Meteorological Office

Climatological Services (M.O.3.)

Hydrological Memoranda, No. 5.

Estimation of standard-period averages for stations with incomplete data

1. Introduction

1.1. This memorandum is concerned entirely with methods which have been used, or are being developed, in the estimation of rainfall averages. It is not unlikely that somewhat similar methods could be applied in connection with other elements.

1.2. During 1958 work was completed on the primary survey of rainfall averages over Great Britain and Northern Ireland for the new standard period 1916-1950. Monthly and annual averages had been obtained for three classes of stations:-

1.2.1. All stations with complete observations during the 35-year period.

1.2.2. A number of stations with nearly complete observations for which the amount of estimation required was small, so that results could not have been seriously affected even with a method of estimation open to criticism or doubt. (It is convenient to include in this class pairs of similar neighbouring stations for which in any pair neither station had complete observations, whilst the pair together covered the whole period with a substantial overlap; some overlapping stations occurring in threes were also merged.)

1.2.3. A very limited number of stations with incomplete observations for which averages were specially required for various purposes (for example, for use in the "Monthly Weather Report"), even though the degree of estimation required might have introduced the possibility of appreciable error. (For this class of stations the method of estimating standard-period averages was similar to that described in section 2 below. It was not considered very satisfactory, but it yielded quick results which could be checked by a more satisfactory method later. This process has now begun; see apendices, paragraph 6.5.6.).

1.3. Averages for more than 1,650 stations were thus obtained and these were subjected to various tests, as a result of which the averages for nearly 400 stations were rejected as unsatisfactory. These tests depended in the main on largely subjective judgements applied to the data plotted on maps (seventeen in all). Wherever possible efforts were made to find a verifiable physical reason for a subjective conclusion that any set of data was unsatisfactory, but it is nevertheless hoped in future work to improve the objective basis of the tests used in primary surveys. It is not the purpose of the present memorandum to discuss this matter, and it is here taken as an assumption that the primary survey yielded completely satisfactory monthly and annual averages for rather more than 1,250 stations, very unevenly distributed over the U.K. The majority of these have been made available already 1,2.

1.4. For drawing maps of average rainfall over all areas in the U.K., with the detail which is required for many important purposes, the 1,250 stations of the primary survey are not adequate. It was therefore necessary to embark on a secondary survey covering the country in greater detail. There are several possible ways of doing this. One would be to investigate in each district the relationship of average rainfall to altitude, exposure, aspect and distance from the sea, and to complete the maps of average rainfall purely on the basis of physical maps. Studies of this kind in various parts of the world have been published 3,4., but though some tentative investigations of the technique applied to districts in Great Britain have been begun, it has not yet been developed far enough for it to be used with confidence everywhere in the U.K. The general application of such work, if it can be successfully accomplished, remains as a task for the future. However successful it may eventually be, there is the disadvantage that this technique alone makes no use at all of the very large amount of information available from stations with incomplete observations.

1.5. The procedure adopted for the secondary survey was that used for the preparation of detailed maps for the rainfall averages of the previous standard period 1881-1915. This was simply to estimate standard period averages for stations with incomplete observations during the period, to subject the estimates to tests for acceptance or rejection, and to cover the whole country in this way in as much detail as possible or necessary, according to the district densities of stations. A disadvantage of this procedure is that there remain some areas, especially upland tracts not yet exploited for water supply or hydro-electric power, where the density of average values, including all possible estimates, is still small. For such areas the average maps must be sketched by largely subjective judgment or, later, by application of the technique mentioned in paragraph 1.4.

1.6. The actual method of estimation used for the 1881-1915 rainfall averages was considered to be open to criticism, and a new method was developed. The purpose of this memorandum is to place on record an account of both mehtods, in sufficient detail to demonstrate the validity of the criticism of the first, to show how the faults are being overcome in the development of the second, and to provide outline instructions for making estimates by the latter. The eventual aim is that the principles applied will be completely objective, so that very similar if not identical results should be obtained by completely independent workers; any piece of work should be capable of being checked without the need for harmonizing subjective judgements or highly developed flair in different individuals. When the eventual aim has been achieved it should be possible to carry out the work by a suitable programme with an electronic computer. In the meantime the object has been to build up the skeleton framework of the final method in such a way that subjective elements which remain, during the early stages of development, do not form an essential part of the system and can be progressively eliminated.

1.7. The following discussion is confined to annual averages. There is no reason apart from lack of time why the same methods should not be applied to averages for the twelve individual months, or to any other set of divisions of the year. It is hoped that when the second method has been developed into a programme for an electronic computer, the estimation of monthly and possibly weekly or five-day averages will be carried out in the same way. Meanwhile, the present method of estimating monthly averages, beginning with satisfactorily checked annual averages, has been dealt with in an earlier memorandum 5.

2. Estimation of rainfall averages, 1881-1915

2.1. For the earlier standard period the work was carried out on a county basis using river basins or parts of river basins as subsidiary divisions within the county boundaries. The system followed the arrangement and grouping of rainfall stations used in the General Table of "British Rainfall" from the time of H.R. Mill to the present day. The areal units are very diverse in size, shape and character. Some information about the methods followed is given in articles in "British Rainfall", for instance in the volumes for 1928 and 1933 6,7.

2.2. In each area a selection was made from the rainfall stations which had complete observations during the 35-year period, with monthly and annual averages tested and accepted as satisfactory. These so-called "ratio stations" normally numbered from 3 to 5 of any area, and in some districts it was necessary to go outside the boundary of an area to make up the number.

G.18695/CJH/9/63/150.

2.3. For a station with observations covering only a part of the 35-year period (henceforth referred to as a short-period station) the procedure was as follows:-

2.3.1. The average annual rainfall for all the years of observations available, y_1 to y_2, was calculated, giving a value r.

2.3.2. For each of the ratio stations the annual rainfall for each year in the same period of years, y_1 to y_2, was expressed as a percentage of the standard-period average for the station.

2.3.3. The percentage rainfall values for all the ratio stations were meaned for the area year by year, and for the entire period, y_1 to y_2, giving a final mean areal value p.

2.3.4. The estimated standard-period average for the short-period station was taken to be:-

$$R = 100\ r/p.$$

2.4. An actual numerical example is given in the appendices, paragraph 6.1. In this case (Leicestershire, River Division 1) only two ratio stations were in fact used. It will be seen that in a number of individual years the percentage values for the two ratio stations differed by considerable amounts, leading to some doubt about the validity of the mean percentages which were applied, irrespective of relative locations, to the data for the short-period stations. Examples of estimates for three short-period stations are included in paragraph 6.1.

