Statistical Hypothesis Tests of Some Micrometeorological Observations[1,2]

S. SethuRaman and Joyce Tichler

Meteorology Group, Brookhaven National Laboratory, Upton, N.Y. 11973

(Manuscript received 9 September 1976, in revised form 21 March 1977)

ABSTRACT

Chi-square goodness-of-fit is used to test the hypothesis that the medium scale of turbulence in the atmospheric surface layer is normally distributed. Coefficients of skewness and excess are computed from the data. If the data are not normal, these coefficients are used in Edgeworth's asymptotic expansion of Gram-Charlier series to determine an alternate probability density function. The observed data are then compared with the modified probability densities and the new chi-square values computed.

Seventy percent of the data analyzed was either normal or approximately normal. The coefficient of skewness g_1 has a good correlation with the chi-square values. Events with $|g_1|<0.21$ were normal to begin with and those with $0.21<|g_1|<0.43$ were approximately normal. Intermittency associated with the formation and breaking of internal gravity waves in surface-based inversions over water is thought to be the reason for the non-normality.

1. Introduction

The assumption of a Gaussian distribution for dispersion calculations in atmospheric studies is common. The justification for the Gaussian assumption is the belief that as the sample size increases, the distribution approaches normality for a stationary, random process. Kolmogorov (1962) and Oboukov (1962) postulated that the logarithm of the energy dissipation rate ϵ is a normally distributed quantity. Stewart *et al.* (1970) and Gibson *et al.* (1970), among others, attempted to verify this hypothesis in the atmospheric surface layer by measuring the short-time average of a velocity derivative and determining its probability densities. Monin and Yaglom (1975) discuss the results of several other investigators. In general, small-scale turbulence expressed as velocity derivatives seems to conform to the log-normal distribution. However, an analysis of the basic parameters of atmospheric flow—longitudinal, lateral and vertical wind fluctuations—for their probability distributions would be helpful in determining the confidence limits of various statistical parameters of the flow. In addition, it would contribute toward a better understanding of the atmospheric processes involved.

An attempt is made in this paper to test the hypothesis that the turbulence in the atmospheric surface layer is normally distributed. The chi-square goodness-of-fit is used to achieve this purpose.

[1] Research performed under auspices of the U. S. Energy and Research Development Administration.

2. Experimental data

Observational data for velocity fluctuations were collected at the south shore of Long Island at heights of 6 and 16 m above the surface (Raynor *et al.*, 1975). A Vector Vane (manufactured by Meteorology Research, Inc.) and a single-sensor, constant temperature hot-wire anemometer (manufactured by Thermo Systems, Inc.) were used for the measurements. The masts were close enough to the land-sea interface to be above the internal boundary layer and exposed to the marine air for onshore flows. The measurements were made in various mean wind speeds ranging from 1 to 10 m s^{-1}. Stable atmospheric conditions were present for most of the cases (SethuRaman *et al.*, 1974). The intensity of turbulence $\sigma_u/\bar{u}$, where σ_u is the standard deviation of the velocity fluctuations and $\bar{u}$ the mean wind speed, varied from 0.018 to 0.14. Output from the Vector Vane corresponding to speed, azimuth and elevation angle and from the hot-wire for horizontal speed were recorded using analog tape recorders. The signals were digitized at 0.5 s intervals after passing through a low-pass filter with the cutoff frequency set at 1 Hz. The Vector Vane has been found to have a frequency response of 1 Hz (SethuRaman and Brown, 1976) and the hot-wire a frequency response of 5 Hz. In order to avoid any effect of the filter, the data were pre-averaged to 1 s means before the analysis. Fifty-two time periods of turbulence data were analyzed. Each time period was about 17 min (1000 observations) in length and was considered an event. Any trends in the data were removed by fitting a least-square line before the χ^2 test was applied. The time period of 17 min was selected since it is sufficiently long to have a large number of observations, but small enough not to be appreciably affected by diurnal trends.

