	3	2
Macon	Georgia	1	0	1	1	1	2	1
Madison	Wisconsin	0	1	0	0	1	0	2
Memphis	Tennessee	0	0	1	0	1	1	1
Miami	Florida	0	0	0	2	2	3	3
Milwaukee	Wisconsin	1	0	1	0	1	1	1
Minneapolis	Minnesota	0	0	0	0	1	0	1
Mobile	Alabama	0	0	1	0	0	1	0
Montgomery	Alabama	0	0	0	1	3	3	1
Nashville	Tennessee	0	0	2	1	3	2	3
New York City	New York	1	1	1	0	2	0	2
New Orleans	Louisiana	0	0	2	2	3	3	3
Oklahoma City	Oklahoma	1	0	1	0	1	1	1
Omaha	Nebraska	0	0	0	0	1	0	2
Orlando	Florida	0	0	0	1	2	2	1
Philadelphia	Pennsylvania	0	0	2	0	3	2	3
Phoenix	Arizona	0	0	0	0	3	2	3
Pittsburgh	Pennsylvania	0	0	2	0	2	2	3
Portland	Oregon	0	0	0	1	2	1	3
Portland	Maine	0	0	0	0	2	2	2
Providence	Rhode Island	0	0	1	0	2	1	1
Raleigh	North Carolina	1	0	1	2	3	2	3
Reno	Nevada	0	0	0	1	3	1	3
Richmond	Virginia	0	0	2	0	3	0	3
Rochester	New York	0	1	0	0	1	1	2
Sacramento	California	0	0	0	1	2	2	1
Saint Louis	Missouri	0	0	2	1	1	2	1
Salt Lake City	Utah	0	0	0	0	2	1	2
San Antonio	Texas	0	0	0	0	3	2	2
San Diego	California	0	0	0	1	2	2	1
San Francisco	California	0	0	0	1	1	2	1
Seattle	Washington	0	0	0	2	2	2	3
Sioux Falls	South Dakota	0	0	0	1	0	0	1
Spokane	Washington	0	0	0	0	1	0	1
Springfield	Missouri	0	0	0	0	1	0	1
Tallahassee	Florida	2	2	0	1	3	2	3
Tampa	Florida	0	0	0	0	2	1	2
Tucson	Arizona	0	0	0	2	2	2	2
Tulsa	Oklahoma	0	0	1	0	1	0	1
Virginia Beach	Virginia	0	0	1	0	3	3	3
Washington	DC	0	0	0	0	2	1	2
Wichita	Kansas	1	1	2	2	3	2	2

Table 8: Clusters based on KMEANS clustering algorithm implemented upon actual differences in averages of the climatic variables before and after the change point in the year 1957.

Cluster	Cities	Cluster characteristic
1	Albany, Allentown, Amarillo, Billings, Boise, Boston, Charleston, Cheyenne, Chicago, Colorado Springs, Concord, Dallas, Dayton, Denver, Des Moines, Eugene, Fargo, Fort Wayne, Hartford, Houston, Indianapolis, Kansas City, Las Vegas, Madison, Memphis, Miami, Milwaukee, Minneapolis, Mobile, Oklahoma City, Omaha, Orlando, Portland ME, Providence, Rochester, Salt Lake City, Sioux Falls, Spokane, Springfield, Tampa, Tucson, Tulsa, Washington DC	No big changes in climatic variables
2	Bakersfield, Fresno, Los Angeles, Phoenix, Portland OR, Reno, Sacramento, San Diego, San Francisco, Seattle	High increases in TMAX90, TAVG, TMAX, and TMIN
3	Atlanta, Austin, Charlotte, Chattanooga, Columbia, Jacksonville, Knoxville, Lexington, Macon, Montgomery, Raleigh	Decrease in TMAX90, TAVG, TMAX, TMIN and increase in TMAX32, PRCP1
4	Albuquerque, Augusta, Birmingham, Brownsville, Burlington, Detroit, El Paso, Greensboro, Little Rock, Nashville, New Orleans, Richmond, San Antonio, Virginia Beach	Small increases in TMAX32, PRCP and small decreases in TAVG, TMAX, TMIN
5	Baton Rouge, New York City, Tallahassee, Wichita	High increases in PRCP1 and PRCP
6	Buffalo, Cincinnati, Cleveland, Columbus, Harrisburg, Louisville, Philadelphia, Pittsburgh, Saint Louis	Big increases in TMAX32, and Cecreases in TMAX90, TAVG, TMAX, TMIN

Table 9: Clusters based on KMEANS clustering algorithm implemented upon actual differences in averages of the climatic variables before and after the change point in the year 1989.

