Rainfall Anomaly Patterns over Southern Africa

T G J DYER
[CLIMATOLOGY RESEARCH GROUP, UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG, 2001]

Abstract

It is reasonable to assume that the year-to-year total variation in rainfall over Southern Africa, as elsewhere, is a composite phenomenon. In this paper the variation in rainfall over both time and space has been decomposed into a number of uncorrelated components using principal components analysis. Each component gives rise to a spatial field or pattern of rainfall anomalies which can be plotted on a map of the sub-continent. The analysis makes available two different pieces of information. Firstly, the principal component loadings indicate the years most strongly represented by each component, that is the rainfall during these years possessed a common behaviour pattern. Secondly, the principal component scores (the rainfall anomalies) show the type of pattern that has tended to prevail during the aforementioned years. Such information can be useful and be applied to a number of relevant practical situations. Because the information is more likely to be used in the solution of a specific rather than a general situation, no attempt is made at this stage to give case studies of its use.

Introduction

The spatial distribution of annual rainfall receipts over a large surface can be represented by mean annual rainfall maps. However, there are instances when this mode of presentation is of little value because some aspects of the variation in this variable can be lost. Additionally, over Southern Africa as elsewhere the rainfall generating systems vary from one part of the country to another. It should be of value to attempt to partition the total variation in rainfall into components. This is what has been done in the work presented herein. Taking a basic data matrix of annual rainfall totals for 68 recording stations spread as evenly as possible over the country, and over the period 1910 – 1977, new variables have been derived which are weighted linear combinations of the original ones. By employing principal components analysis, these new variables are uncorrelated. Each station has a score on each of these variables which can be mapped to provide fields of rainfall anomalies. Furthermore, each variable will relate to specific groups of years. Therefore the fields portray situations that existed over the country for various combinations of years. That is, the rainfall distribution over the country was common for these years. These rainfall fields are of interest in a number of disciplines because they highlight areas receiving high positive rainfall anomalies, and vice-versa. If a large number of years is associated with a given anomaly field, the situation is a common one. Alternatively, a few years related to a field indicates an uncommon to rare event. In hydrology this information could be of value in relation to the siting of dams, whereas in agriculture those areas suitable only to drought resistant crops would become evident. The technique itself could be used in work dealing with weather modification, since anomaly fields based on periods before, and after cloud seeding should indicate if, and how, any improvement in rainfall had been accomplished. These fields can also be used for investigations into inter-relationships between different aspects of rainfall distribution and the general circulation of the atmosphere over and about Southern Africa.

Data

Sixtyeight rainfall stations with records extending over the period 1910 – 1977 have been employed in this analysis (Fig. 1). The reasons for taking fewer stations than in some previous studies, e.g. Dyer (1977), were firstly the ease with which the records could be updated, secondly their continuity over time, and thirdly as uniform a spatial distribution of stations as possible. The data were formed into a matrix of order 68 x 68, cases are stations (rows), and variables the years (columns).

Theory

Principal components analysis consists of determining a unique set of weighted linear combinations of the original data set. These new variables, principal components, are derived such that the first one accounts for as much of the data variance as is possible by a linear combination of the original variables. The second component accounts for as much of the remaining variance as possible, subject to the constraint that these two components are uncorrelated, and so on. The technique provides a mathematical transformation of the data matrix and in this sense it is not a statistical model.

If there are p original variables, there will be p components which together account for the total original variance. However, it is often found in practice that the first few components account for a major portion of this variance. Hence, the dimensionality of the data matrix can be reduced if those components accounting for little variance are discarded as representing noise or unimportant variance. Therefore, a large data matrix can often be accommodated by just a few new variables without much loss in information. How many new variables are used is, of course, under the control of the analyst since he can decide on how many components to retain. This can be done on the basis of some satisfactory percentage of the total variance to be accounted for, or one can take those components with variances greater than unity if the data has been standardised and their resulting covariance matrix becomes that of correlations.

