Statistical Treatment of Rain Gauge Calibration Data

By MRS. A. E. FREENY

(Manuscript received December 2, 1968)

This paper describes the statistical treatment of the calibration data of the capacitance gauges used for measurement of rain rates in the rain gauge network set up in a 160 square kilometer area surrounding Crawford Hill, Holmdel, New Jersey. The expression $R = \alpha e^{-\beta f}$, *where* R *is the rain rate in millimeters per hour,* f *is the frequency of the rain gauge oscillator, and* α *and* β *are fitted coefficients, allows calibration curves to be developed for all 99 of the devices in the system with an overall equivalent standard error of 7.3 millimeters per hour. This paper discusses the distribution of parameters and residuals and gives a refinement which corrects for a fitting bias.*

I. INTRODUCTION

This paper describes the statistical treatment of the calibration data of capacitance gauges used for measurement of rain rate in the rain gauge network set up in a 160 square km area surrounding Crawford Hill, Holmdel, New Jersey. Section II describes the original calibration data plus added observations, Section III discusses the least squares fits to the individual oscillators and their results, and Section IV describes the correction made for bias.

II. CALIBRATION DATA

The calibration data consisted of approximately 20 pairs of readings for each of the 100 gauges.[1,2] These readings, which consist of frequency in KHz and rain rate in mm per hour, were obtained by pouring water via a flowmeter through the gauge at a given rate and recording the corresponding oscillator frequency reading. The data used includes the original calibration readings plus supplementary readings made in the spring of 1967 when a field check on the original

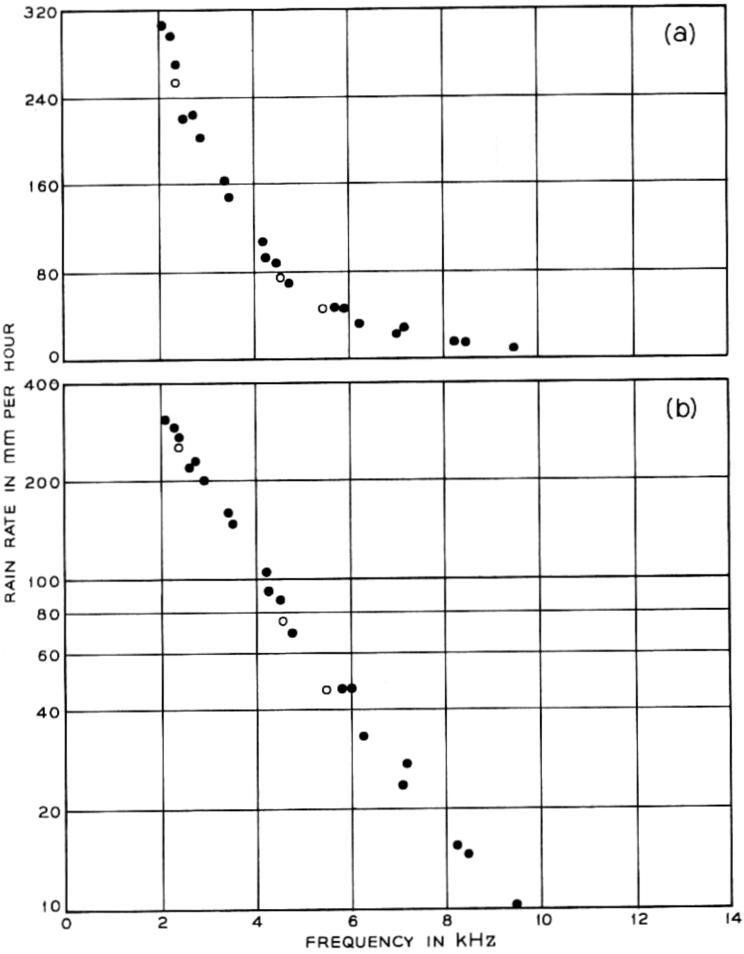

Fig. 1 — Rain rate versus frequency, oscillator 47.

calibrations was made. For oscillator 47, which shows a typical configuration, rain rate R is plotted against frequency f on both a linear and a semilog scale in Fig. 1. Points plotted with a "●" are the original calibration measurements; those with a "○" are added measurements.