3. Theoretical and statistical relations

The chi-square goodness-of-fit test developed by Karl Pearson (1900) is a widely used technique to judge whether a set of N observations of a random variable come from a given distribution. It was proven by Pearson that

$$\chi^2=\sum_{i=1}^{k}\frac{(f_i-F_i)^2}{F_i} \qquad (1)$$

has a chi-square distribution with $k-1$ degrees of freedom for large N, where f_i is the observed frequency in the ith class interval, F_i the expected frequency under the given hypothesis in the ith class interval, and k the number of class intervals into which the data are grouped.

In fitting a theoretical curve to a distribution, it is often necessary to calculate one or more parameters of the curve from the sample data. If the estimates of these parameters are the most efficient ones, the quantity χ^2 still has a chi-square distribution with $k-1-r$ degrees of freedom where r is the number of parameters estimated from the sample data.

The chi-square goodness-of-fit test can be applied in two ways: 1) by selecting class intervals of equal expected frequencies which will result in variable interval width and 2) by selecting class intervals of equal width which will result in differing expected frequencies for the class intervals. Adopting the equal frequency procedure, the data were grouped into 30 class intervals as suggested by Williams (1950) for $N=1000$. Nearly equal expected frequencies of occurrence were maintained in each interval except near the tails where expected frequencies were smaller. This procedure was suggested by Cochran (1952). The choice of 30 class intervals with near equal frequencies in each class was made to optimize the power of the test.

There is still considerable argument about the value of k to be selected for a given number of observations (Dahilya and Gurland, 1973). However, the k selected here is sufficiently large to make the test conservative. The chi-square goodness-of-fit, using Eq. (1), was then applied to the observations with the expected frequencies computed from a normal distribution. The computed χ^2 value was compared with the value for a chi-square distribution with 27 degrees of freedom at the 5% level of significance. Two degrees of freedom were lost in the estimation of mean and variance of the data.

One of the objectives of this study was to determine if the observations were at least "approximately" normal, if not normal, at the same level of significance. The ith moment M_i of a random variable x is given by

$$M_i=\frac{1}{N}\sum_{j=1}^{N}(x_j-\bar{x})^i, \qquad (2)$$

where $\bar{x}$ is the arithmetic mean of the random variable. The higher order moments M_3 and M_4 are useful in describing the statistical characteristics of a sample as they depend on the shape of the frequency curve. Using the commonly adopted procedure, M_3 and M_4 are normalized as follows:

Coefficient of skewness: $g_1=M_3/M_2^{3/2}$ (3)

Coefficient of excess: $g_2=(M_4/M_2^2)-3.$ (4)

M_2 is the variance of the sample. When g_1 and/or g_2 differ significantly (by more than one standard deviation) from their expected values for a normal distribution, a Gram-Charlier series of orthogonal polynomials can be used to improve the fit obtained by a normal curve. Edgeworth's asymptotic expansion of the Gram-Charlier series (Cramer, 1946) is used in the present analysis. The first approximation, $f_1(x)$, is given by

$$f_1(x)=\Phi(x)-\frac{g_1}{3!}\Phi^{(3)}(x), \qquad (5)$$

where $\Phi^{(3)}(x)$ is the third derivative of the normal distribution function $\Phi(x)$. Since g_1 is used in the new calculation of expected frequencies, the number of degrees of freedom for chi-square goodness-of-fit will be reduced by 1.

The second approximation, $f_2(x)$, is given by

$$f_2(x)=\Phi(x)-\frac{g_1}{3!}\Phi^{(3)}(x)+\frac{g_2}{4!}\Phi^{(4)}(x)+\frac{10}{6!}g_1^2\Phi^{(6)}(x), \qquad (6)$$

where $\Phi^{(4)}(x)$ and $\Phi^{(6)}(x)$ are the fourth and sixth derivatives of the normal distribution function. Another degree of freedom is lost by the use of g_2.

4. Results

a. Pretesting of data

One of the basic assumptions in the statistical analysis of a time series is that it is stationary and the events are random and independent. Stationarity in atmospheric data is hard to realize due to diurnal trends. In order to improve the stationarity within the period of analysis, it is customary to remove the trend in the data with the help of a least-squares fitted line. All the 52 events used in the analysis here were tested for stationarity after removing the trend using the method suggested by Bendat and Piersol (1966). Two typical transient data plots are shown in Fig. 1 with the straight line representing the least-squares fitted line.