2.5. As time went on the method was applied increasingly to stations with observations made largely or wholly outside the period 1881-1915. (See examples S_1, S_2, S_3 in paragraph 6.1.). Some stations had observations extending for many years beyond the period, but the convenient term "short-period station" has been retained to mean any station with incomplete observations during the standard period, no matter how long the observations extended beyond it. Many of the original ratio stations lapsed and had to be replaced by new ones. These, increasingly, were themselves stations with estimated and not actual standard-period averages. Eventually many of the ratio stations had observations (on which their estimated averages were based) extending more and more outside the standard period.

2.6. The weaknesses of this method were:-

2.6.1. It did not use all the information available for a particular areal unit, and for the areas surrounding it, but used only a limited subjective selection of the "best" averages in the area.

2.6.2. In grouping together the percentage values for the ratio stations, for the sake of speedy working, it took no account of the possible variation of percentage annual rainfall over the area, no matter what the size, shape and character of the area.

2.6.3. As a result of these two weaknesses combined, it was possible for discontinuities to occur in the distribution of estimated average rainfall values near the boundary between two areas, and such discontinuities were in fact known. In drawing the maps over such boundaries the discontinuities were subjectively smoothed.

2.6.4. In grouping together all the years of observations for a short-period station, instead of investigating them separately, again for the sake of speed, the method yielded only a single estimate of the long-period average, with no means by which an

G.18695/CJH/9/63/150.

estimate of standard error could be readily obtained, and no means of readily applying any objective test to the estimate. One of the main tests which could in fact be applied was: "Does it fit the map?" - the map of average rainfall being drawn. Thus, although revisions of the original map were actually made from time to time, in some districts, the test led to extreme reluctance to embark on anything more than trivial modification of detail until very emphatic evidence in favour of revision had accumulated. The method was calculated to encourage over-cautious conservatism.

2.6.5. Because of the lapsing of original ratio stations and the gradual but steadily increasing substitution of others, some of which had only estimated and not actual standard-period averages, there was the possibility of a slow drift away from the original 1881-1915 standard, with no means of checking the magnitude and direction of any such drift in any district. This was another reason for distrusting revision of the original average map, especially in a district where, rainfall stations having formerly been sparse, there had been a slow "invasion" by new stations linking the area with neighbouring districts, for which there had been more information covering the standard period.

2.7. Some refinements of the method were sometimes attempted, especially where the straightforward application of the quick method yielded an estimate which did not appear to be acceptable. This of course meant a break-down of techniques which had been applied for the sake of speed, since the piecemeal investigation of alternatives was time-consuming in an unorganized and unpredictable way. An estimate of the average for the short-period station could be made by using in turn each of the ratio stations, or more usually subsidiary groups of them, instead of grouping them all together, giving greater weight in the final estimate to the values obtained with the nearer ratio stations. This procedure has much to recommend it, especially where the area covered by the group of ratio stations is long and narrow. Secondly, an estimate of the average for the short-period station could be made for each year of observations taken separately, or more usually for subsidiary groups of years, instead of grouping them all together, and examining the variation of the estimates for successive years or subsidiary periods. Experience with these refinements showed some interesting features, which led to the idea of consistently applying both refinements simultaneously to all short-period stations, and hence, eventually, to the development of the method now being used for the 1916-1960 averages.

3. Estimation of rainfall averages, 1916-1950

3.1. The sole disadvantage of the current method of estimating rainfall averages for short-period stations, compared with the former method, is that it takes much longer. The disadvantage is unfortumately emphasized by the fact that a great part of the additional time is required in the early stages of the work before the production of even a single result can be attempted. It is probable that the difficulty will eventually be overcome by the introduction of a programme for an electronic computer. In the meantime, however, it was necessary to consider very carefully whether the advantages of the method are great enough to justify the extra effort. It was decided that they are, and that the supreme advantage is that once the basic work has been done, both the subsequent estimation of standard-period averages for additional short-period stations, and the revision of earlier estimates with the addition of later data, are very simple and are, automatically, fully consistent with all previous work. There is no prospect of the unsystematic expenditure of extra time similar to that which developed increasingly with the former method.

3.2. The basic work is the preparation of a set of maps. The choice of the scale, the division of the U.K. into suitable areas, and the character of the outline base maps are discussed in the appendices, paragraph 6.2. A specimen outline map, scale 1:625,000 (approximately 1/10 inch to the mile), for the Wear and Tees River Board area, is given as Figure 1. Overlapping of adjacent maps near the boundaries of the chosen areas introduces some duplication of effort in plotting the maps and drawing the isopleths of rainfall distribution on them, but there is no duplication necessary in the tabulation of data and in the arithmetical work, together forming the main part of the effort. As a result of the overlapping, discontinuities in the distribution of estimated averages do not arise.

3.3. The procedure is illustrated by actual examples, taken from the Wear and Tees River Board area, in the appendices, paragraph 6.3, and is as follows:-

3.3.1. An index map is first prepared showing the location of all stations with actual standard-period averages which have been accepted as satisfactory. For each area stations lying beyond the land boundaries in adjacent areas are included.

3.3.2. A second index map is prepared showing the locations of all short-period stations, and all stations for which the standard-period averages have been rejected as unsatisfactory. The latter are treated as short-period stations to see whether any part of the period of observations will yield an acceptable short-period estimate. It may be convenient to have the second index map prepared on a tracing linen base.

3.3.3. Using the first index map for reference a map is plotted and drawn for each year of the period 1916-1950, and the series is continued into the 1951-60 decade, in the earliest work until 1956. These maps show the annual distributions of rainfall expressed in percentage relation to the 1916-1950 station averages. Examples of tabulated percentages are given in paragraph 6.3. and two examples of the maps for the Wear and Tees River Board area are given as Figures 2 and 3. The drawing of the percentage isopleths is of course largely subjective, but it is hoped that eventually an objective method of estimation suitable for use with an electronic computer will be substituted. Meanwhile, an assessment of the errors arising from subjective judgments at this stage (and subsequent stages) is included, later, in the assessment of total errors, by objective determination of the standard errors of the estimated averages.