III. EXPONENTIAL FITS TO INDIVIDUAL OSCILLATORS

The linear character of the semilog plots show that $R = \alpha e^{-\beta f}$, that is, a simple exponential curve, is a reasonable description of the data.

After some consideration the exponential form, rather than the linear alternative $\ln R = \ln \alpha - \beta f$, was chosen so that the residuals would be equally weighted over the frequency range. This minimizes the relative errors at high rain rates. The data for each of the oscillators were put through a nonlinear least squares fitting program; estimates a and b of the parameters α and β were obtained. Figure 2a shows the fit of the data to oscillator 47.

For some of the oscillators, it was clear from the additional readings that the calibration at this time was not the same as it had been at the time the original measurements were made. For these oscillators, the original data were not used and fitting was done on the supplementary data only. Two of these cases were left with only the three additional points, and one had only five points, so the quality of the estimates for these oscillators is poor. In all, six oscillators had less than ten points.

Inspection of the residuals from the fits for the individual oscillators showed very good agreement between the calculated values and the observed values in most cases; however, a significant number of the oscillators showed two types of nonrandom scatter about the fitted curve.

The first type is a pronounced lack of fit of the exponential curve to low rain rate data ($R \leqq 30$ mm per hour) for oscillators 13, 21, and 35.

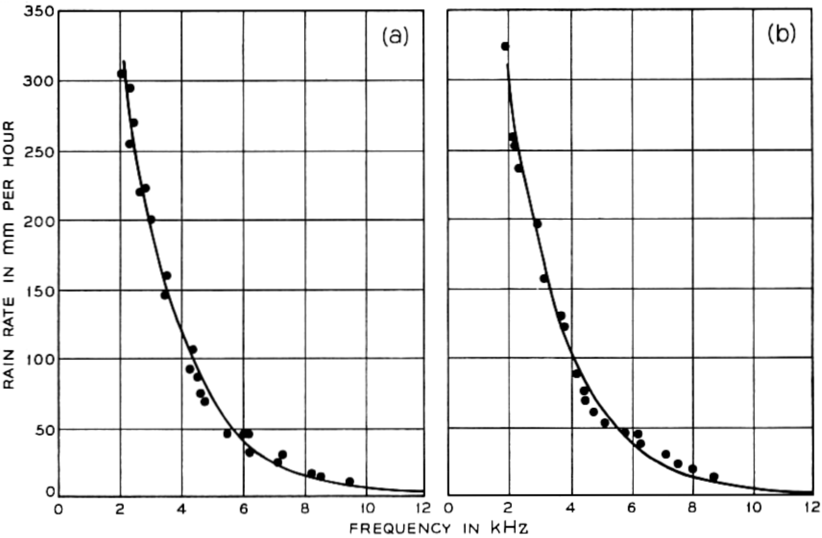

Fig. 2 — Exponential fits: (a) oscillator 47 [$R = 883.84 \exp(-0.516f)$]; (b) oscillator 76 [$R = 849.86 \exp(-0.534f)$].