The first, and succeeding principal components may be written as

$$PC_1 = \alpha_{1.1} X_1 + \alpha_{1.2} X_2 + \alpha_{1.p} X_p \ldots\ldots\ldots\ldots\ldots\ldots \quad (1)$$

$$PC_2 = \alpha_{2.1} X_1 + \alpha_{2.2} X_2 + \alpha_{2.p} X_p \ldots\ldots\ldots\ldots\ldots\ldots \quad (2)$$

$$PC_p = \alpha_{p.1} X_1 + \alpha_{p.2} X_2 + \alpha_{p.p} X_p \ldots\ldots\ldots\ldots\ldots\ldots \quad (P)$$

In these equations the X's are the original variables, and the α's are the principal component coefficients. These coefficients can be scaled to represent the correlations between each original variable and each component, these will be referred to as loadings for convenience. In obtaining the components either the covariance or correlation matrix for the X's may be used and in the present study the latter course has been adopted. Each component has a variance associated with it, and these variances are referred to as eigenvalues. When the correlation matrix is used in the analysis, the sum of the eigenvalues is equal to p the number of variables. The square of a loading for a variable on a particular component gives the proportion of that variable's variance accounted for by that principal component. If there are n cases for each x, then each component has n values called scores and each case has a score on each individual component.

In the present study the loadings represent correlations between years and components. Therefore, if a number of years have high loadings on a given component, this same component is representing something that those years have in common as regards rainfall. Many high loadings indicate a quite normal situation, whilst a few such loadings suggest the opposite.

The principal component scores relate stations to rainfall anomalies over the years represented by a given component. These scores have been standardised in the usual way and for each component the anomalies are divided by the square root of the component's eigenvalue. By plotting station scores per component, on a map, rainfall anomaly fields are defined, and what is more, the fields represented by different components are uncorrelated. These spatial descriptions of rainfall are better seen from the results that follow. The reader is referred to Dyer (1975), Harris (1975) and Maxwell (1977) for further details.

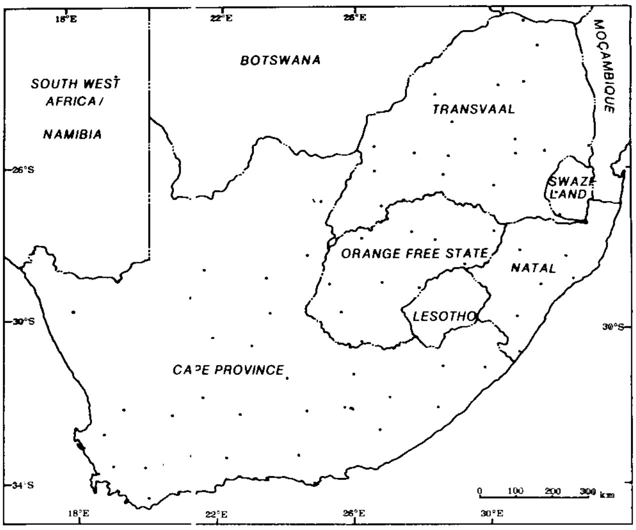

Figure 1
The sixtyeight station locations used in the principal components analysis

Results

Because, quite often, all variables (years in this case) have high loadings (correlations between years and components) on the first component which then represents the mean situation for rainfall over years, the principal components' axes can be rotated to give a new set of linear combinations of the original variables which are of greater use (Press, 1972). Also, the situation sometimes exists where a component has a number of mediocre loadings, that is, it does not represent these years' rainfall very strongly. Hopefully, the rotation leads to less confusion overall in that loadings or correlations on any given component that are mediocre are made either large or close to zero. If the important components, say, m (m $<$ p) are rotated, their combined variance extraction remains unaltered, but this variance is re-distributed across the components. By m important components is meant those of the total number, p, which account for a goodly percentage of the total variance. The remaining (p — m) components are assumed to represent minor effects, or noise in the data.

There are a number of rotative schemes available, and the varimax solution has been used herein. This one maintains orthogonality between the components; that is, each component represents some situation that is uncorrelated with whatever circumstances are represented by the others. Sometimes the characteristics of a particular component can be recognised and then a label may be attached to it. However, it is felt that a label of dubious validity can be misleading and insistence upon attaching labels to components is not advocated. This policy does not detract from the usefulness, or information conveyed by a component's spatial field. There are other rotative systems that permit the components to become correlated but it is felt that the advantages of these are, usually, outweighed by their disadvantages. In the work described within this paper the first five principal components account for a total variance of 91%, and these five have been rotated. Thus instead of considering the original 68 variables, it is possible to use only five new ones, linear combinations of the 68, with a loss of information equal to 9%.