Cluster	Cities	Cluster characteristic
1	Tallahassee	High decrease in PRCP1 and PRCP
2	Albuquerque, Atlanta, Birmingham, Boise, Buffalo, Charleston, Charlotte, Chicago, Columbia, Dallas, Detroit, El Paso, Fresno, Greensboro, Hartford, Houston, Las Vegas, Little Rock, Memphis, Minneapolis, Nashville, Oklahoma City, Phoenix, Portland OR, Portland ME, Reno, Richmond, Salt Lake City, San Antonio, Tampa, Virginia Beach, Wichita	High increase in TMIN
3	Albany, Concord, Dayton, Fargo, Fort Wayne, Knoxville, Madison	High increases in PRCP1 and PRCP
4	Amarillo, Augusta, Bakersfield, Baton Rouge, Billings, Boston, Cheyenne, Cincinnati, Colorado Springs, Denver, Des Moines, Eugene, Jacksonville, Kansas City, Lexington, Los Angeles, Macon, Mobile, Montgomery, New York City, Omaha, Orlando, Rochester, Sacramento, San Diego, San Francisco, Seattle, Sioux Falls, Spokane, Springfield, Tulsa, Washington DC	No big changes in any of the variables
5	Allentown, Burlington, Chattanooga, Cleveland, Columbus, Harrisburg, Indianapolis, Louisville, Milwaukee, Philadelphia, Pittsburgh, Providence, Saint Louis	Decrease in TMAX32, and an increase in TAVG and TMIN
6	Austin, Brownsville, Miami, New Orleans, Raleigh, Tucson	Increase in TMAX90, TAVG, TMAX, TMIN

Table 10: Clusters based on KMEANS clustering algorithm implemented upon actual differences in averages of the climatic variables before and after the change point in the year 2010.

Cluster	Cities	Cluster characteristic
1	Albany, Allentown, Atlanta, Augusta, Boston, Burlington, Charleston, Charlotte, Cleveland, Columbia, Dayton, Fresno, Hartford, Jacksonville, Las Vegas, Lexington, Louisville, Montgomery, New Orleans, Phoenix, Portland ME, Rochester, Sacramento, Salt Lake City, San Diego, Seattle, Tallahassee, Tampa, Tucson, Virginia Beach, Washington DC	Increase in TAVG, TMAX, TMIN
2	Billings, Buffalo, Cheyenne, Chicago, Columbus, Denver, Des Moines, Detroit, Fargo, Fort Wayne, Indianapolis, Kansas City, Little Rock, Los Angeles, Macon, Madison, Memphis, Milwaukee, Minneapolis, Mobile, Oklahoma City, Pittsburgh, Portland OR, Providence, San Francisco, Sioux Falls, Springfield, Tulsa	No significant changes
3	Albuquerque, Amarillo, Austin, Bakersfield, Brownsville, Colorado Springs, Dallas, El Paso, Houston, Miami, Nashville, Orlando, Reno, Saint Louis, Wichita	High increases in TMAX90 and TMAX
4	Birmingham, Boise, Chattanooga, Concord, Greensboro, Harrisburg, Knoxville, New York City, Omaha, Philadelphia, Richmond, San Antonio, Spokane	Increases in TMAX90, TAVG, TMIN
5	Baton Rouge, Cincinnati, Raleigh	Increase in PRCP1
6	Eugene	High increase in TMAX90 and high decrease in PRCP

Table 11: Magnitudes of change for clusters in each change year and for each of the climatic variables representing temperature and precipitation.

Change Year	Cluster	PRCP1	PRCP	TMAX32	TMAX90	TAVG	TMAX	TMIN
		(days)	(mm)	(days)	(days)	$^\circ$C	$^\circ$C	$^\circ$C
1957	1	0.210	0.000	0.011	-0.135	0.000	-0.034	0.013
	2	0	0	0	2.323	0.849	0.403	1.132
	3	0.406	0	1.758	-18.384	-0.749	-0.954	-0.568
	4	0.18	4.653	0.991	0	-0.669	-0.487	-0.692
	5	4.639	254.368	1.366	-5.02	-0.556	-0.459	-0.613
	6	0	0	11.221	-3.03	-0.646	-0.676	-0.402
1989	1	-4.403	-198.544	0.000	14.092	0.597	0.652	0.542
	2	0	0	-1.319	0.226	0.839	0.324	1.177
	3	1.579	111.116	-1.084	-0.654	0.622	0.173	0.912
	4	0.066	2.569	-0.705	-0.398	0.219	0.119	0.266
	5	0	0	-8.491	0	0.967	0.799	1.117
	6	0	0	-0.438	20.025	0.894	0.866	0.921
2010	1	0	0	0	0.108	0.87	0.817	0.86
	2	0	0	0	0.879	0.019	0.054	0.095
	3	0	0	0	16.226	0.824	0.954	0.735
	4	0	0	0	0.656	0.728	0	0.861
	5	3.117	0	0	0	0.716	0	0.82
	6	0	-226.546	0	8.399	0	0.728	0

Table 12: Summary of observations made about changes in climatic variables that occurred in 1957.