3.3.4. For each short-period station estimated percentages are read from the maps for all years of observation covered by the short-period station. It is for this purpose that the second index map on tracing linen (paragraph 3.3.2.) may be convenient. A useful alternative is to prepare a scale on perspex showing a single 100 kilometre square of the national(or Irish) grid, and to mark on it in turn the position of each short-period station with respect to the appropriate major square of the grid. In practice it is quickest to work with small groups of short-period stations(up to five). If the estimated percentage for any station for any year is p and the actual rainfall corresponding to it was r, an estimated standard-period average for that station, from the information for that year alone,is

$$R = 100\ r/p$$

Such an estimate is obtained for each year of observations from the short-period station.

3.3.5. The series of estimates for the short-period station is first examined for any evidence of a steady upward or downward trend, or for any abrupt change from one level to another. Such trends or changes would be evidence of slow changes of exposure, the development of leaks in the rain-gauge or other instrumental faults, sudden changes in the nature of the site and exposure or in the condition of the equipment, a change from a poor observer to a good one or vice versa, and so on. The records are referred to for any evidence which may explain physically any suspected faults of this kind. Subjective judgment will never be eliminated entirely, for <u>all</u> stations, at this stage, withough very

G.18695/CJH/9/63/150.

detailed station histories and unlimited time to study them. But experience has already shown that with a moderate amount of information, and moderate expenditure of time, the work can be made more than 90 per cent objective, at a cautious estimate. The most that can be expected of an electronic computer at this stage, is that it will throw out certain series of estimates which require examination, possibly with a quantitative indication of trends or abrupt changes.

3.3.6. At this point some sub-period of observations for the short-period station may be eliminated as unsatisfactory, though it has been shown in practive that for the large majority of stations the whole period of observations may be, so far, accepted. There remain a series of estimates of the long-period average, fluctuating in a random manner, usually be small amounts, about a mean value. Some criterion must now be applied to see whether there are any individual years for which deviations from the mean seem unacceptably large, in order that the station records may again be referred to for physical evidence in favour of rejecting such anomalous years. The choice of this criterion is discussed in the appendices, paragraph 6.4, with reference also to the possibly controversial question of whether any anomalous value should ever be rejected without physical evidence.

3.3.7. The mean value of the remaining series of yearly estimates is now taken, and in addition the standard error of this mean value may also be calculated as one criterion of the accuracy of the final estimate.

3.4. This account does not include details of certain practical points which are helpful in carrying out the work in accordance with a scheme which is already in use. These are dealt with in the appendices, paragraph 6.5. It is implicit in the account that most of the weaknesses of the former method, which were discussed in paragraph 2.6, have been overcome. However, the method as so far described does not entirely eliminate the subjective test of: "Does the estimate fit the map?" (paragaph 2.6.4.); nor does it dispose completely of the possibility of a future drift away from the original standard, in this case 1916-1950, as discussed for the earlier period in paragraph 2.6.5. With regard to the latter topic, it will merely be stated, at present as an article of faith, that it is believed that the new method minimizes the risk of a slow drift from the standard. It is proposed to give more thought to this matter in the future, when the implications of any drift will become more serious. With regard to the former topic, this test cannot be eliminated altogether from the cartographical use of averages, but can be made more objective by the development of the method mentioned in paragraph 1.4. From systematic study of the regression of average rainfall on altitude and other physical factors, quantitatively expressed, and plotting the departures from the regression equation on maps, it should be possible to investigate the fit of individual values to the average maps on a much sounder basis than hitherto.

3.5. One interesting point which has tended to be glossed over in the past is that it is possible, on any basis of study, to have a few average values (actual or estimated) which in no conceivable way can be made to fit a credible average map: the short-comings of the stations which account for these poor values may or may not be known; yet the faults, whatever they may be, are systematic, so that the data yield apparently trustworthy percentage values, almost invariably year after year, based on quite unreliable absolute measurements. Data for such stations may be published before the situation is appreciated, and sooner or later, when this happens, there are likely to be difficulties. There are sometimes good grounds, until a better solution can be found by practical means, for continuing to use the percentage values whilst rejecting the absolute measurements. But the procedure demands full explanation and probably also requires a provisional estimate of a correction factor for the absolute amounts. For a strictly limited number of stations, for limited periods, it may then be permissible to have two rainfall averages: an estimated "real" average, and an apparent average for use with current data to obtain realistic percentage values. Such a situation existed in recent years with the over-exposed rain-gauge at Lerwick.

G.18695/CJH/9/63/150.

3.6. Finally, a theoretical weakness of the current method of estimating averages for short-period stations remains to be mentioned. When the final estimate is taken as the arithmetic mean of a series of estimates, each based on information for a single year, the result is usually slightly too large. The reason for this and the magnitude of the error are discussed in the appendices, paragraph 6.6. The discrepancy is small compared with the standard error of the final estimate, is very unlikely to be as much as one per cent, and is usually only a small fraction of one per cent. In practice it can be neglected.

4. Concluding remaks and acknowledgment.

This memorandum aims to describe the early stages in the development of a method which seems to be promising. It is not claimed that it describes the final form of a perfected method, which should be rigidly followed in all detail. At several points some possibilities of modifications and some ideas for future work are specifically suggested, and comments on these topics or additional proposals would be welcomed. It is hoped that the general outline of the scheme will prove satisfactory and that future modifications will be improvements in detail at various stages, rather than drastic revision of the fundamental basis of the method. It is hoped that the general procedure and lay-out of the work indicated in the appendices paragraphs 6.3., 6.4., and 6.5. will be retained, so that future work will be broadly consistent and in harmony with the substantial amounts of work which have already been done for a number of areas in England and Wales.

The ideas outlined in this memorandum developed during discussions, over a long period, with several members of the Meteorological Office staff engaged on rainfall work. They have also been modified to some extent as a result of experience gained during the early numerical work. The greater part of this work was under the direct supervision of Miss E. H. Rowsell, who was at first very ably assisted by Miss P.A. Cameron. Later Mr. P. Annette provided useful material with particular relevance to the appendices, paragraphs 6.4. and 6.6.

A. Bleasdale

Meteorological Office, M.O.3.
Bracknell, Berks.

G.18695/CJH/9/63/150.

5. References

1. Meteorological Office. Averages of rainfall for Great Britain and Northern Ireland 1916-1950. H.M.S.O., 1958.

2. Meteorological Office, Climatological Service (M.O.3). Averages of rainfall for Great Britain and Northern Ireland 1916-1950 (Monthly Weather Report and Daily Weather Report stations). Hydrological Memoranda, No. 1.

3. Spreen, W.C. A determination of the effect of topography upon precipitation. Trans. Amer. geophys. Un., 28, 1947, p.285.

