AN APPROACH TO RAINFALL PROBABILITY ANALYSES ON SELECTED CENTRAL OKLAHOMA STATIONS

By

ROBERT BECK DUFFIN

Bachelor of Science
Utah State University
Logan, Utah
1951

Master of Science
Oklahoma State University
Stillwater, Oklahoma
1963

Submitted to the faculty of the Graduate College
of the Oklahoma State University
in partial fulfillment of the requirements
for the degree of
DOCTOR OF PHILOSOPHY
May, 1967

AN APPROACH TO RAINFALL PROBABILITY ANALYSES

ON SELECTED CENTRAL OKLAHOMA STATIONS

Thesis Approved:

James F. Stone

Thesis Adviser

Ralph S. Matlock

Lester W. Reed

Billy B. Tucker

David T. Bee

James E. Garton

N N Durham

Dean of the Graduate College

ACKNOWLEDGMENTS

Grateful acknowledgment is given to the personnel of the Agronomy and Statistics Departments of Oklahoma State University for the facilities, equipment and other resources available for this study.

Appreciation is extended to Dr. John F. Stone for his patience and valuable counsel. Gratitude is also expressed for the time and assistance given by other committee members--Dr. James E. Garton, Dr. Billy B. Tucker, Dr. Ralph S. Matlock, Dr. David E. Bee, and Dr. Lester W. Reed. Special thanks for statistical aid and computer programming assistance is due Dr. Robert D. Morrison, Dr. Dale D. Grosvenor, Bill Accola and other staff members of the Statistics Department. The advice of Herb Kershaw, Jr., O.S.U. Meteorology instructor, is also appreciated.

For encouragement and understanding, deepest thanks is given my wife Beryl and other members of my family.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

One objective of rainfall data analyses is to describe short-term and seasonal rainfall probability functions. Climatologists are finding it difficult to decide which approach to use in the analyses or presentation of rainfall probability data. High frequency "noise" in short-term rainfall probabilities is the major source of this difficulty. The question often arises as to whether rainfall data would be more representative if presented as actual frequencies, or after the "noise" has been smoothed or filtered out. Numerous examples are in evidence of the reporting of data through both approaches (5,7,15,16,17,23,24).

For most purposes, exact statistical measures for characterizing and describing rainfall are lacking. However, a visual comparison of plotted rainfall frequencies for the Stillwater station as reported in the author's M.S. thesis indicates rather well-defined seasonal trends. The M.S. thesis will hereafter be referred to as "the earlier study" whereas the Ph.D. work will generally be referred to as "the present study."

A major purpose of this study was to explore means of providing and evaluating smoothed areal rainfall probability data based on the daily period of record. The availability of long term records based on the various types of individual frontal and non-frontal rainfall events would be considered most desirable. However, this type of record is not available, except for relatively short periods. The daily rainfall record is used in the present study since it is a closer approximation of the actual rainfall event than is rainfall data for longer periods. If rainfall totals are reported on a weekly or monthly basis, the identity of both the daily and individual rainfall event is completely lost.

The earlier study hypothesized that rainfall data for the Stillwater station could be justifiably pooled over periods of the year within which both P_1 and P_2 are considered constant. Results provided support for this hypothesis. This pooling technique seems to provide good estimators of the probability functions, P_1 and P_2.

The first probability (P_1) can be described as the average probability of receiving rainfall on any randomly selected day within the particular grouping of consecutive days. The second probability (P_2) is the average probability that provided rainfall occurs on any randomly selected day within a particular grouping of consecutive days, a specified amount of rainfall will result.

Information obtained at one point of sampling has limited value. The nature of the P_1 and P_2 functions over an area is of particular interest in this study.

The station-to-station comparison data for the present study are obtained by smoothing from all years on record. It is recognized that any year-to-year variation is lost with this type of averaging. It is not the purpose of this study to evaluate year-to-year rainfall fluctuations. The extent of this type of variation, however, must be recognized since it is of great economic concern (9). Palmer, for example, reports Oklahoma and Kansas in an area of maximum standard deviation of summer moisture departures for the United States (19).

CHAPTER II

REVIEW OF LITERATURE

Meteorological Statistics

Adequate methods of statistical treatment for meteorological data are frequently lacking (7,8,17). The assignment of confidence intervals and the estimation of the extent of error in the use of rainfall probability distributions is somewhat hazardous. The mathematical problems which would permit statistical procedures to describe more exactly rainfall probability variation have not yet been completely solved.

Many authors have reported on statistical methods relating to rainfall probability data. A few of these writings relate more directly to this dissertation (2,6,7,8,11,15,17,20,22).

Portig offers two ways that small sample problems can be overcome (20).

> (1) The sample can be arbitrarily broken up into two or more sub-samples of equal size. Each of them must show the calendaricity near the same date. (2) The time analyses (not harmonic analyses!) of adjacent regions must fit together, i.e., it should be possible to trace a calendaricity with some small lag into a neighboring region, from there into another, etc.

These suggestions have particular relevance to the present study. Much of the statistical treatment is involved with testing the reality of the P_1 and P_2 functions throughout the year for each of several selected stations, and then comparing these functions on a station-to-station basis.

The Rainfall Event

Precipitation data recorded as daily rainfall amounts will be considered the basic unit of this study. The daily observations are available over the entire station record whereas the necessary weather maps to assist in properly identifying rainy periods are not available for the entire record.

It is evident that the true rainfall event must be defined in relation to its type of frontal or non-frontal origin. A number of investigations have established significant correlations between consecutive daily rainfall occurrences. In fact, persistence of several days is commonly observed (2,5,6,14,15,16,22,25). The problems of identifying and defining the complex frontal type of activity became apparent in an Oklahoma report by Kershaw and a synoptic precipitation study by Jorgensen (12,13). Still another study submits evidence that broadscale major precipitation trends over the United States are continuous in both time and space (3). These trends were identified through decadel moving averages.

Smoothing Rainfall Data

A commonly acknowledged problem in reporting precipitation data results from high frequency "noise" of the time function being considered (4,5,7,8,11,19,20,22). This "noise" seems to originate partly from year-to-year variability in rainfall frequencies and amounts (9,19).

Several different approaches in reporting rainfall data seem to be emerging. Most workers conclude that sampling fluctuations are responsible for the extremes and that irregularities in plotted data are caused by small sample size (5,7,8,24). It is further contended that some method of smoothing the plotted data is justified. This justification is based on the assumption that the general shape of the low frequency component of the plotted data is meaningful, but the short-term or high frequency component or "noise" irregularities are not (11,18).

The need for smoothing rainfall data is recognized; however, some workers believe that valuable detail might be lost in the process. Precipitation data are therefore sometimes reported both in its "raw" and smoothed condition (5,7,8,11,17). Gleeson regards these two alternatives as being "mutually exclusive but complementary procedures" (8).

Cooperative Weather Bureau and University rainfall probability studies are among those which recognize the need for smoothed rainfall probability values. The outstanding Regional Research projects in this category include: (1) NC - 26, (2) NE - 35, and (3) W - 48. These projects have yielded a number of reports concerned with rainfall probabilities for their respective areas. Most of the workers on these studies fit the incomplete gamma distribution to rainfall data to obtain smoothed probability or rainfall amount values.

The original work in fitting the incomplete gamma distribution to rainfall data is reported by Barger and Thom (1). The more detailed account of this earlier work was published later by Thom (23).

Other approaches to smoothing or filtering of data are reported in the literature (4,5,11,17,18,20,21). A major hypothesis of the present study is that variations in rainfall probability data for Stillwater and other Oklahoma stations can be estimated from non-parametric methods involving smoothing obtained from actual tabulated rainfall events or incremental frequencies.

Reporting Areal Probability Variation

The need for areal reporting of rainfall probabilities is widely recognized. A large majority of both the long record average and individual storm rainfall data reported in the past has related to point rainfall. The major current emphasis is directed toward reporting rainfall frequencies and probabilities on a broad regional basis.

The precipitation or observation day provides the basic data for this thesis. Areal considerations will be limited to long record averages and therefore will not identify synoptic precipitation on an individual storm or on a moving year-to-year basis. Reporting of individual storm and year-to-year synoptic type changes in climatology are areas of interest in which significant advances are being made (3,8,10,12,13,19).

CHAPTER III

METHODS AND MATERIALS

Station Selection

The eight stations selected for this study are located on the rainfall map illustrated in Figure 1.

The length of record for each station is reported in Table I.

The stations for this study were selected from the twenty stations available in central Oklahoma. They were selected in an attempt to obtain an approximate representation of variability from long record rainfall data averages for central Oklahoma. Stations were grouped in an east-west and north-south pattern to approximately represent rainfall variations along and normal to isohyet lines.

Rainfall Probabilities

Rainfall was characterized in this study by evaluating P_1 and P_2 functions within various sized groupings of days throughout the climatological year. The Weather Bureau IBM cards recorded occurrences of either no rain, trace, or a particular amount of rainfall in hundredths of an inch for each day. The week and day numbers of the climatological year are identified with corresponding calendar dates in Table II. Further reference to climatological day 1 through 365 will generally be made without their corresponding calendar identification.

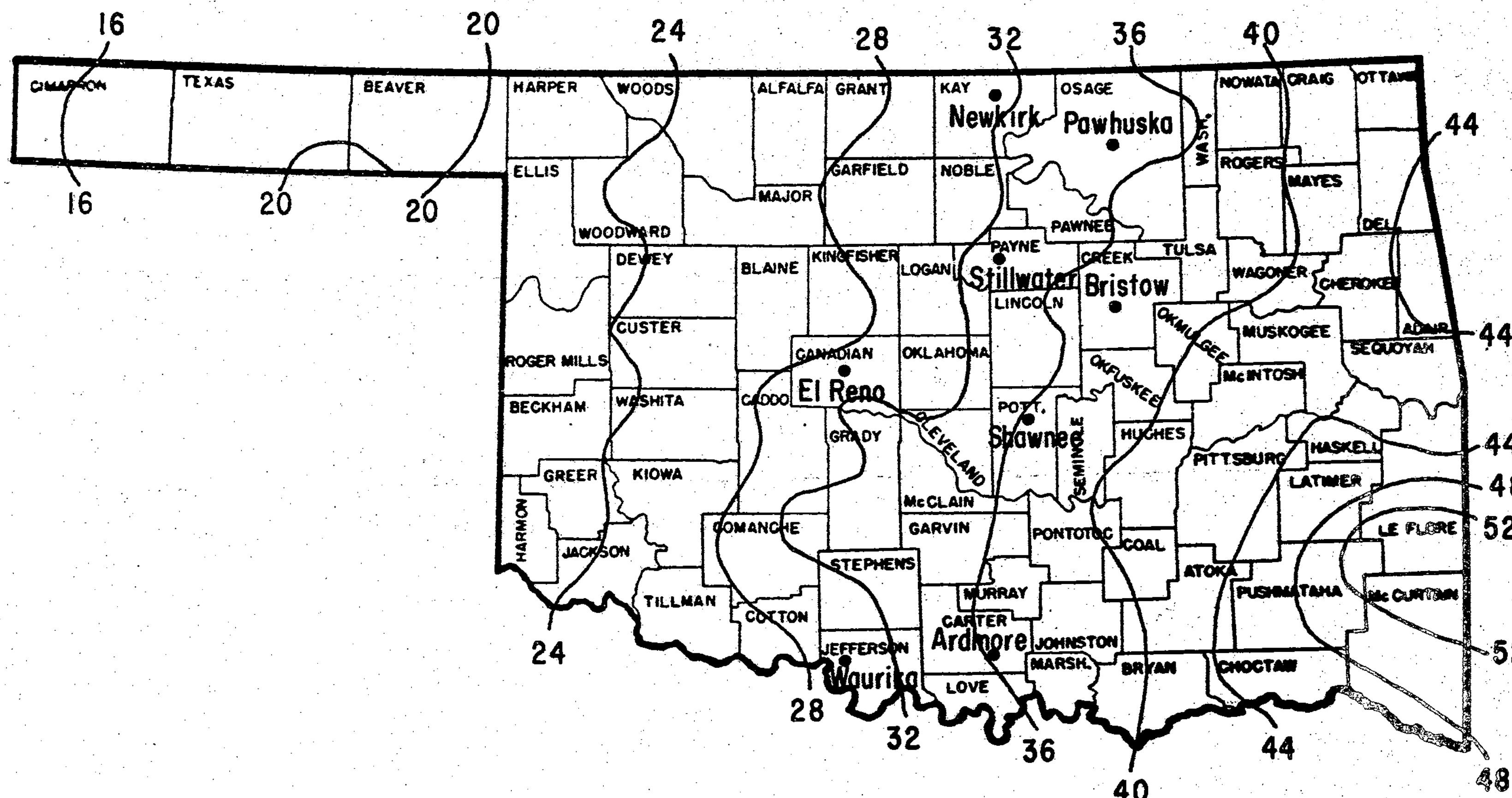

Figure 1. Isohyet Map Locations of Eight Central Oklahoma Stations Selected to Evaluate P_1 and P_2 Functions

TABLE I

ANNUAL PRECIPITATION OF STATIONS AND PERIOD OF RECORD AS APPLIED TO THIS STUDY

Station	Average Annual Precipitation (Inches)*	Number of Years on Record	Period of Record
Newkirk	30.60	68	1897-1964
Pawhuska	34.64	75	1889-1964
Stillwater	32.18	72	1893-1964
Bristow	37.54	50	1915-1964
El Reno	29.08	72	1893-1964
Shawnee	37.22	64	1901-1964
Waurika	31.48	55	1910-1964
Ardmore	37.14	64	1901-1964

*Precipitation amounts are based on averages for the U.S.W.B. normal period, 1931 - 1960.

