

# Classifying rainfall districts: a south Western Australian study

Lynda E. Chambers

Bureau of Meteorology Research Centre, Australia

(Manuscript received February 2000; revised November 2000)

**A discussion of the historical rainfall districts in Western Australia is given and a number of statistical techniques applied to assess the appropriateness of these districts, particularly in the southwest. The districts are often used to summarise climate variables, such as rainfall, on a regional scale. Suggestions are made on alterations to the current district boundaries to give a 'better' regional summary.**

## Introduction

The largest State in Australia, Western Australia, covers an area of approximately 2.5 million square kilometres, extending from latitude 14°S to 35°S and from longitude 113°E to 129°E. The climate of Western Australia is influenced by its large size and latitudinal position. In the north of the State the climate is tropical and dominated by the Australian Monsoon, the south predominantly Mediterranean and the central interior arid. The effect of topography is less pronounced than in most other Australian States, resulting in a general decrease in rainfall with increasing distance from the coast. The major exception to this is the southwest coast where a rapid increase in rainfall occurs from the coastal plain to the higher land at the top of the Darling Range, followed by a marked decrease to the east.

The number of rainfall stations currently recording rainfall in Western Australia and reporting to the Australian Bureau of Meteorology is just under 1000. The length, quality and variability of individual rainfall records often differ substantially and a method of summarising 'like' stations, while taking these factors into account, is highly desirable. One method which is commonly used is based on the Meteorological Districts implemented in 1913 by the Bureau of Meteorology (Jones and Beard 1998; this paper also

gives the rationale for the use of district rainfall). However, as will be discussed in the next section, the development of these districts in Western Australia was not as 'scientific' as that of the eastern States of Australia. This paper explores a number of alternative subdivisions of Western Australia with respect to rainfall, particularly in the southwest corner, and discusses the appropriateness of the current district boundaries.

## History of current rainfall districts

The first recorded meteorological observations in Western Australia were made at Perth around the time of first European settlement in 1829. In 1876 the Meteorological Branch was established as part of the Lands and Surveys Department. Meteorological observations continued to be made under the direction of the Surveyor General until 1893, when the Meteorological Branch became part of the Registry Branch of the Colonial Secretary's Department (Bureau of Meteorology 1960). This was a short association as, in 1896, meteorological measurements became part of the duties of the Astronomer, Mr W.E. Cooke (Bureau of Meteorology 1960). The responsibility for meteorological services in Western Australia was taken over by the Commonwealth Government in 1908, and so remains today.

The early association of the Meteorological Branch with the Lands and Survey Department had

---

*Corresponding author address:* Lynda E. Chambers, Bureau of Meteorology Research Centre, GPO Box 1289K, Melbourne, Vic. 3001, Australia.

a large influence on the boundaries of the current rainfall districts. The original division of Australia into meteorological sections is discussed in Hunt (1910). The primary divisions were made along State boundaries and secondary divisions determined 'as far as a knowledge of physiographical and other factors would permit' (Hunt 1910). However, in the case of Western Australia the physical features of the State were 'neither sufficiently pronounced nor known to enable scientific boundaries to be yet determined' (Hunt 1910) and the State department was left to empirically determine the divisions. To a large extent the isohyets were used as approximate divisions as little was known of the topography of the State (Hunt 1910). However, earlier reports (Cooke 1897) indicate that the basis for classifying the rainfall stations was the official land districts (although these had some relationship to rainfall, in that rainfall was a limiting factor in the expansion of the Colony and the suitability of the land to various farming or other activities).

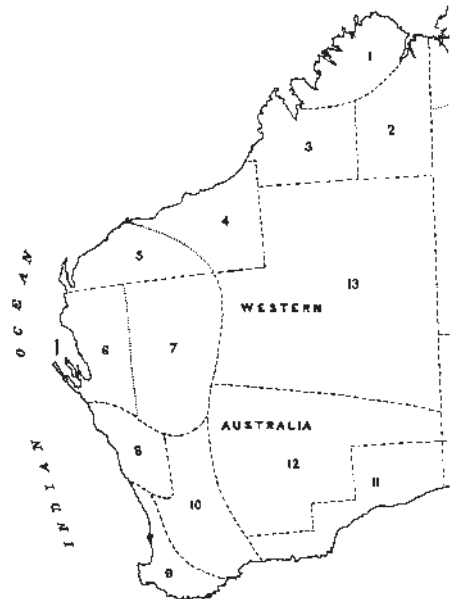
The resulting 13 (Fig. 1) divisions based on Hunt (1910) strongly resemble the currently used meteorological districts (Fig. 2) with a few minor differences, the most notable being the extension of District 9A (current version) to include coastal areas from Albany to east of Esperance.

In the last 30 years there has been considerable interest in (winter) rainfall trends in the southwest of Western Australia (e.g., Gentili 1971; Wright 1971, 1974a,b; Pittock 1983; Sadler et al. 1988; Nicholls and Lavery 1992; Yu and Neil 1993). Since much of the work on the climate of Western Australia has concentrated on this region the majority of this paper will focus on classifying rainfall districts in the southwest of the State. For the purposes of this study the southwest of Western Australia is defined as the region bounded to the north by latitude 27°S and to the east by longitude 122°E (as in Wright (1974a)).

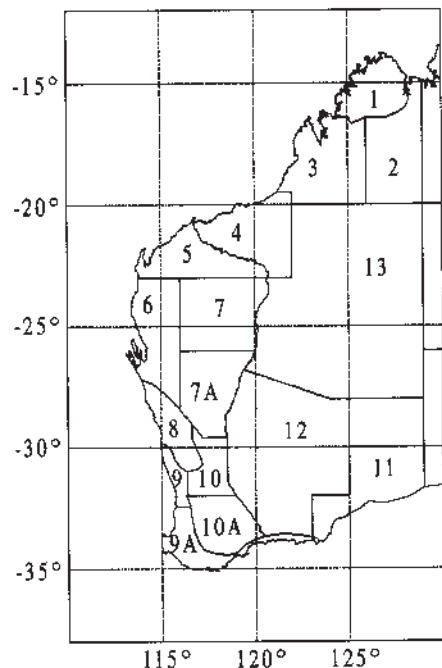
## Data and method of analysis

Two datasets are used in this paper; gridded monthly rainfall completely covering the southwest corner and monthly rainfall data from 67 'high quality' stations scattered throughout the southwest corner. The gridded rainfall dataset was created from the Australian Bureau of Meteorology, National Climate Centre's 0.25 degree spatial dataset (Jones and Weymouth 1997), and has a 0.5 degree spatial resolution. The gridded dataset uses data from January 1950 to 1998. The 'high quality' monthly rainfall data are based on a subset of the whole Australia dataset which includes 191 stations identified

**Fig. 1** Historical meteorological divisions of Western Australia based on Hunt (1910).



**Fig. 2** Current meteorological divisions of Western Australia.



by Lavery et al. (1992). Additional stations were added to the dataset to improve spatial coverage (Lavery et al. 1997). For the southwest corner of Western Australia there were close to 100 years of monthly rainfall data available from 63 rainfall stations (Lavery et al. 1997).

To improve coverage in areas of interest to water resource managers, four additional stations were added to the analysis: Perth Airport, Perth Regional Office, Jarrahdale and Mandurah. Station details are given in Table 1 and the station distribution is shown in Fig. 3.

