Double-Mass Curves

Manual of Hydrology: Part 1. General Surface-Water Techniques

---

GEOLOGICAL SURVEY WATER-SUPPLY PAPER 1541-B

Double-Mass Curves

By JAMES K. SEARCY *and* CLAYTON H. HARDISON

With a section Fitting Curves to Cyclic Data

By WALTER B. LANGBEIN

Manual of Hydrology: Part 1. General Surface-Water Techniques

GEOLOGICAL SURVEY WATER-SUPPLY PAPER 1541-B

Methods and practices of the Geological Survey

UNITED STATES GOVERNMENT PRINTING OFFICE, WASHINGTON : 1960

UNITED STATES DEPARTMENT OF THE INTERIOR

FRED A. SEATON, *Secretary*

GEOLOGICAL SURVEY

Thomas B. Nolan, *Director*

For sale by the Superintendent of Documents, U.S. Government Printing Office
Washington 25, D.C. - Price 20 cents

CONTENTS

ILLUSTRATIONS

TABLES

MANUAL OF HYDROLOGY: PART 1, GENERAL SURFACE-WATER TECHNIQUES

DOUBLE-MASS CURVES

By JAMES K. SEARCY and CLAYTON H. HARDISON

ABSTRACT

The double-mass curve is used to check the consistency of many kinds of hydrologic data by comparing data for a single station with that of a pattern composed of the data from several other stations in the area. The double-mass curve can be used to adjust inconsistent precipitation data.

The graph of the cumulative data of one variable versus the cumulative data of a related variable is a straight line so long as the relation between the variables is a fixed ratio. Breaks in the double-mass curve of such variables are caused by changes in the relation between the variables. These changes may be due to changes in the method of data collection or to physical changes that affect the relation.

Applications of the double-mass curve to precipitation, streamflow, and sediment data, and to precipitation-runoff relations are described. A statistical test for significance of an apparent break in the slope of the double-mass curve is described by an example. Poor correlation between the variables can prevent detection of inconsistencies in a record, but an increase in the length of record tends to offset the effect of poor correlation.

The residual-mass curve, which is a modification of the double-mass curve, magnifies imperceptible breaks in the double-mass curve for detailed study.

Of the several methods of fitting a smooth curve to cyclic or periodic data, the moving-arc method and the double-integration method deserve greater use in hydrology. Both methods are described in this manual. The moving-arc method has general applicability, and the double integration method is useful in fitting a curve to cycles of sinusoidal form.

INTRODUCTION

Hydrologic data generally consist of a sequence of observations of some phase of the hydrologic cycle made at a particular site. The data may be a record of the discharge of a stream at a particular place, or it may be a record of the amount of rainfall caught in a particular rain gage. Although for most hydrologic purposes a long record is preferred to a short one, the user should recognize that the longer the record the greater the chance that there has been a change in the physical conditions of the basin or in the methods of data collection. If these are appreciable, the composite record would represent only a nonexistent condition and not one that existed either before or after the change. Such a record is inconsistent.

An example of changes in physical conditions of the basin is a 25-year gaging-station record that represents 10 years of runoff from an undeveloped drainage basin, 5 years during which a major irrigation or other development was being made, and 10 years of runoff after the development. This 25-year record would not represent the runoff under natural conditions or the future runoff under the existing developed conditions, even though it is an accurate record of the runoff during the 25 years.

An example of changes in methods of data collection is a record from a rain gage operated for several years at a particular location and later moved to a different location because of new observers. Records from the rain gage at the new location might indicate an increase or decrease in average rainfall even though no real change had taken place. Similar false changes may result from variation in exposure of a rain gage or alteration in equipment.

The use of a double-mass curve as described in this manual is a convenient way to check the consistency of a record. Such a check is one of the first steps in the analysis of a long record, except when the scarcity of other old records makes it infeasible. A double-mass curve is a plot on arithmetic cross-section paper of the cumulative figures of one variable against the cumulative figures of another variable, or against the cumulative computed values of the same variable for a concurrent period of time.

Most of the credit for the initial application of the double-mass curve to hydrologic data is due C. F. Merriam, retired, Pennsylvania Water & Power Co. The description of the double-mass curve is a revision of material prepared by W. B. Langbein and others, and used within the U.S. Geological Survey since 1948.

It should not be assumed that all inconsistencies shown by a double-mass curve represent inconsistencies due to change in methods of collecting the data. Further investigation should be made to find the reason for the inconsistency, and to appraise its probable effect on the data. Unless the inconsistency shown by the double-mass curve starts at the same time as a change in method of data collection, and unless the direction of the change shown by the curve could reasonably result from the change in method, no adjustment should be applied to the observed data, because the inconsistency could be due to other causes, such as works of man or vagaries of nature.

A considerable part of this manual is devoted to analyses of precipitation records, because once they have been adjusted for changes in methods of data collection, precipitation records are independent of the works of man and thus provide an index for evaluating changes in streamflow. Such an index is desirable because a decrease in annual runoff due to a decrease in the amount of annual precipitation is not as much of a problem as a decrease due to receiving less runoff for a given amount of precipitation. Sometimes the only way to make an analysis of trends in streamflow is to use precipitation data, because streamflow may be affected by works of man as well as by changes in methods of measurement.

With the coming use of weather modification, precipitation may become less of an index. Perhaps the double-mass curve technique will be useful for detecting artificially-induced changes in precipitation so that proper allowance can be made for such changes

EXPLANATION OF THE DOUBLE-MASS CURVE

The theory of the double-mass curve is based on the fact that a graph of the cumulation of one quantity against the cumulation of an other quantity during the same period will plot as a straight line so long as the data are proportional; the slope of the line will represent the constant of proportionality between the quantities.

A break in the slope of the double-mass curve means that a change in the constant of proportionality between the two variables has occurred or perhaps that the proportionality is not a constant at all rates of cumulation. If the possibility of a variable ratio between the two quantities can be ignored, a break in the slope indicates the time at which a change occurs in the relation between the two quantities. The difference in the slope of the lines on either side of the break in the slope indicates the degree of change in the relation. Changes in the slope can be more accurately discerned if scales for the ordinate and the abscissa are chosen so that the general course of the curve is in a 45° direction relative to the axes.

In hydrologic studies, the use of the cumulations of two measured variables plotted as a double-mass curve may give indefinite results because we may be unable to say which of the variables caused a break in slope. To give more definite results, the cumulations of one of the variables can be plotted against the cumulations of a pattern composed of all similar records in a given area (Merriam, 1937). The pattern, which is composed of the average of many records, is less affected by an inconsistency in the record of any one station.

If a pattern is used, as in checking the consistency of precipitation records, enough stations should be included to insure that the average is not seriously affected by an inconsistency in the record for one of the stations. The number of stations that can be included in a pattern is sometimes limited by the criterion that the area in which the stations are located should be small enough to be influenced by the same general weather conditions. If less than 10 stations are used in the pattern, each record should be tested for consistency by plotting it against the pattern, and those records that are inconsistent should be eliminated from the pattern.

Spurious breaks in the double-mass curve that should be recognized as such are caused by the inherent variability in hydrologic data. Most users recognize that the year-to-year breaks are due to chance and, thus, ignore any break that persists for less than 5 years. Breaks that persist for longer than 5 years are more subtle in that they may be due to chance or they may be due to a real change. Unless the time of the break coincides with a logical reason for the break, statistical methods should be used to evaluate the significance of the break.

APPLICATION TO HYDROLOGIC DATA

The methods for applying the double-mass curve technique to hydrologic data and the way the results are used vary somewhat with the type of data being analyzed. Therefore, the application of the double-mass curve to records of precipitation, runoff, sediment, and precipitation-runoff are treated separately in this manual even though this requires some repetition.

In using the double-mass curve, we assume that the relation between quantities X and Y can be expressed by a line having an equation of the form $Y=bX$, where b is the slope of the double-mass curve. This assumption is substantially correct for relations involving only precipitation data but is not true for many of the relations involving streamflow data or for relations between precipitation and streamflow.

The relations involving streamflow are usually much more complex than those expressed by the simple equation $Y=bX$. Among other forms of the relation are the straight line $Y=a+bX$, where a is the intercept of the straight line on the Y axis, and the curve $Y=a+bX+cX^2$, or $Y=X^b$, where b and c are constants. If the relation is curvilinear or if it is a straight line with an intercept, spurious breaks occur in a double-mass curve. To avoid such spurious breaks, cumulations of Y are plotted against cumulations of Y_c, the correlative estimates of Y corresponding to measured values of X. In other words, when the relation between hydrologic data is complex, the double-mass curve technique requires, as a first step, a definition of the relation between the variables. Once this relation is defined, values of Y_c can be computed according to each recorded value of X.

PRECIPITATION RECORDS

Precipitation, unlike streamflow, is little affected by works of man. Moreover, records of precipitation (in the United States) are often longer than records of other hydrologic data. For these reasons, precipitation records are invaluable in hydrologic studies involving trends. Before precipitation records are used in such studies, they should be tested by the double-mass curve technique to ensure that any trends detected are due to meteorological causes and not to changes in gage location, in exposure, or in observational methods. If the changes detected are not due to meteorological causes, a precipitation record can usually be adjusted by coefficients determined from the double-mass curve.

Factors such as location and exposure affect the consistency with which a rain gage samples the rainfall in a particular area. The consistency is often affected when a gage is moved to the yard of a new observer. Even at the same location, the exposure of a precipitation gage can gradually change through the years, because of the growth of trees and other vegetation in the neighborhood or because of new buildings. Such changes sometimes go unnoted in the station history.

Differences in exposure sometimes occur even at a first-order U.S. Weather Bureau station. For example, moving a gage from one part of a roof to another can affect the catch of the gage greatly, even though the gage is moved only a few feet. Also, construction of tall buildings in the vicinity of the gage changes wind direction and affects thermal air currents, thus influencing the catch in the rain gage. Furthermore, when new post offices or Federal buildings are constructed, the rain gage often is moved from the roof of one building to that of another.

The double-mass curve technique should seldom be used for testing consistency of precipitation data in mountainous areas. The climate within a mountainous area changes with the difference in elevation, and the precipitation at two nearby stations differing greatly in elevation may be due to different meteorological events. Records from areas where the precipitation pattern for one season of the year differs greatly from that of another should be tested by double-mass curves prepared on a seasonal basis rather than a yearly basis.

In using the double-mass curve, some investigators prefer to start with the most recent data and to cumulate and to plot the data for previous years in reverse order to that shown in this report. Such a cumulation is convenient when records with different starting dates are to be tested for consistency, but it does not permit the addition of new data to the table or the graph.