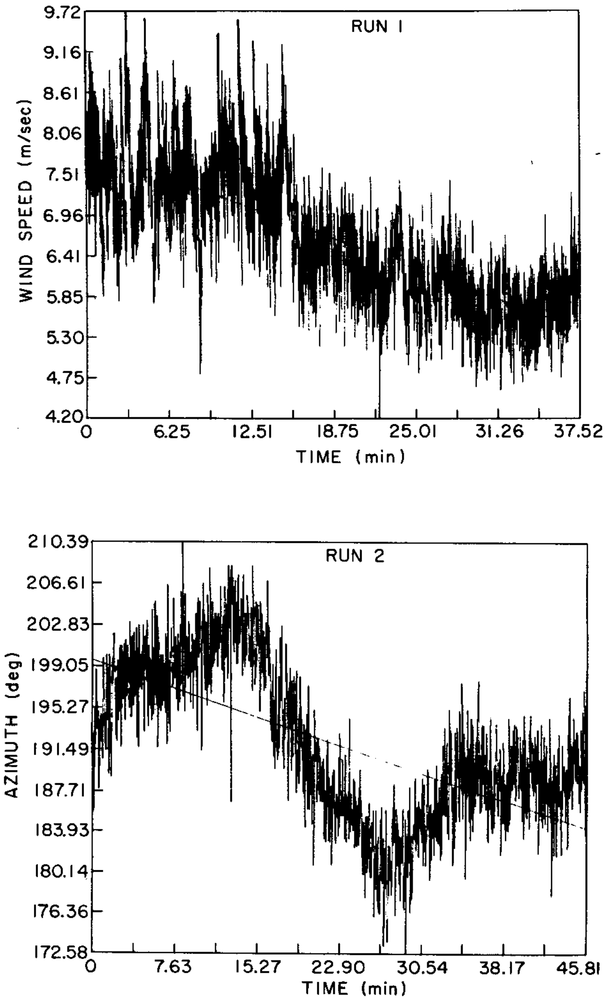

FIG. 1. Transient data for Runs 1 and 2.

Results of the test for stationarity are given in Table 1 which indicate that most of the events analysed here were within one standard deviation, σ from the expected mean.

The data were then tested for randomness without removing the trend using the run test. The run test considers the elements of a sample in the order in which they are taken. The mean is calculated and then the data classified as either above or below the mean. (The median could be used in place of the mean.) A run is defined as a sequence of values all above or below the mean.

TABLE 1. Test for stationarity after removal of trend.

	Percent of events that passed
Within 1 σ	88
Within 2 σ	94
Within 3 σ	100

TABLE 2. Test for randomness with no trend removal showing the percent of events that passed within various standard deviations.

Events	Averaging time (s) 1	5	20	60
Within 1 σ	0	4	8	64
Within 2 σ	0	4	36	84
Within 3 σ	0	16	68	96

Let

m_1 = the number of elements above the mean,
n_1 = the number of elements below the mean,
r_1 = the number of runs.

For $n_1 \geqslant m_1 > 10$, r_1 is approximately normal with the mean and the variance given by (Kenney and Keeping, 1954)

$$\left.\begin{aligned} E(r_1) &= 1 + 2m_1n_1/(m_1+n_1) \\ \operatorname{Var}(r_1) &= \frac{2m_1n_1(2m_1n_1 - m_1 - n_1)}{[(m_1+n_1)^2(m_1+n_1-1)]} \end{aligned}\right\}. \tag{7}$$

The analysis for randomness indicated that none of the events were within 10 standard deviations of the expected mean. But, as the results in Table 2 indicate, the data became more random with increasing averaging time with most of the sets of observations lying within three standard deviations of the expected mean for 1 min averages. This leads to the conclusion that the underlying stochastic process that is being sampled is a discrete-time Markov process of finite order as defined below. A process $\{x_m\}$ is a discrete-time Markov process of order t if the conditional probability

$$\operatorname{Prob}\{x_p = a_p \mid x_m = a_m,\ m < p\} \tag{8}$$

is independent of the values of a_m for $m < p - t$. Billingsley (1961) discusses the justification for using a single time function from a discrete-time Markov process of finite order for chi-square tests.

b. *Chi-square goodness-of-fit test*

The chi-square goodness-of-fit test was made on 52 independent events consisting of longitudinal, lateral

TABLE 3. Chi-square goodness-of-fit test.