4. Levin, A.G. and Myachikov, V.D. Methods of calculating and mapping rainfall. Glav. geof. Obs. T., Leningrad, 64, 1956, p.17. (Typescript translation by J. Grindley.)

5. Bleasdale, A. and Rowsell, E.H. Comparison of rainfall averages 1881-1915 and 1916-1950 for Great Britain and Northern Ireland. Meteorological Office, Climatological Services (M.O.3). Hydrological Memoranda, No. 4.

6. Glasspoole, J. The rainfall of Norfolk. British Rainfall 1928, p.270.

7. Glasspoole, J. The average annual rainfall over the County of London. British Rainfall 1933, p.266.

Additional reference to be considered with items 3 and 4 above:-

8. Hounam, C. Estimation of average annual rainfall over the Port Phillip region of Victoria. Australian Met. Mag., No.21, 1958, p.1.

G.18695/CJH/9/63/150

6. Appendices

6.1. Estimation of rainfall averages, 1881-1915; numerical example.
(See paragraphs 2.1. to 2.5.)

Leicestershire, River Division 1

Year	Percentage rainfall Long-period stations L_1	L_2	Mean	Rainfall in inches Short-period stations S_1	S_2	S_3
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1881	106	112	109			
82	139	136	137			
83	115	127	121			
84	85	85	85			
85	102	110	106			
86	108	115	111			
87	78	71	75			
88	91	101	96			
89	93	108	102			
1890	85	83	84			
91	112	110	111			
92	80	83	81			
93	77	79	78			
94	102	89	95			
95	92	89	91	22.57		
96	91	102	97	25.67		
97	96	98	97	25.05		
98	77	93	85	23.54		
99	94	98	96	24.02		
1900	111	132	121	29.83		
01	95	97	96	23.76		
02	86	79	83	22.27		
03	125	110	117	31.62		
04	85	80	83	-		
05	87	73	80	21.08		
06	107	103	105	29.47		
07	112	111	111	30.29		
08	93	86	89	22.91		
09	108	97	103	26.47		
1910	116	111	113	30.50		
11	82	83	83	22.67		
12	140	138	139	37.02		
13	110	96	103	24.41	25.30	
14	107	105	106	28.09	26.42	
15	113	110	111	29.52	29.47	
16	107	100	103	30.34	30.97	
17	89	93	91		25.64	
18	102	92	97		25.47	24.77
19	120	112	116		32.42	31.47
1920	122	103	113		27.59	28.22
21	68	76	72		18.42	16.26
22	114	121	117		31.23	29.34
23	112	114	113		27.53	28.52
24	131	122	127			29.82
25	104	108	106			25.35
26	106	114	110			28.26
27	129	125	127			32.96
28	109	110	109			27.78
			Totals (inches)	561.10	300.73	302.75
			Number of years	21	11	11
			Mean (inches)	26.72	27.34	27.52
			Mean (percentages)*	101.4	103.8	109.7
			Estimated average (inches)	26.3	26.3	25.1

*From column (4)

6.2. Base maps. (See paragraph 3.2.)

6.2.1. It was decided that for drawing annual percentage maps for the estimation of averages of rainfall 1916-1950, the scale of 1:625,000 (about 1/10 inch to the mile) would be adequate, and that maps on any larger scale might prove to be cumbersome. It has been confirmed in practice that the scale is adequate, though in areas where the density of rainfall stations is greatest (notably the London area) there is slight difficulty. However, such areas are few. The scale has the advantage that it coincides with that of the Ordnance Survey map covering Britain in two sheets, showing topography and the national grid, which has recently been introduced for routine monthly rainfall work. This scale is not adequate for the final drawing of the average rainfall map in inches, when averages for a sufficient number of short-period stations have been estimated. For this purpose the actual and estimated standard-period averages are being plotted on Ordnance Survey gridded maps on a scale of ¼ inch to the mile. The maps are being subsequently reduced, area by area, to the 1:625,000 scale, to cover Great Britain in two sheets as with the published Ordnance Survey version of the 1881-1915 average rainfall map.

6.2.2. For England and Wales it was decided to carry out the secondary survey of rainfall averages on the basis of river board areas. Rectangular base maps were prepared so that any river board area was comfortably enclosed within the map border, the work being extended beyond the river board boundaries in each case, so that by overlapping of adjacent maps any risk of discontinuities was avoided. Where two or more river board areas could be conveniently included on one rectangular map, sometimes by slightly enlarging the rectangle which would have enclosed one area only, combined river board area maps were prepared.

6.2.3. For Scotland some maps were prepared in a similar way for river purification board areas, and some work was done for areas in southern Scotland on this basis. When the work for the North of Scotland Hydro-Electric Board stations was being prepared, it was decided to use a map showing the whole of the Scottish mainland and the Western Isles, but not Orkney and Shetland, though the size of the sheet was rather cumbersome, bearing in mind that forty or more maps must be used simultaneously. Orkney and Shetland will be covered by a separate map, with an overlap with the northern part of the mainland. Northern Ireland can be covered on one sheet.

6.2.4. The outline base maps were prepared in a standard form showing the coast-line, most important rivers and tributaries, river board area boundaries, or major groupings of drainage areas where statutory boundaries do not exist, and the national grid (Irish grid for Northern Ireland). The working copies were printed in grey on white paper (see Figure 1). Copies were also made in black on tracing linen for use as index maps, but not for all areas, since the simpler alternative of a standard perspex scale was found convenient (see paragraph 3.3.4.).

G.18695/CJH/9/63/150.

