2017 IEEE International Conference on Data Mining Workshops

Deriving Data-driven Insights from Climate Extreme Indices for the Continental US

Xinbo Huang
NOAA Southern Regional Climate Center
Louisiana State University
Baton Rouge, LA 70803
Email: xinbo@srcc.lsu.edu

David Sathiaraj
NOAA Southern Regional Climate Center
Dept of Geography & Anth.
Louisiana State University
Baton Rouge, LA 70803
Email: davids@srcc.lsu.edu

Lei Wang & Barry D Keim
Dept. of Geography & Anth.
Louisiana State University
Baton Rouge, LA 70803

***Abstract*—Daily climate data observations from more than 3000 climate measurement sites in the continental U.S. were mined and analyzed to derive insights and trends from climate extreme indices. Daily climate data observations were aggregated by climate divisions and analyzed to derive a new climate extremes indices data set (Threshold Exceedence Frequency, TEF). Each climate division was statistically assessed for the following elements: maximum and minimum temperature, precipitation and snowfall. The climate data time series were divided into 2 time intervals (1946-1980 and 1981-2015) and the occurrence frequencies of various climate extreme indices was statistically examined. Results revealed interesting insights such as an increasing frequency of occurrence of night-time temperatures in South-east US and decreasing frequency of winter temperature and snowfall extremes in northern US. The study also produced a new web-based visualization system to analyze the results of the study. The visualization system included interactive choropleth maps and charts to depict spatiotemporal changes in various climate thresholds over time.**

1. Introduction

Climate extremes are meteorological events that can have significant impacts on human and natural systems. Weather hazards, such as heat waves, drought, heavy thunderstorms, floods, hurricanes, occur frequently, and are a threat to human lives and property. Understanding extreme weather and climate events coupled with increased societal vulnerability to such events, highlights a need to collect and analyze data pertaining to extreme climate events and to discover valuable decision-making insights using data analytics.

The area encompassing the Continental United States is vast with variable climate types and land cover. There are approximately 26,000 climate measurement sites in Continental US. Climate is defined as long-term averages and variations in weather measured over a period of several decades [1]. Climate data conform to a time series nature. This work examines data from nearly

This paper attempts to provide an assessment of climate trends for the continental United States in recent decades by analyzing daily resolution climate data. The climate elements analyzed include the following: maximum temperature, minimum temperature, precipitation, and snowfall. In addition, a data visualization system has also been developed to depict climate trends over the past 70 years. The visualization system is hosted at http://dcat.srcc.lsu.edu.

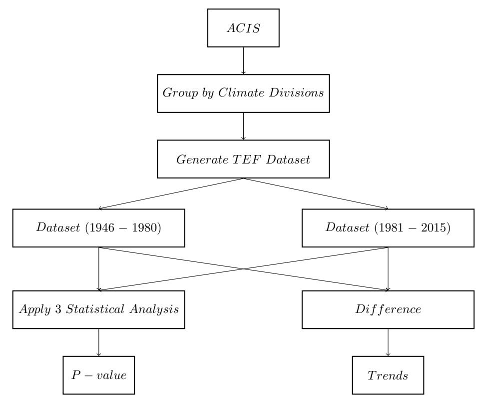

Figure 1. Flowchart for data processing

2. Data and Methodologies

Figure 1 depicts a flowchart for data processing. First, daily climate data are obtained from the Applied Climate Information System (or ACIS) [2]. The data are then grouped by climate divisions, and extreme frequency data are generated by setting some thresholds. Grouping by climate divisions is conducted as follows: An average was computed using data from at least 3 climate measurement sites - each of the sites included data that spanned the time period 1946-2015 and each of the climate measurement sites included less than 10% missing values per year. So for example, to compute for a say New York's climate division 1, annual frequencies of minimum temperatures exceeding 75 °F are collected from at least 3 climate measurement sites and a mean annual frequency value is computed. If less than 3 climate measurement sites were available for a climate

2375-9259/17 $31.00 © 2017 IEEE
DOI 10.1109/ICDMW.2017.46

division (likely due to excessive missing values), then that climate division was excluded from the study.

The time-series extreme frequency data between 1946 and 2015 is divided into two independent samples, which can be compared using non-parametric statistical hypothesis test. Finally, p-value is obtained to be the representation of significance of whether the two time-series climate data have a similar distribution. In addition, the difference between the means of the 2 time-series data for each of the thresholds is also evaluated, to indicate an increasing [1] or decreasing [2] trend.

2.1. Data Source

This research used the daily climate data from the Applied Climate Information System (ACIS) [2], an Internet-based system designed to facilitate the generation and dissemination of climate data products to users. ACIS is developed by the NOAA Regional Climate Centers (RCCs) to manage the complex flow of information from climate data collectors to end users of climate data information.

ACIS provides a web-services framework [3] that accepts time-series search parameters and returns climate information in JavaScript Object Notation (JSON). JSON is a common data format that can be absorbed by many programming languages, including C, C++, Java, JavaScript, Perl, R and Python. For each call, users specify a set of parameter to describe the data being requested. After passing these parameters to the server and accessing these climate data, a climate data product is returned to users.

3210 climate stations in the continental United States were used in this study. ACIS provides access to daily data observations from more than 26000 Global Historical Climate Network's (GHCN-D) [4] climate data measurement sites in the US. However not all the sites span the entire time period of 1946-2015. Additional criteria used for this data analysis and study included the following: allow for less than 10% missing values for a station per year and every climate division should have at least 3 climate measurement sites. Once this criteria was applied, the number of valid stations that fit these criteria reduced to 3210. These 3210 stations are distributed to cover most of land in the continental United States. By analyzing the climate data from these stations, the trends of climate extremes can then be obtained for the continental United States.

2.2. Climate Divisions in the US

The continental United States (U.S.) is subdivided into 344 climate divisions by the National Centers for Environmental Information (NCEI, formerly known as National Climatic Data Center (NCDC)) [5]. Each climate division represents nearly homogenous climatic regions. For each climate division, monthly station temperature and precipitation values are computed from the daily observations [6], and their monthly temperature, monthly water equivalent precipitation, Palmer Drought Severity Index, and Palmer Hydrological Drought Index values have been generated for a period dating back to 1895 ([7] and [8]). Numerous applications have used these divisional climate data, e.g., they are used to monitor the U.S. climate by the NCEI, NOAA's Climate Prediction Center, the National Drought Mitigation Center, and others. These divisional data sets are also used frequently in applied research [8], [9], [10].

```
Initialize array with climate thresholds, ClimateExtremeIndices
for each climate extreme threshold, climext in ClimateExtremeIndices do
    for each climate measurement site stn in a climate division do
        Get climate data observations spanning 1946-2015.
        Analyze climate observations and find frequency of occurrence exceeding threshold climext
    end for
    Group by climate division
    Store each index for each climate division in a key-value store (Redis)
end for
```

Figure 2: Methodology Used to Derive Climate Extreme Indices For Each Climate Division

A similar climate division based approach is used in this spatiotemporal analysis. Climate measurement data from individual sites were grouped to get a climate divisional value.

