Mausam, (1988), **39**, 1, 27-36

551.577.53

Regional precipitation modification by urbanization

A. RAMACHANDRA RAO

School of Civil Engg., Purdue University, West Lafayette, Indiana 47907

and

SRINIVAS G. RAO

Seaburn and Robertson, Inc., P.O. Box 23104, Tampa, Florida 33623

(*Received 27 February 1987*)

सार — शहरीकरण तथा औद्योगिकीकरण, स्थानीय जलवायु तथा वर्षा एवं उससे सम्बन्धित त्रिया में साधारणतया वृद्धि वृष्टिपात के लक्षणों में परिवर्तन लाती है। इस अध्ययन में, शहरी क्षेत्रों में वृष्टिपात के माध्यमूल्यों के परिवर्तन के महत्व का पता लगाने के लिए बहुचर सांख्यिकी परीक्षण का प्रयोग किया गया है। ये परीक्षण, क्षेत्र के बहुत से स्टेशनों के आंकड़ों को विश्लेषित करने में तथा साथ ही शहरी क्षेत्रों में वृष्टिपात के परिवर्तन का अनुमान लगाने में उपयोगी है। बहुचर आंकड़ों की आकाशीय सहसम्बन्ध की संरचना किसी एक स्टेशन के आंकड़ों के विश्लेषण की अपेक्षा वृष्टिपात पर शहरीकरण के प्रभाव के परिमाण को अधिक अच्छे ढंग से बताने में समर्थ है। वर्तमान अध्ययन में T^2 परीक्षण तथा प्रसरण के विश्लेषण पर आधारित परीक्षण का उपयोग वृष्टिपात के माध्य मूल्यों के परिवर्तन के महत्व का परीक्षण करने के लिए किया गया है। संयुक्त राज्य अमेरिका में कनसास नगर के क्षेत्रों, लापोर्ट, सेंट लुइस तथा तुलसा में अनेक स्थानों से वार्षिक वृष्टिपात के आंकड़ों का विश्लेषण किया गया है तथा उनके परिणाम प्रस्तुत किए गए हैं।

ABSTRACT. Urbanization and industrialization change precipitation characteristics by modifying the local climate and generally increasing precipitation and precipitation related activity. Multivariate statistical tests are used in this study to detect the significance of changes in mean values of precipitation in urban areas. These tests are useful to simultaneously analyze the data from a number of stations in a region and to arrive at inferences about changes in precipitation in urban areas. The spatial correlation structure of multivariate data enables better quantification of the effects of urbanization on precipitation than analysis of data from single stations. The T^2-test and a test based on the analysis of variance are used in the present study to test the significance of changes in precipitation mean values. Annual precipitation data from several stations in the LaPorte, St. Louis, Tulsa and Kansas city areas in the United States are analyzed and the results are presented.

1. Introduction

Precipitation and related activity in urban areas have been observed to be different and usually higher than those in the surrounding rural areas. Urbanization and industrialization have been found to modify the local climate and cause these changes in precipitation characteristics. A comparison of rural and urban climates indicate an increase in contaminants (10 times), cloudiness (5 to 10%), precipitation (5 to 10%), and temperature (0.5 to 1° C) and a decrease in relative humidity (2 %), radiation (15 to 20%) and windspeed (20 to 30%) in urban areas or in areas downwind of urban areas (Landsberg 1970 a, b).

Previous investigations of the effect of urbanization on precipitation have indicated that the precipitation amounts and related activity in the LaPorte (Indiana), the St. Louis (Missouri and Illinois), and Tulsa areas (Oklahoma) in the United States have increased over the years. Changnon and his associates (Changnon 1968, Changnon 1969, Changnon 1973, Huff & Changnon 1970, Changnon *et al.* 1977, Ackerman *et al.* 1978, Changnon *et al.* 1979, Changnon 1980 and Changnon 1981) have extensively studied the effects of urbanization on precipitation in the United States. Several statistical methods have been used in the past for analyzing the effects of urbanization on precipitation. These methods include run analysis and intervention analysis.

In this paper, multivariate statistical tests are used to study the covariance structure of multivariate precipitation data set in a region and to detect changes in precipitation mean value observed in urban regions. As multivariate tests make use of the spatial correlation structure of data from several stations in a region, these tests are more reliable than tests involving single stations. Annual precipitation data from several stations in LaPorte, St. Lous, Kansas city and Tulsa area are used.

The basic assumption made in most of these tests is that a multivariate data set arises from a multivariate normal population. Since annual data are approximately normally distributed, multivariate tests are applicable to annual data. Furthermore, these multivariate tests

