DEPARTMENT OF COMMERCE
CHARLES SAWYER, Secretary

WEATHER BUREAU
F. W. REICHELDERFER, Chief

# MONTHLY WEATHER REVIEW

Editor, JAMES E. CASKEY, JR.

Volume 80
Number 8

AUGUST 1952

Closed October 15, 1952
Issued November 15, 1952

## INTERPOLATION OF MISSING PRECIPITATION RECORDS

J. L. H. PAULHUS AND M. A. KOHLER
Hydrologic Services Division, U. S. Weather Bureau, Washington, D. C.

[Manuscript received August 7, 1952]

ABSTRACT

Since 1948, when the Weather Bureau discontinued the practice of publishing interpolated precipitation data, monthly and annual totals have been omitted from its publications whenever any portion of the record was missing. In view of objections by the users of precipitation data, this policy was reviewed for possible revision. A series of tests involving some 1200 storms was conducted to determine if missing records could be satisfactorily estimated by simple procedures which could be applied in the limited time available. Results of these investigations are presented and two methods of interpolation, namely, the normal-ratio and 3-station-average, are selected for use by the Weather Bureau. Procedures to be followed in applying these methods are outlined.

Because of the early deadlines for the issuance of monthly Climatological Data, estimated data, properly identified, will appear in the annual number only, beginning with the 1952 issue. In order to limit the magnitude of the errors of interpolation, annual totals including more than one-third estimated precipitation will not be published.

### INTRODUCTION

In 1948, the Weather Bureau adopted machine methods in the preparation of climatological data for publication, centralizing the activities in seven (now three) Weather Records Processing Centers. One of the most important reasons for the change was to effect earlier publication of monthly data. Incident to this change, the interpolation of precipitation records was discontinued. It was not fully realized at the time what a large amount of summarized data would be thus eliminated from the publications nor how much of a demand there was for interpolation of monthly and annual totals. A survey of the 1950 annual Climatological Data for Colorado revealed that about 25 percent of the stations listed did not have annual amounts published. Moreover, totals for those months with incomplete records were also omitted. An analysis of daily records for 59 of the stations for which totals were not published showed the breakdown given in table 1.

TABLE 1.—*Frequency distribution of length of missing records*

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Number of days with no record | 1-5 | 6-10 | 11-15 | 16-20 | 21-25 | 26-30 | 31-55 |
| Number of stations | 21 | 14 | 10 | 4 | 4 | 2 | 4 |

On many days for which there was no record, observations at nearby stations showed that there was apparently no precipitation in the area. Table 2 shows the probable number of days with precipitation during periods of no record for which estimates would be required to obtain annual totals. On three occasions, the annual amounts were not published although there probably was no precipitation to be estimated, and the majority of the stations probably had 5 or fewer days of precipitation missing during the year.

TABLE 2.—*Frequency distribution of length of missing records with apparent no-rain days eliminated*

| | | | | | | |
|---|---|---|---|---|---|---|
| Number of days requiring estimates | 0 | 1-5 | 6-10 | 11-15 | 16-20 | 21-25 |
| Number of stations | 3 | 33 | 14 | 5 | 3 | 1 |

The omission of annual and some monthly totals for 25 percent of the available stations is a serious handicap to many users of precipitation data, especially in mountainous areas where precipitation varies greatly from station to station and the network is relatively sparse. Continued requests that the Weather Bureau resume publication of estimated data prompted reconsideration of existing policy, and an investigation of practicable interpolation methods was initiated. Because of the relatively large number of interpolations that must be made and the desirability of prompt publication, the investigation of interpolation methods was restricted to simple procedures that could be used by clerical personnel or easily adapted to solutions by machine methods. For these reasons such methods as Horton's [1] inclined plane method of interpolation were not tested.

The purpose of this paper is to present the results of the investigation, and to outline procedures followed by the Weather Bureau in applying the two methods of interpolation for use beginning with July 1952 data.

## REGRESSION EQUATION

A least-squares regression equation of the type

$$P_X=b_1P_1+b_2P_2+b_3P_3 \tag{1}$$

is known to be satisfactory for estimating the precipitation ($P_X$) at Station X from the precipitation ($P_1, P_2$, and $P_3$) at three index stations (1, 2, and 3). However, the least-squares regression is applicable only when reports are available from all the selected index stations and, moreover, the required analysis is quite time consuming. On first thought, these disadvantages appear to make equation (1) undesirable for Weather Bureau purposes, that is, for large scale application.

