The Statistical Analysis of the Sunspot and Lynx Cycles
Author(s): P. A. P. Moran
Source: *The Journal of Animal Ecology*, Vol. 18, No. 1 (May, 1949), pp. 115-116
Published by: British Ecological Society
Stable URL: http://www.jstor.org/stable/1585
Accessed: 14/04/2009 17:55

THE STATISTICAL ANALYSIS OF THE SUNSPOT AND LYNX CYCLES

By P. A. P. MORAN
Institute of Statistics, University of Oxford

(With 1 Figure in the Text)

The regularity and stability of the 'ten-year' lynx cycle in Canada, as revealed in the numbers trapped by the Hudson Bay Company (Elton & Nicholson, 1942) naturally raised the question of it being in some way causally connected with the sunspot cycle. This question has been answered in the negative (Elton & Nicholson, 1942; MacLulich, 1937) and as appears from a careful examination of the figures, rightly so.

However, in view of the extremely interesting nature of these two series it may be worth while to set out here the criteria by which such a question must be decided, particularly in view of the fact that the correct application of statistical theory to such a problem is not at all widely understood.

In Fig. 1, I have graphed Wolfer's sunspot numbers from 1820 to 1934 and the common logarithms of the numbers of lynx caught in the Mackenzie River district of the Northern Department of the Hudson Bay Company (or its equivalent modern area). The sunspot numbers were taken from Yule (1927) for the period 1820–1924 and from the *Monthly Weather Review* for 1925–34. The lynx numbers were taken from Elton & Nicholson (1942) and the Mackenzie River district values were chosen because they gave an uninterrupted sequence from 1820 to 1934, and because taking the values for a single district results in a greater regularity in the cycle. Because of the extremely large values at each peak the common logarithm was taken in order to stabilize the oscillations. The ordinary correlation coefficient between these two series,

$$r = \frac{\sum_1^n (x_i - \bar{x})(y_i - \bar{y})}{\{\sum_1^n (x_i - \bar{x})^2 \sum_1^n (y_i - \bar{y})^2\}^{\frac{1}{2}}},$$

where $\bar{x}$ and $\bar{y}$ are the means of the sets of values $(x_1, \ldots, x_n)$, and $(y_1, \ldots, y_n)$, was found to be 0·1329.

MacLulich (1937) carried out a similar calculation (without taking logarithms) on series based on different data. His conclusion that the two cycles are not causally connected is certainly correct, but he reaches it by making a test of significance on the resulting correlation coefficient and in doing so commits a double error. He says (1937, p. 101): 'If the coefficient is less than the probable error there is no correlation and if it is greater than six times its probable error there is definite correlation.'

In the first place, if the test of significance is valid, significance is attained at the 5% level when the coefficient exceeds 2·91 times its probable error, and at the 1% level when it exceeds 3·82 times its probable error. His limits of significance are therefore too wide. The probable error, a number not now in common use because of certain theoretical reasons, is obtained by multiplying the standard error by 0·6745.

But in any case the test of significance is not valid here. Given any set of pairs of values

$$(x_i, y_i)\ (i = 1, \ldots, n)$$

a correlation coefficient can be calculated and is a valid descriptive measure of the way in which the two series of numbers vary together. But we also usually wish to test the significance of such a coefficient, i.e. to answer the question whether as large or larger a value could have been obtained by chance alone. To do this by the usual methods we must assume that different pairs are independent of each other so that the n pairs may be regarded as a random sample from bivariate population in which the variates are independent. But this condition is not at all true in the present case. For consider (x_i, y_i) and (x_{i+1}, y_{i+1}). Then x_{i+1} is much more likely to be near x_i in value than a randomly chosen x from the whole series, and similarly for y_i, y_{i+1}. We express this by saying that the series $\{x_i\}$ and $\{y_i\}$ are serially correlated in themselves. This fact completely invalidates the above method of testing the significance of the correlation coefficient.

The mathematical problem of setting up a test of significance in such circumstances is one of great difficulty. It may be shown, however (Moran, 1947), that for long series the standard error of the correlation coefficient is approximately equal to the square root of

$$n^{-1}(1 + 2\sum_1^\infty \rho_s \bar{\rho}_s),$$

where n is the length of the series and ρ_s, $\bar{\rho}_s$ are the serial correlation coefficients of the two series. The estimation from the data of these serial correlation coefficients itself presents considerable difficulties.

