The Effect of Likely Biases in Estimating the Variance of Long Time Averages of Climatological Data

ROLAND A. MADDEN

National Center for Atmospheric Research, Boulder, Colorado*

RICHARD H. JONES

School of Medicine, University of Colorado, Denver, Colorado

(Manuscript received 25 April 1996, in final form 27 June 1996)

1. Introduction

Bias in estimating the variance of independent identically distributed (IID) random variables with an unknown mean is well known and readily handled in the time domain by the $n - 1$ factor (e.g., Mood and Graybill 1963, 183). A similar factor can be derived for autocorrelated data (Trenberth 1984) to adjust for bias. Variance can also be determined from the frequency spectrum of a time series, and there are interesting analogous adjustments that can be made to compensate for bias (Trenberth 1984, his appendix). These adjustments depend on the spectral densities near zero frequency, which cannot be estimated directly from the data because they are at timescales longer than the length of the time series. Interestingly, when computing the variance of time averages, these near-zero frequency spectral densities, which must be based on assumptions, become more important as the averaging time becomes longer (Jones 1975). In this paper we consider the effect of various choices or assumptions about the unresolved low-frequency spectral densities on estimates of the variance of averages.

2. Method

Madden and Ramanathan (1980, MR hereafter), in a study directed at detecting climate change due to increasing atmospheric carbon dioxide (''signal''), considered the effect of bias in their estimates of the inherent variability (''noise''), or variability not related to possible changes driven by the increasing CO_2. The data that they examined were 72 yr of monthly and zonally averaged surface temperatures at 60°N. It was assumed that the possible signal was small enough so that the variance of the data could be considered as all noise. The spectrum [$s(f)$] of the data was used to estimate the variance in the following way:

$$\sigma^2 = \int_{f=0}^{0.5} s(f)\, df. \tag{1}$$

Accordingly, the variance of time averages of N values was determined from (Jones 1975)

$$\sigma_N^2 = \int_{f=0}^{0.5} s(f) H_N^2(f)\, df. \tag{2}$$

Here $H_N(f)$ is the amplitude of the frequency response of an average of N observations. It is the Fourier transform of N values each equal to $1/N$, or the weights of the averaging process. That is,

$$H_N^2(f) = \left[\frac{\sin(\pi N f)}{N \sin(\pi f)}\right]^2. \tag{3}$$

Figure 1 presents $H_N^2(f)$ for several averages. Equation (2) and Fig. 1 demonstrate the feature that Jones (1975) stressed, namely that the low frequencies of the spectrum are solely important in estimating variances of time-averaged data.

With only 72 yr of data, MR's spectrum at low frequencies was heavily influenced by $s(f = 0)$, the sum of their arbitrarily truncated autocovariances. Because the variance of time-averaged data was important in their study, MR looked to published paleoclimate data that could give a further clue as to what might be expected for the behavior of the spectrum at periods longer than their 72-yr data record. Kutzbach and Bryson (1974), in their attempt to mesh both instrumental and noninstrumental records, estimated that the average

* The National Center for Atmospheric Research is sponsored by the National Science Foundation.

Corresponding author address: Dr. Roland A. Madden, National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307-3000.

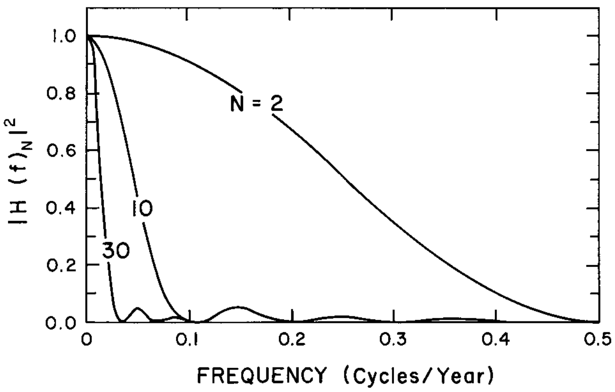

FIG. 1. Square of the amplitude response $|H(f)_N|^2$ for averages of $N = 2$, 10, and 30 yr.

spectral value between 100- and 1000-yr periods is about twice that observed for periods near but shorter than 100 yr. Madden and Ramanathan (1980) then asked what effect underestimating the 100–1000-yr variance by a factor of 2 would have on estimating the variance of time averages. To answer this question, MR simply took a spectrum estimated from the available 72-yr record and increased the low-frequency values by a factor of 2, or

$$S(f) = \begin{cases} s(f), & \text{for } f > .01 \text{ cycles yr}^{-1}; \\ s(f) \times 2, & \text{for } f \leq .01 \text{ cycles yr}^{-1}. \end{cases} \quad (4)$$

Equation (2) was then evaluated for $s(f)$ and $S(f)$, and the results were compared. Madden and Ramanathan (1980) concluded that the bias was probably less than the uncertainties, due to sampling variability for averages of 30 yr or shorter.

