Persistence of Subtropical African Droughts

Richard W. Katz

Environmental and Societal Impacts Group, National Center for Atmospheric Research,[1] *Boulder, CO 80307*

8 December 1977 and 16 February 1978

ABSTRACT

Kraus (1977) has demonstrated that subtropical African droughts exhibit statistically significant persistence. It is emphasized, through a further analysis of annual subtropical African rainfall, that the data are highly variable with only a small degree of persistence. These results have significant implications concerning the appropriate characterization of the likelihood of drought for dissemination to decision-makers.

1. Introduction

In a recent article concerned with the analysis of subtropical rainfall anomalies (Kraus, 1977), it is claimed that subtropical African droughts "tend to be rather persistent" with "unexpectedly long runs of wet or dry years." The question naturally arises as to what is the degree of this persistence since Kraus' results only establish that it is statistically significant. In particular, is the degree of persistence of these rainfall anomalies great enough to be useful for predictive purposes (Kraus makes no claim that his results necessarily have any predictive value) or is it so small that rainfall should still be characterized as highly variable and nearly independent from year to year? Because a similar analysis of Sahelian rainfall (Bunting *et al.*, 1976), based on a smaller number of stations, does not even find statistically significant persistence, it might be expected that this dependence is not very great.

The purpose of this note is to present a further analysis of subtropical African rainfall data in order to obtain estimates of the nature of the year-to-year dependence of rainfall anomalies. These estimates will necessarily be of a qualitative nature, relying in part on some recently introduced techniques in exploratory

[1] The National Center for Atmospheric Research is sponsored by the National Science Foundation.

0027-0644/78/1017-1021$05.00

data analysis. The techniques, nevertheless, do reveal quite clearly important characteristics of the data.

Kraus' analysis is based on annual area-averaged rainfall anomalies. Likewise, in this paper a "drought" shall be considered an "anomaly" or "shortfall" in precipitation (e.g., a substantial negative departure from a long-term mean value). It should, of course, be noted that there are many alternative and possibly more realistic definitions of drought (e.g., Katz and Glantz, 1977).

2. Data analysis

a. Kraus' data and results

The initial data consist of annual rainfall for all available stations in the region 30°W–60°E and 10°N–25°N for the time period 1911–74. The annual rainfall time series for each station are first normalized (or standardized); that is, the station's mean annual rainfall is subtracted from each observation and the resultant deviations are divided by the standard deviation of the station's annual rainfall. For each year these normalized rainfall values are then averaged over all available stations (ranging from 10 to 32 stations depending on the year), yielding one value per year for the region (see Kraus, 1977, for further details). This derived data set, listed in Table 1 of Kraus (1977),[2] shall henceforth be referred to as the *standardized precipitation* data for subtropical Africa.

Kraus presents the results of performing two statistical tests to support his claim that droughts are persistent. He uses rank-order statistics to verify that the long runs of either positive or negative standardized precipitation values which are present in the data would be very unlikely to have happened by chance. This result does not, however, give any indication of the degree of persistence exhibited by the data. It simply suggests that the data do not constitute a completely random series (i.e., a sequence of independent observations).

He also tests for the significance of the first-order autocorrelation coefficient, obtaining a correlation of 0.32 which is significant at the 0.01 probability level. In addition, we have computed a 95% confidence interval for the first-order autocorrelation coefficient based on the assumption of a first-order autoregressive process and making a large-sample approximation for the standard error (see p. 34 of Box and Jenkins, 1976). The resultant confidence interval is (0.09, 0.55), indicating that the true autocorrelation is unlikely to be relatively high but could be quite close to zero.

b. Further analysis

We shall attempt to give a broad characterization of the nature of the African standardized precipitation data, as opposed to simply testing for statistical significance. To do so will require some qualitative procedures, including the use of descriptive statistics and graphs and incorporating some recent developments in exploratory data analysis.

