Fritts, H. C. 1976. *Tree Rings and Climate*. Academic Press, London.
Fritts, H. C. 1991. *Reconstructing Large-scale Climate Patterns from Tree-ring Data*. The University of Arizona Press, Tucson.
Fritts, H. C., Blasing, T. J., Hayden, B. P. and Kutzbach, J. E. 1971. 'Multivariate techniques for specifying tree-growth and climate relationships and for reconstructing anomalies in palaeoclimate', *J. Appl. Meteorol.*, **10**(5), 845–864.
Fritts, H. C., Lofgren, G. R. and Gordon, G. A. 1979. 'Variations in climate since 1602 as reconstructed from tree rings', *Quat. Res.*, **12**(1), 18–46.
Fritts, H. C., Guiot, J., Gordon, G. A. and Schweingruger, F. 1990. 'Methods of calibration, verification and reconstruction', in Cook, E. R. and Kairiukstis, L. (eds), *Methods of Dendrochronology: Applications in the Environmental Sciences*. Kluwer Academic Publishers, Dordrecht, pp. 163–217.
Gittens, R. 1985. *Canonical Analysis*. Springer-Verlag, Berlin.
Glahn, H. R. 1962. An experiment in forecasting rainfall probabilities by objective methods. *Am. Meteorol. Soc.*, **90**, 59–67.
Glahn, H. R. 1968. Canonical correlation and its relationship to discriminant analysis and multiple regression. *J. Atmos. Sci.*, **25**(1), 23–31.
Gordon, G. A. 1982. 'Verification of dendroclimatic reconstructions', in Hughes, M. K., Kelly, P. M., Pilcher, J. R. and LaMarche, V. C., Jr. (eds), *Climate from Tree Rings*. Cambridge University Press, Cambridge, pp. 58–61. London.
Guiot, J. 1981. *Analyse mathmatique de donnes geophysiques, application la dendroclimatologie*. PhD thesis, Institut d'Astronomie et Geophysique, Louvain-la-Neuve.
Guiot, J. 1985. 'The extrapolation of recent climatological series with spectral canonical regression', *J. Climatol.*, **5**, 325–335.
Guttman, L. 1954. 'Some necessary conditions for common-factor analysis', *Psychometrika*, **19**, 149–161.
Helland, I. S. 1987. 'On the interpretation and use of R^2 in regression analysis', *Biometrics*, **43**, 61–69.
Jolliffe, J. T. 1986. *Principal Components Analysis*. Springer-Verlag, New York.
Jones, P. D., Wigley, T. M. L. and Briffa, K. R. 1987. *Monthly Mean Pressure Reconstructions for Europe (back to 1780) and North America (to 1858)*. Technical Report No. 37, US Department of Energy Carbon Dioxide Research Division, 99 pp.
Kaiser, H. F. 1960. 'The application of electronic computers to factor analysis', *Educ. Physchol. Meas.*, **20**, 141–151.
Kelly, P. M. and Sear, C. B. 1984. 'Climatic impact of explosive volcanic eruptions', *Nature*, **311**(5988), 740–743.
Kutzbach, J. E. 1970. 'Large-scale features of monthly mean Northern Hemisphere anomaly maps of sea-level pressure', *Mon. Wea. Rev.*, **98**(9), 708–716.
Lofgren, G. R. and Hunt, J. H. 1982. 'Transfer functions', in Hughes, M. K., Kelly, P. M., Pilcher, J. R. and LaMarche, V. C., Jr. (eds), *Climate from Tree Rings*. Cambridge University Press, Cambridge, pp. 50–56.
Lorenz, E. N. 1956. *Empirical Orthogonal Functions and Statistical Weather Prediction*. M. I. T. Statistical Forecasting Project Report No. 1, Contract AF 19 (604)-1566.
Mitchell, J. M., Jr., Stockton, C. W. and Meko, D. M. 1979. 'Evidence of a 22-year rhythm of drought in the western United States related to the Hale solar cycle since the 17th century', in McCormac, B. M. and Seliga, T. A. (eds), *Solar-Terrestrial Influences on Weather and Climate*. D. Reidel, Dordrecht, pp. 125–143.
Morrison, D. F. 1966. *Multivariate Statistical Methods*. McGraw-Hill, New York.
Nash, J. E. and Sutcliffe, J. V. 1971. 'Riverflow forecasting through conceptual models. 1, a discussion of principles', *J. Hydrol.*, **10**, 282–290.
Preisendorfer, R. W., Zwiers, F. W. and Barnett, T. P. 1981. *Foundations of Principal Component Selection Rules*. SIO Reference Series 81-4, Scripps Institution of Oceanography, La Jolla, California.
Rencher, A. C. and Pun, F. C. 1980. 'Inflation of R^2 in best subset regression', *Technometrics*, **22**(1), 49–53.
Ropelewski, C. F. and Halpert, M. S. 1987. 'Global and regional scale precipitation patterns associated with the El Niño–Southern Oscillation', *Mon. Wea. Rev.*, **115**, 1606–1626.
Sachs, M. S., Webb, T., III and Clark, D. R. 1977. 'Palaeoecological transfer functions', *Ann. Rev. Earth Planet. Sci.*, **5**, 159–178.
Sellers, W. D. 1968. 'Climatology of monthly precipitation patterns in the western United States', *Mon. Wea. Rev.*, **96**(9), 585–595.
Stocker, T. F. and Mysak, L. A. 1992. 'Climatic fluctuations on the century time scale: a review of high-resolution proxy data and possible mechanisms', *Climatic Change*, **20**, 227–250.
Stockton, C. W. and Meko, D. M. 1975. 'A long-term history of drought occurrence in Western United States as inferred from tree rings', *Weatherwise*, **28**(6), 244–249.
Webb, T., III and Bryson, R. A. 1972. Late- and post-glacial climatic change in the northern midwest, USA: quantitative estimates derived from fossil pollen spectra by multivariate statistical analysis. *Quat. Res.*, **2**, 70–111.
Web, T., III and Clark, D. R. 1977. 'Calibrating micropalaeontological data in climatic terms: a critical review', *Annals of the New York Academy of Sciences*, **288**, 93–118.
White, R. M., Cooley, D. S., Derby, R. C. and Seaver, F. A. 1958. The development of efficient linear statistical operators for the prediction of sea-level pressure. *J. Meteorol.*, **15**, 426–434.

