Application of Some Statistical Tests to Investigate the Urban Rainfall Characteristics

RAMACHANDRA A. RAO and R. G. S. RAO
School of Civil Engineering
Purdue University, West Lafayette, Indiana 47907

Abstract

This study reports statistical analyses of the annual rainfall data from urban and surrounding rural areas. The Rank, Median, Mann-Whitney-U, Run and U-tests of significance were used. Annual rainfall data from stations in Indiana and Oklahoma were analyzed. Results indicate that rainfall characteristics of areas adjacent to large metropolitan areas are significantly different from those of surrounding rural areas.

Introduction

Cities have different climates than the surrounding countryside. The difference results from the effects of several interacting variables such as temperature, amounts of water vapor and precipitation, etc. The rainfall in urban as well as adjacent areas is affected by factors such as increased particulation and convection which are essentially due to the presence of these urban areas.

Previous studies (we have selected only a few from the vast literature available on inadvertent weather and precipitation modification by urbanization (2, 4, 5)) have established clearly that relatively significant micro and meso scale changes in weather occur in and near urban areas. The effect of urbanization and industrialization on rainfall characteristics has been investigated by using several methods such as analysis of rainfall records (3), runoff data (9), etc. The statistical analysis of the effects of urban-industrial effects on rainfall, however, has not received much attention. Only a few studies (7) of the effects of urbanization on rainfall appear to have been conducted so far by using statistical methods. Consequently, statistical analyses of the changes brought about by the urbanization on the rainfall characteristics were undertaken. The present paper deals with a part of the preliminary results obtained by the study.

The annual precipitation data were used in the study. Data from two locations in Indiana and in Oklahoma were selected for the study (Fig. 1). The LaPorte, Indiana, area has been shown to receive consistently higher rainfall than the surrounding areas. This phenomenon was called the "LaPorte Anomaly", by Changnon (3) and was attributed to the effects of Chicago-Gary area. The rainfall at and near Tulsa, Oklahoma, was shown by Landsberg (10) to increase with increasing urbanization in Tulsa. Consequently, these two areas were selected for further investigation.

The nature of the raw data plays a significant role in the statistical analysis of data. Main sources of error are, changes in raingage location, changes in methods of observation (human *vs.* automatic), etc. Some of these aspects of the raw data have created considerable controversy (9), *mainly* in the discussion of LaPorte Anomaly. Some of these aspects are discussed in a forthcoming report and are not presently included for lack of space. However, even if we make the rainfall values in the affected period 10% smaller than the reported values for LaPorte, the conclusions presented herein are still valid.