Climatic Variable	Cluster	Region	Increase/Decrease
PRCP1	3 5	Southeastern (3) Eastern half (5)	Increase (3) High increase (5)
PRCP	4 5	South-southeastern (4) Eastern half (5)	Small increase (4) High increase (5)
TMAX32	3 4 6	Southeastern (3) South-southeastern (4) Northeastern (6)	Increase (3) Small increase (4) Big increase
TMAX90	2 3 6	West coast (2) Southeastern (3) Northeastern (6)	High increase (2) Decrease (3) Decrease (6)
TAVG	2 3 4 6	West coast (2) Southeastern (3) South-southeastern (4) Northeastern (6)	High increase (2) Decrease (3) small decrease (4) Decrease (6)
TMAX	2 3 4 6	West coast (2) Southeastern (3) South-southeastern (4) Northeastern (6)	High increase (2) Decrease (3) Small decrease (4) Decrease (6)
TMIN	2 3 6	West coast (2) Southeastern (3) Northeastern (6)	High increase (2) Decrease (3) Decrease (6)

Table 13: Summary of observations made about changes in climatic variables that occurred in 1989.

Climatic Variable	Cluster	Region	Increase/Decrease
PRCP1	1 3	Tallahassee (1) Eastern (3)	High decrease (1) High increase (3)
PRCP	1 3	Tallahassee (1) Eastern (3)	High decrease (1) High increase (3)
TMAX32	5	Northeastern (5)	Decrease (5)
TMAX90	6	Southern (6)	Increase (6)
TAVG	5 6	Northeastern (5) Southern (6)	Increase (5) Increase (6)
TMAX	6	Southern (6)	Increase (6)
TMIN	2 5 6	Throughout (2) Northeastern (5) Southern (6)	High increase (2) Increase (5) Increase (6)

Table 14: Summary of observations made about changes in climatic variables that occurred in 2010.

Climatic Variable	Cluster	Region	Increase/Decrease
PRCP1	5	Baton Rouge, Cincinnati, Raleigh (5)	Increase (5)
PRCP	6	Eugene (6)	High decrease (6)
TMAX32	–	——	——
TMAX90	3 4 6	Central (3) Eastern (4) Eugene (6)	High increase (3) Increase (4) High increase (6)
TAVG	1 4	Eastern or Western (1) Eastern (4)	Increase (1) Increase (4)
TMAX	1 3	Eastern or Western (1) Central (3)	Increase (1) High increase (3)
TMIN	1 4	Eastern or Western (1) Eastern (4)	Increase (1) Increase (4)

Table 15: Region wise representation of changes in temperature and precipitation variables

Region	Temperature variables	Year of change	Increase/decrease
Northeastern	TMAX32 TMAX, TAVG, TMAX, TMIN, TMAX32 TAVG, TMIN, TAVG	1957 1957, 1989 1989, 2010	High increase Decrease Increase
Eastern	PRCP1, PRCP PRCP TMAX90, TAVG, TMAX, TMIN	1957, 1989 1957 2010	High increase Small increase Increase
Southeastern	PRCP1, PRCP, TMAX32 TMAX90, TAVG, TMAX, TMIN	1957 1957	Increase Decrease
Southern	PRCP TAVG, TMAX TMAX90, TAVG, TMAX, TMIN	1957 1957 1989	Small increase Small decrease Increase
Central	TMAX90 TMAX	2010 2010	High increase Increase
West Coast	TMAX90, TAVG, TMAX, TMIN TAVG, TMAX, TMIN	1957 2010	High increase Increase
Throughout	TMIN	1989	High increase

1957, TAVG and TMIN had much higher number of cities that changed in 1989 and 2010 compared to number of cities that TMAX has changed in the same two change years. The same can be observed from the number of change points in each of the climatic variables at each of the 91 cities. This phenomenon should be understood with a deeper understanding of how higher extreme temperatures change compared to average and lower extreme temperature changes.

