Appropriate record lengths for the estimation of mean annual and mean monthly precipitation in southern Africa†

SD Lynch* and MC Dent

Department of Agricultural Engineering, University of Natal, PO Box 375, Pietermaritzburg 3200, South Africa

Abstract

Mean annual precipitation (MAP) and mean monthly precipitation (MMP) are two of the most widely used variables in hydrological design, water resources planning and agrohydrology. The use of short-term records can bias estimates of MAP and MMP significantly. Such bias is more pronounced in semi-arid regions which cover much of South Africa. The objective of this study was to estimate appropriate record lengths for good estimates of MAP and MMP in 24 regions covering South Africa.

The results of this study are presented in tabular form for one of the regions and also as a map of South Africa on which appear the 24 regions. A regional delimitation of the number of years of record required in order to produce an estimate of MAP which is within approximately 10 per cent of the long-term mean, 90 per cent of the time is presented on this map.

Introduction

The importance of mean annual precipitation (MAP) as an index in hydrological applications may be gauged by the numerous examples mentioned below. MAP is an index which is used to characterise overall climate and moisture status of a catchment. It has a major influence on both soil conditions and their drainage characteristics (Bedient *et al.,* 1978) and is a dominant factor influencing type and condition of vegetation. MAP was used by Pitman (1980) as a basis for the regionalisation of Namibia into homogeneous meteorological zones. He also related the frequency of occurrence of severe storms and amount of rainfall per event to MAP. Midgley and Pitman (1969) used MAP in the estimation of mean annual runoff (MAR) and Midgley and Pitman (1978) used MAP in their depth-duration-frequency relationships for point rainfall in South Africa. In this capacity alone MAP must surely be used many hundreds of times per year in southern Africa. Pitman (1973) expressed the rainfall input to his well-known monthly streamflow estimation model in terms of MAP and Pitman and Stern (1981) used MAP in the same role in Namibia. MAP was used by Seeber (1983) in an extreme-value expression for the determination of the cumulative distribution of point rainfall rate, in a model for use in the telecommunications industry.

Schulze (1984) used the streamflows from Pitman *et al.* (1981) and related MAR to MAP through regression analysis in 21 runoff regions in Natal to then map MAR using MAP as the only predictor. An underestimation of 10 per cent in a catchment MAP of 1 300 mm was shown by Midgley and Pitman (1969) to underestimate MAR by 26 per cent in the Drakensberg. Boughton (1981) found that MAR varied from 100 mm to 530 mm for an MAP variation from 850 mm to 1 700 mm in the Upper Condamine River Basin, Australia and that variations of 20 per cent in MAR for 10 per cent change in MAP are common.

In the field of agrohydrology, MAP has been used extensively. *Eragrostis curvula* yield was expressed as a function, *inter alia* of MAP by Jones (1982). MAP and other variables were used by Brockett (1982) to obtain a function for kikuyu yield. Jones *et al.* (1980) expressed a minimum percentage fodderbank accumulation in terms of MAP and other variables. Schulze (1983) used recommendations from the Department of Agriculture and Fisheries (1981) and then developed regional regression equations which included MAP in order to enable the mapping of average first burning dates of veld in Natal. MAP was incorporated as one of their criteria for determining optimum growing areas for pineapples by Nield and Boshell (1976), whilst Schulze (1983) obtained correlation coefficients typically of 0,84 when he related first planting dates for maize to MAP in Natal. MAP, altitude, temperature and duration of soil moisture deficit were used by Schönau (1982) in order to determine criteria for optimum growth of various timber species.

It is evident from the above that good estimates of MAP and mean monthly precipitation (MMP) are of fundamental concern in a number of fields. The choice of the appropriate period and record lengths for the estimation of MAP and MMP in South Africa is therefore important. This study was instituted with the objective of answering these questions. Such findings were to be used ultimately in a project to map MAP and MMP in South Africa.