Two methods of developing rainfall ‘districts’ are explored in this paper: principal component analysis (PCA) and cluster analysis (CA); methodological details of these methods can be found in Morrison (1990) and Everitt (1993). Both methods have been widely cited in the meteorological literature (e.g., Wright 1974a; Drosowsky 1993; Gong and Richman 1995; Mason 1998).

Prior to the principal component analysis, the monthly rainfall datasets were standardised at each

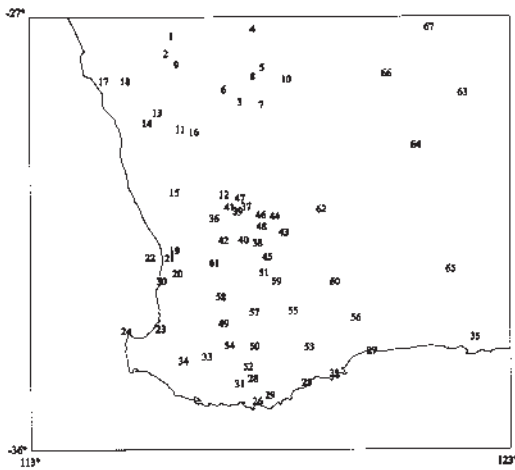
**Table 1. High quality Australian rainfall station information (taken from Lavery et al. (1997) and with three additional stations). The bracketed numbers in the first column are the station numbers referred to in the text.**

<i>Station number</i>	<i>Latitude</i>	<i>Longitude</i>	<i>Elevation (m)</i>	<i>First year</i>	<i>Name</i>
06048 (1)	-27.38	115.97	400	1904	Twin Peaks (Twin Peaks)
06055 (2)	-27.75	115.83	300	1904	Woolgorong (Woolgorong)
07011 (3)	-28.78	117.37	400	1903	Paynes Find (Burnerbinmah)
07014 (4)	-27.25	117.65	500	1903	Cue (Coodardy)
07057 (5)	-28.07	117.85	426	1895	Mount Magnet (Mt Magnet P.O.)
07063 (6)	-28.53	117.03	350	1909	Yalgoo (Muralgarra Stn)
07090 (7)	-28.85	117.83	450	1905	Paynes Find (Wydgee)
07095 (8)	-28.23	117.65	400	1899	Mount Magnet (Yoweragabbie)
07096 (9)	-27.98	116.03	300	1910	Yalgoo (Yuin)
07197 (10)	-28.28	118.32	400	1896	Mount Magnet (Challa)
08013 (11)	-29.35	116.17	320	1907	Perenjori (Bowgada)
08066 (12)	-30.70	117.05	300	1908	Wongan (Kokardine)
08079 (13)	-29.03	115.65	290	1907	Manarra
08088 (14)	-29.20	115.43	153	1896	Mingenew (Mingenew P.O.)
08091 (15)	-30.65	116.00	203	1898	Moora (Moora Shire)
08106 (16)	-29.37	116.42	260	1907	Perenjori (Perangery)
08141 (17)	-28.38	114.45	110	1907	Northampton (Willi Gully North)
08147 (18)	-28.32	114.95	270	1910	Nabawa (Yuna)
09021 (19)	-31.93	115.97	15	1944	Belmont (Perth Airport)
09023 (20)	-32.33	116.05	230	1882	Serpentine (Jarrahdale)
09034 (21)	-31.95	115.87	19	1876	Perth (Perth Regional Office)
09038 (22)	-32.00	115.50	46	1880	Rottnest (Rottnest Island L.H.)
09503 (23)	-33.48	115.73	37	1898	Capel (Boyanup P.O.)
09519 (24)	-33.53	115.02	97	1904	Busselton (Cape Naturaliste)
09520 (25)	-34.60	118.73	110	1897	Albany (Cape Riche)
09551 (26)	-35.02	117.75	10	1902	Albany (Grassmere)
09557 (27)	-33.95	120.12	15	1907	Hopetoun (Hopetoun P.O.)
09561 (28)	-34.48	117.63	262	1905	Mount Barker (Kendenu P.O.)
09564 (29)	-34.95	117.88	22	1904	Albany (King River)
09572 (30)	-32.50	115.77	15	1889	Mandurah (Mandurah Park)
09591 (31)	-34.63	117.38	300	1900	Mount Barker (Pardelup)
09594 (32)	-34.43	119.35	30	1904	Bremer Bay (Peppermint Grove)
09616 (33)	-34.10	116.67	200	1903	Boyup Brook (Westbourne)
09619 (34)	-34.15	116.20	251	1901	Manjimup (Walgarrup)
09739 (35)	-33.65	122.32	200	1902	Esperance (Bankina Farm)
10024 (36)	-31.17	116.85	305	1910	Goomalling (Casuarina Vale)
10032 (37)	-31.00	117.40	360	1908	Wyalkatchem (Cowcowing)
10037 (38)	-31.73	117.77	269	1903	Kellerberrin (Cuttening)
10039 (39)	-31.02	117.20	200	1910	Dowerin (Doodarding Well)
10041 (40)	-31.63	117.45	250	1910	Tammin (Doongin Peak)
10045 (41)	-31.02	117.13	300	1908	Dowerin (Ejanding)
10091 (42)	-31.63	117.02	300	1900	Meckering (Meckering P.O.)
10092 (43)	-31.48	118.28	315	1903	Merredin (Merredin S.C.)

Table 1. Continued.

Station number	Latitude	Longitude	Elevation (m)	First year	Name
10112 (44)	-31.18	118.10	292	1907	Nungarin (Nungarin P.O.)
10123 (45)	-32.00	117.93	427	1910	Bruce Rock (The Granites)
10126 (46)	-31.12	117.80	290	1910	Trayning (Trayning P.O.)
10133 (47)	-30.83	117.32	350	1909	Koorda (Walk Walkin)
10149 (48)	-31.35	117.80	300	1903	Trayning (Codg Codgin)
10505 (49)	-33.35	117.03	272	1891	Arthur River (Arthur River Comp.)
10525 (50)	-33.85	117.63	328	1891	Broomehill (Broomehill P.O.)
10536 (51)	-32.33	117.87	295	1910	Corrigin (Corrigin P.O.)
10537 (52)	-34.30	117.55	255	1891	Cranbrook (Cranbrook P.O.)
10541 (53)	-33.85	118.82	350	1907	Jerramungup (Dalgleish)
10582 (54)	-33.83	117.15	305	1885	Kojonup (Kojonup Composite)
10592 (55)	-33.10	118.47	286	1909	Lake Grace (Lake Grace)
10611 (56)	-33.28	119.78	310	1932	Lake King (Truro)
10636 (57)	-33.17	117.67	400	1907	Dumbleyung (The Oaks)
10658 (58)	-32.83	116.98	300	1905	Cuballing (Wonnaminta)
10668 (59)	-32.47	118.15	280	1910	Kondinin (Koorikin)
10670 (60)	-32.50	119.37	310	1906	Hyden (Lake Carmody)
10795 (61)	-32.12	116.87	200	1909	Beverley (Avondale Res. Stn.)
12011 (62)	-30.98	119.12	400	1912	Southern Cross (Bullfinch)
12013 (63)	-30.75	121.77	380	1897	Kalgoorlie (Bulong)
12052 (64)	-29.68	121.03	426	1897	Menzies (Menzies P.O.)
12065 (65)	-32.20	121.78	277	1897	Norseman (Norseman P.O.)
12067 (66)	-28.20	120.43	400	1911	Sandstone (Pinnacles)
12108 (67)	-27.13	121.33	500	1911	Wiluna (Wonganoo)

Fig. 3 Locations of the 'high quality' rainfall stations. Note the sparseness of stations in the east and the relatively low number of coastal stations.



grid (or station) point by removing the monthly mean and dividing by the monthly standard deviation, as described in Drosdowsky (1993). PCA was

performed in the S-mode (Drosdowsky 1993), with the EOFs rotated using the Varimax orthogonal rotation (Richman 1986). The use of raw, square root or cube root transformed rainfall data (not shown) gave very similar results to that of the standardised monthly rainfall with minor differences in the magnitude of the loadings.