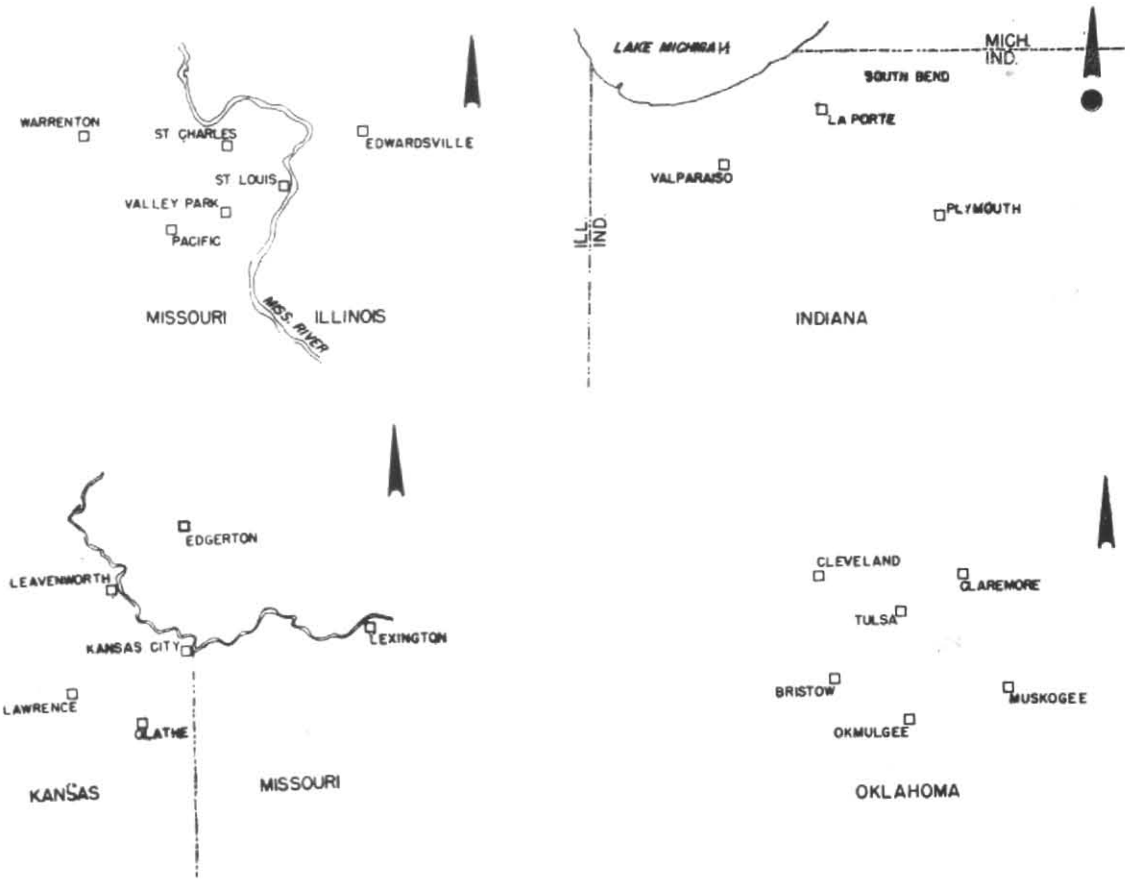

Fig. 1. Location of stations

TABLE 1

Elementary statistics of data used in the study

Station	Beginning year	Ending year	Break point	Observations			Unaffected		Affected		Entire		% change in mean
				Unaffected	Affected	Entire	Mean	S.D.	Mean	S. D.	Mean	S. D.	
LAP	1915	1968	1929	14	40	54	36.90	5.800	44.59	8.085	42.60	8.17	20.84
SB	1915	1968	1929	14	40	54	32.89	5.008	35.88	5.621	35.11	5.53	9.09
PLY	1915	1968	1929	14	40	54	34.82	6.284	36.12	4.897	35.78	5.95	3.73
VALP	1915	1968	1929	14	40	54	37.04	6.045	37.67	5.205	37.50	5.89	1.70
STLS	1916	1970	1941	25	30	55	24.66	7.532	36.63	8.670	35.74	8.08	5.71
EDW	1916	1970	1941	25	30	55	37.85	7.128	40.46	7.827	39.27	7.49	6.90
VALL	1916	1970	1941	25	30	55	36.57	7.212	38.20	8.093	37.46	7.61	4.46
WAR	1916	1970	1941	25	30	55	37.93	7.557	34.81	6.712	37.75	6.98	—8.23
STCH	1916	1970	1941	25	30	55	37.60	6.570	36.64	6.794	35.81	6.63	—2.55
KAN	1906	1970	1931	25	40	65	37.02	5.030	35.33	8.195	36.00	7.18	4.56
LEX	1906	1970	1931	25	40	65	38.02	8.127	37.67	9.922	37.80	9.02	—0.92
OLA	1906	1970	1931	25	40	65	35.64	6.664	36.83	9.358	36.38	8.32	3.34
LAW	1906	1970	1931	25	40	65	36.07	7.210	36.00	9.072	36.03	8.17	—0.19
LEA	1906	1970	1931	25	40	65	34.47	5.160	36.06	8.898	35.45	7.54	4.61
TULS	1916	1970	1931	15	40	55	40.21	8.538	37.61	8.557	38.34	8.28	—6.47
CLA	1916	1970	1931	15	40	55	40.03	7.729	37.11	8.769	37.91	8.45	—7.29
MUSK	1916	1970	1931	15	40	55	41.74	7.434	41.50	10.06	41.57	9.27	—0.57
CLE	1916	1970	1931	15	40	55	36.87	8.103	35.12	9.424	35.60	8.96	—4.75

LAP: LaPorte; SB: South Bend; PLY : Plymouth ; VALP : Valparaiso ; STLS: St. Louis ; EDW: Edwardsville ; VALL: Valley Park ; WAR: Warrenton; STCH: St. Charles; KAN: Kansas City; EDG: Edgerton; LEX: Lexington; OLA: Olathe; LAW: Lawrence, LEA : Leavenworth; TULS: Tulsa; CLA: Claremore; MUSK: Muskogee; CLE: CLeveland.

% change = [Affected period mean—unaffected period mean]/unaffected period mean.

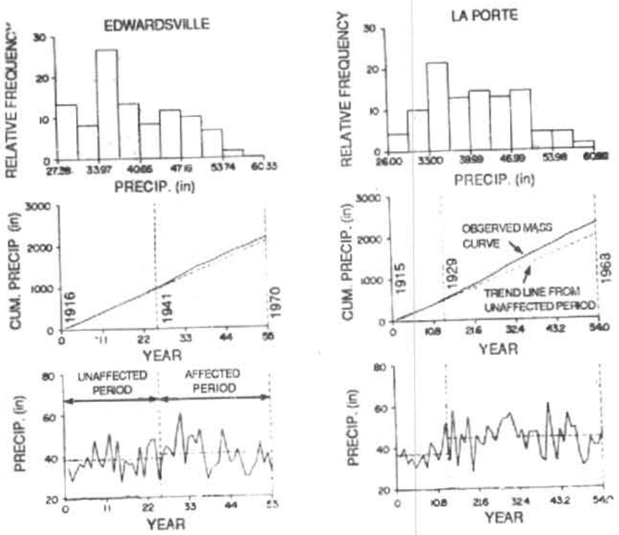

Fig. 2. Typical histograms, mass curves and data traces

are relatively insensitive to departure from normality provided all the data are similarly affected. It is necessary to consider this question since changes in precipitation can be due to non-urban effects such as the natural or random variability of the data itself. Statistical testing helps to objectively determine the significance of changes brought about by urbanization.