CHECKING CONSISTENCY

The use of the double-mass curve for checking the consistency of precipitation records is explained by the following example in which the annual records of five precipitation stations, designated *A*, *B*,

534972 O—60——2

C, *D*, and *E*, in or near a given drainage basin are used. First the annual precipitation data for each year are tabulated and then cumulated in chronological order as in table 1. The mean of the cumulative precipitation shown in the last column of table 1 is the pattern for testing the individual station records. The cumulative precipitation for each station is then plotted against the cumulative precipitation of the pattern as shown in figure 1. For simplicity, the only double-mass curves shown in figure 1 are those for stations *A* and *E*.

The pattern used in figure 1 could be refined by testing each station against the pattern and eliminating from the average those stations, such as station *E*, whose double-mass curves show significant breaks in slope. If similar breaks occur at several stations, however, the

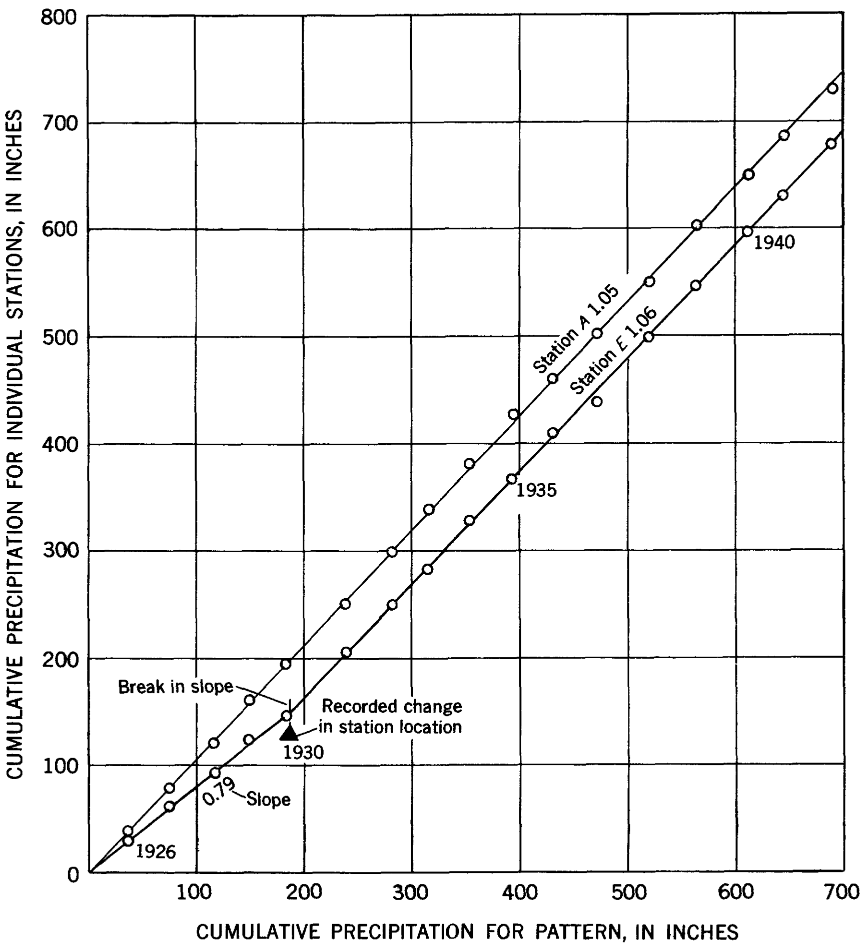

FIGURE 1.—Double-mass curve of precipitation data.

TABLE 1.—*Annual precipitation, in inches, for double-mass curve*

Year	Annual precipitation for stations indicated						Cumulative annual precipitation for stations indicated					
	A	*B*	*C*	*D*	*E*	Mean	*A*	*B*	*C*	*D*	*E*	Mean
1926	39. 75	45. 70	30. 69	37. 36	32. 85	37. 27	39. 75	45. 70	30. 69	37. 36	32. 85	37. 27
1927	39. 57	38. 52	40. 99	30. 87	28. 08	35. 61	79. 32	84. 22	71. 68	68. 23	60. 93	72. 88
1928	42. 01	48. 26	40. 44	42. 00	33. 51	41. 24	121. 33	132. 48	112. 12	110. 23	94. 44	114. 12
1929	41. 39	34. 64	32. 49	39. 92	29. 58	35. 60	162. 72	167. 12	144. 61	150. 15	124. 02	149. 72
1930	31. 55	45. 13	36. 72	36. 32	23. 76	34. 70	194. 27	212. 25	181. 33	186. 47	147. 78	184. 42
1931	55. 54	53. 28	62. 35	36. 61	58. 39	53. 23	249. 81	265. 53	243. 68	223. 08	206. 17	237. 65
1932	48. 11	40. 08	47. 85	38. 61	46. 24	44. 18	297. 92	305. 61	291. 53	261. 69	252. 41	281. 83
1933	39. 85	29. 57	32. 74	26. 89	30. 34	31. 88	337. 77	335. 18	324. 27	288. 58	282. 75	313. 71
1934	45. 40	41. 68	36. 13	32. 44	46. 78	40. 49	383. 17	376. 86	360. 40	321. 02	329. 53	354. 20
1935	44. 89	48. 13	30. 73	41. 56	38. 06	40. 67	428. 06	424. 99	391. 13	362. 58	367. 59	394. 87
1936	32. 64	39. 48	35. 40	31. 32	42. 82	36. 33	460. 70	464. 47	426. 53	393. 90	410. 41	431. 20
1937	45. 87	44. 11	39. 16	44. 14	37. 93	42. 24	506. 57	508. 58	465. 69	438. 04	448. 34	473. 44
1938	46. 05	38. 94	43. 27	50. 62	50. 67	45. 91	552. 62	547. 52	508. 96	488. 66	499. 01	519. 35
1939	49. 76	41. 58	49. 85	41. 09	46. 85	45. 83	602. 38	589. 10	558. 81	529. 75	545. 86	565. 18
1940	47. 26	49. 66	47. 86	39. 01	50. 52	46. 86	649. 64	638. 76	606. 67	568. 76	596. 38	612. 04
1941	37. 07	31. 92	32. 15	34. 45	34. 38	33. 99	686. 71	670. 68	638. 82	603. 21	630. 76	646. 03
1942	45. 89	38. 16	52. 39	47. 32	47. 60	46. 27	732. 60	708. 84	691. 21	650. 53	678. 36	692. 30

geographical location of the stations should be studied by plotting the stations on a map. If their grouping discloses a regional climatic anomaly, the breaks due to this anomaly do not necessarily indicate inconsistent records, but may indicate which stations should be grouped in a pattern.

The double-mass curve for station *E* shows a break in slope at the year 1930. The double-mass curve gives no clue as to the reason for the break but an examination of the station history reveals that, in 1930, a new observer was hired and the gage was moved to the yard of the new observer some distance away. By convention breaks such as this one that have been definitely identified with changes in location are shown by a distinctive symbol. As station histories are sometimes incomplete, especially for the earlier years, breaks may be due to changes in gage location even though the station record does not indicate that these changes took place.

The double-mass curve for station *A*, which is virtually an unbroken straight line with a slope of 1.05, indicates that this record is consistent although the points scatter slightly on both sides of the line.

ADJUSTMENT OF PRECIPITATION RECORDS

When the double-mass curve of precipitation data from a particular station indicates a break in slope and the reason for the break is determined, the record for one set of conditions may be adjusted to what it would have been if it had been collected under the other set of conditions. The period of record to be adjusted depends upon the use that is to be made of the records. The investigator should be conservative when adjusting precipitation records. Underadjustment is preferable to overadjustment.

The theory of the double-mass curve suggests the method of adjusting an inconsistent record. For example, in figure 1, the double-mass curve of station *E* would plot as a straight line if the observation conditions for 1926–30 and 1931–42 were the same; in other words, the slope of the double-mass curve would be the same for the two periods. The observed data for 1926–30 are adjusted by multiplying them by the ratio of the slope of the double-mass curve for 1931–42 to the slope for 1926–30, or

$$P_a=\frac{b_a}{b_o}P_o$$

where

P_a=adjusted precipitation
P_o=observed precipitation
b_a=slope of graph to which records are adjusted
b_o=slope of graph at time P_o was observed

Adjusted precipitation data for station E (1926–30)

[$b_a : b_o$ as 1.06:0.79]

Year	P_o	P_a
1926	32.85	44.08
1927	28.08	37.68
1928	33.51	44.96
1929	29.58	39.69
1930	23.76	31.88

The mass-curve technique is not suitable for adjusting daily precipitation or storm precipitation.

MISSING RECORD

Most analyses involving precipitation data are complicated by incomplete data for some of the years. For example, of the 128 precipitation stations in the climatic summary for Kentucky, 1931 through 1952, about 75 percent have 1 or more years for which a yearly total was not given. This percentage was computed without including incomplete years at the beginning or at the end of a record. Although records for one State selected at random cannot be considered typical, they do show that the problem of missing data is serious in a hydrologic analysis.

The practice of the U.S. Weather Bureau is to interpolate missing or incomplete monthly precipitation data by using the data from three adjacent stations (Paulus and Kohler, 1952). Where the normal annual precipitation at each of the three adjacent stations differs by less than 10 percent from that of the interpolation station (station with the incomplete record), the precipitation for the period of missing record is estimated as the straight average of the concurrent precipitation at the three adjacent stations. Where, as in a mountainous area, the normal annual precipitation at any of the three stations differs by more than 10 percent from that of the interpolation station, the normal-ratio method is used. In this method, the precipitation at each of the three stations is multiplied by the ratio of the normal annual precipitation at the interpolation station to the normal annual precipitation at each station. The weighted precipitation of the three stations is averaged to obtain the estimate for the interpolation station.

The double-mass curve can also be used to estimate missing precipation data, but the method is generally more laborious and no more accurate than the U.S. Weather Bureau method. To estimate the missing record at station A using the record of station B, multiply the data at B, corresponding to the missing period, by the ratio of the slope of the double-mass curve of station A to the slope of the double-mass curve for station B. Independent estimates are made from several adjacent stations and the estimates are averaged to obtain the estimate for station A, sometimes weighting the individual estimates when they vary considerably. The estimator should not calculate from base estimates on distant long-term stations for which double-mass curves are available rather than using nearby short-term stations that might provide a more accurate estimate.

STREAMFLOW RECORDS

Streamflow records are among the most valuable of all hydrologic data because they represent an integration of all hydrologic factors. Furthermore, the flow of streams is a sensitive indicator of climatic variations, because runoff is the residual of precipitation after the demands for evapotranspiration are satisfied. With evapotranspiration losses fairly constant from year to year in a given area, variations in annual runoff are much greater in percentage than variations in annual precipitation.