Mapping long-term rainfall statistics

The statistics MAP and MMP are really useful when they are presented as areal estimates. Such areal estimation of long-term rainfall, from point measurements, is therefore an important prerequisite to the use of this fundamental hydrological variable in, for example, distributed modelling and agricultural land-use planning. It is therefore necessary to estimate and then extrapolate and interpolate these variables in space.

It is not only essential to have a good estimate of the population statistic, for example of the MAP, at a few rainfall stations but it is also desirable to have a comprehensive spatial and representative altitudinal distribution of stations with an acceptable record length, before attempting to achieve a good areal estimate of long-term rainfall. It is reasonable to expect that increasing the length of the rainfall record at each station will improve the spatial distribution of rainfall stations which meet any record length criterion. This point is illustrated aptly in Fig. 1, which depicts the location

† Revised paper. Originally presented at the Hydrological Sciences Symposium, Rhodes University during September 1987.

*To whom all correspondence should be addressed.

Received 3 February 1989; accepted in revised form 27 December 1989.

of stations which would qualify for inclusion, should 10, 20 or 40 years' minimum record length limits be set. In addition to the improved two-dimensional spatial distribution shown in Fig. 1 the altitudinal representation should also be considered. Fig. 2 depicts the cumulative frequency distribution of the rainfall station altitudes and the grid point altitudes for the area shown in Fig. 1. The dearth of stations in the higher altitudes of, for example the south-western Cape region, is noticeable and is of concern when regression techniques to estimate MAP using *inter alia* altitude are applied, particularly where the regression equations are used to estimate MAP beyond the altitude range of rainfall stations for which the regression equation was developed in a region.

The rainfall record may be lengthened beyond and prior to the period of active recording at the station and gaps in the record filled in to produce the same effect on the distribution of stations which meet a record length criterion as would the lowering of the record length criterion itself.

