International Workshop on the Correction of Precipitation Measurements
1 - 3 April 1985 ZURICH

VERIFICATION OF RAINFALL DATA QUALITY AND HOMOGENEITY THROUGH THE METHOD OF ANNUAL RAINFALL INDICES VECTOR

Y. L'HÔTE

Laboratoire d'Hydrologie
Institut Français de Recherche Scientifique pour le Développement en Coopération
(O . R . S . T . O . M.) FRANCE

ABSTRACT

The author points out the important phases of the method of annual rainfall indices vector (G. Hiez 1977 and Y. Brunet-Moret-1979). This method is similar to the double-mass method, but here annual observations are not only compared to some other stations, but to a single index using all the annual observations of a homogeneous area to be calculated. We call "regional vector" the series of annual indices.

Some results are given, concerning six studies made in large intertropical and mediterranean zones, over the last five years.

Then we present some advantages of the method, such as the extension of annual means and the estimation of missing data, calculated with computer. Finally the vector of annual indices is a good synthesis for studying long term variations, namely dry and wet sequences. The obtained vectors in diversified climates can be successfully compared.

1 - REQUIREMENTS FOR A CONTROL - FIRST AND SECOND DEGREE HOMOGENIZATION

Prior to any simple statistical approach (like computation of mean) upon a long period of rainfalls data, mostly obtained through non recording gauges, it is advisable (for eventually correct or delete) to detect :

a- important mistakes which can occur during measurement (loss of water, missing or unqualified observer etc...) or during data report (neglect, false report etc...).

b- systematical errors resulting from unnoticed change of place, or from an equipment discrepancy (discordance of test-tube and bucket diameter), etc...

If these errors are not detected, there is a risk to get some non homogeneous series including samples of several statistical populations, and the final result will be wrong.

In a classical way, for a specific climatic zone, we can check the homogeneity of annual rainfall series through the double-mass analysis.

When the study is conducted at a wide scale (a country survey, great basin monographs..), the homogenization is considered as definitive, and we follow two successive steps of which we give definitions here after, as accepted by French speaking people.

- Homogénéisation du 1er ordre (first degree homogenization) :

Insure that the sample is really a part of a single population. We consider generally that the additions, corrections and deletions resulting from this operation, as well as the ones obtained by other comparisons are part of the first degree homogenization.

- Homogéneisation du 2ème ordre (second degree homogenization) :

Through various methods, computation for a common length of time (as long as possible) of statistical values, specially: annual means, standard deviation,... estimation of missing data, etc...

2 - THE METHOD OF ANNUAL RAINFALL INDICES VECTOR

For a wide region including many stations, the double-mass method becomes a drudgery, even with automatic computation. To eliminate one of inconveniences, G. Hiez (1977) suggested to compare each station to a single sequential group of rainfall indices (Called REGIONAL VECTOR), calculated with the observed data at all the stations.

We are briefly presenting here the main points of a similar method programmed by Y. Brunet-Moret (1979) :

The index zi of a year i is computed through the rainfalls of the n stations (n fluctuating from one year to the other) :

$$zi = \sum_{a=1}^{a=n} (Pai/\bar{P}a) \;/\; n$$

Where Pai is the annual rainfalls during the year i at the station a

and $\bar{P}a$ the annual rainfalls mean at station a

The output of computer page 3, copied on table 1 shows the values of indices zi (1.0481 0.7774 ...), for the successive years (1964, 1965...); the indices are calculated with 5, 6...stations. The mean of zi is 1. Index approach 1 for average years (1964, 1977.....). The wet years get high indices (1970..) and vice versa (1971 1980..).

Table 2 shows the deviation Eai for each station a and each year i :

$$Eai = (Pai/\bar{P}a) - zi$$

81195

The 118 (in this case) deviations represent a normal distribution about their mean (zero), and with a standard-deviation of $\sqrt{MT2}$ (point out on table number 2). Through this table, we can detect :

- some years too much dry or wet, where deviations Eai exceed 2.5(3.0...) times the standard-deviation, (i.e. 1% (0. 1%..) of the observations). For example : year 1966 at station 1020,
- sequences of years with high positive or negative deviations, showing possible systematical errors. For example : years from 1975 to 1981, station 1640, and years 1980-1981, at station 1050.

With other pages of listing (eight pages on the whole), and an automatical graph, we can check the double-mass curves between the regional vector and each station. So we can confirm or not the detected errors. In this case we have confirmed :

- equipement discrepancies about stations 1640 and 1050. These have been observed on the field,
- errors during report at station 1020, year 1966.