It is important to test the covariance structure of multivariate data to ensure that the covariance matrix is nondiagonal. If the covariance matrix is diagonal, then the multivariate tests are not applicable. The equality of covariance matrices in the affected and unaffected periods must also be tested as this property affects the results of multivariate tests. Therefore, these two aspects are investigated before applying multivariate tests to analyze the effect of urbanization on annual rainfall characteristics.

It is emphasized that analysis of the cause and effect relationships among the many complex and interactive climatic variable is outside the scope of this study. The present study is concerned with determining the statistical significance of the observed change in mean precipitation at a station or group of stations.

2. Preliminary statistical analysis of data

Observed annual precipitation data from several stations in the LaPorte, St. Lous, Tulsa, and Kansas city areas are used in this study. These stations (Fig. 1) are listed in Table 1. The precipitation data in these stations are judged to be good and consistent based on the U.S. Weather Bureau data records. To quantify the changes in precipitation characteristics attributed to the effect of urbanization, the approximate time at which the precipitation characteristics have changed should be determined first. Mass curves, double mass curves, and comparison of averages of split samples of data are used in the present study to estimate the time at which rainfall characteristics have started to change. Examples of mass curves are shown in Fig. 2 along with approximate break points in them. These break points are also compared with the results obtained by moving average analysis. Dotted lines in Fig. 2 indicate continuation of trend lines of mass curves before the break. Double mass curves are also computed for various combinations of stations and trends are examined. The precipitation data acquired before the year in which changes in precipitation characteristics are evident are called the *unaffected data* and those from the period after the changes are apparent, the *affected data.*

TABLE 2

Chi-square test for testing the normality of annual data

Station	Computed $\chi^2_{0\cdot05}$	Critical $\chi^2_{0\cdot05}$	Degrees of freedom	Decision*
STLS	6.299	12.59	6	A
STCH	2.589	12.59	6	A
VALL	2.033	11.07	5	A
EDW	10.416	12.59	6	A
TULS	8.190	12.59	6	A
CLA	10.231	12.59	6	A
CLE	12.491	14.07	7	A
MUSK	8.252	12.59	6	A
KAN	2.611	12.59	6	A
LEA	1.440	12.59	6	A
LEX	1.420	12.59	6	A
LAW	1.839	12.59	6	A
OLA	4.537	12.59	6	A
LAP	0.835	12.59	6	A
PLY	0.926	12.59	6	A
SB	3.296	12.59	6	A
VALP	5.874	12.59	6	A
WAR	4.537	14.07	7	A

*A=Accept the hypothesis that the data are normally distributed

Some of the elementary statistics of observed annual precipitation are given in Table 1 for the affected, unaffected and entire periods. Also shown in Table 1 are the percentage changes in mean precipitation between the affected and unaffected periods. The percentage change is within $\pm$ 10% for all stations except LaPorte, for which it is about + 21%, thereby indicating a substantial increase in the annual precipitation mean at LaPorte. In some cases the change in precipitation mean values are well within a standard deviation of the unaffected period data, which indicates that the observed change for these cases may be statistically insignificant.

Examples of histograms of annual data are shown in Fig. 2. As the number of annual data values is small, histograms of annual data represent rough approximations to probability densities of annual precipitation. The frequency distributions of precipitation, even in stations which are close to each other, show very different characteristics. The χ^2 goodness of fit test results given in Table 2 indicate that the annual data may be considered to be approximately normally distributed. Autocorrelation and power spectra of annual precipitation data, an example of which is shown in Fig. 3, indicate that the annual data can be considered to be essentially uncorrelated.

