STATISTICAL ANALYSIS OF AUSTRALIAN PRESSURE DATA

By S. C. Das*

[*Manuscript received February* 8, 1956]

Summary

We give a statistical analysis of the pressure data along the east coast of Australia, to test a hypothesis which arose out of Deacon's (1953) work, that there was a shift in the mean high pressure belt. It is shown that the belt is slowly moving southward.

I. Introduction

In the present paper we consider the statistical analysis of Australian pressure data in order to test the following hypothesis which arose out of Deacon's (1953) work. Deacon compared the mean daily maximum temperatures for the summer seasons between two 30 year periods 1881–1910 and 1911–40 for some Australian inland localities. Table 1 of his paper gives temperature differences for the two periods for 14 stations. All these show a consistently lower temperature in period 2 than in period 1. He also found that for the period 1911–50, the summer rainfall over much of the southern part of Australia was considerably greater than in the previous 30 years. He conjectured that these changes were due to a climatic trend, and by noting the difference in the character of the annual variation in atmospheric pressure between these two periods he suggested that the cause of the phenomena might be a shift in the mean high pressure belt.

To investigate how far the conjecture is correct we shall make an analysis of the pressure data. The way in which this is done is briefly indicated below. By taking the sea-level pressure along a line across the maximum of high pressure, it is possible to determine by appropriate curve fitting the position and the intensity of the maximum at any time. Both the position and the intensity are expected to show a variation with season, but the interesting point is to examine whether any of these elements are showing a climatic trend. With this in view we proceed to determine

(1) the annual mean position of the maximum,
(2) the annual mean intensity of the maximum,
(3) the amplitude of the annual cycle of position,
(4) the amplitude of the annual cycle of intensity,

and to examine what trends, if any, are present in any one of these.

From practical considerations it was suggested that the line should incorporate the following stations : Cairns, Townsville, Rockhampton, Brisbane,

* Australian National University, Canberra, A.C.T. ; present address : Ravenshaw College, Orissa, India.

Sydney, Moruya Heads, Gabo I., Launceston, and Hobart. Other stations for which data are available could not be included as they are either far east or west of this line and inclusion of data from them would have confused the results because of lateral pressure differences, or because they are far above sea-level and another type of error would be likely to be introduced.

II. Source of the Data

All these stations have records dating back to 1909 which provided 46 years of data. For some stations and for some years these were available from published records " Results of Rainfall Observations " made in different states. The figures given in these records are, generally, the mean of 9 a.m. and 3 p.m. pressures in inches corrected and reduced to 32 °F., M.S.L., and standard gravity. The rest of the data were supplied by the Division of Meteorological Physics, C.S.I.R.O., with a few adjustments here and there to make the values as complete and homogeneous as possible. As an example we give the adjustments made for Cairns. For a few months here, no values are available and the values for Cairns aerodrome have been inserted instead. For some other months only 9 a.m. values are available ; these were adjusted by assuming that the pressure variation between 9 a.m. and 3 p.m. at Cairns aerodrome is the same as at the post office, i.e. from the 9 a.m. values at the post office have been subtracted half the difference between the values at 9 a.m. and 3 p.m. at the aerodrome.

III. Analysis of the Data

For every month, we have the nine station values of the position (latitude) and intensity of the monthly average pressures. The first step in the analysis is to take the latitude as the independent variable and the intensity as the dependent variable, and to fit an appropriate curve to determine the intensity and position of the monthly mean maximum pressures. But the appropriate curve to be fitted depends upon the relationship between position and intensity of the monthly average pressures, which is unfortunately not known. Consequently, as a sort of pilot method, the data for a sample of months were plotted, and smooth graphs were drawn by eye through these plotted points to give some indication of this relationship. The graphs varied so much from month to month and from year to year that it was very difficult to find one or two suitable equations which would fit the data satisfactorily. The only objective method was to fit orthogonal polynomials in successive stages, the goodness of fit of consecutive terms of higher degree being observed and tested at each stage. This would have required a huge amount of labour and would not have given much more information than the subjective method followed. Besides there is the following difficulty. In certain summer months the maximum lies south of Hobart, and any reasonable curve fitted to the nine observations would not have a relative maximum, so that some subjective judgment must necessarily be used.