Cluster analysis of the 'high quality' station data, hereafter referred to as the station data, was achieved using a number of different distance measures (linkages). Cluster analysis is a useful tool for dealing with classification problems when it is not known beforehand from which subgroup the observations originate. The methods used in this paper involved hierarchical clustering, i.e. initially each object is assumed to form a separate group with objects or groups close to one another being successively merged. Some additional details will be given in the Results section, and further information on cluster analysis can be found in Everitt (1993).

If similar station groupings can be obtained using principal component analysis and cluster analysis then there is the suggestion that the resulting regions are natural divisions of rainfall in southwestern Australia (Drosdowsky 1993); this criterion will be used to obtain 'revised' rainfall district boundaries.

Where ‘winter’ rainfall is mentioned in this paper, we are referring to monthly rainfall totals accumulated over the period May to October. Rainfall in the southwest is winter dominated, and it is rainfall during this period which is most important for agriculture, particularly cereal crops (Wright 1974a).

## Results

### Principal component analysis

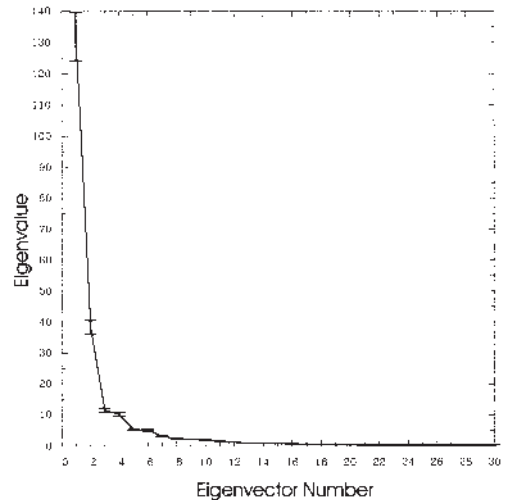
The eigenvalue spectrum (Fig. 4) for the monthly gridded rainfall data suggests breaks at four and at six components. In order to retain more of the variance (and a larger number of ‘districts’), the larger number of components was selected. The six rotated components (Fig. 5) reveal a reasonably simple structure, with minimal overlap. Taking a loading of 0.55 as the district boundary, the EOFs neatly classify the southwest into six largely non-overlapping regions (Fig. 6).

The first component has highest loadings in the northeast, and accounts for around 21 per cent of the total rainfall variance. The highest loadings occur in a region of variable rainfall, and fall almost entirely within the current District 12. The second component, with highest loadings in the southwest corner, explains around 16 per cent of the variance. The region of highest loadings within this component encompasses the current District 9, most of 9A (except the coastal strip to the east of Albany), and the western side of District 10A. This region has a Mediterranean climate with winter rainfall maxima.

The current District 7 and the northern half of District 8 are represented by the third principal component (explaining 14.6 per cent of the variance). The fourth component accounts for 12.1 per cent of the variance and has its highest loadings along the south-eastern coastline. This principal component incorporates the eastern sections of District 9A and 10A and the southern region of District 12. The fifth principal component encompasses most of District 10 and the southern section of District 8, explaining around 13 per cent of the variance. The final principal component, component 6, accounts for approximately 11 per cent of the variance and has its highest loadings in a region surrounding Kalgoorlie. This region lies entirely within the current District 12.

The northwest/southeast elongation of the districts is most likely a reflection of the atmospheric dynamics. Both Jones and Trewin (2000) and Seaman (1982) comment that atmospheric variability is rarely isotropic, with mid-latitude temperature (and geopotential) covariance functions having the slowest decay in a northwest/southeast direction, as appears to be the case with rainfall in the present study.

**Fig. 4** Scree plot of eigenvalue versus eigenvector number for the first 30 eigenvectors of the correlation matrix of the standardised monthly anomalies of the gridded rainfall dataset. Also shown are the estimated error bars for the eigenvalues based on the criterion of North et al. (1982).



### Cluster analysis

The ‘Splus’ agglomerative hierarchical clustering method ‘agnes’ (S-PLUS 2000) was used to define clusters of rainfall stations. The algorithm commences by computing the dissimilarity matrix (a measure of the ‘difference’ between the rainfall stations), if one was not provided. Initially each rainfall station is considered as a separate cluster. The two clusters with the smallest between-cluster dissimilarities are then merged, and the dissimilarity between the new cluster and all the remaining clusters computed. These steps are repeated until the desired number of clusters is obtained.