Streamflow records to be used in any analyses involving the record as a whole should be checked for consistency, and the cause of any inconsistency investigated. Sometimes the record must be adjusted to consistent conditions. Inconsistency can be due to a change in the method of collecting streamflow data or to major changes in water use, storage, or evapotranspiration in the basin.

The double-mass curve can be used to check for inconsistencies in streamflow records in much the same manner as it is used to check for inconsistencies in precipitation records, but the assumption that a constant ratio exists between a given record and a group of records in the area may not be valid. In an analysis of annual discharge, such an assumption is usually justified, even although the ratio of monthly discharge is not constant, because annual discharge is composed largely of streamflow not affected by variable basin conditions. If there is any question, a plot of the variables should be made, and, if the ratio is found to vary greatly with discharge, a double-mass curve of observed and computed values of the discharge should be used.

CHECKING CONSISTENCY

The method of using the double-mass curve to check the consistency of the streamflow record by comparing it with the records from a group of other stations in the same area is presented in the following example. In order to compute the group mean, or pattern, the streamflow data are first converted to a comparable basis, such as runoff in inches, in cubic feet per second per square mile, or in percent of mean flow; otherwise, the flow of a large stream in the group would have more effect on the pattern than would several small streams. In this example, annual runoff in inches is used as shown in table 2. The

TABLE 2.—*Annual runoff, in inches, for double-mass curve*

Water year	Stream A		Pattern		Water Year	Stream A		Pattern	
	Yearly	Cumulative	Yearly	Cumulative		Yearly	Cumulative	Yearly	Cumulative
1921	19.73	19.73	19.61	19.61	1934	9.35	244.16	7.94	189.54
1922	15.80	35.53	12.29	31.90	1935	32.75	276.91	25.58	215.12
1923	17.52	53.05	8.12	40.02	1936	7.57	284.66	4.06	219.18
1924	16.58	69.63	14.39	54.41	1937	19.72	304.38	13.76	232.94
1925	5.33	74.96	3.53	57.94	1938	28.33	332.71	28.64	261.58
1926	16.45	91.41	13.80	71.74	1939	15.04	347.75	10.41	271.99
1927	30.67	122.08	24.03	95.77	1940	13.65	361.40	10.68	282.67
1928	21.22	143.30	12.40	108.17	1941	17.42	378.82	30.15	312.82
1929	21.96	165.26	19.70	127.87	1942	17.82	396.64	21.60	334.42
1930	19.34	184.60	18.10	145.97	1943	9.41	406.05	8.96	343.38
1931	9.87	194.47	5.13	151.10	1944	21.13	427.18	20.01	363.39
1932	24.81	219.28	18.30	169.40	1945	37.85	465.03	40.25	403.64
1933	15.53	234.81	12.20	181.60					

data for the individual stations comprising the pattern are omitted from this table and only their averages are shown.

When the cumulative runoff from table 2 is plotted on figure 2, a decided break in slope is noted at the 1938 point. The mean annual runoff, in inches, of periods on both sides of the break is as follows:

Period	*Stream A*	*Pattern*
1921–38	18.48	14.53
1939–45	18.90	20.29

The later period is shown to be somewhat wetter than the earlier period, but the increase is less marked on stream *A* than on the streams

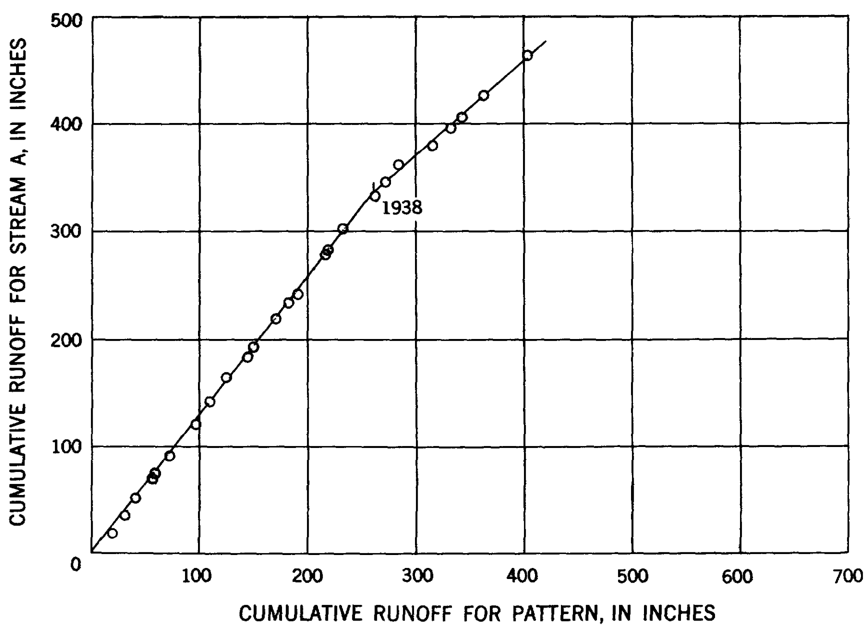

FIGURE 2.—Double-mass curve of streamflow data.

comprising the pattern. Furthermore, during the earlier period the mean for stream A is higher than that of the pattern, and during the later period the relation is reversed. These contrasts suggest an inconsistency in the record for stream *A*.

The next step would be to examine the record for stream *A* and try to find a reason for the inconsistency. Sometimes the record should be recomputed or adjusted as explained in the next section.

If the break in the double-mass curve cannot be explained by error in the base data, it would be wise to make the statistical test for significance described in a later section. This test shows the likelihood of the break occurring by chance alone as compared with a real reason for the break. The statistical test for significance also furnishes a valuable guide to conclusions as to the effect on streamflow of known changes in stream regimen.

If the statistical test shows that the break is unlikely to have occurred by chance, further hydrologic analysis is required to find the reason for the break and to decide if the change in conditions is permanent or temporary.

ADJUSTING STREAMFLOW RECORDS

The double-mass curve, which is useful in detecting inconsistencies, should seldom be used to adjust streamflow records. Instead, the adjustment should be based on other analyses made after a probable source of the inconsistency has been found. If the inconsistency is due to a change in method, such as an improvement in the high water rating, the record would be recomputed on the basis of revised ratings. If the inconsistency is due to diversions out of the basin, the amount of the diversion would be estimated and added to the observed discharge. If a significant inconsistency is attributed to changes in land use or land management practices and if a hypothetical record for one of the periods is desired, estimates should usually be obtained through correlation methods. The reason for not adjusting streamflow records by the percentage shown by the break in the double-mass curve is that the adjustment likely varies throughout the range in discharge and, thus, would have a different average for a series of wet years than for a series of dry years.

SEDIMENT RECORDS

The double-mass curve is used for studying trends in sediment yield and in detecting the effect of watershed practices on sediment yield.

Figure 3, from Guy's (1957) report, is a double-mass curve of cumulative sediment plotted against the cumulative runoff. The slope of the double-mass curve defines the mean concentration during the period. The double-mass curve for storm runoff and sediment dis-

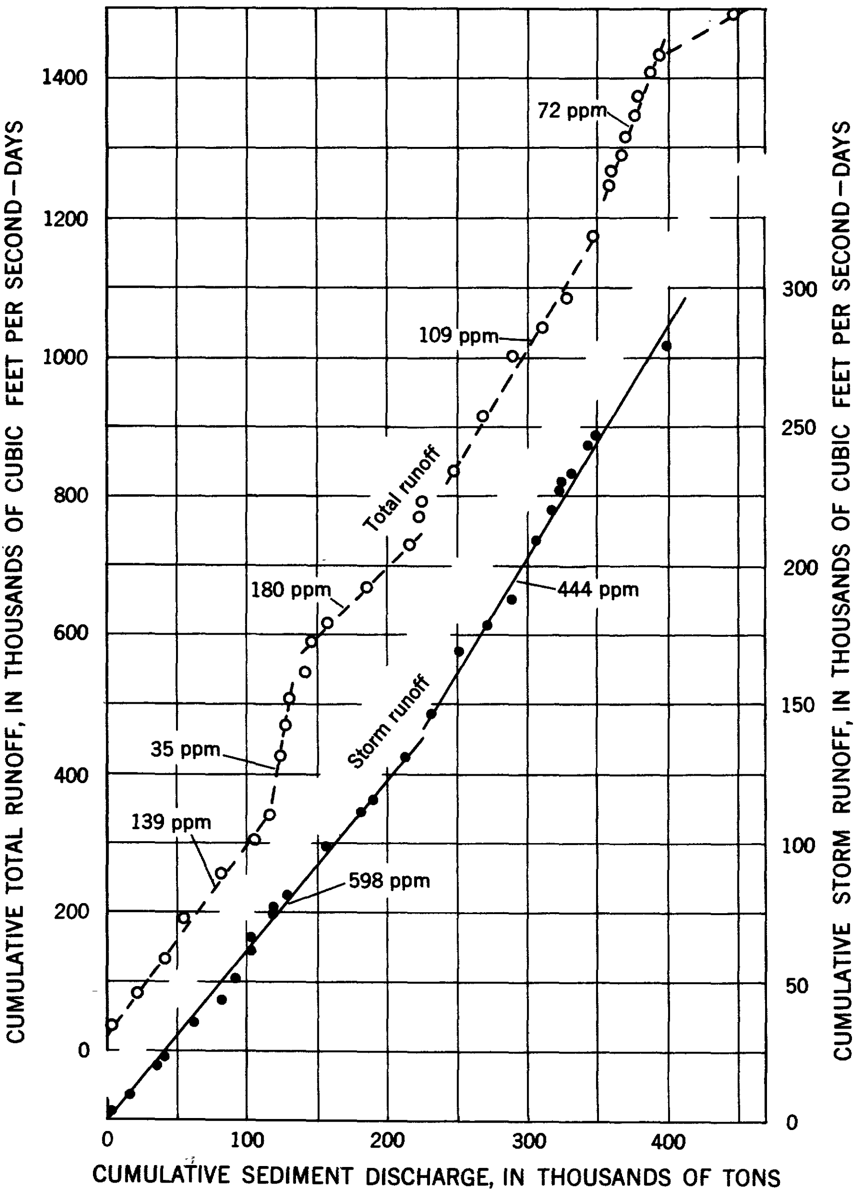

FIGURE 3.—Cumulative sediment versus water discharge by 3-month increments for storm runoff and total runoff, Brandywine Creek at Wilmington, Del., January 1947 to September 1955.

charge can be represented by two straight lines with a break at a cumulation of about 225,000 tons. The cumulation before the break has a mean sediment concentration of 598 ppm and the cumulation after the break has a mean concentration of 444 ppm.