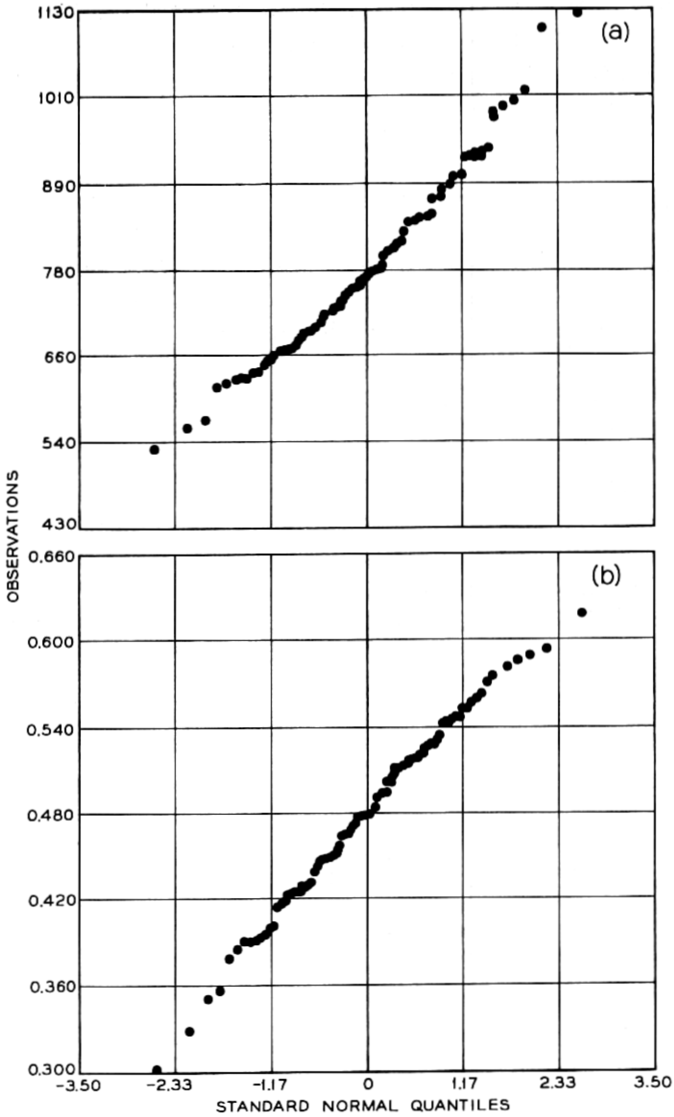

Fig. 3 — Normal probability plots of estimates from exponential fits to each oscillator: (a) alpha, (b) beta.

The fitted curve overestimates the rain rate by as much as 20 mm per hour at very low rates; but the overestimate disappears by about 30 mm per hour.

The second type of scatter is a tendency of the observed rain rate to lie below the curve for $4 \lesssim f \lesssim 5$ KHz, and above the curve for $5 \lesssim f \lesssim 9$ KHz for some oscillators. Figure 2b shows the fit of the curve to the data from oscillator 76 which follows this pattern. Several other straightforward functions were tested on a few of these oscillators but did not result in improvement of the residual patterns. However, the simple exponential equation expresses the relationship between rain rate and frequency quite well.

Of the 100 oscillators, 89 had an error mean square of less than 100, and 11 had an error mean square greater than 100. Inspection of the individual fits to the oscillators indicated that there was an outlier in the data of the oscillators having two of the 11 worst fits. These observations were removed and new fits were calculated. These have error mean squares less than 100. Of the remaining nine worst fits, one was to oscillator 56 which was not being used in the rain gauge network, so the data from this oscillator were excluded from further calculations. After removal of this data, the average value of the error mean square is 57.75. The mean square of the pooled residuals is 53.55, which is equivalent to a standard error of about 7.3.

An informal assessment of the distribution of α and β may be made from normal probability plots of a and b which show the values of a and b plotted against standard normal quantiles for a sample of size 99.[3] These plots are shown in Fig. 3. The plot of a shows an upper tail which is too long to be normally distributed. The plot of b has both tails too short for a normal distribution. It is also possible to interpret these plots as showing a slight noticeable curvature upward for a and downward for b. This suggests that the estimates from the equation in the form $\ln R = \mathrm{A} - \beta f$ would have been more nearly normally distributed.

Figure 4 is a normal plot of the pooled residuals for all oscillators. The outstanding feature of this plot is a change in the slope toward the upper tail of the distribution which could be produced by the residuals being from different distributions (hopefully of the same shape but with different parameters). This could occur if the oscillators were not all from the same population.