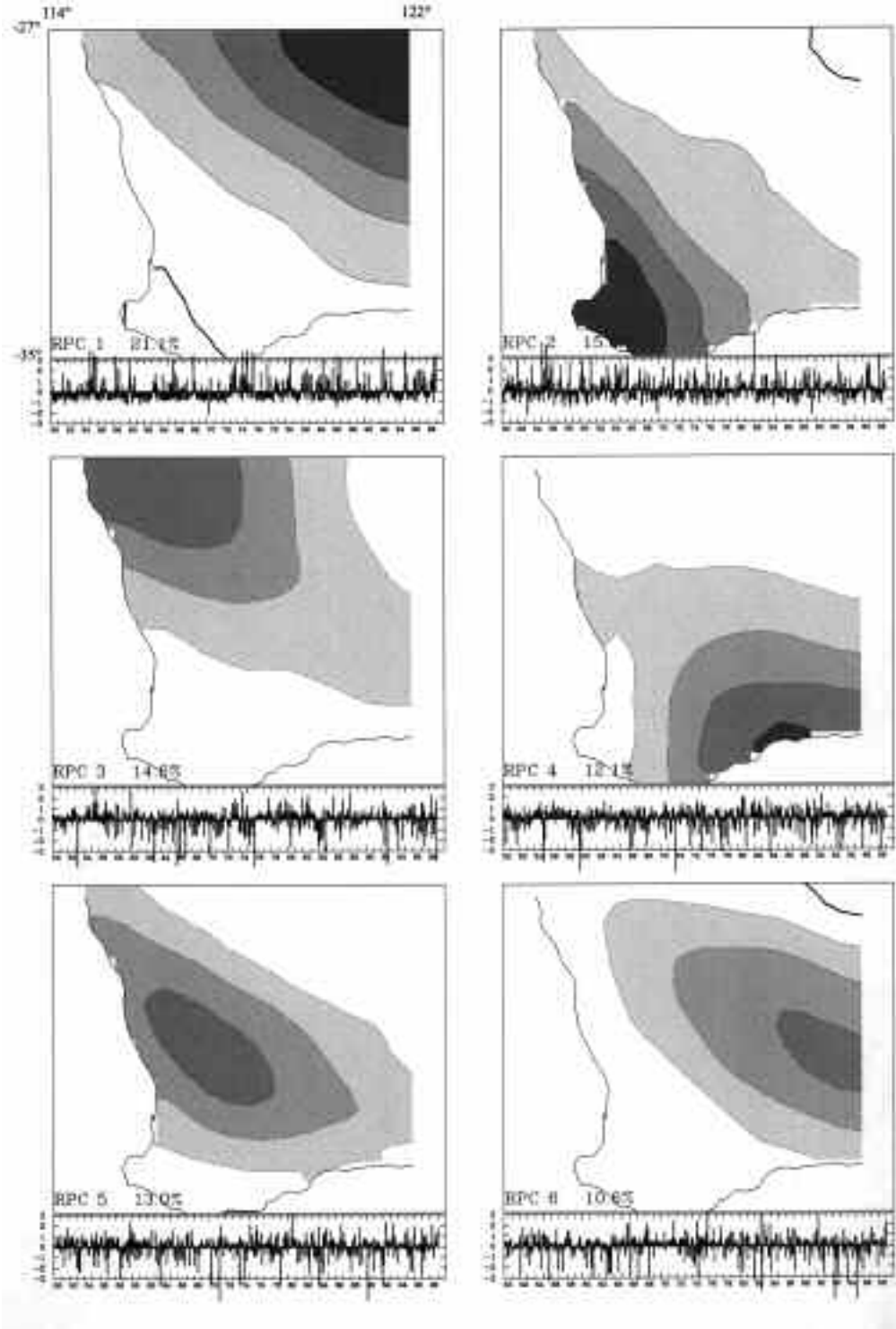
In this paper the Euclidean distance measure is used to calculate the dissimilarity  $d$  between rainfall stations or clusters of stations. The dissimilarity between rainfall locations  $i$  and  $j$  is then given by

$$d(i,j) = \sqrt{\sum_{f=1}^p (x_{if} - x_{jf})^2}$$

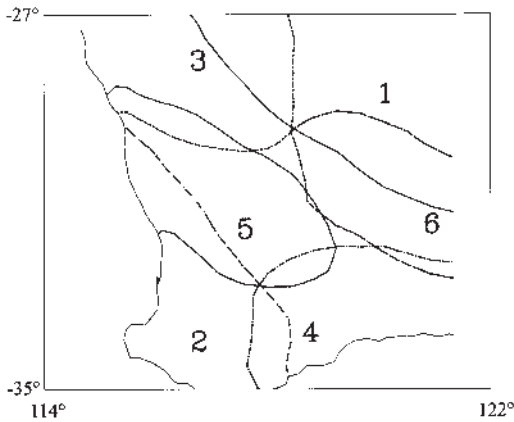
where  $p$  is, in this case, the number of time points (months or seasons).

There are many ways of defining new clusters based on the dissimilarities, four of which were considered in this paper: Ward’s, single link, complete link and average link (Everitt 1993; S-PLUS 2000). Ward’s method is also known as the sum of squares method and the trace method, and selects groups for

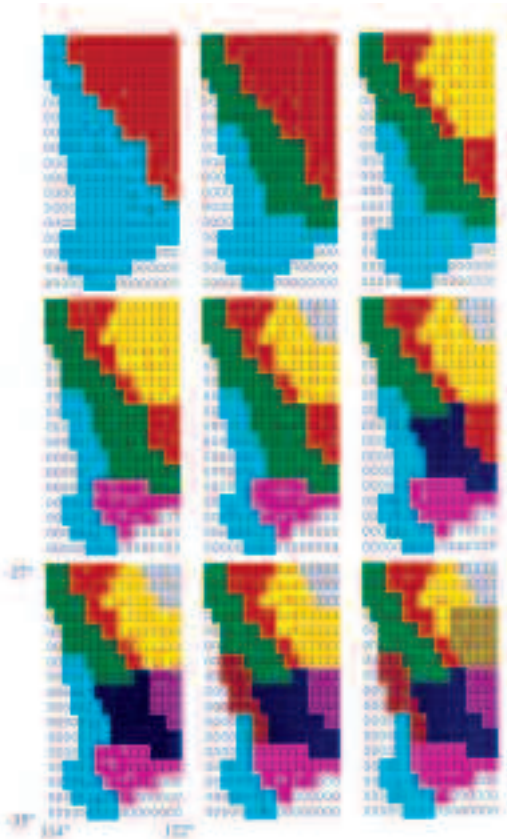
Fig. 5 Spatial pattern of loadings and associated scores (time series) of the first six gridded rainfall VARIMAX rotated components of the standardised monthly anomalies for the dataset. Contour interval is 0.2, with zero contour heavy, negative contours dashed and areas above +0.2 and below -0.2 shaded.



**Fig. 6** The six regions defined by the principal component analysis, with boundaries specified by the 0.55 loading.



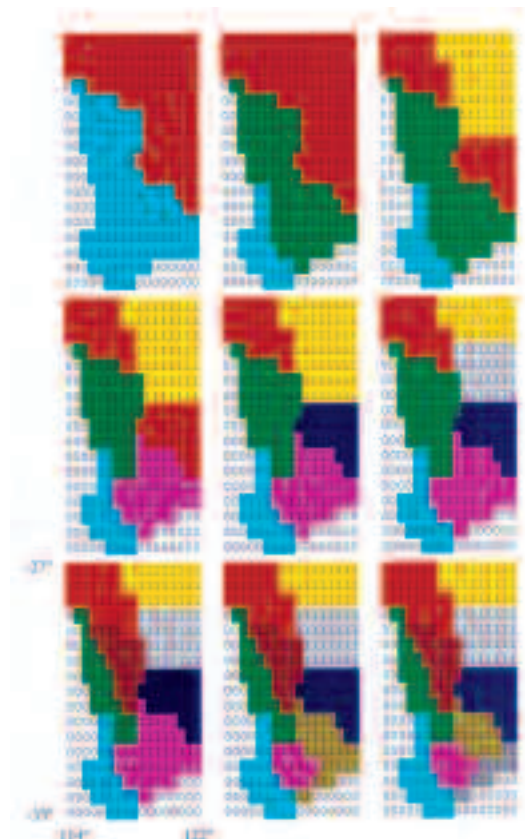
**Fig. 7** Clusters of raw gridded rainfall data, with the number of clusters varying from 2 to 10, determined using Ward’s method.



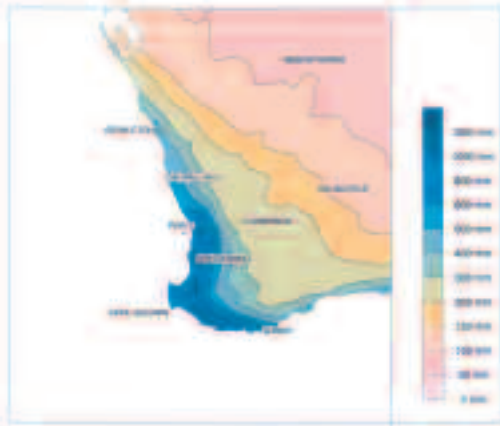
merging based on the smallest increase in the sum of within-group squares. The single link method is also known as the nearest neighbour method, with the ‘closest’ groups being merged based on the smallest distance between any two members from different groups. The complete link method is also known as the compact or farthest neighbour method. Groups are merged, under this method, based upon minimising the largest distance between any two members from different groups. The average link method is based upon the average distance between the objects in one group and the objects in the other group.