TABLE II

THE CLIMATOLOGICAL YEAR

Week and Day Numbers with Inclusive Dates

Week Number	Day Numbers	Beginning Date	Ending Date	Week Number	Day Numbers	Beginning Date	Ending Date
01	1- 7	Mar. 1	Mar. 7	27	183-189	Aug. 30	Sept. 5
02	8- 14	Mar. 8	Mar. 14	28	190-196	Sept. 6	Sept. 12
03	15- 21	Mar. 15	Mar. 21	29	197-203	Sept. 13	Sept. 19
04	22- 28	Mar. 22	Mar. 28	30	204-210	Sept. 20	Sept. 26
05	29- 35	Mar. 29	Apr. 4	31	211-217	Sept. 27	Oct. 3
06	36- 42	Apr. 5	Apr. 11	32	218-224	Oct. 4	Oct. 10
07	43- 49	Apr. 12	Apr. 18	33	225-231	Oct. 11	Oct. 17
08	50- 56	Apr. 19	Apr. 25	34	232-238	Oct. 18	Oct. 24
09	57- 63	Apr. 26	May 2	35	239-245	Oct. 25	Oct. 31
10	64- 70	May 3	May 9	36	246-252	Nov. 1	Nov. 7
11	71- 77	May 10	May 16	37	253-259	Nov. 8	Nov. 14
12	78- 84	May 17	May 23	38	260-266	Nov. 15	Nov. 21
13	85- 91	May 24	May 30	39	267-273	Nov. 22	Nov. 28
14	92- 98	May 31	June 6	40	274-280	Nov. 29	Dec. 5
15	99-105	June 7	June 13	41	281-287	Dec. 6	Dec. 12
16	106-112	June 14	June 20	42	288-294	Dec. 13	Dec. 19
17	113-119	June 21	June 27	43	295-301	Dec. 20	Dec. 26
18	120-126	June 28	July 4	44	302-308	Dec. 27	Jan. 2
19	127-133	July 5	July 11	45	309-315	Jan. 3	Jan. 9
20	134-140	July 12	July 18	46	316-322	Jan. 10	Jan. 16
21	141-147	July 19	July 25	47	323-329	Jan. 17	Jan. 23
22	148-154	July 26	Aug. 1	48	330-336	Jan. 24	Jan. 30
23	155-161	Aug. 2	Aug. 8	49	337-343	Jan. 31	Feb. 6
24	162-168	Aug. 9	Aug. 15	50	344-350	Feb. 7	Feb. 13
25	169-175	Aug. 16	Aug. 22	51	351-357	Feb. 14	Feb. 20
26	176-182	Aug. 23	Aug. 29	52	358-364	Feb. 21	Feb. 27

Averaging Daily Rainfall Occurrences

The different size groupings of days in which P_1 and P_2 were obtained include 365 consecutive 1-day periods and 365 overlapping 3-, 5-, 7-, 15- and 29-day consecutive periods throughout the climatological year. The basic computer program to obtain P_1 values was devised for 1-day values. Modifications of this 1-day program were made to obtain the 3-, 5-, 7-, 15- and 29-day period data through an equally weighted running average scheme.

Data from running averages were obtained by including 1, 2, 3, 7, and 14 consecutive days on both sides of a consecutive moving midpoint day throughout the year for the respective 3-, 5-, 7-, 15- and 29-day equally weighted moving average data. The P_1 record for each grouping always began with climatological day 1 (March 1) as midpoint day and ended with climatological day 365 (February 28) as the ending midpoint day. Periods of different length were used for the purpose of evaluating P_1 and P_2 functions at various degrees of smoothing. The retention of greater detail might be expected by using periods of relatively short length. While this feature might be considered desirable for some purposes, the high frequency "noise" of rainfall frequencies derived from short periods is evidence of the need for data smoothing.

Computer programs were devised to both calculate and plot P_1 values for the different length periods for each station. This programming was done for the IBM Computers, Type 1410 and Type 7040. The IBM Sorter Type 083 was used to arrange the WB-1009 daily cards into a chronological order by years with March 1 as the beginning day for each year. Missing data cards were prepared to provide for proper consideration of missing observation days.

IBM punch cards with daily recordings provided the following types of basic data: (1) rainfall amounts, (2) trace, (3) missing data, or (4) no rain. Rain-day cards were recorded into increment classes up to 2.99 inches. The rainfall increments from greater than trace to 2.99 inches were .01-.09, .10-.19, .20-.29, .30-.39, .40-.49, .50-.59, .60-.69, .70-.79, .80-.89, .90-.99, 1.00-1.24, 1.25-1.49, 1.50-1.74, 1.75-1.99, 2.00-2.24, 2.25-2.49, 2.50-2.74, and 2.75-2.99 inches. Rainfall amounts over 2.99 inches were recorded individually for each station with their respective dates.

Basic data for obtaining P_1 and P_2 values for various size groupings throughout the year includes the number of rain days, missing data, trace, and no-rain cards. P_1 values for a particular size grouping were obtained by the following relationship:

$$P_1 = \frac{\text{number of days with rain}}{\text{days on record}} = \frac{\text{total} - (\text{missing data} + \text{trace} + \text{no rain})}{\text{total} - \text{missing data}}$$

The P_2 values for various size groupings throughout the year were calculated and recorded from the previously described data:

$$P_2 = \frac{\text{number of rain days within a specified rainfall increment}}{\text{total rain days}}$$

Slope Value Determination

In the earlier study, a procedure of semi-log graphing of P_2 values was used in an attempt to evaluate variation in P_2. A similar procedure was proposed in the present work as one of two independent methods for evaluating the P_2 function. It involved the computer calculation of semi-log datum point values for P_2 (percent of total rainfall occurrences in a given rainfall increment) in relation to the X values (incremental rainfall amounts). It was intended that each P_2 value used in slope determinations be adjusted by dividing by its corresponding 0.1-inch or 0.25-inch class size. Due to a programming error, the P_2 values were divided by their respective class number. The classes were numbered from 1 through 18 as indicated in Table III. Since the sensitivity to evaluating the P_2 function by the resulting slope values does not seem to be altered, reprogramming was not considered necessary. In calculating slopes, the zero rainfall occurrences within incremental rainfall classes were ignored or passed over in the computer program. A straight line was fitted to each set of semi-log versus P_2 values by the least squares method.

For each frequency distribution being considered, the computer calculated the slope of the straight line representing the semi-log distribution. An understanding of the physical significance of the slopes is important for the proper evaluation of results. Upon examination of the examples provided in Table III and Figure 2, it becomes evident that the semi-log plotting of data points is approximately linear. This holds true quite well throughout the year. Because of the heavy grouping of small rains, this is not entirely true, however. The comparison of absolute slope values obtained from different P_2 distributions is used to evaluate the P_2 function since slope values will change when the relative proportions of the different incremental rainfall amounts in relation to the total rainfall occurrences changes. If the slope values are a valid test, they should provide a measure as to whether the percentages of larger or smaller rains are increasing or decreasing from one time of the year to another.

For example, consider two consecutive 29-day periods for a given station. If in the first period a greater percentage of the total rain-

TABLE III

SLOPE DETERMINATION DATA FOR MIDPOINT DAY 188 OF 29-DAY GROUPING FOR PAWHUSKA*

Observation Number	Rainfall Class (Inches)	Adjusted P_2 Values** (Y)	$\text{Log}_{10}Y$
1	.10	2.1176	.32585
2	.20	.88235	-.05436
3	.30	.35294	-.45230
4	.40	.14706	-.83251
5	.50	.11176	-.95170
6	.60	.07353	-1.13354
7	.70	.09663	-1.01485
8	.80	.03308	-1.48032
9	.90	.04575	-1.33960
10	1.00	.02873	-1.54157
11	1.25	.04597	-1.33746
12	1.50	.02682	-1.57154
13	1.75	.02627	-1.58050
14	2.00	.01580	-1.80122
15	2.25	.00255	-2.59274
16	2.50	.00229	-2.63850
17	2.75	.00000	Ignored
18	3.00	.00000	Ignored

Intercept (b) = -.2477, Slope (M) = -.9571

* Data as recorded by I.B.M. Number 1410 in slope calculation program

** P_2 values as herein recorded were obtained as follows:

$$\frac{\text{rain-day frequency for a particular class}}{\text{rain-day frequency for all classes}} \Big/ \text{observation number X 10}$$

TABLE III

SLOPE DETERMINATION DATA FOR MIDPOINT DAY 188 OF 29-DAY GROUPING FOR PAWHUSKA*

Observation Number	Rainfall Class (Inches)	Adjusted P_2 Values** (Y)	$Log_{10}Y$
1	.10	2.1176	.32585
2	.20	.88235	- .05436
3	.30	.35294	- .45230
4	.40	.14706	- .83251
5	.50	.11176	- .95170
6	.60	.07353	-1.13354
7	.70	.09663	-1.01485
8	.80	.03308	-1.48032
9	.90	.04575	-1.33960
10	1.00	.02873	-1.54157
11	1.25	.04597	-1.33746
12	1.50	.02682	-1.57154
13	1.75	.02627	-1.58050
14	2.00	.01580	-1.80122
15	2.25	.00255	-2.59274
16	2.50	.00229	-2.63850
17	2.75	.00000	Ignored
18	3.00	.00000	Ignored

Intercept (b) = -.2477, Slope (M) = -.9571

* Data as recorded by I.B.M. Number 1410 in slope calculation program

** P_2 values as herein recorded were obtained as follows:

$$\frac{\text{rain-day frequency for a particular class}}{\text{rain-day frequency for all classes}} \Big/ \text{observation number X 10}$$

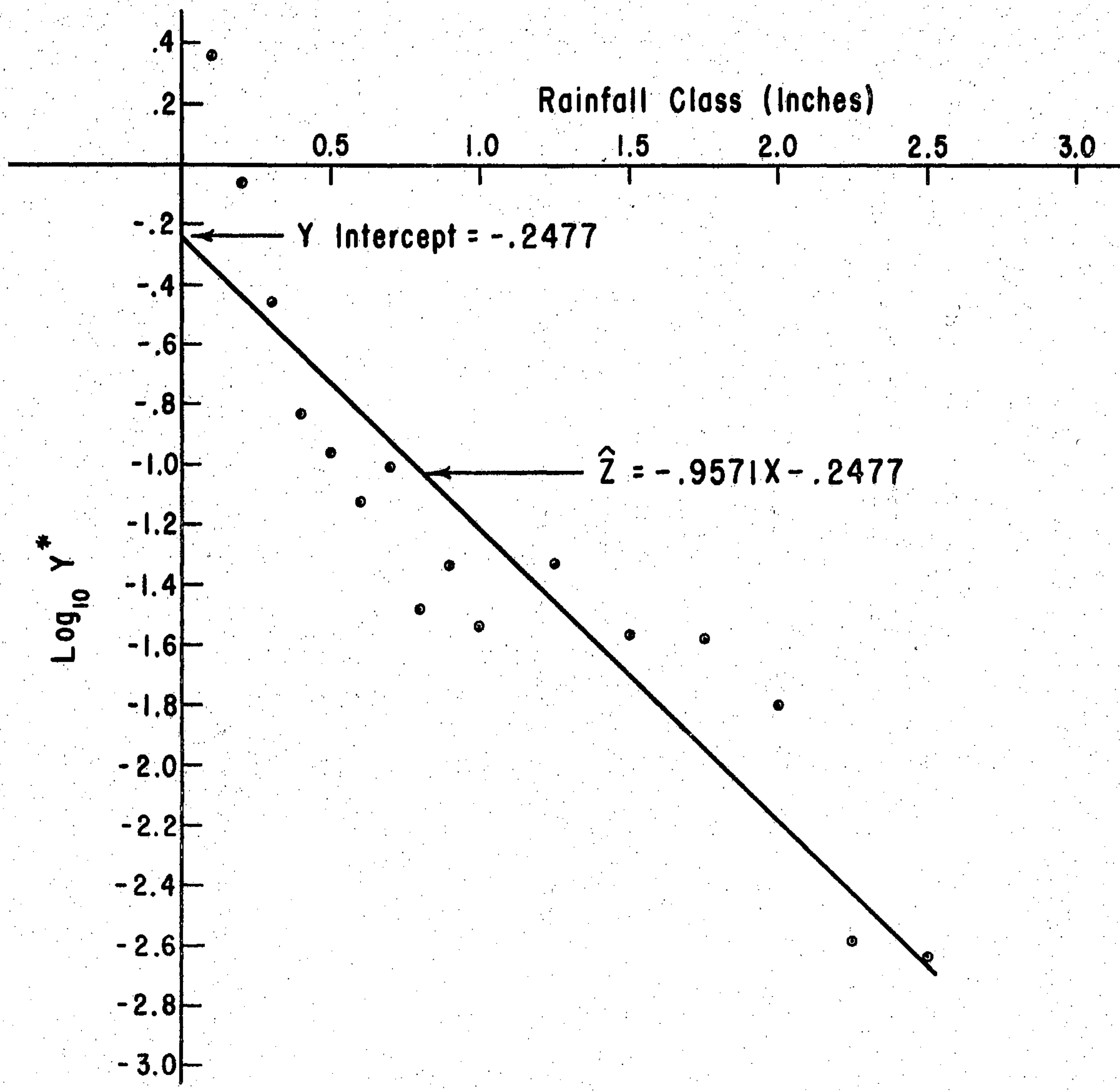

Figure 2. Plotting of Slope Determination Data for Midpoint Day 188 in a 29-Day Grouping of the Pawhuska Station

$$*Y = \frac{\text{rain-day frequency for a particular class}}{\text{rain-day frequency for all classes}} \Big/ \text{observation number X 10}$$

fall occurrences consist of larger incremental rainfall amounts than is true for the second period, this difference could be indicated by comparing the slopes of the plotted data. All slope values are negative. In the first period, with its greater comparative percentage of larger rains, the line would have a smaller absolute slope value than in the second period.