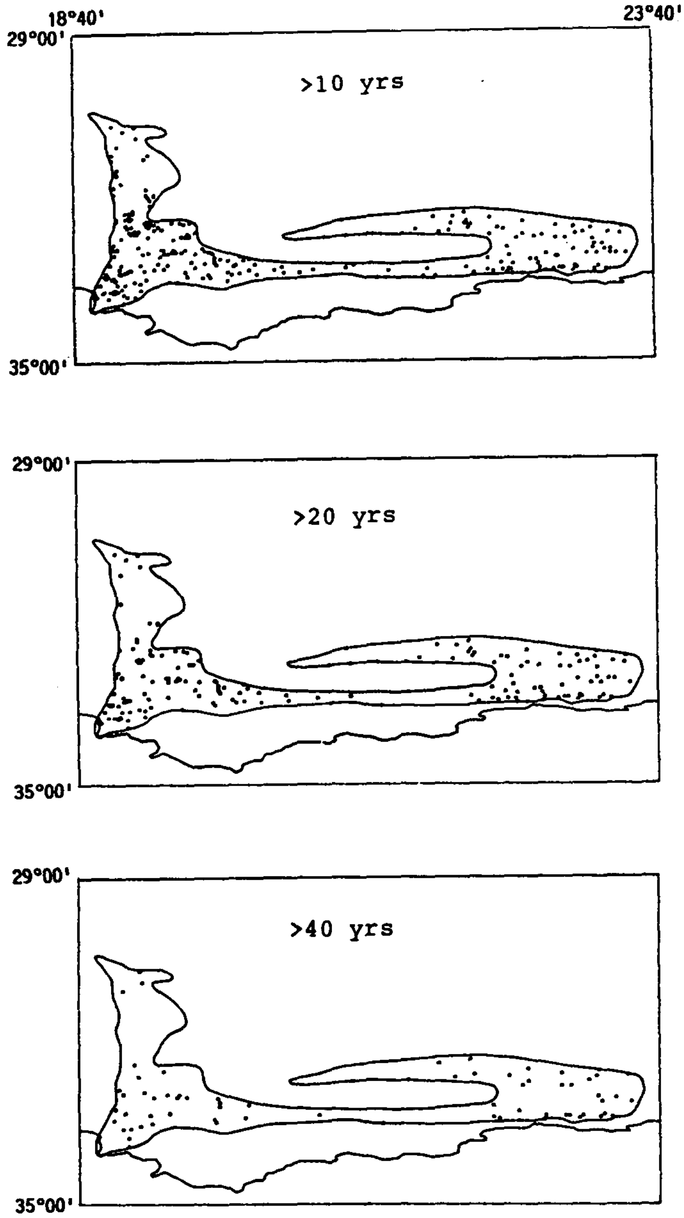

Figure 1

Spatial distribution of rainfall stations in the south-western Cape with record lengths of more than 10, 20 and 40 years.

The choice of the period of rainfall record and its length is not a straightforward matter. There are three aspects to this question which require detailed consideration, viz.:

- the length of the base period;
- the start and end years of the base period, or alternatively; and
- the acceptability of having no fixed period as long as the span of the record is long enough.

Whilst reviewing these questions it is important to reflect on the following extracts from Dyer and Tyson (1977), the World Meteorological Organisation (1974) and Dunne and Leopold (1978). Dyer and Tyson (1977) state that: "No evidence can be found to show that South African rainfall has increased or decreased progressively throughout the period of meteorological record. Instead quasi-periodic fluctuations appear to have persisted throughout the period" (p. 145). This may be interpreted as implying that it does not matter when one takes one's period of record so long as it is of sufficient duration to have spanned enough of the "quasi-periodic fluctuations" that their influence does not affect critical statistics. The World Meteorological Organisation (1974) states that the choice of the optimum base period for MAP estimates is "a period long enough to represent a good sample of record through time. Too short a period may be unduly influenced by a particularly dry or rainy period. Too long a period will require too much synthesis of record" (p. 5.3) and will most probably result in having to use a poor network of rainfall stations.

Dunne and Leopold (1978) made a significant contribution to the discussion on the base period: "By international agreement, the average climate of the world used to be defined as the average value of precipitation, temperature and other variables for the period 1931 to 1960. As the period of observation was extended forward and backward in time, it became apparent that the period 1931 to 1960 had one of the most abnormal climatic regimes in the last several hundred years. The average climate of a place is now generally updated each decade to another 30-year period, 1941 to 1970, 1951 to 1980, and so on. Many published records and atlas maps, however, are based upon data from 1931 to 1960 and they should be checked against older and more recent information wherever possible" (p. 71).

On the existing 1:250 000 isohyetal maps of southern Africa it is stated that the MAP is for the period 1921-1960. A study of the rainfall record length and period for stations in several regions revealed that the records at these stations must have been "normalised" in order to allow the station to be used for the isohyetal mapping exercise, bearing in mind the 40-year base period from 1921 to 1960. Such a normalisation technique generally involved the determination of the ratio between the annual rainfall at the station to be normalised and a long-term station (covering the base period). This ratio was then used to estimate annual totals for the "missing" years at the station with the short record. Such a technique, which is a rudimentary form of linear regression, has several shortcomings, notably that the variance of the synthetic annual values is considerably reduced as stated by Matalas and Jacobs (1964) and Zucchini and Hiemstra (1984). The reduced variance of the synthetic data impairs the use of these data in regression analyses where the variance of the annual totals is used in the weighting of the input variable MAP. In addition the usefulness of one of the most important by-products of this study, namely the synthetic monthly rainfall data, would have been reduced, had the normalising technique been applied, since the variance of the

