Comparison of Nonparametric Regression and Correlation Methods for Precipitation Time Series

Elisabeth G. Ingram[1] and John J. Ramirez-Avila[2]

[1]Doctoral Student, Watersheds and Water Quality Research Lab, Bagley College of Engineering, Mississippi State University, Mississippi State, MS, USA.
[2]Assistant Professor, Watersheds and Water Quality Research Lab, Civil and Environmental Engineering Department, Bagley College of Engineering, Mississippi State University, Mississippi State, MS, USA.

***

Abstract – *It is important for society to have an understanding of the changing trends of precipitation, which impact surface water and groundwater available for potable use, irrigation and industries. Precipitation trends vary across the globe, and they are beneficial for planning and future economic development of any given location. There are numerous methods available for determining trends and predicting changes in precipitation. In this study, four different nonparametric regression and correlation methods were applied to annual precipitation data sets for 33 stations within the Mobile – Tombigbee Basin, each having at least 40 complete seasonal years of data available from the time period of October 1, 1940 to September 30, 2015. The Mann-Kendall, Kendall's tau, Spearman's rho and Kendall-Theil Robust Line methods were used for analysis of the annual precipitation data. Each method was applied to all stations, and the results of the different nonparametric regression and correlation methods were compared for analysis. The results show that for the majority of the stations, the trend was found to be the same regardless of which nonparametric regression method was used for analysis. Only one station had varying results between the different nonparametric methods. This research provides four nonparametric regression and correlation coefficient methods that can be applied to data sets of annual precipitation data, which can be used to predict how precipitation can be expected to change for any area.*

Key Words: Nonparametric, Mann-Kendall, Kendall's tau, Spearman's rho, Kendall Theil Robust Line, Precipitation Trends, Mobile - Tombigbee

1.INTRODUCTION

Understanding the trends of precipitation is an essential component of future planning for society and ecosystems, as both are dependent on the availability of water for daily use [1, 2]. Researchers are analyzing the changing climate, including precipitation, in many ways using a variety of different methods and data filters. An accurate prediction of precipitation trends can play an important role in a country's future economic development [3, 4], especially in areas where agriculture is predominant, and industries may be more susceptible to changes in climate. Changing trends of precipitation can impact surface water and groundwater available for potable use, irrigation and industries. Water availability, which determines what types of plants and animals can survive in a location, is also impacted by changes in precipitation [5]. Precipitation trends can be used in trend analysis in flooding and droughts, which impact areas throughout the world each year. Unfortunately, precipitation trends have been found to vary across the globe and within the United States [6]. The U.S. Global Change Research Program, in the report, *Global Climate Change Impacts in the United States* [6], divides the contiguous United States into six (6) regions for analysis. However, due to the variability of climate within a region, it is recommended to further divide the regions into smaller areas for precipitation trend analysis [6, 7].

According to Rahman and Begum, predicting precipitation trends using time series data is much more difficult than predicting temperature trends [8]. Melillo, Richmond, and Yohe [9] noted that models are in complete agreement in showing decreases in precipitation in the subtropics and increases in precipitation at higher altitudes. Consistently, in the Fifth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC) published in 2013 [7], it was reported that zonal mean precipitation will very likely increase in high latitudes, as well as some of the mid latitudes, and it will likely decrease in the subtropics. The frequency and intensity of heavy precipitation events over land was determined to likely increase on average in the near future. However, as noted in the Fifth IPCC report, the large-scale trends contain considerable spatial variability and can be considered to vary within regions [7]. The contrast between wet and dry areas is expected to increase both in the U.S. and globally, consistent with the motto that in the future, the wet areas will get wetter and the dry areas will get drier [9].

Precipitation trends have been studied for many areas around the globe, using a variety of different statistical test methods, classified as parametric and nonparametric tests. Parametric tests are hypothesis tests which assume the data has a normal distribution, allowing for a simple linear or bell curve regression analysis. Helsel and Hirsch [10] noted that parametric tests are of questionable value when applied to water resources data, as the data is typically not normal or symmetric. Nonparametric tests use independent data, making no assumptions regarding the data distribution type and better tolerate any outliers in the data. Therefore, nonparametric tests are widely used in water resources data analysis [10, 11]. Several nonparametric tests are available for working with time series trends, as nonparametric tests can determine if there is a monotonic trend in a time series data set, as well as how strong the relationship is between the variables [10]. Monotonic relationships are present when there is a gradual change over time that is consistent in any direction, and a monotonic trend is present when the variable being studied, such as precipitation, is determined to increase or decrease as time increases. Step trends, in which there is an abrupt shift at a specific point in time, are unlikely in precipitation time series data sets, due to the fact that precipitation changes are not impacted by any man-made operation or structure, such as a dam. Various nonparametric regression and correlation coefficient calculations have been used for precipitation trend series analysis, and it is common to use more than one method for comparison purposes.

The Mobile - Tombigbee Basin includes areas in Mississippi and Alabama and consists primarily of forested and agricultural land. The basin contains rivers which are used for navigation, wildlife mitigation and recreational activities, such as fishing, swimming and boating [12, 13]. Located within the Mobile - Tombigbee Basin, the Tombigbee River provides one of the principal routes of commercial navigation in the southern United States, as it is navigable along much of its length through locks and dams. As part of the Tennessee-Tombigbee Waterway, it is heavily traveled to transport manufactured goods [13, 14]. Many industries along the rivers and within the basin area rely on the waterways to remain navigable, as trade along the waterways continues to increase each year [15], and changes in precipitation impact the navigability of the rivers.