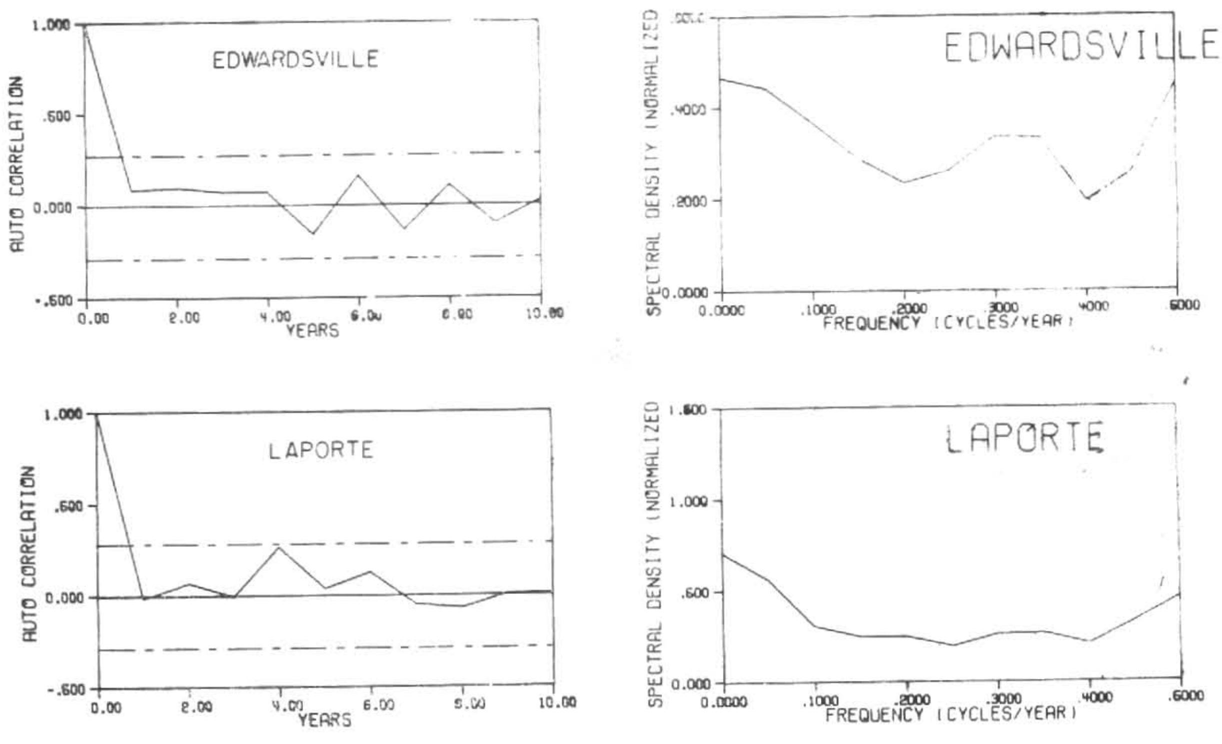

Fig. 3. Typical correlograms and power spectra of annual precipitation data

TABLE 3

Lag zero cross correlation coefficients of annual data

	Unaffected					Affected					Entire				
	STLS	VALL	WAR	STCH	EDW	STLS	VALL	WAR	STCH	EDW	STLS	VALL	WAR	STCH	EDW
STLS	1.000	0.899	0.753	0.872	0.543	1.000	0.803	0.568	0.913	0.671	1.000	0.843	0.638	0.897	0.658
VALL		1.000	0.858	0.869	0.839		1.000	0.628	0.801	0.625		1.000	0.720	0.831	0.715
WAR			1.000	0.795	0.788			1.000	0.717	0.564			1.000	0.741	0.649
STCH				1.000	0.823				1.000	0.848				1.000	0.840
EDW					1.000					1.000					1.000
		LAP	PLY	SB	VALP		LAP	PLY	SB	VALP		LAP	PLY	SB	VALP
LAP		1.000	0.806	0.893	0.649		1.000	0.636	0.603	0.711		1.000	0.685	0.614	0.664
PLY			1.000	0.814	0.646			1.000	0.732	0.692			1.000	0.739	0.678
SB				1.000	0.649				1.000	0.726				1.000	0.703
VALP					1.000					1.000					1.000
	KAN	LAW	LEX	OLA	LEA	KAN	LAW	LEX	OLA	LEA	KAN	LAW	LEX	OLA	LEA
KAN	1.000	0.673	0.776	0.736	0.675	1.000	0.848	0.792	0.899	0.715	1.000	0.797	0.779	0.846	0.782
LAW		1.000	0.682	0.789	0.660		1.000	0.740	0.932	0.876		1.000	0.723	0.891	0.816
LEX			1.000	0.577	0.570			1.000	0.801	0.725			1.000	0.737	0.677
OLA				1.000	0.574				1.000	0.869				1.000	0.803
LEA					1.000					1.000					1.000
		TULS	CLA	CLE	MUSK		TULS	CLA	CLE	MUSK		TULS	CLA	CLE	MUSK
TULS		1.000	0.907	0.714	0.800		1.000	0.674	0.660	0.486		1.000	0.735	0.675	0.543
CLA			1.000	0.637	0.676			1.000	0.820	0.679			1.000	0.782	0.671
CLE				1.000	0.598				1.000	0.675				1.000	0.658
MUSK					1.000					1.000					1.000

TABLE 4

Results from test for independence

Station	Entire period g_1	Entire period Decision	Unaffected period g_1	Unaffected period Decision	Affected period g_1	Affected period Decision
St. Louis	202.54	R	241.38	R	196.54	R
Edwardsville	170.20	R	215.87	R	159.27	R
LaPorte	120.79	R	173.688	R	123.62	R
Kansas City	240.33	R	180.51	R	295.72	R
Leavenworth	224.48	R	138.49	R	285.63	R
Tulsa	127.82	R	181.17	R	129.14	R

R=Reject the null hypothesis that the data set is uncorrelated.

The preliminary data anlysis thus indicates that the precipitation has increased in the affected period in comparison with the unaffected period in several stations located in and around urban areas. The significance of these changes are analyzed by using several multivariate tests in section 4, following an investigation of covariance properties of data in section 3.

3. Tests on covariance properties of data

The lag-zero correlation matrix of the observed multivariate data are computed by using Eqn. (1) :

$$\hat{\rho} = \frac{\sum_{i=1}^{N} [P_1(i) - \overline{P_1}]\,[P_2(i) - \overline{P_2}]}{\left[\sum_{i=1}^{N} \{P_1(i) - \overline{P_1}\}^2 \sum_{i=1}^{N} \{P_2(i) - \overline{P_2}\}^2\right]^{\frac{1}{2}}} \qquad (1)$$

In Eqn. (1), $P_1(i)$ and $P_2(i)$ are respectively the annual precipitation at stations 1 and 2. $\overline{P_1}$ and $\overline{P_2}$ are the mean values of $P_1(i)$ and $P_2(i)$. $\hat{\rho}$ is the estimate of the correlation coefficient.

The correlation coefficients $\hat{\rho}$ between different stations for the affected, unaffected and entire periods are given in Table 4. The correlation coefficients corresponding to the affected and unaffected periods are substantially different from each other. For example, the unaffected period correlation coefficients for data from St. Louis and Valley Park is 0.899; it is 0.893 for LaPorte and South Bend and it is 0.673 for Kansas city and Lawrence. The corresponding affected period correlation coefficients are respectively 0.803, 0.603 and 0.848. Although some of this variability may be explained by the different number of observations in the affected and unaffected periods the differences between the correlation coefficients in the two periods are substantial, which indicates some change in the precipitation characteristics in the affected period. The maximum correlation coefficient between any two stations in Table 4 is underscored.