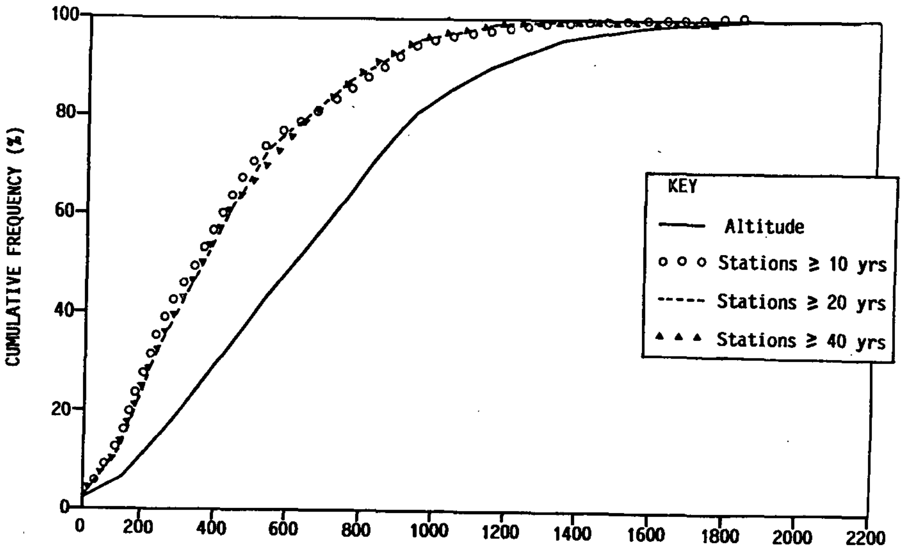

Figure 2
Cumulative frequency distributions of rainfall station altitude and grid point altitude for rainfall stations in the south-western Cape with record lengths of more than 10, 20 and 40 years.

monthly rainfall record is vital, if used for example in any rainfall/runoff modelling for the purposes of assessing water resources.

Several preliminary studies revealed the paucity of rainfall stations with 30 or more years of record and thus made it imperative to review the concept of a fixed long-term base period. It was decided that the best possible estimate of the population statistic at a rainfall station was to be gained by using the data at that station, irrespective of the start and end years of the rainfall record period, provided of course that the record was sufficiently long.

It therefore remained to establish a minimum record length of annual totals in order to calculate a reasonable estimate of the long-term MAP and MMP. Furthermore it was necessary to determine whether this minimum record length varied from region to region. Knisel *et al.* (1979) suggest that the use of short-term records can bias estimates of MAP significantly and that this bias is more pronounced in semi-arid regions. Large parts of South Africa are semi-arid. Numerous studies on South Africa's annual rainfall have shown the apparent persistence of high and low periods (for example Dyer, 1976; Hall, 1976; Dyer and Tyson, 1977; Dyer and Gosnell, 1978). Rawson (1908), cited in Dyer and Tyson (1977) and discussed by Louw (1982), reported a 19-year oscillation in the position of the Atlantic and Indian Ocean anti-cyclones for the period 1841 to 1906. Nevill (1908) found an 18 to 20-year oscillation to be present in Natal for the period 1850 to 1908. Hall (1976) analysed the growth rings of a tree felled near Pietermaritzburg in 1910 and suggests that a quasi 20-year oscillation was evident from the width of growth rings from 1910 back to 1750. Dyer and Gosnell (1978) found a significant long-term oscillation with a mean wavelength of 19,2 years in the sugarcane belt of Natal. Dyer (1976) used the quasi 20-year oscillations in order to estimate expected future rainfall in selected parts of South Africa. Whitmore (1962) reported "a significant degree of persistence" in the annual rainfall of South Africa. Louw (1982) found oscillations of 18 years and shorter in the South African annual rainfall record. Research conducted by Bramley (1980) in Zimbabwe on annual rainfall showed a 19 to 20-year periodicity in Harare, an 18-year periodicity in Gweru and a 27-year periodicity in Bulawayo.