6.3. Estimation of rainfall averages, 1916-1950; numerical example:
(See paragraph 3.3.)

6.3.1. Long-period stations

	L_1		L_2	
N.G.R.	NZ(45)409471		NZ(45)264416	
Average	27.89 in.		26.31 in.	
1916	29.72	106.5	27.26	103.6
17	27.74	99.5	26.02	98.9
18	25.98	93.2	22.60	85.9
19	29.38	105.3	26.05	99.0
1920	31.91	114.4	25.73	97.8
21	22.37	80.2	19.30	73.3
22	29.23	104.8	26.88	102.2
23	29.63	106.2	24.85	94.4
24	32.18	115.4	28.64	108.8
25	30.62	109.8	26.12	99.3
26	26.77	96.0	24.81	94.3
27	33.33	119.5	29.02	110.3
28	28.03	100.5	23.56	89.5
29	22.89	82.1	20.18	76.7
1930	35.62	127.7	34.88	132.6
31	29.59	106.1	27.82	105.7
32	26.44	94.8	26.06	99.0
33	24.93	89.4	24.38	92.6
34	28.35	101.6	27.18	103.3
35	31.43	112.7	29.89	113.6
36	30.19	108.2	29.16	110.8
37	29.23	104.8	31.16	118.4
38	24.75	88.7	25.56	97.1
39	27.05	97.0	31.47	119.6
1940	27.83	99.8	25.90	98.4
41	26.51	95.1	26.80	101.9
42	22.80	81.7	20.26	77.0
43	25.18	90.3	22.45	85.3
44	27.48	98.5	29.03	110.3
45	24.21	86.8	24.83	94.4
46	29.53	105.9	28.45	108.1
47	23.12	82.9	24.52	93.2
48	30.28	108.6	29.76	113.1
49	17.80	63.8	18.59	70.7
1950	34.18	122.5	31.66	120.3
51	31.27	112.1	29.36	111.6
52	24.37	87.4	22.48	85.4
53	21.67	77.7	20.57	78.2
54	28.63	102.6	30.64	116.5
55	19.78	70.9	19.08	72.5
56	25.53	91.5	24.74	94.0

G.18695/CJH/9/63/150.

6.3.2. Short-period stations

	S_1			S_2				S_3				S_4			
N.G.R.	NZ(45)096549			NZ(45)585176				NY(35)968195				NZ(45)128222			
Average	30.15 in., 16 years			30.87 in., 20 years				37.98 in., 18 years				30.26 in., 20 years			
1930								41.87	113.0	~~37.05~~	/X M	37.09	126.0	29.44	
31								38.04	99.5	~~37.05~~	/X M	30.49	108.0	28.23	
32								35.99	93.2	38.61		25.76	88.0	29.27	
33								31.36	82.2	38.15		27.79	82.5	33.68	
34								37.08	97.0	38.23		28.86	97.2	29.69	
35								41.54	112.0	37.09		34.51	117.0	29.49	
36				34.15	111.0	30.76		39.54	105.0	37.66		36.74	116.0	31.67	
37				34.75	107.5	32.32		31.54	84.0	37.55		33.30	103.0	32.33	
38				31.48	100.8	31.23		44.06	115.7	38.08		31.13	102.2	30.46	
39				32.23	104.5	30.84		41.83	111.4	37.54		34.83	115.8	30.09	
1940				(31.20)	107.5	29.02		32.29	85.0	37.99		27.07	87.5	30.94	
41	29.18	95.5	30.55	30.08	96.7	31.11		29.96	79.0	37.92		26.30	84.4	31.16	
42	21.49	76.5	28.09/	(25.33)	82.5	30.70		29.76	78.5	37.91		24.95	72.0	~~34.65~~/XM	
43	26.77	86.5	30.95	22.31	77.0	28.97		40.27	105.0	38.35		26.80	85.0	31.53	
44	28.75	102.0	28.19	33.78	107.5	31.42		40.61	106.5	38.13		33.79	105.3	32.09	
45	25.32	88.5	28.61	~~[22.05]~~	~~91.8~~	~~24.02~~/	X n.p. low	36.29	96.5	37.61		35.30	97.5	36.20	
46	27.16	94.3	28.80	31.98	105.5	30.32		38.01	100.0	38.01		42.39	106.0	39.99	
47	32.51	100.8	32.25	26.90	90.5	29.72		42.81	112.0	38.22		37.60	107.0	35.14	rim
48	33.85	107.0	31.63	27.31	94.4	28.93		41.13	106.5	38.62 /		41.11	105.0	39.15	damaged
49	23.39	72.0	32.49/	23.85	69.5	34.32/		35.74	94.0	38.02		30.48	82.0	37.17	
1950	31.95	113.0	28.27	35.31	114.0	30.97		41.04	111.0	36.97		35.39	111.5	32.17	rim
51	35.12	111.2	31.58	40.65	126.5	32.13		40.27	112.5	35.79	falling	33.08	117.5	28.15	repaired,
52	26.68	84.0	31.76	29.90	100.0	29.90		27.73	75.1	36.92	off	23.54	78.0	30.18	1950
53	22.95	75.8	30.28	23.77	79.2	30.01		28.42	77.0	36.91	c.w.	22.73	79.5	28.60	
54	33.84	117.2	28.87	36.07	115.0	31.36		46.22	127.0	36.39	adj.	32.74	118.2	27.70 /	
55	21.20	70.9	29.90	25.31	78.3	32.32		27.02	76.0	35.55	gauge	22.77	74.2	30.68	
56	31.83	105.4	30.19	29.90	96.0	31.14		36.70	103.3	35.53		32.14	107.5	29.90	
		n	16		n	21	20		n	20	18		n	21	20
		Total	482.41		T	641.50	617.48		T	757.79	683.69		T	639.93	605.38
		Mean	30.15		M	30.55	30.87		M	37.89	37.98		M	30.47	30.26
		Range	4.40		R	10.30			R	1.57			R	6.95	
		S.D.	1.18		S.D.	2.58			S.D.	.39			S.D.	1.74	
			(32.51			(35.71				(38.69				(33.95	
		M±2 S.D.	(27.79		M±2 S.D.	(25.39			M±2 S.D.	(37.09			M±2 S.D.	(26.99	

G.18695/CJH/9/63/150.

6.4. Criterion for anomalous individual values of the estimated standard-period average. (See paragraph 3.3.6.)

6.4.1. After the elimination of any sub-periods of defective observations (see paragraph 3.3.5.), the series of individual estimates of the standard-period average requires examination for isolated anomalous values. The criterion to be used for selecting suspect values has not yet been finally settled. The criterion should probably be based on the standard deviation of the series of estimates, values deviating from the mean by more than (say) twice the standard deviation being considered for rejection. Such a criterion would allow for the fact that in a district with a sparse distribution of stations with actual averages, the standard deviation of the estimates can be expected to be greater than in districts with denser distributions. The disadvantage is that, pending the full introduction of machine methods, the calculation of standard deviations, even with an electrical calculating machine, is laborious and slow.

6.4.2. Estimated standard deviations based on the extreme range of the series of estimates have been used. The method is not strongly recommended, since although it works fairly well on the average, it is at times very erratic. It is most erratic for very short-period estimates when a sound estimate of the standard deviation is of most importance. In particular the random occurrence of very skew extremes in very small samples (and occasionally in larger ones) is liable to bias this estimate. It can be used with caution, but an alternative should be available if it appears to give a poor representation of the standard deviation. The table attached to this appendix as paragraph 6.4.9. gives the relationship between extreme range and standard deviation for various values of n, the number of estimates in the series.