APPENDIX

validation

Verification tests of dendroclimatic reconstructions

The sign test. The sign test (ST) is a simple non-parametric test of the similarity between series based on a count of the number of agreements and disagreements in sign. If the number of agreements exceeds the number of disagreements by greater than that expected by chance alone, the reconstruction passes. For $n < 50$, a test based on the binomial distribution must be used or tabulated critical values can be obtained

ST <- function(x,y) {
length(sign(x) = sign(y))}

(e.g. Fritts, 1976). For $n > 50$, a normal approximation can be made yielding the test statistic for an α-level of 0·05, where n is the number of years in the verification period. The ST does not make any assumptions about the underlying distribution of the data. Hence, it is robust to any violation of the normality constraint should this occur. The ST can be applied either to the departures from the mean of the actual and reconstructed data or, as used here, to the sign changes associated with the first-differences of each series. The first-differences are computed simply as

$$\Delta Y_t = Y_t - Y_{t-1}$$

The ST based on first transforming each series to differences (as above) is a very strict measure of the high-frequency agreement between series and is essentially immune to the affects of any trend or positive autocorrelation in the original series.

Product means test. The first-difference ST measures only the agreement in the direction of change from one year to the next. As such it throws away all information about the magnitudes of the yearly changes. In contrast, the product means (PM) test takes into account both the signs and the magnitudes of the changes when testing for agreement between series. It is calculated from the cross-products of the acual and estimated yearly departures from their respective mean values. If the signs of the departures are the same, the cross-products are positive; if the signs of the departures are different, the cross-products are negative. The difference between the mean positive cross-products and the mean negative cross-products is tested using a standard t-test

$$t = \frac{m_1 - m_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$$

where m_1 is the mean of the positive products, m_2 is the mean of the negative products, s_1^2 and s_2^2 are the corresponding sample variances, and n_1 and n_2 are the number of positive and negative cross-products (Fritts, 1976). If the mean positive cross-product is significantly greater than the mean negative cross-product based on the t-test, the reconstruction passes this verification test.

Product-moment correlation coefficient. The product-moment correlation coefficient (r) is found in virtually all introductory statistics books. It is a powerful statistic for testing the association between two variables. One of its assumptions is that the series being cross-correlated are bivariately normally distributed, which is a rather restrictive assumption. As a consequence, the product-moment correlation coefficient is not very robust. However, its use in dendroclimatology is widespread, so we include it here. It is calculated as

$$r = \frac{\sum_{i=1}^{n} (x_i - \bar{x}_v)(\hat{x}_i - \bar{\hat{x}}_v)}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x}_v)^2 \sum_{i=1}^{n} (\hat{x}_i - \bar{\hat{x}}_v)^2}}$$

where and $\bar{x}_v$ and $\bar{\hat{x}}_v$ are the means of the actual and estimated data, respectively, for the verification period. Under the null hypothesis $H_0: r = 0$, the significance of the calculated r is easily tested using a t-test (Fritts, 1976). Note that both means are removed when calculating r. As a consequence, the correlation coefficient measures association in relative terms only and, thus, is not sensitive to any discrepancy between the actual and estimated means in the verification period.