The fact that the existing 1:250 000 scale isohyetal maps employed a 40-year base period combined with the apparent tendency for a periodicity around 20 years in the annual rainfall led to the decision that, initially, a minimum record length of 40 years be considered. However, it became apparent during the study that parts of the country are not represented realistically with rainfall data, if only the stations with 40 years or more of record are considered. It was evident that some relaxation of this record length criterion was necessary in order to achieve an acceptable spatial distribution of stations. However, such a reduction was not to lead to a poorer estimate of the mean. It was felt that the resultant inclusion of vital stations would certainly lead to more realistic isohyetal estimates than if those stations had been omitted.

Determining the upper limits of the percentage error of the short-term mean

The problem of estimating the error of the estimates of the mean may be approached from a purely statistical point of view if it is assumed that the distribution of errors of estimating the mean is a function of the t-distribution. Under such an assumption the standard error of the estimate of the mean (sem) would be

$$\text{sem} = \sqrt{s^2/N}$$

where:

s^2 = variance of the observations
N = number of observations

In this study no assumptions were made regarding the distribution and the analysis proceeded as outlined below.

A first step in the proposed relaxation of record length was to

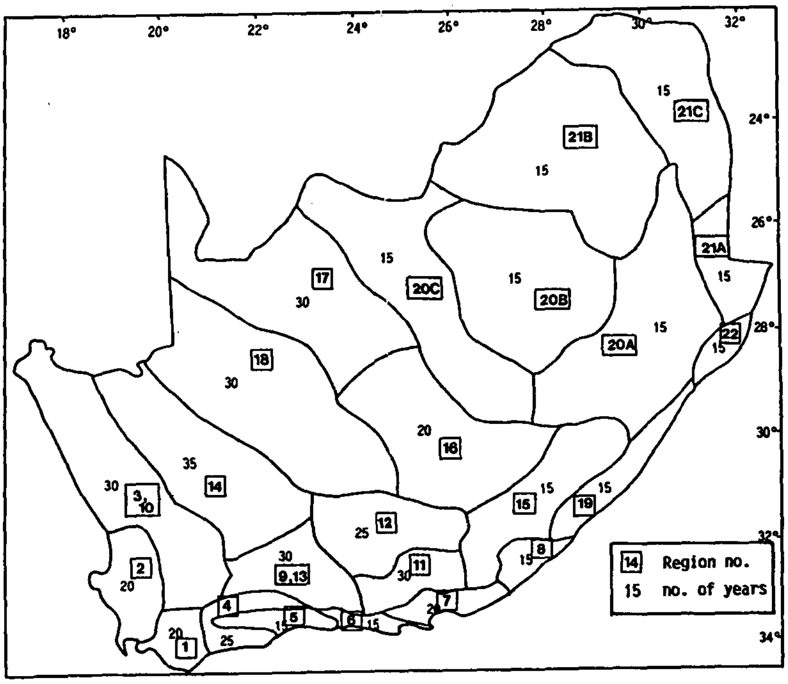

Figure 3
Record lengths required to ensure that the mean of annual rainfall estimates to within 10 per cent of the long-term mean, 90 per cent of the time

determine the possible error, in the estimate of the long-term mean, which would be introduced if shorter term means were to be used. The rainfall stations were grouped according to the 24 major rainfall regions of southern Africa proposed by Welding and Havenga (1974) and shown in Fig. 3. Thereafter all the stations in each region which had more than 40 years of record were selected. Moving "windows" of 5, 10, 15, 20, 25, 30, 35 and 40 years were passed through the data and the MAP and MMP were calculated for each of these windows. The difference between this short-term (window) mean and the long-term (>40 year) mean was then expressed as a percentage of the long-term mean. The 80 and 90 percentile limits of these errors were established for each of the windows, i.e. the percentage error, both positive and negative that may be expected to be exceeded on 20 and 10 per cent of occasions, respectively, if a shorter record were to be used to estimate the long-term mean. A table similar to that shown in Table 1 was compiled for each of the 24 major Welding and Havenga (1974) regions. These tables are presented in Dent *et al.* (1989).