By least squares procedure the slope (m) is estimated from the model $\text{Log}_{10}P_2 = mX + b$. For convenience, let $Z = \text{Log}_{10}P_2$. The sample slope ($\hat{m}$) as seen in the example provided in Figure 2 is obtained from the sample regression of equation $\hat{Z} = \hat{m}X + \hat{b}$.

Tabulations of slope values for different sized consecutive groupings of days throughout the year were used in this study as partial criteria for the evaluation of P_2 variation. Consecutive non-overlapping period slope values were readily obtained from equally weighted running average data. Periods of shorter length than the 15- and 29-day provided such erratic data that it was considered of questionable value in this study.

Probability Ratio Determination

In a further attempt to characterize P_2, the probability ratio determination was developed. The purpose of introducing this method was to compare its results for equivalence with corresponding slope values.

Since both slope and probability ratio methods are independently derived, the comparison between their results should be of interest. The probability ratio description is as follows:

Let P_{1k} = probability of a rain event on day k where $1 \leq k \leq 365$.

$P_{1k} = \frac{N^*_k}{N_k}$ where N^*_k is the number of days in the set N_k on which precipitation was reported.

$P_{2ki} = \frac{N^*_{ki}}{N^*_k}$ where N^*_{ki} is the number of days with rainfall in N^*_k rain days where the amount recorded fell in the ith increment of precipitation amount. The ith increment corresponds to observation numbers from 1 through 18 as reported in the example in Table III.

If $\bar{a}_i$ is the mean amount of rain in the ith increment, then we perform

$$\frac{\sum \bar{a}_i N^*_{ki}}{N^*_k} = \sum \bar{a}_i P_{2ki} \quad . \tag{1}$$

Working with the left hand member $\frac{1}{N^*_k} \sum_i \bar{a}_i N^*_{ki} = \frac{1}{N^*_k} T_k$, where T_k is the total water input on all the kth days of the record.

Hence from equation (1)

$$\sum_i \bar{a}_i P_{2ki} = \frac{1}{N^*_k} T_k \quad .$$

Multiplying by N_k

$$N_k \sum_i \bar{a}_i P_{2ki} = \frac{N_k}{N^*_k} T_k \tag{2}$$

So

$$\sum_i \bar{a}_i P_{2ki} = \frac{1}{N_k} \frac{T_k}{P_{1k}} \quad .$$

The term on the right is referred to as the probability ratio in this study.

If we can assume $\bar{a}_i$ can be approximated by the value of the center of the precipitation increment i, then for a given day k, the probability ratio would seem to be a function of P_{2k} alone. If the P_{2ki} function does not change from day k to, say k + 1, then $\frac{T_k}{P_{1k}}$ would be expected to equal $\frac{T_{k+1}}{P_{1(k+1)}}$. However, constancy of $\frac{1}{N_k} \frac{T_k}{P_{1k}}$ does not guarantee constancy of P_{2k} over k since it is conceivable that P_{2k} could vary in such a manner that $\Sigma \bar{a}_i P_{2ki}$ would not change with k. Note that the $\bar{a}_i$ weighting factor tends to bias the results toward the larger values of the i increments, so that small random changes could produce large changes in the terms $\Sigma \bar{a}_i P_{2ki}$. On the other hand, if $\frac{1}{N_k} \frac{T_k}{P_{1k}}$ changes significantly, it must be because of a change in the P_{2k} function.

A summary on P_{2k} changes follows:

(1) Changes of P_{2k} are conceivable without corresponding $\frac{1}{N_k} \frac{T_k}{P_{1k}}$ changes.

(2) Changes of P_{2k} are possible which change $\frac{1}{N_k} \frac{T_k}{P_{1k}}$.

It seems unlikely that $\frac{1}{N_k} \frac{T_k}{P_{1k}}$ could change without a corresponding change in $\Sigma_i \bar{a}_i P_{2ki}$.

Statistical Tests

The runs-test (21), student's t-test, and visual comparisons of plotted data were adopted as statistical methods to assist in evaluating plots of the P_1 and P_2 functions. Various combinations of paired P_1 and P_2 data comparisons at specific climatological periods of time were made by these methods. Paired data comparisons included combinations resulting from separation of the within-station P_1 and P_2 data into various sub-samples such as even years versus odd years and first half of the record versus last half of the record. Further reference to these comparisons of within-station sub-samples will be made. Portig (20) suggests that if data are relevant they should stand the test of separation into arbitrary groups which should still produce the approximate same result. Paired data comparisons of sub-samples were made to assist in the evaluating P_1 and P_2 values obtained from corresponding dates. These within-station comparisons of P_1 and P_2 functions were made as opposed to those derived from various station-to-station comparisons.

Within-station paired data comparisons of P_1 and P_2 functions should provide some basis for evaluating station-to-station comparisons. There is no apparent reason to believe that for any particular station P_1 and P_2 values for the odd years of record would be any different from those of even years. Therefore, differences in P_1 or P_2 at corresponding calendar dates of comparison would be expected to originate from random sources. There could be some basis, however, for contending that the first half of the record might generate different probability values than the latter half. Long term, broadscale climatic changes might conceivably result in average probability for the first half of the record that are actually somewhat different than for those of the last half. However, these within-station comparisons are expected to provide some measure of random P_1 and P_2 differences which will in turn assist in evaluating station-to-station variability.

A further approach was used to evaluate random type P_1 and P_2 differences from within-station and station-to-station variability of t-test and runs-test values. More than one beginning date was chosen for the midpoint day of the first consecutive non-overlapping period for each station. Beginning dates were usually climatological days 1 and 15 for 15-day groupings, and climatological days 1, 8, 15 and 22 for 29-day groupings. It was not necessary to obtain t- and runs-test data for all these beginning dates for all stations. The approximate range in test results originating from different beginning dates was obtained with independent calculations for each station, thus different sets of relatively independent P_1 and P_2 paired data values were generated for P_1 and P_2.

A second consideration by Portig also has application to this study (20). He reports that an areal change should at most introduce phase lag of the function being considered if the distribution of the function is relatively constant with area. Various combinations of station-to-station paired data comparisons were made by the same three statistical methods used in evaluation of within-station paired data comparisons. For example, the statistical tests made on station-to-station paired data involve the comparing of P_1 and P_2 values of two stations at corresponding time periods of the year. In this manner, any areal phase lag of the P_1 and P_2 functions was studied.

For a set of paired values to be considered different, the following conditions are necessary: (1) the number of runs must prove to be other than random, and subsequently, (2) the mean difference between data pairs must prove different as determined by the t-test at a specified significance level.

The following example in Table IV of paired data comparisons is presented here as data from which determinations of t- and runs-test values were made.

For purpose of this example, paired comparisons of P_1 data for the 29-day groupings of Stillwater and Pawhuska stations is used. Paired values for 12 consecutive non-overlapping periods are used for the comparison using climatological day 1 as the midpoint day for the first period. Data representing non-overlapping periods were used to obtain a higher degree of independence between P_1 and P_2 values throughout the time period being considered.

The symbol D in Table IV represents the difference between Stillwater and Pawhuska's P_1 values for each of the 12 consecutive non-over-

TABLE IV

RUNS- AND t-TEST P_1 DETERMINATION DATA FOR THE STILLWATER VERSUS PAWHUSKA 29-DAY PAIRED COMPARISON

X_1*	X_2*	$D = X_1 - X_2$
.181 **	.191	-.010
.221	.210	.011
.270	.279	-.009
.318	.320	-.002
.235	.237	-.002
.201	.197	.004
.198	.202	-.004
.197	.200	-.003
.179	.177	.002
.157	.161	-.004
.143	.136	.007
.151	.144	.007

$\Sigma D = -.003$
$(\Sigma D)^2 = .000009$
$D^2 = .000469$

$S^2 = .00004$
$S = .0063$
$t = -0.14$

Runs = 8

*X_1 = Stillwater P_1 data for consecutive non-overlapping 29-day periods

X_2 = Pawhuska P_1 data for consecutive non-overlapping 29-day periods

**Climatological day 1 is the 29-day grouping midpoint day for this beginning period P_2 value lapping 29-day periods. Future references to D will be made in reporting both P_1 and P_2 results.

The t-test is not proposed as a unique measure of probability variation in this study. It seems obvious, however, that it does provide a good comparative basis from which overall magnitudes of difference can be evaluated. One should observe that it is possible for identical t-values to be calculated from two sets of widely divergent paired precipitation data. For example, the t-value from one paired set of data might represent a situation with no data cross-overs, while the second paired data set having cross-overs could yield an identical t-test value. The runs-test should assist in evaluating the extent of crossing over of paired data.

The null hypothesis, H_o, that $\sum D = 0$ is accepted in the Stillwater-Pawhuska example since the t-test value is -0.14 and $t_{.05}(11) = 2.201$. This Stillwater-Pawhuska t-test value is very small when compared to paired data from other stations. When this t-test information is considered with runs-test values which indicate random crossing over, we must conclude that the paired P_1 values for Stillwater and Pawhuska are both derived from a very similar population of rainfall events. At the 5 percent significance level, runs-test values ranging from 3 - 11 would be considered random under the hypothesis that the two stations provide random P_1 samples from a common population of P_1 values. The runs-test value of 8 in our example further supports the contention that there is no real difference between chronologically paired P_1 values for the Stillwater and Pawhuska stations.

CHAPTER IV

RESULTS AND DISCUSSION

Smoothing Probability Functions

A typical plotting of 1- and 29-day P_1 values is illustrated with the relatively long record Pawhuska data in Figure 3.

This extreme scattering of adjacent 1-day P_1 data values as compared with the relatively smooth 29-day P_1 data is characteristic of all eight stations included in this study. However, the stations with shorter records, such as Bristow, had somewhat greater P_1 variability than was observed in the longer record stations for all corresponding sized periods. As would be expected, the degree of smoothing increased as the length of period increased. The low and high extremes of P_1 smoothing as used in this study are illustrated in Figure 3. It is evident that the 29-day plotted values removed the extreme variability of "noise" found most prevalent in the 1-day grouping. The low frequency change is somewhat evident even in the 1-day values. However, the high frequency "noise" introduces considerable uncertainty.

Smoothing of P_1 values intermediate between the 1-day and the 29-day extremes was plotted for all stations. Results of these groupings of 3-, 5-, 7- and 15-day were not all readily explained. An intermediate-frequency cyclic tendency was in evidence from the 3- to the 7-day groupings for all stations. This tendency is most prominent with the

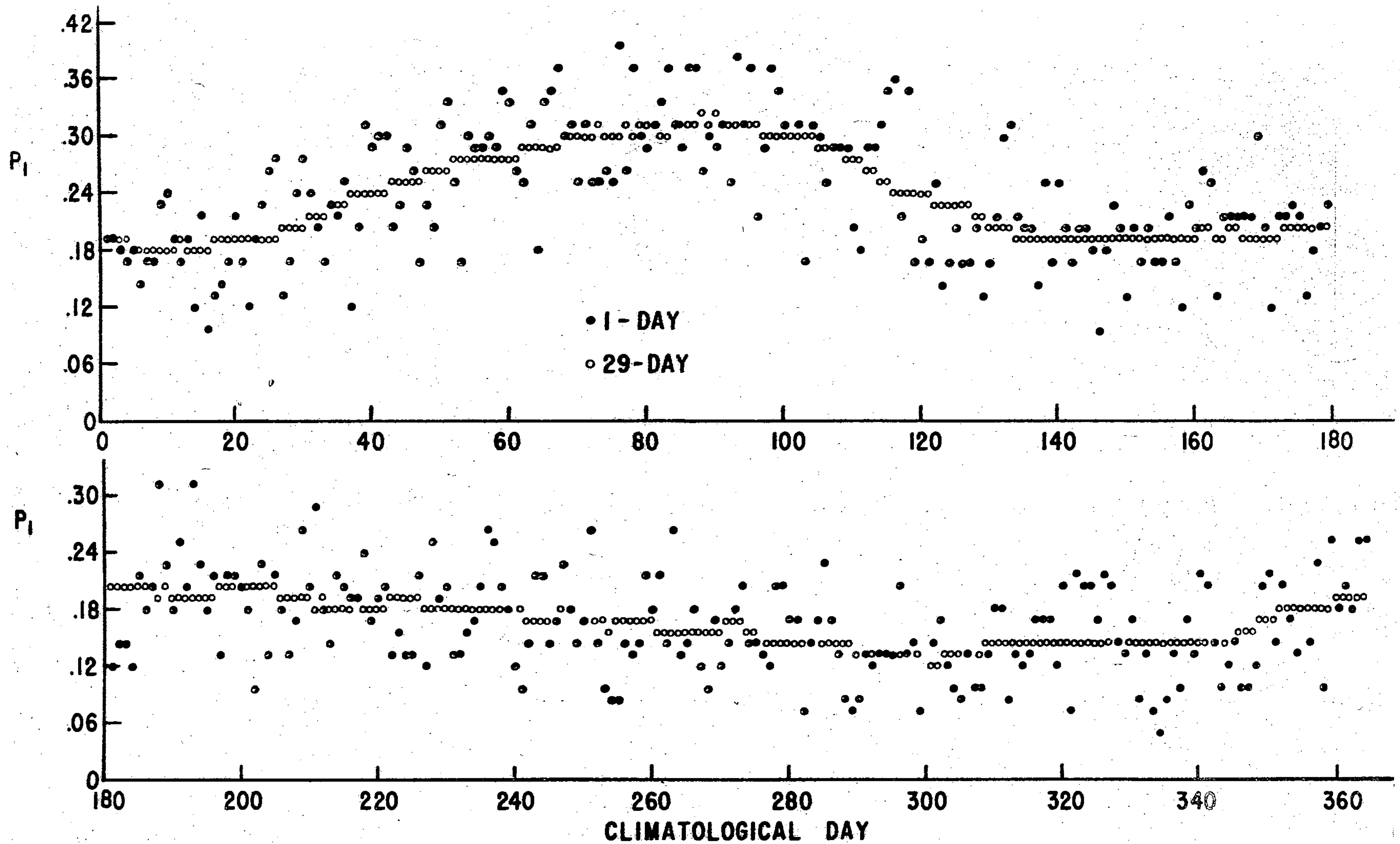

Figure 3. A comparison of 1-Day Versus 29-Day Grouping P_1 Data Values for the Pawhuska Station

5- and 7-day groupings. The example provided in Figure 4 illustrates 5-day P_1 data for both the Stillwater and the Pawhuska stations. This plotting is typical of the cyclic effect that becomes evident in intermediate frequency plottings of all eight stations. The similarity between this Stillwater and Pawhuska data results suggests a coupling of individual rainfall events for the two stations. The cyclic effect is probably largely introduced by the equally weighted moving average technique used in obtaining the plotted data values (7,11).