The NCEI climate divisions dataset for Continental US comprises of 327 out of the 344 climate divisions. Hence for this study, climate measurement data from 3210 data sites were grouped into 327 climate divisions to derive a climate extreme indices data set. We call this the Threshold Exceeding Frequencies (TEF) Dataset. To group a set of climate measurement sites into a climate division, a minimum of 3 climate measurement stations were set as a requirement and the frequency measurement (number of days exceeding a threshold) was averaged for the climate division. Each of the climate data sites in a climate division was also required to have less than 10% missing values per year. Figure 2 provides a pseudocode or general methodology of how the TEF data set is generated.

2.3. Threshold Exceeding Frequencies (TEF) Dataset

The daily climate data from 3210 climate measurement sites was used to generate the Threshold Exceeding Frequency (TEF) data set. This derived data set was generated by computing the annual number of days when the climate observation exceeded a given threshold. The thresholds used for this study was based on a combination of thresholds

1. In this work, ***increasing trend*** denotes that the more recent time period (1981-2015) is experiencing higher mean frequency of days for the threshold being analyzed, when compared to the prior time period (1946-1980)

2. ***decreasing trend*** denotes that the more recent time period (1981-2015) is experiencing lower mean frequency of days for the threshold being analyzed, when compared to the prior time period (1946-1980)

TABLE 1. THRESHOLDS TO GENERATE TEF DATASET

Element	Thresholds
Maximum Temperature	≥105, ≥100, ≥95, ≥85
Minimum Temperature	≥80, ≥75, ≥70, ≥65, ≤36, ≤32, ≤28, ≤24, ≤15, ≤10, ≤5, ≤0
Precipitation	≥2, ≥4, sum
Snowfall	sum

used in the *CLIMDEX - Datasets for Indices of Climate Extremes* [11] and the Southeast chapter of the *US National Climate Assessment* document, released in 2014 [12]. For each threshold and a given climate division, there would be 70 values, corresponding to the 1946-2015 study period.

The thresholds used in this study included the following: maximum temperatures greater 105 °F, 100 °F, 95 °F, or 85 °F, minimum temperatures greater than 80 °F, 75 °F, 70 °F, or 65 °F, minimum temperatures lower than 36 °F, 32 °F, 28 °F, 24 °F, 15 °F, 10 °F, 5 °F, 0 °F, precipitation values greater than 2 inches or 4 inches, total annual precipitation (in inches), and total annual snowfall (in inches), as shown in Table 1. In addition, to ensure a serially complete data set, climate divisions should have at least 3 climate measurement sites, each of which, having no more than 10% of missing values per year.

2.4. Analysis using Non-Parametric Statistical Tests

The determination of the distribution form which a sample is drawn is an important problem in many statistical applications [13]. If the distribution is not known, or is known to not follow a particular form, then non-parametric statistics are appropriate. 3 non-parametric tests - Wilcoxon Signed-Rank Test, Mann-Whitney U Test and Kolmogorov-Smirnov Test - were used to compare 2 time-series comprising of the time periods 1946-1980 and 1981-2015. These 3 tests were applied on each of the climate indices and for each climate division in the Continental US. For this study, the non-parametric tests were applied to answer the question - are the frequency of occurrence of days exceeding a given threshold different for the 2 time periods and is the difference statistically significant? So in other words, as an example, if one is looking at minimum temperatures exceeding a threshold of 75, are the 2 time series representing annual frequency of days exceeding this threshold, statistically different or are they the same. The difference between the means of the annual frequencies for the 2 time series was also analyzed. A positive difference indicates that there are more number of days where the minimum temperature exceeds the threshold of 75 for the time period 1981-2015 as compared to 1946-1980. If the p-value obtained from the non-parametric test indicates a value that is less than 0.05, then the 2 time series are considered statistically significant and different.

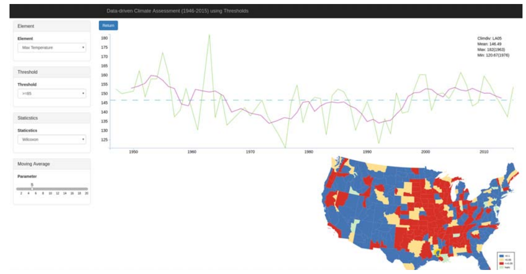

Figure 3: Screenshot of data visualization system with chart of maximum temperature

3. Data Visualization

3.1. Structure of Data Visualization System

To demonstrate the results of this study and to intuitively help users access the climate extremes frequencies data set, a data visualization system was developed. Typical users can span domains such as climate science, agriculture, finance and economics (commodity markets), recreation site managers (such as those in the ski resort industry) and actuarial science. This system will help users to query the climate threshold exceedence frequencies derived in this study and provides an interface to depict the statistical analysis conducted.

The data visualization system contains a low latency, robust memory database, a flexible real-time query system, and a user-friendly web interface. The visualization system includes the following: the querying options in the interface receive a set of meta data and threshold parameters from the user and this information is transmitted to the database. The database stores the threshold exceeding frequencies (TEF) data set. Once the data is retrieved from the database, it is sent to the interface that includes a map and chart-based visualizations.

Figure 3 displays the interface of the data visualization system. It has a query panel where one can set input parameters and provides a choropleth map to show the distribution of p-values for all the climate divisions in the Continental US. Clicking on a climate division produces a line chart that provides trends on whether the frequency of days exceeding a certain climate frequency is showing an increasing trend in the more recent time period (1981-2015) or a decreasing trend.

The query panel of the visualization tool includes the following: 1. an element selector which can select climate elements including maximum temperature, minimum temperature, precipitation, and snow, 2. a climate extreme threshold selector, 3. a statistic selector which can indicates the non-parametric test used (Wilcoxon, Mann-Whitney, Kolmogorov-Smirnov or K-S and Difference in mean frequency of days between 1946-1980 and 1981-2015, 4. and a moving average parameter slider which can select integers from 1 to 20. These query panels are interactive and data is refreshed based on the user input.