Obviously, the disadvantages of equation (1) would be eliminated if the regression constants $b_1$, $b_2$, and $b_3$ could be obtained by some simple procedure instead of by the relatively cumbersome least-squares method. It was believed that this might be accomplished (fig. 1) by expressing $b$ as a function of effective area (as indicated by

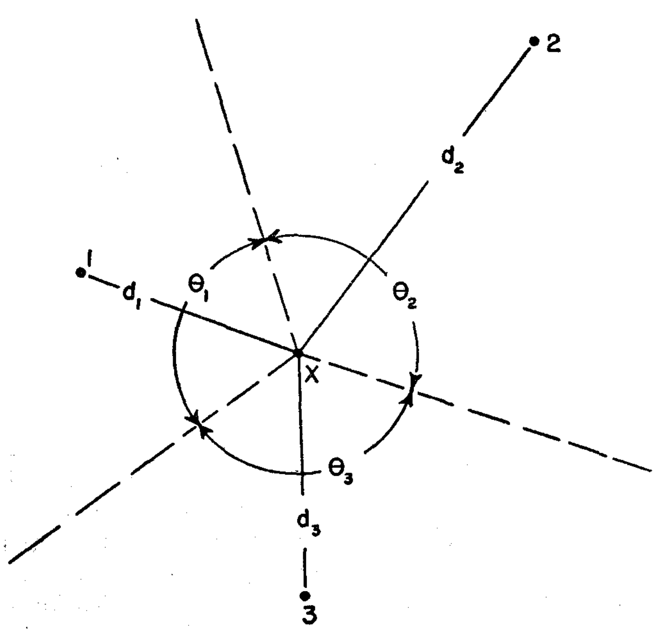


FIGURE 1.—Angle $\theta$ is formed by rays from Station $X$ bisecting the angles between lines from $X$ to Stations 1, 2, and 3, respectively. The relative distance between $X$ and an index station is $d_n/\Sigma d$.

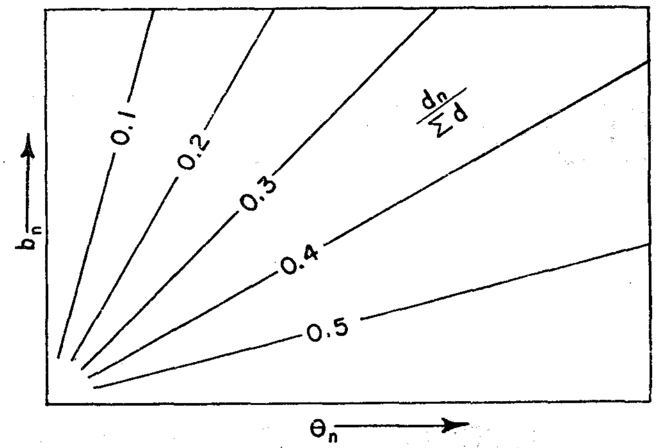


FIGURE 2.—Possible relation of regression coefficient ($b_n$) to effective area ($\theta_n$) and relative distance $d_n/\Sigma d$.

the intercepted angle $\theta$ between rays from Station $X$ bisecting the angles between lines from $X$ to index stations 1, 2, and 3) and relative distance ($d_n/\Sigma d$) between $X$ and 1, 2, and 3, respectively, or

$$b_n=f\left(\theta_n, \frac{d_n}{\Sigma d}\right). \tag{2}$$

By first finding the values of $b_n$ in equation (1) for a number of different situations by the least-squares method, it would then be possible to develop a relation between $b_n$ and $\theta_n$ and $d_n/\Sigma d$ such as

$$b_n=k_1\frac{d_n}{\Sigma d}+k_2\theta_n+k_3\frac{d_n}{\Sigma d}\theta_n \tag{3}$$

which is shown graphically in figure 2. From this graph, it would then be possible to determine quickly the proper "$b$" values for any combination of stations and then use equation (1) for estimating the missing rainfall.