Mississippi and Alabama are both impacted by many natural disasters, including floods, droughts, hurricanes, tropical storms and tornadoes, primarily due to their close proximity to the Gulf of Mexico [16, 17]. Not only can the natural disasters harm humans, they pose serious risks to society and create challenges for the economic development. A prediction for changes in future precipitation would assist in planning and allow adjustments to procedures and policies for industries within the Mobile - Tombigbee Basin, to ensure that rivers remain navigable and industries are prepared for any changes caused by varying precipitation.

During research of precipitation trend analysis in the Mobile - Tombigbee Basin, various nonparametric regression and correlation methods were evaluated. No clear determination could be found as to which nonparametric method should be applied to time series precipitation data to analyze precipitation trends. Therefore, the nonparametric regression and correlation methods were performed on the same data sets to compare results and determine how the results differed among the various methods. This study can be used to assist managers, planners and others in determining a nonparametric regression analysis and correlation method for performing trend analysis of precipitation time series data for an area.

2. METHODOLOGY

This study investigates the variability of results using four different nonparametric tests on precipitation time series trend analyses for 33 stations within the Mobile - Tombigbee Basin. Daily precipitation data for a 75-year time period (1940-2015) was used to review and compare the results from performing different nonparametric regression and correlation analyses. Data used for this study was obtained from the National Oceanic and Atmospheric Administration (NOAA) [18]. Each station included in the study had a minimum of 40 complete seasonal years of data available for the time period from October 1, 1940 through September 30, 2015. Figure 3 shows the NOAA stations within the Mobile - Tombigbee Basin used for this study. Table 1 includes the information for each station shown in Figure 3.

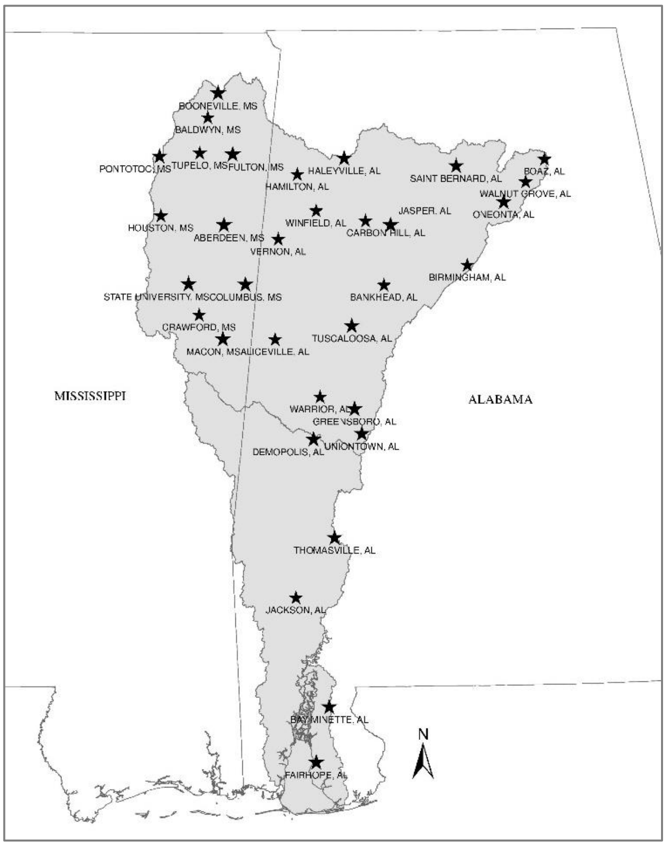

Fig -1: NOAA Stations Included in Study

Table -1: Summary of NOAA Stations Included in this Study

NOAA Station Number(s)[1]	NOAA Station Location	Number of Seasonal Years with Complete Precipitation Data Available
10178	ALICEVILLE, AL	67
220021	ABERDEEN, MS	40
220378	BALDWYN, MS	62
14855; 10505	BANKHEAD, AL	63
10583	BAY MINETTE, AL	66
13876	BIRMINGHAM, AL	75
10957	BOAZ, AL	58
220955	BOONEVILLE, MS	68
11377	CARBON HILL, AL	52
221870; 221880	COLUMBUS, MS	69
22046	CRAWFORD, MS	66
12240; 12245	DEMOPOLIS, AL	66
12813	FAIRHOPE, AL	66
223208	FULTON, MS	64
13511	GREENSBORO, AL	59
13620	HALEYVILLE, AL	66
13644; 13645	HAMILTON, AL	72
224265	HOUSTON, MS	54
14193; 14192	JACKSON, AL	60
14225; 14226	JASPER, AL	48
225361; 225366	MACON, MS	56
16121	ONEONTA, AL	62
227106	PONTOTOC, MS	69
17157	SAINT BERNARD, AL	68
228374	STARKVILLE, MS	66
18178; 23802	THOMASVILLE, AL	71
18385	TUSCALOOSA, AL	70
229000; 93862	TUPELO, MS	61
18446	UNIONTOWN, AL	57
18517	VERNON, AL	48
18648	WALNUT GROVE, AL	42
12742	WARRIOR, AL	62
18998	WINFIELD, AL	61

[1] *Stations with more than one station number listed indicate that the monitoring station was relocated and assigned a new station number during the 75-year time period from October 1, 1940 to September 30, 2015.*

2.1 Study Area. Mobile - Tombigbee Basin

The Mobile - Tombigbee Basin encompasses approximately 57,000 square kilometers and is a region defined by the United States Geological Survey (USGS) as Hydrologic Unit Code (HUC) 0316 [19], as shown in the map in Figure 2.

