

## NOTE ON CORRELATION COEFFICIENTS DERIVED FROM CUMULATIVE DISTRIBUTIONS WITH REFERENCE TO GLACIOLOGICAL STUDIES

By J. T. ANDREWS, B. D. FAHEY and D. ALFORD

(Institute of Arctic and Alpine Research, University of Colorado, Boulder, Colorado 80302, U.S.A.)

**ABSTRACT.** In many areas of glaciology, cumulative degree days, either positive or negative, are regressed against another cumulative value, such as ablation or lake-ice growth. Very strong functional relationships are frequently found with high correlation coefficients. This note shows that, if pairs of random numbers are cumulated, the resulting correlation coefficients are extremely high with a Fisher transformed mean of  $r = 0.986$  and standard error of  $\pm 0.001$  (based on 50 individual computations of  $r$  which in turn were based on 20 cumulated pairs of random numbers between 0 and 99). These results indicate that caution must be exercised in the physical interpretation of data of this kind.

**RÉSUMÉ.** Note sur les coefficients de corrélation calculés à partir de distributions cumulatives en rapport avec les études glaciologiques. Dans beaucoup de domaines de la glaciologie, des degrés-jours cumulés, positifs ou négatifs, sont analysés par régression en fonction d'autres valeurs cumulées, telles que l'ablation ou la croissance de la glace de lac. Des relations fonctionnelles très étroites sont fréquemment trouvées avec de hauts coefficients de corrélations. Cette note montre que, en cumulant les valeurs d'un couple de nombres choisis au hasard, les coefficients de corrélation résultants sont extrêmement élevés avec une moyenne transformée de Fisher de  $r = 0,986$  et une erreur standard de  $\pm 0,001$  (basé sur 50 calculs individuels de  $r$  qui est à son tour basé sur 20 jours cumulés de nombres choisis au hasard entre 0 et 99). Ces résultats mettent en évidence les précautions à prendre dans l'interprétation physique de données de cette sorte.

**ZUSAMMENFASSUNG.** Bemerkung zu Korrelationskoeffizienten aus kumulativen Verteilungen mit Bezug auf glaziologische Untersuchungen. In vielen Gebieten der Glaziologie werden kumulative Zeitspannen, entweder positiv oder negativ, in Regression gegenüber anderen kumulativen Werten wie etwa der Ablation oder dem Wachstum des See-Eises gesetzt. Häufig ergeben sich dabei sehr starke funktionale Beziehungen mit hohen Korrelationskoeffizienten. Hier wird gezeigt, dass bei einer Kumulierung von Paaren zufälliger Zahlen die sich ergebenden Korrelationskoeffizienten extrem hoch sind mit einem nach Fisher transformierten Mittel von  $r = 0,986$  und einem mittleren Fehler von  $\pm 0,001$ . (Diese Werte beruhen auf 50 einzelnen Berechnungen von  $r$ , die ihrerseits auf 20 kumulative Paare von zufälligen Zahlen zwischen 0 und 99 gestützt sind.) Diese Ergebnisse zeigen, dass bei der physikalischen Interpretation von Daten dieser Art besondere Vorsicht geboten ist.

### INTRODUCTION

Measurement of the components of the surface-energy flux as they influence the growth or melt of ice in its various forms has yet to become a standardized technique. A number of indices of this energy flux, as it influences glaciological processes, have been proposed. Perhaps the most widely used of these is the "cumulative degree-day" index which is based upon the easily, if not always accurately, measured parameter of air temperature. In terms of this index, the cumulative increase or decrease of the glaciological parameter in question is plotted as some function of the cumulative departures of some aspect (i.e. the term is not standardized) of the daily temperature regime from a selected reference temperature, normally  $0^{\circ}\text{C}$ . The departure of the daily mean temperature is most often used, although other aspects of the diurnal temperature cycle have been used (Gartska and others, 1958). The strong functional relationship which generally results from these plots is taken as an indication that air temperature is, in fact, a valid index of the energy-exchange process as it influences the rate of change of the ice property in question.

The impetus for this short note was provided by the extremely high correlation coefficients,  $r$ , which several of us at the Institute of Arctic and Alpine Research had obtained in plotting a variety of bivariate cumulative frequency distributions involving the relationship of air temperature to ice accretion or ablation. The cumulative relationships commonly produced  $r$  values of  $>0.95$ . At the same time, values of  $r$  for non-cumulative values were frequently non-significant. We undertook, therefore, a study of the statistical significance of cumulative data by determining values of  $r$  produced by a bivariate cumulative frequency-distribution analysis of pairs of random numbers. It is the purpose of this note to discuss our findings and comment briefly on the "cumulative degree-days" concept as an index of the surface-energy flux.