6.4.3. An alternative estimate of standard deviation is given by three-quarters of the inter-quartile range of the series of estimates. This value is not greatly affected by the asymmetry of randomly skew extremes, nor is it appreciably affected by the slight skewness of the distribution of the series of estimates discussed in paragraph 6.6. A weakness is that possibly tricky interpolation is required for small samples, whilst it may prove rather laborious picking out the values above and below the upper and lower quartiles of large samples.

6.6.4. A very simple criterion, which works quite well, is merely to pick out the largest and smallest values, one, two or three values at each extreme according to sample size, and examine them carefully, especially if there is any marked skewness of the extremes, when most attention is given to the values deviating most from the mean.

6.4.5. Attempts are being made to devise an empirical criterion which is a slight extension of this. In carrying out the secondary survey over a number of areas in England and Wales a first step was to select a number of stations with observations throughout the whole standard period and treat them as short-period stations. That is, the annual percentage maps were first plotted and drawn without the data for these stations, and percentages for each year for each such station were estimated from the maps. (The data for these stations were then plotted, and the maps redrawn as necessary, before proceeding with the work for actual short-period stations.) A body of data was thus obtained, covering stations in different types of country, with various densities of actual observations in their neighbourhood, for which both actual percentages and independently estimated percentages are available. It is intended to add to this body of data and eventually to use it to obtain information relevant to the accuracy of this method of estimating averages.

G.18695/CJH/9/63/150.

6.4.6. In the meantime, from a rather small sample of data, it appears that the variability of individual estimates in a series for a short-period station does not depend to any great extent on the nature of the country, the altitude of the station, or the magnitude of the average rainfall. It depends primarily, and in fact almost entirely, on the density of actual averages in the neighbourhood, on which the estimates are based. It is re-assuring that this is so. The fact is to a large extent a justification of the use of percentage rainfall maps for this and for some other purposes. The quantitative result is not yet firmly established, but as a rough guide, to take in conjunction with the very simple criterion suggested in paragraph 6.4.4. the permissible range from the mean is approximately:-

Number of stations with actual averages, within 20 kilometres	Permissible range (per cent)
10 or more	± 5
6 to 9	± 7
5 or less	± 10

In some areas of Scotland and Northern Ireland the permissible range may need to be extended further. Values outside the permissible range should be considered for rejection.

6.4.7. When the anomalous values have been selected for consideration it is sometimes possible to find physical reasons for suspecting the observations for that year. Many such anomalous values may actually have been marked for examination in advance (see paragraph 6.5.5.). Whether previously marked or not, they should be rejected. In other cases re-examination of the relevant percentage map may show a careless estimate of the percentage value or a slip in writing it down. Amendments in cases of this type should not be made unless there is clear evidence of a bad blunder, and not merely a slight error. Otherwise it would be possible to go on indefinitely making slight adjustments to the original estimates. A third possibility is that an apparently erratic percentage may be due to an irregular distribution of percentage values in the particular locality for the year in question, arising perhaps from an individual thunderstorm or a thundery month. If this is established the anomalous percentage value should not be rejected, even though its retention increases the standard error of the final estimate. This type of result is a natural consequence of erratic thunderstorm distributions and any attempt to smooth out the irregularity would replace actual statistics by incorrect subjective estimates. In contradiction to what has already been said in paragraph 6.4.6., there does in fact appear to be a tendency to slightly smaller standard deviations of the estimates (expressed as percentage of the means) in districts with high annual rainfall compared with relatively dry districts. It is not certain from the small sample of data so far examined whether the tendency is statistically significant, but if it is it must be due to the greater percentage effect of individual erratic thunderstorm distributions in districts with low average rainfall.

6.4.8. When the three possibilities of paragraph 6.4.7. have been looked into, there sometimes remain odd anomalous values which cannot be accounted for. Should these be rejected or retained in the series of estimates to be used for the final mean estimate? Whilst it is tempting to suggest that no value should be rejected unless a physical reason for rejection can be found, it is nevertheless true that a value can be so absurd as to be impossible, with no available evidence for the reason. In this case present practice is to reject it, especially as it is usually an odd value deviating in one sense from the mean, without any compensating value with a comparable deviation in the opposite sense. But the value so rejected is specially marked as having been rejected without physical justification (see paragraph 6.5.9.) so that it remains clear in any later checking of the work what has been done. As there is no precise numerical criterion of "absurdity" the same practice is at present extended to all values picked out by the rather arbitrary numerical criterion. The values so rejected are not numerous and the practice makes little difference to the final results.

6.4.9. Ratio of mean sample range to standard deviation in a sample of size n.

n	ratio	n	ratio	n	ratio	n	ratio	n	ratio
		11	3.34	21	3.99	31	4.34	41	4.57
2	1.13	12	3.43	22	4.03	32	4.37	42	4.59
3	1.71	13	3.51	23	4.07	33	4.40	43	4.61
4	2.10	14	3.59	24	4.11	34	4.42	44	4.63
5	2.39	15	3.66	25	4.15	35	4.44	45	4.65
6	2.61	16	3.73	26	4.19	36	4.47	46	4.67
7	2.80	17	3.79	27	4.22	37	4.49	47	4.69
8	2.97	18	3.85	28	4.25	38	4.51	48	4.71
9	3.11	19	3.90	29	4.28	39	4.53	49	4.73
10	3.24	20	3.95	30	4.31	40	4.55	50	4.74

Ratio of standard deviation to mean sample range

n	ratio	n	ratio	n	ratio	n	ratio	n	ratio
		11	.399	21	.251	31	.230	41	.219
2	.885	12	.292	22	.248	32	.229	42	.218
3	.585	13	.285	23	.246	33	.227	43	.217
4	.476	14	.279	24	.243	34	.226	44	.216
5	.418	15	.273	25	.241	35	.225	45	.215
6	.383	16	.268	26	.239	34	.224	46	.214
7	.357	17	.264	27	.237	37	.223	47	.213
8	.337	18	.260	28	.235	38	.222	48	.212
9	.332	19	.256	29	.234	39	.221	49	.211
10	.309	20	.253	30	.232	40	.220	50	.211

G.18695/CJH/9/63/150.