Measuring 644 kilometers long, the Tombigbee River is an important feature within the Mobile - Tombigbee Basin. As part of the construction of the Tennessee-Tombigbee Waterway, the Tombigbee River was impounded and became navigable for travel by barges and recreational boats. It receives water from the Black Warrior River and later joins the Alabama River approximately 48 kilometers (30 miles) north of Mobile, to form the Mobile River [12]. Figure 3 shows the stream networks within the Mobile - Tombigbee Basin, provided by the USGS [19].

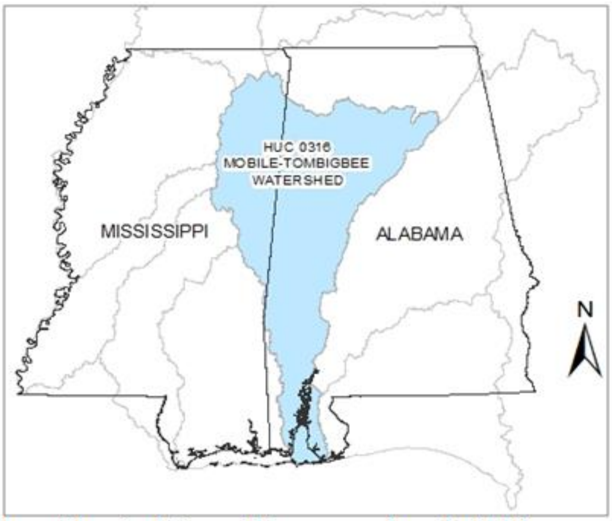

Fig -2: Map of the Mobile - Tombigbee Basin, defined by HUC 0316, located in Alabama and Mississippi.

Fig -3: Provided by the USGS, this map shows the stream networks located within the Mobile - Tombigbee Basin, shown outlined in the green line and denoted by HUC codes 031601 and 031602 [19].

2.2 Nonparametric Regression and Correlation Coefficient Calculations

Each of the four nonparametric regression and correlation methods discussed in this section were applied to all 33 stations included in this study for analysis of the annual precipitation during the 75-year time period. Results are summarized for comparison and analysis.

2.2.1 Mann-Kendall Test

The Mann-Kendall test [20, 21, 22] appears to be the most common nonparametric test used by researchers for studying hydrologic time series trends [3], perhaps because the test does not require the data to meet an assumption of normality, in which the data roughly fits a bell curve shape [10, 20]. The test compares the relative magnitudes of sample data rather than the data values themselves [21], and it can be used to analyze data collected over a long period of time for consistently increasing or decreasing monotonic trends of a Y-value when the X variable is time [10, 20, 22]. Results of the Mann-Kendall test, referred to as the Mann-Kendall Statistic (S), can be used in conjunction with a variance of S (VAR(S)) and standardized test statistics (Z), to provide an indication of whether a trend exists and whether the trend is positive or negative. However, it does not indicate the magnitude of the slope or estimate the trend line itself, even when a trend is present. When using the Mann-Kendall test, a larger number of data points will allow the test to find a true trend, if one exists, as opposed to a trend found by chance if only a small number of data points are used [20, 22].

The null hypothesis (H_0) for the Mann-Kendall test is that there is no monotonic trend in the series. The alternate hypothesis (H_1) for the study of precipitation trend over time is that there is an increasing or decreasing trend in annual precipitation. To calculate S, the data must first be listed in the order in which it was collected over time: $x_1, x_2, ..., x_n$, where x_i is the datum at time i. Next, the sign of all possible differences $x_j - x_k$, where $j > k$ are determined. The differences are $x_2 - x_1, x_3 - x_1, \ldots, x_n - x_1, x_3 - x_2, x_4 - x_2, \ldots, x_n - x_{n-2}, x_n - x_{n-1}$, and the mathematical equations used for calculating the differences are shown below:

$$sign(x_j - x_k) = \begin{cases} 1 \; if \; x_j - x_k > 0 \\ 0 \; if \; x_j - x_k = 0 \\ -1 \; if \; x_j - x_k < 0 \end{cases}$$

Once the differences are calculated, the net result of all such increments and decrements yields the final value of S, which is therefore calculated by:

$$\mathrm{S} = \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} sign(x_j - x_k)$$

An increasing trend is indicated with a very high positive value of S, and a decreasing trend is indicated by a very low negative value of S. It is necessary to compute the probability associated with S and the sample size, n, to statistically quantify the significance of the trend. VAR(S) is used for data sets with more than 10 values, provided that the data set does not have many tied values [22]. VAR(S) can be calculated using the following equation:

$$\mathrm{VAR(S)} = \frac{1}{18}\left[n(n-1)(2n+5) - \sum_{p=1}^{g} t_p(t_p - 1)(2t_p + 5)\right]$$

In which n is the number of data points, g is the number of tied groups in which a set of sample data has the same value, and t_p is the number of data points in the p^{th} group. To determine the significance of the results, a 95% confidence level was recommended, equivalent to a 5% level of significance, also noted as alpha (α). Data sets with less than a 95% confidence level are not considered to be statistically significant, and therefore no statistically significant trend is determined to be present [10]. H_0 cannot be rejected when the probability value (p-value) is greater than the significance level of 0.05, meaning that if the p-value is greater than 0.05, no statistically significant trend is considered to be present. For this study, the precipitation data was input into the software XLSTAT, which calculated S, VAR(S), and the p-value of the data set.