Slope and probability ratio values derived from periods less than 29-day origin are too erratic to be considered of value in this study.

Within-Station Probability Variation

P_1 Results

Table V lists the t- and runs-test results for P_1 within-station comparisons of even versus odd years.

Under the H_o that there is no difference between even and odd year P_1 data, the runs-test values in Table V for both 15- and 29-day groupings are almost entirely within random expectations at the 5 percent significance level. With the 24 observations for each of the 15-day groupings, the expected range in runs values would vary from 7 to 19. With the 29-day groupings and 12 observations per set, the expected range of runs values would vary from 3 to 11.

Results of the t-tests for the two different starting dates from the 15-day grouping seem to have no greater variation than those for the 29-day grouping. However, this is not true for the runs-test results. These values indicate much more random crossing over for the 15-day grouping as compared to the 29-day grouping.

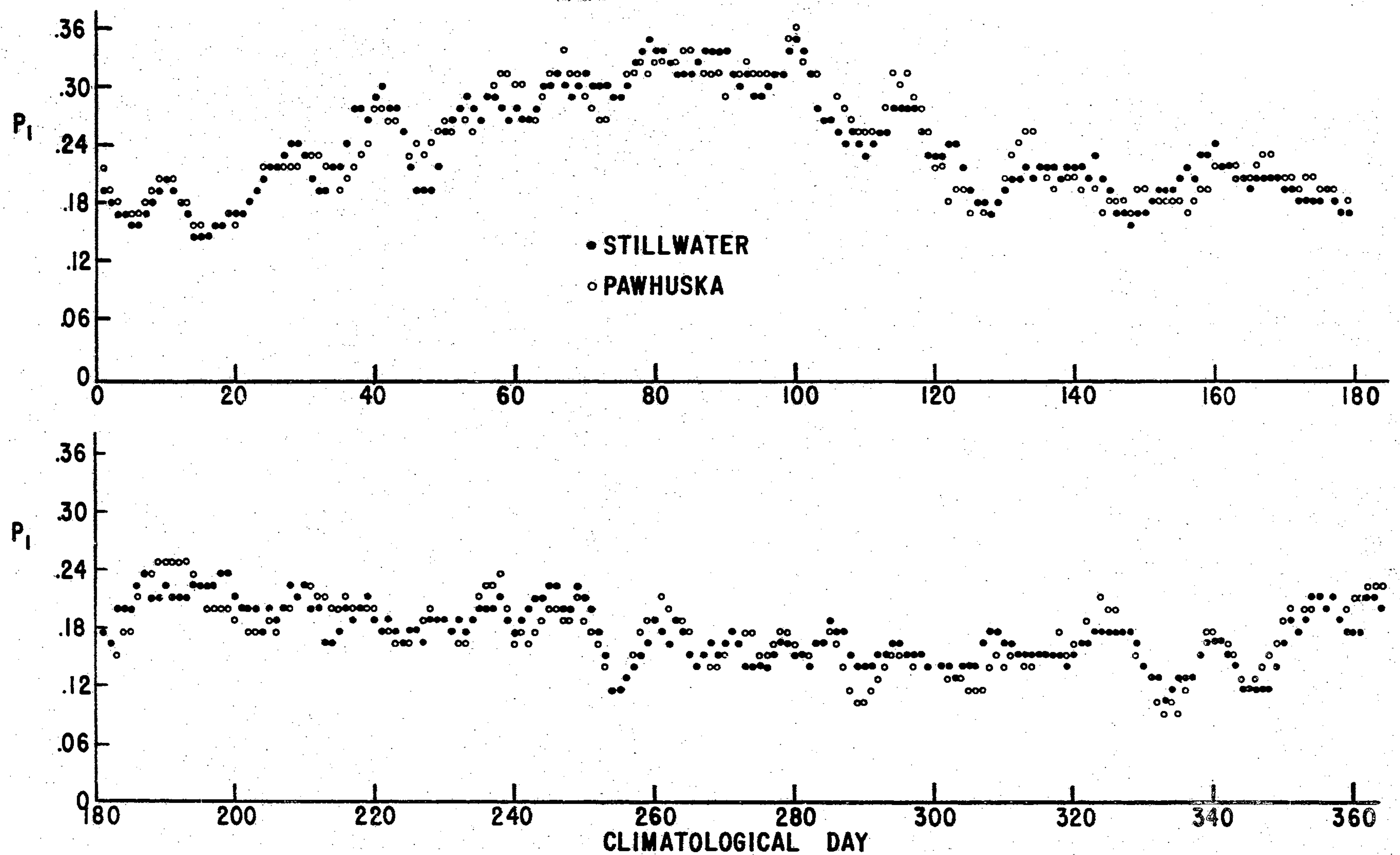

Figure 4. A Comparison of 5-Day P_1 Data Values for the Pawhuska Versus Stillwater Stations

The range in t- and runs-test values from one station to another is clearly greater than that originating from within-station comparisons of different starting dates.

The maximum range of differences in t-test values of Table V is observed by examining the Bristow and Stillwater data.

The t-test values for the Stillwater station in Table V, ranging from -0.47 to 0.61 for different starting dates, are the lowest for all stations. Bristow's comparable t-test data are the highest for all stations. They range from -2.39 to -4.31. The difference or spread in values between the Stillwater and Bristow stations probably represents a random type variation and not significant differences. Even or odd year data for a given station represent only one half the record as compared to the total record of the same station. The smaller sample of rainfall events associated with sampling only a portion of the total record is a contributing source of random variation. Under the H_o that there is no difference between the even year and odd year P_1 data, Stillwater's t-test values of Table V ranging from -0.47 to 0.61 are not unusual. With 11 degrees of freedom just 5 percent of the t values would be expected to lie outside the range $-2.201 < t < 2.201$. This same H_o for the Bristow, Shawnee and Ardmore data would probably still be accepted even though t-test values are slightly beyond the 5 percent ± 2.201 t-test value since the runs-test data generally indicate random crossing over.

Table VI is also presented here to further compare t- and runs-tests for portions of total station P_1 records. It is of interest that the first half versus last half year comparisons for the Pawhuska station yielded t- and runs-data indicating more similarly paired data than was true for the Pawhuska even versus odd year results. The reverse of this

TABLE V

RESULTS OF t- AND RUNS-TESTS ON P_1 VALUES FOR EIGHT CENTRAL OKLAHOMA STATIONS INVOLVING PAIRED COMPARISONS OF EVEN VERSUS ODD YEARS ON RECORD

	Even Years Versus Odd Years					
	15-Day Grouping		29-Day Grouping			
Station	1*	8	1	8	15	22
Newkirk	-1.87**	-1.96	-1.90	-1.87	-1.80	-1.36
	8***	10	4	4	4	6
Pawhuska	-2.37	-2.38	-1.82	-2.28	-1.48	-1.70
	6	6	4	4	4	4
Stillwater	-0.53	-0.71	-0.40	-0.47	0.61	0.05
	10	8	4	4	6	6
Bristow	-3.52	-3.29	-4.31	-3.57	-2.63	-2.39
	8	6	2	4	4	2
El Reno	0.36	-0.87	-0.78	-0.80	-0.69	-0.54
	16	8	6	8	8	6
Shawnee	-1.94	-2.47	-2.58	-2.49	-2.36	-2.55
	8	8	4	4	4	4
Waurika			-2.57	-1.92	-2.31	-1.99
			4	4	2	4
Ardmore	-3.35	-2.40	-2.75	-2.96	-3.07	-2.83
	12	12	6	4	2	4

* Climatological beginning midpoint date for the particular sized period.

** t-test value (at the 5 percent significance level critical t-test values for the 15-day grouping are ± 2.069 and for the 29-day grouping are ± 2.201).

*** Runs-test value (at the 5 percent significance level runs-test values which indicate random crossing over for the 15-day grouping range from 7 to 19, and from 3 to 11 for the 29-day grouping).

TABLE VI

RESULTS OF t- AND RUNS-TESTS OF P_1 VALUES FOR THE PAWHUSKA AND STILLWATER STATIONS INVOLVING PAIRED COMPARISONS OF FIRST HALF VERSUS LAST HALF YEARS ON RECORD

Station	15-Day Grouping		29-Day Grouping			
	1*	8	1	8	15	22
Pawhuska	0.14**	0.21	-0.04	-0.05	-0.19	0.24
	16***	16	6	6	6	6
Stillwater	-3.89	-3.09	-4.56	-3.90	-3.79	-3.05
	8	8	2	2	2	4

* Climatological beginning midpoint date for the particular sized period.

** t-test value

*** Runs-test value

situation is true for the Stillwater station. These relationships would tend to support the idea that even versus odd year and first half versus last half year differences occur randomly and have no apparent significance.

Figure 5 illustrates the Pawhuska even versus odd year comparisons of the 29-day grouping. This is provided to visually compare P_1 variation originating from within a station with the corresponding t- and runs-test data of Table V.

Slopes and Probability Ratios

A listing of t- and runs-test values for within-station comparisons of slope and probability ratio determinations is reported in Table VII and VIII respectively. Assuming H_o that even year and odd year data are the same, the runs values from both Tables VII and VIII are mostly within the expected range. The 15-day slope data grouping at different starting dates is not available for this report. The 29-day grouping t- and runs-test results reported in Table VII seem to indicate a random crossing over of even versus odd year data. The range of within-station data variability occurrence of Table VII is approximately the same as exists with station-to-station total record comparisons reported in Table VIII.

The statistical information on the probability ratio data in Table VIII assumes a somewhat less variable pattern than the slope data reported in Table VII. All t-test values except one for both 15- and 29-day groupings indicate, at the 5 percent significance level, that within each station the even versus odd year sampled data are the same.

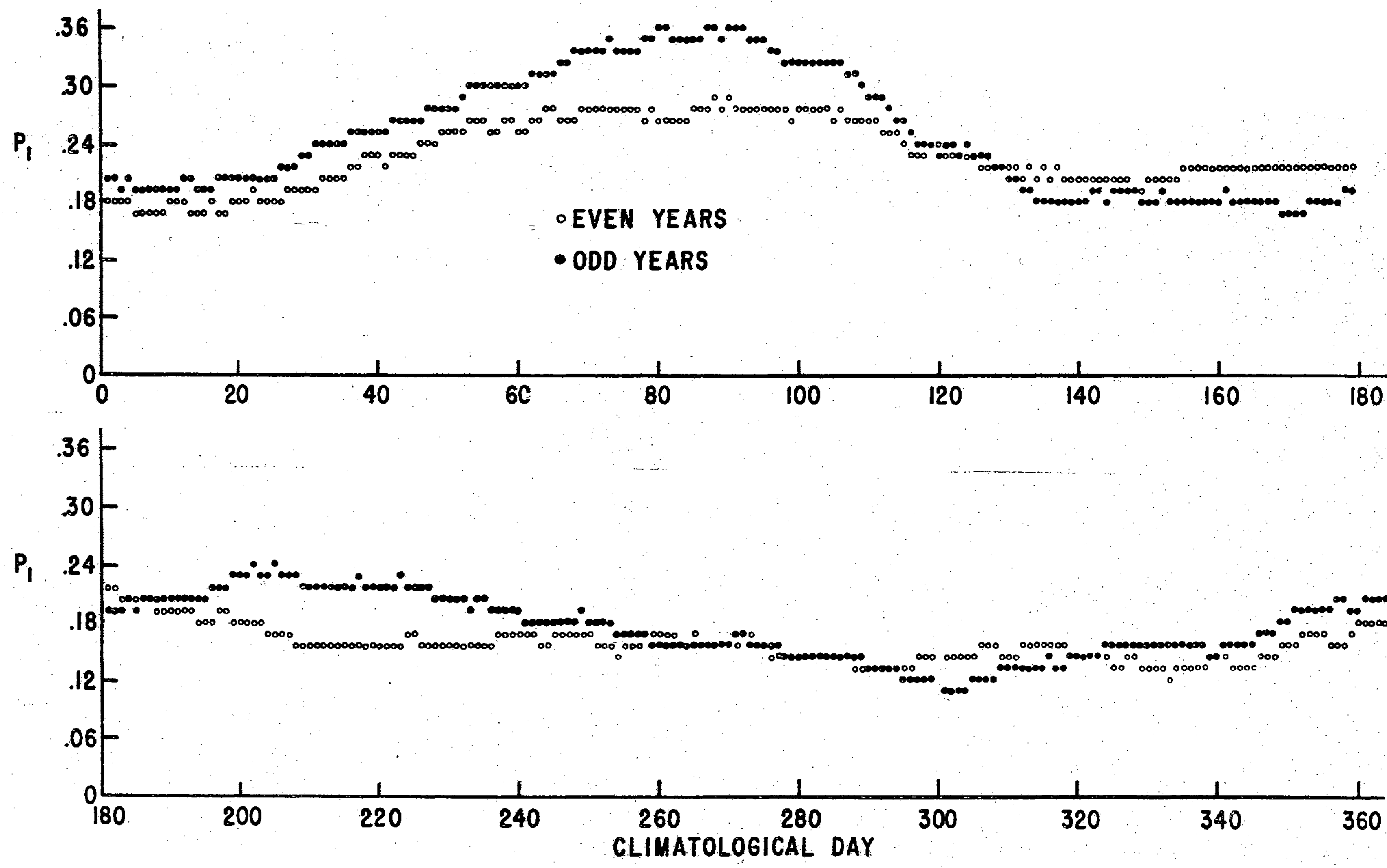

Figure 5. A Pawhuska Station Comparison of 29-Day P_1 Data Values for Even Versus Odd Years