6.5. Notes on practical procedure. (See paragraph 3.4.)

6.5.1. Two types of working sheet are used for tabulating the data, for estimating rainfall averages for the period 1916-1950. Blank specimens of the actual sheets, A and B, are in the envelope attached to this appendix and numerical examples as set out on these sheets are given in paragraph 6.3.

6.5.2. The first step is that, for the area in question, data for all stations with acceptable actual averages for the period are entered on sheets A. After entering station particulars, including the actual annual average in inches, the annual values from 1916 to the latest year of record available are entered in the first column (from the set of three columns for each station). At present the sheets are ruled to take all years up to 1960. As the work progresses into the next decade, it may be possible to use current annual percentage maps, based on 1916-1950 averages, plotted and drawn as a routine. It may therefore not be necessary to draw up a new sheet A for years beyond 1960.

6.5.3. The annual values are converted to percentages, to the first place of decimals, and the percentages are entered in the second column. From these percentages, percentage maps for each year from 1916 onwards are plotted and drawn. For each area data for stations beyond the boundaries of the area are plotted, and in drawing the maps overlapping parts of adjacent maps are harmonized. The index map of stations with actual averages (paragraph 3.3.1.) is used at this stage. For both the index map and the annual map it is an advantage to complete the smaller grid squares of the base map (Figure 1) in light pencil as illustrated in Figures 2 and 3.

6.5.4. Briefly coded notes on the data may be entered in the third column from each station on sheets A. It is sometimes necessary to question individual values at this stage, or in drawing the maps, and to apply a further check for consistency to the data, although the averages have already been accepted.

6.5.5. On sheets B data are entered for all short-period stations for which estimates of the averages are to be attempted, beginning with station particulars (excluding "average" which is put near the head of the columns for later use), and continuing with annual values in inches in the first column, for all available years of record. The second and third columns are used later. In the fourth column briefly coded notes should be made about inspections, suspected deficiencies of the data, construction of turf walls, provision of new equipment and so on. The following conventions have also been adopted:

6.5.5.1. An annual value which is suspected of being in error is enclosed in round brackets.

6.5.5.2. An annual value expected to be in error is enclosed in square brackets. (This includes values rejected from publication in annual volumes of "British Rainfall".)

6.5.5.3. A known change of site, instrument or observer which is expected to have produced an appreciable effect on homogeneity is indicated by a continuous red line drawn across the columns below the last unaffected year.

6.5.5.4. A known change which may not have produced serious effects is indicated in a similar way by a broken red line.

6.5.6. Generally speaking short-period stations with less than five complete years of observations are not included, and lapsed short-period stations are neglected. These rules may be varied in areas where the distribution of stations is very sparse and it is necessary to estimate

G.18695/CJH/9/63/150.

as many averages as possible. Stations with averages rejected during the primary survey (paragraph 1.3. and 3.3.2.) are included as short-period stations. Strictly speaking, stations in the classes referred to in paragraphs 1.2.2. and 1.2.3. should also be included, but in order to save time, the present practice is to include only those stations in the class of paragraph 1.2.3. which have less than eighteen complete years of observations. Checking of estimates for such stations is already yielding encouraging results as shown by the following comparisons:-

	Former estimate	New estimate
Winchester	31.97 in.	32.15 in.
Luddington	33.64 in.	33.65 in.
Wellesbourne	25.03 in.	25.03 in.

6.5.7. Using the set of forty or more percentage maps, percentage values for the short-period stations are estimated, for all available years of record, and entered in the second column. At this stage the second index map on tracing linen or a perspex scale is used (paragraphs 3.3.2. and 3.3.4.). In making the estimates individual plotted percentages should be ignored, and the estimated percentages should be estimated purely by interpolation between the isopleths. If any inidividual plotted value does not fit the lines as drawn this value must have been distrusted to some extent in drawing the lines; it should be distrusted to the same extent in estimating percentages.

6.5.8. Using the values in inches in the first column and the estimated percentages in the second, estimated averages, to two places of decimals in inches, are calculated and entered in the third column for each year of observations. The series of estimated values is then considered, as described in paragraphs 3.3.5. to 3.3.7., using any information entered according to paragraph 6.5.5. as a guide. It has already been found that "untrustworthy" values entered in round or even square brackets (paragraph 6.5.5.1. and 6.5.5.2.) sometimes turn out to be quite reliable on any test that can be devised, whilst some "accepted" values prove to be more erratic.

6.5.9. After the elimination of any anomalous individual estimates, a revised mean estimate (based on rather fewer years of observations) is entered at the foot of the fourth column, as shown in paragraph 6.3. Values eliminated for sound physical reasons are deleted in red ink. Values eliminated without any good physical reason have been found are deleted in pencil and marked "M". It is helpful to mark of in pencil "X" against any value which is considered for rejection, with "/" for the highest and lowest values, adding the "M", or red ink deletions, if rejection is confirmed. The revised mean estimate should also be entered at the head of the columns.

6.5.10. Finally all acceptable actual averages and all estimated averages are plotted on maps ($\frac{1}{4}$ inch to the mile, showing topography) to one place of decimals only (the position of the decimal showing the position of the station). In drawing the maps some of the estimates may be rejected (see paragraphs 3.4. and 3.5.). A note on this, with a very approximate map estimate of the average in red ink, should be added to the final mean estimate entered at the head of the columns on sheet B.

6.5.11. As work progresses in future years it will be necessary to draw up a new form of sheet B for use beyond 1960. By this time both actual and estimated averages will be in regular use for plotting current percentage maps, and this work provides a further check on the validity of averages which may lead to the entry of further notes on the working sheets for both actual and estimated averages. However, if work of this kind indicates

G.18695/CJH/9/63/150.

that some kind of error is arising, the result does not necessarily imply that the average value itself is untrustworthy. It may equally well mean that current data for the station in question are becoming unreliable. For this reason the subject is not pursued in this memorandum, though the test has in fact already been applied to all actual averages and some estimated ones, by plotting annual percentage maps based on the 1916-1950 averages for a few years in the present decade, before the 1916-1950 averages were introduced for regular use.

G.18695/CJH/9/63/150.

6.6. <u>Bias of the estimated standard-period average.</u> (See paragraph 3.6.)

6.6.1. From the information for any individual year the estimate of the standard-period average for a short-period station is:-

$$R = 100\, r/p \quad \text{(see paragraph 3.3.4.)}$$

The value of p is an estimate of the true but unknown value P, so that

$$R = (100\, r / P).\ (P/p)$$
$$= A.\ P / p$$

where A is the true but unknown average for the short-period station (a constant).