In testing the feasibility of this hypothesis, equation (1) was solved for 20 sets of four stations, with 60 rainy periods for each set, making a total of 1,200 rainy periods. In order to make the test as rigid as possible, all rainy periods were selected in the May-September season, when the areal distribution of precipitation is relatively "spotty". The lengths of the rainy periods used varied from 2 to 12 days.

Unfortunately, it was not possible to determine a satisfactory relation between the 60 values of $b_n$, obtained as in the preceding paragraph, and the corresponding values of $\theta_n$ and $d_n/\Sigma d$. The method, however, appears to merit further study.

## THREE-STATION-AVERAGE METHOD

A test of equation (1) with all values of $b$ equal to ⅓ was made next. The equation thus used is nothing more than an expression of an estimate based on the average precipitation at three stations, or

$$P_X=\frac{1}{3}(P_1+P_2+P_3) \tag{4}$$

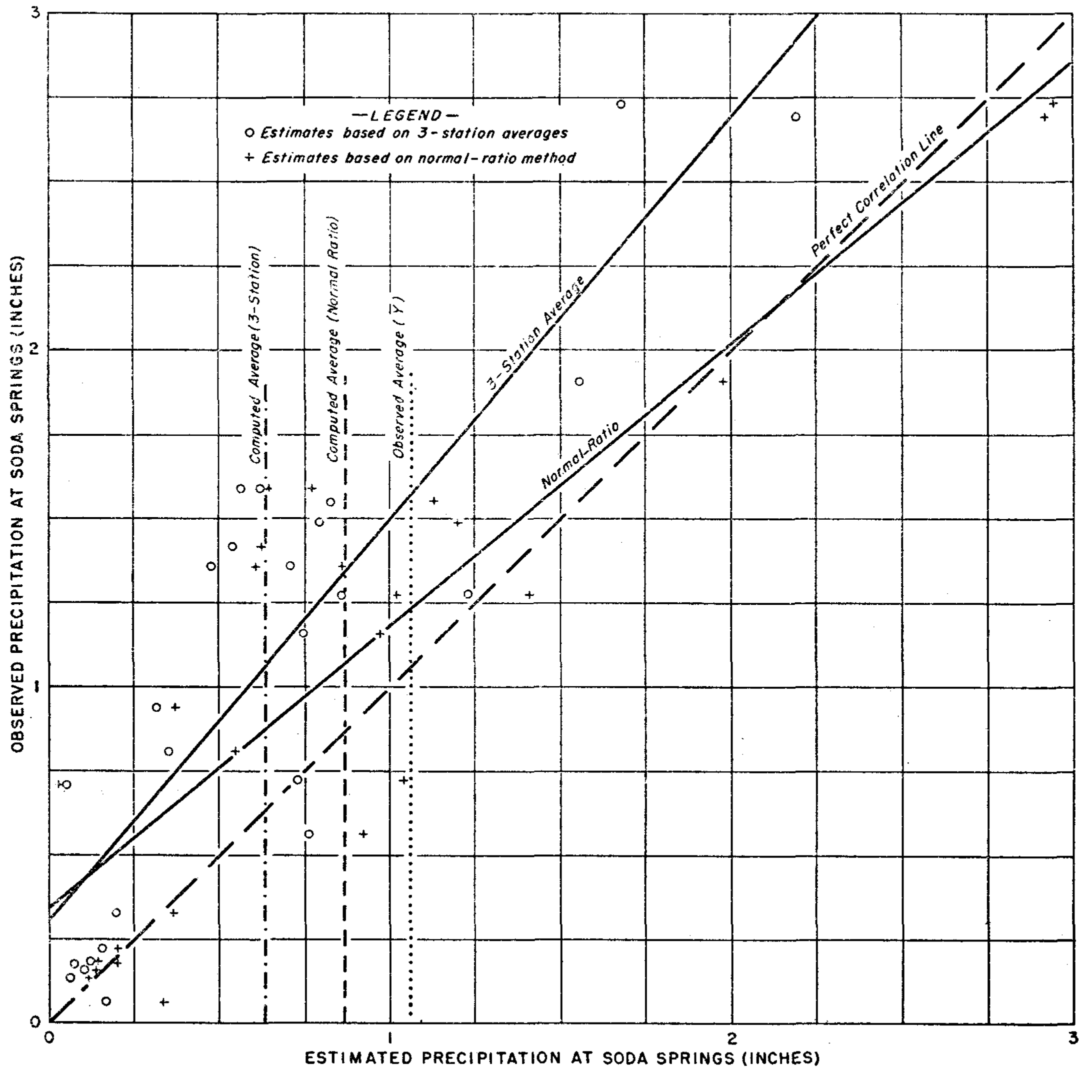


FIGURE 3.—Comparison of results of 3-station-average and normal-ratio interpolation methods in estimating precipitation for Soda Springs, Calif.

Ten sets of four stations were used in the test. Care was exercised to effect a fairly even distribution of the index stations around the interpolation station, a satisfactory distribution being considered achieved if the three index stations were located in alternate 60-degree sectors of a circle centered at the interpolation station. The estimates obtained by this method were considered to be satisfactory, though less accurate than those of equation (1) determined by the least-squares method. Horton, also, had found the three-station-average to be reasonably satisfactory.

Using the same central stations of the 10 sets used in testing equation (4), another 10 sets having four surrounding stations instead of three were selected to test whether a four-station-average would yield appreciably better results than equation (4). The resulting improvement obtained was considered insufficient to justify the added effort.