TABLE VII

RESULTS OF t- AND RUNS-TEST ON SLOPE VALUES FOR EVEN VERSUS ODD YEAR PAIRED COMPARISONS OF FIVE CENTRAL OKLAHOMA STATIONS

Station	29-Day Grouping			
	1*	8	15	22
Pawhuska	2.01** 6***	1.76 4	1.64 6	2.52 4
Stillwater	2.38 4	1.81 4	1.43 4	0.90 4
Bristow	0.90 6	1.16 6	-0.85 6	0.78 4
El Reno	2.22 4	1.07 6	1.26 6	0.40 6
Waurika	1.35 2	2.25 2	0.21 2	0.98 6

* Climatological beginning midpoint date for the particular sized period.

** t-test value

*** Runs-test value

TABLE VIII

RESULTS OF t- AND RUNS-TESTS ON PROBABILITY RATIO VALUES FOR PAIRED COMPARISONS FOR EIGHT CENTRAL OKLAHOMA STATIONS FOR SPECIFIED SUB-SAMPLE PORTIONS OF THE TOTAL YEARS ON RECORD

	Even Years Versus Odd Years			First Half Versus Last Half Years		
	15-Day Grouping	29-Day Grouping		15-Day Grouping	29-Day Grouping	
Station	1*	1	15	1	1	8
Newkirk	0.40**	0.21	0.42			
	10***	6	6			
Pawhuska	0.43	0.83	0.76	1.79	1.58	1.49
	16	8	8	10	4	2
Stillwater	-1.11	-0.81	-1.10	1.27	1.72	1.23
	10	4	2	12	2	2
Bristow	0.86	0.76	0.53			
	12	6	6			
El Reno	-0.29	-0.15	-0.02			
	14	4	6			
Shawnee	0.27	0.31	-0.33			
	16	6	8			
Waurika	-0.05	-0.47	-1.13			
	12	6	6			
Ardmore	1.75	2.08	1.89			
	12	6	4			

* Climatological beginning midpoint date for the particular sized period.
** t-test value
*** Runs-test value

Station-to-Station Probability Variation

P_1 Results

Figure 6 is presented here to report in one illustration the P_1 data for all eight stations under consideration so that the reader might visually compare all stations. The t- and runs-test values for station-to-station P_1 comparisons are reported in Table IX. It would seem that the t- and runs-test values of Table IX provide at least partially sensitive indicators for station-to-station evaluation of the P_1 function. However, they are not represented as statistical methods that completely describe the within-station or station-to-station P_1 function of interest in this study.

Two extremes of the Table IX data are evident. The most obvious one is the Shawnee-El Reno comparison. Values for t vary from 17.92 to 30.19, and with only one run these values are indeed unusual if $\Sigma D = 0$. This hypothesis obviously must be rejected.

Figure 7 includes plotted P_1 data values for El Reno, Shawnee, Newkirk and Waurika. The information derived from Table IX for the El Reno-Shawnee comparison shows: (1) no crossing over of data and (2) a maximum magnitude in P_1 differences as verified in the plotted data of Figure 6. Upon examination of the plotted data, it becomes evident that a maximum displacement exists between the Shawnee-El Reno P_1 values.

The plotted comparisons of Figure 7 also assist in establishing other significant relationships. A comparison of both the Newkirk and Waurika data with those of the discussed Shawnee-El Reno comparisons seems appropriate in the further defining of the areal P_1 function. The Newkirk-El Reno results of Table IX indicate similar P_1 differences with

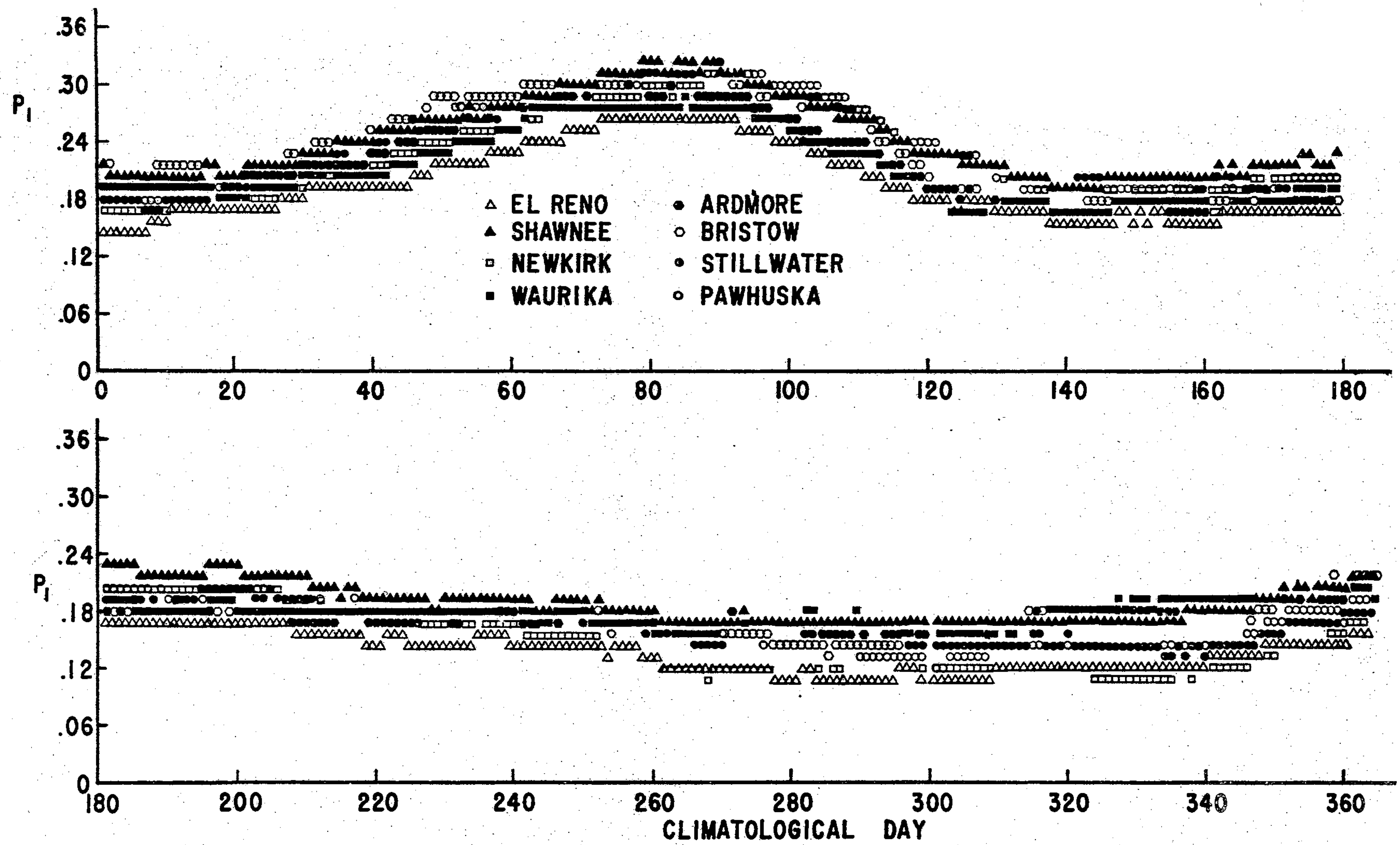

Figure 6. A Comparison of 29-Day P_1 Data Values for Eight Central Oklahoma Stations

TABLE IX

RESULTS OF t- AND RUNS-TESTS ON P_1 PAIRED COMPARISONS FOR EIGHT CENTRAL OKLAHOMA STATIONS

Station	15-Day Grouping		29-Day Grouping			
	1*	8	1	8	15	22
Newkirk-Pawhuska			-4.04**		-3.05	
			1***		4	
Newkirk-Stillwater			-2.46		-2.65	
			4		4	
Newkirk-Bristow			-1.79		-1.87	
			2		2	
Newkirk-El Reno			4.63		2.08	
			2		4	
Newkirk-Shawnee			-4.67		-3.65	
			2		2	
Newkirk-Waurika	-1.00	-1.06	-0.59	-0.58	-0.60	-0.60
	4	8	2	4	2	2
Newkirk-Ardmore	-1.30	-1.37	-0.80	-0.69	-0.85	-0.86
	6	4	2	2	2	2
Pawhuska-Bristow			-0.48		0.09	
			2		2	
Pawhuska-El Reno			12.22		7.74	
			1		1	
Pawhuska-Shawnee			-3.54		-2.53	
			2		2	
Pawhuska-Waurika			1.11		0.89	
			2		2	
Pawhuska-Ardmore			1.09		0.89	
			2		2	
Stillwater-Pawhuska	-0.29		-0.14		-0.13	
	12		8		6	
Stillwater-Bristow			-0.20		-0.27	
			4		2	
Stillwater-El Reno	13.76	11.44	16.55	11.67	9.43	10.65
	1	1	1	0	1	0
Stillwater-Shawnee			-3.75		-4.09	
			2		2	
Stillwater-Waurika	0.26		1.28		0.87	
	2		2		2	
Stillwater-Ardmore			1.21		0.80	
			4		2	

TABLE IX (Continued)

Station	15-Day Grouping 1*	8	29-Day Grouping 1	8	15	22
Bristow-El Reno			7.21 1		7.01 1	
Bristow-Shawnee			-2,24 2		-2.48 2	
Bristow-Waurika			1.77 4		1.39 4	
Bristow-Ardmore			1.84 4		1.84 8	
Shawnee-El Reno	24.33 1	24.91 1	26.77 1	30.19 0	26.08 0	17.92 0
Shawnee-Waurika			4.35 2		3.62 4	
Shawnee-Ardmore			4.60 1		4.31 4	
El Reno-Waurika			-5.27 1		-3.89 1	
El Reno-Ardmore			-6.45 1		-6.32 1	
Waurika-Ardmore			0.06 4		-0.71 4	

* Climatological beginning midpoint date for the particular sized period.

