

NOTES AND CORRESPONDENCE

A Quantitative Estimate of the Effect of Aliasing in Climatological Time Series

ROLAND A. MADDEN

National Center for Atmospheric Research, Boulder, Colorado*

RICHARD H. JONES

University of Colorado Health Sciences Center, and the Geophysical Statistics Project, National Center for Atmospheric Research, Boulder, Colorado*

7 July 2000 and 2 May 2001

1. Introduction

The phenomenon of aliasing is well known and covered in most basic textbooks on time series analysis. For many climate studies we need to minimize aliasing effects. For example, to better understand the dependence of climate variables on differing forcings, we need to be able to determine how large variability is at various timescales. Estimates of variance that occurs at low frequencies can be compromised by variance aliased from higher frequencies.

There is extensive literature estimating aliasing effects on data associated with what might be referred to as weather timescales (short), and some pointing out how specific variations might alias to climate timescales (e.g., Gray and Madden 1986; Wunsch 2000). However, to our knowledge, there is none providing general estimates of the possible size of aliased variations onto climate timescales. Results from simple calculations estimating the size of aliasing effects with averaging and sampling strategies often used in climate analysis surprised us by their magnitude. This note describes these results.

A parameter that we use to summarize the likely effects of aliasing in an analysis is the ratio of averaging length (AL) to sampling interval (SI), AL/SI. We stress that for $AL/SI < 1$, considerable aliased variance can appear at all, even the lowest, frequencies. A survey of articles published during the past four years in the *Jour-*

nal of Climate revealed one that showed spectra with $AL/SI = 1/2$ (6-month averages sampled once per year), one with $AL/SI = 5/12$ (5-month averages sampled once per year), and three with $AL/SI = 1/4$ (3-month averages sampled once per year). It is likely that all these spectra were badly aliased at all frequencies. Recent interest in decadal variability makes it important that the aliasing problem be fully appreciated.

The theoretical example to follow is based on the spectra of first-order autoregressive processes (AR1), and the empirical one is based on spectra of an observed time series of daily pressure data averaged over different lengths of time and sampled at different time intervals.

2. Theoretical example

A spectrum analysis of a time series resolves variance from low frequencies out to a high “cutoff” (“Nyquist,” or “folding”) frequency $f_c = 0.5$ cycle per sampling interval. Frequencies aliased to a resolved frequency f ($0 \leq f \leq f_c$) are

$$(2f_c \pm f), (4f_c \pm f), \dots, (2nf_c \pm f)$$

(e.g., Bendat and Piersol 1970, p. 229). The unaliased spectrum of a process that resolves all variance at its true frequency then “folds” about the cutoff frequency to produce the aliased spectrum (see Blackman and Tukey 1958, p. 120; Bendat and Piersol 1971, p. 230). We use that principle to compute aliasing effects for an AR1 process.

A zero mean AR1 process is

$$y(t) = \phi y(t-1) + \varepsilon_t, \quad (1)$$

where we take t to be time in integer days, ϕ is the lag-1-day autocorrelation, and ε_t is a random input with zero mean and variance σ^2 . The spectrum is

* 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.

E-mail: ram@ucar.edu

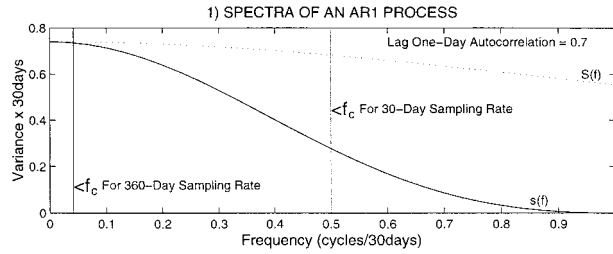


FIG. 1. Unaliased spectrum of an AR1 process with a lag-1-day autocorrelation of 0.7 [$S(f)$ dotted], and the unaliased spectrum of 30-day averages of the same AR1 process [$s(f)$ solid]. The folding frequency (f_c) for sampling every 30 days and every 360 days is indicated by the vertical lines.