(a) *Likelihood test for independence*

In the analysis of covariance structures of k-dimensional multivariate populations such as data from k stations, if the $k(k-1)/2$ population correlations are all equal to zero, then the results from analyses based upon sample covariance matrices would be unacceptable consequently, whether the population is covariance.

(b) *Test for equality of covariance matrices*

In this test the covariance matrices of data from the affected and unaffected period Σ_a and Σ_u are tested for equality. The null and alternate hypotheses used in the test are given below and the two populations in the affected and unaffected periods are assumed to be normally distributed with means $\hat{\mu}_i$ and covariances $\hat{\Sigma}_i$.

$$H_0 : \underset{\sim}{\Sigma_u} = \underset{\sim}{\Sigma_a}$$

$$H_a : \underset{\sim}{\Sigma_u} \neq \underset{\sim}{\Sigma_a}$$

Let $\hat{\Sigma}_u$ and $\hat{\Sigma}_a$ be the estimates of $\underset{\sim}{\Sigma_u}$ and $\underset{\sim}{\Sigma_a}$ with n_u and n_a degrees of freedom. If $\hat{\Sigma}_u$ and $\hat{\Sigma}_a$ are respectively computed by using independent samples of sizes m and n then n_u=m-1 and n_a=n-1. Let $\underset{\sim}{\hat{\Sigma}}$ be the pooled estimate of $\underset{\sim}{\Sigma}$ under H_0.

$$\underset{\sim}{\hat{\Sigma}} = \frac{n_u \underset{\sim}{\hat{\Sigma}_u} + n_a \underset{\sim}{\hat{\Sigma}_a}}{n_u + n_a}$$

The test statistic corresponding to the modified likelihood ratio test is designated M and is given in Eqn. (2):

$$M = (n_u + n_a) \ln |\underset{\sim}{\hat{\Sigma}}| - n_u \ln |\underset{\sim}{\hat{\Sigma}_u}| - n_a \ln |\underset{\sim}{\hat{\Sigma}_a}| \qquad (2)$$

The statistic g_2 $(=MC^{-1})$ has been shown by Box (1949) to be approximately χ^2 distributed with n_d degrees of freedom (n_d=10 in the present case) as n_u and n_a tend to infinity, where C^{-1} is defined in Eqn. (3) :

$$C^{-1} = 1 - \frac{(2k^2 + 3k - 1)}{6(k+1)(k-1)} \left[\sum_{i=1}^{2} \frac{1}{n_i} - \frac{1}{\Sigma n_i}\right] \qquad (3)$$

$$n_1 = n_u,\ n_2 = n_a$$

The decision rule for the test is given below:

$$\text{If } g_2 \begin{cases} \leqslant x_a^2(10) \rightarrow \text{Accept } H_0 \\ > x_a^2(10) \rightarrow \text{Reject } H_0 \end{cases} \qquad (4)$$

matrix $\underset{\sim}{\Gamma}$ is a diagonal matrix is tested as discussed below. The null (H_0) and alternate (H_a) hypotheses used in the test are :

$$H_0 : \underset{\sim}{\Gamma} = \underset{\sim}{I}$$

$$H_a : \underset{\sim}{\Gamma} \neq \underset{\sim}{I}$$

where $\underset{\sim}{\Gamma}$ and $\underset{\sim}{I}$ are respectively the population correlation and identity matrices. If the alternate hypothesis is accepted, then at least one cross correlation coefficient or covariance is significantly different from zero.

The generalized likelihood test is used to test the null hypothesis. Let the number of observations available from each station be N, which consitute the N vectors, $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_N$. Let these observations be drawn independently from k dimensional multinormal population with parameters mean μ and covariance matrix $\underset{\sim}{\Sigma}$,

TABLE 5

Results from test for equality of covariance matrices

Station	g_2	Decision
St. Louis	14.56	A
Eswardsville	14.78	A
LaPorte	15.10	A
Kansas City	15.86	A
Leavenworth	15.65	A
Tulsa	17.42	A

A=Accepted the null hypothesis that covariance matrices in the unaffected and affected periods are equal.

TABLE 6

Results from Hotelling's T^2-test

Station	g_3	Decision
St. Louis	0.7291	A
Edwardsville	0.8980	A
LaPorte	3.3090	R
Kansas City	1.9580	A
Leavenworth	0.8310	A
Tulsa	0.5480	A

TABLE 7

Confidence limits for $(\mu_u - \mu_a)$

Station	Upper limit	Lower limit	Decision
LaPorte	5.030	10.368	R
South Bend	5.181	4.803	R
Plymouth	—0.442	3.058	A
Valparaiso	—0.298	2.562	A

A(R)=Accept (reject) the null hypothesis that there is no change in mean

whose estimates are indicated by $\hat{\mu}$ and $\hat{\underset{\sim}{\Sigma}}$. Under these conditions, the likelihood ratio statistic $\hat{\lambda}$ is given by Eqn. (5) where $|\hat{\underset{\sim}{\Gamma}}|$ represents the determinant of the correlation matrix :

$$\hat{\lambda} = \frac{|\hat{\underset{\sim}{\Sigma}}|^{N2}}{(\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2)^{N/2}} = |\hat{\underset{\sim}{\Gamma}}|^{N/2} \quad (5)$$

$\hat{\sigma}_1, \hat{\sigma}_2, \ldots \hat{\sigma}_k$ are the estimates of standard deviations of k sets data. The statistic g_1 in Eqn. (6) has been shown by Bartlett (1954) to be χ^2 distributed with $k(k\text{-}1)/2$ degrees of freedom :

$$g_1 = -(N - 1 - \frac{2k+5}{6}) \ln |\hat{\underset{\sim}{\Gamma}}| \quad \chi^2[k(k-1)/2] \quad (6)$$

The decision rule for this test is given below:

$$\text{If } g_1 \begin{cases} < \chi^2\{k(k-1)\}/2 \rightarrow \text{Accept } H_0 \text{ at significance level } \alpha \\ \geqslant \chi^2\{k(k-1)/2\} \rightarrow \text{Reject } H_0. \end{cases}$$

The statistic g_1, computed by using the entire unaffected and affected period data sets, are shown in Table 3. The data sets consisted of stations are indicated in Table 3. These results conducted at 5% level indicate that the cross correlations in annual rainfall between stations are significantly different from zero. Consequently, the covariance matrices can be further used for tests.