** t-test value

*** Runs-test value

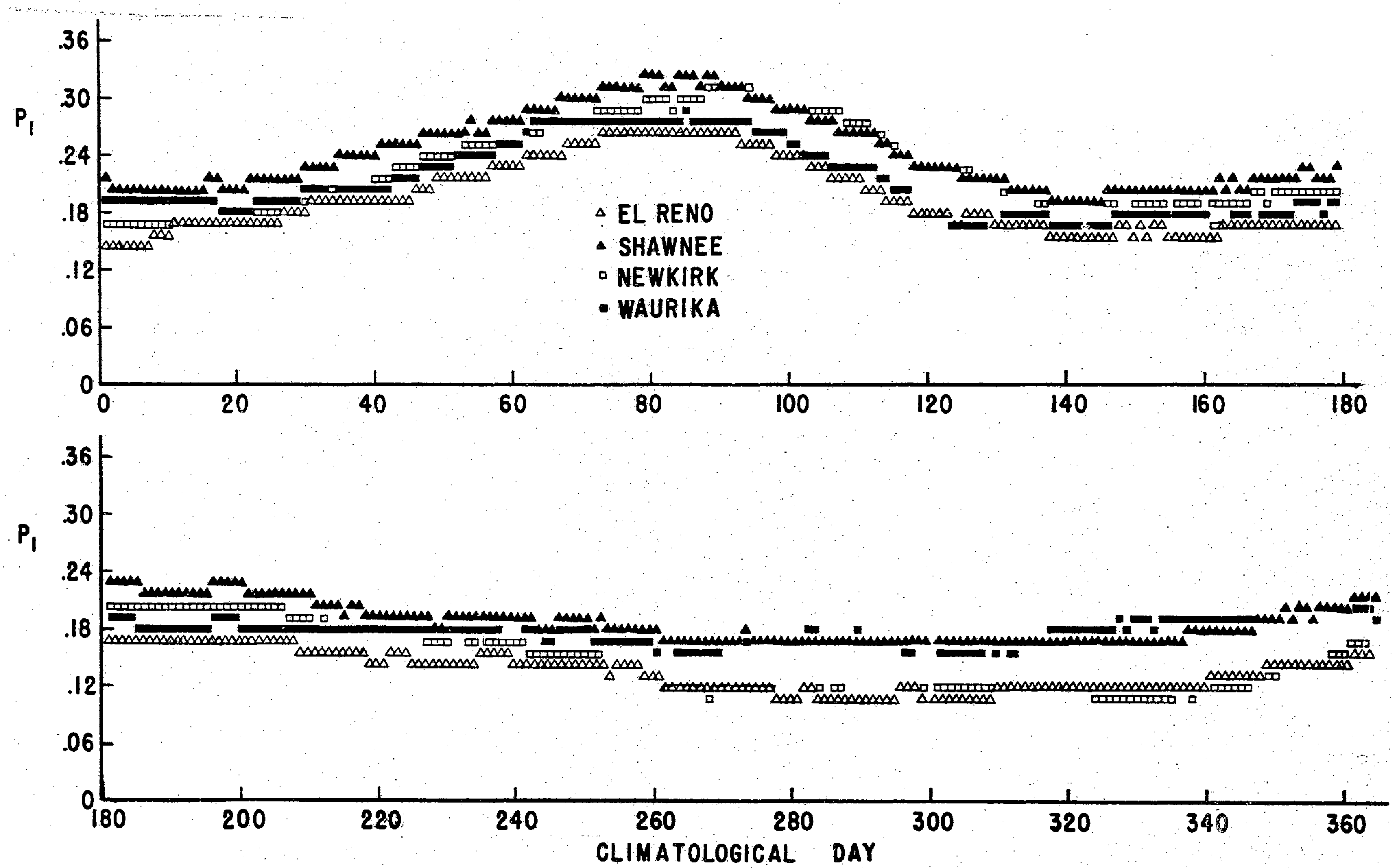

Figure 7. A Comparison of 29-Day P_1 Data Values for the Newkirk, El Reno, Shawnee and Waurika Stations

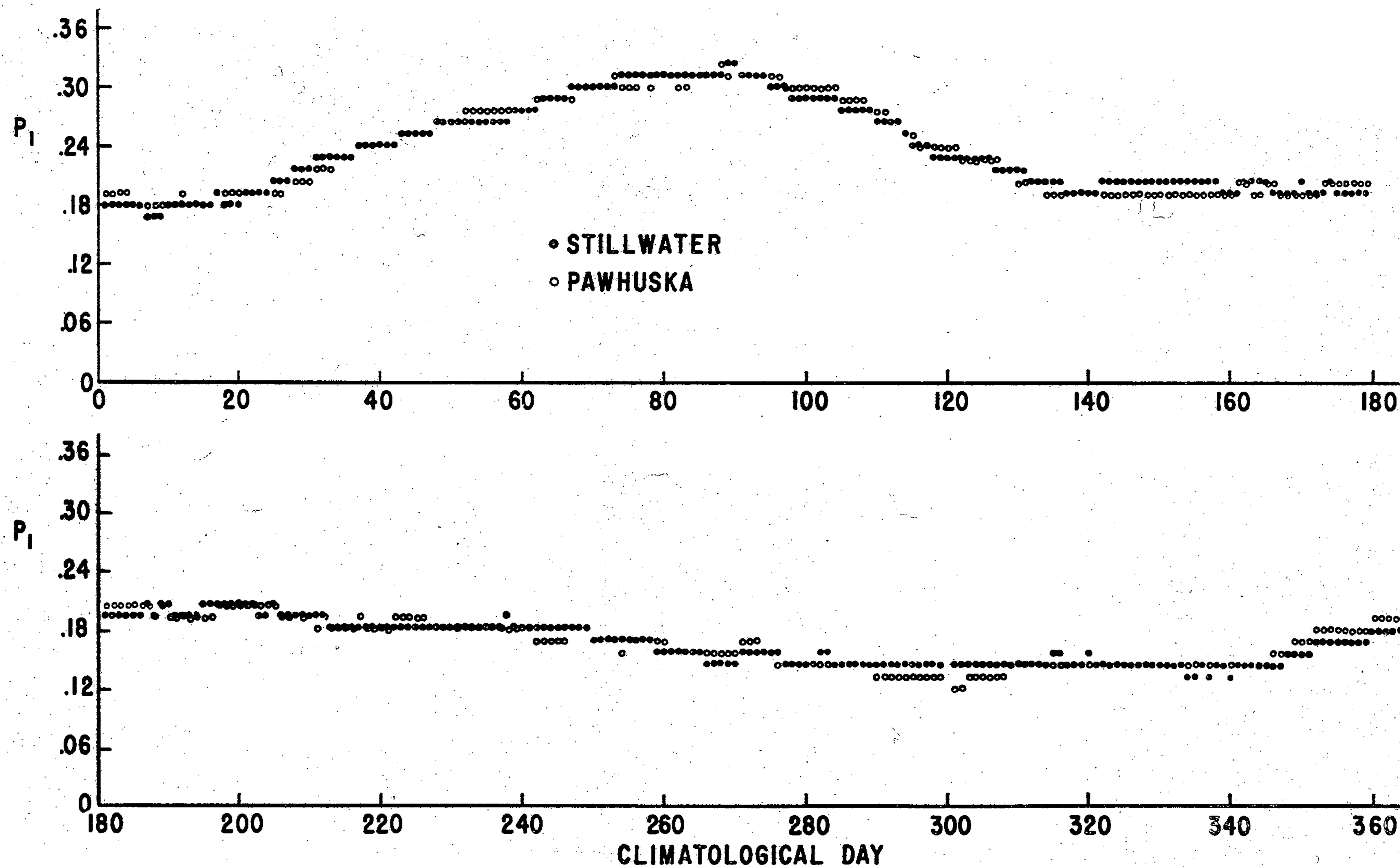

Figure 8. A Comparison of 29-Day P_1 Data Values for Pawhuska Versus the Stillwater Station

those in the Shawnee-Waurika comparison. However, the plotted data reveal certain patterns of area wide P_1 change that are not in evidence in the data of Table IX.

From day 250 through 40, the Newkirk-El Reno P_1 values are similar. From day 40 through 250, however, Newkirk P_1 values are consistently higher than corresponding El Reno values. Similar P_1 values are observed for Shawnee and Waurika from about day 240 through 40. This range of differences lies well above the Newkirk-El Reno comparisons previously referred to which represented approximately the same period. From day 40 through 240 Shawnee P_1 values are consistently above those corresponding to the Waurika station. During this period the Newkirk-Shawnee station P_1 values are similar.

Figure 8, which compares the plotted 29-day grouping P_1 data for Pawhuska and Stillwater, is presented as the outstanding example of a set of paired data which remains nearly identical throughout the entire year. The Pawhuska-Stillwater comparison could probably be considered identical for all dates. Table VIII results for Stillwater-Pawhuska certainly verify the similarities in evidence with the plotted data. Both the 29- and 15-day grouping t-test values are comparatively small. The 29-day grouping t-test values of -0.14 and -0.13 are the lowest of any station comparisons. The runs-test values for the 29-day grouping of the Stillwater-Pawhuska comparison are 8 and 6 for starting dates 1 and 15 respectively. These values further support the hypothesis that Stillwater's P_1 data is the same as that for Pawhuska. Other paired station-to-station data with t- and runs-test values which indicate quite similar data throughout the year are Waurika-Ardmore, and Bristow-Ardmore. The Waurika-Ardmore paired data as shown in Figure 6 indicate a slightly higher range in P_1 values for Ardmore from climatological day 20 through 130. For the remainder of the year, these two stations seem to have nearly identical values.

The station comparisons found in Figure 6, 7 and 8 seem to describe the following area P_1 change:

(1) The Shawnee station maintains consistently higher P_1 values compared to El Reno's comparatively low P_1 values throughout the year. Shawnee and El Reno seem to represent the upper and lower P_1 range of values as compared to the other six stations on most dates of the year.

(2) Comparable P_1 data from Waurika and Ardmore, the two southern-most stations of this report, is quite similar throughout the year. Both stations maintain high P_1 values in the range of Shawnee's from day 240 through 20. From day 20 through 240 these two stations have relatively low P_1 values approaching those of El Reno and also some P_1 values intermediate between Shawnee and El Reno.

(3) From day 20 through 120, Ardmore maintains slightly higher P_1 values as compared to Waurika. During the remainder of the year, days 120 through 20, the P_1 values are very similar.

(4) Newkirk and Pawhuska, the two northern-most stations of this report, are also similar throughout the year. They are nearly identical from day 100 through 240. During this time period, the P_1 data of these two stations is in a range with values nearly as high as those for Shawnee. From day 240 through 100, Newkirk P_1 values are moderately lower than those for Pawhuska. However, during most of this time the P_1 values for both of these stations are more nearly in the range of the relatively low El Reno values.

(5) The above described areal P_1 changes tend to identify a trend

for the southern-most stations to maintain relatively high P_1 values from about day 240 through 30 during which time Newkirk's values of the northern-most stations are in the lower range, approximately with the El Reno values. In this study the east-west station P_1 value comparisons crossing isohyet lines, only partly demonstrate the higher total rainfall amounts to the east. There does not seem to be any predictable pattern of P_1 change for the stations sampled across isohyet lines. Ardmore and Waurika data proved to be similar except for slightly higher P_1 values for Ardmore from day 20 through 30 while those for the Shawnee-El Reno comparison resulted in a maximum but near uniform spread. The Newkirk-Pawhuska comparison shows Newkirk to have moderately lower P_1 values than Pawhuska from day 240 through 100. During the remainder of the year, their P_1 values appear to be the same.

Stillwater-Pawhuska is the only paired combination of stations along equal isohyet lines that resulted in nearly identical P_1 values. Ardmore and Shawnee are located close to the 36-inch isohyet line. Their P_1 values are very similar from day 240 through 30. For the remainder of the year, Shawnee P_1 values are moderately higher. With Bristow's 37-inch total annual rainfall, it is nearly in the same isohyet as Shawnee and Ardmore. Bristow P_1 values resemble those for Shawnee, except Bristow's are moderately lower from days 110 through 300.

Newkirk, El Reno and Waurika are all approximately on the 31-inch rainfall isohyet. The previous evaluation of P_1 differences for these three stations has already indicated that their P_1 change pattern does not necessarily conform to isohyet lines.

Slope Results

The purpose of Figure 9 is to illustrate the general pattern of plotted slope values. The stations which have 29-day groupings available

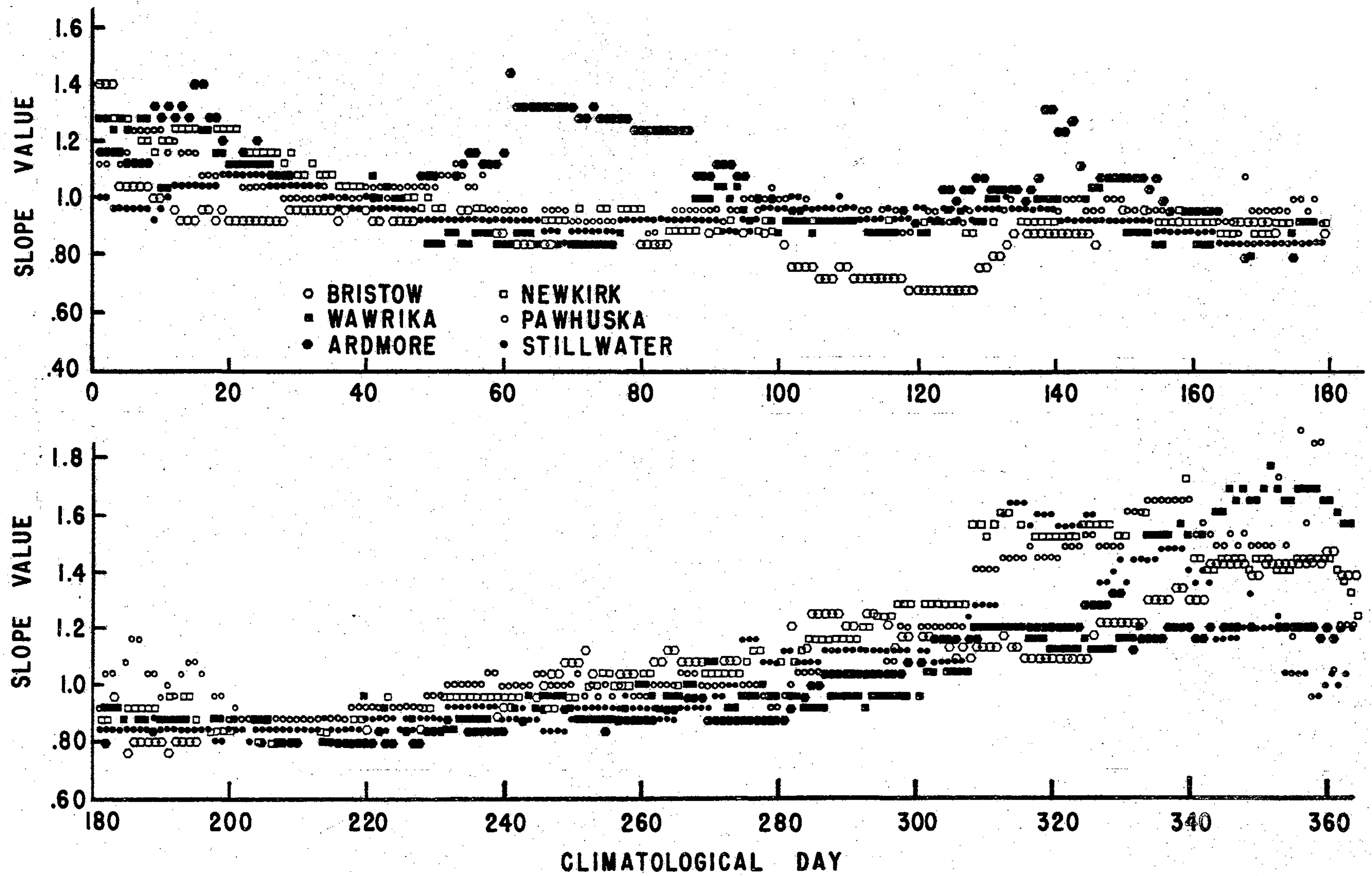

Figure 9. A Comparison of 29-Day Slope Data Values for Six Central Oklahoma Stations

are Stillwater, Pawhuska, Newkirk, Waurika, Bristow and Ardmore. Plotted observations for all these stations have certain characteristics in common that should be useful in evaluating P_2 change throughout the year. For example, slope values of each of the six stations reported in Figure 9 assume a similar pattern from day 60 through 240. At approximately day 240, slope values for all stations begin a gradual increase through about day 350 and then undergo a gradual decrease to day 60.