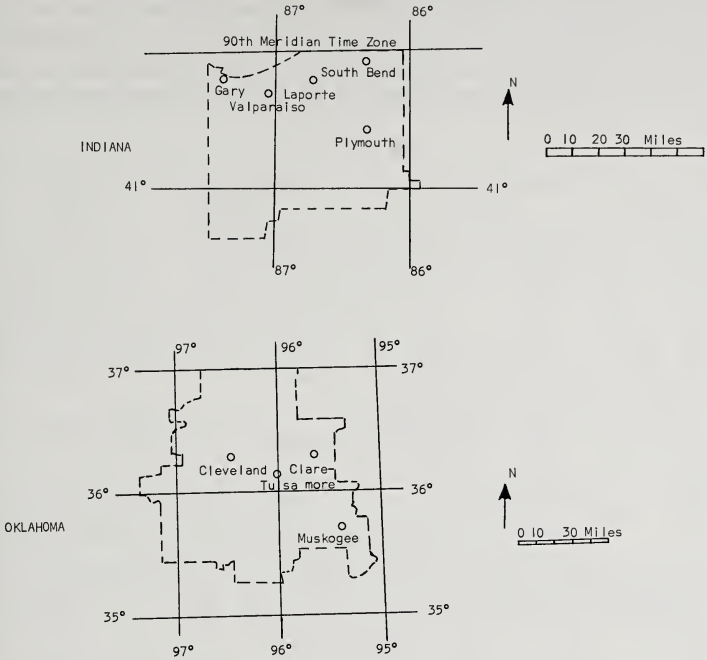

FIGURE 1. *Locations of stations.*

The annual and mean rainfall at a few stations near LaPorte, Indiana, and Tulsa, Oklahoma, are shown in Figure 2. The mean annual rainfall at the Indiana and Oklahoma stations were shown to change at 1929 and 1931, respectively, by using double mass curve and other techniques. For the data from LaPorte and South Bend, the mean annual rainfall *increased* at about 1929, with the mean showing a larger increase for the data from LaPorte than for the data from South Bend. For the data from Oklahoma, the mean annual rainfall *decreased* at Tulsa and Claremore starting from 1931. In the present paper the period before 1929 and 1931, respectively, for Indiana and Oklahoma, is referred to as the *unaffected* period and the period after is called the *affected* period. The mean and standard deviation of data from the affected and the unaffected periods and the percentage change in mean values are also shown in Figure 2.

From the data presented in Figure 2, it is clear that the mean values of annual rainfall have increased for stations such as LaPorte and South Bend, Indiana, and have decreased for Tulsa, Oklahoma. The main objective of the present paper is to test whether these changes are statistically significant and not to correlate the rainfall with urbanization indicators. The statistical tests, then, must indicate whether the changes in the rainfall characteristics observed at stations such as La Porte and Tulsa are significant. If these changes in means, medians, etc., of the data from the affected and unaffected periods are significant, then they *may be* attributed to the effects of urbanization. We would like to emphasize, however, that the investigation of the effects of urbanization must not only be statistical but also be reinforced by studies of the "Metromex" type (6). The data on the urbanization of the Chicago-Gary area, and Tulsa, may be found among the references cited.

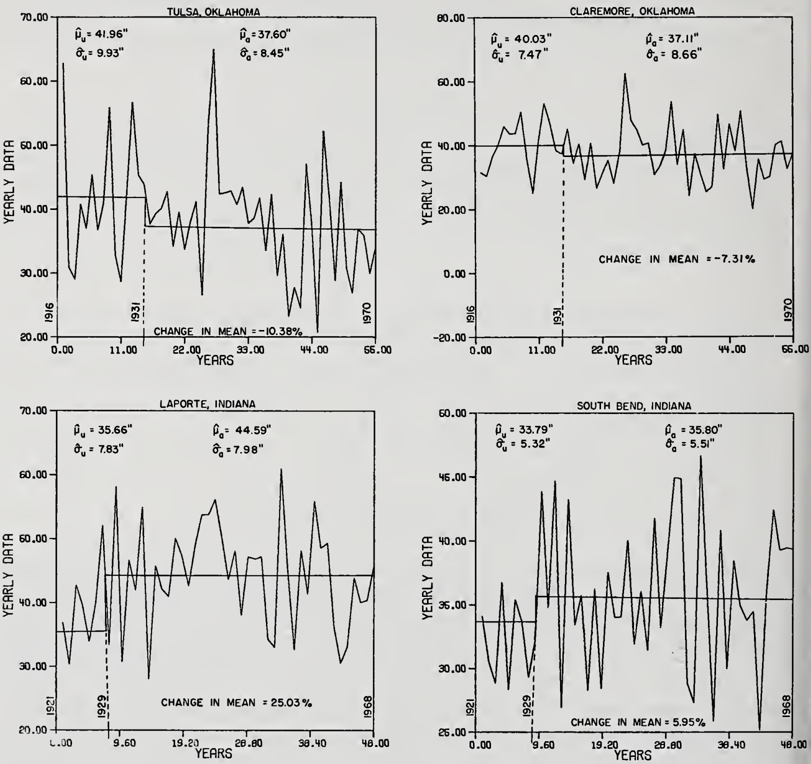

FIGURE 2. *Variation in annual rainfall at some representative stations.*

Statistical Tests Used in the Study

The statistical tests used in the study are of the nature of hypothesis testing. In the following discussion, the null and alternative hypotheses are represented as H_o and H_a, and the annual precipitation values in the affected and unaffected periods are respectively designated as P_i and P_j. The total number of P_i and P_j values are n and m, and $N = n+m$. The significance level at which the null hypothesis is rejected is represented by α and α was fixed at 5% for all the tests. The details of all the tests may be found in standard reference works (1).