6.6.2. Using S for summation and n for the number of years for which acceptable observations for the short-period station are available, the final mean estimate of the average R_n, obtained as described in paragraphs 3.3.4 to 3.3.6., is:-

$$R_n = (A/n).\ S(P/p)$$

6.6.3. For the manner of estimation of the values of p, it can be expected that for sufficiently large samples the statistical distribution of p/P will be(almost) symmetrical, with mean value (very near to) 1, and with small standard deviation. It should not be significantly different from a normal distribution with mean 1. The distribution of P/p, however, will not be normal or even symmetrical. The distribution will be positively skew because a pair of symmetrical values $(1 - a)$ and $(1 + a)$ in the distribution of p/P become an unsymmetrical pair $(1 + a + a^2 + ..)$ and $(1 - a + a^2 - ...)$ in the distribution of P/p, with the positive deviation from the mean greater than the corresponding negative. The mode of the distribution of P/p is approximately $(1 - 2s^2)$, and the mean is approximately $(1 + s^2)$, where s is the standard deviation of p/P.

6.6.4. Hence the expected value of the estimate:-

$$R_n = A/n.\ S(P/p)$$
$$= A(1 + s^2)$$

instead of A. That is the estimate given by R_n is positively biassed, on the average, by $100s^2$ per cent. Since, however, the value of s is not usually as much as 0.05, the magnitude of the bias in the estimate R_n is not likely to be as much as 0.25 per cent, on the average.

6.6.5. This has been confirmed empirically, For a total of 60 estimates the apparent positive bias of the estimate R_n was about 0.2 per cent on the average. It was much greater in individual cases, and maximum values for varying periods of years n, on which the estimates were based, were approximately:-

n	5	10	15	20	25	30
percentage bias	1.2	0.8	0.6	0.4	0.4	0.3

For the shorter periods the random effects which are to be expected with small samples led to a larger scatter of individual values than for the longer periods, and there were even two cases of apparent negative bias of as much as 0.3 per cent. The two largest values, 1.2 per cent for a 5-year estimate and 0.8 per cent for a 10-year estimate, were in absolute values:-

G.18695/CJH/9/63/150.

Estimate, R_n	Standard Deviation	Estimated Bias of R_n
67.0 in.	0.9 in.	0.8 in.
27.8 in.	0.5 in.	0.2 in.

In general the bias of the estimate is probably much smaller than the standard deviation of the estimate, and probably wholly negligible. But the occurrence of a value as large as 1 per cent, for a very short-period estimate, suggests that the matter may be worth looking into further, with the object of devising a simple criterion and correction of the bias for the few cases when it may be significant. This is being pursued.

WEAR and TEES RIVER BOARD AREA

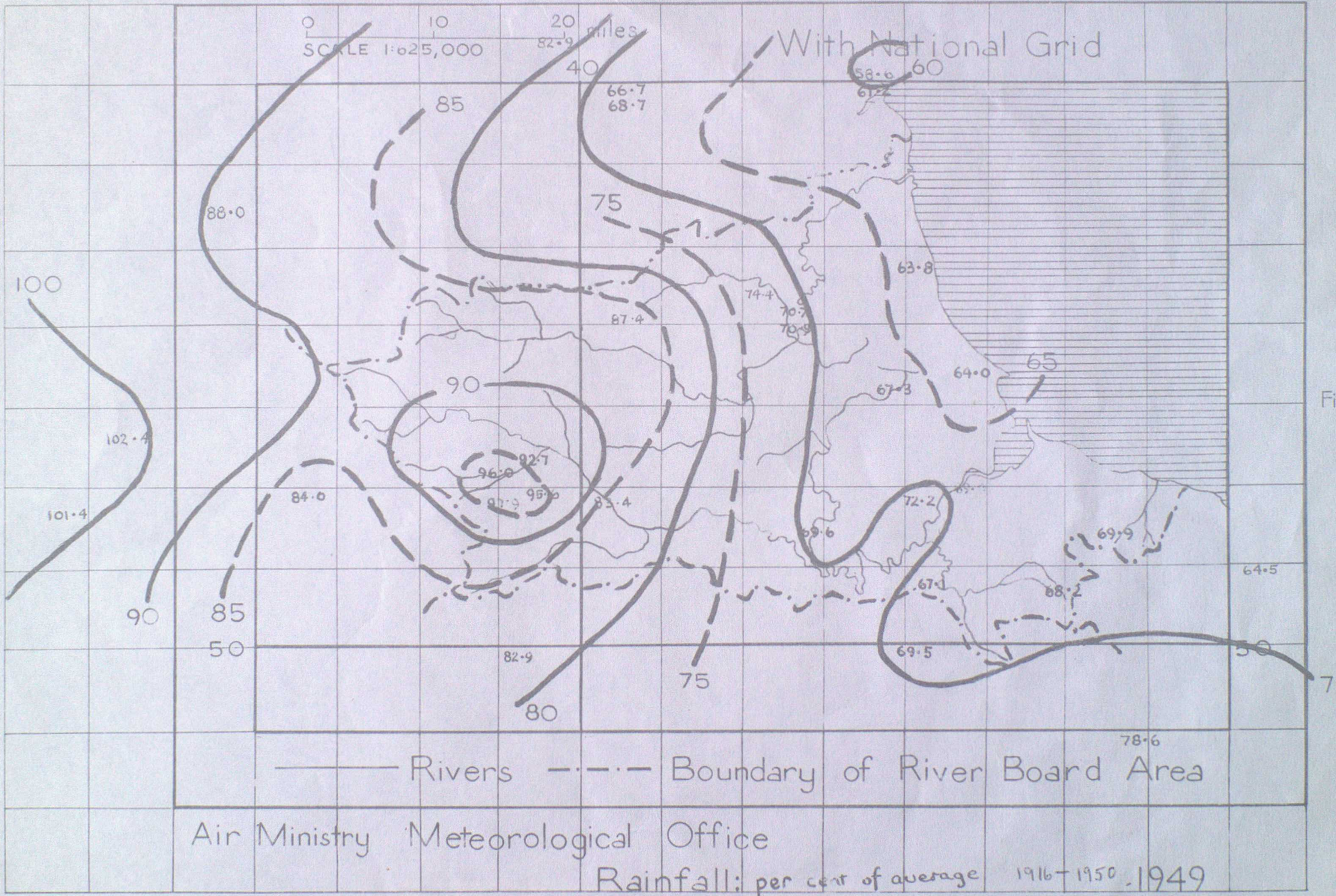

Fig.2