3. A further look at bias effects

a. Estimates of bias by truncating a long series

Madden and Ramanathan (1980) somewhat arbitrarily assumed that their spectrum underestimated the variance at periods longer than 100 yr by a factor of 2. In this section we examine the effects of bias on the variance of time averages by considering a time series that is more than 300 yr long and, therefore, is one for which at least some of the very low frequency variance is known. The time series is that of central England temperatures originally published by Manley (e.g., 1974) and updated by Parker et al. (1992). Figure 2 shows the spectrum of the full 336-yr series (dashed line). The average was removed, and the autocovariance was estimated out to a maximum of 71 lags. A Parzen window was applied to these autocovariances, resulting in spec-

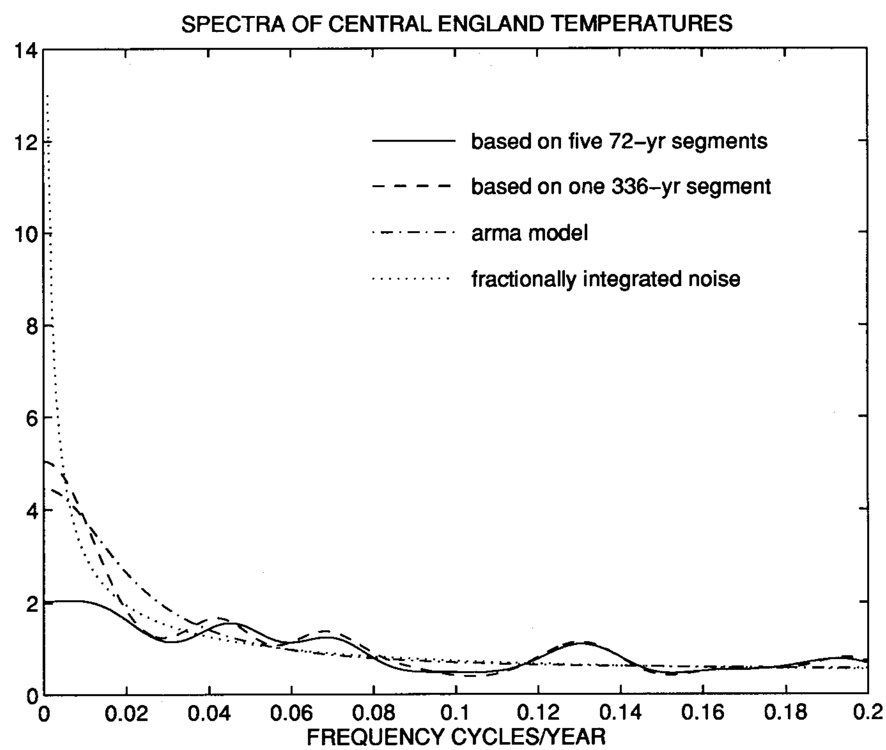

FIG. 2. Biased spectrum based on five 72-yr segments (solid), a less biased spectrum based on one 336-yr segment (dashed), the ARMA (1,1) spectrum (dash–dot), and the fractionally integrated noise spectrum ARIMA (1,d,0). Spectral density units are (°C)^2 × yr. Areas are proportional to variance.