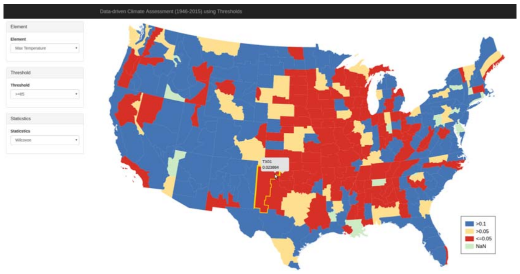

Figure 4: Choropleth map of p-values using Wilcoxon for Maximum Temperatures greater than 85 °F

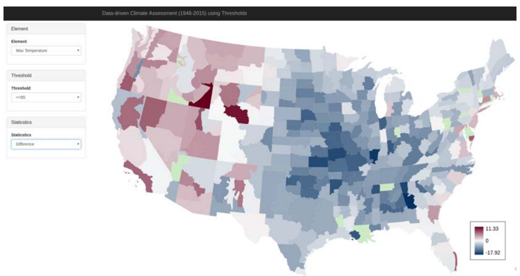

Figure 5: Screenshot of data visualization system with difference choropleth map when maximum temperature is greater than 85 °F

3.2. Choropleth Map

A choropleth map is a thematic map in which areas are shaded or patterned in proportion to the measurement of the statistical variable being displayed on the map [14]. The visualization system displays the results of the non-parametric tests as a choropleth map. Each climate division in the continental United States is shaded on the basis of the p-value (corresponding to the application of Wilcoxon, Mann-Whitney or K-S tests) or the difference of mean frequency of days exceeding a certain threshold between the 2 time periods. For example, as shown in Figure 4, there is the choropleth map about wilcoxon test p-value of the climate dataset of climate divisions whose maximum temperature is greater than 85 °F. In the choropleth map, the blue represents that the p-value is greater than 0.1 and the 2 time series are not statistically significant, the yellow color represents climate divisions that had a p-value between 0.05 and 0.1 and red represents climate divisions that had a p-value that was lower than 0.05 and the difference between the 2 time series was statistically significant, and the green represents climate divisions that had too many missing observations.

Figure 5 is the choropleth map for difference when minimum temperature is greater than 85 °F. The red areas means the frequencies in these places increase and the blue areas means that in these decrease.

3.3. Line Chart

When a climate division in the choropleth map is clicked, the line chart about the time-series extreme climate data for the climate division is displayed to help users access more detailed information. Figure 6 shows interface about the chart, the green line represents climate extremes frequencies data of the climate division, the blue dash line represents the mean of the data, the red line represents moving average line and users can set the parameter of moving average line in the left panel.

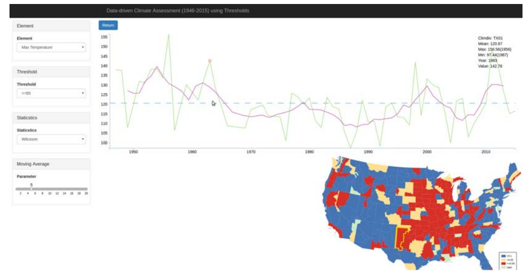

Figure 6: Screenshot of line chart with the climate division TX01 when maximum temperature is greater than 85 °F

When a mouse is moved over the chart, the closest value is shown so that users can check all the value in the climate extremes datasets.

4. Results and Analysis

4.1. P-value

In statistics, the p-value is a function of the observed sample results (a test statistic) relative to a statistical model, which measures how extreme the observation is [15]. The p-value is defined as the probability of obtaining a result equal to or ”more extreme” than what was actually observed, when the null hypothesis is true [16]. A small p-value (typically ≤ 0.05) indicates strong evidence against the null hypothesis, so you reject the null hypothesis. A large p-value (> 0.05) indicates weak evidence against the null hypothesis, so you fail to reject the null hypothesis [17], [18], [19].

Table 2, Table 3, and Table 4 displays the number of climate divisions with different statistic test p-value ranges, including p-value $\leq$ 0.05, 0.05 $<$p-value$\leq$ 0.1, and p-value$>$ 0.1. The left-most column in each of the tables provides the names of each of the climate extreme indices (Maximum temperature has been abbreviated as *tx*, *tn* represents minimum temperature, *pc* represents precipitation, and *sw* represents snow). The percentages provided in parenthesis indicate the percentage of climate divisions that show an increasing trend of occurrence of a climate extreme (in other words, the difference between the extreme frequency of occurrence for the 2 time periods studied was positive). In other words, a high percentage value indicates that there are more number of climate divisions where the frequency of occurrence of a given climate extreme is greater in the recent time period of 1981-2015 as compared to the time period of 1946-1980. Similarly, a small percentage value for climate extreme thresholds such as minimum temperatures less than 10 indicates a declining trend for the most recent time period of 1981-2015.

TABLE 2. ANALYSIS OF RESULTS USING WILCOXON - BREAKDOWN OF CLIMATE DIVISIONS BASED ON P-VALUES AND THE PERCENTAGE OF CLIMATE DIVISIONS WITH AN INCREASING TREND (THE LABEL OF 'MISSING' INDICATES CLIMATE DIVISIONS FOR WHICH P-VALUES COULD NOT BE COMPUTED DUE TO UNAVAILABILITY OF DATA)

	<=0.05	>0.05	>0.1	Missing
tx>=105	7 (57%)	10 (8%)	306 (43%)	4
tx>=100	33 (39%)	19 (63%)	271 (39%)	4
tx>=95	50 (30%)	25 (32%)	248 (38%)	4
tx>=90	64 (12%)	39 (26%)	220 (4%)	4
tx>=85	124 (1%)	43 (19%)	156 (38%)	4
tn>=80	39 (95%)	4 (75%)	280 (51%)	4
tn>=75	106 (95%)	21 (95%)	196 (66%)	4
tn>=70	124 (97%)	27 (93%)	172 (77%)	4
tn>=65	107 (96%)	25 (80%)	191 (80%)	4
tn<=36	161 (2%)	29 (14%)	133 (29%)	4
tn<=32	168 (2%)	30 (7%)	125 (32%)	4
tn<=28	132 (2%)	41 (7%)	150 (27%)	4
tn<=24	132 (1%)	41 (0%)	150 (21%)	4
tn<=15	113 (0%)	35 (0%)	175 (16%)	4
tn<=10	100 (0%)	35 (0%)	188 (17%)	4
tn<=5	78 (0%)	30 (0%)	215 (21%)	4
tn<=0	69 (0%)	22 (0%)	232 (20%)	4
pc>=2	56 (98%)	39 (100%)	235 (79%)	0
pc>=4	13 (100%)	16 (94%)	301 (61%)	0
pcsum	87 (99%)	32 (94%)	211 (78%)	0
swsum	96 (4%)	23 (26%)	211 (3%)	0