## LITERATURE

A number of workers in the field of glaciology have proposed empirical equations relating ice accretion or ablation to air temperatures. Negative degree-days have been related to the rate of lake-ice formation (Andrews and McCloughan, 1961; Ragle, 1963; Jones, 1969), with the formation of sea ice (Zubov, 1945), with the thermal expansion and contraction of a lake-ice cover (Wilson and others, 1954) and with river-ice formation (Stevens, 1940). Negative degree-days form the basis for equations predicting the rate of front penetration into the soil (Carlson, 1952; Carlson and Kersten, 1953; Aldrich, 1956). Positive degree-days, as an index of the energy required for ice ablation, have been used by Clyde (1929), Church (1942), Zingg (1951), Schytt (1955), Gartska and others (1958), Keeler (1964), and Outcalt and MacPhail (1965). The general application of the technique has been discussed by Ingersoll and others (1949) and Linsley and others (1949). In every case, a significant statistical functional relationship was found to exist between cumulative air temperature, expressed in either positive or negative degree-days, and the rate of ice accretion or ablation. At the same time, there is little agreement among the various workers as to the quantitative relationship between the two. As reported by Gartska (1964), values of the point melt rate of snow vary by an order of magnitude among the various studies. There is a similar lack of agreement among the empirical equations derived for ice-growth prediction derived from cumulative degree-day studies.

## DATA AND ANALYSIS

To check the statistical significance of cumulative data, 50 groups of random numbers, each comprising 20 pairs, were summed and the data regressed for  $y$  on  $x$  ( $y = a + bx$ ). The random numbers ranged in value from 0 to 99, which we feel is realistic in terms of the length of time and spread for actual data.\* The regressions were computed using an Olivetti 101 Programma and checked on a CDC 6400 series computer. The 50 runs provided distributions for  $r$ ,  $a$  and  $b$ . The respective means and standard deviations are shown in Table I.† The results show, as expected, that  $a \rightarrow 0$  and  $b \rightarrow 1$ . Correlation coefficients are extremely high, with a sample mean of 0.986. On the other hand, if the random numbers are run as non-cumulative samples, the mean value of  $r$  is 0.236 ( $N = 10$  and  $r$  value not significant at the 95% level).

TABLE I. DESCRIPTIVE STATISTICS FOR SAMPLE DISTRIBUTIONS OF  $r$ ,  $a$  AND  $b$ 

$N$ 50	$r$	$r \ddagger$	$a$	$b$
Mean	0.984	0.986	-30.31	1.06
Range	0.042	—	472.08	1.35
Best estimate of $\sigma$	0.009	—	81.05	0.30

‡ Established from tables of Fisher's  $\mathcal{Z}$  (Arkin and Colton, 1963, p. 127).

Some idea of the statistical significance of  $r$  values derived from cumulative distributions can be gained by computing the standard error of the mean,  $\sigma_m$ .

For an  $N$  of 50, the standard error of Fisher's  $\mathcal{Z}$  is 0.0403. Re-converting back to  $r$  provides confidence limits which lie between 0.988 and 0.984.

## CONCLUSIONS

The degree-day index has been used extensively as a substitute for more detailed energy-flux data. While we do not question the probable physical relationship that exists between air temperature and ice accretion/ablation, we consider that the computed values of  $r$  for this, or any other cumulative distribution have to be examined very critically. Our study suggests that virtually any two parameters can be plotted as cumulative functions of each other and produce high correlation coefficients, irrespective of the existence of any physical relationship. At the 95% level, the confidence limits for pairs of cumulative random numbers are 0.984 and 0.988. Any value of  $r$  less than this can probably be explained

\* Values for  $r$  will be sensitive to the range of incremented values. If the numbers are large ( $> 100$ ) and the record short ( $N \approx 10$ ),  $r$  values could be less than those presented here.

† Samples involving high  $r$  values tend to be non-normally distributed. Thus, observations of  $r$  were transformed to Fisher's  $\mathcal{Z}$  (Arkin and Colton, 1963, p. 16), the distribution of which is normal even when the population value of  $r$  is large and the sample small.

largely in terms of the "smoothing" introduced by the analytical technique itself. This would suggest that the cumulative degree-day index of net surface-energy flux should not be used in correlation studies.