It should be observed that it is not necessary to correct all that seems to be unreliable. In fact, at the end of the study, it may remain for example 5% of values out of 2 standard-deviations (but approaching). In this case, where 2 standard-deviations equal 0.18 (before correcting) six values may remain approximatively.

3 - APPLICATIONS UNDER INTERTROPICAL AND MEDITERRANEAN CLIMATES

Table 3 shows quantitative characteristics about six studies on different size zones. We can note that the number of regions (with different vectors) is not related to the wideness of the study area, but rather :

- to the density of stations,
- to annual and monthly weather changes, more than the annual rain amount.

For example, about Basse-Terre of Guadeloupe (mountainous island), we have gathered stations with annual rainfalls ranging from 1000 to 8000 millimeters, but located on the same side of the mountain (windward or leeward).

On the other hand, G. Hiez and L. Rancan (1983) have defined about the whole Brazil, regions delineated by imbricated hexagons of 12500 km², which can been divided or grouped.

About the efficiency of the method, table 4 shows that we can get many corrections, specially with long time series.

4 - ADDED ADVANTAGES OF THE METHOD

With vector method, and chiefly the computing program, we can obtain the following operations :

a) - take into consideration all the stations, including the ones with few broken records: 5, 6 years (adjustable), not necessary continuous,

b) - obtain immediatly a long term average upon the vector period,

c) - eventually improve this extension upon a longer period, by taking into consideration the other vectors of the study,

d) - if necessary, calculate the missing annual rainfalls, by preserving the standard-deviation of the observed sample. So the estimated annual values may be used :

- in a hydrological model,
- to improve the annual statistical study of the stations with short period in particular, even if the series are gaussian or not.

Finally, the annual indices vector is a good synthetic tool to point out the sequences of dry and wet years. For example, we shall select the Chad Lake Basin survey (Y. Brunet-Moret and a. 1979) -about 1000 000 km²- for which we indicate (graph 1) the annual isohyets (1946 - 1976) and the boundaries of the sixteen regions.

The grah number 2 shows for the septentrional regions 1 and 5, and for the southern region 9, the variation of the cumulated differences between indices zi and their mean (equal 1.0) c/f table 1, column 4 :

$$Ci = \sum_{i=1}^{i=n} (zi - 1)$$

Where i is the sequential number of zi series

and n the number of years in the vector

As an example, we read the deviations of Ci about zone 5 as follows :

- a sequence of wet years from 1949 to 1959, during which Ci remains increasing,
- a group of relatively average years from 1960 to 1967, where Ci is quite constant,
- a sequence of dry years from 1968 to 1976 (Ci decreasing).

The graph 2 well shows the small dissimilarities between regions 1 and 5 (i. e. year 1951, 1962 to 1967...), but specially the lower ranges of dry and wet sequences observed in region 9 (showing a more wet average).

With such synthetic tools, we expect to compare climatic trends upon large scale regions, like continents.

5 - CONCLUSION

In spite of the interesting advantages showed as above, it should be observed that, like with double-mass method, the utilization of vectors is a long and wearying task. The six studies referenced above have required from 2 to 8 months of engineers' time, all data ready for processing.

The major difficulties occured in data are the lacks of history of stations, for instance the data of change in location, alteration of equipment, etc...

However, with some quite accurate data, we can expect through vector method to detect in the first step, and then to correct carefully :

- the obvious mistakes in data collections,

Table 1

OUTPUT OF COMPUTER

VECTEUR Z APRES 3 ITERATIONS E=ECARTS A Z PAGE TROIS 1

NBRE STATIONS	ANNEE	VECTEUR	CUMUL 0.00	MOYENNE	MOMENT 2	
5	1964	1.0481	0.05	0.0000	0.000754	1964
6	1965	0.7774	-0.17	0.0000	0.005633	1965
7	1966	1.4021	0.23	-0.0469	0.025969	1966
5	1967	0.8826	0.11	0.0000	0.008305	1967
7	1968	0.8681	-0.02	0.0000	0.003229	1968
7	1969	1.0629	0.04	0.0000	0.002280	1969
7	1970	1.5967	0.64	0.0000	0.002430	1970
7	1971	0.6159	0.25	0.0000	0.003138	1971
7	1972	1.0672	0.32	0.0000	0.002691	1972
7	1973	0.7619	0.08	0.0000	0.005208	1973
7	1974	1.0205	0.10	0.0000	0.002789	1974
7	1975	0.7653	-0.13	0.0000	0.003760	1975
7	1976	0.8346	-0.30	0.0000	0.003482	1976
7	1977	1.0000	-0.30	-0.0275	0.010487	1977
7	1978	0.9127	-0.38	-0.0252	0.007130	1978
6	1979	1.3994	0.02	-0.0457	0.022095	1979
6	1980	0.6818	-0.30	0.0000	0.005326	1980
6	1981	1.3028	-0.00	0.0000	0.025647	1981