The plot of residuals versus rain rate (Fig. 5a) shows the tendency of the residuals to be positive for $10 \lesssim R \lesssim 30$ mm per hour, nega-

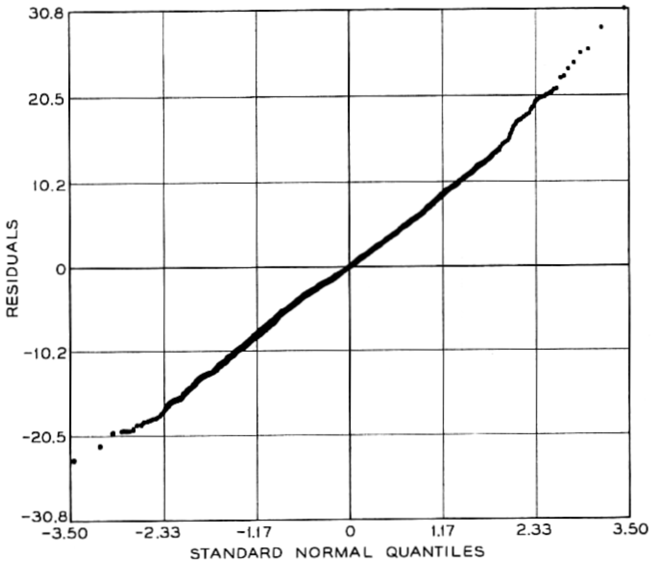

Fig. 4 — Normal probability plot of residuals from exponential fits.

tive for $50 \lesssim R \lesssim 100$ mm per hour, and positive again for $120 \lesssim R \lesssim 200$ mm per hour. This agrees roughly with the tendency toward nonrandom scatter of the residuals about the individual exponential fits and is not unexpected.

The plot of residuals versus frequency (Fig. 5b) shows the nonhomogeneity of fit over the frequency range of the oscillators, that is, that the spread of the residuals about the fitted curve is not the same for all values of the frequency. The plot also demonstrates the tendency of the residuals toward oscillation about 0.

IV. CORRECTION FOR CALIBRATION BIAS

It was decided to try to remove the calibration bias shown as a cyclic trend in Fig. 5a. Accordingly a function of the form

$$r_1 = A \sin (\alpha R^{\beta} + \varphi),$$

where r_1 is the residual from the exponential fit and R is the associated rain rate, was fit to the pooled residuals.

The resulting equation is

$$r_1 = 3.813 \sin (1.450 R^{0.3894} - 2.560)$$

with an error mean square of 46.33.