TABLE 3. ANALYSIS OF RESULTS USING MANN-WHITNEY TEST - BREAKDOWN OF CLIMATE DIVISIONS BASED ON P-VALUES AND THE PERCENTAGE OF CLIMATE DIVISIONS WITH AN INCREASING TREND (THE LABEL OF 'MISSING' INDICATES CLIMATE DIVISIONS FOR WHICH P-VALUES COULD NOT BE COMPUTED DUE TO UNAVAILABILITY OF DATA)

	<=0.05	>0.05	>0.1	Missing
tx>=105	24 (75%)	13 (77%)	220 (52%)	70
tx>=100	40 (38%)	26 (50%)	250 (42%)	11
tx>=95	55 (27%)	23 (35%)	249 (38%)	0
tx>=90	66 (11%)	39 (26%)	222 (4%)	0
tx>=85	127 (9%)	43 (19%)	157 (38%)	0
tn>=80	57 (91%)	17 (94%)	167 (69%)	86
tn>=75	118 (93%)	25 (92%)	159 (74%)	25
tn>=70	135 (94%)	28 (93%)	157 (79%)	7
tn>=65	113 (93%)	24 (79%)	188 (81%)	2
tn<=36	165 (2%)	28 (14%)	134 (28%)	0
tn<=32	173 (2%)	28 (7%)	126 (32%)	0
tn<=28	135 (1%)	42 (7%)	150 (27%)	0
tn<=24	136 (1%)	41 (0%)	150 (21%)	0
tn<=15	120 (0%)	34 (0%)	170 (16%)	3
tn<=10	104 (0%)	38 (3%)	179 (17%)	6
tn<=5	82 (0%)	28 (0%)	205 (22%)	12
tn<=0	74 (0%)	25 (4%)	210 (21%)	18
pc>=2	59 (98%)	39 (97%)	231 (80%)	1
pc>=4	27 (89%)	19 (79%)	253 (68%)	31
pcsum	87 (99%)	32 (94%)	211 (78%)	0
swsum	96 (4%)	26 (23%)	206 (31%)	2

TABLE 4. ANALYSIS OF RESULTS USING K-S TEST - BREAKDOWN OF CLIMATE DIVISIONS BASED ON P-VALUES AND THE PERCENTAGE OF CLIMATE DIVISIONS WITH AN INCREASING TREND (THE LABEL OF 'MISSING' INDICATES CLIMATE DIVISIONS FOR WHICH P-VALUES COULD NOT BE COMPUTED DUE TO UNAVAILABILITY OF DATA)

	<=0.05	>0.05	>0.1	Missing
tx>=105	7 (57%)	4 (50%)	312 (44%)	4
tx>=100	24 (38%)	19 (47%)	280 (41%)	4
tx>=95	44 (25%)	19 (58%)	260 (37%)	4
tx>=90	51 (18%)	28 (18%)	244 (38%)	4
tx>=85	111 (8%)	42 (29%)	170 (35%)	4
tn>=80	31 (97%)	5 (60%)	287 (53%)	4
tn>=75	99 (95%)	19 (84%)	205 (69%)	4
tn>=70	98 (97%)	21 (86%)	204 (80%)	4
tn>=65	85 (95%)	28 (89%)	210 (81%)	4
tn<=36	149 (2%)	24 (4%)	150 (28%)	4
tn<=32	158 (2%)	29 (7%)	136 (30%)	4
tn<=28	132 (1%)	28 (14%)	163 (25%)	4
tn<=24	118 (1%)	31 (0%)	174 (18%)	4
tn<=15	103 (0%)	20 (0%)	200 (14%)	4
tn<=10	91 (0%)	24 (0%)	208 (15%)	4
tn<=5	70 (0%)	19 (0%)	234 (20%)	4
tn<=0	52 (0%)	20 (0%)	251 (18%)	4
pc>=2	42 (98%)	27 (100%)	261 (81%)	0
pc>=4	11 (100%)	8 (100%)	311 (62%)	0
pcsum	77 (100%)	38 (89%)	215 (79%)	0
swsum	86 (9%)	30 (17%)	214 (29%)	0

Some key observations from this analysis can be summarized as follows:

- Thresholds corresponding to minimum temperature (night-time temperatures) had a large number of climate divisions that indicate significant change. For example, for $tn >= 75$, using the Wilcoxon Test (see Table 2), 118 out of 324 climate divisions had a p-value of less than 0.05 and 95% of these 118 climate divisions indicated an increasing trend (an increasing frequency of occurrence of minimum temperatures greater than 75). This analysis indicates that in recent time periods, there have been more days with higher night-time temperatures.
- Similarly there is a decreasing trend in the frequency of occurrence of low minimum temperatures (winter temperatures). In other words, there are fewer number of days with minimum temperatures less than a given threshold. For example, Table 3 lists for $tn <= 10$, 91 climate divisions that have a statistically significant p-value of less than 0.05. More importantly, all 91 climate divisions indicate a declining trend of minimum winter temperatures.
- Extreme high precipitation events show an increasing trend as indicated in Tables 2, 3 and 4.
- There is a declining trend in snowfall totals (as indicated in Tables 2, 3 and 4).

7 climate extreme thresholds - maximum temperature $\geq$ 95°F, minimum temperature ≥ 75°F, minimum temperature $\leq$ 32°F, minimum temperature $\leq$ 0°F, precipitation $\geq$ 2 inches, annual total precipitation, annual total snowfall - are now analyzed further.