## NORMAL-RATIO METHOD

Up to this point all tests had been restricted to precipitation in the Plains region. Realizing that a straight average of precipitation at surrounding stations would not always yield satisfactory estimates in mountainous regions, it was decided to test the ratio of the normal annual precipittaion ($N_X$) at the interpolation station to that at the respective index stations as a weighting factor, or

$$P_X=\frac{1}{3}\left[\left(\frac{N_X}{N_1}\right)P_1+\left(\frac{N_X}{N_2}\right)P_2+\left(\frac{N_X}{N_3}\right)P_3\right], \qquad (5)$$

Obviously, equation (5) yields results similar to those that would be obtained by interpolation from a map of storm precipitation expressed in terms of percentage of the mean annual. It is also obvious that equation (5) is another form of equation (1), with each $b$ value being equal to

TABLE 3.—*Stations used in interpolation tests for mountainous regions*

| Set No. | State | Interpolation station | Index stations | Elevation (ft. msl) | Distance (mi.) | Normal annual precipitation (in.) |
|---|---|---|---|---|---|---|
| 1 | California | Soda Springs | | 6,752 | | 46.11 |
| | | | Nevada City | 2,600 | 35 | 52.17 |
| | | | Sierraville Ranger Station | 4,975 | 20 | 23.69 |
| | | | Truckee Ranger Station | 6,000 | 12 | 27.38 |
| 2 | Oregon | Medford | | 1,457 | | 20.26 |
| | | | Prospect | 2,473 | 30 | 38.99 |
| | | | Ashland | 2,050 | 14 | 19.98 |
| | | | Grants Pass | 940 | 24 | 29.75 |
| 3 | Washington | Cedar Lake | | 1,560 | | 103.30 |
| | | | Snoqualmie Falls | 430 | 9 | 55.19 |
| | | | Lake Cle Elum | 2,200 | 32 | 34.56 |
| | | | Buckley | 685 | 22 | 46.34 |
| 4 | Idaho | Lowman | | 3,870 | | 22.95 |
| | | | Garden Valley | 3,147 | 21 | 22.14 |
| | | | Deadwood Dam | 5,300 | 17 | 32.87 |
| | | | Atlanta | 5,560 | 30 | 24.51 |
| 5 | Colorado | Silverton | | 9,401 | | 25.37 |
| | | | Telluride | 8,756 | 12 | 23.87 |
| | | | Hermit | 8,912 | 31 | 17.35 |
| | | | Mesa Verde | 6,960 | 62 | 19.22 |

one-third the ratio of the normal precipitation at Station X to that at the corresponding index station.