Component 1

The loadings on this component, correlations between years and the component, are given in Fig. 2. It will be seen that the rainfall for 34 years is well represented by this new variable. By this is meant that one of the factors that go into making up the total rainfall variation occurred in a similar form during 34 of the 68 years considered. In this instance all the loadings or correlations are of the same sign so the component represents the sum of rainfall anomalies over the years having these high loadings. For clarity, only those loadings with values greater than 0,5 have been displayed.

Having established the years whose partial rainfall, a fraction of the total, is represented by the first component, it now remains to determine how this fraction was distributed over the country as a whole. This is accomplished by plotting the component scores on a map for each rainfall station. These scores represent rainfall, but they have been standardised in the usual manner by subtracting the mean and dividing by the standard deviation. So, for this component (and the others) there are 68 scores, that is, one per station. Say the scores, rainfall, are represented by Y_i $(1 < i < 68)$ and let the mean of this score be $\bar{y}$, with standard deviation s. Then each score is transformed by calculating the values $(Y_i - \bar{y})/s$ (it will be appreciated that we are dealing with rainfall anomalies). These scores will have zero mean and unit variance, therefore the strength of the anomalies for each component lend themselves to easy comparison. Apart from the above property, these standardised scores tend to be normally distributed and it is common policy to attach most importance to only those with values greater than unity, having ignored their signs.

The rainfall field represented by this first component is shown in Fig. 3 which is the plot of the standardised station scores. These scores have been multiplied by 10 for ease of presentation. Relative to the means for each year with high loadings (see Fig. 2) positive anomalies across the Transvaal were accompanied by negative ones along the Cape coast and southwestern Cape. In fact this component represents a situation whereby the country is divided roughly in half, with the northern section having positive rainfall anomalies and the southern part oppositely so. There was a very strong negative pocket in the southwestern Cape, but note the lack of relief over much of the Cape Province.

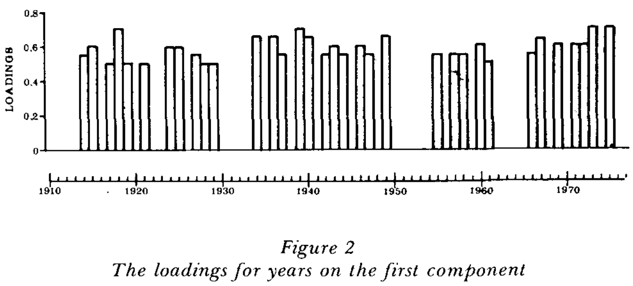

Figure 2
The loadings for years on the first component

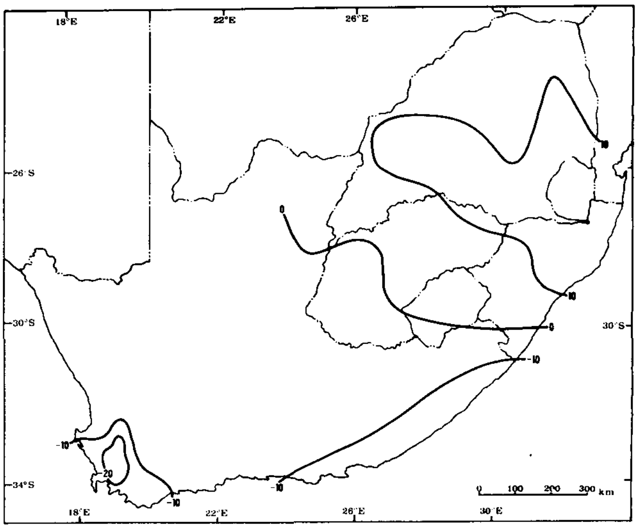

Figure 3
The scores for stations on the first component

If attention is returned to Fig. 2, indicating the years when this situation tended to be in existence, it will be noted that, on the whole, this particular field was present, or absent, over runs of years. However, ignoring the absences within a period of high loadings (denoting the presence of this field), it will be seen that the runs of years associated with this particular field were longer than those associated with its absence. So although this particular field of anomalies was present for 34 of the 68 years of record, it did not occur in a completely random way. Again, if the occasional absent year in a run of affirmative years is ignored, then the Cape appeared to suffer runs of negative anomalies covering periods up to about 16 years in length. Of course, for the same periods or runs of years, the northern section of the country had positive rainfall anomalies. It should be remembered that the first component accounts for only part of the total variation in rainfall, therefore, these anomalies are fractions of the total anomalies or variations about the long term individual station rainfall means.