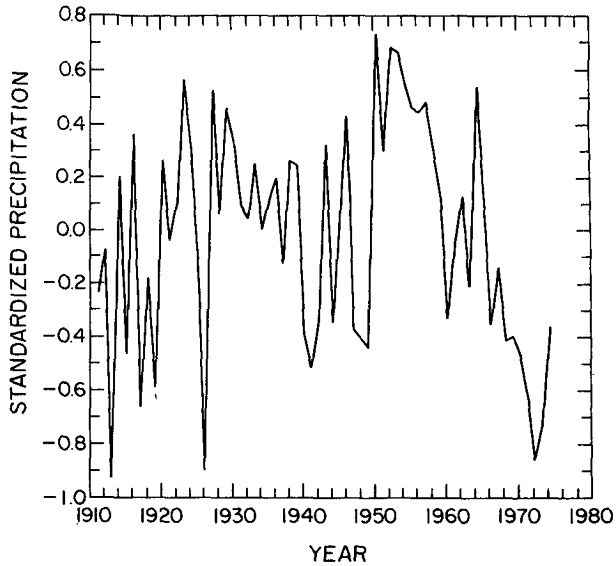

FIG. 1. Subtropical African annual standardized precipitation, 1911–74.

Fig. 1 shows a plot of the African standardized precipitation data. It is evident that the standardized precipitation values are highly variable from year to year and that there is certainly not a high degree of persistence present in the data.

Kraus classifies years as dry or wet according to the sign of the standardized precipitation value; that is, less than zero ("dry") and greater than zero ("wet"). We investigate the degree of persistence of such dry and wet years by considering the conditional distribution of standardized precipitation given that the preceding year is either dry or wet. Descriptive statistics which characterize the unconditional (or overall) distributions are given in Table 1 (first three columns). In addition, these distributions are graphically displayed in Fig. 2a in the form of boxplots (see Appendix A). A boxplot is essentially a streamlined histogram. It simplifies making comparisons among several distributions.

While there is a definite shift in the two conditional distributions in the direction of persistence, the spreads of the conditional distributions are roughly the same as for the unconditional distributions (as judged by the lengths of the boxes in Fig. 2a). In comparison with this variability, the differences between these two conditional distributions are relatively small. It is apparent that, while the shifts in conditional distributions may be statistically significant, there is still a large uncertainty in predicting the next year's precipitation on the basis of whether the current year is dry or wet.

[2] The entries in Table 1 of Kraus (1977) for Africa contain two printing errors: $a_{1955} = +0.461$ and $a_{1961} = -0.028$ (Kraus, personal communication).

TABLE 1. Descriptive statistics for conditional distributions of subtropical African standardized precipitation.

		Conditional distributions given preceding year				
	All data	Less than zero	Greater than zero	Less than lower quartile	Between quartiles	Greater than upper quartile
Sample size	63	29	34	16	31	16
Mean	−0.015	−0.165	0.114	−0.199	−0.038	0.214
Standard deviation	0.426	0.460	0.353	0.472	0.378	0.385
Minimum	−0.930	−0.930	−0.663	−0.857	−0.930	−0.663
Lower quartile	−0.374	−0.461	−0.127	−0.512	−0.329	0.071
Median	0.071	−0.181	0.125	−0.373	0.007	0.300
Upper quartile	0.308	0.200	0.436	0.200	0.251	0.461
Maximum	0.739	0.739	0.685	0.739	0.567	0.685

In an attempt to more realistically characterize drought in terms of precipitation, we have also classified years according to three states: less than lower quartile ("very dry"), between lower and upper quartiles ("normal"), and greater than upper quartile ("very wet"). We now consider the conditional distribution of standardized precipitation given each of these three possible states for the preceding year. Descriptive statistics for the three conditional distributions are listed in Table 1 (last three columns) with the corresponding boxplots shown in Fig. 2b. While it does appear that at least one of the conditional distributions (i.e., given very wet) is less variable than the unconditional distribution, the results support essentially the same conclusion as in the two-state case. The shifts in the three conditional distributions, although consistent with a persistence effect, are again small relative to the variability of the distributions.

Finally, in addition to examining the conditional distribution based on an arbitrary classification according to a few states, we shall consider the conditional distribution of standardized precipitation as a continuous function of the level of precipitation on the preceding year. A graphical technique involving the use of moving statistics (see Appendix B) is applied to the ordered pairs of values of standardized precipitation on consecutive years. This technique is basically a generalization of the concept of a moving average as employed in time series analysis and can also be thought of as a continuous analogue of the boxplots given in Fig. 2.