$$S(f) = \frac{\sigma^2}{(1 + \phi^2 - 2\phi \cos 2\pi f)}, \quad (2)$$

where f is the frequency. We assume that (2) gives us the unaliased spectrum of daily data. The spectrum of time averages is

$$s(f) = |H(f)|^2 S(f). \quad (3)$$

Here, $H(f)$ is the transfer function of the averaging process (Blackman and Tukey 1958, p. 25 and p. 112). For a simple running average of T values, $H(f)$ is the Fourier transform of T weights equal to $1/T$, or

$$H(f) = \frac{\sin(T\pi f)}{T \sin(\pi f)}. \quad (4)$$

Figure 1 shows the spectrum $s(f)$ of a time series of 30-day averages of an AR1 process with a lag-1-day autocorrelation of 0.7. It was determined by (3) with $\phi = 0.7$ in (2) and $T = 30$ in (4). The spectrum can be scaled so that areas under it are equal to variance. Although only plotted to 1 cycle $(30 \text{ day})^{-1}$ the variance is distributed in the frequency range $0 \leq f \leq f_c$, where $f_c = 15 \text{ cycles } (30 \text{ day})^{-1}$ ($0.5 \text{ cycle day}^{-1}$). Here $s(f)$ is the unaliased spectrum (we assume there is no aliasing from frequencies higher than $0.5 \text{ cycle day}^{-1}$). No matter how the data were to be sampled, and f_c correspondingly changed, the total variance would continue to be that given by the unaliased spectrum. If the sampling interval were once every 30 days (approximately equivalent to a time series of monthly data) instead of once every day, then f_c would equal $0.5 \text{ cycle } (30 \text{ day})^{-1}$ as indicated by the middle vertical line in Fig. 1. All the variance depicted by the area under $s(f)$ to the right of, or at higher frequencies than, f_c would fold or alias into the new resolved frequency range $0 \leq f \leq f_c = 0.5 \text{ cycle } (30 \text{ days})^{-1}$. In this case the largest aliasing is near f_c and there is no aliasing near zero frequency because of the zero in the unaliased spectrum at $1 \text{ cycle } (30 \text{ days})^{-1}$.

If the sampling interval were every 360 days (approximately equivalent to a time series of, say Januaries), then all variance at frequencies higher than $f_c = 0.5 \text{ cycle } (12 \times 30 \text{ days})^{-1}$ would fold into the resolved frequency range $0 \leq f \leq f_c$. This f_c is indicated by a vertical line toward the left in Fig. 1. It is clear that

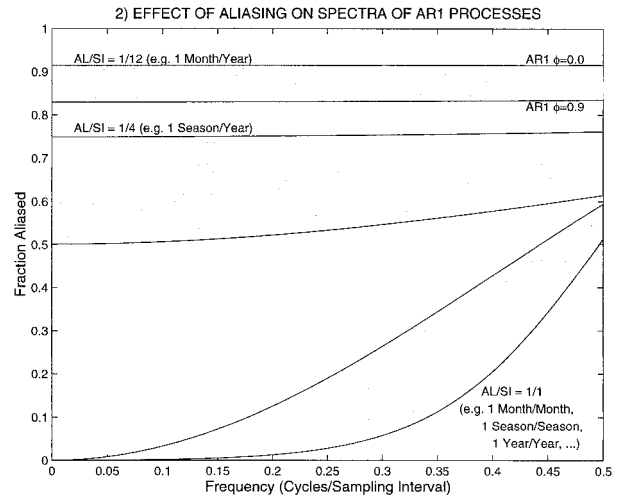


FIG. 2. Fraction of aliased variance (ordinate) found in the resolved frequency range of zero to 0.5 cycle per sampling interval (abscissa) for three ratios of (AL/SI) . Bottom band is for $AL/SI = 1/1$, e.g., 1 month (season, or year), sampled every month (season, or year). Middle band is for $AL/SI = 1/4$, e.g., 1 season sampled once per year. Top band indicates range determined by first-order autoregressive (AR1) models for daily data with lag-1-day correlation $\phi = 0.0$ (white noise; most aliasing) and 0.9 (highly persistent; least aliasing). In the white noise case it is assumed that there is no variance at frequencies greater than $0.5 \text{ cycle day}^{-1}$. An $AL = 30$ was used to determine the bands. For AL greater than 30, the bands are narrower because the highly persistent cases have more aliasing, and they fall within the bands presented here.

there will be large aliasing at all resolved frequencies, even those much lower than f_c .