	Events that passed Number	Percent
Total analyzed	52	100
Normal [Eq. (1)]	27	52
Normal or skewed normal [Eq. (5)]	30	58
Normal or normal after second approximation [Eq. (6)]	34	65
Normal [Eq. (1)] or approximately normal [Eqs. (5) & (6)]	37	71

TABLE 4. Chi-square computation for Run 1.

i	f	F [Eq. (1)]	f_1 [Eq. (5)]	f_2 [Eq. (6)]
1	6	6.2	5.3	5.1
2	28	33.9	33.0	32.9
3	36	33.5	33.3	33.4
4	36	32.1	31.6	31.7
5	26	30.0	29.4	29.5
6	25	35.4	35.2	35.3
7	44	33.7	34.4	34.4
8	38	37.2	37.9	37.9
9	31	32.3	32.9	32.8
10	42	34.3	34.8	34.8
11	37	36.0	36.5	36.4
12	47	37.5	37.9	37.8
13	38	38.7	39.0	38.9
14	41	39.4	39.6	39.5
15	38	39.8	39.9	39.7
16	47	39.8	39.7	39.6
17	35	39.4	39.2	39.1
18	32	38.7	38.4	38.3
19	37	37.5	37.1	37.0
20	33	36.0	35.5	35.5
21	28	34.3	33.8	33.7
22	32	32.3	31.7	31.7
23	42	37.2	36.5	36.5
24	29	33.7	33.0	33.1
25	30	35.4	35.6	35.7
26	25	30.0	30.6	30.7
27	40	32.1	32.6	32.8
28	35	33.5	33.7	33.9
29	34	33.9	34.8	34.7
30	8	6.2	7.1	7.0
χ^2 value		22.7	20.8	20.8

Mean wind speed = 7.45 m s^{-1}, g_1 = 0.061.
Standard deviation = 0.66 m s^{-1}, g_2 = − 0.017.

and vertical turbulence measured on 10 different days during various meteorological conditions. Results of the tests given in Table 3 indicate that about 52% of the events analyzed were normal. Three additional events became approximately normal when the first approximation was used as in Eq. (5) with a term involving g_1 and seven when the second approximation given by Eq. (6) was applied with terms involving g_1 and g_2, bringing the total to 70%.

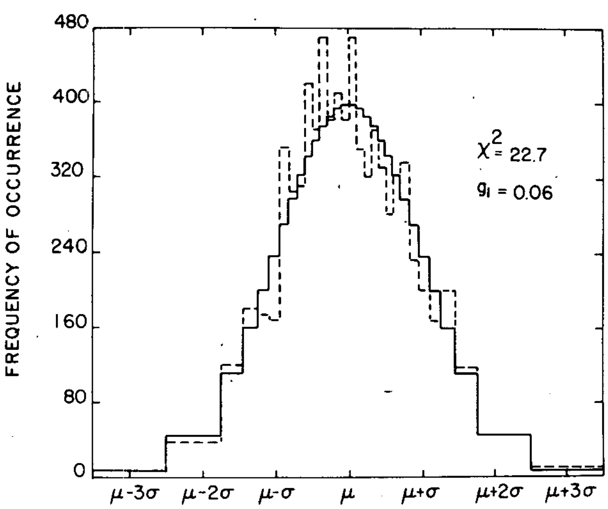

FIG. 2. Frequency histograms of observed (dotted line) and normal (solid line) distributions for Run 1.

In all cases, by "were normal" and "became normal" it is meant to at least the 5% level of significance.

Typical calculations for a set of observations (Run 1 in Fig. 1) that passed the chi-square test with the hypothesis that the distribution was normal is shown in Table 4. The first 17 min of 1 s averaged data were used. There was very little trend for this time period. Chi-square values with 27 degrees of freedom for 5% and 95% significance are 40.1 and 16.2, respectively. When the data were normal, successive approximations did not change the value of χ^2 appreciably. Frequency histograms of the observed frequencies and expected frequencies if the data were normally distributed, are shown in Fig. 2. The area under each interval would correspond to the respective frequencies.