Cluster analysis was initially performed on the gridded monthly rainfall data described above. Ward’s method was used with the objective being to determine the effect of both standardising the variables and varying the number of clusters. The clusters obtained are shown in Fig. 7 for the raw (unstandardised) monthly rainfall data and in Fig. 8 for the rainfall anomalies (standardised data). When the number of

**Fig. 8** Clusters of standardised (‘normalised’) gridded rainfall data, with the number of clusters varying from 2 to 10, determined using Ward’s method.



**Fig. 9** 'Winter' (May to October) mean rainfall (mm) in southwestern Australia using historical rainfall totals from the period 1913 to 1995.



clusters is low (less than six), the positions of the clusters are similar, with the only real differences being the cluster boundary placements. However, once more than five clusters are formed, differences occur in the locations of the clusters between the two sets. When the number of clusters is less than five the cluster locations, for both the standardised and raw rainfall data, are very closely tied to the overall winter rainfall pattern (Fig. 9). However, when five or more clusters are used it is harder to relate the additional clusters to observed rainfall patterns in Fig. 9.

Winter rainfall data from 'high quality' rainfall stations were then used to illustrate the effect of both differing cluster sizes and method of determining cluster membership. All four methods (Ward's, single, complete and average linkage), using the Euclidean distance measure, grouped the same rainfall stations when forming two clusters. When applied to three clusters, all but Ward's method (Fig. 10) placed station 20 (Jarrahdale) in its own cluster, indicating that there may be data quality issues associated with this station. The station grouping for three clusters under Ward's method closely resembles that for four clusters using the other three methods, ignoring the single station (20) cluster. However, when considering four clusters Ward's method starts to deviate significantly from the other three methods by splitting the inland cluster in the northeast corner of the domain in two. This does occur in the other methods but only when the number of clusters is much greater.

The complete linkage method (Fig. 11 top left, for the case of six clusters) differs from the single linkage (Fig. 11 top right, six clusters) and average linkage (Fig. 11 bottom left, six clusters) methods, once six clusters are formed, by splitting the inland cluster into two regions rather than isolating station 32 (Bremer Bay) as a separate cluster. The average and single linkage methods give the same clusters up to the maximum number of clusters considered (nine).

The selection of the optimum number of clusters is more difficult than the number-of-factors problem (SAS 1989) with no single criterion being available to make an objective decision (Everitt 1993; Wolter 1992). A subjective choice of around ten per cent of the total number of stations has been suggested (Torok 1996) which results in six to seven clusters in this study (as shown in Fig. 11).

## Discussion

The choice of the dissimilarity measure can affect the final clustering solution. Comparing Fig. 11 bottom right, in which the similarity measure is based on correlations of monthly rainfall totals, and Fig. 10, where the dissimilarity measure is Euclidean, we see that there is more 'banding' in the former with the clusters more closely resembling the rainfall isopleths. Another difference between the two measures is the isolation of station 20 when using the Euclidean measure. Although the Euclidean dissimilarity measure is the mostly widely used in cluster analysis (Everitt 1993) its apparent failure to produce 'reasonable' clusters in this example is most likely due to the underlying assumption that the variables be uncorrelated. For many of the possible rainfall station pairs this assumption is invalid, particularly for neighbouring stations. Greater agreement between the currently defined rainfall districts, the Principal Component 'clusters' and the cluster analysis results occurs when the distances are based upon correlations rather than Euclidean measures.

Although the single linkage and average linkage methods gave the same station clusters (for up to nine clusters), these linkage methods tended to isolate individual stations more than the other methods considered. Consequently these linkage methods were deemed unsatisfactory for the rainfall analysis. The single, average and complete linkage methods have a tendency to cluster together, at a relatively low level, stations linked by a series of intermediates (Everitt 1993). This phenomenon is known as 'chaining', the failure to resolve relatively distinct clusters when a small number of stations lie between them.

Fig. 10 Rainfall clusters, with the number of clusters (K) varying from 3 to 8, using ‘winter’ rainfall station data and Ward’s method.