Figure 2 is presented to summarize aliasing effects as a function of resolved frequency for AR1 processes. Two extreme AR1 models were considered for daily data: lag-1-day autocorrelation = 0.0 (white noise model), and = 0.9 (very persistent model). For both models, it was assumed that there was no variance at frequencies higher than $0.5 \text{ cycle day}^{-1}$. Theoretical spectra of the models for daily sampling were multiplied by the square of the frequency response of 91- (seasonal), 30- (monthly), and 365-day (yearly) averages to produce unaliased spectra of averaged data. Differing sampling intervals were then considered and the unaliased spectra of averaged data were folded about the corresponding folding frequency to determine the fraction of aliased variance for a given sampling interval. Results for the white noise model produced the largest fractions and they are nearly the same for 365-, 91-, and 30-day averaging. The persistent model produces the smallest fractions and, when plotted against cycles per sampling interval, the 30-day averaging has less than the 91-day averaging, which has less than the 365-day averaging.

To summarize, three bands are highlighted in Fig. 2. From bottom to top the bands represent an SI equal to the AL, or $AL/SI = 1/1$, $AL/SI = 1/4$, and $AL/SI = 1/12$, respectively. The top of each band is the worst

case (most aliasing) based on the white noise model, and the bottom of each band is the best case (least aliasing) and is based on the persistent model with 30-day averages. Results of the persistent model for 91- and 365-day averages lie within the band, as do those for the AR1 model of the 30-, 91-, and 365-day averages with smaller autocorrelations. Because the lag-1-day autocorrelation of 0.9 is larger than found in most meteorological data, it is likely that actual aliasing for variables behaving like the AR1 model will, on average, fall within the bands. One can estimate results for any AL/SI from the figure. Values for AL/SI = 1/2 (e.g., 6-month averages sampled once per year), for example, lie between the bottom and middle bands.

From Fig. 2, we can conclude that aliased variance for monthly (seasonal or yearly) averaged data from an AR1 process sampled monthly (seasonally or yearly) is less than 10% at frequencies less than 0.2 cycle month⁻¹ (cycles per season or cycles yr⁻¹) or periods longer than 5 months (seasons or years). For seasonally averaged data sampled yearly (or other AL/SI = 1/4) the aliasing exceeds 50% at all resolved frequencies, and for monthly data sampled yearly (or other AL/SI = 1/12) it exceeds 80%.

We argue that Fig. 2 provides a first-order estimate for the fraction of aliased variance as a function of frequency, averaging length, and sampling interval. However, it should be noted that the *fraction* of aliased variance can be somewhat less if true low-frequency variance exceeds that of an AR1 process. It can also be somewhat more if there are spectral peaks at relatively high frequencies that will alias into resolved frequencies.

3. Empirical example

Daily sea level pressure (SLP) data based on the Historical Weather Maps (U.S. Weather Bureau 1899–1939) and continuing daily maps for the period 1 January 1899–31 December 1998 were used. Digitized gridpoint values were originally obtained from the National Oceanic and Atmospheric Administration, the Massachusetts Institute of Technology, the U.S. Navy, the National Meteorological Center (now National Centers for Environmental Prediction), and the National Climatic Data Center. The entire record is available at the National Center for Atmospheric Research (available online at <http://www.scd.ucar.edu/dss/datasets/ds010.0.html>). Data from a grid point at 55°N, 30°W, in the center of the North Atlantic was selected. Even in the early years there were relatively frequent observations from ships of opportunity there. The exact nature of the data used to illustrate the effects of aliasing is not important. Any observed or modeled data that behave like a meteorological time series are adequate.

Data were first normalized with respect to their first and second moments:

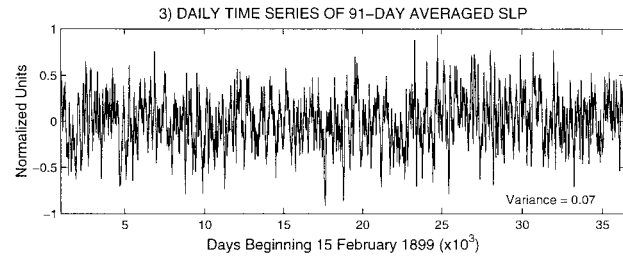


FIG. 3. The 91-day moving averages of normalized sea level pressure from 55°N, 30°W. Data are plotted for each day for 36 434 days beginning 15 Feb 1899. Variance of the series is indicated in the lower right.