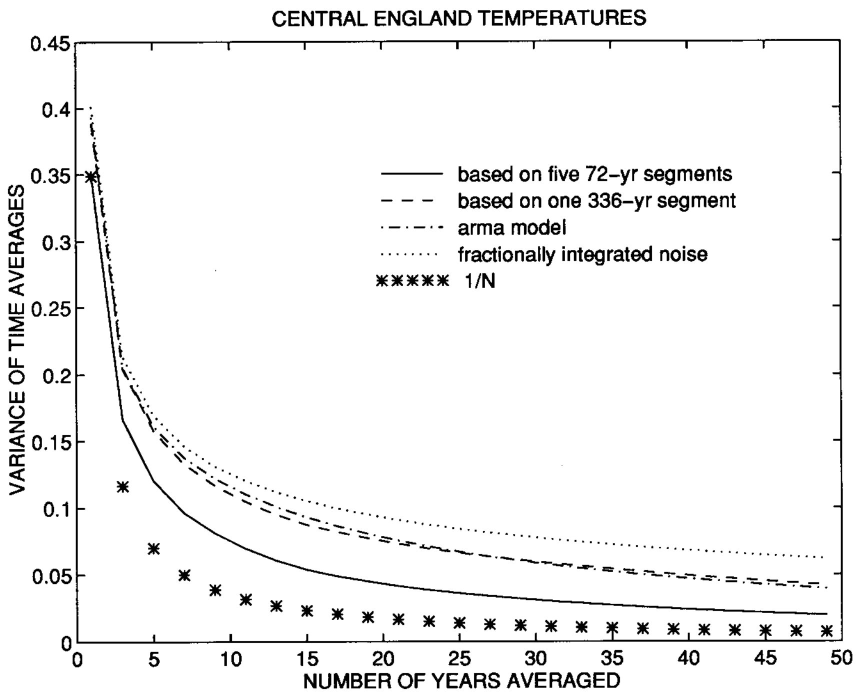

FIG. 3. Variance of time averages in (°C)^2 determined from (2) and the biased spectrum (solid), less biased (dashes), ARMA (1,1) (dash–dot), and fractionally integrated spectra of Fig. 2 (dots). Asterisks denote the reduction of variance if data is independent, $1/N$.

tral estimates with approximately 17.6*df* and a bandwidth of 0.026 cycles yr^{-1} (Jenkins and Watts 1968, 252). The spectrum resolves variance to about 0.013 cycles yr^{-1} or 77-yr periods. This spectrum will be referred to as the ''less biased'' spectrum.

This particular time series is extraordinarily long. More common are time series of lengths similar to the one studied by MR, namely a 72-yr record. To estimate the effects of bias if only such a short record of central England temperatures were available, the 336-yr time series was broken up into five 72-yr segments (a 6-yr overlap was necessary between each segment). A spectrum was then computed by repeating the above procedure for each segment, with a maximum again of 71 lags, and then averaging over all five segments. This is represented by the solid line in Fig. 2. The bandwidth is still about 0.026, and the *df* are $5 \times 3.71 \approx 18.8$. The apparent increase in *df* results from the six data point overlap between segments. This spectrum is the ''biased'' spectrum.

Figure 3 shows the variance of time averages based on (2) and the biased spectrum of Fig. 2 (solid line), and on the less biased spectrum of Fig. 2 (dashed line). Integrations were approximated using the trapezoidal rule and a frequency interval of 0.001. The variances of the yearly data ($N = 1$) of 0.35 C^2 from the biased (72-yr segment) and 0.39 C^2 from the less biased spectrum (336-yr segment) are the areas under the respective spectral curves in Fig. 2. The difference is almost completely accounted for by the area between the dashed and solid lines near zero frequency. Independence is often assumed between values, and the resulting decline in variance with averaging, decreasing as $1/N$, is indicated by the asterisks in Fig. 3.

b. Estimates of bias by assuming a long-memory statistical model

As described above, variance based on the 336-yr central England temperature record is less biased than that based on a shorter subset. Without measurements, we do not know what the true very low frequency (lower than 336^{-1} cycles yr^{-1}) variance is. Our approach has resulted in a spectral estimate at zero frequency that is equal to the sum of a somewhat arbitrarily chosen number of Parzen-weighted autocovariances. In this section we model the data as a fractionally integrated Autoregressive Integrated Moving Average (ARIMA) (p, d, q) process, discussed by Hosking (1981). Bloomfield (1992), in a study of the uncertainties in estimates of trends, also modeled his data as a fractionally integrated ARIMA(p, d, q) process. This process does not force the spectral density to be bounded at $f = 0$. It is an example of a long-memory process. Its spectrum takes the form (e.g., Brockwell and Davis 1991, 522)

$$s(f) \approx c/f^{2d} \text{ as } f \to 0 \quad \text{and } -.5 < d < .5, \tag{5}$$

where c is a constant. For $0 < d < 0.5$, the zero-frequency spectral estimate is infinite, but the integral of $s(f)$ over f, and correspondingly the variance of the process, are bounded.