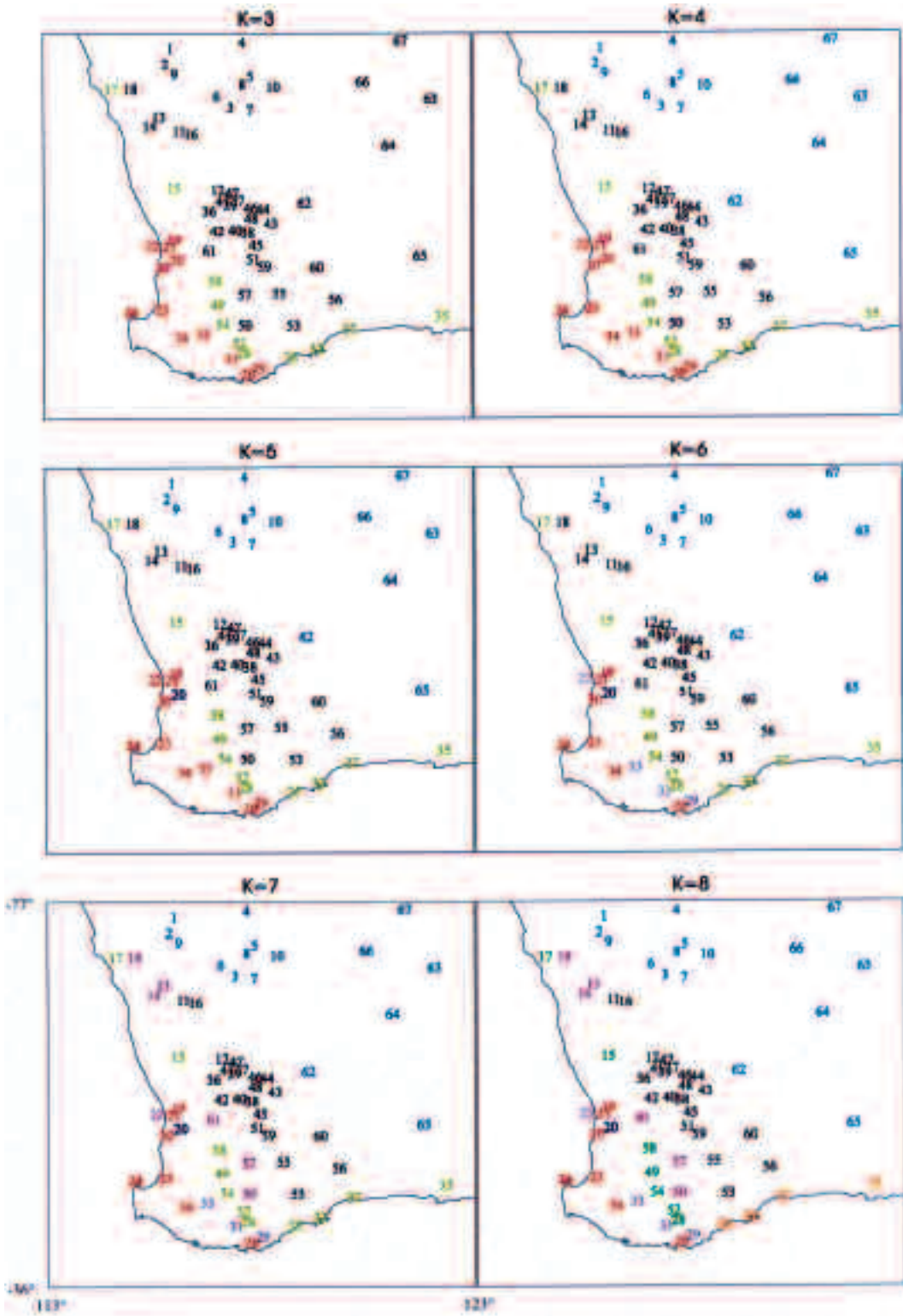
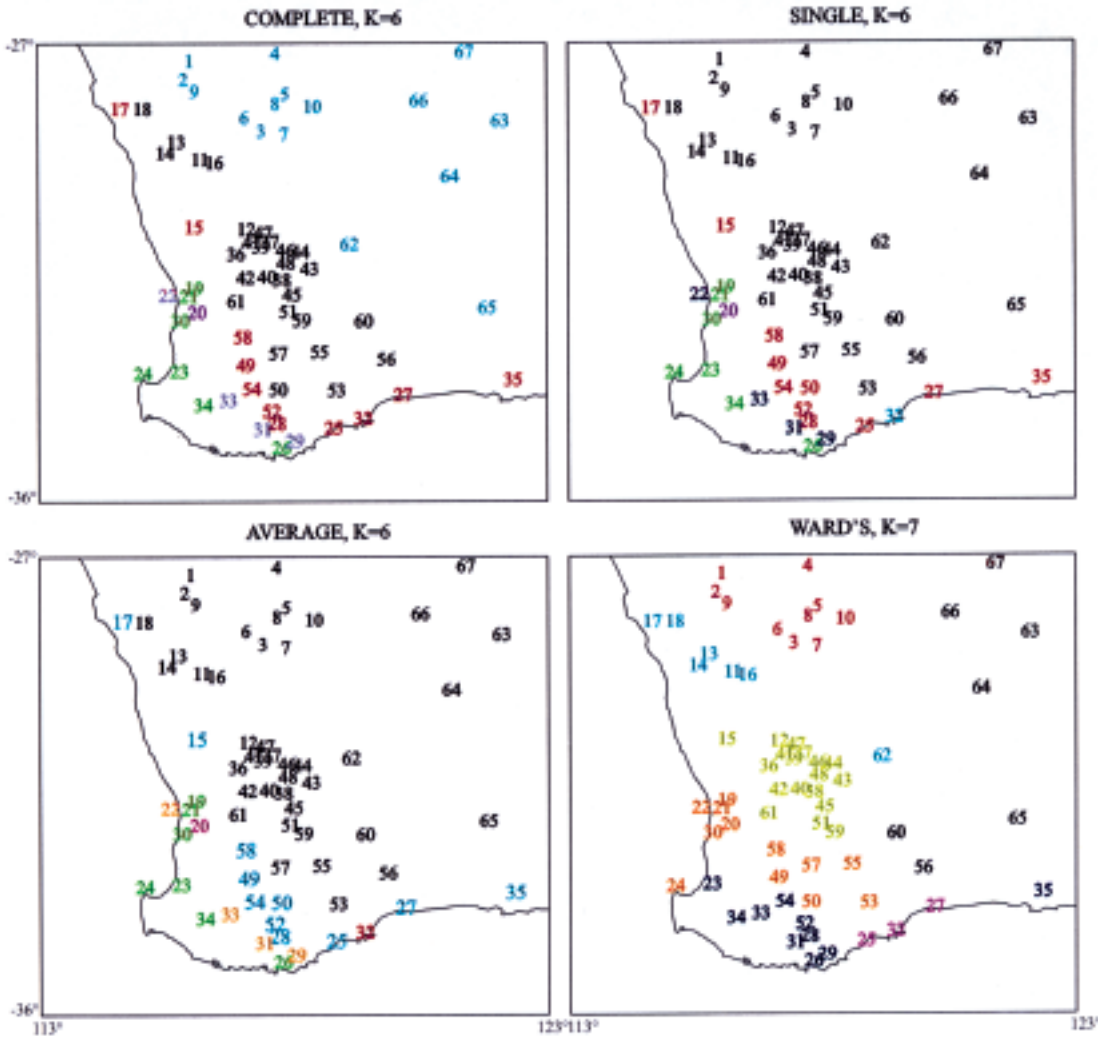


Fig. 11 Top left: ‘winter’ rainfall, six clusters, using station rainfall and the complete linkage method. Top right: ‘winter’ rainfall, six clusters, using station data and the single linkage method. Bottom left: ‘winter’ rainfall, six clusters, using station data and the average linkage method. Bottom right: ‘winter’ rainfall, seven clusters determined using Ward’s linkage method and correlation distances.



Ward’s linkage method is considered to have a bias towards producing ‘spherical’ clusters (Everitt 1993). However, when used in conjunction with the Euclidean distance measure, this was not evident in the district rainfall analysis (Fig. 10). Studies by others have suggested that the ‘best’ linkage methods include Ward’s, group average and the complete linkage (Everitt 1993). However, there is no real agreement on which method is the most appropriate overall. Gong and Richman (1995) suggest that Ward’s method is useful for the analysis of precipitation data,

particularly for the central and eastern North American region. As mentioned above, when the Euclidean distance measure is used, none of the linkage methods appeared to produce useful clusters.

The gridded rainfall analyses gave the most complete classification of the rainfall in the southwest of Australia, with greater and more even coverage. The non-gridded rainfall data suffered from very uneven station distribution with very few stations to the east and in coastal regions and an (over) abundance of stations from District 10A. This led to considerable dif-

ferences in the district and gridded cluster analysis results (Figs 7, 8 and 11 bottom right). The sparseness of rainfall stations to the east resulted in the formation of a single large cluster, over most of the eastern half of the analysis domain. When more data were provided in the form of a gridded analysis, the single large cluster disappeared.