Summary and conclusions

It is evident from Fig. 3 that the length of record required varies from region to region. As may be expected, the length of record required in the more arid regions of southern Africa is longer than that required in the wetter regions. However, it should be noted that the expression of the possible error in per cent may be somewhat misleading. For example when consulting Table 1 the percentage possible error at the 90 per cent confidence level of 13 per cent for the 10-year mean may seem high. However, this possible error should be seen in the light of the improvement in the estimate of the mean that can be expected from a 40-year record. The improvement would only be 9 per cent. The corresponding reduction in the spatial and altitudinal coverage of stations would, on the other hand, be drastic in most cases.

The use of percentages to express the "error in the estimate" has a disadvantage in that a 10 per cent error in MAP in the Karoo may only be equivalent to 20 mm whilst a 10 per cent error in the Drakensberg may be 150 mm. The only "solution" to this dilemma is to bring the above-mentioned points to the attention of the reader so that the results of these analyses are interpreted correctly. In some of the drier months the percentage errors show anomalous trends with respect to the record lengths. This "apparent error" may be attributed to the high variability of rainfall in the dry months.

After completion of the above-mentioned studies and having gained experience in mapping the statistics of rainfall it became apparent that the question of the minimum length of record was one which cannot be answered unequivocally. Generally the objective was to use an estimate which, at the 90 per cent confidence limits, yielded a probable error of less than 10 per cent.

The results of this study have many practical applications since the rainfall index MAP has numerous uses which have been outlined in the introduction to this paper. The major purpose of the study, however, was to provide, on a regional basis, guidelines for the minimum acceptable record length for rainfall stations whose data were used subsequently to map MAP and other rainfall statistics over southern Africa. The techniques and results of that mapping study are reported by Dent *et al.* (1989).

Acknowledgements

The Water Research Commission is acknowledged gratefully for funding this research. The following institutions are also acknowledged with gratitude: Department of Environment Af-

TABLE 1
UPPER LIMITS OF THE PERCENTAGE ERROR OF THE SHORT-TERM MEAN: AN EXAMPLE OF THE ERROR ANALYSIS FOR THE WELDING AND HAVENGA (1974) REGION 20B