Table X reports the t- and runs-test values for station-to-station slope comparisons. The H_o that $\sum D = 0$ is not rejected for any station-to-station combination. The t-test data does not support the hypothesis that real station-to-station P_2 differences exist. A few examples of extreme variability of runs-test data for different starting dates are evident within paired data results. These include the Newkirk-Pawhuska, Newkirk-Bristow and the Stillwater-Pawhuska comparisons. The consistent and low value 29-day runs-test values for Newkirk-Waurika and Newkirk-Ardmore seem to indicate a crossing over that is not random. However, after examining the comparable 15-day slope data for these pairs of stations and the 29-day grouping data for all other stations, it would indeed be of questionable merit to attempt a station-to-station differentiation of slope differences. Both the Stillwater-Pawhuska t- and runs-test results for different beginning dates are indicative of the extreme variability of the data. Within-station P_2 values equal to zero seem to be responsible for part of the extreme differences in adjacent slope values for a given station. Both the zero P_2 values and general variability of rainfall with time are undoubtedly responsible for erratic statistical results for slope values. General features of the plotted curves should be helpful characteristics to be used in examining P_2 for change.

TABLE X

RESULTS OF t- AND RUNS-TESTS ON SLOPE AND PAIRED PROBABILITY RATIO COMPARISONS FOR EIGHT CENTRAL OKLAHOMA STATIONS

Station	Slope Comparisons						Probability Ratio	
	15-Day Grouping		29-Day Grouping				29-Day Grouping	
	1*	8	1	8	15	22	1	15
Newkirk-Pawhuska	0.88**	0.94	0.34	-0.73	0.84	0.11	-0.82	-0.28
	10***	12	2	4	6	6	6	8
Newkirk-Stillwater			1.56		2.41		2.54	2.25
			6		4		4	4
Newkirk-Bristow	1.60	2.01	1.45	2.21	3.18	2.18	-0.65	-0.38
	8	12	8	4	4	4	6	4
Newkirk-El Reno							0.25	0.95
							10	6
Newkirk-Shawnee							0.66	1.02
							4	4
Newkirk-Waurika	1.52	0.32	-0.10	-0.01	-0.13	-0.12	2.21	2.59
	10	14	2	2	2	2	4	4
Newkirk-Ardmore	0.76	-0.02	0.72	0.31	0.31	0.63	-3.02	-1.75
	10	4	2	2	2	2	2	2
Pawhuska-Bristow			1.16		0.86		0.06	-0.26
			4		4		4	2
Pawhuska-El Reno							1.18	1.31
							4	4
Pawhuska-Shawnee							1.24	1.21
							4	7
Pawhuska-Waurika			-0.21		-0.50		2.74	2.93
			2		2		4	4

TABLE X (Continued)

Station	Slope Comparisons: 15-Day Grouping 1*	15-Day Grouping 8	29-Day Grouping 1	29-Day Grouping 8	29-Day Grouping 15	29-Day Grouping 22	Probability Ratio 29-Day Grouping 1	Probability Ratio 29-Day Grouping 15
Pawhuska-Ardmore			0.69 2		-0.56 4		-1.99 4	-1.84 2
Stillwater-Pawhuska	0.29 10	1.58 12	-1.48 4	-3.59 2	-1.33 4	0.07 8	-3.31 4	-2.86 2
Stillwater-Bristow			0.91 8		1.51 6		-2.24 2	-2.09 2
Stillwater-El Reno							-1.59 6	-1.64 6
Stillwater-Shawnee							-1.14 4	-0.92 6
Stillwater-Waurika			-0.52 4		-0.79 2		-0.97 6	0.04 4
Stillwater-Ardmore			-0.10 2		-0.80 8		-4.97 1	-3.62 2
Bristow-El Reno							1.02 4	1.31 2
Bristow-Shawnee							1.47 4	1.84 4
Bristow-Waurika			-0.77 2		-1.26 2		2.96 2	3.67 2
Bristow-Ardmore			-0.66 6		-1.53 4		-1.85 2	-1.65 2
Shawnee-El Reno							-0.58 6	-0.79 4

TABLE X (Continued)

Station	Slope Comparisons: 15-Day Grouping 1*	15-Day Grouping 8	29-Day Grouping 1	29-Day Grouping 8	29-Day Grouping 15	29-Day Grouping 22	Probability Ratio 29-Day Grouping 1	29-Day Grouping 15
Shawnee-Waurika							3.08	2.07
							2	4
Shawnee-Ardmore							-3.94	-3.44
							2	2
El Reno-Waurika							2.63	3.27
							4	2
El Reno-Ardmore							-2.91	-3.17
							4	4
Waurika-Ardmore	-0.93	-0.86	0.54	0.53	0.33	0.54	-6.38	-5.48
	10	10	4	4	4	2	1	1

* Climatological beginning midpoint date for the particular sized period.

** t-test value

*** Runs-test value

Probability Ratios

An evaluation of the total plotted data for the various stations reported provides a useful comparison of probability ratio changes, as they relate to comparable slope changes. From the mathematical nature of the probability ratio, it seems unlikely that changes would not be meaningful if they occurred in both slope and probability values at the same points in time.

Figure 10 is presented to enable an evaluation of visual comparisons of plotted probability ratio values with corresponding slope values of Figure 9.

Only the general features of the slope and the probability ratio plotted data are of interest since no attempt will be made in this discussion to define possible station-to-station P_2 differences.

The probability ratio data of Table X are somewhat less variable than the slope data of the same Table. This might be expected since the P_1 value is the primary variable in the ratio denominator. As was reported earlier, the P_1 station-to-station comparison assumed well-defined patterns. The numerator of the ratio, however, is the appropriate total rainfall value for the years on record within either the 15- or 29-day grouping. The total rainfall value, due to infrequent occurrences of large rainfall amounts, could be a source of variability. In calculating slope values, rainfall increments were included up to 3 inches. High rainfall amounts which could tend to bias slope values were not included.

The t- and runs-test data of Table X, with few exceptions, indicate that there is relatively little difference between stations. With the limited methods for statistical treatment of this data, it is doubtful that any of the values for various stations at a particular date can be

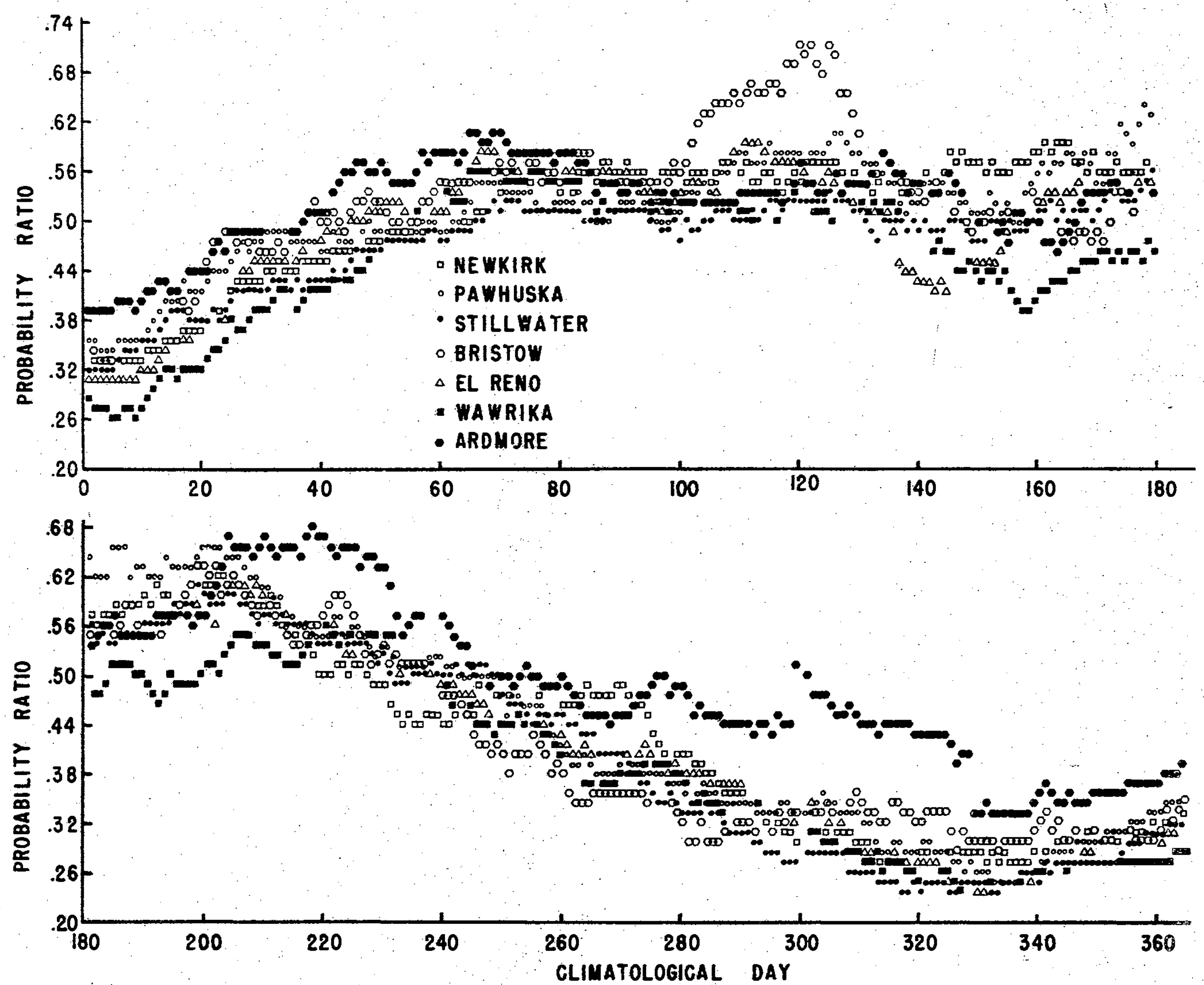

Figure 10. A Comparison of 29-Day Probability Ratio Data Values for Seven Central Oklahoma Stations

considered different. However, the Waurika-Ardmore values are unusual if H_o is $\sum D = 0$. Under this same hypothesis, the Waurika and Ardmore comparisons with other stations yield some unusual values.

Several concurrent slope versus probability ratio changes are in evidence by comparing the composite data for all stations. The absolute slope values of Figure 9 reach a maximum, and the probability ratio values are at a minimum for all stations at approximately day 340. From about day 340 through 70 slopes decrease in absolute value while probability ratio values increase. If slope and probability ratio values are useful in examining P_2 for change, then it is expected that concurrent changes in values resulting from both of these two methods can be related to P_2 change. For example, if a relatively high proportion of the total rainfall occurrences of a distribution exists as small incremental events, this condition should be reflected by a relatively high absolute slope value. The probability ratio value corresponding to this relatively high slope value would likely be relatively low. This is true since a relatively high proportion of small rainfall occurrences in a distribution would generally result in a relative increase in the numerical P_1 value of the probability ratio denominator with a consequent lowering of the probability ratio value. Since both of these methods used in examining P_2 for change reflect a general concurrent change for all stations, it is assumed that P_2 is likewise a changing function from day 340 through 70. Increasing numerically from day 70, both slopes and probability ratios are relatively constant. Slope values remain in this condition to approximately day 240. Ratio values for the various stations are also approximately constant from day 70 through 240 except for the tendency to peak from day 160 through day 240, with the peak centering at about day 200. This peak occurs during a period in which slope values can be considered constant. Since there is possible bias introduced by fitting a straight line to the semi-log plottings, constancy of slope values does not necessarily assure a constant P_2 situation. Since both probability ratios and slopes do not vary concurrently during days 160 through 240, P_2 constancy might be considered somewhat uncertain. However, high occurrence of unusually large rainfall events during this time seems responsible for the tendency of probability ratio data to peak.

Constant Probability Periods

The Earlier P_1 and P_2 Study

An objective of the earlier study was to characterize rainfall through defining P_1 and P_2 functions for the Stillwater station. Constancy of these two probabilities was used as the criterion by which the year was divided into periods.