Rank Test

The null hypothesis H_o to be accepted or rejected by the application of the rank test is that the affected and unaffected period populations are identical.

The test statistic Z_o is given by Equation 1, in which T_a is the sum of rank of precipitation values in the affected period.

$$Z_o = \frac{T_a - \bar{T}}{\sigma} \qquad [1]$$

The precipitation values P_i and P_j are combined and ranked in increasing order such that (if there are no ties) the smallest and largest values have ranks 1 and N. The value of T_a is estimated by adding the ranks of precipitation values in the affected period. $\bar{T}$ and σ are given in Equation 2.

$$\left.\begin{aligned} \bar{T} &= \frac{m(n+m+1)}{2} = \text{expected value of } T_a. \\ \sigma^2 &= m(n)\ (n+m+1)/12 \end{aligned}\right\} \qquad [2]$$

The decision rule for the Rank test is given below.

If $Z_O > N_{1-a}$: reject H_O (Affected and unaffected period populations are not identical).

If $Z_O \leqslant N_{1-a}$: accept H_O (Affected and unaffected period populations are identical).

N_{1-a} is the value of the unit normal variate at a significance level α.

The results of the rank test given in Table 1 indicate that the null hypothesis can be accepted for the data from all the stations. Consequently, for the data from all, no statistically significant changes may be claimed to be present in the populations from the affected and unaffected periods.

TABLE 1. *Summary of results of rank test.*

Station	Test Statistic Z_O	Critical Value N_{1-a}	Is $Z_O > N_{1-a}$?
Indiana			
LaPorte	—1.1484	1.645	No
South Bend	1.2180	1.645	No
Valparaiso	— .2204	1.645	No
Plymouth	.2436	1.645	No
Oklahoma			
Tulsa	—1.3796	1.645	No
Claremore	—1.3229	1.645	No
Cleveland	— .6992	1.645	No
Muskogee	0.1701	1.645	No

Mann-Whitney U-test

The null and alternate hypotheses H_O and H_a tested by the Mann-Whitney U-test are: H_O: P_i and P_j have the same distribution; H_a: the location parameter of P_j is greater than the location parameter P_i. The test statistic w is defined in Equation 3.

$$w = \frac{U_M - \bar{U}_M}{\sigma_M} \quad [3]$$

In Equation 3, U_M is the number of times P_j values are larger than P_i values. U_M is estimated by using Equation 4, in which T_a is the sum of ranks of precipitation values in the affected period (see Rank Test).

$$U_M = nm + \frac{m(m+1)}{2} - Ta \quad [4]$$

$$\bar{U}_M = \frac{nm}{2} \; ; \; \sigma_M^2 = \frac{nm(n+m+1)}{12} \quad [5]$$

The decision rule for the test is that if $w > N_{1-a}$, then the null hypothesis is rejected. From the results shown in Table 2, it can be seen that the null hypothesis must be accepted for the data from all the stations. The implication is that the probability distributions of the data from unaffected and affected periods are not statistically significant. Although the Rank and the Mann-Whitney U-tests are similar, the Mann-Whitney U-test is more powerful.

TABLE 2. *Summary of results of Mann-Whitney U-test.*

Station	Test Statistic W	Critical Value N_{1-a}	Is $W > N_{1-a}$?
Indiana			
LaPorte	1.1484	1.645	No
South Bend	—1.2180	1.645	No
Valparaiso	.2204	1.645	No
Plymouth	— .2436	1.645	No
Oklahoma			
Tulsa	1.3796	1.645	No
Claremore	1.3229	1.645	No
Cleveland	.6992	1.645	No
Muskogee	— .1701	1.645	No

Run Test

The null hypothesis H_O in the run test is similar to that of the Rank test and is given below.