534972 O—60——3

The double-mass curve of total runoff and sediment discharge is a series of lines differing in slope. The variable relation between total runoff and sediment discharge is understandable because, during the period of the study, storm-runoff periods accounted for 89 percent of the sediment but for only 19 percent of the total runoff.

PRECIPITATION-RUNOFF RELATIONS

When the double-mass curve is used to study trends or possible changes in precipitation-runoff relations (Oltman and Tracy, 1951) or when it is used with precipitation to check the consistency of streamflow records (Merriam, 1937), the cumulative measured runoff should be plotted against the cumulative computed runoff, taken from a precipitation-runoff relation. A double-mass curve of cumulative measured runoff and cumulative precipitation should not be used because the relation between precipitation and runoff is seldom a constant ratio even during a period when there was no change in the relation. As explained earlier in this report a variable ratio would violate the theory of the double-mass curve.

AREAL PRECIPITATION

The average amount of precipitation on a particular river basin is obtained by averaging the precipitation measured at the rain gages in and near the basin. Methods used to compute the average rainfall on the basin are:

1. Using the catch of a single gage for small areas (1 square mile or less).
2. Computing the arithmetic mean of the rain-gage records.
3. Weighting the precipitation at individual stations by the Thiessen method. (Horton, 1924.)
4. Drawing isohyetal lines and computing a weighed average.

The application of the various methods is discussed in Water-Supply Paper 846 (Williams and others, 1940) and in standard texts on hydrology.

Precipitation in a given area generally increases with the altitude, therefore the average precipitation on a basin with a wide range in elevation should be adjusted (Williams and others, 1940) for this increase.

EFFECTIVE PRECIPITATION

The amount of precipitation that fell the previous year is one of the factors that affect the relation between precipitation and runoff and cause the points on a graph of annual precipitation plotted against annual runoff to assume a "shotgun" pattern. Generally the scatter of the points can be reduced by plotting an effective precipitation instead of the observed precipitation.

Use of an effective precipitation is one way of making allowance for the variable amount of water carried over from year to year as ground water storage in the basin. The effective precipitation (P_e) commonly used is that proportion of the current year's precipitation (P_o) and the proportion of the preceding year's precipitation (P_1) that furnishes the current year's runoff, or

$$P_e = aP_o + bP_1$$

When the carryover effect extends beyond the previous year, portions of additional years ($P_2 \ldots P_n$) with their coefficients are added to the effective precipitation formula. The sum of the coefficients, a, b, c, etc., must equal unity. The coefficients a and b can be determined by rank correlation, as shown in table 3.

In table 3, the yearly runoff (column 3) is assigned a number according to rank beginning with the highest runoff as number 1 as shown in column 2. When two values are identical, they are both assigned the average of the two order numbers they would have if they were slightly different from each other. The yearly runoffs are cumulated in column 4.

The yearly average precipitation (P_o) over the basin is tabulated in column 6 and order numbers are assigned as was done for the yearly runoff and are shown in column 5. The difference in rank or order number between the precipitation and the runoff is squared and noted in column 7. The sum of the squares of the difference in rank is obtained.

A formula for effective precipitation is assumed and values of effective precipitation are computed. In the first trial, tabulated in columns 8–10, the following formula is assumed:

$$P_e = 0.8P_o + 0.2P_1.$$

The values of effective precipitation are assigned order numbers and the difference in rank between these numbers and the rank of the runoff for the same year are squared. The sum of the squares of the difference in rank, 515, is compared with that for the current year's precipitation, 775. In the first trial the effective precipitation agrees with the runoff better than the current year's precipitation as shown by a lower sum of squares of the difference in rank.

A second assumption for effective precipitation is made where:

$$P_e = 0.6P_o + 0.4P_1.$$

The procedure of the foregoing paragraph is repeated; the sum of squares of the difference in rank, 556, is found to be greater than that of the first trial.

TABLE 3.—*Correlation of annual precipitation and annual runoff, Colorado River near Grand Canyon, Ariz.*

[The yearly runoff for 1921 and for 1922 is estimated. Columns 2, 5, 8, 11, and 14 are assigned numbers according to order of rank beginning with the highest runoff as No. 1]

Water year	Runoff (millions of acre-feet): Relative rank	Runoff: Yearly (Y)	Runoff: Cumulative	Precipitation (inches), Original data: Relative rank	Original data: Yearly average precipitation	Original data: Column $(5-2)^2$	Effective precipitation assumptions[1], First assumption: Relative rank	First assumption: $0.8P_o + 0.2P_1$	First assumption: Column $(8-2)^2$	Second assumption: Relative rank	Second assumption: $0.6P_o + 0.4P_1$	Second assumption: Column $(11-2)^2$	Third assumption: Relative rank	Third assumption: $0.7P_o + 0.3P_1$ (X)	Third assumption: Column $(14-2)^2$	Cumulative $0.7P_o + 0.3P_1$	X^2	Y^2	XY
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1920					16.68														
1921	1	21.27	21.27	4	16.52	9	3	16.55	4	2	16.58	1	3	16.57	4	16.57	274.56	452.41	352.44
1922	3	17.84	39.11	8	14.16	25	7	14.63	16	7	15.10	16	7	14.87	16	31.44	221.12	318.27	265.28
1923	6	17.05	56.16	5	16.30	1	5	15.87	1	6	15.44	0	4	15.66	4	47.10	245.24	290.70	267.00
1924	14	13.01	69.17	23	11.56	81	[2]20.5	12.51	42	13	13.46	1	16	12.98	4	60.08	168.48	169.26	168.87
1925	18	11.74	80.91	6	15.15	144	[2]8.5	14.43	90	12	13.71	36	12	14.07	36	74.15	197.96	137.83	165.18
1926	11	14.42	95.33	13	13.78	4	12	14.05	1	10	14.33	1	11	14.19	0	88.34	201.36	207.94	204.62
1927	[2]4.5	17.26	112.59	1	20.36	12	1	19.04	12	1	17.73	12	1	18.39	12	106.73	338.19	297.91	317.41
1928	[2]9.5	15.63	128.22	18	12.49	72	11	14.06	2	5	15.64	20	8	14.85	2	121.58	220.52	244.30	232.11
1929	2	19.43	147.65	2	18.43	0	2	17.24	0	3	16.06	1	2	16.65	0	138.23	277.22	377.52	323.51
1930	13	13.42	161.07	[2]9.5	14.00	12	6	14.89	49	4	15.77	81	6	15.33	49	153.56	235.01	180.10	205.73
1931	25	6.74	167.79	26	9.94	1	25	10.55	0	25	11.56	0	25	11.16	0	164.72	124.55	45.16	75.22
1932	8	15.97	183.76	12	13.85	16	15	13.07	49	21	12.29	169	19	12.68	121	177.40	160.78	255.04	202.50
1933	21	10.01	193.77	22	12.12	1	22	12.47	1	19	12.81	4	20	12.64	1	190.04	159.77	100.20	126.53
1934	26	4.66	198.43	25	10.09	1	26	10.50	0	26	10.90	0	26	10.70	0	200.74	114.49	21.72	49.86
1935	20	10.22	208.65	15	13.12	25	[2]20.5	12.51	0	23	11.91	9	22	12.21	4	212.95	149.08	104.45	124.79
1936	16	12.32	220.97	16	12.72	0	17	12.80	1	18	12.88	4	18	12.84	4	225.79	164.87	151.78	158.19
1937	15	12.41	233.38	19	12.47	16	19	12.52	16	20	12.57	25	21	12.54	36	238.33	157.25	154.01	155.62
1938	[2]9.5	15.63	249.01	11	13.99	2	14	13.69	20	15	13.38	30	14	13.53	20	251.86	183.06	244.30	211.47
1939	22	9.62	258.63	24	11.14	4	24	11.71	4	22	12.28	0	23	12.00	1	263.86	144.00	92.54	115.44
1940	24	7.44	266.07	21	12.31	9	23	12.08	1	24	11.84	0	24	11.96	0	275.82	143.04	55.35	88.98
1941	7	16.94	283.01	3	16.94	16	4	16.01	9	8	15.09	1	5	15.55	4	291.37	241.80	286.96	263.42
1942	[2]4.5	17.26	300.27	14	13.65	90	10	14.31	30	9	14.97	20	9	14.64	20	306.01	214.33	297.91	252.69
1943	19	11.43	311.70	7	14.63	144	[2]8.5	14.43	110	11	14.24	64	10	14.34	81	320.35	205.64	130.64	163.91
1944	12	13.53	325.23	17	12.51	25	16	12.93	16	16	13.36	16	15	13.15	9	333.50	172.92	183.06	177.92
1945	17	11.87	337.10	[2]9.5	14.00	56	13	13.70	16	14	13.40	9	13	13.55	16	347.05	183.60	140.90	160.84
1946	23	9.09	346.19	20	12.39	9	18	12.71	25	17	13.03	36	17	12.87	36	359.92	165.64	82.63	116.99
Total		346.21				775			515			556		359.92	480		5,064.48	5,022.89	4,946.52
Average		13.32												13.84					

[1] P_1 is the preceding year's precipitation and P_o is the current year's precipitation in the effective precipitation assumptions.
[2] When two values are identical, they are both assigned the average of the two order numbers they would have if they were slightly different from each other.

A third assumption for effective precipitation is made where:

$$P_e = 0.7P_o + 0.3P_1.$$

The sum of squares of the difference in rank, 480, is the lowest of the four trials. The relation

$$P_e = 0.7P_o + 0.3P_1$$

is thus adopted as the effective precipitation that will correlate best with the yearly runoff. It is noted in table 3 that the sums of squares decrease with increasing percentages of the previous year's precipitation until a low point, in this example $0.3P_1$, is reached. Beyond this point, the sums of squares increase with increasing percentages of the preceding year's precipitation.

COMPUTED RUNOFF

Computed runoff as used in this report is the runoff taken from a curve of relation of precipitation and runoff. It represents the most probable amount of runoff for a given amount of effective precipitation under conditions such as existed during the period used to define the curve.

In figure 4 is an example of a curve used to obtain computed runoff. The yearly values of effective annual precipitation are plotted against the annual runoff, and a curve is drawn to average the shotgun pattern of plotted points. In this example the points can be averaged by a straight line.