2.2.3 Kendall's tau Correlation

Following the Mann-Kendall calculations, subsequent calculation of Kendall's tau (τ) coefficient permits a comparison of the strength of correlation between the X and Y variables [10, 23]. Developed in 1938 by Maurice Kendall [23], the value of tau ranges from -1 to 1 and demonstrates the determined trend, as well as the strength of the trend, with stronger trends being further from zero. Essentially, τ is a scaled measure of S [24]. As with other nonparametric methods, the τ correlation coefficient is not impacted by the magnitude of extreme values, and typically can be used with as few as 10 observations within a dataset [25].

If S has been calculated for a data set, calculating τ involves just one formula, using the value of S and the number of observations within the data set, n. The equation to be used is shown below:

$$\tau = \mathrm{S} \div [n(n-1)/2]$$

τ can also be calculated without the value of S. To perform this calculation, the time series data sets must first be listed in chronological order, at which time the number of concordances and discordances can be determined, as well as the number of ties, in the paired observations. Two observations are considered concordant if they are in the same order with respect to each variable. For example, using two observations in the data set (x_i, y_i) and (x_j, y_j), as time (considered as the x-value) increases in the data set, the y-value will also increase, meaning that the observations are concordant. The same observations would be considered discordant if the y-value is in reverse order from the x-value, such as when x increases, y decreases. Lastly, the two observations are considered tied if $x_i = x_j$ and/or $y_i = y_j$. The total number of pairs that will be generated for a sample size is calculated by

$$N = \binom{n}{2} = \frac{1}{2}n(n-1)$$

N can be broken down into these five quantities:

$$N = P + Q + X_0 + Y_0 + (XY)_0$$

in which P is the number of concordant pairs, Q is the number of discordant pairs, X_0 is the number of pairs tied only with the x variable, Y_0 is the number of pairs tied only with the y variable, and $(XY)_0$ is the number of pairs tied on both X and Y. Using time series data, with daily precipitation totals, the x variable should not be tied at any point, resulting in no values for X_0 and $(XY)_0$. The value of τ can be calculated to measure the association between the *x* and *y* variables, calculated using the following formula:

$$\tau = \frac{P - Q}{\sqrt{(P + Q + X_0)(P + Q + Y_0)}}$$

The result of this equation will be a value between -1 and 1. Values of zero, or near to zero, will signify no trend for the data series. Statistical significance was determined and incorporated into the analysis of the results for each data set. Both τ and the p-value were calculated using the XLSTAT software, which required the precipitation time series data to be input into the software for each station being analyzed.

2.2.3 Spearman's Correlation

Spearman's rank correlation coefficient rho (ρ) is a nonparametric coefficient that can be used to test for monotonic trends and measure the strength of association between two variables, much like τ. It is a common numerical measure of the degree of linear association between two variables [26]. A value of 1 would indicate a perfect positive correlation, and a value of -1 would indicate a perfect negative correlation. Values near zero are calculated when no trend is present in the data set. An advantage of the Spearman's rho rank correlation coefficient method is that a normal distribution of the *x* and *y* values in the data set is not required. Also, the *x* and *y* values can be continuous or ordinal. Spearman's rho calculations assume no seasonal trends are present, which typically require more sophisticated evaluations, and should only be used on data sets with at least 20 observations [10]. Calculation of ρ is often used in combination with the Mann-Kendall test for comparison purposes, and it is less commonly used independently to access the significance of trends in hydro-meteorological data [27].

In order to calculate ρ, each observation (x, y) within the data set with a sample size *n*, is converted to a rank, and assigned a rank value, $rg(x_i)$, $rg(y_i)$. The difference (d_i) in the ranks of each observation are calculated as shown below.

$$d_i = rg(x_i) - rg(y_i)$$

ρ can then be calculated using the following formula:

$$\rho = 1 - \frac{6\sum d_i^2}{n\,(n^2 - 1)}$$

The statistical significance of the data set should be determined and included in the analysis of the results of the ρ correlation coefficient calculation. This calculation is not sensitive to any outliers in the data set and is best used with a data set containing at least ten observations [24].

A calculator for finding the value of ρ, given the precipitation time series data for each station, was provided online by Social Science and used for this study [26]. The software also calculated the p-value, allowing for interpretation to determine if the results could be considered as statistically significant.

2.2.4 Kendall-Theil Robust Line

The Kendall-Theil Robust Line Method (KTR), also referred to as the Theil-Sen Estimator, Sen's Slope Estimator or Kendall Robust Line-fit method, was developed originally in 1950 by Henri Theil [28]. Later, in 1968, the process was refined by Pranab K. Sen [25], and it was again further developed later by Maurice Kendall in 1975 [22]. Theil originally introduced the method using a set of two-dimensional points to determine the median of the slopes, as calculated by all pairs of sample points within a data set [28]. Sen later extended the definition of the nonparametric test to handle the case in which two data points have the same x-coordinate [25]. As the method was developed, it has also become more frequently used with the development of computer software available to perform the analysis. The KTR line is resistant to the effects of outliers in the dataset, as well as non-normality. Helsel and Hirsch [10] note that robust methods, such as the KTR line, are not sensitive to values at the tails of the distributions, and therefore, minimize the effect of assumptions about data below detection limits and the effect of outliers on the determination of relations between variables. The KTR line method should be used to determine how steeply a trend is decreasing or increasing over time. However, the results do not indicate if a statistically significant trend is present, so this test should be used in conjunction with another nonparametric test, such as the Mann-Kendall, to determine if a trend is present in the data available, as well as the significance level and p-value of the data set results.