We shall now discuss results from cluster analysis based on magnitudes of change presented in Tables 8-10 and Table 11. There are six clusters in the change year 1957 (Table 8), and the magnitudes of change for these six clusters are presented in Table 11. Clearly, there are identifiable differences between the clusters. Cities in Cluster 2 (Bakersfield, Fresno, Los Angeles, Phoenix, Portland OR, Reno, Sacramento, San Diego, San Francisco, Seattle; darker orange) belonging to the west coastal region of the US have shown high increases in TMAX90 (2.323 days), TAVG (0.849°C), TMAX (0.403°C), and TMIN (1.132°C). All cities in Cluster 3 (Atlanta, Austin, Charlotte, Chattanooga, Columbia, Jacksonville, Knoxville, Lexington, Macon, Montgomery, Raleigh) belong to the southeastern region, and these cities have shown increased average change in PRCP1 (0.406 mm), TMAX32 (1.758 days), and significantly decreased changes in TMAX90 (-18.384 days), TAVG (0.749°C), TMAX (0.954°C), and TMIN (0.568°C). Cities in Cluster 6 (Buffalo, Cincinnati, Cleveland, Columbus, Harrisburg, Louisville, Philadelphia, Pittsburgh, Saint Louis) are all in the northeastern region, and cities in this cluster have very high average increase in TMAX32 (11.221 days) and decreases in TAVG (0.646°C), TMAX (0.676°C), and TMIN (0.402°C). Cities in Cluster 4 (Albuquerque, Augusta, Birmingham, Brownsville, Burlington, Detroit, El Paso, Greensboro, Little Rock, Nashville, New Orleans, Richmond, San Antonio, Virginia Beach) are mostly seen in south-southeastern parts of the US and these cities have experienced small increases in TMAX32 (0.991 days), PRCP (4.653 mm), and small decreases in TAVG (0.669°C), TMAX (0.487°C), and TMIN (0.692°C). Cluster 5 (Baton Rouge, New York City, Tallahassee, Wichita) has only four cities in it and these cities are located only in the eastern half of the US map and these cluster of cities may be characterized to show high increases in PRCP1 (4.639 days) and PRCP (254.368 mm). Finally, Cluster 1 (rest of the cities), which as most number of cities these cities have no significant changes, and are all spread evenly throughout the US. Overall, it is clear from Table 8 that there have been more decreasing tends in the temperature variables, and thus the change year 1957 can be viewed as indicative of the beginning of a cooling period. It is also worth noting the very large increase of 254.368 mm of precipitation in PRCP at cluster 5 cities.

Moving on to change year 1989, there are again six clusters in this change year as well (Table 9, Table 11). Among these, cities in Cluster 3 (Albany, Concord, Dayton, Fargo, Fort Wayne, Knoxville, Madison) are spread in the eastern part of the US, cities in Cluster 5 (Allentown, Burlington, Chattanooga, Cleveland, Columbus, Harrisburg, Indianapolis, Louisville, Milwaukee, Philadelphia, Pittsburgh, Providence, Saint Louis) are all clustered in northeastern part of the US, and cities in Cluster 6 (Austin, Brownsville, Miami, New Orleans, Raleigh, Tucson) are all lined up in the southern part of the US. Among the remaining two clusters, Cluster 1 has only one city (Tallahassee) with high decrease in PRCP1(-4.403 days) and PRCP (-198.544 mm), and cluster 2 consisting of large number of cities can be characterized as having large increase in TMIN (1.177°C). Cluster 4 has the largest number of cities and the cities in this cluster show no significant change in their averages. Cities in Cluster 3 have high increases in PRCP1 (1.579 days) and PRCP

(111.116 days); Custer 5 cities show a decrease in TMAX32 (-8.491°C), and an increase in TAVG (0.967) and TMIN (1.117), and cluster 6 cities showed increase in TMAX90 (20.025 days), TAVG (0.894°C), TMAX (0.866°C), and TMIN (0.921°C). Overall, Table 11 makes it clear that the magnitudes of change in this cluster are mostly positive, particularly for temperature variables and thus the change year 1989 can be seen as ending the cooling period that began in 1957 and that there is a transition into the beginning of warmer periods.