The possibility of smoothing data by obtaining pooled P_1 and P_2 values within each of these periods was hypothesized as an advantage of this system. If constancy of both P_1 and P_2 exists within a particular period of time, there should be a valid argument for pooling total incremental rainfall occurrences into one frequency distribution. This hypothesis seemed to be supported by results of the earlier work, since pooling did result in smoothed P_1 and P_2 distributions.

One purpose of the present study is to provide further statistical evidence that these periods of the earlier study did in reality represent different P_1 and P_2 situations.

The slope procedure evaluation of the function P_2 in the pooling study had many similarities to that described for the present study. Slope values derived from semi-log graphing versus their corresponding P_2 values were used to evaluate the P_2 function. Periods of relatively constant P_2 throughout the year were developed on slope values for the 28-day consecutive non-overlapping periods. An evaluation of the P_1 function was made from 28-day non-overlapping period data.

Periods of constant P_2 as determined by slope values obtained from 14- and 28-day groupings are as follows:

Period 1 includes weeks 51 through 8.
Period 2 includes weeks 9 through 36.
Period 3 includes weeks 37 through 50.

Slope values for different groupings within each of these periods are considered relatively constant. Consecutive P_1 values within each of the three P_2 periods sometimes varied more than 10 percent from one 28-day grouping to another. When this occurred, the P_2 periods were further divided into periods within which both P_1 and P_2 were considered constant.

Table XI lists a contingency table of the frequency distributions for incremental rainfall occurrences within each of the three P_2 periods of the earlier study. The H_o that corresponding P_2 values for all three periods are the same is rejected, even at the one percent level. With 18 degrees of freedom, the calculated X^2 value is 41.90. The $X^2_{(0.01)}$ is 32.00.

The three P_2 periods in Table XI were further divided into eight P_1 periods. Within each of these eight periods, P_1 was considered to be constant. Table XII lists the frequency distributions of incremental rainfall occurrences for each of the eight P_1 periods.

The H_o is made that frequencies for each of the eight P_1 periods of

TABLE XI

CONTINGENCY TABLE OF FREQUENCY DISTRIBUTION OF INCREMENTAL RAINFALL AMOUNTS WITHIN THREE P_2 PERIODS OF THE STILLWATER STATION

P_2 Period	Incremental Rainfall Amount (Inches)									
	.01-.09	.10-.19	.20-.29	.30-.39	.40-.49	.50-.59	.60-.69	.70-.79	.80-.89	.90-.99
1 (Weeks 51-8)	292	141	97	70	49	37	44	27	18	16
2 (Weeks 9-36)	799	399	273	208	165	127	110	97	94	60
3 (Weeks 37-50)	395	147	120	93	60	39	25	32	18	13

TABLE XII

CONTINGENCY TABLE OF FREQUENCY DISTRIBUTION OF INCREMENTAL RAINFALL AMOUNTS WITHIN EIGHT P_1 PERIODS OF THE STILLWATER STATION

P_1 and P_2 Periods			Incremental Rainfall Amount (Inches)									
P_2 Period	P_1 Period	Weeks	.01-.09	.10-.19	.20-.29	.30-.39	.40-.49	.50-.59	.60-.69	.70-.79	.80-.89	.90-.99
1	1	51-52	65	26	20	15	11	4	6	4	1	3
	2	01-04	90	52	32	26	15	19	9	13	9	6
	3	05-08	137	63	45	29	23	14	29	10	8	7
2	4	09-12	162	78	45	36	29	21	18	19	20	15
	5	13-16	127	78	55	36	27	18	25	18	28	12
	6	17-20	108	58	39	32	32	19	13	14	12	7
	7	21-36	402	185	134	104	77	69	54	46	34	26
3	8	37-50	395	147	120	93	60	39	25	32	18	13

Table XII are the same. With 63 degrees of freedom, the calculated X^2 value is 94.68. Obviously, this H_o can be rejected, even at the one percent level.

In Table XII three P_1 periods are listed for P_2 period 1. For P_2 period 2, four P_1 periods are listed. The H_o is made that the rainfall frequencies for each of the three P_1 divisions of P_2 period 1 are the same. A similar H_o is made for the four P_1 periods of P_2 period 2. With 18 degrees of freedom, the calculated X^2 for P_2 period 1 is 20.71. This leaves a 10 to 25 percent probability of a greater X^2 value. Therefore, the H_o of constancy of P_1 periods within P_2 period 1 can be rejected only at the 10 to 25 percent significance level. The calculated X^2 for P_2 period 2 with 27 degrees of freedom is 28.09. In repeated sampling, there is a 25 to 50 percent chance of greater value. Hence, an argument for lack of constancy of P_1 periods within P_2 period 2 is supported only at a 25 to 50 percent significance level.

Present Study

The characteristics of the probability ratios and semi-log slope values have been tested for use in examining P_2 for change. Based on the data of these two independent approaches, P_2 of the periods from about day 60 through 240 could be considered constant. This period corresponds very closely to period 2 of the earlier study. From day 240 through 60, P_2 is a variable function; however, both semi-log slope values and probability ratios report the same general direction or pattern of variability for corresponding dates. Periods 1 and 3 of the earlier study represent a nearly identical time period as has been described from day 240 through 60. The 29-day equally weighted moving average semi-log and probability ratio values for different stations gradually change over most of this period. Based on this pattern of P_2 change from day 240 through 60, several constant P_2 periods could be established on the premise that P_2 constancy within each period makes possible such periods throughout the year.

In describing rainfall input, periods of both P_1 and P_2 constancy must be obtained if the advantages of pooling rainfall frequencies are realized for each of several periods throughout the year. It is not within the purpose of the present study to divide each station into such periods. However, judging from the P_1 and P_2 functions obtained for various stations this division and consequent pooling seems tenable as a future objective.

CHAPTER V

SUMMARY AND CONCLUSIONS

A major interest of this study was to evaluate the need for smoothing rainfall frequencies with central Oklahoma stations. Eight stations were selected for this purpose. Computer techniques were employed for obtaining all tabulated and plotted data. Two rainfall probabilities were suggested to assist in defining rainfall input. P_1 is the probability of any amount of rainfall. P_2 is the conditional probability of receiving specified amounts of rainfall provided rainfall occurs.

Students' t- and runs-tests and visual observation of plotted data were used to assist in the comparison of P_1 and P_2 functions for the eight stations at different degrees of smoothing.

Low frequency components of the P_1 and P_2 functions were found to be useful for making t- and runs-tests of within-station and station-to-station paired data comparisons and visual comparisons. The cyclic appearance of plotted P_1 data for 3- to 7-day equally weighted moving averages is largely introduced by this method of smoothing. The longer non-overlapping 15- and 29-day periods were used in this study. These longer periods eliminated extreme scattering and cyclic effects associated with the shorter groupings.

Even year versus odd year and first half versus last half year probability tests were made to characterize within-station variability of P_1 and P_2. Differences in probabilities resulting from these comparisons seem to be of random origin. There is no particular reason to believe that data from the first half, or even years, of record represent a different population of rainfall events than does data from last half, or odd years, of record.

The P_1 function assumes a pattern with well-defined seasonal and station-to-station differences. El Reno has consistently low and Shawnee consistently high P_1 values throughout the year. This east-west P_1 differentiation assumes a less predictable pattern with other combinations of stations. The southern-most stations, Waurika and Ardmore, maintain relatively high P_1 values during the approximate time from day 240 through 20 with quite intermediate to low values through the remainder of the year. The northern-most stations, Newkirk and Pawhuska, have comparatively high P_1 values from about day 60 through 240 with intermediate to low values during the remainder of the year. In the Waurika-Ardmore and Newkirk-Pawhuska comparisons, the lower P_1 values of the western-most stations are not as evident as with El Reno-Shawnee.

Two independent methods were used to characterize the P_2 functions. One of these methods consisted of comparing semi-log slope values of straight lines located by P_2 datum points. The second method, termed the probability ratio, was derived from a total rainfall value as the numerator with its corresponding P_1 value as the denominator. These two methods provide concurrent evidence for locating points in time at which significant P_2 changes occur. P_2 can probably be assumed constant from about day 60 through 240. From day 240 through 60, both methods show P_2 as a changing function.

Relatively long periods of similar rainfall probabilities were established for the Stillwater Station (4). The X^2 test was used to provide additional evidence of the reality of seasonal probability change. High frequency fluctuations were experienced in short-term rainfall probabilities. This difficulty led to procedures for averaging out the high frequency variation components. These procedures primarily consisted of pooling rainfall frequencies within the relatively long periods of constant probabilities.

Describing rainfall input by future pooling of rainfall occurrences throughout the year in periods of both P_1 and P_2 constancy would seem to be a tenable consequence of this present study.

LITERATURE CITED

1. Barger, G. L. and H. C. S. Thom. Evaluation of Drought Hazard. Agron. J. 41:519-526. 1949.

2. Caskey, Jr., J. E. A Markov Chain Model for the Probability of Precipitation Occurrence in Intervals of Various Length. Monthly Weather Review 91:298-301. 1963.

3. Dingle, A. N. Patterns of Change of Precipitation in the United States. J. Meteor. 11:220-225. 1955.

4. Duffin, R. B. An Analysis of Precipitation Data for Testing the Validity of Pooling Rainfall Occurrences Within Periods of Similar Rainfall Probabilities. M. S. Thesis, Oklahoma State Univ. 1963.

5. Eichmeier, A. H. and W. D. Baten. Rainfall Probabilities During the Crop Season in Southern Lower Michigan. Monthly Weather Review. 90:277-281. 1962.

6. Feyerherm, A. M. and L. D. Bark. Statistical Methods for Persistent Precipitation Patterns. Paper for Sixth National AMS Conf. on Agric. Meteor. 1964.

7. Friedman, D. G. and J. E. Janes. Estimation of Rainfall Probabilities. Conn. (Storrs) Agri. Expt. Sta. Bull. 332:1-22. 1957.

8. Gleeson, T. A. Some Tests and Developments of a Statistical Theory of Synoptic-Scale Meteorology. J. Appl. Meteor. 1:18-19. 1962.

9. Hershfield, D. M. A Note on the Variability of Annual Precipitation. J. Applied Meteor. 1:575-578. 1962.

10. Huff, F. A. and J. C. Neill. Areal Representativeness of Point Rainfall. Trans. Amer. Geophys. Union 38:341-346. 1957.

11. Holloway, Jr., J. L. Smoothing and Filtering of Time Series and Space Fields. Advances in Geophy. Academic Press:351-389. 1958.

12. Jorgensen, D. L. A Computer Derived Synoptic Climatology of Precipitation from Winter Storms. J. Applied Meteor. 2:226-234. 1963.

13. Kershaw, Jr., H. A Comparison Between Frontal and Non-frontal Precipitation in Stillwater, Oklahoma. Unpublished Research of Meteor. Dept. O.S.U. 1965.

14. Kotz, S. and J. Neumann. On the Distribution of Precipitation Amounts for Periods of Increasing Length. J. Geophy. Res. 68:3635-3640. 1963.

15. ______________________. On the Distribution of Precipitation Amounts for Periods of Increasing Length. Addendum. J. Geophy. Res. 69:800-801. 1964.

16. ______________________. Shorter Contributions-Autocorrelations in Precipitation Amounts. J. Meteor. 16:683-685. 1959.

17. Lorenz, N. L. Meteorological Statistics and Data Handling. Trans. Amer. Geophys. Union 44:394-395. 1963.

18. Morrow, C. T. Averaging Time and Data-Reduction Time for Random Vibration Spectra. J. Acoustical Soc. Amer. 30:456-461. 1958.

19. Palmer, W. C. Moisture Variability and Drought Severity. U.S. Weather Bureau Paper for Sixth National AMS Conf. on Agric. Meteor. 1964.

20. Portig, W. H. A Search for Rainfall Calendaricities. J. Atmos. Sci. 21:462. 1964.

21. Siegel, Sidney. Nonparametric Statistics for the Behavioral Sciences. McGraw-Hill Book Company, Inc. New York, New York. 1956.

22. Slutzky, E. The Summation of Random Causes as the Source of Cyclic Processes. Econometrica 5:105-146. 1937.

23. Thom, H. C. S. A Note on the Gamma Distribution. Monthly Weather Review 86:117-122. 1958.

24. Topil, A. G. Precipitation Probability at Denver as Related to Length of Period. Monthly Weather Review 91:293-297. 1963.

25. Wiser, E. H. Modified Markov Probability Models of Sequences of Precipitation Events. Monthly Weather Review 93:511-516. 1965.

VITA

Robert Beck Duffin

Candidate for the Degree of

Doctor of Philosophy

Thesis: AN APPROACH TO RAINFALL PROBABILITY ANALYSES ON SELECTED CENTRAL OKLAHOMA STATIONS

Major Field: Soil Science

Biographical:

Personal Data: Born at Aberdeen, Idaho, August 11, 1926, the son of Wallace H. and Thelma A. Duffin.

Education: Graduated from Aberdeen High School, Aberdeen, Idaho, 1944; received the Bachelor of Science Degree in Agronomy from Utah State University, 1950; received the Master of Science degree in Soil Science from Oklahoma State University in 1963; completed the requirements for the Doctor of Philosophy degree in May, 1967.

Experience: Reared on irrigated farm at Aberdeen, Idaho; Soil Scientist for Bureau of Reclamation, summer, 1950; Assistant, Associate, and Extension Irrigation Specialist, Oklahoma Agricultural Extension Service, 1951-1967.

Member of: Epsilon Sigma Phi, American Society of Agricultural Engineers, Sprinkler Irrigation Association, and National Reclamation Association.