When Theil first used the method to estimate the slope of a nonlinear line, a slope $\Delta y/\Delta x$ was computed for each comparison of data pairs, and the median of all possible pairwise slopes is taken as the nonparametric estimated slope. The fitted line generated always goes through the point $(x_{\text{median}}, y_{\text{median}})$ [28].

To calculate the estimated slope of the linear line using the KTR method, the median (m) of the slopes for each set of points (x_i, y_i) is calculated using the formula

$$m_{ij} = \frac{(y_j - y_i)}{(x_j - x_i)} \quad \text{for } i = 1 \text{ to } n - 1 \text{ and } j = 2 \text{ to } n.$$

The number of possible slopes (N_p) between data pairs is calculated using the equation

$$N_p = \frac{n(n-1)}{2}$$

After all slopes are calculated, the slope estimates (m_{ij}) are sorted and ranked from lowest to highest. If N_p is an odd number, the median slope is selected as the middle value of the array. In situations where N_p is an even number, the median is calculated as the arithmetic average of the two center points. Once the slope m has been determined, the y-intercept (b) can be calculated, as the median of all values.

$$b = y_i - mx_i$$

It should be noted that slopes are calculated only for points having distinct x values. Situations in which two data points have the same x values should be removed. The slope determined by the KTR Line Method should be significantly different from zero when τ is significantly different from zero [26].

The USGS offers a software for calculating the equation of the Kendall-Theil Robust Line and other applicable values for the time series data sets. This software, and the techniques and methods involved, is available for download online, and details regarding the process and calculations are found in the publication written by Gregory E. Granato [29]. The KTR Software developed by USGS was used for this study.

3. RESULTS

The preliminary analysis for this study included determining S, as well as its variance, τ, ρ and the slope of the KTR Line for the time series of annual precipitation from October 1, 1940 through September 30, 2015, for each NOAA station listed previously in Table 1. For each calculation, it was noted if the result was an increasing or decreasing trend in the annual precipitation, and the p-value was also noted in the results for the Mann-Kendall, Kendall's tau and Spearman's rho tests. Results were considered significant when the p-value was less than or equal to 0.05.

Table 2 lists the results of the Mann-Kendall test, including the S, VAR(S), the p-value and the trend found in the total annual precipitation, increasing or decreasing. Of the 33 stations used for analysis in this study, nine (9) of them had statistically significant results. These stations were Aliceville, AL, Baldwyn, MS, Bay Minette, AL, Haleyville, AL, Hamilton, AL, Jasper, AL, Starkville, MS, Walnut Grove, AL and Winfield, AL.

The results of Kendall's tau test are listed in Table 3, which includes the calculated value of τ, as well as the p-value and the apparent trend of the annual precipitation. These calculations for τ resulted in the same determined trends as the Mann-Kendall test, and the same nine (9) stations with a p-value less than 0.05.

Table -2: Summary of Results of the Mann-Kendall Test

NOAA Station Location	Mann-Kendall Statistic (S)	Variance of S VAR(S)	p-value	Trend (Increasing or Decreasing)
Aberdeen, MS	47	34147.67	0.799	Increase
Aliceville, AL	216	7366.67	0.012	Increase
Baldwyn, MS	381	27104.33	0.021	Increase
Bankhead, AL	297	28427.00	0.078	Increase
Bay Minette, AL	415	32651.67	0.022	Increase
Birmingham, AL	299	47791.67	0.171	Increase
Boaz, AL	215	22223.67	0.149	Increase
Booneville, MS	248	35688.67	0.189	Increase
Carbon Hill, AL	12	16059.33	0.925	Increase
Columbus, MS	282	37275.33	0.144	Increase
Crawford, MS	277	32651.67	0.125	Increase
Demopolis, AL	289	32651.67	0.110	Increase
Fairhope, AL	203	32651.67	0.261	Increase
Fulton, MS	316	29792.00	0.067	Increase
Greensboro, AL	42	23382.67	0.784	Increase
Haleyville, AL	511	32651.67	0.005	Increase
Hamilton, AL	598	42316.00	0.004	Increase
Houston, MS	108	17966.00	0.420	Increase
Jackson, AL	232	24583.33	0.139	Increase
Jasper, AL	224	12658.67	0.047	Increase
Macon, MS	156	20020.00	0.270	Increase
Oneonta, AL	-95	27104.33	0.564	Decrease
Pontotoc, MS	179	37274.33	0.354	Increase
Saint Bernard, AL	164	35688.67	0.385	Increase
Starkville, MS	437	32651.67	0.016	Increase
Thomasville, AL	-109	40588.33	0.588	Decrease
Tupelo, MS	171	38908.33	0.386	Increase
Tuscaloosa, AL	108	25823.33	0.502	Increase
Uniontown, AL	16	21102.67	0.912	Increase
Vernon, AL	-186	12658.67	0.100	Decrease
Walnut Grove, AL	291	8514.33	0.001	Increase
Warrior, AL	321	27104.33	0.051	Increase
Winfield, AL	347	25822.33	0.031	Increase

The results of the Spearman's rho calculation and the corresponding p-value are summarized in Table 4. The same nine stations discussed above resulted in p-values less than 0.05, determined as rejecting the null hypothesis.

Table 5 summarizes the results of the KTR Line test, and the KTR equation calculated for each data set. The level of confidence and significance were not calculated with the KTR software.