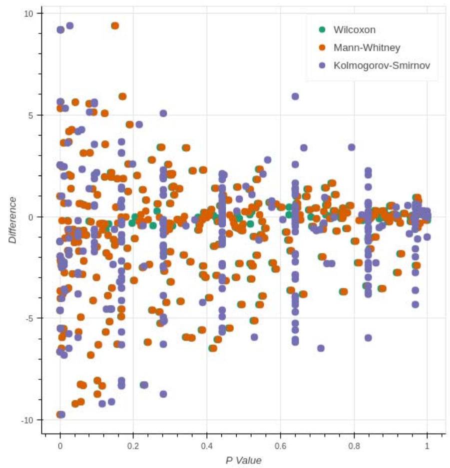

Figure 7: Distribution of difference and p-value of climate divisions in continental U.S. when maximum temperature is greater than 95 °F, for the 2 time series - 1946-1980 and 1981-2015

4.2. Maximum Temperature is Greater than 95 °F

Figure 7 is the scatter plot of all climate divisions in the continental U.S. for the number of days when maximum temperature is greater than 95°F. x axis represents p-value, y axis represents the difference between the means of two time-series (1946 - 1980 and 1981 - 2015). Green points in the scatter plot represent p-values from the Wilcoxon Test, purple points represent Mann-Whitney Test and brown points represent Kolmogorov-Smirnov (K-S) Test.

To interpret the scatterplot, p-values close to 0 means that the difference between the 2 time-series is statistically significant. A positive difference means the frequencies are indicating an increasing trend and a negative difference indicates a decreasing trend.

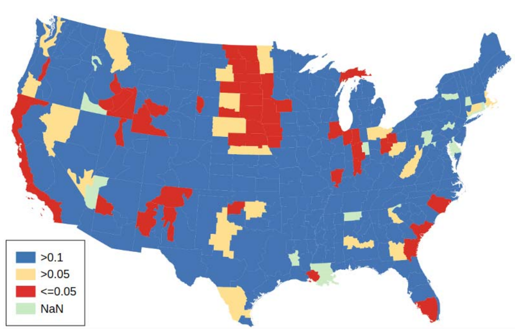

Figure 8: Distribution of Wilcoxon test p-value when maximum temperature is greater than 95 °F

For this threshold ($tx >= 95$), based on the scatter plot (Figure 7) and spatial maps (Figures 8,9,10,11), apart from a few climate divisions in the High plains (in states such as North and South Dakota, Minnesota, Nebraska) and coastal California, that indicated statistical significant increasing trends, overall there was not much evidence of an increasing or decreasing trend for this climate threshold.

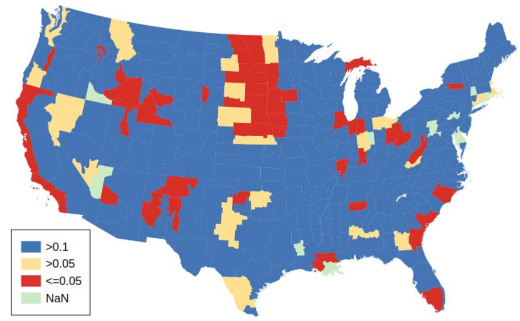

Figure 9: Distribution of Mann-Whitney test p-value when maximum temperature is greater than 95 °F

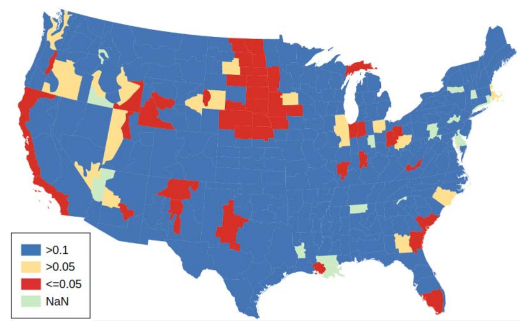

Figure 10: Distribution of Kolmogorov-Smirnov test p-value when maximum temperature is greater than 95 °F

4.3. Minimum Temperature is Greater than 75 °F

Figure 12 is the scatter graph of the all climate divisions in the continental U.S. for frequencies when minimum temperature is greater than 75 °F. In the figure, most of p-values are close to 0, and most of differences between the 2 time series are greater than 0. This indicates that the frequency of occurrence of high minimum temperature is showing an increasing trend for the time period of 1981-2015.

Figure 13, Figure 14, Figure 15, and Figure 16 display p-values and differences in frequencies when minimum temperature is greater than 75 °F. Based on these maps, it is clear that vast areas in the South-east, South-central and South-west parts of the Continental US are showing an

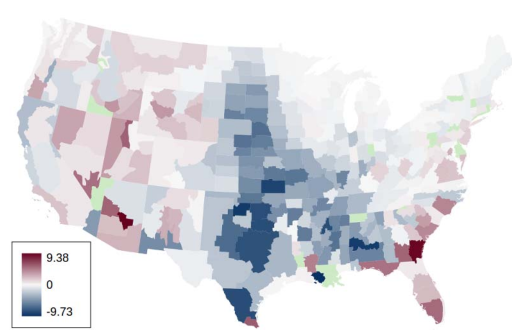

Figure 11: Distribution of difference when maximum temperature is greater than 95 °F

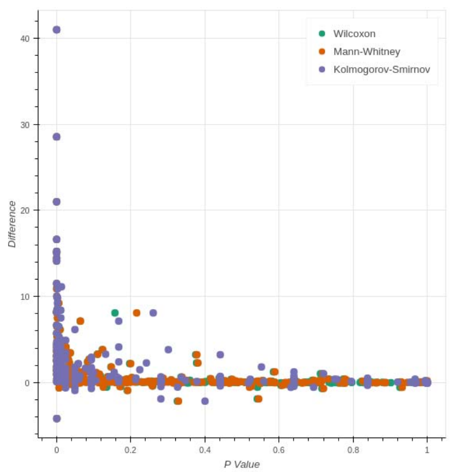

Figure 12: Distribution of difference and p-value of climate divisions in continental U.S. when minimum temperature is greater than 75 °F

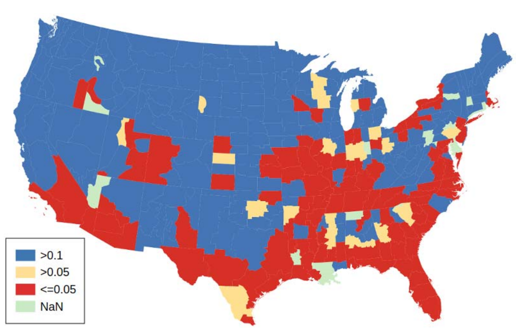

Figure 13: Distribution of Wilcoxon test p-value when minimum temperature is greater than 75 °F