Fig. 3 shows the conditional distribution in the form of smoothed moving quartiles (i.e., lower quartiles, medians, and upper quartiles), as the level of precipitation on the preceding year is gradually changed. All three curves show a gradual increase consistent with persistence. This change, however, is again quite small relative to the variability of the conditional distribu-

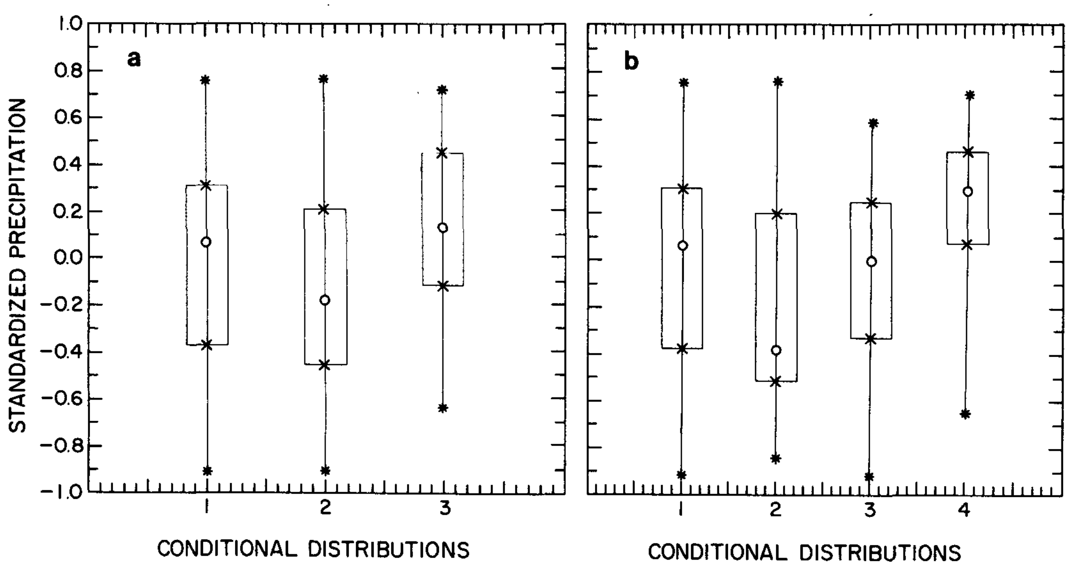

FIG. 2. Conditional distributions of standardized precipitation in the form of boxplots showing minima and maxima (*), lower and upper quartiles (X) and medians (O). Part a illustrates the two-state case where 1 is the unconditional distribution, 2 the conditional distribution given dry preceding year, and 3 conditional given wet. Part b is the three-state case where 1 is the unconditional distribution, 2 the conditional distribution given very dry preceding year, 3 conditional given normal, and 4 conditional given very wet (see text).

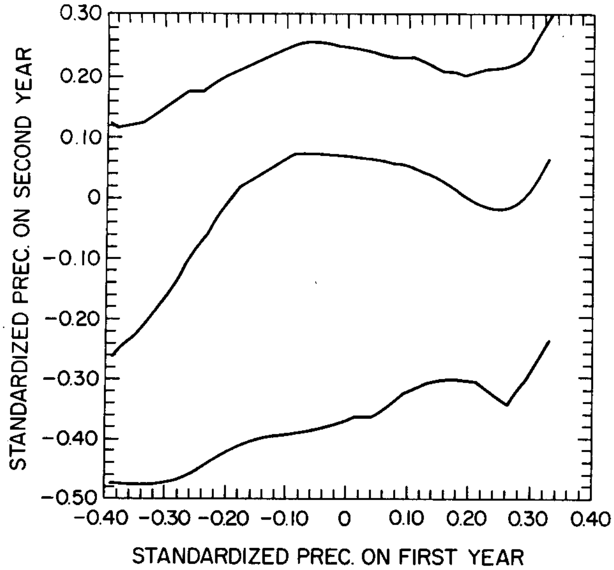

FIG. 3. Smoothed moving quartiles for standardized precipitation on consecutive years (bottom curve for lower quartiles, middle curve for medians and top curve for upper quartiles), based on sliding batches of size 20 and smoothed using moving averages of order 10 (see Appendix B).

tions as indicated by the interquartile ranges (differences between upper and lower quartiles).

In summary, the relationship between values of subtropical African standardized precipitation on consecutive years can best be characterized as highly variable with only a small degree of year-to-year persistence. More complicated predictive schemes (e.g., conditioning on two or more preceding years), of course, could be considered but the results would still be essentially the same.