Component 2

This component represents a rainfall factor, uncorrelated with the one represented by the first component, which was in evidence at the beginning and end of the record. There are two exceptions to this, namely the years 1932 and 1935. Strong loadings were present for a total of sixteen years (Fig. 4) or about 0,25 of the time. However, again it is seen that the years (Fig. 4) tend to be in blocks. The actual rainfall anomaly fields represented by this component can be seen from a plot of the component scores (Fig. 5). In this case a factor exists that tends to divide the country in an east-west sense. Quite strong gradients were present from the Natal and Cape coasts and about three quarters of the interior suffered negative rainfall anomalies. During the years with high loadings on this component it will be seen that the Natal coast had very high positive rainfall anomalies indeed, possibly sufficient to cause flooding but further research is necessary to establish the truth of this possibility. Whilst component one may well represent strong anti-cyclonic conditions in the north and associated adverse, for the southwestern Cape, movements of the south-atlantic high it is perhaps likely that component two represents rainfall in Natal, particularly, arising from frontal situations.

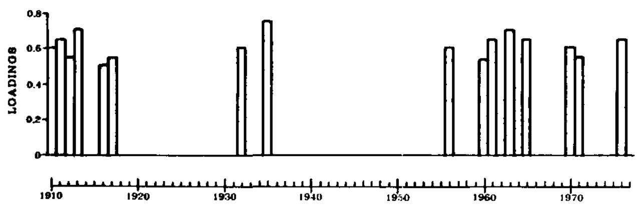

Figure 4
The loadings for years on the second component

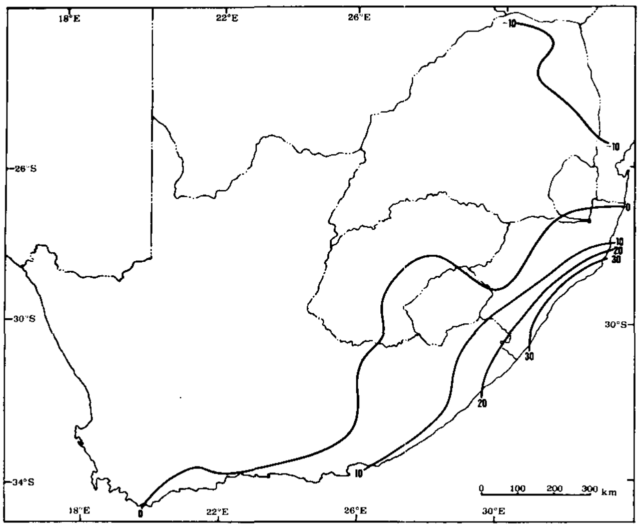

Figure 5
The scores for stations on the second component

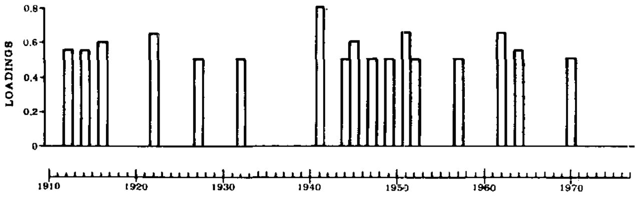

Figure 6
The loadings for years on the third component

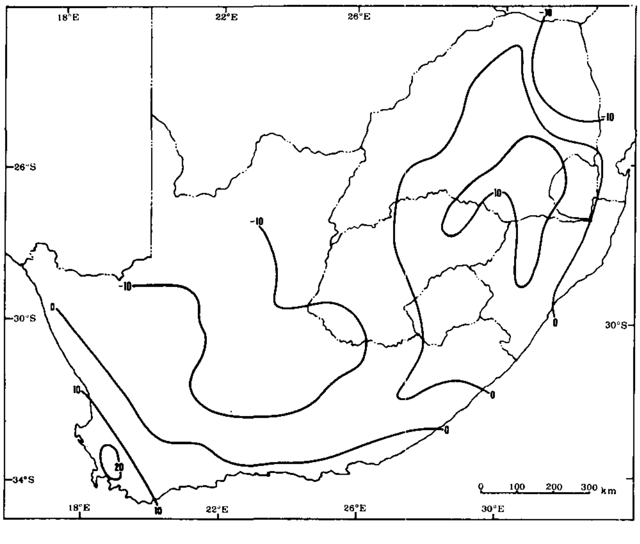

Figure 7
The scores for stations on the third component

Component 3

Seventeen years rainfall anomalies are represented by this component, virtually the same number as for the previous component. However, the distribution of the highly loaded or correlated years as indicated in Fig. 6, is quite different from that in Fig. 4. The type of situation that presented itself over the country during the years highlighted in Fig. 6 are shown from the plot of the component scores in Fig. 7. There was the tendency for an undulating rainfall situation over the subcontinent. The rainfall anomalies were particularly high and positive over the southwestern Cape, low and negative over the Karoo, quite high over the southeastern Transvaal, and again negative over the far northern part of this same state. Points of interest shown by the field in Fig. 7 are the adjacent high positive and high negative anomalies over the southwestern Cape and interior respectively, and a similar situation over northern areas of the country.

Component 4

About a third of the years rainfall is associated with this component (Fig. 8). To a certain extent, the years tended to block and this was particularly so during the twenties, and fifties. This component or rainfall factor is very much associated with the northeastern Transvaal (Fig. 9). With the exception of a small pocket of positive anomalies in the southwestern Cape, the contours over the remainder of the country show little relief. This component may be representative of conditions associated with cyclones which tend to occur and inflict considerable damage over that part of the country, that is, the northeastern Transvaal.