Among clusters in change year 2010 (Table 10), Eugene , OR identifies itself as Cluster 6. This city on the west coast can be identified with large drop in PRCP (-226.546 mm) and a large increase in TMAX90 (8.399 days). A large drop in average precipitation together with a large increase in the number of extremely hot days implies that Eugene might have begun undergoing impactful climatic change in 1989, moving towards drought like conditions. Next Cluster 5 (Baton Rouge, Cincinnati, Raleigh) stands out as a cluster with strong increase in PRCP1 (3.117 days). Cities in Cluster 3 (Albuquerque, Amarillo, Austin, Bakersfield, Brownsville, Colorado Springs, Dallas, El Paso, Houston, Miami, Nashville, Orlando, Reno, Saint Louis, Wichita), located mostly in the central region of the US have undergone large increases in TMAX90 (16.226 days) and TMAX (0.954°C), essentially showing increases in extremely hot conditions, both in duration and intensity. Cluster 4 (Birmingham, Boise, Chattanooga, Concord, Greensboro, Harrisburg, Knoxville, New York City, Omaha, Philadelphia, Richmond, San Antonio, Spokane) with cities located mostly on the eastern region began undergoing moderately large increases in temperature variables TMAX90, TAV, and TMIN. Cluster 1 (Albany, Allentown, Atlanta, Augusta, Boston, Burlington, Charleston, Charlotte, Cleveland, Columbia, Dayton, Fresno, Hartford, Jacksonville, Las Vegas, Lexington, Louisville, Montgomery, New Orleans, Phoenix, Portland ME, Rochester, Sacramento, Salt Lake City, San Diego, Seattle, Tallahassee, Tampa, Tucson, Virginia Beach, Washington DC) with cities located mostly in either east coast or west coast has also undergone increasing trends in TAVG, TMAX and TMIN variables. Cluster 2 with large number of cities located throughout US showed no significant changes in any of the variables. With the exception of Eugene that showed large drop in PRCP, a striking feature of this change year is that there are no negative changes in any of the averages across all clusters and all variables. The increases are all in temperature variables only, and thus, the change year 2010 can be seen as a shift towards even warmer conditions that began in 1989.

The discussion of results will be enriched much more through a proper compilation of various observations made about changes that occurred in the years 1957, 1989, and 2010. We have done such a compilation of observations for each of 1957, 1989, and 2010, and these compilations are presented in Tables 12-14, respectively. There is much to learn from a proper understanding of the information contained in each of these tables. We begin with a careful look at Table 12 where observations are summarized about changes that occurred in 1957. The Precipitation variables PRCP1 and PRCP had moderate increases in the southeastern region, and high increases in the eastern half of the US. As for temperature variables, there is much similarity in the changes that occurred in TMAX90, TAVG, TMAX and TMIN variables. All of these four temperature variables have undergone high increase in the west coast and a decrease in southeastern as well as northeastern parts of the US. Only TMAX32 variable has undergone an increase in southeastern and northeastern regions. The summary from Table 13 for the change year 1989 reveals that there was a high decrease in the two precipitation variables PRCP1 and PRCP at Tallahassee, and high increase in PRCP1 and

PRCP in the eastern region. Among temperature variables even though TMAX32 decreased in the northeastern region, other variables TMAX90, TAVG, TMAX and TMIN have all increased in the southern region, and the temperature has increased in the northeastern for TAVG and TMIN also. As for changes in the year 2010, Table 14 shows the precipitation variable PRCP1 increased at Baton Rouge, Cincinnati, and Raleigh whereas there was sharp decrease PRCP at Eugene. Among temperature variables no changes were observed in TMAX32. Similar changes occurred in the three variables TAVG, TMAX, and TMIN with increasing temperatures seen in western and eastern regions. The TMAX90 temperature variable has undergone high increases in central region while significant increases occurred in the eastern region.

Region wise representation of changes presented in Table 15 also allow us to further understand the nature of the changes that occurred in both temperature and precipitation variables. Changes in Precipitation variables, PRCP1 and PRCP, occurred in eastern, southeastern, and southern areas of the US. All the changes in both the precipitation variables occurred in either 1957 or 1989, and moreover, all changes have led to varying levels (small to high) of increases only. In particular, increases in PRCP occurred in all three regions, whereas increases in PRCP1 occurred only in eastern and southeastern regions, that too in 1957. As for temperature changes, there were both decreases and increases in the temperature variables. The decreases were limited to northeastern, southeastern, and southern regions and the decreases in temperature variables occurred mostly in 1957 only. All changes that occurred in temperature variables in 2010 have been increases only, and these increases have occurred in northeastern, eastern and central regions. The year 1989 saw decreases in the northeastern region and otherwise increases in southern region while high increases occurred throughout in TMIN only.

Finally, we have computed overall magnitudes of change for each climatic variable over the 75-year long sampling period 1948-2023. The computed overall average changes are: PRCP1: 0.193 days; PRCP: 5.559 mm; TMAX32: -0.166 days; TMAX90: 0.660 days; TAVG: 0.333°C; TMAX: 0.186°C; TMIN: 0.429°C. Clearly, at an overall level, there were no significant changes in the averages of the two precipitation variables PRCP1 and PRCP as well as the two discrete temperature variables TMAX32 and TMAX90. The overall changes in TAVG, TMAX and TMIN are of much interest. These changes observed over the 75-year period can be better compared with previous works in the literature if we convert these average changes into °C/100 years. Upon doing so we find the changes in averages as – TAVG: 0.444°C /100 yr; TMAX: 0.248°C /100 yr; TMIN: 0.572°C /100 yr.