Welding and Havenga region 20B		Possible percentage error made by using a short record, 412 station n												
		JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANN
5-YEAR MEAN AT	90% CONFD.	36,5	39,5	42,5	59,5	74,5	100,5	95,5	100,5	88,5	42,5	41,5	36,5	17,5
	80% CONFD.	27,5	30,5	32,5	45,5	59,5	79,5	79,5	76,5	71,5	33,5	31,5	27,5	13,5
	SAMPLE SIZE	24262	24247	24234	24096	24026	23974	23938	23972	24023	24097	24091	24115	23812
10-YEAR MEAN AT	90% CONFD.	24,5	25,5	28,5	46,5	51,5	80,5	75,5	80,5	58,5	29,5	27,5	24,5	12,5
	80% CONFD.	18,5	19,5	22,5	35,5	39,5	64,5	58,5	61,5	46,5	23,5	20,5	18,5	9,5
	SAMPLE SIZE	22202	22187	22174	22036	21966	21914	21878	21912	21963	22037	22031	22055	21752
15-YEAR MEAN AT	90% CONFD.	19,5	19,5	21,5	39,5	39,5	63,5	59,5	63,5	38,5	23,5	20,5	19,5	10,5
	80% CONFD.	14,5	15,5	17,5	31,5	30,5	49,5	45,5	50,5	29,5	18,5	15,5	14,5	7,5
	SAMPLE SIZE	20142	20127	20114	19976	19906	19854	19818	19852	19903	19977	19971	19995	19692
20-YEAR MEAN AT	90% CONFD.	16,5	16,5	17,5	35,5	34,5	51,5	50,5	51,5	31,5	19,5	16,5	15,5	8,5
	80% CONFD.	12,5	12,5	13,5	27,5	27,5	38,5	38,5	42,5	24,5	14,5	12,5	11,5	6,5
	SAMPLE SIZE	18082	18067	18054	17916	17846	17794	17758	17792	17843	17917	17911	17935	17632
25-YEAR MEAN AT	90% CONFD.	13,5	14,5	14,5	30,5	30,5	47,5	40,5	45,5	28,5	16,5	14,5	13,5	7,5
	80% CONFD.	10,5	10,5	11,5	24,5	23,5	34,5	31,5	36,5	21,5	12,5	10,5	9,5	5,5
	SAMPLE SIZE	16022	16007	15994	15856	15786	15734	15698	15732	15783	15857	15851	15875	15572
30-YEAR MEAN AT	90% CONFD.	11,5	11,5	13,5	27,5	27,5	41,5	35,5	41,5	23,5	13,5	12,5	11,5	6,5
	80% CONFD.	9,5	8,5	10,5	21,5	21,5	31,5	28,5	33,5	17,5	10,5	9,5	8,5	4,5
	SAMPLE SIZE	13962	13947	13934	13796	13726	13674	13638	13672	13723	13797	13791	13815	13512
35-YEAR MEAN AT	90% CONFD.	10,5	10,5	11,5	24,5	24,5	36,5	32,5	36,5	20,5	11,5	10,5	9,5	5,5
	80% CONFD.	7,5	7,5	8,5	19,5	19,5	27,5	25,5	29,5	14,5	8,5	7,5	7,5	3,5
	SAMPLE SIZE	11902	11887	11874	11736	11666	11614	11578	11612	11663	11737	11731	11755	11452
40-YEAR MEAN AT	90% CONFD.	9,5	8,5	10,5	21,5	22,5	31,5	28,5	33,5	19,5	10,5	8,5	8,5	4,5
	80% CONFD.	7,5	6,5	7,5	17,5	17,5	23,5	21,5	27,5	14,5	7,5	6,5	6,5	3,5
	SAMPLE SIZE	9842	9827	9814	9676	9606	9554	9518	9552	9606	9677	9671	9695	9392

fairs, South African Weather Bureau; Department of Agriculture and Water Supply; Department of Environment Affairs, South African Forestry Research Institute; South African Sugar Association Experiment Station; Computing Centre for Water Research; and University of Natal, Computer Services Division.

References

BEDIENT, PB, HUBER, WC and HEANEY, JP (1978) Drainage density as an index of watershed development. *Journ.Irrig.Drain.Div.* **104** 373-387.

BOUGHTON, WC (1981) Rainfall variability in rainfall-runoff modelling. Civil Engineering Transactions, The Institution of Engineers, Australia. 68-73.

BROCKETT, GM (1982) Personal communication. Cedara Agricultural Research Institute, Cedara.

BROMLEY, K (1980) Seasonal rainfall prediction. *The Farmer, Zimbabwe* **51** 15-16.

DENT, MC, LYNCH, SD and SCHULZE, RE (1989) Mapping mean annual and other rainfall statistics over southern Africa. Dept. of Agricultural Engineering, University of Natal. ACRU Report 27. Water Research Commission, Pretoria. Report No. 109/1/89.

DEPARTMENT OF AGRICULTURE AND FISHERIES (1981) Guide to Soil Conservation Committee members, Natal region. Cedara Agricultural Research Institute, Cedara. Mimeographed.

DUNNE, T and LEOPOLD, LB (1978) *Water in Environmental Planning.* Freeman, San Francisco 71-72.

DYER, TGJ (1976) Expected future rainfall over selected parts of South Africa. *South African Journ.Sci.* **72** 237-239.

DYER, TGJ and GOSNELL, JM (1978) Long-term rainfall trends in the South African sugar industry. *Proc. South African Sugar Technol. Assn.* 237-239.

DYER, TGJ and TYSON, PD (1977) Estimating above and below normal rainfall periods over South Africa 1972-2000. *Journ.Appl.Meteor.* **16** 145-147.