Component 5

Only four years rainfall are well correlated with this component (Fig. 10), and the situation occurring during these years may therefore be looked upon as constituting a rare event. The actual years when the field in Fig. 11 existed were 1931, 1950, 1963 and 1974. A ridge of positive rainfall anomalies extends inland from the northeastern Cape coast and southern Natal coast. Negative anomalies existed over much of the southern area of the country, northeastern Natal, and Swaziland. A centre of positive anomalies was present over the northeastern Transvaal. To a certain extent the rainfall field represented by this component is undulating across the country in the same sense that existed for component three.

Conclusion

Using principal components analysis, five uncorrelated rainfall anomaly fields have been defined which together account for approximately 91% of the total variance in rainfall for 68 records spread over the country. The technique applied to the data provides a mathematical transformation thereof, that is, it is not a statistical frolic! Therefore, the rainfall anomaly fields are real, and being based on past data they actually occurred. Each field represents situations common to a number of years, which are indicated by the loadings for years on the various components. In describing rainfall patterns over the country as a whole, it is possible to determine those areas of the country that commonly suffer droughty conditions. The use of the fields in this sense is flexible because the person concerned may define what would be considered as a serious anomaly score and note the appropriate distribution of such anomalies. The fields lend themselves to applications in the siting of dams. For this purpose it would be necessary to determine areas of the country with large positive rainfall anomalies, coupled with the frequency of occurrence of such values which are shown in the figures giving loadings for years. Of course, in some areas of research the grid chosen may be too coarse. For example, the technique is in the process of being applied to an assessment of the effect of cloud seeding for hail suppression over the Nelspruit area. This particular exercise requires a much finer grid of rainfall recording stations and apart from the fact that the work is still in progress, to fully report on it here would seriously overburden the paper. Finally, another way in which it might be necessary to refine the grid system would be to consider monthly, or less, rainfall totals. In this respect perhaps such a course of action would be more appropriate when dealing with the various phenological stages in crop development over a growing season. The same would be true if one were investigating irrigation requirements. It is therefore concluded that the description of past rainfall behavioural patterns has many possible useful applications, and these are presently being actively pursued. It needs hardly be stated that the results obtained in this paper, together with their applications could equally well apply to other variables. Those that come to mind are temperature, pressure fields, soil moisture deficits, and the like.