Reduction of error. The reduction of error (RE) statistic was first described by Lorenz (1956) as a way of determining if a meteorological forecast was better than climatology (i.e. the mean of the meteorological data in the calibration period). It is calculated as

$$\mathrm{RE} = 1{\cdot}0 - \left[\frac{\sum_{i=1}^{n} (x_i - \hat{x}_i)^2}{\sum_{i=1}^{n} (x_i - \bar{x}_c)^2}\right]$$

where $\bar{x}_c$ is the mean of the actual data in the calibration period. The value of RE ranges from $+1{\cdot}0$ to $-\infty$, with an RE $= 0$ being no better than climatology. Thus, RE > 0 indicates that the reconstruction is better than the calibration period mean. However, there is no way of testing the RE for statistical significance. Note that $\bar{x}_c$ in the denominator will not produce the true corrected sum-of-squares unless it is identical to the verification period mean. Consequently, a large difference in the calibration and verification means can lead to an RE greater than the square of the product-moment correlation coefficient. This occasional odd behaviour suggests that the RE should be interpreted cautiously when the data contain trends or are highly autocorrelated. Fritts (1976) notes that RE is an extremely rigorous verification statistic because it has no lower bound. As a consequence, only a few bad tree-ring estimates are needed to result in a negative RE even though most of the reconstruction is of useful quality. Because of this lack of robustness, positive REs can be very difficult to obtain in practice.

Coefficient of efficiency. The coefficient of efficiency (CE), introduced into dendroclimatology only recently (Briffa *et al.*, 1988b), was first described in the hydrology literature as an expression of the true R^2 of a regression equation when it is applied to new data (Nash and Sutcliffe, 1971). The CE is calculated as

$$\mathrm{CE} = 1{\cdot}0 - \left[\frac{\sum_{i=1}^{n} (x_i - \hat{x}_i)^2}{\sum_{i=1}^{n} (x_i - \bar{x}_v)^2}\right]$$

where $\bar{x}_v$ is the mean of the actual data in the verification period. Like the RE, the CE ranges from $+1{\cdot}0$ to $-\infty$. Note that the only difference between the RE and CE lies in the denominator. Although this difference appears to be trivial, in fact large differences in the RE and CE can occur. When $\bar{x}_v = \bar{x}_c$, CE = RE. However, when $\bar{x}_v \neq \bar{x}_c$, RE will be greater than the CE by a factor related to that difference. This follows by noting that for the CE, the sum-of-squares in the denominator is fully corrected because $\bar{x}_v$ is the proper mean. However, for the RE, the denominator sum-of-squares will not be corrected fully unless the calibration period mean is fortuitously identical to the verification period mean. When this is not the case, the denominator sum-of-squares of the RE will be larger than that of the CE and, therefore, RE > CE. Because the CE will usually be less than the RE, it is even more difficult to pass. Like the RE, there is no significance test for the CE, but as a 'rule of thumb', a CE > 0 indicates some useful information in the climatic reconstruction.

:RNATIONAL JOURNAL OF CLIMATOLOGY, VOL. 14, 403–414 (1994) 551.577.21 : 77:551.502.3

ESTIMATING CONTINENTAL AND TERRESTRIAL PRECIPITATION AVERAGES FROM RAIN-GAUGE NETWORKS

CORT J. WILLMOTT

Center for Climatic Research, Department of Geography, University of Delaware, Newark, Delaware 19716, USA

SCOTT M. ROBESON

Department of Geography, Indiana University, Bloomington, Indiana 47405, USA

AND

JOHANNES J. FEDDEMA

Department of Geography, University of California, Los Angeles, California 90024, USA

Received 13 November 1992
Accepted 14 July 1993

ABSTRACT

Influences of varying rain-gauge networks on continental and terrestrial precipitation averages (derived from data observed on those networks) are evaluated. Unsystematically and systematically designed station networks are considered, the latter being represented by the NCAR World Monthly Surface Station Climatology, which contains hand-picked but time-varying networks that date back to the 1800s. Biases arising from spatially uneven and temporally variable precipitation-observing networks can be significant.

For all the continents, except South America, sparse rain-gauge networks produce overestimates of continental mean precipitation. Mean precipitation for South America, in contrast, is underestimated substantially by low densities of observing stations. Sampling errors tend to be large in areas of high precipitation and in regions with strong spatial precipitation gradients (e.g. in the Sahel). These patterns occur whether the station network has been selected systematically (as in the NCAR network) or unsystematically.

Systematic sampling of mean precipitation (at the NCAR station locations), however, suggests that many yearly NCAR station networks are adequate for estimating continental average precipitation. As early as 1890, NCAR networks for Australia resolve continental average precipitation accurately. Not until 1960, however, do NCAR networks for South America begin to resolve continental mean precipitation adequately. Regional and continental NCAR network errors also tend to cancel one another, often giving accurate yearly estimates of terrestrial mean precipitation.

KEY WORDS Precipitation averages Rain-gauge networks

INTRODUCTION

Continental and terrestrial averages of precipitation are key components of the global hydrological cycle, yet completely reliable estimates of these spatial means are not available (Willmott and Legates, 1991). Even less is known about the seasonal and interannual variability in continental and terrestrial precipitation averages. Although global climate model (GCM) simulations and remotely sensed estimates of large-scale precipitation are improving (Legates and Willmott, 1992), historical rain-gauge records and networks continue to comprise the bases for the most credible estimates (Legates and Willmott, 1990; Hulme, 1992). Large-scale spatial averages made from historical data, in other words, comprise our best understanding of the spatial, seasonal and interannual variability in continental and terrestrial averages of precipitation. The reliability and variability of continental and terrestrial precipitation averages that have been derived from the historical rain-gauge record are the subject of this paper.

CCC 0899-8418/94/040403-12