POUR LES Z DU VECTEUR NBRE D ANNEES 18 MOYENNE 1.
VARIANCE 0.0698038 COEF ASY 0.706 COEF VAR 0.2642
PROB AU DEPT TEST HOMOGENEITE 0.793 STRICT COEF AUTOCORRELATION -0.501

Table 2

OUTPUT OF COMPUTER

MATRICE DES 118 E=ECARTS A Z:PRECIPITATION/MOYENNE-VECTEUR PAGE CINQ 1
*1720*1020*1040*1640*1050*1740*1750*

1964	****	-.05	0.02	****	0.01	0.01	0.01	****	****	****	****	****	****	****	1964
1965	****	-.02	-.11	0.14	0.02	-.05	0.02	****	****	****	****	****	****	****	1965
1966	0.03	-.39	0.12	0.00	-.04	-.10	0.05	****	****	****	****	****	****	****	1966
1967	-.06	-.13	****	0.10	-.01	****	0.10	****	****	****	****	****	****	****	1967
1968	-.09	-.05	-.03	0.08	0.02	0.07	0.01	****	****	****	****	****	****	****	1968
1969	0.03	-.07	-.00	0.07	-.02	0.04	-.05	****	****	****	****	****	****	****	1969
1970	-.04	0.10	-.01	0.02	0.00	-.08	-.00	****	****	****	****	****	****	****	1970
1971	-.08	0.04	0.01	0.10	-.03	0.01	-.06	****	****	****	****	****	****	****	1971
1972	-.03	-.04	0.00	0.11	-.00	0.03	-.07	****	****	****	****	****	****	****	1972
1973	-.06	0.05	-.08	0.00	0.14	0.02	-.07	****	****	****	****	****	****	****	1973
1974	-.04	-.08	-.05	0.06	0.05	0.04	0.02	****	****	****	****	****	****	****	1974
1975	-.05	0.02	-.04	-.09	0.11	0.05	-.00	****	****	****	****	****	****	****	1975
1976	0.00	-.04	-.03	-.10	0.09	0.03	0.05	****	****	****	****	****	****	****	1976
1977	-.06	0.10	-.04	-.23	0.05	-.03	0.02	****	****	****	****	****	****	****	1977
1978	0.02	-.01	-.03	-.21	0.04	0.02	-.00	****	****	****	****	****	****	****	1978
1979	0.06	0.06	0.01	-.35	-.03	-.02	****	****	****	****	****	****	****	****	1979
1980	0.11	0.01	0.06	-.07	-.11	-.00	****	****	****	****	****	****	****	****	1980
1981	0.17	0.11	0.09	-.25	-.20	0.09	****	****	****	****	****	****	****	****	1981

*1720*1020*1040*1640*1050*1740*1750*

MOMENTS DES E MT1 -.82365Q-02 MT2 0.76750Q-02 MT3 -.11460Q-02 MT4 0.45051Q-03

$$\sigma = \sqrt{MT2} = 0{,}088$$

Table 3

Study	Area surveyed km²	Number of years (end of study)	Number of selected stations	Number of regions
Chad Lake basin (project)	1 000 000	31	101	16
Republic of Togo (atlas)	56 000	49	72	9
Central Tunisia (monograph)	9 700	55	83	6
Alaotra Lake-Madagascar (model)	7 000	35	30	2
Guadeloupe Island (monograph)	1 520	50	105	5
Tahiti Island (monograph)	1 000	15	78	10

Table 4

STUDY	NUMBER OF STATIONS				Remarks
	Beginning of the study	Deleted (too many errors)	With equipement discrepancies	With heterogeneous series	
ALAOTRA	45	15	11	9	35 years
GUADELOUPE	131	26	21	71	50 years
TAHITI	80	2	1	0	15 years. Storage raingauges