$$\hat{p}(i, j) = \frac{p(i, j) - \bar{p}(j)}{\bar{\sigma}(j)}. \quad (5)$$

In Eq. (5) $p(i, j)$ is the SLP at the point with i the year index from 1899 to 1998; j is the day index from 1 to 365 (366 during leap years). There are 36 524 total days. This normalization eliminates the seasonal variation in the mean and variance of the daily data whose expected values become zero and one, respectively. The smooth seasonally varying average and standard deviation are, respectively, $\bar{p}(j)$ and $\bar{\sigma}(j)$. The smoothing for $\bar{p}(j)$ was accomplished by first computing the average over the 100-yr record for each day of the year, Fourier transforming the averages, retaining the first three annual waves, and doing an inverse transform. For $\bar{\sigma}(j)$, variances were calculated across the 100 years, logarithms of the variances were then Fourier transformed, three annual waves retained, and antilogs of the inverse transform were taken. The logarithmic transform is used to stabilize, or make more uniform, the variance of the estimated variance, whose untransformed value varies with the size of the variance itself.

Figure 3 presents a 91-day moving average of the data (55°N, 30°W). There are 36 434 (36 524 – 90) values plotted in Fig. 3. These 91-day averages, sampled every day are considered to be the unaliased data. That is, data sampled daily resolve variations of frequencies less than 0.5 day⁻¹, or periods longer than 2 days, and we assume variations on timescales shorter than 2 days do not alias significantly into our daily sampled data. The 91-day averaging reduces the variance from one for the daily, normalized data to 0.07 normalized units squared for the averaged data. The importance of the 0.07 variance is the fact that no matter how we sample data from Fig. 3 the expected variance is 0.07.

The unaliased spectrum was computed by the following:

- 1) Fourier transforming the daily sampled, 91-day moving averaged data of Fig. 3;
- 2) squaring the resulting coefficients at each frequency to give the periodogram; and
- 3) smoothing by a running average across frequency of 11 adjacent periodogram estimates.

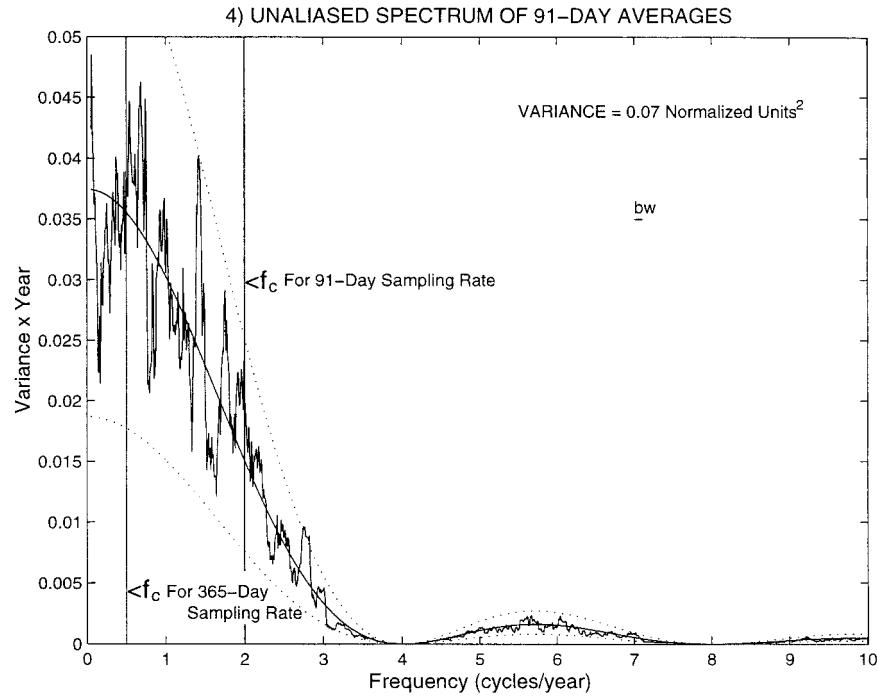


FIG. 4. Unaliased spectrum of the 91-day moving averaged data sampled daily that is shown in Fig. 3. Bandwidth (bw) is indicated and is $0.11 \text{ cycle yr}^{-1}$ [$(11/36\ 434) \times 365$]. The smooth line is an AR1 null hypothesis with a lag-1-day autocorrelation of 0.72. Dotted lines enclose the 95% significance interval about the null. Vertical lines mark the folding frequency (f_c) if sampling were once per year ($0.5 \text{ cycle yr}^{-1}$), or 4 per year (2 cycles yr^{-1}).