Table -3: Summary of Results of Kendall's tau

NOAA Station Location	Kendall's tau (τ)	p-value	Trend (Increasing or Decreasing)
Aberdeen, MS	0.021	0.799	Increase
Aliceville, AL	0.277	0.012	Increase
Baldwyn, MS	0.201	0.021	Increase
Bankhead, AL	0.152	0.078	Increase
Bay Minette, AL	0.193	0.022	Increase
Birmingham, AL	0.108	0.171	Increase
Boaz, AL	0.130	0.149	Increase
Booneville, MS	0.109	0.189	Increase
Carbon Hill, AL	0.009	0.925	Increase
Columbus, MS	0.120	0.144	Increase
Crawford, MS	0.129	0.125	Increase
Demopolis, AL	0.135	0.110	Increase
Fairhope, AL	0.095	0.261	Increase
Fulton, MS	0.157	0.067	Increase
Greensboro, AL	0.025	0.784	Increase
Haleyville, AL	0.238	0.005	Increase

Hamilton, AL	0.234	0.004	Increase
Houston, MS	0.075	0.420	Increase
Jackson, AL	0.131	0.139	Increase
Jasper, AL	0.199	0.047	Increase
Macon, MS	0.101	0.270	Increase
Oneonta, AL	-0.050	0.564	Decrease
Pontotoc, MS	0.076	0.354	Increase
Saint Bernard, AL	0.072	0.385	Increase
Starkville, MS	0.204	0.016	Increase
Thomasville, AL	-0.044	0.588	Decrease
Tupelo, MS	0.071	0.386	Increase
Tuscaloosa, AL	0.059	0.502	Increase
Uniontown, AL	0.010	0.912	Increase
Vernon, AL	-0.165	0.100	Decrease
Walnut Grove, AL	0.338	0.001	Increase
Warrior, AL	0.170	0.051	Increase
Winfield, AL	0.190	0.031	Increase

After reviewing and comparing the determined trend directions using the four nonparametric regression and coefficient methods, it was noted that for 32 of the 33 station locations used in the analysis, the determined trend for annual precipitation was the same for all methods and tests used. Only one station, Uniontown, AL, had different trend directions using different nonparametric tests. The data set of annual precipitation for this station showed an increasing trend in annual precipitation using the Mann-Kendall, Kendall's tau and KTR Line methods, but it resulted in a decreasing trend in annual precipitation when using the Spearman's rho calculation. The p-value for this station was high for all methods/tests, determined to be 0.912 for the Mann-Kendall and Kendall's tau tests and 0.993 for the Spearman's rho test. With such a high p-value, almost to the maximum value of 1.0, it should be interpreted that the H_0 cannot be rejected, meaning that no trend is present in the data set analyzed. The high level of the p-value for the Uniontown, AL station could indicate that the trend is very close to static, meaning little or no change in total annual precipitation over time. It could also signify that the trend is increasing or decreasing at such a low rate it cannot be consistently calculated. If the data set for Uniontown, AL were removed from the analysis, all remaining data sets for the stations used were in agreement for the resulting trend of the annual precipitation, regardless of the result of the p-value for the data set.

Table -4: Summary of Results of Spearman's rho

NOAA Station Location	Spearman's rho (ρ)	p-value	Trend (Increasing or Decreasing)
Aberdeen, MS	0.02049	0.86926	Increase
Aliceville, AL	0.4516	0.00344	Increase
Baldwyn, MS	0.290977	0.02177	Increase
Bankhead, AL	0.21237	0.09473	Increase
Bay Minette, AL	0.25828	0.03627	Increase
Birmingham, AL	0.15673	0.17933	Increase
Boaz, AL	0.17358	0.19254	Increase
Booneville, MS	0.15378	0.21055	Increase
Carbon Hill, AL	0.01708	0.90437	Increase
Columbus, MS	0.16368	0.17898	Increase
Crawford, MS	0.19194	0.1226	Increase
Demopolis, AL	0.20927	0.09173	Increase
Fairhope, AL	0.12556	0.31511	Increase
Fulton, MS	0.22756	0.07054	Increase
Greensboro, AL	0.03761	0.77732	Increase
Haleyville, AL	0.32976	0.00685	Increase
Hamilton, AL	0.19098	0.13062	Increase
Houston, MS	0.12438	0.37021	Increase
Jackson, AL	0.1825	0.16283	Increase
Jasper, AL	0.28561	0.04909	Increase
Macon, MS	0.1473	0.27865	Increase
Oneonta, AL	-0.06248	0.62952	Decrease
Pontotoc, MS	0.08827	0.47075	Increase
Saint Bernard, AL	0.08085	0.5122	Increase
Starkville, MS	0.28246	0.02157	Increase
Thomasville, AL	-0.06767	0.57497	Decrease
Tupelo, MS	0.09387	0.43956	Increase
Tuscaloosa, AL	0.7472	0.5671	Increase
Uniontown, AL	-0.00133	0.99217	Decrease
Vernon, AL	-0.22764	0.11968	Decrease
Walnut Grove, AL	0.48043	0.00128	Increase
Warrior, AL	0.24751	0.05243	Increase
Winfield, AL	0.2865	0.02519	Increase

4. CONCLUSIONS

This study compared the variability of annual precipitation trends for 33 stations within the Mobile - Tombigbee Basin using four nonparametric regression and coefficient correlation methods (Mann-Kendall Statistic, Kendall's tau, Spearman's rho and Kendall Theil Robust Line equation). After performing the calculations of the four methods using data sets of the annual precipitation totals for stations within the Mobile - Tombigbee Basin, it was determined that in the majority of cases (97%), the same trend and level of significance was determined, regardless of the method used. The need for an additional study is present, to further investigate the inconsistency in the results for the one station that did not have agreement in the results between the various methods.