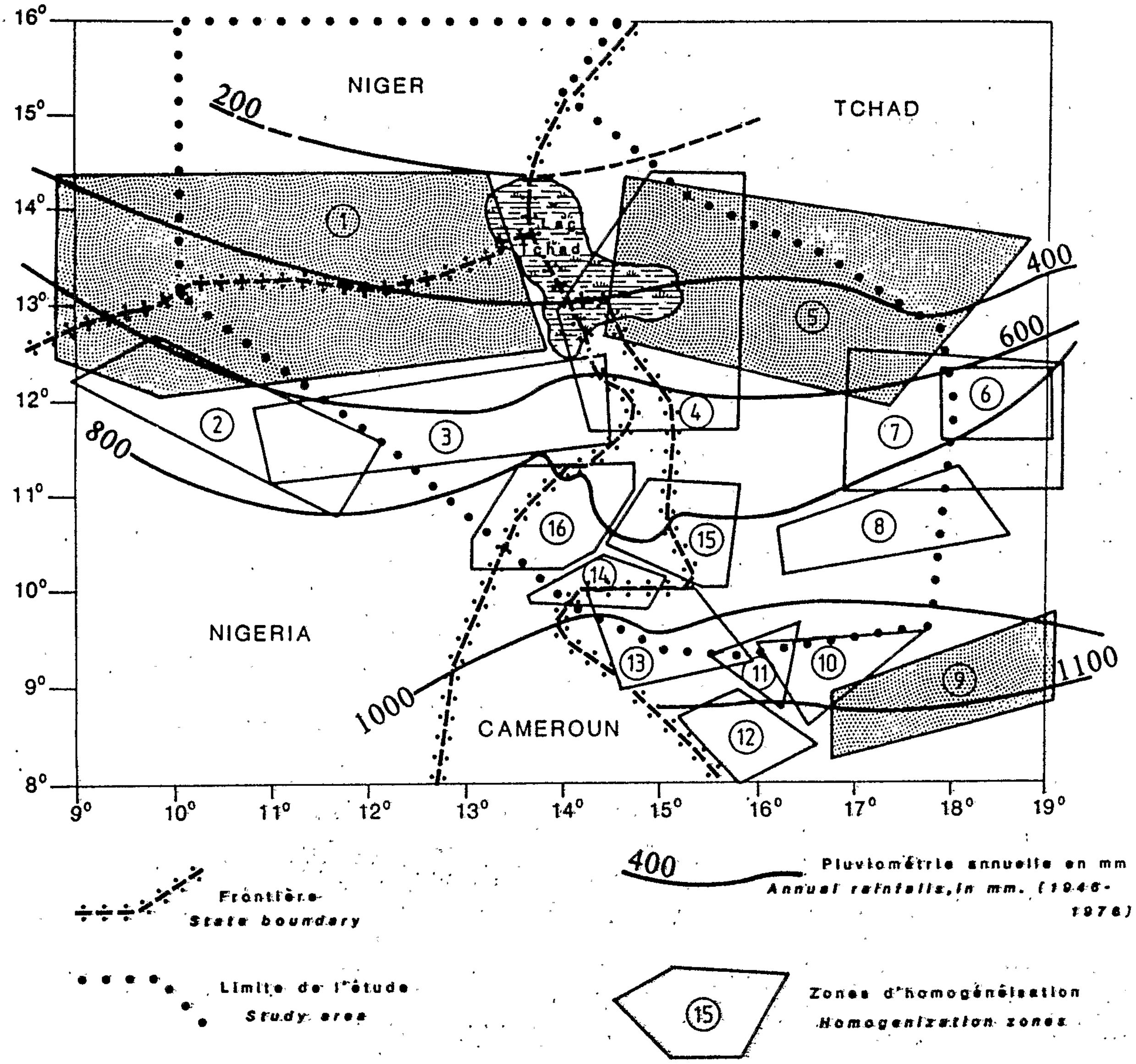
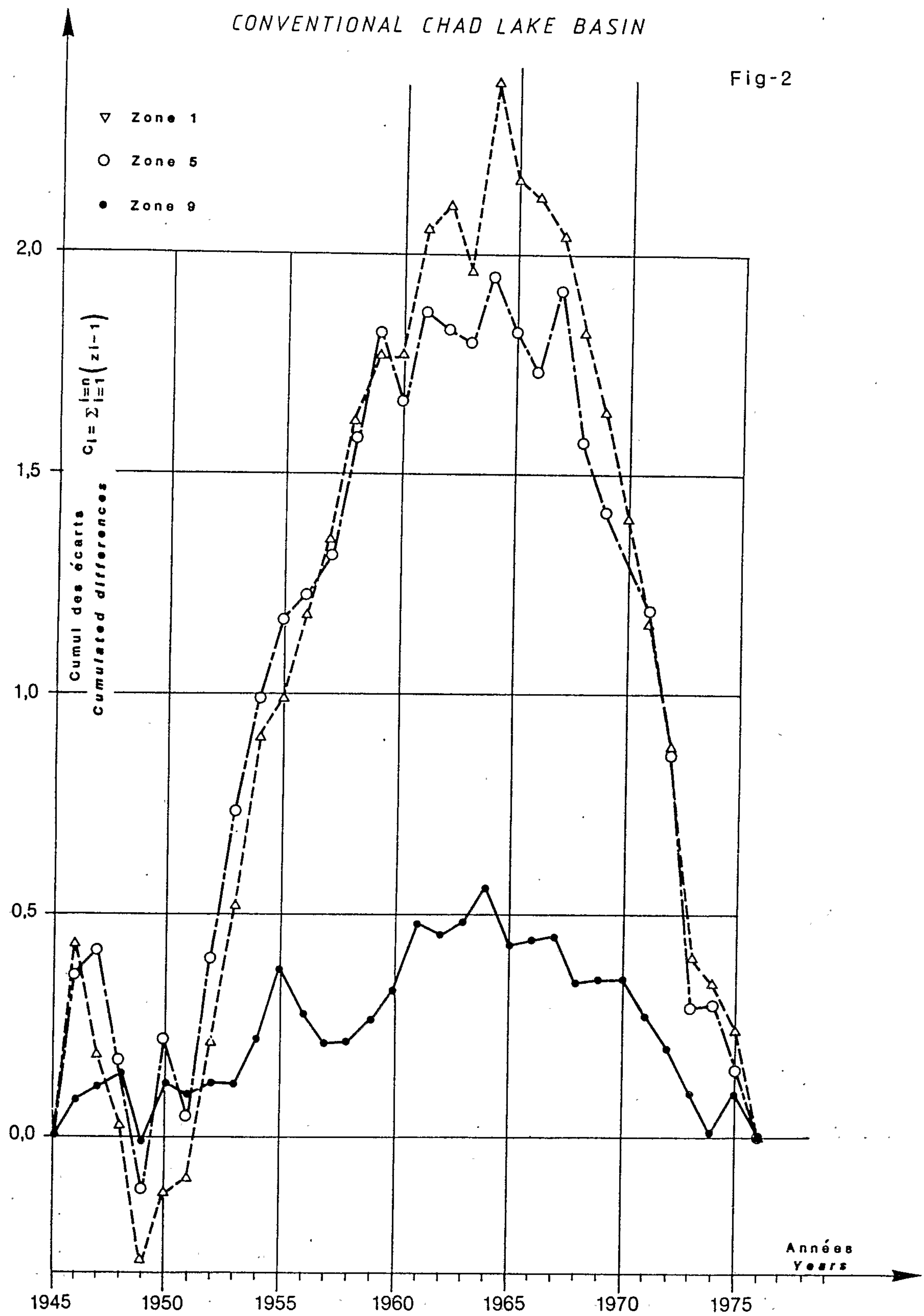