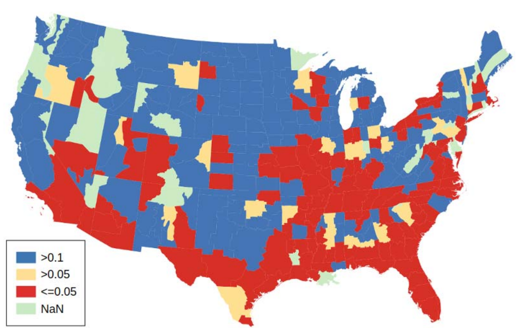

Figure 14: Distribution of Mann-Whitney test p-value when minimum temperature is greater than 75 °F

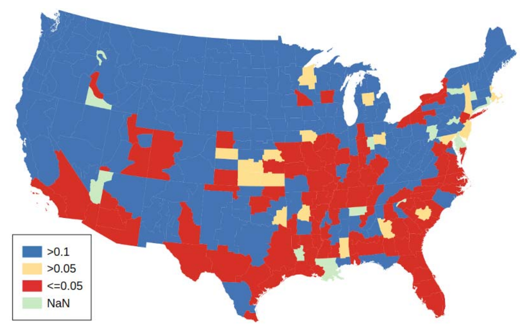

Figure 15: Distribution of Kolmogorov-Smirnov test P-value when minimum temperature is greater than 75 °F

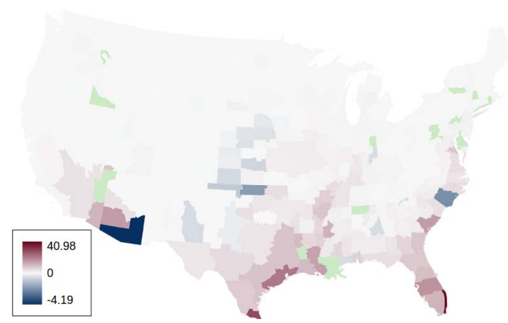

Figure 16: Distribution of difference when minimum temperature is greater than 75 °F

increasing frequency trend in minimum temperatures greater than or equal to 75.

4.4. Minimum Temperature is Lower than 32 °F

Figure 17 is the scatter graph of the all climate divisions in the continental U.S. for frequencies when minimum temperature is lower than 32 °F. Figures 18, 19,20, 21

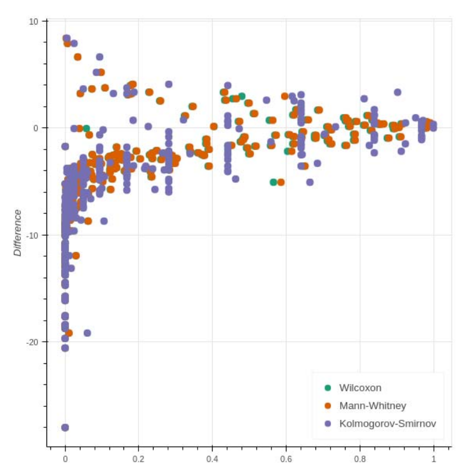

Figure 17: Distribution of difference and p-value of climate divisions in the continental U.S. when minimum temperature is lower than 32 °F

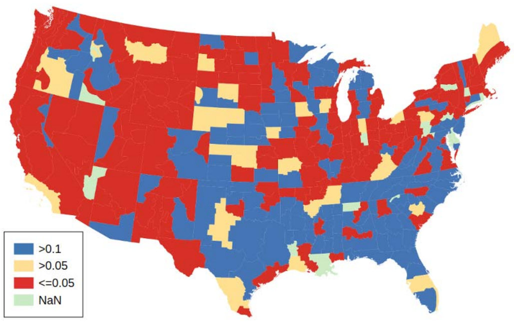

Figure 18: Distribution of Wilcoxon test p-value when minimum temperature is lower than 32 °F

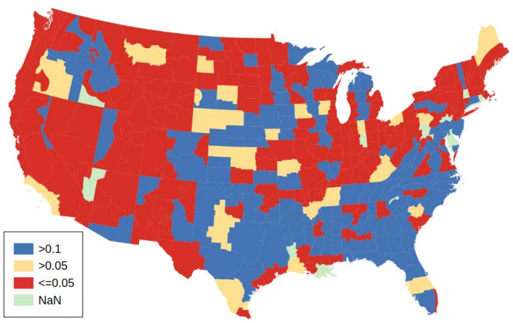

Figure 19: Distribution of Mann-Whitney test p-value when minimum temperature is lower than 32 °F

show areas where days with minimum temperatures below freezing point are decreasing in the recent time period and the statistical difference between the 2 time series. Vast portions of the regions west of the rockies, the mid-west and the north-east portions of the US experienced fewer days with less than 32 °F in the most recent time period of 1981-2015 as compared to 1946-1980.

4.5. Minimum Temperature is Lower than 0 °F

Figure 22 is the scatter graph of the all climate divisions in the continental U.S. for frequencies when minimum temperature is lower than 0 °F. In the figure, it is shown that

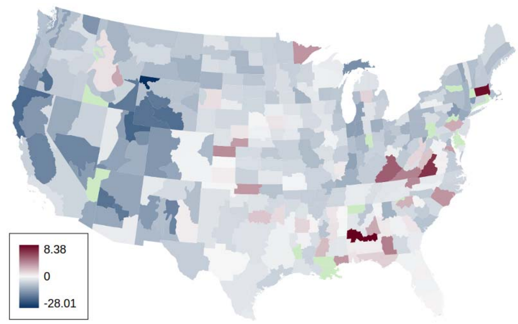

Figure 21: Distribution of difference when minimum temperature is lower than 32 °F

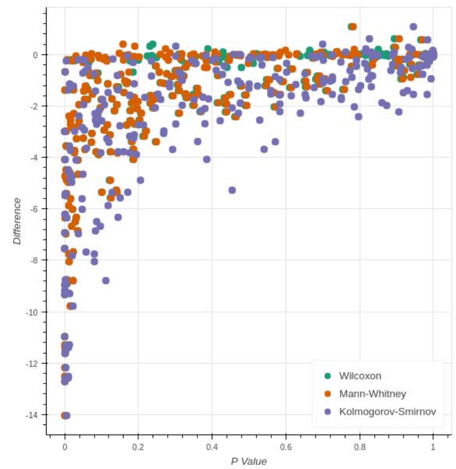

Figure 22: Distribution of difference and p-value of climate divisions in the continental U.S. when minimum tempertaure is lower than 0 °F

the number of days with extreme low minimum temperature decreases sharply.