Table -5: Summary of Results of the KTR Line Method

NOAA Station Location	KTR Line Equation	Trend (Increasing or Decreasing)
Aberdeen, MS	$Y = 53.630 + 0.014 * X$	Increase
Aliceville, AL	$Y = 40.907 + 0.268 * X$	Increase
Baldwyn, MS	$Y = 45.564 + 0.160 * X$	Increase
Bankhead, AL	$Y = 50.483 + 0.127 * X$	Increase
Bay Minette, AL	$Y = 57.252 + 0.191 * X$	Increase
Birmingham, AL	$Y = 51.854 + 0.064 * X$	Increase
Boaz, AL	$Y = 52.920 + 0.081 * X$	Increase
Booneville, MS	$Y = 51.133 + 0.088 * X$	Increase
Carbon Hill, AL	$Y = 56.361 + 0.007 * X$	Increase
Columbus, MS	$Y = 49.061 + 0.107 * X$	Increase

Crawford, MS	Y = 48.940 + 0.086 * X	Increase
Demopolis, AL	Y = 50.605 + 0.096 * X	Increase
Fairhope, AL	Y = 60.834 + 0.096 * X	Increase
Fulton, MS	Y = 51.779 + 0.114 * X	Increase
Greensboro, AL	Y = 52.996 + 0.015 * X	Increase
Haleyville, AL	Y = 52.515 + 0.178 * X	Increase
Hamilton, AL	Y = 49.797 + 0.182 * X	Increase
Houston, MS	Y = 52.765 + 0.062 * X	Increase
Jackson, AL	Y = 52.493 + 0.084 * X	Increase
Jasper, AL	Y = 50.593 + 0.218 * X	Increase
Macon, MS	Y = 52.619 + 0.075 * X	Increase
Oneonta, AL	Y = 56.844 + -0.044 * X	Decrease
Pontotoc, MS	Y = 53.336 + 0.056 * X	Increase
Saint Bernard, AL	Y = 54.085 + 0.048 * X	Increase
Starkville, MS	Y = 48.688 + 0.149 * X	Increase
Thomasville, AL	Y = 57.264 + -0.038 * X	Decrease
Tupelo, MS	Y = 52.563 + 0.052 * X	Increase
Tuscaloosa, AL	Y = 50.738 + 0.032 * X	Increase
Uniontown, AL	Y = 50.643 + 0.009 * X	Increase
Vernon, AL	Y = 65.572 + -0.224 * X	Decrease
Walnut Grove, AL	Y = 43.893 + 0.255 * X	Increase
Warrior, AL	Y = 47.391 + 0.119 * X	Increase
Winfield, AL	Y = 52.216 + 0.174 * X	Increase

The results of the trend analysis methods indicate that increasing or decreasing trends in precipitation time series data sets agree between the various methods. Researchers should therefore be confident that time series trend analysis can be performed using any of the four nonparametric regression and coefficient correlation methods used, and findings among the methods would be consistent. However, the use of at least two methods in any study is recommended, for confirmation of results and comparison of not only the resulting trend for the data set, but the rate at which the trend is increasing or decreasing. The level of confidence and statistical significance were also discussed and compared between the various methods. Again, the results were consistent between methods, with the same stations resulting in a level of p-value less than 0.05 for each method. It is at the researcher's discretion the interpretation of the p-value, but typically in time series trends, the value of 0.05 for alpha is considered for use, and any variations of this should be noted prior to testing of the data sets. Further incorporation of other nonparametric regression and correlation methods could be added to compare with those used in this study for additional confirmation or investigation into the results and analysis of a trend in the time series data sets.

REFERENCES

[1] D.R. Easterling, K.E. Kunkel, J.R. Arnold, T. Knutson, A.N. LeGrande, L.R. Leung, R.S. Vose, D.E. Waliser, and M.F. Wehner, 2017: Precipitation change in the United States. In: *Climate Science Special Report*: Fourth National Climate Assessment, Volume I [Wuebbles, D.J., D.W. Fahey, K.A. Hibbard, D.J. Dokken, B.C. Stewart, and T.K. Maycock (eds.)]. U.S. Global Change Research Program, Washington, DC, USA, pp. 207-230, doi: 10.7930/J0H993CC.

[2] A.K. Misra, "Climate change and challenges of water and food security." International Journal of Sustainable Built Environment, Vol. 3, No. 1, pp. 153-165, 2014.

[3] I. Aahmad, D. Tang, T.F. Wang, M. Wang, B. Wagan, "Precipitation trends over time using Mann-Kendall and Spearman's rho tests in Swat River Basin, Pakistan," Advances in Meteorology, vol. 2015, 2015, Article ID 431860. doi:10.1155/2015/431860.

[4] P.J. Edwards, K.W. Williard and J.E. Schoonover, Fundamentals of Watershed Hydrology. Journal of Contemporary Water Research & Education, 154: pp. 3-20, 2015. doi:10.1111/j.1936-704X.2015.03185.x.

[5] U.S. Environmental Protection Agency (EPA), "Climate Change Indicators in the United States", Fourth Edition. EPA 430-R-16-004, 2016.

[6] U.S. Global Change Research Program (USGCRP), *Global Climate Change Impacts in the United States*. T.R. Karl, J. M. Melillo, and T. C. Peterson (eds.). United States Global Change Research Program, Cambridge University Press, New York, NY, USA, 2009.