The above average changes per century are highly influenced by the cooling period that began in 1957 and continued till 1989. Hence, in order to understand more recent trends in temperature changes, it is better to compute the overall magnitudes of changes in TAVG, TMAX and TMIN for the period 1990-2023, a 33 year period. We found these 33-year period changes in averages as – TAVG: 0.595°C; TMAX: 0.404°C; TMIN: 0.699°C. Assuming present temperature trends would continue till the end of the century, the same changes when projected as °C/100 yr are – TAVG: 1.803°C /100 yr; TMAX: 1.224°C /100 yr; TMIN: 2.118°C /100 yr. Of course, the assumption that current temperature trends would continue till the end of the century can be seen to be unrealistic and in this sense the above °C/100 yr increases should be viewed as being conservative.

Comparing the changes in temperature variables with existing literature, even if global in scope, Hawkins and Jones (2013) remarked that more recent analyses support average temperature increases at the rate of 0.500°C/100 yr, first projected by Callendar (1938). In comparison, our current study projects change in average temperature for the US as 0.444°C /100 yr. Based upon a change point modeling, Lee *et al.* (2014) concluded that monthly maximum had a mean change of 0.47°C/Century while the mean change for the monthly minimum was 1.65°C/Century. Results for our monthly maximum TMAX and monthly minimum TMIN showed increases in both extremes. For the whole data period 1948-2023, the increase in TMAX is 0.248°C /100 yr and the same for TMIN is 0.572°C /100 yr. However, if we consider increases for the data period 1990-2023, then the increases in the two extremes are much higher with the increase in TMAX at 1.224°C /100 yr and the increase in TMIN at 2.118°C /100 yr.

6. Concluding remarks

In this study, we have applied recently developed method of high dimensional change point analysis for identifying changes in temperature and precipitation variables based upon data from 91 stations from contiguous United States for the period 1948-2023. A total of seven climatic variables have been considered for studying changes and among these, one precipitation variable and four temperature variables represent extremes. The analysis has identified changes occurring in the years 1957, 1989, and 2010. The magnitudes of changes in the variables and relevant areas where changes have occurred has all been discussed in sufficient detail in the previous section. Here, we shall focus briefly on reasons behind the changes identified by the methodology. First, it is important to note that the change point methodology applied in this study only enables to identify changes but doesn't dwell into reasons behind any of the changes identified by the method. Thus, we need to collect such information from published literature. Changes in climatic variables can occur due to anthropogenic factors or due to various natural phenomena including volcanic eruptions, solar radiation fluctuations, ocean fluctuations such as Pacific Decadal Oscillations (PDO) *etc.* Abrupt changes in climatic variables can also occur due to undocumented causes such as changes in measuring instrumentations that do not get recorded, unrecorded shifts in station locations, *etc.* Anthropogenic causes are those human activities such as industrialization pollution, deforestation, urbanization, *etc.*, that lead to emitting harmful greenhouse gases into the atmosphere.

Wild *et al.* (2005) discuss about evidence of solar dimming caused by air pollution between the period 1958-1985 and the reversal of solar dimming to solar brightening subsequent to 1985. It is possible that the solar dimming between 1958-1985 may have induced the temperature declines that our analysis has identified between the years 1959-1989, a time period that closely matches with solar dimming period. Since solar dimming is a global phenomenon, it is possible that the temperature declines during the identified period may not be limited to the United States alone. Also, the solar dimming apparently does not impact uniformly throughout the United States since the temperature declines have been noticed predominantly in the northeastern, southeastern and southern regions of the US. Conversely, the solar brightening that began after 1985 might explain the observed increases uniformly in all temperature variables subsequent to the year 1989. Greater increases in temperature variables observed since 2010 require further investigation.

Acknowledgements:

The authors thank the reviewer(s) and the Editors for constructive comments and suggestions that have led to improved discussions in the paper.

References

Armal, S., Devineni, N., and Khanbilvardi, R. (2018). Trends in extreme rainfall frequency in the contiguous united states: Attribution to climate change and climate variability modes. *Journal of Climate*, **31**, 369–385.

Bai, J. (1994). Least squares estimation of a shift in linear processes. *Journal of Time Series Analysis*, **15**, 453–472.

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.

Bhattacharjee, M., Banerjee, M., and Michailidis, G. (2017). Common change point estimation in panel data from the least squares and maximum likelihood viewpoints. *arXiv preprint arXiv:1708.05836*.