Figure 23, Figure 24, Figure 25, and Figure 26 display p-values and differences of frequencies when minimum temperature is lower than 0 °F. The frequencies of extreme low minimum temperature mostly show declines.

Frequencies of extreme minimum temperature below 0 decrease sharply when the two time periods are compared.

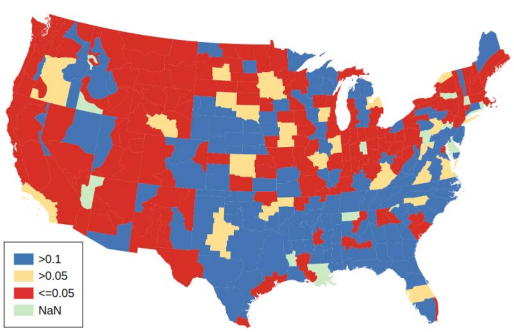

Figure 20: Distribution of Kolmogorov-Smirnov test p-value when minimum temperature is lower than 32 °F

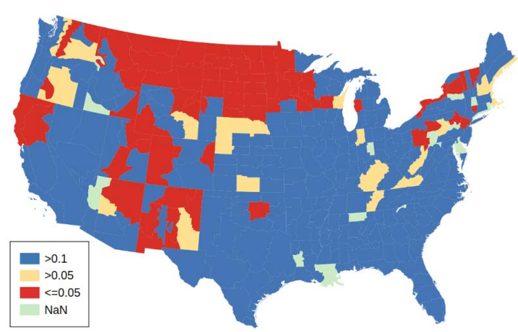

Figure 23: Distribution of Wilcoxon test p-value when minimum temperature is lower than 0 °F

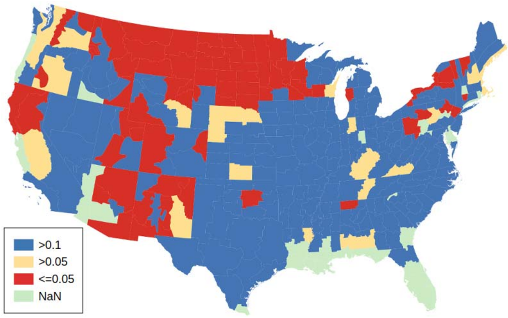

Figure 24: Distribution of Mann-Whitney test p-value when minimum temperature is lower than 0 °F

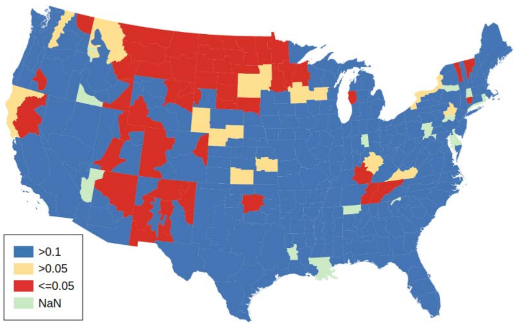

Figure 25: Distribution of p-values from Kolmogorov-Smirnov when minimum temperature is lower than 0 °F

However, inlands areas shows little change for extreme low minimum temperature. Frequencies of extreme low minimum temperature in the west coastal area decreases more rapidly than those in the east coastal area.

4.6. Annual Total Precipitation

Figure 27 is the scatter plot of the all climate divisions in the continental U.S. for the total annual precipitation. In the figure, it is shown that total annual precipitation amounts is increasing in most of the country.

Figure 28 displays the differences in total annual precipitation between the 2 time series. As shown, precipitation amounts have increased in inland areas from Oklahoma to

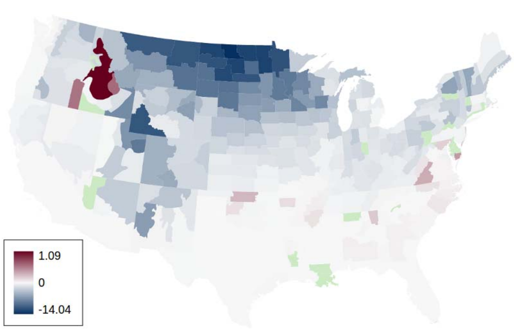

Figure 26: Distribution of difference when minimum temperature is lower than 0 °F

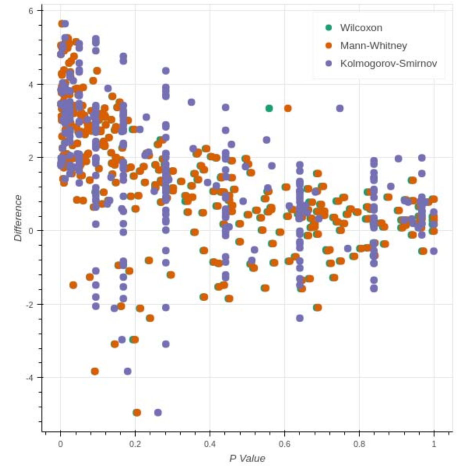

Figure 27: Distribution of difference and p-value of total annual precipitation in climate divisions in the Continental U.S.

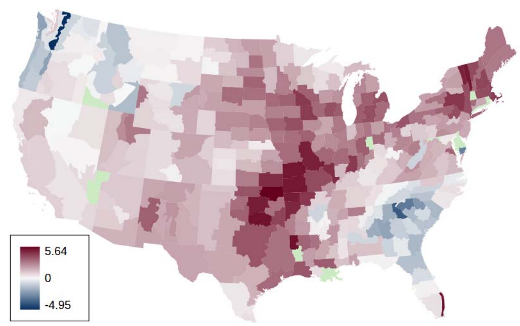

Figure 28: Distribution of difference of total annual precipitation

Missouri and in the northeast region. In contrast, annual precipitation levels decreased in northwest and southeast regions.

Overall, the frequency of occurrence of extreme high precipitation events and total annual precipitation amounts show an increasing trend. Compared to the western coastal areas and southeast region, precipitation in inland areas and northeast region show an increasing trend.

4.7. Annual Total Snowfall

Figure 29 is the scatter graph of the all climate divisions in the continental U.S. for the total annual snowfall. In the figure, it is shown that total annual snowfall decreases in the most recent time series.

Figure 30 displays differences of total annual snowfall sum. As shown, annual snowfall amounts show a decreasing trend and snowfall amounts in the Pacific Northwest, Nevada and Montana shows an increasing trend.