H_O: The samples from the affected and unaffected periods arise from the same distribution.

The statistic U_R used in the run test is defined in Equation 6 in terms of the expected value $\bar{r}$ and standard deviation σ_r of run lengths. A run is defined as an unbroken sequence of values of P_i or P_j.

$$U_R = \frac{r - \bar{r}}{\sigma_r} \qquad [6]$$

If the samples P_i and P_j are from the same population, then the observations in the unaffected and affected periods will be well mixed and the number of runs will be large. Mean $\bar{r}$ and the standard deviation σ_r of runs are computed by Equations 7 and 8.

$$\bar{r} = \frac{2nm}{N} \qquad [7]$$

$$\sigma_r^2 = \frac{2nm(2nm-m-n)}{N^2(n+m-1)} \qquad [8]$$

The decision rule is that if $U_R > N_{1-\alpha}$, then the null hypothesis is rejected at a level of significance α. From the results of Run Test presented in Table 3, the null hypothesis can be rejected for data from all the stations except LaPorte and Plymouth.

TABLE 3. *Summary of results of run test.*

Station	Test Statistic U_R	Critical Value $N_{1-\alpha}$	Is $U_R > N_{1-\alpha}$?
Indiana			
LaPorte	2.64	1.645	Yes
South Bend	1.301	1.645	No
Valparaiso	— .357	1.645	No
Plymouth	1.873	1.645	Yes
Oklahoma			
Tulsa	0.283	1.645	No
Claremore	1.097	1.645	No
Cleveland	1.096	1.645	No
Muskogee	0.407	1.645	No

Median Test

The median test is used to test the significance of the change in the median values of the data from the affected and unaffected periods. The null hypothesis for this test is given below.

H_0: There is no change in the median values of the data from the affected and unaffected periods.

The data from the affected and unaffected periods are combined and arranged in the increasing order of magnitude. The number of P_i and

P_j values above and below the combined median values are determined. Let the number of P_i values above and below the combined median by n_{1a} and n_{1b} and the number of P_j values above and below the combined median be n_{2a} and n_{2b}. The test statistic M is defined as in Equation 9.

$$M = \frac{(\,|\,2n_{1a} - (n_{1a}+n_{1b})\,|\, - 1)^2}{n} + \frac{(\,|\,2n_{2a} - (n_{2a}+n_{2b})\,|\, - 1)^2}{m} \qquad [9]$$

If $M > \chi_{1-a}(1)$ then the null hypothesis that there is a change in the median values of data from the affected and unaffected periods is rejected. The results obtained by applying the median test to the data from several stations in Indiana and Oklahoma shown in Table 4 indicate that there is no statistically significant change in the median values from affected and unaffected periods.

TABLE 4. *Summary of results of median test.*

Station	Test Statistic M	Critical Value $\chi^2_{1-a}(1)$	Is $M > \chi^2_{1-a}(1)$?
Indiana			
LaPorte	1.6879	3.841	No
South Bend	1.6489	3.841	No
Valparaiso	0.0074	3.841	No
Plymouth	0.4220	3.841	No
Oklahoma			
Tulsa	.2917	3.841	No
Claremore	0.2917	3.841	No
Cleveland	0.2917	3.841	No
Muskogee	0.0250	3.841	No

U-test

The statistical significance of the change in the mean value of precipitation in the unaffected and affected periods is tested by the U-test in which the following assumptions are made: 1) The precipitation data are normally distributed; 2) the precipitation values are uncorrelated; 3) the population parameters of precipitation are known. The first two assumptions are approximately valid for the data considered in the present study, whereas the third assumption cannot be verified.

The test statistic U_O is defined in terms of the estimated precipitation mean values in the unaffected (μ_U) and affected (μ_a) periods, and the standard deviations in the unaffected (σ_U) and affected (σ_a) periods. The test statistic U_O is defined in Equation 10.