Figure 4 is the resulting unaliased spectrum of the 91-day moving averages. The frequency response [$H(f)$] of the 91-day averaging has zeroes at 4 cycles yr^{-1} ($1/91 \text{ cycle day}^{-1}$), 8 cycles yr^{-1} ($2/91 \text{ cycles day}^{-1}$), . . . , resulting in corresponding zeroes in the spectrum at those frequencies. Although the spectrum of these daily sampled, 91-day averaged data is plotted only to $10 \text{ cycles yr}^{-1}$ in Fig. 4, $f_c = 0.5 \text{ cycle day}^{-1}$ or about $182 \text{ cycles yr}^{-1}$.

The spectrum is scaled so that areas on Fig. 4 give the variance. The smooth line represents an AR1 null hypothesis with the lag-1-day autocorrelation of the data (0.72). The dotted lines enclose the 95% significance interval about the null, assuming 22 degrees of freedom (DOF). The area under the spectrum can be approximated by the area of the triangle $0.036 (\text{Variance} \times \text{Year}) \times 3.8 (\text{cycles yr}^{-1})/2 = 0.07$, close to the variance of data in Fig. 3. Just as in the time domain, the variance and thus the area under our spectral estimate must always equal about 0.07 no matter where the folding frequency is. If sampling were once per year, as with a time series of [Jun-Jul-Aug (JJA)] or [Dec-Jan-Feb (DJF)] averages, the resulting aliased spectral estimates will extend from 0.0 to $0.5 \text{ cycle yr}^{-1}$. They will average about 0.14 normalized units squared, well above the unaliased spectral values of Fig. 4, in order that their integral will continue to equal 0.07. If data were sampled

each season (i.e., 4 times per year), then the folding frequency is 2 cycles yr^{-1} and the average aliased spectrum must be near 0.035 to reflect a total variance of 0.07 and the aliasing is considerably less.

To illustrate, the unaliased spectrum of Fig. 4 is compared with the spectra of the traditional seasonal averages sampled once per year and once per season in Fig. 5. For the case of sampling every season (Fig. 5b) aliasing is most serious near f_c (2 cycles yr^{-1}), but it is a problem at all resolved frequencies for sampling once per year (Fig. 5a).

It is interesting to note that Fig. 5a reveals DJF averages have twice as much variance (0.10 normalized units squared) as do JJA averages (0.05 normalized units squared). While the normalization of Eq. (5) makes the variance of daily data near 1 year-round, it does not affect the seasonal variation of lagged autocorrelations or the persistence. The DJF daily data are more persistent than JJA data with characteristic times between independent estimates of near 7 and 5 days, respectively [e.g., see Madden 1979, his Eq. (2)], resulting in a slower decrease in variance with averaging length in DJF than in JJA. This seasonal difference in variance reflects nonstationarity (or cyclostationarity), an added complication in interpreting spectra, which we neglect in considering effects of aliasing.