We did not analyze in detail the significance of the correlation produced by plotting cumulative values of one parameter against incremental changes in the other. One instinctively feels that less procedural bias is introduced by this second method but only an extended analysis similar to that described here will determine if this is correct. Calculation indicates that  $r$  based on random numbers is  $\approx 0.2$  for such procedures.

A cursory check of the relevant literature has failed to produce any results similar to those presented here, although we suspect that the findings are not unknown amongst statisticians.

#### ACKNOWLEDGEMENTS

The authors would like to thank Dr Roger G. Barry for his contributions to discussions on this topic and his critical review of a draft of this manuscript. This paper was prepared in part under research for contract DA-ARO-D-31-124-G1163 (U.S. Army Research Office, Durham, U.S.A.).

MS. received 17 July 1970

#### REFERENCES

- Aldrich, H. 1956. Frost penetration below highway and airfield pavements. *Bulletin. [U.S.] Highway Research Board*, 135, p. 124-44.
- Andrews, J., and McCloughan, C. 1961. Patterns of lake ice on Knob Lake, 1954-60. *McGill Sub-Arctic Research Papers*, No. 11, p. 64-90.
- Arkin, H., and Colton, R. 1963. *Tables for statisticians. Second edition.* New York, Barnes and Noble, Inc.
- Carlson, H. 1952. Calculation of depth of thaw in frozen ground. (In Frost action in soils: a symposium. *U.S. Highway Research Board. Special Report No. 2*, p. 192-222. ([U.S.] National Academy of Sciences—National Research Council Publication 213.))
- Carlson, H., and Kersten, M. 1953. Calculation of depth of freezing and thawing under pavements. *Bulletin. [U.S.] Highway Research Board*, 71, p. 81-95.
- Church, J. E. 1942. Snow and snow surveying; ice. (In Meinzer, O. E., ed. *Hydrology*. New York, McGraw-Hill Book Co., p. 83-148) (Physics of the Earth, Vol. 9.)
- Clyde, G. 1929. Snow-melting characteristics. *Technical Bulletin. Utah Agricultural Experiment Station*, No. 231.
- Gartska, W. 1964. Snow and snow survey. (In Chow, V. T., ed. *Handbook of applied hydrology*. New York, McGraw-Hill Book Co., p. 10-33, 34.)
- Gartska, W., and others. 1958. *Factors affecting snowmelt and streamflow*, by W. Gartska, L. Love, B. Goodell and F. Bertle. Washington, D.C., U.S. Government Printing Office.
- Ingersoll, L., and others. 1949. *Heat conduction with engineering and geological applications*, by L. Ingersoll, O. Zobel and A. Ingersoll. New York, McGraw-Hill Book Co.
- Jones, J. 1969. The growth and significance of white ice at Knob Lake, Quebec. *Canadian Geographer*, Vol. 13, No. 4, p. 354-72.
- Keeler, C. 1964. Relationship between climate, ablation and runoff on the Sverdrup Glacier, 1963, Devon Island, N.W.T. *Arctic Institute of North America. Research Paper No. 27*.
- Linsley, R., and others. 1949. *Applied hydrology*, by R. Linsley, M. Kohler and J. Paulhaus. New York, McGraw-Hill Book Co.
- Outcalt, S. I., and MacPhail, D. 1965. A survey of neoglaciation in the Front Range of Colorado. *University of Colorado Studies. Series in Earth Sciences*, No. 4.
- Ragle, R. 1963. Formation of lake ice in a temperate climate. *U.S. Cold Regions Research and Engineering Laboratory. Research Report 107*.
- Schytt, V. 1955. Glaciological investigations in the Thule Ramp area. *U.S. Snow, Ice and Permafrost Research Establishment. Report 28*.
- Stevens, J. 1940. Winter over-flow from ice gorging on shallow streams. *Transactions. American Geophysical Union*, 1940, Pt. 3, p. 973-78.
- Wilson, J., and others. 1954. A study of ice on an inland lake, by J. Wilson, J. Zumberge and E. Marshall. *U.S. Snow, Ice and Permafrost Research Establishment. Report 5*.
- Zingg, T. 1951. Beitrag zur Kenntnis des Schmelzwasserabflusses der Schneedecke, Schnee und Lawinen in der Schweizer Alpen, Winter 1949-50. *Winterberichte des Eidgenössischen Institutes für Schnee- und Lawinenforschung*, No. 14.
- Zubov, N. N. 1945. *L'dy Arktiki*. Moscow, Izdatel'stvo Glavsevmorputi. [English translation: *Arctic ice*. San Diego, U.S. Navy Electronics Laboratory, [1963?].]