A typical case (Run 2) where the normal fit was improved by successive approximations is given in Table 5. This corresponds to the measurement of directional fluctuations measured in degrees presented in

TABLE 5. Chi-square computation for Run 2.

i	f	F [Eq. (1)]	f_1 [Eq. (5)]	f_2 [Eq. (6)]
1	15	6.2	11.7	14.1
2	30	33.9	39.0	40.9
3	24	33.5	34.9	41.5
4	28	32.1	35.1	33.3
5	26	30.0	33.4	32.3
6	36	35.4	36.7	36.0
7	46	33.7	29.8	29.7
8	28	37.2	33.1	33.4
9	23	32.3	29.1	29.6
10	30	34.3	31.3	32.0
11	33	36.0	33.3	34.3
12	32	37.5	35.2	36.5
13	40	38.7	37.0	38.4
14	42	39.4	38.4	38.9
15	47	39.8	39.4	41.0
16	36	39.8	40.2	41.7
17	39	39.4	40.4	41.9
18	38	38.7	40.4	41.7
19	44	37.5	39.8	41.0
20	57	36.0	48.7	39.7
21	40	34.3	37.3	38.1
22	25	32.3	35.5	36.1
23	46	37.2	41.3	41.6
24	44	33.7	37.6	37.5
25	31	35.4	34.1	33.5
26	28	30.0	26.6	25.5
27	26	32.1	29.1	27.2
28	34	33.5	32.1	28.6
29	29	33.9	28.8	30.6
30	3	6.2	0.7	3.2
χ^2 value		55.0	45.6	35.7

Mean angle = 190°, g_1 = −0.358.
Standard deviation = 3.5°, g_2 = 0.526.

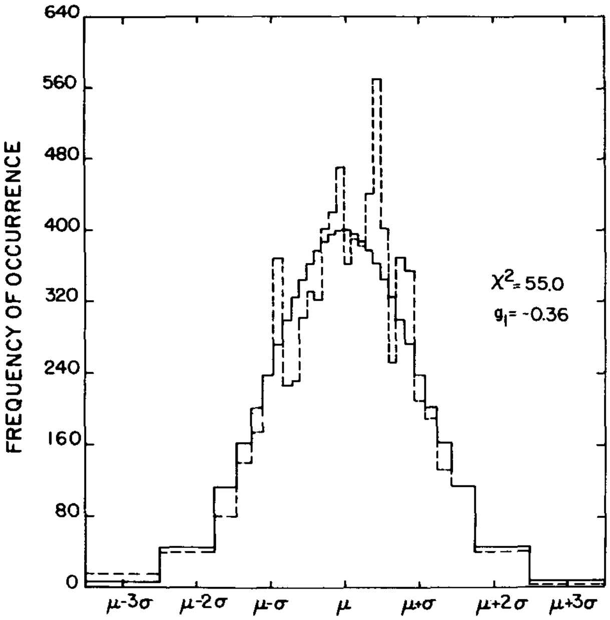

FIG. 3. As in Fig. 2 except for Run 2.

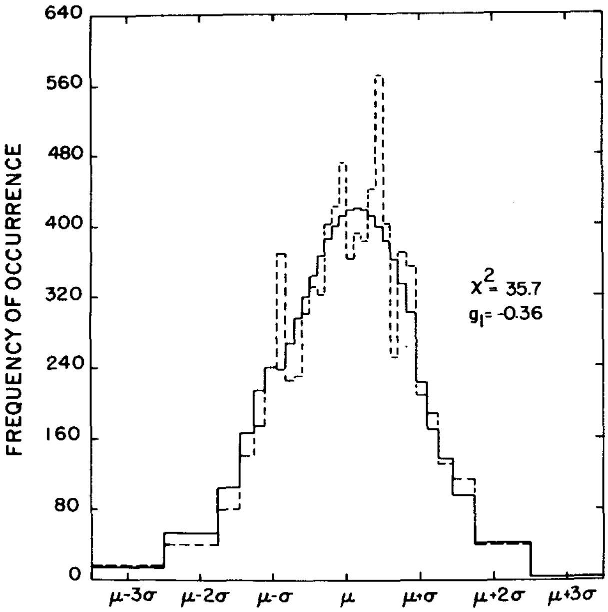

FIG. 5. As in Fig. 4 except for the second approximation (solid line) to normal using coefficients of skewness and excess.