Acknowledgements

To Mr PJ Stickler for the final form of the diagrams, referees for constructive criticisms that made the paper more readily readable, and to the Water Research Commission for financial support toward the cost of travel to visit establishments of relevance in the United Kingdom.

References

DYER, T.G.J. (1975) The assignment of rainfall stations into homogeneous groups: an application of principal components analysis. *Quart. J. Roy. Meteor. Soc.* **101** 1005–1013.

DYER, T.G.J. (1977) On the application of some stochastic models to precipitation forecasting. *Ibid* **103** 177–189.

HARRIS, R.J. (1975) *A Primer of Multivariate Statistics.* Academic Press, London, 332 pp.

MAXWELL, A.E. (1977) *Multivariate Analysis on Behavioural Research.* Chapman & Hall, London. 157 pp.

PRESS, J.S. (1972) *Applied Multivariate Analysis.* Holt, Rinehart, and Winston, London. 521 pp.

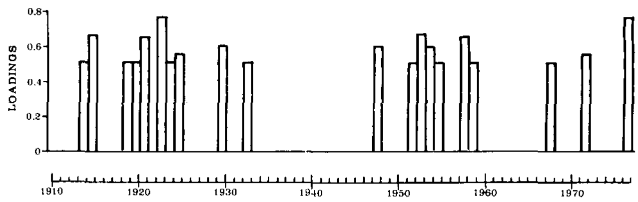

Figure 8
The loadings for years on the fourth component

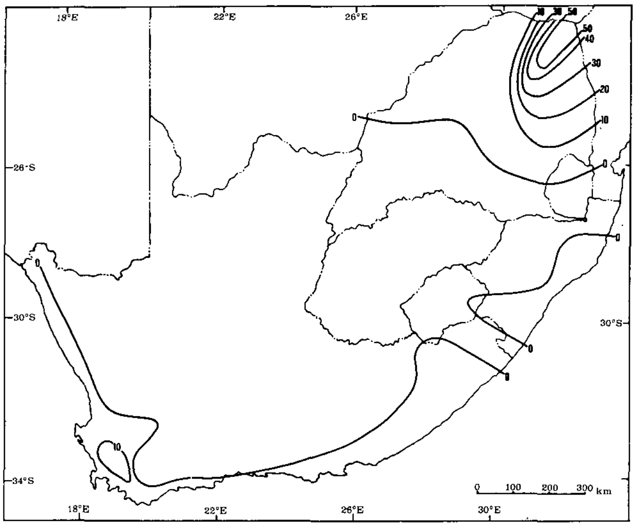

Figure 9
The scores for stations on the fourth component

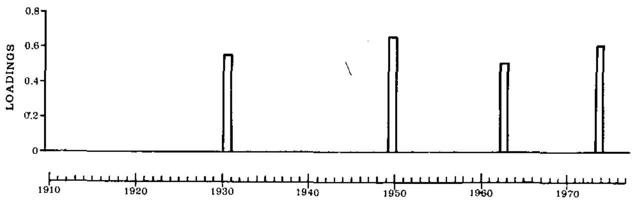

Figure 10
The loadings for years on the fifth component

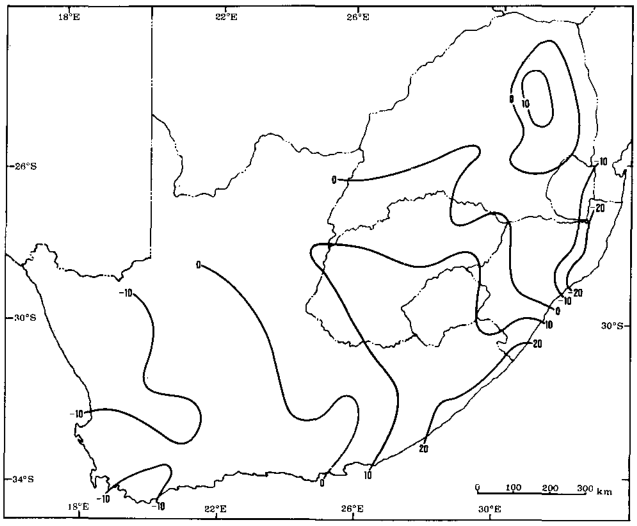

Figure 11
The scores for stations on the fifth component