The fractionally integrated ARIMA process was fitted to the data by maximum likelihood, with d constrained to be in the interval $0 < d < 0.5$ so that the process was stationary. The maximum likelihood estimation also estimated a constant mean together with its standard deviation, giving a maximum likelihood estimate of the variance of the time average based on the fitted error model. Several error models were fit, and the best fit was determined (by Akaike's information criterion; Akaike 1974) to be an ARIMA (1, d, 0), with d estimated as 0.317. The maximum likelihood estimate of the standard deviation of the estimated 336-yr mean was 0.1956. Fitting ARMA models to the data did not provide as good a fit as the fractionally integrated ARIMA model. The best fit was an ARMA (1,1) model, and the estimate of the standard deviation of the estimated mean was 0.0804.

The differences in these two estimated standard deviations of the mean can be attributed to the very different shape of the spectral densities near zero frequency. The fractionally integrated ARIMA process goes off to infinity at zero, while the ARMA spectral density is finite. The spectral density calculated from the estimated parameters in the fractionally integrated ARIMA model is

$$s(f) = \frac{0.668}{|1 + 0.1312e^{2\pi if}|^2|1 - e^{2\pi if}|^{.634}},$$

while the estimated spectral density for the ARMA (1, 1) model is

$$s(f) = \frac{0.673|1 - 0.673e^{2\pi if}|^2}{|1 - .873e^{2\pi if}|^2}.$$

Both spectra are included in Fig. 2.

To evaluate (2) numerically for the fractional ARIMA model, the spectral density near zero can be integrated in closed form using the asymptotic expression (5). The integral from zero to f is

$$\frac{c}{1 - 2d} f^{1-2d}.$$

The constant is chosen so that the asymptotic form of the spectral density matches the actual value at the first frequency that is calculated, f_1. This gives

$$c = s(f_1) f_1^{2d},$$

so the integral of the spectral density from zero to f_1 is

$$s(f_1) f_1^{2d} \int_0^{f_1} f^{-2d}\, df = \frac{f_1 s(f_1)}{1 - 2d}.$$

To integrate the rest of the spectral density numerically for frequencies f_1 to 1/2, the trapezoidal rule is again used.

The variances of time averages based on these model spectra are also shown in Fig. 3. Values for the ARMA (1,1) process, shown as the dash–dot curve, are nearly identical to those derived from the less biased (336-yr segment) spectrum. They become slightly smaller for averages longer than about 30 yr because of the larger zero-frequency spectral estimate of the less biased spectrum. Values for the ARIMA (1,d,0) are largest of all (dotted curve in Fig. 3) because the near-zero-frequency spectrum approaches infinity. Difficult to see in the figure is the fact that the estimated variance of 1-yr averages is slightly above 0.4 for the fractionally integrated spectrum and slightly under 0.4 for the less biased spectrum.

4. Conclusions

The instrumental record typically only allows reasonable estimates of the spectrum to periods of less than 100 yr. It is variability on this and longer timescales that determines the variance of long time averages. We do not know what this very low frequency variability is. Here we have attempted to underscore the problem of the resulting bias in our estimates of the variance of time averages by examining the near-zero spectral estimate that can be determined from an extraordinarily long time series. We have also considered, as an extreme case, results for a fractionally integrated model, which does not constrain the zero-frequency spectral estimate to be finite. Although we have approached the problem of bias in the frequency domain, results apply equally to variances computed in the time domain.

To summarize the results, let us take the less biased spectrum to be the "true" spectrum. Figure 4 presents the ratio (times 100%) of the variances determined from the biased spectra to those from the less biased spectra, or the ratio of variances one would expect to compute from a 72-yr segment to those from a 336-yr segment. For very long time averages the ratio approaches 40%, which is the ratio of the respective zero-frequency spectral estimates. An important thing to note is that a serious underestimation of the variance occurs for averages approximately 15 yr long and longer. The underestimate if data are assumed independent is considerably worse, as indicated by the asterisks in Fig. 4. If the fractionally integrated spectrum is closer to reality, then the percentage of the true variance determined from the biased spectrum (dotted curves in Fig. 4) is less than if the less-biased spectrum were true, although not a great deal less. It is near 50% for 15-yr averages and less than 40% for 40-yr averages. We cannot argue that the underestimates of variance indicated in Fig. 4 are necessarily representative of a wide range of time series, but these results do suggest that, for long time averages, the bias effect may often be important.