Fig. 1. Again, only the first 17 min of 1 s averages have been used. There is an appreciable trend in the data which was removed before the chi-square test. The normal fit to the data improved with the χ^2 value changing from 55.0 to 35.7. The value at the 5% level of significance with 25 degrees of freedom is 37.7. Frequency histograms of the observed frequencies and expected frequencies for a normal distribution are shown in Fig. 3. Figs. 4 and 5, respectively, compare the observed frequencies with the modified expected frequencies after first and second approximations.

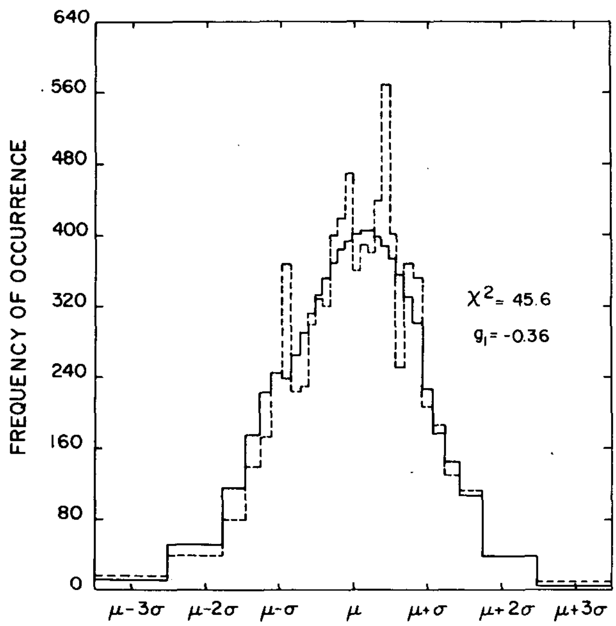

FIG. 4. Frequency histograms of observed (dotted line) and first approximation (solid line) to normal using coefficient of skewness. Run 2.

c. Coefficient of skewness

The coefficient of skewness g_1 for the sets of observations analyzed here varied from -1.1 to $+1.2$. Twenty-one events had negative skewness which is 40% of the total number. There seems to be a good correlation between g_1 and the level of significance α where

$$\alpha = \mathrm{Prob}[\chi_n^2 > \chi_{n;\alpha}^2], \tag{9}$$

χ_n^2 is the chi-square value with n degrees of freedom and $\chi_{n;\alpha}^2$ are the tabulated chi-square values. The ratio $\chi^2/(\chi^2)_E$ where $(\chi^2)_E$, the expected value for a normal distribution, is another indicator of the level of

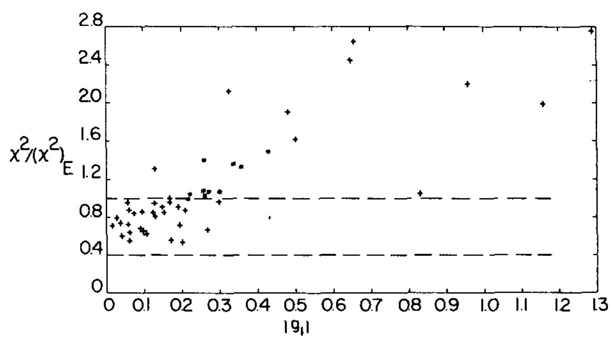

FIG. 6. Variation of $\chi^2/(\chi^2)_E$ vs $|g_1|$. Asterisks indicate events that were approximately normal. The dotted lines represent the 95 and 5% levels of significance. (Five events with the ratio greater than 2.8 are not shown in this figure.)