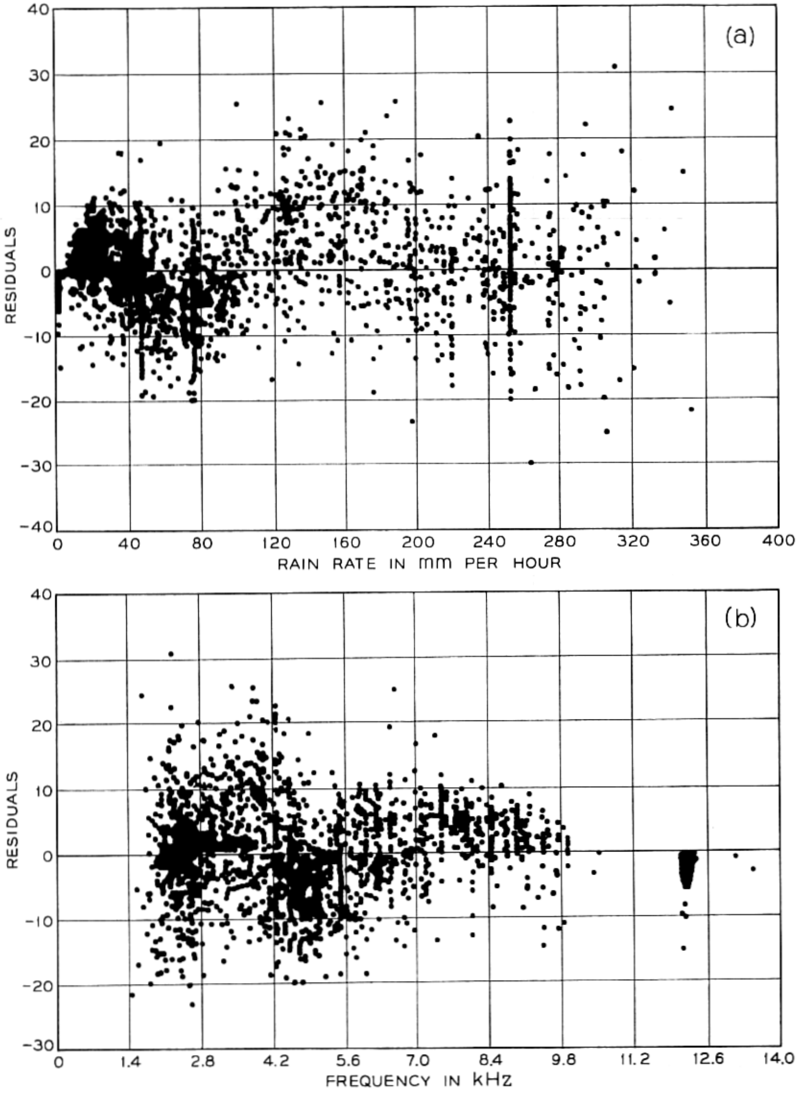

Fig. 5 — Residuals from exponential fits versus (a) rain rate and (b) frequency for all oscillators.

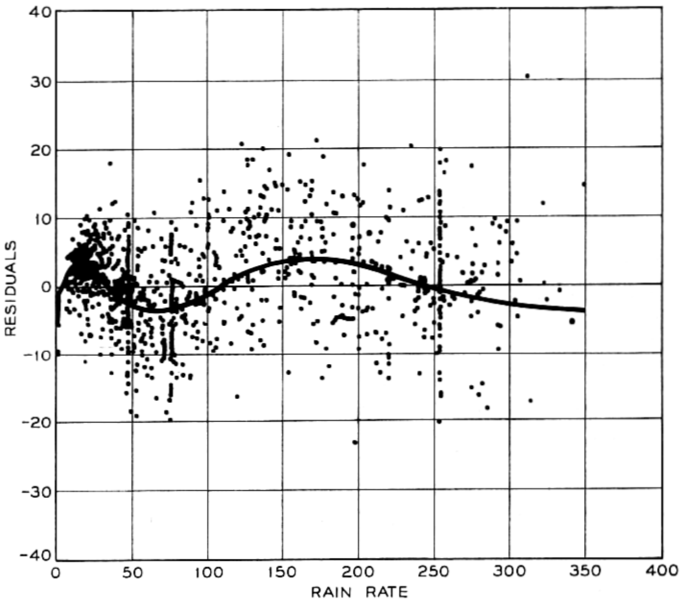

Fig. 6 — Sine fit of residuals from exponential fits to rain rate.

Figure 6 plots this fit and shows half of the residuals selected at random. The residuals from this fit (r_2) are plotted against rain rate in Fig. 7a and against frequency in Fig. 7b.

Figure 7 shows that the cyclic trend has indeed been removed by the sine function fit leaving a more random scatter of the residuals r_2. The spread of these residuals is different at different values of rain rate and frequency as was the spread of the residuals r_1. A normal plot of the residuals r_2 looks substantially like that of Fig. 4, with the same change in slope toward the upper tail, and somewhat more curvature in the extreme tails.

V. CONCLUSIONS

A sine corrected exponential fit to this calibration data produces a pooled standard error of less than 7 mm per hour in the rain rate. This is not very much better than the pooled standard error of about 7.3 mm per hour associated with the exponential fits alone. However, the sine correction has removed the average bias present in the exponential fits and will produce a small average error in the rain rates between