Acknowledgments. We thank P. D. Jones for providing the central England temperatures. R. H. Jones is supported by the Geophysical Statistics Project at the National Center for Atmospheric Research under Grant

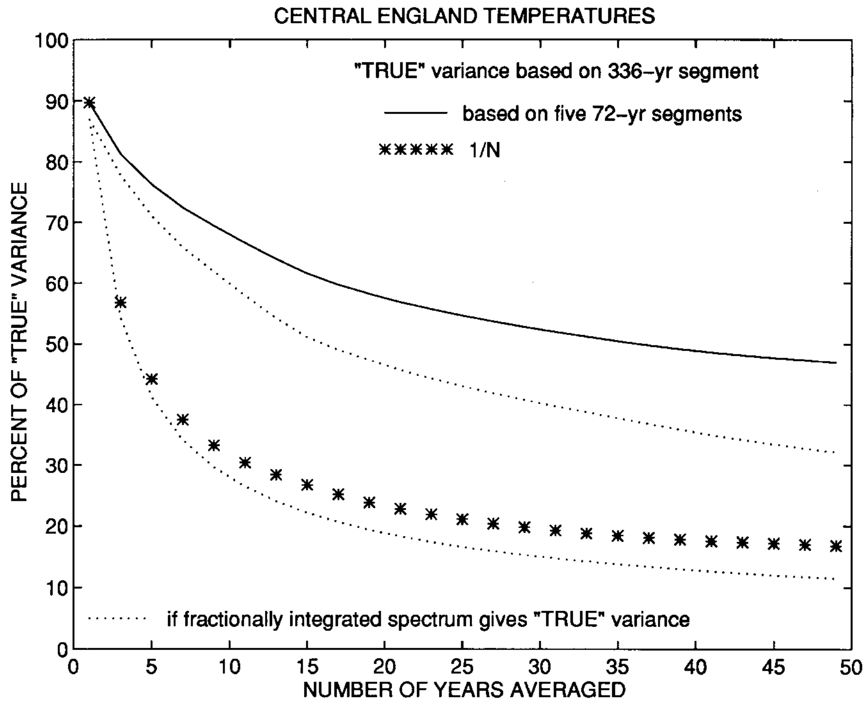

FIG. 4. Ratio (in percent) of the variance of time averages estimated from the biased spectrum (solid line of Fig. 3), or from assuming a white noise spectrum (asterisks), to that estimated from the less biased spectrum (dashed line of Fig. 3). Dotted curves are similar ratios, but with values from the fractionally integrated spectrum (dotted line of Fig. 3) in the denominator.

NSF DMS9312686. L. Stephens expertly assembled this manuscript and one for a preprint volume that was its forerunner.

REFERENCES

Akaike, H., 1974: A new look at the statistical model identification. *IEEE Trans. Autom. Control,* **AC–19,** 716–723.

Bloomfield, P., 1992: Trends in global temperature. *Climate Change,* **21,** 1–16.

Brockwell, P. J., and R. A. Davis, 1991: *Timeseries: Theory and Methods.* 2d ed. Springer-Verlag, 577 pp.

Hosking, J. R. M., 1981: Fractional differencing, *Biometrika,* **68,** 165–176.

Jenkins, G. M., and D. G. Watts, 1968: *Spectral Analyses and its Applications.* Holden Day, 525 pp.

Jones, R. H., 1975: Estimating the variance of time averages. *J. Appl. Meteor.,* **14,** 159–163.

Kutzbach, J. E., and R. A. Bryson, 1974: Variance spectrum of holocene climatic fluctuations in the North Atlantic sector. *J. Atmos. Sci.,* **31,** 1958–1963.

Madden, R. A., and V. Ramanathan, 1980: Detecting climate change due to increasing carbon dioxide. *Science,* **209,** 763–768.

Manley, G., 1974: Central England temperatures: Monthly means 1659 to 1973. *Quart. J. Roy. Meteor. Soc.,* **100,** 389–405.

Mood, A. M., and F. A. Graybill, 1963: *Introduction to the Theory of Statistics.* McGraw-Hill, 443 pp.

Parker, D. E., T. P. Legg, and C. K. Folland, 1992: A new daily central England temperature time series, 1772–1991. *Int. J. Climatol.,* **12,** 317–342.

Trenberth, K. E., 1984: Some effects of finite sample size and persistence on meteorological statistics. Part I: Autocorrelations. *Mon. Wea. Rev.,* **112,** 2359–2368.