significance. The relationship between $\chi^2/(\chi^2)_E$ and the absolute value of g_1 is shown in Fig. 6 where $(\chi^2)_E$ is the chi-square value at the 5% level of significance. A regression analysis of the variables gave a relation of the form

$$\chi^2/(\chi^2)_E = 1.6|g_1| + 0.66 \tag{10}$$

with a correlation coefficient of 0.7. The coefficient of skewness g_1 that would correspond to $\chi^2/(\chi^2)_E = 1$ is 0.21. Events became approximately normal and passed the chi-square test for $0.21 < |g_1| < 0.43$ beyond which they were appreciably skewed. The sample mean and the variance of g_1 are given by (Cramer, 1946)

$$\left.\begin{aligned} E(g_1) &= 0 \\ M_2(g_1) &= 6(N-2)/[(N+1)(N+3)] \end{aligned}\right\}, \tag{11}$$

with $N=1000$, $M_2(g_1)=0.006$ giving a standard deviation equal to 0.077. It was found that 96% of the data that had g_1 within three standard deviations of the expected value, i.e., $|g_1|<0.23$, passed the chi-square test with the assumption that the distribution was normal.

d. Coefficient of excess

The coefficient of excess g_2 varied from -0.5 to 5.8. Twenty-six events had negative excess which is 50% of the total number analyzed. The magnitude of the positive coefficients of excess is larger than the negative ones. The range of values was more symmetric for coefficient of skewness. Although the correlation between $|g_2|$ and $\chi^2/(\chi^2)_E$ was low, the data with $-0.3 < g_2 < 0.3$ corresponding to two standard deviations from the expected value of g_2 were either normal or approximately normal. Most of the data that did not pass the chi-square test had large positive g_2 values.

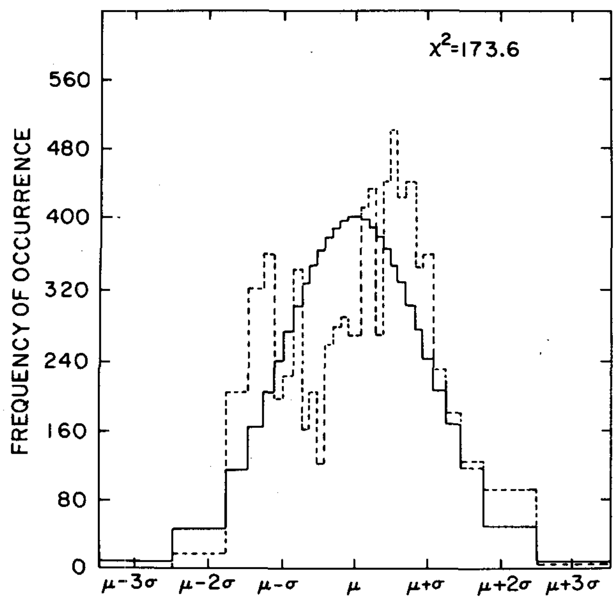

FIG. 7. Frequency histogram of observed (dotted line) and normal (solid line) distributions for typical longitudinal fluctuation data with strong intermittency. This event had a g_1 of 0.07 and a g_2 of -0.7.

5. Discussion of results

The observations analyzed indicate that about 50% of them are normal to begin with based on the chi-square test at the 5% level of significance; 20% more became approximately normal when the coefficients of skewness and excess were included in the underlying distribution. In order to determine why the other 30% of the observations were non-normal, a careful analysis of the coefficient of skewness g_1 and the coefficient of excess g_2 was made. It revealed that the observations were non-normal for $g_1>0.43$. For these observations, values of g_2 were appreciably greater than 1. The positive value of g_2 indicates that the sample distribution is less flat than the normal distribution. The probability distribution of the sample will have a higher central part and more elongated tails as compared with a normal distribution of the same variance. For a large excess, the values of the random variable tend to concentrate in separate regions, near the mean and a few standard deviations away from the mean. This is one of the signs of the phenomena of intermittency in atmospheric turbulence—alternating periods of small and large velocity fluctuations.

All the observations reported here are for onshore flows with the atmospheric stability ranging from slightly stable to strong surface-based inversions. These conditions are conducive to the formation and breaking of internal gravity waves of different harmonics causing intermittency in turbulence. A recent case study (SethuRaman, 1976) of these internal waves showed them to be comprised of wavelengths ranging from small to large scales (with corresponding periods varying from a few seconds to a few minutes). These, in turn, cause intermittencies of differing degrees depending on several meteorological parameters, *viz.*, wind speed, atmospheric stability, etc. Monin and Yaglom (1975) also indicate that small-scale turbulence is intermittent and with increasing Reynolds number, the intermittency extends even to energy-containing eddies.