The statistic g_2 is shown in Table 5. The statistic g_2 is well below the critical value in all cases except for stations around Tulsa for which it is closer to the critical value. Consequently, we may consider the covariance matrices in the unaffected and affected periods to be equal.

4. Testing changes in precipitation characteristics

(a) *Multivariate T^2 test for equality of means*

Consider the k-dimensional annual precipitation series $P(k \times N)$ which are assumed to be multivariate normal with mean vector $\hat{\underset{\sim}{\mu}}$ and covariance matrix $\hat{\underset{\sim}{\Sigma}}$. The null and alternate hypotheses to test the changes in these random variables are given ahead.

$$H_0 : \underset{\sim}{\mu_u} = \underset{\sim}{\mu_a}$$

$$H_a : \underset{\sim}{\mu_u} \neq \underset{\sim}{\mu_a}$$

where $\underset{\sim}{\mu_u}$ and $\underset{\sim}{\mu_a}$ represent the unaffected and effected mean vectors $(\mathbf{k} \times \mathbf{l})$. Let Σ be the common but unknown covariance matrix of full rank k. The estimates $\hat{\mu}_u$, $\hat{\mu}_a$ and $\hat{\underset{\sim}{\Sigma}}$, the pooled covariance estimate of the common covariance matrix are computed by using the observed data. The multivariate T^2 test is used to test H_0. The test statistic, Hotelling's-T^2, is computed by Eqn. (7).The decision rule for testing the null hypothesis is given in Eqn. (8) (Anderson 1954):

$$T^2 = \frac{nm}{n+m} (\underset{\sim}{\mu_u} - \underset{\sim}{\mu_a})^T \; \Sigma(\underset{\sim}{\mu_u} - \underset{\sim}{\mu_a}) \quad (7)$$

$$\text{If } g_3 \begin{cases} \leqslant F_\alpha\,(k, N-k-1) \rightarrow \text{Accept } H_0 \\ > F_\alpha\,(k, N-k-1) \rightarrow \text{Reject } H_0 \end{cases} \quad (8)$$

where $g_3 = (N-k+1)\;T^2/(N\text{-}2k)$. The F distribution values used in the decision rule are readily available.

The results of the T^2-test are given in Table 6. For the data from all groups of stations except the group from LaPorte, the null hypothesis that the mean vectors from the unaffected and affected period are equal can be accepted. For the data from LaPorte and neighbouring stations, the null hypothesis must be rejected.

In order to find out the station(s) which leads to the rejection of the null hypothesis, the upper and lower confidence limits are constructed for the elements of the mean vector. If $\mathbf{a}$ is non-null vector (which is the assumed to be a unit vector in the present study) the upper and lower confidence limits for the vector difference $(\hat{\mu}_u - \hat{\mu}_a)$ are given by Eqns. (9) and (10) respectively.

$$\underset{\sim}{\hat{\mu}} = \underset{\sim}{\mu_u} - \underset{\sim}{\mu_a}\{\mathbf{a}^T \underset{\sim}{\mu} \pm \sqrt{\frac{1}{N_E}}\, \mathbf{a}^T \underset{\sim}{\Sigma}\, \mathbf{a} T_\alpha(k, N-k-1)\} \quad (9)$$

$$N_E = (n+m)\,/\,nm \quad (10)$$

If the confidence limits include zero, then the null hypothesis is accepted for that set; otherwise it is rejected.

TABLE 8

Results of test based on the analysis of variance by using annual data from a single station and all the stations in a region

Station	Data from a single station		Data from a group of stations in the same region					
			Entire data		Unaffected		Affected period data	
	Test statistic F_0 (df)	DEC	Test statistic F_0 (df)	DEC	Test statistic F_0 (df)	DEC	Test statistic F_0 (df)	DEC
STLS	2.7345 (1.530)	A	1.1293 (3.216)	A	1.1553 (3.96)	A	0.3052 (3.116)	A
EDW	1.6430 (1.53)	A	2.1036 (3.216)	A	1.0482 (3.86)	A	1.4474 (3.156)	A
LAP	37.7421 (1.53)	R	15.4100 (3.212)	R	1.68 (3.52)	A	16.8940 (3.156)	R
TULS	2.5350 (1.52)	A	4.2110 (3.216)	R	0.9936 (3.56)	A	3.5228 (3.156)	R
KAN	2.8351 (1.63)	A	0.6845 (3.256)	A	0.61 (3.96)	A	0.3133 (3.156)	A
LEAV	0.5630 (1.63)	A	0.9192 (3.256)	A	1.1495 (3.96)	A	0.2835 (3.156)	A

df = Degrees of freedom DEC = Decision

A(R) = Accept (reject) the null hypothesis that there is no change in mean

The upper and lower confidence limits for the stations in the LaPorte group are given in Table 7. The confidence limits for the difference in means in unaffected and affected periods do not span zero for LaPorte and South Bend data, thereby indicating that the data from LaPorte and South Bend have caused the rejection of the null hypothesis in the LaPorte group of stations. The data from Plymouth and Valparaiso have not contributed to the rejection of the null hypothesis. These results indicate that the observed mean precipitation at LaPorte and South Bend have increased.

(b) *The test based on the analysis of variance*

This test is based on the one-way classification of the precipitation data. In one-way classification, the data from a station is considered as a group and several such groups of data in a region are used to detect the change in mean. In order to use this test, there should be at least two groups of observations. Data from unaffected and affected periods may be used as two groups and hence a change in mean in the precipitation at a single station can be detected. Or serveral data sets from a region may be used to test whether there is any nonhomogeneity in the data (or a change in mean has occurred) from a region.