5. Conclusion

In summary, climate data from more than 3000 climate stations in the continental U.S. between 1946 and 2015 ahve

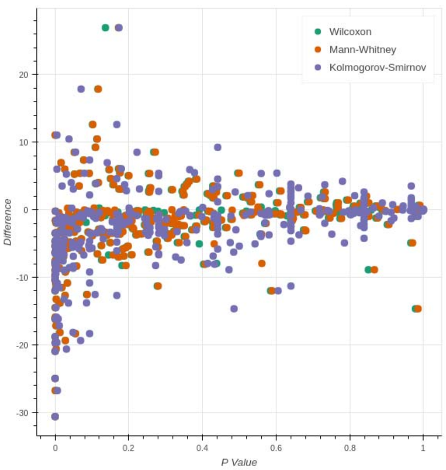

Figure 29: Distribution of difference and p-value of total annual snowfall in climate divisions in the continental U.S.

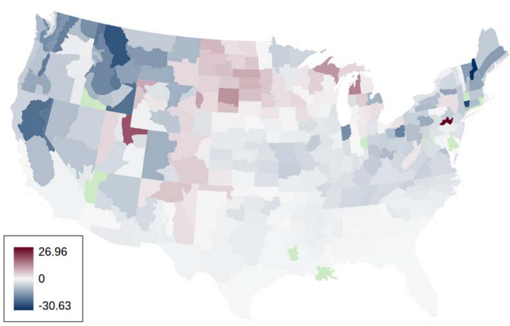

Figure 30: Distribution of difference of total annual snowfall

been analyzed to develop a new climate extremes indices dataset (TEF). Extensive statistical analysis was done to compare the frequency of occurrence of climate extremes between 2 time periods (1946-1980 and 1981-2015) and the conclusions are summarized as follows:

- The frequency of occurrence of extreme high temperatures did not yield any statistically significant trends for most of the climate divisions analyzed.
- The occurrence frequency of extreme high minimum temperatures shows an increasing trend in the most recent time period of 1981-2015.
- The frequency of occurrence of extreme low minimum temperatures shows a decreasing trend for the recent time period of 1981-2015.
- Due to the frequency of occurrence of high maximum temperatures not changing much and high minimum temperatures showing a sharp increasing trend, the diurnal temperature range narrows down across continental United States.
- The occurrence frequency of high precipitation events shows an increasing trend for 1981-2015. Total annual precipitation also shows an increasing trend, especially in inland areas and in the northeast region.
- In the Northern US states, total annual snowfall amounts show a declining trend for the time period of 1981-2015.

References

[1] J. M. Melillo, T. T. Richmond, and G. Yohe, "Climate change impacts in the united states," *Third National Climate Assessment*, 2014.

[2] R. C. Centers, "Applied Climate Information Systems," http://www.rcc-acis.org, 2004, [Online; accessed 1-July-2017].

[3] A. T. DeGaetano, W. Noon, and K. L. Eggleston, "Efficient access to climate products using acis web services," *Bulletin of the American Meteorological Society*, vol. 96, no. 2, pp. 173–180, 2015. [Online]. Available: https://doi.org/10.1175/BAMS-D-13-00032.1

[4] M. J. Menne, I. Durre, R. S. Vose, B. E. Gleason, and T. G. Houston, "An overview of the global historical climatology network-daily database," *Journal of Atmospheric and Oceanic Technology*, vol. 29, no. 7, pp. 897–910, 2012. [Online]. Available: https://doi.org/10.1175/JTECH-D-11-00103.1

[5] N. B. Guttman and R. G. Quayle, "A historical perspective of us climate divisions," *Bulletin of the American Meteorological Society*, vol. 77, no. 2, pp. 293–303, 1996.

[6] T. Karl and W. J. Koss, *Regional and national monthly, seasonal, and annual temperature weighted by area, 1895-1983*. National Climatic Data Center, 1984.

[7] B. D. Keim, A. M. Wilson, C. P. Wake, and T. G. Huntington, "Are there spurious temperature trends in the united states climate division database?" *Geophysical Research Letters*, vol. 30, no. 7, 2003.

[8] J. Allard, B. D. Keim, J. E. Chassereau, and D. Sathiaraj, "Spuriously induced precipitation trends in the southeast united states," *Theoretical and Applied Climatology*, vol. 96, no. 1-2, pp. 173–177, 2009.

[9] B. D. Keim, G. E. Faiers, R. A. Muller, J. M. Grymes, and R. V. Rohli, "Long-term trends of precipitation and runoff in louisiana, usa," *International Journal of Climatology*, vol. 15, no. 5, pp. 531–541, 1995.

[10] D. J. Leathers, A. J. Grundstein, and A. W. Ellis, "Growing season moisture deficits across the northeastern united states," *Climate Research*, vol. 14, no. 1, pp. 43–55, 2000.

[11] "Climdex - datasets for indices of climate extremes," http://www.climdex.org/indices.html, 2014.

[12] "2014 national climate assessment," http://nca2014.globalchange.gov/highlights/regions/southeast, 2014.

[13] H. J. Khamis, "The δ-corrected kolmogorov-smirnov test for goodness of fit," *Journal of statistical planning and inference*, vol. 24, no. 3, pp. 317–335, 1990.

[14] S. Bell, "Introduction to scientific research methods in geography," *The Canadian Geographer/Le Géographe canadien*, vol. 51, no. 3, pp. 402–403, 2007.

[15] B. Bhattacharya and D. Habtzghi, "Median of the p value under the alternative hypothesis," *The American Statistician*, vol. 56, no. 3, pp. 202–206, 2002.

[16] R. Hubbard, "Alphabet soup blurring the distinctions betweenp's anda's in psychological research," *Theory & Psychology*, vol. 14, no. 3, pp. 295–327, 2004.

[17] T. Sellke, M. Bayarri, and J. O. Berger, "Calibration of ρ values for testing precise null hypotheses," *The American Statistician*, vol. 55, no. 1, pp. 62–71, 2001.

[18] V. E. Johnson, "Revised standards for statistical evidence," *Proceedings of the National Academy of Sciences*, vol. 110, no. 48, pp. 19 313–19 317, 2013.

[19] D. Colquhoun, "An investigation of the false discovery rate and the misinterpretation of p-values," *Open Science*, vol. 1, no. 3, p. 140216, 2014.