Fig-1 ETUDE DU BASSIN CONVENTIONNEL DU LAC TCHAD
STUDY OF THE CONVENTIONAL CHAD LAKE BASIN

BV CONVENTIONNEL DU LAC TCHAD

CONVENTIONAL CHAD LAKE BASIN



VARIATION DU CUMUL DES ECARTS DU VECTEUR A SA MOYENNE (1.000)

VARIATION OF THE CUMULATED DIFFERENCES BETWEEN VECTOR AND ITS MEAN (1.000)

- flagrant equipment discrepancies,
- surroundings or location changes (when they are noticed, and in a small number for a station).

6 - REFERENCES

. BRUNET-MORET (Y.), 1979 - Homogénéisation des précipitations.

Cahiers ORSTOM, série Hydrologie, vol.XVI, n° 3-4, pp. 147-170

. BRUNET-MORET (Y.), COLOMBANI (J.),VUILLAUME (G.), 1979 - Etude des potentialités du bassin conventionnel du Lac Tchad.

ORSTOM, Paris.

. HIEZ (G.L.G.), 1977 - L'homogénéité des données pluviométriques.

Cahiers ORSTOM, série Hydrologie,vol.XIV, n° 2, pp. 129-172

. HIEZ (G.L.G.), RANCAN (L.), 1983 - Aplicacao do metodo do vetor regional no Brasil.

4ème Symposium brésilien des ressources hydriques,

Novembre 1983, Blumenau (Brésil).