Five sets of four stations (one interpolation station and three index stations fairly evenly distributed about it) were selected so as to provide large ranges of elevations within each set. To assure wide geographic and climatic coverage, one set was selected for each of the five States, California, Oregon, Washington, Idaho, and Colorado (table 3). Twenty-five rainy periods were selected from each set, and the precipitation at the interpolation station was estimated by both the three-station-average and normal-ratio methods. Figure 3 shows the results of these two types of estimates compared against the observed precipitation in the California set. The two solid lines on this figure were determined by least-squares analysis. Table 4 gives the average bias, that is, the difference between average computed and average observed precipitation, for the five sets. The normal-ratio method yielded better results than did the three-station-average method for all five sets. In four of the sets, the average error of estimate of the normal-ratio method was less than one-half that of the three-station-average.

TABLE 4.—*Average bias of estimates by 3-station-average and normal-ratio methods*

| Set No. | Average precipitation observed (in.) | Average bias | | | |
|---|---|---|---|---|---|
| | | 3-station-average | | Normal-ratio | |
| | | Inches | Percent | Inches | Percent |
| 1 | 1.06 | −0.42 | 40 | −0.20 | 19 |
| 2 | .65 | +.13 | 20 | −.10 | 15 |
| 3 | 1.62 | −.95 | 59 | −.18 | 11 |
| 4 | .42 | −.04 | 10 | +.01 | 2 |
| 5 | .67 | −.13 | 19 | 0 | 0 |

Other tests were made with ratios of normal monthly precipitation substituted for the ratios of normal annual in equation (5) but the results were not appreciably different.

## PRINCIPLES FOR APPLYING THE METHODS

The results of these tests led to the decision that the normal-ratio method will be used by the Weather Bureau whenever the normal annual precipitation at any of the index stations differs from that of the interpolation station by more than 10 percent. If the largest difference is 10 percent or less, the normal-ratio method is still applicable, but the simpler three-station-average method may be used. The plan adopted by the Weather Bureau is limited to interpolation of incomplete or missing monthly totals.

With the adoption of the two methods of interpolation, principles to govern their application were established. Those of interest to users of precipitation data are listed below.

1. Interpolated precipitation amounts are not published in current monthly Climatological Data, but monthly and annual amounts containing estimated precipitation appear in the annual section summaries, and are identified by the letter "E", indicating that the amounts are wholly or partially estimated. All interpolations are made from the published Climatological Data.

2. If the total of all estimates during one calendar year is over one-third of the established annual total or the gap in record exceeds six months, the annual total and all monthly totals containing more than one-third estimated precipitation are not to be published. Whenever the annual total is published, all monthly totals are also published, even when completely estimated. In determining what percentage of a record is estimated, only that portion of the record actually missing is considered. It should be emphasized that all restrictions apply to amounts and not to number of days for which estimates were made.

3. As far as possible, interpolations are based on comparison of adjacent records covering whole storm periods that produce general precipitation. Individual daily amounts are not estimated unless the time period of the general disturbance that produced precipitation is less than 48 hours. No effort is made to interpolate daily amounts when conditions have produced sporadic or scattered rainfall.

4. Monthly totals may be interpolated regardless of the nature of the precipitation-producing processes that characterized the month.

5. Partial records of monthly rainfall are brought up to monthly totals by the procedures outlined herein, except that no interpolations whatever are made for stations with less than one year of record.

6. In the case of snow at a station with both water equivalent and snowfall measurement missing, the water equivalent is estimated from the water equivalent at three index stations, as for rain, provided the water equivalent at each of the three index stations is a measured and not an estimated value. However, if the station with the missing water equivalent has recorded a snowfall measurement, the water equivalent will continue to be estimated as in the past by assuming 10-percent density. This assumption is known to be badly in error at times, but on the average it gives results as accurate as the average interpolation from surrounding stations, and, until an improved technique of estimating water equivalent is found, the 10-percent rule will be used.

7. Snowfall or depth of snow on ground is not interpolated by the methods described herein nor estimated from water equivalent.

8. Interpolation of missing precipitation records begins with the July 1952 data, and it is anticipated that estimates for the latter portion of 1952 can be included in the annual issue of Climatological Data for 1952. However, whenever it can be undertaken without undue hardship, the interpolation of missing records for the entire year of 1952 is being encouraged.