Figure 6 is presented to further quantify the amount

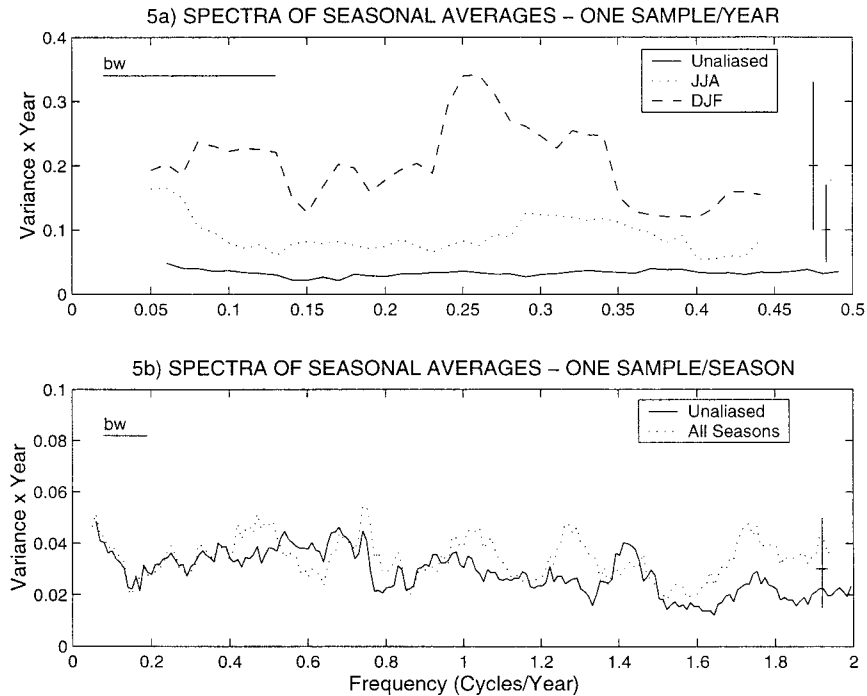


FIG. 5. (a) Solid line is the low-frequency part of the unaliased spectrum from Fig. 4. Dashed (dotted) line is the aliased spectrum computed from a time series of DJF (JJA) seasonal means. The folding frequency $f_c = 0.5$ cycle yr^{-1} . To give an idea of likely sampling variability in these aliased spectra, we find for 22 DOF that the 95% limits for a null hypothesis of 0.2 (0.1) range from 0.10 (0.05) to 0.33 (0.17) (indicated on the right). (b) Solid line is the low-frequency part of the unaliased spectrum from Fig. 4. Dotted line is an aliased spectrum computed from a time series of all nonoverlapping seasonal means (DJF, MAM, JJA, . . .). The folding frequency $f_c = 2$ cycles yr^{-1} . The 95% limits for a null hypothesis of 0.03 range from 0.015 to 0.05 (indicated on the right) for 22 DOF.

of aliased variance in the empirical results, and to relate it to the theoretical results of Fig. 2. The fractions of aliased variance for the pressure data for time series of 100 DJF and JJA averages (1 season per year, $AL/SI = 1/4$) are plotted in Fig. 6b. They are given by the difference between the solid line of Fig. 5a (estimated unaliased spectrum) and the dashed and dotted lines, respectively, divided by the dashed or dotted line (aliased spectra of the sampled data). The shaded band is taken from the theoretical band of Fig. 2.

The dotted line in Fig. 6c has the similar result for the time series of all 400 seasons (1 sample per season, $AL/SI = 1/1$). It is derived from the two spectra of Fig. 5b. Negative fractions occur when a spectral value of the particular sample of seasonal averages is smaller than the estimate unaliased spectrum. Also plotted in Fig. 6c are fraction aliased variance for a sample of 1200 monthly averages sampled every month (solid line) and that for 100 annual averages sampled every year (dashed line). Similarly, Fig. 6a shows the fractions of aliased variance for a time series of 100 Januarys and Julys (1 sample per year, $AL/SI = 1/12$). The spectra of these latter four cases are not shown. The shaded bands are from Fig. 2.

4. A partial remedy

To minimize aliasing we need to minimize variance at $f > f_c$ before sampling. Many of our longest time series are in the form of monthly averages. In that case, a convenient way to minimize variance at all $f > f_c$ is to first compute a new time series of running 3-month averages. The spectrum of the resulting 3-month averaged or seasonal data sampled monthly takes the character of that of Fig. 4 with a zero (nearly) at 4 cycles yr^{-1} and $f_c = 6$ cycles yr^{-1} (0.5 cycle month^{-1}). Figure 4 shows that only a very small amount of variance at $f > 6$ cycles yr^{-1} survives the 3-month averaging. Figure 7 shows the estimated unaliased spectrum from Fig. 4 and the spectrum of 3-month running means sampled monthly. The y axis is logarithmic so that the 95% confidence limits can be readily shown. At frequencies less than about 3 cycles yr^{-1} differences between the two spectra are small relative to these limits implying very little aliasing. The two spectra begin to diverge at $f > 3$ cycles yr^{-1} , because the unaliased spectrum is zero at $1/91$ cycle day^{-1} (close to 4 cycles yr^{-1}) and the monthly sampled spectrum is not since 3-month running averages range in length from 89 to 92 days.