Bhattacharjee, M., Banerjee, M., and Michailidis, G. (2020). Change point estimation in a dynamic stochastic block model. *The Journal of Machine Learning Research*, **21**, 4330–4388.

Bieniek, P. A., Walsh, J. E., Thoman, R. L., and Bhatt, U. S. (2014). Using climate divisions to analyze variations and trends in alaska temperature and precipitation. *Journal of Climate*, **27**, 2800–2818.

Cai, H. and Wang, T. (2023). Estimation of high-dimensional change-points under a group sparsity structure. *Electronic Journal of Statistics*, **17**, 858–894.

Callendar, G. S. (1938). The artificial production of carbon dioxide and its influence on temperature. *Quarterly Journal of the Royal Meteorological Society*, **64**, 223–240.

Chen, Y., Wang, T., and Samworth, R. J. (2023). Inference in high-dimensional online changepoint detection. *Journal of the American Statistical Association*, **119**, 1461–1472.

Cho, H. (2016). Change-point detection in panel data via double cusum statistic. *Electronic Journal of Statistics*, **10**, 2000–2038.

Cho, H. and Fryzlewicz, P. (2015). Multiple-change-point detection for high dimensional time series via sparsified binary segmentation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **77**, 475–507.

Cho, H. and Kirch, C. (2022). Bootstrap confidence intervals for multiple change points based on moving sum procedures. *Computational Statistics & Data Analysis*, **175**, 107552.

Csorgo, M. and Horváth, L. (1997). *Limit Theorems in Change-Point Analysis.* John Wiley & Sons Chichester.

Eichinger, B. and Kirch, C. (2018). A MOSUM procedure for the estimation of multiple random change points. *Bernoulli*, **24**, 526 – 564.

Enikeeva, F. and Harchaoui, Z. (2019). High-dimensional change-point detection under sparse alternatives. *The Annals of Statistics*, **47**, 2051 – 2079.

Fotopoulos, S. B., Jandhyala, V. K., and Khapalova, E. (2010). Exact asymptotic distribution of change-point mle for change in the mean of gaussian sequences. *The Annals of Applied Statistics*, **4**, 1081–1104.

Fryzlewicz, P. (2014). Wild binary segmentation for multiple change-point detection. *The Annals of Statistics*, **42**, 2243–2281.

Gaffen, D. J. and Ross, R. J. (1998). Increased summertime heat stress in the us. *Nature*, **396**, 529–530.

Griffiths, M. L. and Bradley, R. S. (2007). Variations of twentieth-century temperature and precipitation extreme indicators in the northeast united states. *Journal of Climate*, **20**, 5401–5417.

Grundstein, A. and Dowd, J. (2011). Trends in extreme apparent temperatures over the united states, 1949–2010. *Journal of Applied Meteorology and Climatology*, **50**, 1650–1653.

Hawkins, E. and Jones, P. D. (2013). On increasing global temperatures: 75 years after callendar. *Quarterly Journal of the Royal Meteorological Society*, **139**, 1961–1963.

Jackson, B., Scargle, J. D., Barnes, D., Arabhi, S., Alt, A., Gioumousis, P., Gwin, E., Sangtrakulcharoen, P., Tan, L., and Tsai, T. T. (2005). An algorithm for optimal partitioning of data on an interval. *IEEE Signal Processing Letters*, **12**, 105–108.

Jandhyala, V., Fotopoulos, S., MacNeill, I., and Liu, P. (2013). Inference for single and multiple change-points in time series. *Journal of Time Series Analysis*, **34**, 423–446.

Jirak, M. (2015). Uniform change point tests in high dimension. *The Annals of Statistics*, **43**, 2451–2483.

Kassambara, A. and Mundt, F. (2021). Factoextra: extract and visualize the results of multivariate data analyses, R package version 1.0. 7. 2020.

Kaul, A., Fotopoulos, S. B., Jandhyala, V. K., and Safikhani, A. (2021). Inference on the change point under a high dimensional sparse mean shift. *Electronic Journal of Statistics*, **15**, 71–134.

Kaul, A., Jandhyala, V. K., and Fotopoulos, S. B. (2019). An efficient two step algorithm for high dimensional change point regression models without grid search. *Journal of Machine Learning Research*, **20**, 1–40.

Kaul, A., Zhang, H., Tsampourakis, K., Michailidis, G., Kaul, A., Zhang, H., Tsampourakis, K., and Michailidis, G. (2023). Inference on the change point under a high dimensional covariance shift. *Journal of Machine Learning Research*, **24**, 1–68.