The frequency distribution for longitudinal velocity fluctuations for a typical event with a strong surface-based inversion and breaking internal gravity waves is shown in Fig. 7. Values of the statistical parameters associated with this event were $g_1=0.06$, $g_2=-0.7$, $\chi_E^2=41.3$, $\chi_{obs}^2=173.6$, where χ_{obs}^2 is the observed chi-square value.

Here, again, the observations tend to concentrate at separate regions, close to the mean and away from the mean.

Thus the medium-scale turbulence in the atmospheric surface layer seems to be normal or approximately normal except when the atmosphere becomes strongly stable resulting in the formation and breaking of internal waves leading to the phenomenon of intermittency. A correlation of the normality or non-normality of the data with the occurrence of meteorological variables conducive for the formation of gravity waves will be attempted as a continuation of this work.

Acknowledgments. The authors would like to thank P. Michael, Head, Atmospheric Sciences Division, and C. S. Kao, Department of Applied Mathematics, for helpful discussions.

REFERENCES

Bendat, J. S., and A. G. Piersol, 1966: *Measurement and Analysis of Random Data*. Wiley, 407 pp. (see Chap. 4).

Billingsley, P., 1961: Statistical methods in Markov chains. *Ann. Math. Statis.*, **32**, 12–40.

Cochran, W. G., 1952: The χ^2 test of goodness of fit. *Ann. Math. Statis.*, **23**, 315–345.

Cramer, H., 1946: *Mathematical Methods of Statistics*. Princeton University Press, 575 pp. (see Chap. 17).

Dahilya, R. C., and J. Gurland, 1973: How many classes in the Pearson Chi-square test?, *J. Amer. Statis. Assn.*, **68**, 707–712.

Gibson, C. H., G. R. Stegan and R. B. Williams, 1970: Statistics of the fine structure of turbulent velocity and temperature fields measured at high Reynolds number. *J. Fluid Mech.*, **41**, 153–167.

Kenney, J. K., and E. S. Keeping, 1954: *Mathematics of Statistics*. Van Nostrand, 348 pp. (see Part 1, Chap. 13).

Kolmogorov, A. N., 1962: A refinement of previous hypothesis concerning the local structure of turbulence in a viscous incompressible fluid at high Reynolds number. *J. Fluid Mech.*, **13**, 82–85.

Monin, A. S., and A. M. Yaglom, 1975: *Statistical Fluid Mechanics. Mechanics of Turbulence*, Vol. 2. 874 pp. (see Chap. 8).

Oboukov, A. M., 1962: Some specific features of atmospheric turbulence. *J. Fluid Mech.*, **13**.

Pearson, K., 1900: On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Phil. Mag.*, Series 5, **50**, 157–172.

Raynor, G. S., P. Michael, R. M. Brown and S. SethuRaman, 1975: Studies of atmospheric diffusion from a near-shore oceanic site. *J. Appl. Meteor.*, **14**, 1080–1094.

SethuRaman, S., R. M. Brown and J. Tichler, 1974: Spectra of atmospheric turbulence over the sea during stably stratified conditions. *Preprints Symp. Atmospheric Diffusion and Air Pollution*, Santa Barbara, Calif., Amer. Meteor. Soc., 71–76.

——, 1976: An observation of atmospheric internal gravity waves in a turbulent coastal environment. *Preprints Conf. Coastal Meteorology*, Virginia Beach, Amer. Meteor. Soc., 55–59.

——, and R. M. Brown, 1976: A comparison of turbulence measurements made by a hot-film probe, a bivane, and a directional vane in the atmospheric surface layer. *J. Appl. Meteor.*, **15**, 138–144.

Stewart, R. W., J. R. Wilson and R. W. Burling, 1970: Some statistical properties of small scale turbulence in atmospheric boundary layer. *J. Fluid Mech.*, **41**, 141–152.

Williams, C. A., Jr., 1950: On the choice of the number and width of classes for chi-square test of goodness of fit. *J. Amer. Statis. Assn.*, **45**, 77–86.