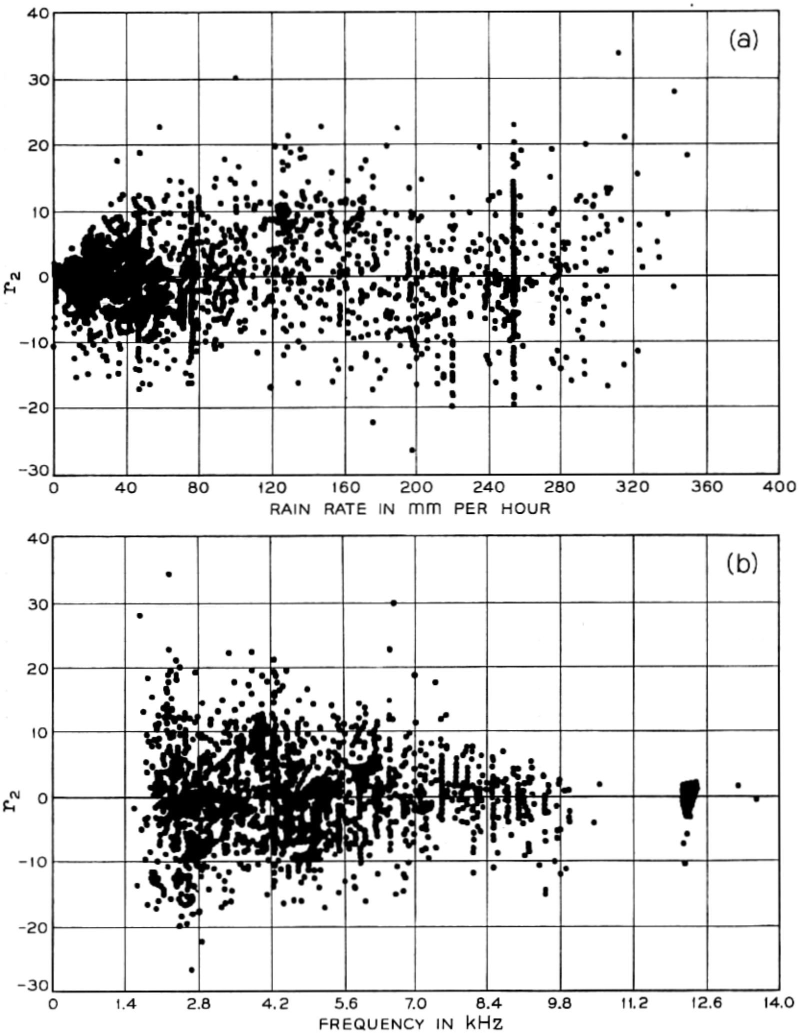

Fig. 7 — Residuals from sine fit versus (a) rain rate and (b) frequency.

10 and 100 mm per hour which is where the majority of the data of interest lies. The normal plots of the pooled residuals from the exponential fits and the residuals from the sine corrected fit are sufficiently similar so as not to give preference to either the plain or corrected exponential fit.

In processing the rainfall data collected by this network in 1967, the individual exponential coefficients for each gauge were used to calculate rain rates from the recorded frequencies. The sine correction was then applied to the rain rates from all the gauges. If the correction produced a negative value, 0 was substituted. No further corrections were made to the fits to the three gauges which showed overestimation of rain rates $\leqq$ 30 mm per hour. These rain-rate values were recorded for future analysis.

VI. ACKNOWLEDGMENT

The writer wishes to thank J. D. Gabbe for guidance throughout this analysis, R. Gnanadesikan for his remarks on the interpretation of probability plots, and C. L. Mallows, D. C. Hogg, and R. A. Semplak for their comments on an earlier draft of this paper.

REFERENCES

1. Semplak, R. A., "A Gauge for Continuously Measuring Rate of Rainfall," Rev. Sci. Instruments, *37*, No. 11 (November 1966), pp. 1554–1558.
2. Semplak, R. A., and Keller, H. E., "A Dense Network for Rapid Measurement of Rainfall Rate," B.S.T.J., this issue pp. 1745–1756.
3. Wilk, M. B., and Gnanadesikan, R., "Probability Plotting Methods for the Analysis of Data," Biometrika, *55*, No. 1 (June 1968), pp. 1–18.