HALL, MJ (1976) Dendroclimatology, rainfall and human adaption in the later Iron Age in Natal and Zululand. *Ann. Natal Museum* **22** 693-703.

JONES, RI (1982) Personal communication. Cedara Agricultural Research Institute, Cedara.

JONES, RI, ARNOTT, JK and KLUG, JR (1980) Fodder production planning. Report N13/80. Cedara Agricultural Research Institute, Cedara. pp 27 and appendices.

KNISEL, WG, RENARD, KG and LANE, LJ (1979) Hydrologic data collection: How long is long enough? *Proceedings of the Speciality Conference on Irrigation and Drainage in the Nineteen Eighties,* Am.Soc.Civ.Eng. Albuquerque, New Mexico.

LOUW, WJ (1982) Rawson se siklus. *Weerburo Nuusbrief* **373** 103-105.

MATALAS, NC and JACOBS, B (1964) A correlation procedure for augmenting hydrologic data. Statistical studies in hydrology. Geological Survey Professional Paper 434-E. United States Govt. Printing Office, Washington, DC.

MIDGLEY, DC and PITMAN, WV (1969) Surface water resources of South Africa. University of Witwatersrand, HRU Report 2/69.

MIDGLEY, DC and PITMAN, WV (1978) A depth-duration-frequency diagram for point rainfall in Southern Africa. University of Witwatersrand, HRU Report 2/78.

NEVILL, E (1908) The rainfall of Natal. *Agric.Journ.Natal* **11** 531-1533.

NIELD, RE and BOSHELL, F (1976) An agroclimatic procedure and survey of the pineapple production potential of Columbia. *Agric.Meteor.* **17** 81-92.

PITMAN, WV (1973) A mathematical model for generating monthly river flows from meteorological data in South Africa. University of Witwatersrand, HRU Report 2/73.

PITMAN, WV (1980) Analysis of large area storms in SWA/Namibia. University of Witwatersrand, HRU Report 1/80.

PITMAN, WV, MIDDLETON, BJ and MIDGLEY, DC (1981) Surface water resources of South Africa. University of Witwatersrand, HRU Report 9/81.

PITMAN, WV and STERN, JA (1981) Design flood determination in SWA/Namibia. University of Witwatersrand, HRU Report 14/81.

RAWSON, HE (1908) The anticyclonic belt of the Southern Hemisphere. *Quart.Journ.Royal.Meteor.Soc.* **34** 165-185 (Cited by Dyer and Tyson, 1977).

SCHÖNAU, APG (1982) Personal communications. Institute for Commerical Forestry Research, University of Natal, Pietermaritzburg.

SCHULZE, RE (1983) *Agrohydrology and Climatology of Natal.* Water Research Commission, Pretoria. pp 138.

SCHULZE, RE (1984) An assessment of surface water resources of Natal. Dept. Agricultural Engineering, University of Natal, Pietermaritzburg, ACRU Report 20. Natal Town and Regional Planning, Pietermaritzburg. Report 63. pp 44.

SEEBER, RJ (1983) A Southern African extreme-value model for the determination of the cumulative distribution of point rainfall rate. NITR, CSIR, Pretoria. Report TEL: 192. pp 148.

WELDING, MC and HAVENGA, CM (1974) The statistical classification of rainfall stations in the Republic of South Africa. *Agrochemophysica* **6** 5-24.

WHITMORE, JS (1962) On evidence of "persistence" in annual rainfall in the Republic of South Africa. Dept. Water Affairs, TR 25. pp 11.

WORLD METEOROLOGICAL ORGANISATION (1974) *Guide to Hydrological Practices.* WMO, Geneva, Switzerland No. 168, edn. 3.

ZUCCHINI, W and HIEMSTRA, LAV (1984) Part 1: Augmenting hydrological records. Water Research Commission, Pretoria. Report No. 91/3/84.