3. Remarks

It should be apparent from the above discussion that a distinction needs to be made between the statistical significance and the practical significance of tests of hypotheses. While the subtropical African standardized precipitation data are certainly not a purely random time series, there is not necessarily a great enough degree of year-to-year dependence to be of much, if any, value for predictive purposes. It should be added, nevertheless, that such statistically significant results may be of some value in suggesting the need to search for underlying physical mechanisms which might explain this dependence.

This distinction is critical with regard to the appropriate characterization of the likelihood of drought for dissemination to decision-makers. Perceptions of drought and its likelihood can be quite erroneous and these misperceptions can lead to inappropriate action, even to excuses for the lack of any action, to combat the impact of drought (e.g., Katz and Glantz, 1977). Since year-to-year precipitation amounts can be highly variable, a decision-maker should be made aware, for instance, that a wet year does not imply that there is little or no chance that the next year will be dry.

In this regard, perhaps too much effort has been devoted to the search for persistence, trends or cycles in precipitation and other meteorological data. It is not clear that the exhaustive analysis of Sahelian precipitation data with these purposes in mind has led to any useful information concerning the likelihood of drought in the Sahel next year.

APPENDIX A

Boxplot

A boxplot [for more details see Chapter 2 of Tukey (1977)] is constructed by first plotting vertically the minimum, lower quartile, median, upper quartile and maximum for the given sample of data. A line is then drawn through these points, connecting the minimum and maximum, to show the range of the data. Finally, a box is drawn to emphasize the interquartile range (or middle half of the data). The two horizontal sides of the box are drawn through the lower and upper quartiles, making its vertical length equal to the interquartile range.

APPENDIX B

Moving Statistics

The idea of enhancing scatterplots with *moving statistics* is due to Cleveland and Kleiner (1975). We assume that observations $\{(X_i,Y_i): i=1, \ldots, n\}$ are given for two random variables X and Y, where the X_i's are arranged in increasing order carrying along the concomitant Y_i's. The data are divided into sliding batches of size r denoted

$$B_x(i;r)=\{X_i, X_{i+1}, \ldots, X_{i+r-1}\},$$

$$B_y(i;r)=\{Y_i, Y_{i+1}, \ldots, Y_{i+r-1}\}, \qquad i=1, \ldots, n-r+1.$$

Let T_0 denote the midmean statistic (i.e., mean of middle half of data). This statistic is computed for each x-batch $B_x(i;r)$. Let T_1, T_2 and T_3 denote the lower quartile, median and upper quartile statistics, respectively. These statistics are computed for every y-batch $B_y(i;r)$. The graphical display consists of plots of the moving quartiles for the y-batches versus the moving midmeans for the x-batches; namely,

$$T_j[B_y(i;r)] \text{ vs } T_0[B_x(i;r)], \quad j=1, 2, 3.$$

Usually, before plotting these statistics, the moving quartiles (T_1,T_2,T_3) are smoothed across batches (i.e., as a function of i). This smoothing is performed using a moving average of, say, order s. See Katz (1977) for further details concerning the use of moving statistics.

REFERENCES

Box, G. E. P., and G. M. Jenkins, 1976: *Time Series Analysis: Forecasting and Control* (rev.). Holden-Day, 575 pp.

Bunting, A. H., M. D. Dennett, J. Elston and J. R. Milford, 1976: Rainfall trends in the West African Sahel. *Quart. J. Roy. Meteor. Soc.*, **102**, 59–64.

Cleveland, W. S., and B. Kleiner, 1975: A graphical technique for enhancing scatterplots with moving statistics. *Technometrics*, **17**, 447–454.

Katz, R. W., 1977: Techniques for detecting dependence between meteorological variables. *Preprints Fifth Conf. Probability and Statistics*, Las Vegas, Amer. Meteor. Soc., 101–105.

——, and M. H. Glantz, 1977: Rainfall statistics, droughts and desertification in the Sahel. *Desertification: Environmental Degradation in and around Arid Lands*, M. H. Glantz, Ed., Westview Press, 81–102.

Kraus, E. B., 1977: Subtropical droughts and cross-equatorial energy transports. *Mon. Wea. Rev.*, **105**, 1009–1018.

Tukey, J. W., 1977: *Exploratory Data Analysis*. Addison-Wesley, 688 pp.