The selection of index stations for making the interpolations is governed by the following principles:

1. The three index stations are selected as close to the missing station and as evenly distributed about it as possible. Ideally, each station should fall into an alternate sector of a circle centered at the station and divided into six 60-degree sectors. For the normal-ratio method, stations with normals showing large departures from that of the interpolation station are to be avoided. Such a situation might prevail when the interpolation station is on a high ridge and one or more of the index stations is to the lee side of the range. In such a case it is preferable to disregard the criterion governing even distribution of index stations about the interpolation station and to select instead index stations on the same side of the ridge as the interpolation station.

2. For coastal stations, two stations with similar exposures may be used, one on either side of the missing station. If the normal or average annual precipitation at either of the two stations differs by more than 10 percent from the normal or average annual precipitation at the missing station the normal-ratio method is used. For two stations this is expressed as

$$PX=\frac{1}{2}\left[\left(\frac{N_X}{N_1}\right)P_1+\left(\frac{N_X}{N_2}\right)P_2\right]$$

where $N_X$ is the annual normal at the missing stations, $N_1$ and $N_2$ are the annual normals at the two nearby stations, and $P_1$ and $P_2$ are the respective totals for the two stations.

If a station entering into the interpolation procedure has too short a record to establish a normal, an estimate of the normal is obtained according to the following principles:

1. If one of the index stations 1, 2, or 3 has no established normal but has at least 10 years of record, the average for the period of record may be used in lieu of the normal. Normals will not be estimated for stations with less than 3 years of record. The normal for a station having only 3 to 9 years of record is estimated by extrapolation based on at least three surrounding stations. Thus, the estimate of the normal, $N_1$, at index Station 1 with a record of $n$ years where $3 \leq n \leq 9$, is expressed by the formula

$$N_1=\frac{A_1}{3}\left[\frac{N_a}{A_a}+\frac{N_b}{A_b}+\frac{N_c}{A_c}\right]$$

where the subscripts $a$, $b$, and $c$ refer to the nearest three stations surrounding Station 1 that have normals, $N_a$, $N_b$, and $N_c$, and a record for the same $n$ years as Station 1; $A$, with subscript, is the average precipitation at the respective stations for the same $n$-year period. In the selection of Stations $a$, $b$, and $c$, two of them usually may be chosen from the group $X$, 2, and 3; however, the entire original group should not be used because Stations $X$, $B$ and $C$ would all be on one side of Station 1.

2. If the station for which an estimated precipitation amount is required has no normal but has at least one year of record, the average annual precipitation for the corresponding period at the index stations is substituted for the normals. Thus, if an estimate is required for a station having only two complete years of record, the average annual precipitation for those two years only is used in determining what interpolation method should be used, and, if the normal-ratio method is indicated, is used in deriving the weighting factors. No estimates are made for stations that have less than one year of record.

3. Estimated normals are used as a last resort, that is, only when there are no stations with actual normals that meet the location criteria. If there is a choice between a nearby station with an estimated normal and a station at a greater but reasonable distance with an actual normal, the latter station is used. Wherever possible, the sectors are oriented in such a manner that the need for estimated normals is kept at a minimum.

The portion of record missing at a station determines the selection of period for estimates, according to the following principles:

1. When the record for Station $X$ is missing for an entire month, the monthly amounts at Stations 1, 2, and 3 are used as the basis of the estimate. When the record at Station $X$ is missing for only a portion of a month, the amounts used for estimating are for the period actually missed or for a few days longer, depending on whether the missing period begins and ends between storms or during a storm or storms.

2. When the missing record extends into two months, first an estimate is made of the missing amount for the total storm, without regard to month. Next, an estimate of the portion of the missing record in one of the two months is made using only precipitation for the corresponding days at stations having the same observation time as the missing station. Finally, the portion for the other month is obtained by subtracting the two values obtained in the preceding steps.

## ACKNOWLEDGMENTS

Among the several members of the Weather Bureau's Hydrologic Services Division who took part in this study, L. L. Weiss and A. L. Shands deserve special mention—the former, for several tests that he either made personally or supervised, and the latter, for critically reviewing the manuscript.

## REFERENCE

1. R. E. Horton, "Rational Study of Rainfall Data Makes Possible Better Estimates of Water Yield," *Engineering News-Record*, Vol. 79, July-December 1917, p. 212.