The null hypothesis (H_0) in the test is stated as "the groups of data are homogeneous," *i.e.*, they come from populations which are identical as far as the mean is concerned. The alternate hypthesis (H_a) is that "the groups of data are not homogeneous," *i.e.*, there is a change in the mean.

The asumptions made in this test are that (1) the observations of each group are randomly chosen from a parent population corresponding to that group, (2) within each group the variation is normal with a common variance. The population corresponding to different groups can differ, if at all, only in their mean values. All the observations are assumed to be normally distributed with the same variance. The test is relatively insensitive to non-normality if all the observations are similarly affected.

Let the precipitation from the ith station during jth year be denoted by P_{ij}. Let the number of groups of stations be k and let N_i be the total number of observations availabe at the ith station. Also let μ_T be the mean of the observations from all stations, μ_i be the mean of the observations at the ith station. Let N_T be the total number of observations available from all the stations.

TABLE 9

Test based on the analysis of variance (pairs of stations)

Station 1 (affected station)	Station 2 (unaffected station)	Entire period F_0 (df)	DEC	Unaffected F_0 (df)	DEC	Affected period data F_0 (df)	DEC
STLS	Vally	1.289 (1, 108)	A	.830 (1, 48)	A	.519 (1, 58)	A
	War	1.905 (1, 108)	A	2.342 (1, 48)	A	.228 (1, 58)	A
	St Ch	2.000 (1, 108)	A	.0053 (1, 48)	A	0.000 (1, 58)	A
EDW	Vally	1.554 (1, 108)	A	.399 (1, 48)	A	1.202 (1, 58)	A
	War	1.190 (1, 108)	A	.0016 (1, 48)	A	2.304 (1, 58)	A
	St. Ch	6.463 (1, 108)	R	2.426 (1, 48)	A	4.071 (1, 58)	R
LAP	SB	30.591 (1, 108)	R	3.862 (1, 26)	A	31.335 (1, 78)	R
	Ply	25.796 (1, 108)	R	.864 (1, 26)	A	31.035 (1, 78)	R
	Valp	13.567 (1, 108)	R	.0032 (1, 26)	A	19.136 (1, 78)	R
TULS	Cla	0.072 (1, 108)	A	0.007 (1, 28)	A	.067 (1, 78)	A
	Cle	2.665 (1, 108)	A	1.262 (1, 28)	A	1.528 (1, 78)	A
	Musk	3.561 (1, 108)	R	.246 (1, 28)	A	3.422 (1, 78)	R
KAN	Law	.004 (1, 128)	A	0.321 (1, 48)	A	.119 (1, 78)	A
	Lex	1.5572 (1, 128)	A	0.245 (1, 48)	A	1.3242 (1, 78)	A
	Ola	0.762 (1, 128)	A	0.732 (1, 48)	A	.586 (1, 78)	A
LEA	Law	2.503 (1, 128)	A	3.387 (1, 48)	A	.582 (1, 78)	A
	Lex	.167 (1, 128)	A	.815 (1, 48)	A	.111 (1, 78)	A
	Ola	.430 (1, 128)	A	.481 (1, 48)	A	.143 (1, 78)	A

DEC = Decision F_0 = Test statistic df = Degrees of freedom

A(R)=Accept (reject) the null hypothesis that there is no change in mean

$$\mu_T = \frac{\sum_{i=1}^{k}\sum_{j=1}^{N_i} P_{ij}}{N_T}\ ;\ \mu_i = \frac{\sum_{j=1}^{N_i} P_{ij}}{N_i}\ ;\ N_T = \sum_{i=1}^{k} N_i \quad (11)$$

The total variation in all the observations can be written as Eqn. (12):

$$\sum_{i=1}^{k}\sum_{j=1}^{N_i}(P_{ij}-\mu_T)^2 = \sum_{i=1}^{k}\sum_{j=1}^{N_i}\{(P_{ij}-\mu_i)+(\mu_i-\mu_T)\}^2$$

$$= \sum_{i=1}^{k}\sum_{j=1}^{N_i}(P_{ij}-\mu_i)^2 + \sum_{i=1}^{k}\sum_{j=1}^{N_i}(\mu_i-\mu_T)^2 \quad (12)$$

Eqn. (12) can be written as (Eqn. 13):

$$\sum_{i=1}^{k}\sum_{j=1}^{N_i}(P_{ij}-\mu_T)^2 = \sum_{i=1}^{k}\sum_{j=1}^{N_i}(P_{ij}-\mu_i)^2 + \sum_{i=1}^{k} N_i(\mu_i-\mu_T)^2 \quad (13)$$

Eqn. (13) is the usual decomposition of the total variance in the data. Variance of individual observations about the mean of the whole is equal to the variance of observations about the respective group means *plus* variance of the group means about the mean of the whole. The total variance is equal to the sum of the variances between groups and the variance within the groups. These three variances in Eqn. (13) are distributed as χ^2 with $(N\text{-}1)$, $(N\text{-}k)$, and $(k\text{-}1)$ degrees of freedom, respectively. Then the statistic F_0 given in Eqn. (14) is F-distributed with $(k\text{-}1)$ and $(N\text{-}k)$ degrees of freedom.

$$F_0 = \frac{\hat{S}_B^2}{\hat{S}_N^2} \quad (14)$$

where,

$$\hat{S}_B^2 = \frac{1}{k-1}\sum_{i=1}^{k} N_i(\mu_i-\mu_T)^2$$

and

$$\hat{S}_N^2 = \frac{1}{(N-k)}\sum_{i=1}^{k}\sum_{j=1}^{N_i}(P_{ij}-\mu_i)^2.$$

The decision rule for the test is as follows:

$$\text{If } F_0 \begin{cases} \leqslant F_\alpha(k-1,\ N-k) \rightarrow \text{Accept } H_0 \\ > F_\alpha(k-1,\ N-k) \rightarrow \text{Reject } H_0 \end{cases} \quad (15)$$

The test is carried out by using the unaffected and affected data and the results are given in Table 8. In Table 8, the decision is given as accept (A) or reject (R) the null hypothesis. The null hypothesis is rejected only for LaPorte, implying that there is a change in mean only at LaPorte.