[7] B. Kirtman, S.B. Power, J.A. Adedoyin, G.J. Boer, R. Bojariu, I. Camilloni, F.J. Doblas-Reyes, A.M. Fiore, M. Kimoto, G.A. Meehl, M. Prather, A. Sarr, C. Schär, R. Sutton, G.J. van Oldenborgh, G. Vecchi and H.J. Wang., "Near-term Climate Change: Projections and Predictability. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change" [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013.

[8] M.D. Rahman, and M. Begum, "Application of nonparametric test for trend detection of rainfall in the largest Island of Bangladesh," ARPN Journal of Earth Sciences, Vol. 2, No. 2, pp. 40-44, 2013.

[9] J.M. Melillo, T.C. Richmond, and G.W. Yohe, Climate Change Impacts in the United States: The Third National Climate Assessment. U.S. Global Change Research Program, 2014, 841 pp. doi:10.7930/J0Z31WJ2.

[10] D.R. Helsel and R.M. Hirsch, Statistical methods in water resources. Studies in Environmental Science 49. New York: Elsevier, 2002.

[11] R. Morton, and B.L. Henderson, Estimation of nonlinear trends in water quality: An improved approach using generalized additive models, Water Resources Research, vol. 44, 2008, doi:10.1029/2007WR006191.

[12] M.T. Calloway, M.E. Roberts, and C. Taylor, Temporal and Spatial Patterns of Fish Distribution and Diversity in the Noxubee River, Mississippi and Alabama. Copeia. 105. 100-107. 10.1643/CI-16-428, 2017.

[13] S.R. Green, Environ. Geol. Water Sci Vol. 7, 1985, pp. 9-13. https://doi.org/10.1007/BF02509879.

[14] J.J. Ramirez-Avila, W.H. McAnally, S.L. Ortega-Achury, A Study on Regional Frequency Analysis of Annual and Seasonal Rainfall in Mississippi and Alabama. In Proceedings 2018 World Environmental and Water Resources Congress. [S. Kamojjala (ed)], 2018.

[15] N. Bankhead, A. Simon and D. Klimetz, Analysis of Streambank Erosion along the Lower Tombigbee River, Alabama, USDA-ARS National Sedimentation Laboratory, Research Report Number 62, 2018.

[16] H.H, Jeffcoat, J.B. Atkins and D.B. Adams, U.S. Geological Survey; "General Climatology" section by Steven F. Williams, Assistant Alabama State Climatologist. U.S. Geological Survey Water-Supply Paper 2375, pp. 163-170, 1989.

[17] R.E. Southard, and E.J. Tharpe, U.S. Geological Survey; "General Climatology" section by Charles L. Wax, State Climatologist; "Water Management" section by Charles A. Branch, Mississippi Bureau of Land and Water Resources. U.S. Geological Survey Water-Supply Paper 2375, p. 353-360, 1989.

[18] J. McEnery, J. Ingram, Q. Duan, T. Adams, and L. Anderson, NOAA'S ADVANCED HYDROLOGIC PREDICTION SERVICE: Building Pathways for Better Science in Water Forecasting, Bulletin of the American Meteorological Society, vol. 86, 32017, 2005, https://doi.org/10.1175/BAMS-86-3-375.

[19] X. Jian, D. Wolock, and H. Lins. WaterWatch—Maps, Graphs, and Tables of Current, Recent, and Past Streamflow Conditions. Fact Sheet 2008-3031, United States Geological Survey (USGS), 2008.

[20] H. Mann, Nonparametric Tests Against Trend, Econometrica, vol. 13, 1945, pp. 245-259. doi:10.2307/1907187.

[21] R.O. Gilbert, Statistical Methods for Environmental Pollution Monitoring, Wiley, NY, 1987.

[22] M.G. Kendall, Rank Correlation Methods, 4th Edition, Charles Griffin, London, 1975.

[23] M.G. Kendall, A New Measure of Rank Correlation. Biometrika, vol. 30, 1938, pp. 81-89, JSTOR 2332226, doi:10.1093/biomet/30.1-2.81.

[24] Interstate Technology & Regulatory Council (ITRC), “Groundwater Statistics and Monitoring Compliance, Statistical Tools for the Project Life Cycle. GSMC-1. Washington, D.C.: Interstate Technology & Regulatory Council, Groundwater Statistics and Monitoring Compliance Team,” 2013.

[25] P.K. Sen, "Estimates of the regression coefficient based on Kendall's tau", Journal of the American Statistical Association, vol. 63, 1968, pp.1555-1574, doi:10.2307/2285891.

[26] C. Spearman, General Intelligence, Objectively Determined and Measured, The American Journal of Psychology, vol: 15 (2), pp. 201–292, 1904.

[27] S. Yue, P. Pilon, G. Cavadias, “Power of the Mann-Kendall and Spearman’s rho tests for detecting monotonic trends in hydrological series,” Journal of Hydrology, vol. 259, 2002, pp. 254-271.

[28] H. Theil, A rank-invariant method of linear and polynomial regression analysis, Nederl. Akad. Wetensch., vol. 53, 1950, pp. 386–392, doi: 10.1007/978-94-011-2546-8_20.

[29] G.E. Granato, “Kendall-Theil Robust Line (KTRLine--version 1.0)-A visual basic program for calculating and graphing robust nonparametric estimates of linear-regression coefficients between two continuous variables, Techniques and Methods of the U.S. Geological Survey," book 4, chap. A7, U.S. Geological Survey, pp. 31, 2006.

BIOGRAPHY

Elisabeth Ingram obtained her bachelor’s and master’s degrees in Civil Engineering from Auburn University. She is currently a Ph. D. student under the supervision of Dr. John Ramirez-Avila, focusing her research on precipitation trends and patterns.