Killick, R., Fearnhead, P., and Eckley, I. A. (2012). Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association*, **107**, 1590–1598.

Kunkel, K. E., Liang, X.-Z., and Zhu, J. (2010). Regional climate model projections and uncertainties of us summer heat waves. *Journal of Climate*, **23**, 4447–4458.

Lai, Y. and Dzombak, D. A. (2019). Use of historical data to assess regional climate change. *Journal of Climate*, **32**, 4299–4320.

Lee, J., Li, S., and Lund, R. (2014). Trends in extreme us temperatures. *Journal of Climate*, **27**, 4209–4225.

Lee, S., Seo, M. H., and Shin, Y. (2016). The lasso for high dimensional regression with a possible change point. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **78**, 193–210.

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K., Studer, M., Roudier, P., and Gonzalez, J. (2013). Package 'cluster'. *Dosegljivo na*, **980**.

Martinez-Villalobos, C. and Neelin, J. D. (2023). Regionally high risk increase for precipitation extreme events under global warming. *Scientific Reports*, **13**, 5579.

Nazarian, R. H., Brizuela, N. G., Matijevic, B. J., Vizzard, J. V., Agostino, C. P., and Lutsko, N. J. (2024). Projected changes in mean and extreme precipitation over northern mexico. *Journal of Climate*, **24**.

Nazarian, R. H., Vizzard, J. V., Agostino, C. P., and Lutsko, N. J. (2022). Projected changes in future extreme precipitation over the northeast united states in the na-cordex ensemble. *Journal of Applied Meteorology and Climatology*, **61**, 1649–1668.

Oswald, E. M. (2018). An analysis of the prevalence of heat waves in the united states between 1948 and 2015. *Journal of Applied Meteorology and Climatology*, **57**, 1535–1549.

Oswald, E. M. and Rood, R. B. (2014). A trend analysis of the 1930–2010 extreme heat events in the continental united states. *Journal of Applied Meteorology and Climatology*, **53**, 565–582.

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

Robinson, W. A. (2021). Climate change and extreme weather: A review focusing on the continental united states. *Journal of the Air & Waste Management Association*, **71**, 1186–1209.

Rupp, D. E., Hawkins, L. R., Li, S., Koszuta, M., and Siler, N. (2022). Spatial patterns of extreme precipitation and their changes under˜ 2° c global warming: a large-ensemble study of the western usa. *Climate Dynamics*, **59**, 2363–2379.

Venkatraman, E. S. (1992). *Consistency Results in Multiple Change-Point Problems*. Stanford University.

Vershynin, R. (2019). *High-Dimensional Probability*, volume NA. Cambridge, UK: Cambridge University Press.

Villarini, G., Smith, J. A., and Vecchi, G. A. (2013). Changing frequency of heavy rainfall over the central united states. *Journal of Climate*, **26**, 351–357.

Wang, D., Yu, Y., and Rinaldo, A. (2020). Univariate mean change point detection: Penalization, cusum and optimality. *Electronic Journal of Statistics*, **14**, 1917–1961.

Wang, D., Yu, Y., and Rinaldo, A. (2021). Optimal change point detection and localization in sparse dynamic networks. *The Annals of Statistics*, **49**, 203 – 232.

Wang, T. and Samworth, R. J. (2018). High dimensional change point estimation via sparse projection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **80**, 57–83.

Wang, X., Huang, G., Lin, Q., Nie, X., and Liu, J. (2015). High-resolution temperature and precipitation projections over ontario, canada: a coupled dynamical-statistical approach. *Quarterly Journal of the Royal Meteorological Society*, **141**, 1137–1146.

Wang, X. L., Chen, H., Wu, Y., Feng, Y., and Pu, Q. (2010). New techniques for the detection and adjustment of shifts in daily precipitation data series. *Journal of Applied Meteorology and Climatology*, **49**, 2416–2436.

Wild, M., Gilgen, H., Roesch, A., Ohmura, A., Long, C. N., Dutton, E. G., Forgan, B., Kallis, A., Russak, V., and Tsvetkov, A. (2005). From dimming to brightening: Decadal changes in solar radiation at earth's surface. *Science*, **308**, 847–850.

Zhao, X. and Chu, P.-S. (2010). Bayesian changepoint analysis for extreme events (typhoons, heavy rainfall, and heat waves): An rjmcmc approach. *Journal of Climate*, **23**, 1034–1046.

Zhou, X., Huang, G., Baetz, B. W., Wang, X., and Cheng, G. (2018). Precis-projected increases in temperature and precipitation over canada. *Quarterly Journal of the Royal Meteorological Society*, **144**, 588–603.