The data from several stations in a region (the groupings are shown in Table 9) are used and for each of the cases of entire, unaffected and affected periods the test statistics are given in Table 8. With the entire data, the null hypothesis is rejected only for LaPorte and Tulsa data. With only the unaffected period data, the null hypothesis is accepted for all data sets as it should be. With only the affected period data, the null hypothesis is once again rejected for LaPorte and Tulsa data.

In order to find out the data from the unaffected stations which are deviating the most from the data from the affected station in mean value in each group, data from affected station and individual unaffected stations are analyzed pairwise and the results are given in Table 9. Once again entire, unaffected and affected data are considered separately. The entire data for St. Louis, Tulsa, Kansas city and Leavenworth do not show any significant change in mean value with any unaffected station whereas LaPorte and Tulsa data show changes in mean with all unaffected stations considered separately. Also, for only the unaffected period, data are homogeneous as expected for all the cases. When only the affected period data is used, a significant change in mean value is present with Edwardsville (with Edwardsville-St. Charles pair of data), Tulsa (with Tulsa-Muskogee pair), and LaPorte (with all pairs of data) data.

5. Summary and conclusions

In this study, multivariate statistical tests are employed to test the significance of observed changes in annual precipitation brought about by urbanization. Annual precipitation data from several stations in the LaPorte, St. Louis, Tulsa, and Kansas city areas are used to quantify the effects of urbanization on precipitation in and around urban areas. Mass curves and moving averages of precipitation are used to establish the time at which urbanization has started affecting the precipitation characteristics. The observed precipitation data at each station is divided into two periods corresponding to the unaffected and affected periods. Data from several stations in each urban region are pooled to form multivariate data sets. Equality of covariance matrices are tested. The statistical significance of the observed changes in the mean value of precipitation is tested by using multivariate T^2-test and a test based on the analysis of variance. Based on the results from the study, the following conclusions are presented :

(*i*) The multivariate T^2-test and analysis of variance test indicate that the increase in mean annual precipitation in the precipitation mean value at LaPorte and at South Bend is significant.

(*ii*) The test based on the analysis of variance indicates that the observed increase in the mean annual precipitation is significant at Edwardsville and Tulsa. The observed increase in mean annual precipitation at Edwardsville is significant.

(*iii*) According to these tests the observed changes at St. Louis and Kansas, Leavenworth are not significant.

(*iv*) Based on the observed changes in mean precipitation, the test based on the analysis of variance seems to be more sensitive to changes than the T^2-test.

References

Ackerman, B.S., Changnon, S.A., Dzurisin, Jr., G., Gatz, D.F., Grosh, R.C., Hilberg, S.D., Huff, F.A., Mansell, J.W. Ochs, H.T., Peden, M.E., Schickedanz, P.T., Semonin. R.G. and Vogel, J.L., 1978, Summary of METROMEX **2**, Causes of Precipitation Anomalies, "Illinois State Water Survey Bull. 63. Urbana, 395 pp.

Anderson, T.W., 1954, "*An Introduction of Multivariate Statistical Analysis,* John Wiley and Sons, New York.

Bartlett, M.S., 1954, ,'A Note on the Multiplying Factors for Various Chi-square Approximations," *Royal Statistical Society,* Series B, **16**, pp. 296-298.

Box, G.E.P., 1949, "A General Distribution Theory for a Class of Likelihood Criteria," *Biometrika*, **36**, pp. 317-346.

Changnon, S.A., Jr., 1968, "The LaPorte Weather Anomaly Fact or Fiction," *Bull. Am. met. Soc.*, **49**, 1. pp. 4-11.

Changnon, S.A., Jr., 1969, Recent Studies of Urban Effect on Precipitation in the United States," *Bull. Am. met. Soc.*, **50**, 6, pp. 411-421.

Changnon, S.A., Jr., 1973, "Inadvertent Weather and Precipitation," *Irrigation and Drainage Division*, Proc. of the ASCE, **99**, No. IRI, pp. 27-41.

Changnon, S.A., Jr., 1980, "More on the LaPorte Anomaly: A Review," *Bull. Am. met. Soc.*, **61**, 702-711.

Changnon, S.A., Jr., Ed., 1981, "Metromex: A Review and Summary", Am. Met. Soc., **18**, Monograph No. 40, 181 pp.

Changnon, S.A., Jr., Huff, F.A., Schickedanz, P.T. and Vogel, J.L., "Summary of METROMEX, Vol. 1, Anomalies and, Impacts," Illinois State Water Survey, Bull. 62, Urbana, 260 pp.

Changnon, S.A., Jr., Jameson, A.R., Dzurisin, G.L., Scott, R.W. and Grosh, R.C., 1979, "Studies of Urban and Lake Influences on Precipitation in the Chicago Area, Illinois State Water Survey, Urbana, 190 pp.

Huff, F.A., and Changnon, S.A. Jr., 1970, "Urban Effects of Daily Rainfall Distribution", preprints of papers presented at the Am. Met. Soc., II National Conf. on Weather Modification, Santa Barbara, California, pp. 215-220.

Landsberg, H.E., 1970(a), "Climates and Urban Planning", in *Urban Climates*, WMO (1970), pp. 364-374.

Landsberg, H.E., 1970(b), "Man-made Climatic Changes, Science, Assn for Adv. of Sci., Washington, D.C. pp. 1265-1274.

Rao, A.R. and Rao, S.G., 1974, "Analysis of the Effects of Urbanization on Rainfall Characteristics", Technical Report. Purdue University, Water Resources Research Center, West Lafayette, Indiana.

Schikcedanz, P.T., "Application of Factor Analysis in Weather Modification", Proc. of the Fifth Conference on Probability and Statistics, Las Vegas, Nevada.