An account of their use in investigating the structure of a series will be found in Yule (1927), and there is a considerable literature on their estimation in recent papers in *Biometrika*, and in the *Journal* and *Supplement of the Royal Statistical Society*. See also Moran (1947). A coefficient may be considered to be significantly different from zero at the 5 % level if it is known that its distribution is approximately normal, about zero mean, and if it exceeds twice its standard error.

In general most time series are of such nature that the presence of the serial correlation coefficients in formula (1) results in the standard error being larger than would be the case if the successive pairs of observations were independent. In the present case $r = 0{\cdot}1329$, which would not be considered significant at the 5 % level if successive pairs of observations were independent. It would therefore be very unlikely to be judged significant by the more exact test using formula (1). The above described error of testing the significance of a correlation coefficient by ordinary methods when successive pairs of observations are not independent is so common in economic and other writing (e.g. Beveridge, 1944, Appendix D, p. 410) that it seems valuable to direct attention to it.

Although statistically significant correlation cannot be deduced from the data in Fig. 1, we can, nevertheless, say, on the other hand, that the possibility of sunspots being a major factor in causing the lynx cycle is decisively ruled out. For the oscillations of the latter are so regular that they cannot be explained by another regular cyclical phenomenon with which it is clearly in phase at some times and out of phase at others. Thus from 1820 to 1850 the cycles are very nearly in phase, whilst from 1850 to 1890 very nearly in opposite phase and so on. The regularity of the oscillation in both cases is so great that it is clear that there can be no causal relation between the two series.

Although not quite relevant in the present problem, two other sources of fallacy in the treatment of correlation between such series deserve mention. In the first place if there exist trends in the two series, large positive or negative correlations are to be expected but are not to be regarded as evidence of causal connexion, for clearly any two phenomena which show a pronounced trend throughout time will give such high values when correlated together.

Another common procedure when looking for causal connexions between such series is to lag one series one, two, or more units of time with respect to the other and calculate a correlation coefficient for each lag, picking out the largest. This may give suggestive results but once again the test of significance is invalidated, because we have chosen the largest of a set of correlation coefficients.

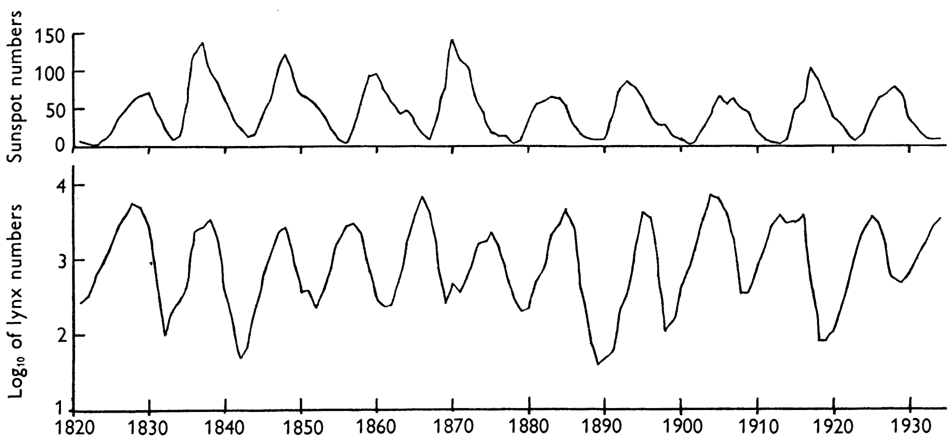

Fig. 1. Cycles in numbers of sunspots and trapped lynx.

SUMMARY

The statistical testing of correlation between cyclical time series is discussed. The fallacy of applying the ordinary test of significance to the correlation coefficient between series which are not serially independent in themselves, is pointed out, and other sources of fallacious conclusions mentioned. Examination of the sunspot and Canadian lynx cycles, however, rules out the possibility of a causal connexion between them.

REFERENCES

Beveridge, Lord. (1944). 'Full employment in a free society.' London.

Elton, C. & Nicholson, M. (1942). J. Anim. Ecol. 11: 215–44.

MacLulich, D. A. (1937). Univ. Toronto Stud. biol. 43: 1–136.

Moran, P. A. P. (1947). Biometrika, 34: 281–91.

Yule, G. U. (1927). Philos. Trans. A, 226: 267–98.