One form of the equation of a straight line is:

$$Y = a + bX$$

in which, for the least squares method:

$$a \text{ (the intercept on the } Y\text{-axis)} = \overline{Y} - b\overline{X}$$

$$b \text{ (the slope of the line)} = \frac{\Sigma XY - N\overline{X}\overline{Y}}{\Sigma X^2 - N\overline{X}^2}$$

X = An annual effective precipitation value

Y = An annual runoff value

and

Σ = sum of

$\overline{X}, \overline{Y}$ = average of the X's or of the Y's

N = number of pairs of items

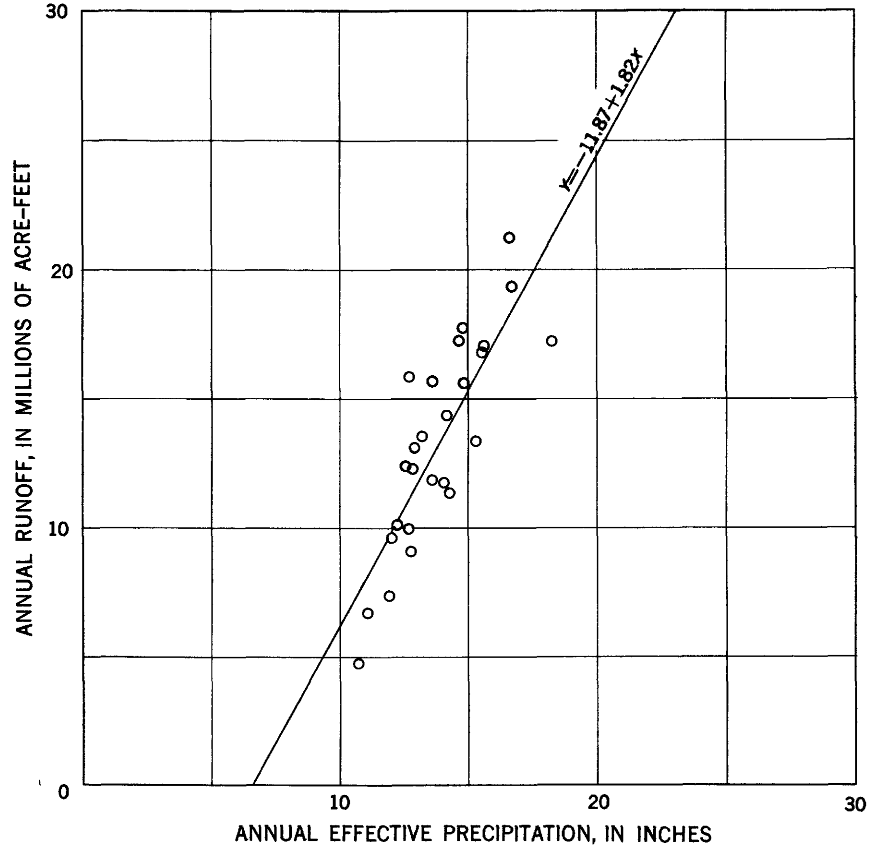

FIGURE 4.—Relation of effective annual precipitation to annual runoff, Colorado River near Grand Canyon, Ariz., 1921–46.

Using this formula for the data in figure 4 and substituting the values from columns 18 to 20 of table 3:

$$b=\frac{4{,}946.52-26(12.84)(13.32)}{5{,}064.48-26(13.84)^2}=\frac{153.45}{84.29}=1.82$$

$$a=13.32-1.82(13.84)=-11.87$$

Thus $Y=-11.87+1.82X$

The computed runoff given by Y is usually indicated by the symbol Y_c or R_c.

If the points in figure 4 had indicated a curved relation, a line would not have been fitted by least squares; instead, a smooth curve to average the points would have been drawn.

The computed runoff is the runoff expected from a given amount of precipitation if there is no change in the precipitation-runoff relation during the period. The double-mass curve of measured runoff versus computed runoff thus would reveal changes in the precipitation-runoff relation as well as possible inconsistencies in the records.

In the example, an expected annual runoff (Y_c) is computed for each year from the relation $Y_c = -11.87 + 1.82X$, where X is the effective annual precipitation. The expected annual runoff is listed in column 4 of table 4 and cumulated in column 5. The annual and cumulated measured runoff is transferred from table 3. The double-mass curve of measured runoff versus computed runoff is plotted in figure 5.

The double-mass surves for the period 1921–26 and for the period 1927–46 are about parallel but are offset between 1926 and 1927. The offset is explained by the large difference between the measured and the expected runoff in a single year, 1927, and should not be interpreted as a break.

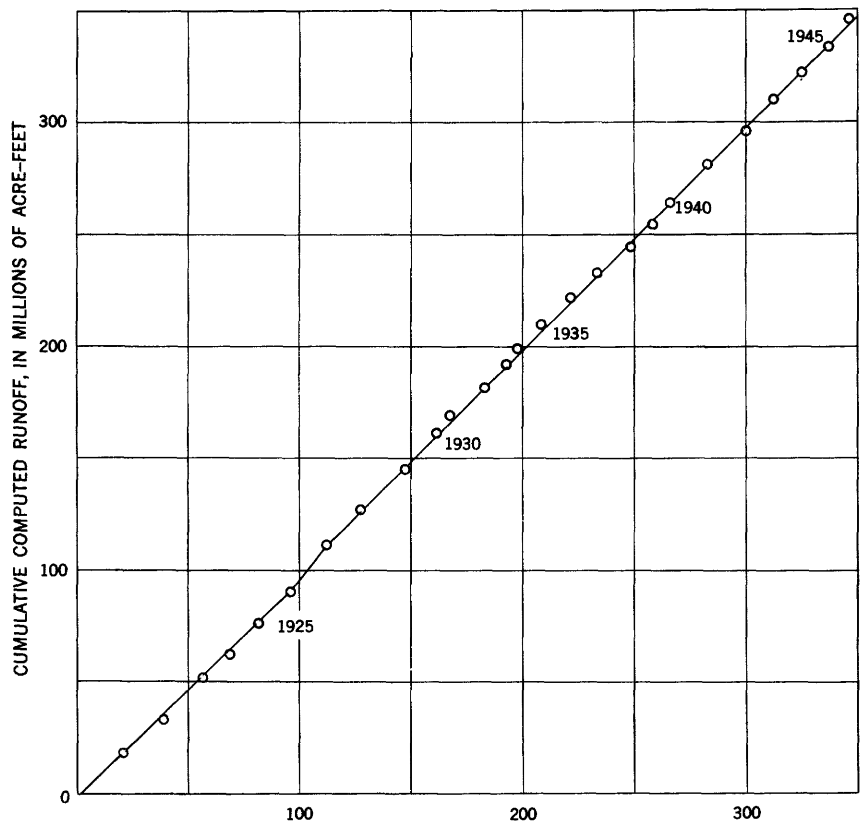

FIGURE 5.—Double-mass curve of measured runoff versus computed runoff, Colorado River near Grand Canyon, Ariz., 1921–46.

TABLE 4.—*Measured annual runoff and computed annual runoff, Colorado River near Grand Canyon, Ariz.*

[Y for 1921–22 is estimated]

Water Year	Runoff (millions of acre-feet)					
	Measured		Expected		Residual	
	Y	Cumulative	Y_e	Cumulative	$Y-Y_e$	Cumulative
1921	21.27	21.27	18.29	18.29	+2.98	+2.98
1922	17.84	39.11	15.19	33.48	+2.65	+5.63
1923	17.05	56.16	16.63	50.11	+.42	+6.05
1924	13.01	69.17	11.75	61.86	+1.26	+7.31
1925	11.74	80.91	13.74	75.60	−2.00	+5.31
1926	14.42	95.33	13.96	89.56	+.46	+5.77
1927	17.26	112.59	21.60	111.16	−4.34	+1.43
1928	15.63	128.22	15.16	126.32	+.47	+1.90
1929	19.43	147.65	18.43	144.75	+1.00	+2.90
1930	13.42	161.07	16.03	160.78	−2.61	+.29
1931	6.74	167.79	8.44	169.22	−1.70	−1.41
1932	15.97	183.76	11.21	180.43	+4.76	+3.35
1933	10.01	193.77	11.13	191.56	−1.12	+2.23
1934	4.66	198.43	7.60	199.16	−2.94	−.71
1935	10.22	208.65	10.35	209.51	−.13	−.84
1936	12 32	220.97	11.50	221.01	+.82	−.02
1937	12.41	233.38	10.95	231.96	+1.46	+1.44
1938	16.53	249 01	12.75	244.71	+2.88	+4.32
1939	9.62	258.63	9.97	254.68	−.35	+3.97
1940	7.44	266.07	9.90	264 58	−2.46	+1.51
1941	16.94	283.01	16.43	281.01	+.51	+2.02
1942	17.26	300.27	14.77	295.78	+2.49	+4.51
1943	11.43	311.70	14.23	310.01	−2.80	+1.71
1944	13.53	325.23	12.06	322.07	+1.47	+3 18
1945	11.87	337.10	12.79	334 86	−.92	+2.26
1946	9.09	346.19	11.55	346.41	−2.46	−.20

When a break in the double-mass curve occurs, the statistical test for significance, described in a later section, furnishes a valuable guide to the decision as to whether a change has occurred in the precipitation-runoff relation.

RESIDUAL-MASS CURVE

Minor breaks that are hidden by the smoothing of a double-mass curve can be magnified for detailed study by using a residual-mass curve. The example of a residual-mass curve in figure 6 is obtained by subtracting the computed runoff from the measured runoff (table 4, col. 6) and cumulating the residuals (table 4, col. 7) and plotting them against the year of occurrence.

The residual-mass curve has the following characteristics:

1. Changes in the relation between two variables are magnified.
2. A break in the double-mass curve corresponds to a maximum or a minimum point on the residual-mass curve.

Figure 6 reveals many minor breaks in the double-mass curve but, in general, the residuals increase during the period 1921–26 and fluctuate without a particular trend after 1927. This conclusion corresponds with the two lines on the double-mass curve and supports the conclusion that there has been no major change in the relation between rainfall and runoff during the period 1921–46.

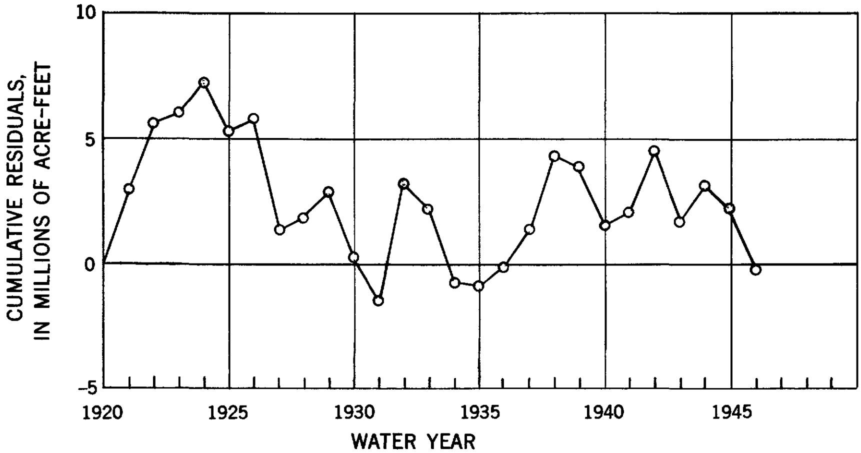

FIGURE 6.—Residual-mass curve of measured runoff minus computed runoff, Colorado River near Grand Canyon, Ariz.