In view of this, the pressure data for individual months of each year for the nine given stations were plotted as ordinates against the latitudes of these stations as abscissae. By drawing smooth graphs by eye through these plotted points, the position and the intensity of the mean monthly maximum pressures were

TABLE 1

CALCULATED ANNUAL VALUES FROM METEOROLOGICAL DATA

Year	Annual Mean Position of the Maximum (deg. of lat.)	Amplitude of the Annual Cycle of Position (deg.)	Annual Mean Intensity of the Maximum (in.)	Amplitude of the Annual Cycle of Intensity (in.)
1909	31·79	2·92	30·005	0·076
1910	31·71	4·93	30·040	0·100
1911	31·50	4·49	30·028	0·107
1912	31·12	5·15	30·051	0·108
1913	31·71	5·66	30·052	0·130
1914	32·17	0·80	30·106	0·132
1915	29·21	3·40	30·010	0·084
1916	30·96	4·64	30·094	0·296
1917	30·83	5·82	30·001	0·080
1918	31·79	5·68	30·053	0·089
1919	32·12	5·54	30·078	0·111
1920	32·12	4·06	30·032	0·073
1921	32·42	4·37	30·066	0·070
1922	31·83	3·09	29·985	0·125
1923	29·12	5·74	30·010	0·094
1924	30·96	2·33	30·021	0·159
1925	32·25	4·50	30·050	0·123
1926	29·54	3·82	30·022	0·086
1927	32·00	3·88	30·123	0·230
1928	30·83	5·43	30·025	0·093
1929	30·37	4·55	30·005	0·133
1930	31·96	4·63	30·093	0·107
1931	31·25	4·15	30·050	0·097
1932	32·30	2·17	30·029	0·114
1933	34·83	2·89	30·027	0·124
1934	35·28	6·61	30·059	0·126
1935	32·50	5·08	30·014	0·100
1936	32·80	5·65	30·038	0·117
1937	34·46	2·77	30·055	0·154
1938	33·96	3·79	30·029	0·151
1939	31·58	5·63	30·033	0·082
1940	32·04	0·68	30·067	0·166
1941	31·87	3·32	30·066	0·105
1942	33·25	4·08	30·034	0·078
1943	33·04	2·02	30·005	0·090
1944	31·58	1·15	30·045	0·172
1945	32·71	4·48	30·048	0·100
1946	32·12	5·66	30·003	0·092
1947	32·83	5·45	30·048	0·105
1948	32·08	5·21	30·032	0·092
1949	32·71	4·80	30·049	0·157
1950	35·83	5·85	30·065	0·135
1951	31·75	4·81	30·028	0·096
1952	31·58	1·45	30·003	0·097
1953	31·75	3·89	30·047	0·060
1954	33·46	1·67	30·087	0·114

determined. Thus 12 values for the intensity and 12 values for the position were obtained for each of the years from 1909 to 1954.

The means (from the 46 yearly observations) for each month (for intensity and position) were then calculated. These showed a seasonal movement which was well fitted by a simple sine curve. By fitting such a sine curve the annual values for (1), (2), (3), and (4) were obtained ; these are given in Table 1.

In order to examine whether there is any trend in the annual values, a polynomial regression line of the form $Y=A+B'\xi_1'+C'\xi_2'$ has been fitted to each column of Table 1. Following the accepted notation used in Fisher and Yates's tables, ξ_1' and ξ_2' stand for orthogonal polynomials of degree one and two in x where x is the number of years measured from the centre point of the series. The results of computations are given in Table 2.

TABLE 2

REGRESSION COEFFICIENTS

Annual Values	B'	C'	$t_{B'}$	$t_{C'}$
Annual mean position of the maximum pressure ..	0·02054	—0·00044	2·94	0·19
Amplitude of the annual cycle of position ..	—0·00750	—0·00114	0·89	0·40
Annual mean intensity of the maximum	—0·00001	0·00001	0·02	0·04
Amplitude of the annual cycle of intensity ..	—0·00009	—0·00008	0·38	1·02

The first two columns give the values of the first and the second regression coefficients for each of the annual values. To test whether a regression coefficient is significantly different from zero we have computed the corresponding t by dividing the particular regression coefficient in question by its estimated standard error. These t values are given in the third and the fourth columns of the table. For 43 degrees of freedom none of the regression coefficients are significant except B' for the annual mean position of the maximum pressure. The regression equation obtained for the mean position in degrees is

$$Y=32\cdot106+0\cdot04108x,$$

and shows that the high pressure belt is shifting slowly towards the south. The regression coefficients for the annual mean intensity of the maximum and the amplitude of its oscillation are of course so far from significant that nothing can be said about them. While the coefficient of the first term in the regression of the amplitude of the cycle (in the position) on time is not significant, it is possible that the amplitude of the cycle is decreasing so that not only is the mean position of the maximum further south, but the oscillation about this mean position is shallower and the maximum will not tend to come as far north. Figure 1 shows the annual mean position and the above regression line.