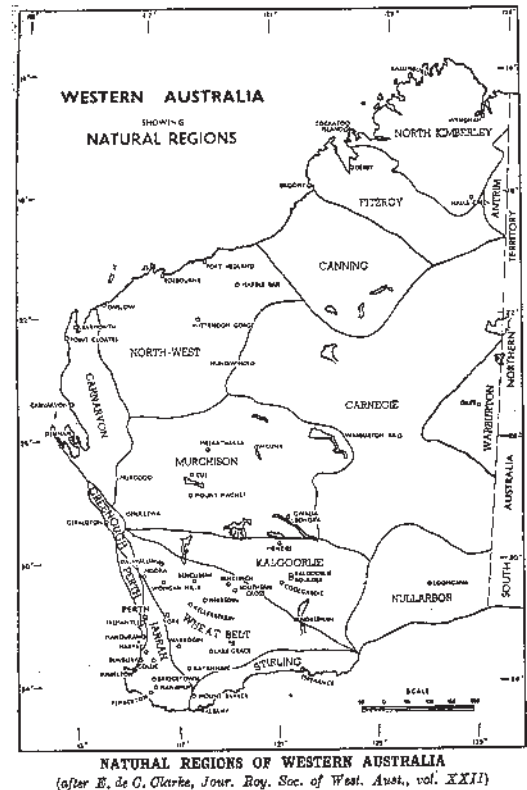
The raw and standardised gridded rainfall cluster analyses gave very similar results for a low number of cluster sizes (Figs 7 and 8), indicating that for large-scale summaries of rainfall either method could be used. However, for finer subdivisions real differences exist between the raw and standardised analyses. The 'better' description of rainfall is probably given by the standardised analysis. The clusters resulting from this analysis are very similar to those obtained from the principal component analysis (PCA). The main difference between the PCA and cluster analysis is that the divisions produced from the PCA have overlapping solutions. The problem is how to deal with the stations in the overlapping regions. Possible solutions include: excluding these stations from both clusters, entering them into more than one cluster, or placing them in the cluster which is 'closest' (e.g. has the closest cluster centroid).

### Comparison of results with previously proposed divisions

One of the earlier attempts at defining 'natural regions' in Western Australia was by Clarke in 1926. In his paper Clarke (1926) initially divided the State into two physical regions, a coastal strip from near Albany to Broome and a tableland in the interior of the State. Geological boundaries were then used to further subdivide the State into eight regions. Finally rainfall characteristics were used to create additional subdivisions, resulting in a total of 15 regions. Seven of these regions fall within the analysis domain of this study. A further region, 'Stirling', along the southern coastline, was later added in a report by Prider (1960) in the *Official Yearbook of Western Australia* (Fig. 12).

In 1959 a tourism and publicity booklet was published by the Western Australian Government on the weather and climate of Western Australia (Gentilli 1959). The first climatic division was based on three major climatic regions: north, with a wet summer and dry winter; southwest, with a wet winter and dry summer; and the intermediate zone, with scanty erratic rains. Further regional subdivisions were made using annual and monthly rainfall and temperature data. Seven of Gentilli's climatic regions (Fig. 13) occur within the analysis domain. There are considerable differences in the regions proposed by Clarke (Fig. 12) and Gentilli (Fig. 13). Both Clarke and Gentilli's

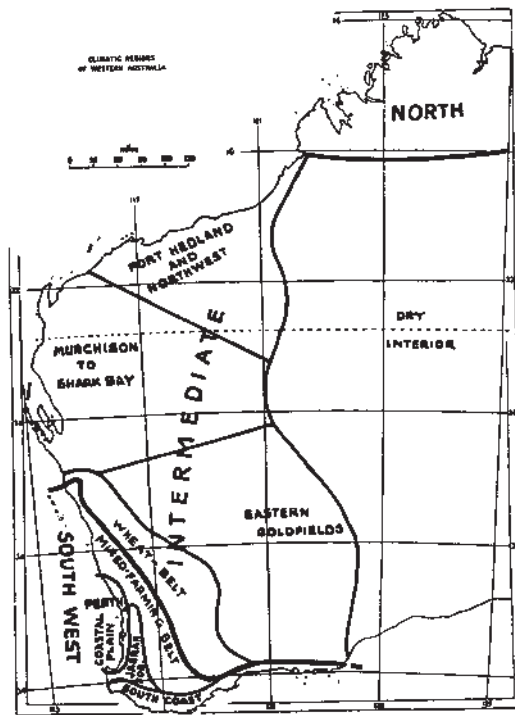
Fig. 12 'Natural' regions of southwestern Australia based upon Clarke (1926) (but adapted from Prider (1960)).



climate regions differ from the official Australian Bureau of Meteorology rainfall districts (Fig. 2) and from the rainfall analyses of this paper (for example, the PCA analysis, Fig. 6). It is possible that some of the differences could be attributed to the use of temperature or other non-rainfall data in defining regions, particularly in the case of Gentilli (1959), and also to the quality and quantity of rainfall data used in the studies. Neither Clarke (1926) nor Gentilli (1959) note the number or quality of rainfall stations used in their analyses.

In 1971 McBoyle (1971) published a proposed climatic classification of Australia that used factor analysis (Morrison 1990) to define regions based on twenty climatic variables including: temperature, relative humidity, precipitation, and the number of rain-days. When ten nationwide divisions are defined, the southwest corner is divided into three areas: northern, southwestern coastal and eastern coastal. When fourteen nationwide divisions are considered the analysis

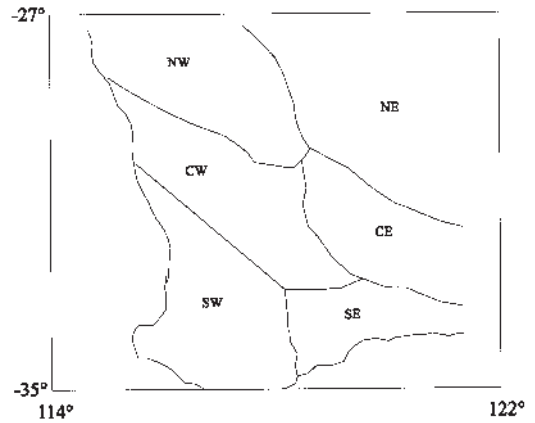
Fig. 13 Climatic regions of southwestern Australia adapted from Gentilli (1959).



domain is split into five areas with the northern section subdivided in the northeast and the southwestern section split in an east-west direction. A word of caution when interpreting McBoyle's results for the southwest region: as the number of stations used is very small in this region (eight) each station has, potentially, a very large impact on the division boundaries. With such a small number of stations one would have to be very careful of potential outliers, possibly in the form of low quality climatic records. There were simply not enough stations in the analysis domain in McBoyle's study to allow valid comparisons with the 'clusters' found in this study.

Other studies, such as Wright (1974a) and Yu and Neil (1993), define a small number of regions (three or four) based mainly on isohyets. Yu and Neil's and Wright's first three regions are similar and both resemble the three cluster solution in Fig. 7. Stern et al. (2000) applied a modified Köppen classification to Australian climate data resulting in seven regions in the southwest of Australia. Most of the regions proposed by Stern et al. differ greatly from those of this study, most likely due to the inclusion of many additional variables such as temperature and humidity.

Fig. 14 Proposed 'new' rainfall districts for southwestern Australia.



## Conclusions

Over the years many different rainfall/climate regions have been proposed for the southwestern corner of Australia. As this paper has shown the classification of individual rainfall stations or rainfall grid-points can vary greatly according to the method of analysis used. It is not surprising, therefore, that so many different classifications have been proposed.

The use of gridded rainfall data is recommended over individual station data as it minimises the impact of poor data density, reduces observation and representation errors, and avoids single station clusters. This is particularly important when searching for large-scale regions and when there may be questions of data quality. The gridded dataset also has the advantage of the 'data' being evenly spread over the entire area of interest while the rainfall station data, particularly in this study, tends to be unevenly distributed.