STATISTICAL TEST FOR SIGNIFICANCE

A statistical test of significance of an apparent break in slope of a double-mass curve is usually the last step of a double-mass-curve analysis. Breaks in the double-mass curves that are due to changes in method of computation or to known diversion of water, not returned to the stream, will have been removed so that the remaining breaks are due either to chance or to a change in the relation, but we do not know which. Statistical methods are used to help answer this question.

The significance of an apparent break in the double-mass curve can be determined by an analysis-of-variance test (covariance in this example). Variance sometimes called the mean square, is a measure of the scatter of individual values about the mean of the values and is computed as the sum of the squares of the differences between the mean and the individual values divided by one less than the number of values. Covariance is a mean of the cross product of the deviations in the same way as variance is a mean square, and is used in computing the variance about a line of regression. Details of the analysis-of-variance method are found in textbooks on statistics, and only the application of the F test (variance-ratio test) to the data in table 2 is discussed here. For the purpose of this example it is assumed that the difference in slope that occurred in 1938 cannot be removed by recomputing the records or by adjusting for known diversions. A graphical method for evaluating the significance of breaks in the double-mass curve is given by Weiss and Wilson (1953).

Since the double-mass curve between stream A and the pattern seems to indicate that a change in the relation between that stream and the pattern occurred in 1938, an analysis of variance test is applied to the data for the periods 1921–38 and 1939–45. The statistics used in this test is F, the ratio of the among-periods variance to the within-periods variance. The larger the apparent shift in the relation the larger this ratio becomes.

Most collections of statistical tables contain a table of the F distribution to use in testing for equality of variance. The test consists of computing F, the variance ratio, from the data being tested and comparing the computed value with the tabulated value of the F distribution for the level of significance selected (generally 5 percent). The chances of the computed value of F equaling the tabular value at the 5-percent point by chance alone are 1 in 20. When, as in this example, the computed value of F exceeds the tabular value for the 5-percent level, the probability of obtaining such a result by chance alone becomes less and the result is significant at the 5-percent level (sometimes stated as the 95-percent level). As applied to the data in table 2, a computed value of F that exceeds the value tabulated for the 5-percent level would support the presumption that the break in the double-mass curve is caused by a change in the relation between station A and the pattern and not by the vagaries of sampling. The test does not prove that the relation is different in the two periods, but only that a break this great would be unlikely to occur were there really no difference in the periods.

The data in table 2 are analyzed by the variance-ratio test (F test) in the following example in which these symbols are used:

N=25 years—total period (1921–45).
N_1=18 years—period 1 (1921–38).
N_2=7 years—period 2 (1939–45).
X=An annual runoff for the pattern.
X_1=An annual runoff during period 1 for the pattern.
X_2=An annual runoff during period 2 for the pattern.
Y=An annual runoff for station A.
Y_1=An annual runoff during period 1 for station A.
Y_2=An annual runoff during period 2 for station A.
Σ=Total for period.
x, x_1, x_2=Deviations—difference between an X value and the period mean.
y, y_1, y_2=Deviations—difference between a Y value and the period mean.
$\frac{\Sigma x^2}{N-1}$ or $\frac{\Sigma y^2}{N-1}$=Variance of X or Y.
$\frac{\Sigma xy}{(d.f.)}$=Covariance of X and Y.

d=Residual—the vertical (Y) difference between the line through the plotted points and the plotted point.

$d.f.$=Degrees of freedom. In general, the degrees of freedom equal the number of items minus the number of constraints imposed on the system. Degrees of freedom for this example are discussed in the explanation of table 5.

In this example, shortcut statistical methods are used in the special application of the F-test to a specific problem—that of determining the significance of a break in the slope of a double-mass curve. The shortcut methods obscure what is being done but gives an example that can be followed easily as a model when data of the same type are to be tested.

The annual runoff data in table 2 are prepared for the F-test as follows:

Step 1

Add the X and Y values for period 1 and period 2 and for the entire period.

$$\Sigma X_1 = 261.58 \qquad \Sigma Y_1 = 332.71$$
$$\Sigma X_2 = 142.06 \qquad \Sigma Y_2 = 132.32$$
$$\Sigma X = 403.64 \qquad \Sigma Y = 465.03$$

Step 2

Square the individual X and Y values and obtain the sum of squares by period and for the entire period.

$$\Sigma X_1^2 = 4{,}711.9 \qquad \Sigma Y_1^2 = 7{,}132.6$$
$$\Sigma X_2^2 = 3{,}698.8 \qquad \Sigma Y_2^2 = 3{,}001.2$$
$$\Sigma X^2 = 8{,}410.7 \qquad \Sigma Y^2 = 10{,}133.8$$

Step 3

Square the total for each period and divide by the number of years in the period.

$$(\Sigma X_1)^2 \div N_1 = 3{,}801.3 \qquad (\Sigma Y_1)^2 \div N_1 = 6{,}149.8$$
$$(\Sigma X_2)^2 \div N_2 = 2{,}883.0 \qquad (\Sigma Y_2)^2 \div N_2 = 2{,}501.2$$
$$(\Sigma X)^2 \div N = 6{,}517.0 \qquad (\Sigma Y)^2 \div N = 8{,}650.1$$

Step 4

Multiply the X and corresponding Y values and sum the products for each period, and for the entire period.

$$\Sigma X_1 Y_1 = 5{,}710.8$$
$$\Sigma X_1 Y_2 = 3{,}243.1$$
$$\Sigma XY = 8{,}953.9$$

Step 5

Multiply the total X for each period by the total Y for that period and divide by the number of years in the period.

$$\Sigma X_1 \Sigma Y_1 \div N_1 = 4{,}835.0$$
$$\Sigma X_2 \Sigma Y_2 \div N_2 = 2{,}685.3$$
$$\Sigma X \Sigma Y \div N = 7{,}508.2$$

The analysis of variance and covariance is summarized in table 5.

TABLE 5.—*Analysis of variance and covariance of data in table 2.*

	1	2	3	4	5	6	7
	Degrees of freedom	Sum of x^2	Sum of y^2	Sum of xy	Degrees of freedom	Sum of d^2	Mean d^2
(a) Among periods	1	167.3	0.9	12.1	$N_1=1$	87.7	87.7
(b) Within periods	23	1,726.4	1,482.8	1,433.6	$N_2=22$	292.3	13.3
(c) Total	24	1,893.7	1,483.7	1,445.7	23	380.0	

$F=\frac{87.7}{13.3}=6.59$; from table for $N_1=1$ and $N_2=22$, $F=4.30$ at 5 percent level and 7.94 at 1-percent level.

The items in table 5 are computed as follows:

Column 1 (degrees of freedom)

(a) Among periods:

The degrees of freedom among periods = the number of periods (2) minus 1 = 1.

(b) Within periods:

The degrees of freedom within periods = the number of items (25) minus 2 (a mean for period 1 and a mean for period 2 has been computed) = 23.

(c) Total:

The degrees of freedom for the total equal the number of items (25) minus 1 (lost through computing a mean for the total period) = 24.

Column 2 (sum of x^2)

(a) Among periods:

$$\sum x^2=\frac{(\sum X_1)^2}{N_1}+\frac{(\sum X_2)^2}{N_2}-\frac{(\sum X)^2}{N}$$

From step (3):

$$3{,}801.3+2{,}883.0-6{,}517.0=167.3$$

(b) Within periods:

$$\sum x^2=\sum X^2-\left[\frac{(\sum X_1)^2}{N_1}+\frac{(\sum X_2)^2}{N_2}\right]$$

From steps (2) and (3):

$$8{,}410.7-(3{,}801.3+2{,}883.0)=1{,}726.4$$

(c) Total:

$$\sum x^2 = \sum X^2 - \frac{(\sum X)^2}{N}$$

From steps (2) and (3):

8,410.7−6,517.0=1,893.7

Column 3 (sum of y^2)

(a) Among periods:

$$\sum y^2 = \frac{(\sum Y_1)^2}{N_1} + \frac{(\sum Y_2)^2}{N_2} - \frac{(\sum Y)^2}{N}$$

From step (3):

6,149.8+2,501.2−8,650.1=0.9

(b) Within periods:

$$\sum y^2 = \sum Y^2 - \left[\frac{(\sum Y_1)^2}{N_1} + \frac{(\sum Y_2)^2}{N_2}\right]$$

From steps (2) and (3):

10,133.8−(6,149.8+2,501.2)=1,482.8

(c) Total:

$$\sum y^2 = \sum Y^2 - \frac{(\sum Y)^2}{N}$$

From steps (2) and (3):

10,133.8−8,650.1=1,483.7

Column 4 (sum of xy)

(a) Among periods:

$$\sum xy = \frac{\sum X_1 \sum Y_1}{N_1} + \frac{\sum X_2 \sum Y_2}{N_2} - \frac{\sum X \sum Y}{N}$$

From step (5):

4,835.0+2,685.3−7,508.2=12.1

(b) Within periods:

$$\sum xy=\sum XY-\left(\frac{\sum X_1\sum Y_1}{N_1}+\frac{\sum X_2\sum Y_2}{N_2}\right)$$

From steps (4) and (5):

$$8{,}953.9-(4{,}835.0+2{,}685.3)=1{,}433.6$$

(c) Total:

$$\sum xy=\sum XY-\frac{\sum X\sum Y}{N}$$

From steps (4) and (5):

$$8{,}953.9-7{,}508.2=1{,}445.7$$

The analysis of variance and covariance in table 5 (columns 5–7) compares the mean of the sum of the squares of the residuals (sum of d^2) from the regression line through all the items for the total period with the mean of the sum of the squares of the residuals from separate but parallel regression lines through the points for each part of the period.

Column 5 (degrees of freedom)

One additional degree of freedom is lost from (b) Within periods, and from (c) Total, by reason of computing a slope of the regression line.

Column 6 (sum of d^2)

The sum of squares of the residuals is computed by the formula:

$$\sum d^2=\sum y^2-\frac{(\sum xy)^2}{\sum x^2}$$

(c) Total:

$$\sum d_t^2=1{,}483.7-\frac{(1{,}445.7)^2}{1{,}893.7}=380.0$$

The total dispersion of the Y values about the means for the total period minus the variance due to the regression is computed from the formula given for Σd^2 using the values in columns 2–4, table 5.

(b) Within periods:

$$\sum d_w^2=1{,}482.8-\frac{(1{,}433.6)^2}{1{,}726.4}=292.3$$

The sum of squares of the residuals about a regression line through the dispersion of the points about their period means is computed from the formula given for Σd^2 using the values in columns 2–4, table 5.