The graph for the mean position suggests that there is, apart from the long-period trend, a certain persistence in successive values. Accordingly, we have

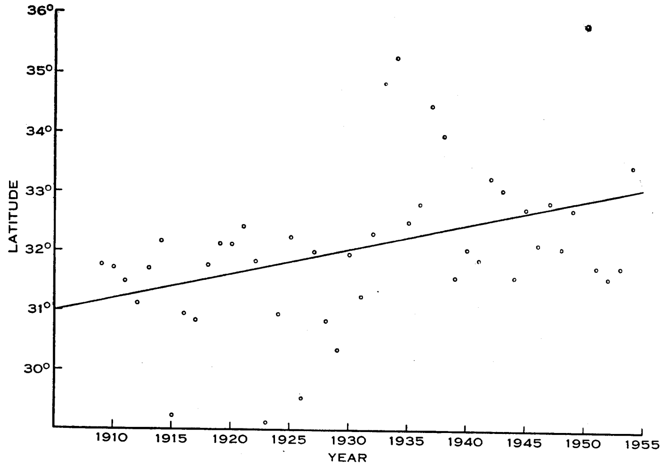

Fig. 1.—The annual mean position of the maximum pressure.

computed the first five serial correlations of the residuals, Z_t, from the regression on the first two orthogonal polynomials. The formula used was

$$r_j = \sum_{1}^{46-j} Z_t Z_{t+j} \Big/ \sum_{1}^{46} Z_t^2.$$

This formula will bias the estimates downwards slightly but this will not matter for our purpose. The first five r_j's were

$$r_1 = 0 \cdot 226, \quad r_2 = -0 \cdot 015, \quad r_3 = 0 \cdot 081, \quad r_4 = 0 \cdot 030, \quad r_5 = -0 \cdot 010\,;$$

clearly the last four are not significant. The first may be tested by computing

$$d = 2(1 - r_1) - (Z_1^2 + Z_{46}^2) \Big/ \sum_{1}^{46} Z_t^2 = 1 \cdot 54,$$

and referring this to the tables of Durbin and Watson (1951). These tables give bounds for the 5 per cent. significance point for d. For a one-sided test against positive serial correlation the bounds show that the significance point lies between $1 \cdot 43$ and $1 \cdot 62$. The statistic d falls between these limits. However, Hannan (1957) has shown that for orthogonal polynomials the true significance point for d is very near to the upper bound as tabulated by Durbin and Watson. The value d is therefore significant and indicates positive serial correlation, the first serial correlation being about $0 \cdot 23$.

From the remaining r_j's it appears that a reasonable model of the process generating the residuals is a simple Markoff process. A test of goodness of fit of this model may be obtained by forming

$$R_j = r_j - 2r_1 r_{j-1} + r_1^2 r_{j-2}, \quad j \geqslant 2.$$

On the hypothesis of a simple Markoff process the R_j's are asymptotically normal and independent with zero mean and variance which may be estimated by

$$S^2=(1-r_1^2)^2/46.$$

Then $\chi^2_{p-1}=\sum_2^p R_j^2 \Big/ S^2$ is asymptotically distributed as χ^2 with $p-1$ degrees of freedom. Using the first 5 r_j's only, we obtain

$$\chi^2_4=0\cdot748,$$

which is not significant at any reasonable level.

This test is due to Quenouille (1947).

The presence of serial correlation naturally affects the test of significance of the regression coefficients, but it is apparent in this case that the effect will not appreciably alter the significance of the first term. In fact the variance of the regression coefficients will be increased in the ratio $(1+\rho)/(1-\rho)$ (Bartlett 1935), if the true residuals are Markovian with serial correlation ρ. Even allowing for a substantial downwards bias in r_1 as an estimate of ρ, this factor is still less than 2. No exact test of significance for the regression coefficients is available when serial correlation is present in the residuals, but an approximate procedure is to treat the test statistic $t_{B'}$ as having the same type of distribution as in the case of independent residuals but with an increased variance. Using the (conservative) figure 2 for the ratio in which the variance is increased, this procedure results in the treatment of the statistic $t_{B'}$ as having come from only half the number of observations, that is 23. This would leave the first coefficient significant at the 1 per cent. point. Furthermore, Grenander (1954) has shown that, asymptotically, for the orthogonal polynomials the straightforward least square procedure is as efficient as the best linear unbiased procedure (which could be used only if the nature of the process generating the true residuals were known). Thus there is no point in recomputing the regression to take account of the serial correlation.

IV. Acknowledgments

The author wishes to thank Mr. E. L. Deacon, Division of Meteorological Physics, C.S.I.R.O., for suggesting the problem; Mr. E. J. Hannan for some helpful discussions; and Mr. H. E. Gunther of the Geography Department of the Australian National University for drawing the graph.

V. References

Bartlett, M. S. (1935).—*J.R. Statist. Soc.* **98**: 536.
Deacon, E. L. (1953).—*Aust. J. Phys.* **6**: 209–18.
Durbin, J., and Watson, G. S. (1951).—*Biometrika* **38**: 159–78.
Grenander, U. (1954).—*Ann. Math. Statist.* **25**: 252–72.
Hannan, E. J. (1957).—*Biometrika* (in press).
Quenouille, M. H. (1947).—*J.R. Statist. Soc.* **110**: 123.