Principal component analysis (Fig. 6) and cluster analysis using Ward's method, combined with distances based on correlations, produced similar rainfall clusters (Figs 7 and 8, with six clusters). Although there were basic similarities between these methods and the official District boundaries there were also some major differences. The results of this study (Figs 6-8) suggest the following:

1. Official District 12 should be divided into northern and southern sections (Fig. 14; NE and CE);
2. Three rainfall regions occur along the western coastal areas (Fig. 14; NW, CW, SW) with one major difference from the official districts: the southern most region should not extend along the eastern coastline as it does in Official District 9A (Fig. 2);
3. There should be a distinct southeastern region, encompassing part of Official Districts 9A and 10A (Fig. 14; SE);

#### 4. Other minor modifications of official district boundaries.

Figure 14 graphically illustrates the proposed district boundaries.

Of the official districts, District 6, was not isolated as a distinct region in this study. This is most likely due to the small number of grid-points/stations from within this district that were used in this analysis. With either more datapoints in this region or by expanding the northern boundary of the study area, a further cluster resembling District 6 may result. However, rainfall in this region tends to be concentrated in northwest cloudbands which elongate NW-SE. Hence there is little reason to expect rainfall variations in District 6 to be different from those in District 7A.

A question which arises from this study is the physical interpretation of the revised rainfall regions. Rather than speculating on physical interpretations of why the regions are different it is suggested that detailed teleconnection studies should be conducted to aid the understanding of climate in this region.

As mentioned earlier, many of the current rainfall district boundaries were not determined scientifically, due to lack of knowledge of the regions. With nearly 90 years worth of extra data, perhaps now is the time to re-examine these district boundaries to provide more representative regional rainfall summaries.

## Acknowledgments

The author would like to acknowledge D.A. Jones, R. Fawcett and W. Drosdowsky, of the Bureau of Meteorology, for helpful comments made on early drafts of this paper.

## References

- Bureau of Meteorology 1960. Climate and Meteorology, in *Official Yearbook for Western Australia 1960 No. 2*, R.J. Little (ed.), Commonwealth Bureau of Census and Statistics, WA Offices, 34-47.
- Clarke, E. de C. 1926. Natural regions in Western Australia. *J. Roy. Soc. Western Australia, Vol. XII, No. 14*, 117-32.
- Cooke, W.E. 1897. Meteorological Observations made at the Perth Observatory and other places in Western Australia, during the year 1897, under the direction of W. Ernest Cooke, M.A., F.R.A.S., Government Astronomer. *Meteorological Report No. 46*, Western Australia, p. 8.
- Drosdowsky, W. 1993. An analysis of Australian seasonal rainfall anomalies: 1950-1987. I: Spatial patterns. *Int. J. Climatol.*, 13, 1-30.
- Everitt, B.S. 1993. *Cluster Analysis*. Edward Arnold, London, 170 pp.
- Gentilli, J. 1959. *Weather and Climate in Western Australia*. Western Australian Government Tourist and Publicity Bureau, 24 pp.
- Gentilli, J. (ed.) 1971. *Climates of Australia and New Zealand, World Survey of Climatology, Volume 13*, Elsevier, Amsterdam, 405 pp.
- Gong, X. and Richman, M.B. 1995. On the application of cluster analysis to growing season precipitation data in North America east of the Rockies. *Int. J. Climate*, 8, 897-931.
- Hunt, H.A. 1910. *Australian Monthly Weather Report and Meteorological Abstract Vol. 1(5)*, 149-93.
- Jones, D.A. and Beard, G. 1998. Verification of Australian monthly district rainfall totals using high resolution gridded analyses. *Aust. Met. Mag.*, 47, 41-54.
- Jones, D.A. and Weymouth, G. 1997. An Australian monthly rainfall dataset. *Technical Report 70*, Bur. Met., Australia, 19 pp.
- Jones, D.A. and Trewin, B. 2000. The spatial structure of monthly temperature anomalies over Australia. *Aust. Met. Mag.*, 49, 261-76.
- Lavery, B.M., Kariko, A.P. and Nicholls, N. 1992. A historical rainfall dataset for Australia. *Aust. Met. Mag.*, 40, 33-9.
- Lavery, B.M., Joung, G. and Nicholls, N. 1997. An extended high-quality historical rainfall dataset for Australia. *Aust. Met. Mag.*, 46, 27-38.
- Mason, S.J. 1998. Seasonal forecasting of South African rainfall using a non-linear discriminant analysis model. *Int. J. Climatol.*, 18, 147-64.
- McBoyle, G.R. 1971. Climatic classification of Australia by computer. *Aust. Geographic Studies* 9, 1-14.
- Morrison, D.F. 1990. *Multivariate Statistical Methods*. McGraw-Hill, Singapore, 495 pp.
- Nicholls, N. and Lavery, B. 1992. Australian rainfall trends during the twentieth century. *Int. J. Climatol.*, 12, 153-63.
- North, G.R., Bell, T.L., Cahalan, R.F. and Moeng, F.J. 1982. Sampling errors in the estimation of empirical orthogonal functions. *Mon. Weath. Rev.*, 110, 699-706.
- Pittock, A.B. 1983. Recent climatic changes in Australia: implications for a CO<sub>2</sub>-warmed earth. *Climatic Change*, 5, 321-40.
- Prider, R.T. 1960. Part 6 – Natural Regions, in *Official Yearbook of Western Australia 1960, No. 2*, Commonwealth Bureau of Census and Statistics, WA Offices, p. 77.
- Richman, M.B. 1986. Rotation of principal components. *J. Climatol.*, 6, 293-335.
- Sadler, B.S., Mauger, G.W. and Stokes, R.A. 1988. The water resource implications of a drying climate in south-west Western Australia, in Pearman, G.I. (ed.), *Greenhouse: Planning for Climate Change*. CSIRO, Melbourne, pp. 296-311.
- SAS Institute Inc. 1989. *SAS/STAT User's Guide, Version 6, Fourth Edition, Volume 1*, Cary, NC: SAS Institute Inc., 943 pp.
- Seaman, R.S. 1982. A systematic description of the spatial variability of geopotential and temperature in the Australian region. *Aust. Met. Mag.*, 30, 133-41.
- S-PLUS 2000. Guide to Statistics, Volume 2, Data Analysis Products Division, MathSoft, Seattle, WA.
- Stern, H., de Hoedt, G. and Ernst, J. 2000. Objective classification of Australian climates. *Aust. Met. Mag.*, 49, 87-96.
- Torok, S.J. 1996. The development of a high quality historical temperature data base for Australia, Unpublished Ph.D. Thesis, University of Melbourne.
- Wolter, K. 1992. Cluster analysis in climate variability research – recent developments. *Proceedings of the 5th International Meeting on Statistical Climatology*, Toronto, Canada, June 1992.
- Wright, P.B. 1971. Spatial and temporal variations in seasonal rainfall in south-western Australia. *Miscellaneous Publication 71/1*, Institute of Agriculture, University of Western Australia, Agronomy Department, 89 pp.
- Wright, P.B. 1974a. Seasonal rainfall in southwestern Australia and the general circulation. *Mon. Weath. Rev.*, 102, 219-32.
- Wright, P.B. 1974b. Temporal variations in seasonal rainfalls in southwestern Australia. *Mon. Weath. Rev.*, 102, 233-43.
- Yu, B. and Neil, D.T. 1993. Long-term variations in regional rainfall in the south-west of Western Australia and the difference between average and high intensity rainfalls. *Int. J. Climatol.*, 13, 77-88.