H_O: There is a change in the mean value

$$U_O = \frac{\mu_a - \mu_U}{\sigma_U/\sqrt{m}} \qquad [10]$$

If the value of the statistic U_O is larger than N_{1-a} then the null hypothesis H_O is accepted. The results of application of the U-test to the data from Indiana and Oklahoma (Table 5) indicate that the null hypothesis can be accepted only for the data from LaPorte. In other words, the change in the mean values in the data from LaPorte are statistically significant.

TABLE 5. *Summary of results of U-test.*

Station	Test Statistic U_O	Critical Value N_{1-a}	Is $U_O > N_{1-a}$?
Indiana			
LaPorte	7.2142	1.645	Yes
South Bend	0.391	1.645	No
Valparaiso	0.1564	1.645	No
Plymouth	0.4863	1.645	No
Oklahoma			
Tulsa	—1.4248	1.645	No
Claremore	—1.518	1.645	No
Cleveland	— .8658	1.645	No
Muskogee	— .2061	1.645	No

Discussion and Conclusions

Of the five statistical tests discussed previously, the rank test, the Mann-Whitney U-test, and the run test deal with testing the significance of the changes in annual rainfall probability distributions in the affected and unaffected periods. The median test and the U-test are used to test the significance of changes in the median and mean values of annual rainfall in the unaffected and affected periods.

The results from the rank, and the run tests indicate that there may be significant changes in the probability distribution of annual precipitation at LaPorte, whereas the result from Mann-Whitney U-test gives the opposite result. Since the Mann-Whitney U-test is more powerful than the other two tests, the probability distributions of the precipitation data from the affected and unaffected periods may be assumed to be the same. This aspect, however, needs further study. The median values do not appear to have changed significantly from the unaffected to the affected period. The mean values of the data from LaPorte, on the other hand, appear to have increased from the unaffected to the affected period. Further analyses, which are not included here, also indicated a significant increase in the mean value in the data from LaPorte.

In view of these considerations the following conclusions may be drawn.

1) There is a statistically significant increase in mean annual rainfall at Laporte, Indiana, and Bristow, Oklahoma, which may be attributed to the effects of urbanization and industrialization.

2) The probability distributions and the median values of annual precipitation data do not appear to have been affected significantly.

Acknowledgments

This research was supported by funds from the Office of Water Resources Research, U. S. Dept. of the Interior, under the grant OWRR-B-025-IND. We are grateful to Dr. D. Wiersma, Director, Water Resources Research Center for administrative assistance.

Literature Cited

1. BROWNLEE, K. A. 1961. Statistical theory and methodology in science and engineering. John Wiley and Sons, Inc., New York, N.Y. 570 p.
2. CHANGNON, S. A. 1962. Air over cities. SEC Tech. Rept. A62-5, Pub. Health Serv., Cincinnati, O. 413 p.
3. ————. S. A. 1968. The LaPorte weather anomaly—fact or fiction? Bull. Amer. Meteorol. Soc. 50:4-11.
4. ————. 1969. Recent studies of urban effects of precipitation in the United States. Bull. Amer. Meteorol. Soc. 50:411-421.
5. ————. 1973. Inadvertent weather and precipitation modification by urbanization. ASCE J. Irrig. Drain. Div. 99:27-41.
6. ————, F. A. HUFF, and R. G. SEMONIN. 1971. Metromex: and investigation of inadvertent weather modification. Bull. Amer. Meteorol. Soc. 52:958-961.
7. ————, and P. T. SCHICKEDANZ. 1971. Statistical studies of inadvertent modification of precipitation. Proc. Interna. Symp. on Prob. and Stat. in the Atmos. Sci., Amer. Meteorol. Soc., Honolulu, Hawaii. p. 137-42.
8. HOLZMAN, B. G. 1971. LaPorte precipitation fallacy. Science 171:847.
9. HIDORE, J. J. 1971. The effects of accidental weather modification on flow of Kankakee River. Bull. Amer. Meteorol. Soc. 52:99-103.
10. LANDSBERG, H. E. 1956. The climate of towns. p. 586-606. *In* William L. Thomas, Jr. [ed.] Man's Role in Changing the Face of the Earth. Univ. Chicago Press, Chicago, Ill. 1193 p.