(a) Among periods:

The reduction in the sum of squares of the residuals between the "Total" and the "Within periods" line is attributable to the dispersion of the means, or $380.0-292.3=87.7$.

F (6.59) is computed by dividing the among-periods mean d^2 (87.7), by the within-periods mean d^2 (13.3).

Interpretation of the test.—The computed value of F (6.59) exceeds the tabular value of F (4.30) at the 5-percent level. Therefore, the break in the double-mass curve is significant at the 5-percent level. Thus, from a purely statistical evaluation, the break in the curve between 1938 and 1939 can be attributed to a change in the relation between station A and the pattern.

EFFECT OF THE DEGREE OF CORRELATION BETWEEN STATION RECORDS

Poor correlation between the station being tested and the pattern can obscure the detection of significant differences between periods. To illustrate this point, the data in table 2 have been changed to decrease the correlation between station A and the pattern without changing the means for the two periods. The changed data are listed in table 6.

TABLE 6.—*Annual runoff, in inches, for double-mass curve with poor correlation between the station and the pattern*

Water year	Stream A Y	Pattern X	Water year	Stream A Y	Pattern X
1921	21.73	17.61	1934	7.35	9.94
1922	13.80	14.29	1935	34.75	23.58
1923	19.52	6.12	1936	5.75	6.06
1924	14.58	16.39	1937	21.72	11.76
1925	7.33	1.53	1938	26.33	30.64
1926	14.45	15.80	1939	17.04	8.41
1927	32.67	22.03	1940	11.65	12.68
1928	19.22	14.40	1941	19.42	28.15
1929	23.96	17.70	1942	15.82	23.60
1930	17.34	20.10	1943	11.41	6.96
1931	11.87	3.13	1944	19.13	22.01
1932	22.81	20.30	1945	37.85	40.25
1933	17.53	10.20			

The analysis of variance and covariance in table 7 was computed in the manner described for table 5 in the preceding section. It will be noted that the computed F (2.33) is less than the tabular value (4.30) at the 5-percent level. Thus the F test indicates that the break is not significant. As shown in the following section, the effect

TABLE 7.—*Analysis of variance and covariance of data in table 6*

	Degree of freedom	Sum of x^2	Sum of y^2	Sum of xy	Degree of freedom	Sum of d^2	Mean d^2
Among periods	1	167.3	0.9	12.2	$N_1=1$	65.8	65.8
Within periods	23	1,826.4	1,589.7	1,330.0	$N_2=22$	621.2	28.2
Total	24	1,993.7	1,590.6	1,342.2	23	687.0	----------

$F=\frac{65.8}{28.2}=2.33$; from table for $N_1=1$ and $N_2=22$, $F=4.30$ at 5-percent level and 7.94 at 1-percent level.

of poor correlation can be overcome by extending the length of record.

EFFECT OF THE LENGTH OF RECORD

An increase in the length of record tends to offset the effect of variability of data. The effect of increasing the length of record can be shown by adding to the second period two 7-year periods, both identical with the data for 1939–45 (table 6). The analysis of the data with the 14 added years is given in table 8.

TABLE 8.—*Analysis of variance and covariance of data in table 6, after adding 14 years of record*

	Degrees of freedom	Sum of x^2	Sum of y^2	Sum of xy	Degrees of freedom	Sum of d^2	Mean d^2
Among periods	1	321.7	3.9	24.4	$N_1=1$	112.7	112.7
Within periods	37	3,528.1	2,551.8	2,429.3	$N_2=36$	879.1	24.4
Total	38	3,849.8	2,555.7	2,453.7	37	991.8	----------

$F=\frac{112.7}{24.4}=4.62$; from table for $N_1=1$ and $N_2=36$, $F=4.11$ at 5-percent level and 7.39 at 1-percent level.

The computed F (4.62) exceeds the tabular value (4.11) at the 5-percent level. Thus we can state that the break in the double-mass curve is significant.

By comparing the result of the analysis of the variable data (table 7) with that of the less variable data (table 5), it is evident that conclusions drawn from short periods of high variability are more likely to be in error than conclusions based on less variable data for the same length of period. With data of high variability (table 7) the length of period must be greater (table 8) to arrive at accurate conclusions.

SUMMARY

Although the double-mass curve is a valuable tool in hydrologic studies, a good understanding of its theory and application to various types of hydrologic problems is necessary if misuse is to be avoided. What may be a good practice when working with precipitation records may be a bad practice when working with streamflow records or precipitation-runoff relations. Furthermore, if the variables do not bear a fairly constant ratio to each other throughout the range of data, the double-mass curve should be based on observed data and data computed from a curve of relation rather than on a cumulation of the variables themselves.

For precipitation records, the double-mass curve can be used to detect changes in the consistency of precipitation records and to determine the amount of adjustment to be applied to make them consistent. Precautions should be taken to see that the record is not being tested against records that are in themselves inconsistent. A plausible cause for any apparent inconsistency should be found before a record is said to be inconsistent. If the cause of the inconsistency is in doubt, adjustments should not be made unless the inconsistency is statistically significant.

For streamflow records, the double-mass curve can be used to detect possible changes in consistency but should not be used to determine the amount of adjustment. The date of the possible change in consistency is a clue to possible changes in method of collecting the records or to possible diversions from or to the basin. If the change is due to a change in methods the records should be recomputed from the adjusted basic data, and if the change is due to diversions, adjustments for diversion should be applied. Streamflow records should seldom be adjusted by the application of a correction factor.

For precipitation-runoff relations, the precipitation records being used should first be tested for consistency by the double-mass curve method and the runoff records should be examined for changes in location and method. The records should be made as consistent as possible, and then be plotted against each other to define a relation. The double-mass curve should be based on observed and computed values of runoff, and should be analysed statistically before a break in slope is said to be significant.

FITTING CURVES TO CYCLIC DATA

By W. B. Langbein

A common problem in hydrology is to fit a smooth curve to cyclic or periodic data. This section describes two methods that are applicable to those problems where the length or period of the cycle is known beforehand—as, for example, a day, a year, or a meander length. Curve-fitting by freehand drawing offers a simple and direct course where the data are closely aligned and a fit is clearly evident.

In many problems, however, a satisfactory fit is far from obvious, and analytical methods may be necessary. Among the analytical methods available for curve fitting are:

1. Moving arithmetic average.
2. Moving arc.
3. Fitting polynomial by least squares.
4. Double integration.
5. Fitting Fourier (sine and cosine) series.

Of these methods, only the moving arithmetic average is in common use in hydrologic practice. The moving arithmetic average method will greatly attenuate and modify the curve of data that conform, for example, to a simple sine curve. In principle, the moving average fits a series of straight lines to a set of data. If the graph of the data is a curve, a series of straight lines does a poor job of fitting the plotted points. Fitting curves by the polynomial and by the Fourier series methods is exceedingly complex and, in view of the inaccuracies inherent in hydrologic data, it is not practical for most hydrologic purposes. For these reasons the moving arc and the double integration methods deserve consideration in fitting curves to hydrologic data. Furthermore, these methods are theoretically sound and eminently practical.

MOVING ARC

PARABOLIC ARCS

Moving arcs, which successfully follow a curve, are described in the actuarial literature (for example, Whittaker and Robinson (1932), and Sasuly (1934)). Fitting moving parabolic arcs seems simplest. The respective weights for parabolic arcs of various lengths are given in table 9.

TABLE 9.—*Weights for n-point least-square parabolas*

[Weights are symmetrical about central values in boldface. Whittaker and Robinson (1932, p. 295)]

Number of points, n								
5	7	9	11	13	15	17	19	21
−3	−2	−21	−36	−11	−78	−21	−136	−171
12	3	14	9	0	−13	−6	−51	−76
17	6	39	44	9	42	7	24	9
12	**7**	54	69	16	87	18	89	84
−3	6	**59**	84	21	122	27	144	149
	3	54	**89**	24	147	34	189	204
	−2	39	84	**25**	162	39	224	249
		14	69	24	**167**	42	249	284
		−21	44	21	162	**43**	264	309
			9	16	147	42	**269**	324
			−36	9	122	39	264	**329**
				0	87	34	249	324
				−11	42	27	224	309
					etc.	etc.	etc.	etc.
35	21	231	429	143	1,105	323	2,261	3,059

To illustrate how the above weights can follow a curve, consider the following values from a sine curve:

x	y
1	0
2	.500
3	.866
4	1.000
5	.866
6	.500
7	0

A straight arithmetical average of 5 points centered on the 4th item above, gives a value of 0.75 compared with 1.000 for the known points. Applying the 5-point parabolic weights gives 0.995 as a result.

Table 10 illustrates the application of a 5-point moving parabolic arc to the residual monthly deviations from a correlation of the flows of two streams in Utah. In this problem, the smoothed values of the monthly deviations are probably superior to the deviations as originally computed, because logically there should be a uniform variation in these deviations among the months. It is desired, however, not to destroy the intrinsic character of the variations.

The procedure is as follows: The monthly average deviations as given are listed under y in table 10 and plotted on figure 7. Each

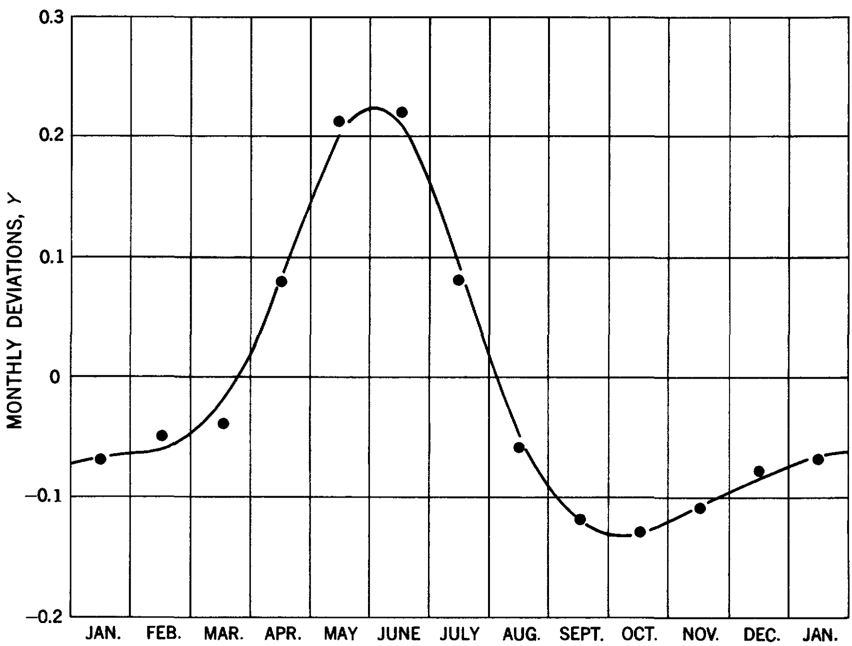

FIGURE 7.—Illustration of smoothing by 5-point moving arc.

TABLE 10.—*Illustration of use of 5-point moving parabolic arc*

Month	y	Product of y times indicated weight			5-point total	5-point average	Month
		−3	+12	+17			
January	−0.07	+0.21	−0.84	−1.19			
February	−.05	[+.15]	−.60	−.85			
March	−.04	+.12	[−.48]	−.68	−0.74	−0.021	March
April	+.08	−.24	+.96	[+1.36]	+2.89	+.083	April
May	+.21	−.63	[+2.52]	+3.57	+7.05	+.20	May
June	+.22	[−.66]	+2.64	+3.74	+7.16	+.205	June
July	+.08	−.24	+.96	+1.36	+3.01	+.086	July
August	−.06	+.18	−.72	−1.02	−1.77	−.050	August
September	−.12	+.36	−1.44	−2.04	−4.23	−.12	September
October	−.13	+.39	−1.56	−2.21	−4.55	−.13	October
November	−.11	+.33	−1.32	−1.87	−3.82	−.109	November
December	−.08	+.24	−.96	−1.36	−2.98	−.085	December
January	−.07	+.21	−.84	−1.19	−2.30	−.066	January
February	−.05	+.15	−.60	−.85	−2.17	−.062	February
March	−.04	+.12	−.48	−.68			

value of y is then multiplied successively by each weight; the value for January, −0.07, is multiplied by −3, by 12, and by 17. To obtain the 5-point total for April, for example, the weighted values for February, March, April, May, and June (bracketed figures in table 10) are totaled: $+0.15-0.48+1.36+2.52-0.66=+2.89$. The 5-point average is found by dividing +2.89 by the algebraic sum of the weights, 35, to yield +0.083.

In using the moving arc, how many points are to be used should be decided. A short arc will provide close local fits, but will provide little averaging. On the other hand, long arcs may tend to alter the intrinsic character of the cycle. A useful compromise would be to use at least one-fourth but no more than one-half the number of points available for defining the cycle. An odd number of items in the arc is advised because it provides a definite central value. To provide sufficient averaging of random components, at least 5 points should be used. This requirement in turn, necessitates at least 10 points in the cycle for use of a parabolic arc.

QUARTIC ARCS

If there are fewer than 10 points in the cycle, the 5-point parabola will not make a satisfactory fit. In this situation, a 7- or 9-point quartic arc should be used with the weights given in table 11. The quartic arcs could be used in place of the parabolic arcs for all problems, although this entails more work.

The arithmetic work can be simplified by rounding off the central item in table 11 by subtracting or adding one as required. The weights are then divisible by 5. For example, approximate weights for a 7-point quartic arc are 1, −6, 15, 26, 15, −6, and 1. Their sum is 46.

TABLE 11.—*Weights for least square quartic arcs*

7 point	9 point	11 point
+5	+15	+18
−30	−55	−45
+75	+30	−10
+131	+135	+60
+75	**+179**	+120
−30	+135	**+143**
+5	+30	+120
	−55	+60
	+15	−10
		−45
		+18
+231	+429	+429

DOUBLE INTEGRATION

Double integration can be used where the cyclic data approximate the simple form:

$$y=a+b\sin\frac{2\pi x}{p}\pm\epsilon \tag{1}$$

in which a represents the mean, b the amplitude, and p the length of the cycle; x represents the position in the cycle, and ϵ the random component. If a, the mean, is subtracted from each of these values, and the progressive totals, or first integral, of these differences are taken, the result is a series of the form:

$$-\frac{bp}{2\pi}\cos\frac{2\pi x}{p}\pm\Sigma\epsilon \tag{2}$$

A second integration will give a series of the form:

$$-b\left(\frac{p}{2\pi}\right)^2\sin\frac{2\pi x}{p}\pm\Sigma\Sigma\epsilon \tag{3}$$

The term $\Sigma\Sigma\epsilon$ will, in general, be small, since the positive chance variations in ϵ will tend to balance the negative terms. The second integration, though opposed in sign, is in phase with the original data and substantially all random variations will have been averaged out. The second integration provides adequate smoothing and so the integrating process need go no further. Although minor cyclic components tend to be averaged out, the second integration may depart appreciably from a true sine curve if the original data so require.

The double integration method is, therefore, peculiarly adapted to those problems where a sinusoid may be considered a close approximation of the curve sought. The process is illustrated by the following

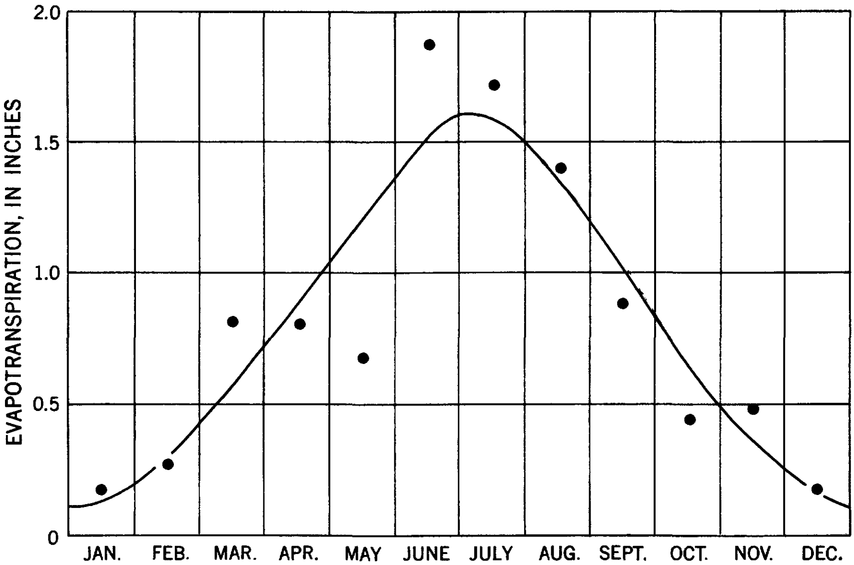

FIGURE 8.—Result of fitting a cycle graph by double integration.

TABLE 12.—*Example of double integration applied to monthly data*

[Note that entries under the 1st integration, representing sums to the end of each month, are moved down one line; thus they are on a line midway between the months. Entries under the 2d integration are again moved down one line; thus they are on a line even with the months]

Month	Computed evapotranspiration	Deviation from mean	1st integration	Deviation from mean	2d integration	Deviation from mean	Deviation −(12÷2π²)	Smoothed values
	1	2	3	4	5	6	7	8
January	0.18	−0.63						
			−0.63	−0.57				
February	.27	−.54			−0.57	+1.91	−0.52	0.29
			−1.17	−1.11				
March	.82	+.01			−1.68	+.80	−.22	.59
			−1.16	−1.10				
April	.80	−.01			−2.78	−.30	+.08	.89
			−1.17	−1.11				
May	.68	−.13			−3.89	−1.41	+.39	1.20
			−1.30	−1.24				
June	1.88	+1.07			−5.13	−2.65	+.73	1.54
			−.23	−.17				
July	1.72	+.91			−5.30	−2.82	+.77	1.58
			+.68	+.74				
August	1.40	+.59			−4.56	−2.08	+.57	1.38
			+1.27	+1.33				
September	.88	+.07			−3.22	−.74	+.20	1.01
			+1.34	+1.40				
October	.44	−.37			−1.82	+.66	−.18	.63
			+.97	+1.03				
November	.48	−.33			−.79	+1.69	−.46	.35
			+.64	+.71				
December	.18	−.63			−.08	+2.40	−.66	.15
			+.01	+.08				
January					0	+2.48	−.68	.13
Means	.81	0	−.06	0	−2.48	0	0	.81

computations. The data shown on figure 8 and in table 12 are for monthly evapotranspiration from ground water as determined from a hydrologic budget for Beaverdam Creek in eastern Maryland. Evapotranspiration follows a sinusoidal variation from month to month through the course of the year. The erratic variations in the given data represent errors that arise because the figures are the residuals between relatively large quantities.

The given values of computed evapotranspiration are listed in the first numbered column in table 12. The deviations from the mean of the data are listed in the second column. The third column shows the cumulative summation of the deviations, and completes the first integration. The process is repeated in the fourth and fifth columns, and a second integration is made as shown in the fifth column. Deviations from the mean of the second integration are shown in the sixth column. These figures are divided by the constant $-(12 \div 2\pi)^2$ in which the numerator 12 corresponds to p, the number of items in the cycle. The divisor $-(12 \div 2\pi)^2$ converts the data for equation (3) to that for equation (1). The quotients listed in column 7 represent the smoothed deviations from the mean of the data, in this example, 0.81. Adding these to the mean gives the final smoothed results, shown in column 8 and on figure 8, which represent a reasonable interpretation of the original data.

SUMMARY

The moving arc methods are quite general in application. Of the moving arc methods, the parabolic arcs have a slight advantage of simpler arithmetic. However, when parabolic weights extend beyond a reversal in curvature, quartic weights should be used instead. The double integration method is a simple means of fitting a curve when the cycle is of a sinusoidal form.

LITERATURE CITED

Guy, H. P., 1957, The trend of suspended-sediment discharge of the Brandywine Creek at Wilmington, Del., 1947–1955: U.S. Geol. Survey open-file report, p. 14, fig. 3.

Horton, R. E., 1924, Determining the mean precipitation of a drainage basin: New England Water Works Assoc., v. 38, p. 1–47.

Merriam, C. F., 1937, A comprehensive study of the rainfall on the Susquehanna valley: Am. Geophys. Union Trans., pt. 2, p. 471–476.

Oltman, R. E., and Tracy, H. J., 1951, Trends in climate and in precipitation-runoff relation in Missouri River basin: U.S. Geol. Survey Circ. 98, 113 p.

Paulhus, J. L. H., and Kohler, M. A., 1952, Interpolation of missing precipitation records: U.S. Weather Bur. Monthly Weather Rev., v. 80, p. 129–133.

Sasuly, Max, 1934, Trend analysis of statistics: Washington D.C., Brookings Institute.

Weiss, L. L., and Wilson, W. T., 1953, Evaluation of significance of slope changes in double-mass curves: Am. Geophys. Union Trans., v. 33, p. 893–896.

Whittaker, E. T., and Robinson, G., 1932, The calculus of observations: London, Blackie & Son.

Williams, G. R., and others, 1940, Natural water loss in selected drainage basins: U.S. Geol. Survey Water-Supply Paper 846, p. 6–11.