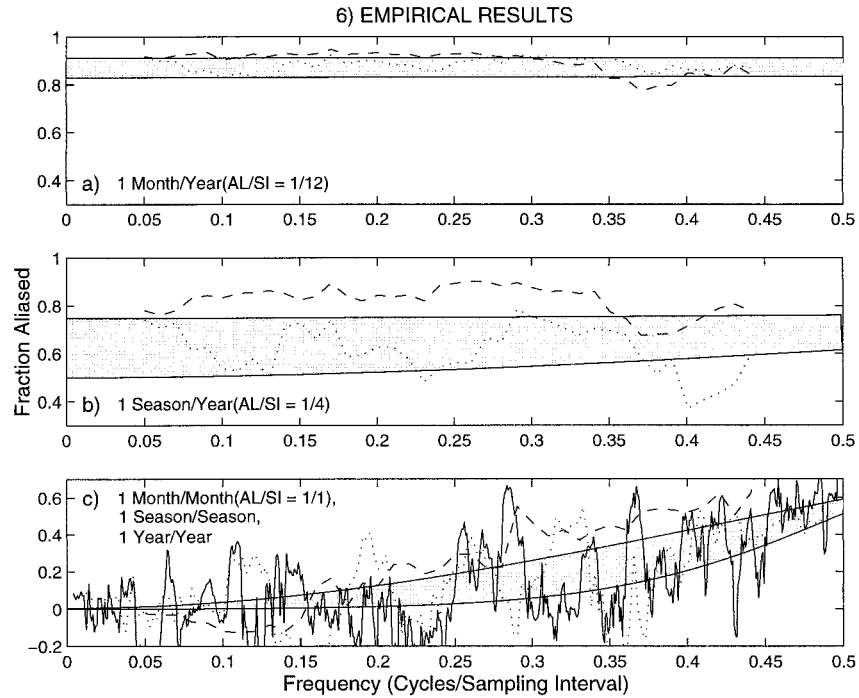


FIG. 6. (a) Shaded band is from $AL/SI = 1/12$ of Fig. 2. Dashed (dotted) line is the fraction of aliased variance in the spectrum of a time series of Jan (Jul) means of $55^{\circ}N, 30^{\circ}W$ SLP. (b) Shaded band is from $AL/SI = 1/4$ of Fig. 2. Dashed (dotted) line is the fraction of aliased variance in the spectrum of a time series of DJF (JJA) means of $55^{\circ}N, 30^{\circ}W$ SLP. (c) Shaded band is from $AL/SI = 1/1$ of Fig. 2. Solid, dotted, and dashed lines are the fractions of aliased variance in the spectra of monthly, seasonal, and yearly averaged data sampled monthly, seasonally, and yearly, respectively.

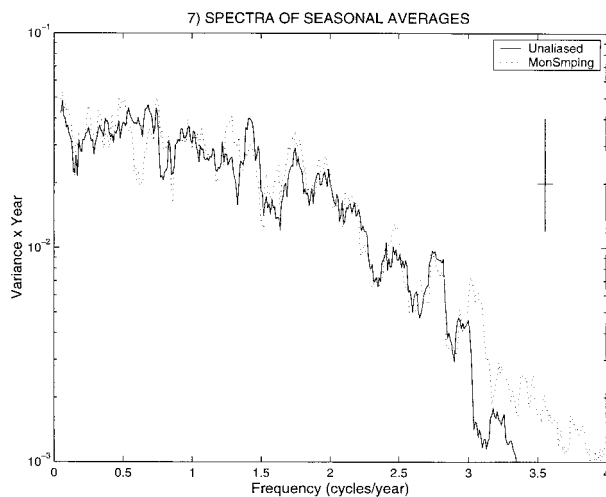


FIG. 7. Unaliased spectrum of the 91-day moving averaged data sampled daily from Fig. 4 (solid), and the spectrum of 3-month moving averages sampled monthly (dotted). Spectral values are plotted on a logarithmic scale versus linear frequency only out to 4 cycles yr^{-1} . Although not shown, the unaliased spectrum resolves variance to 182 cycles yr^{-1} and the monthly sampled spectrum to 6 cycles yr^{-1} . The 95% limits for 22 DOF and the bw of the analyses are indicated by the cross.

The above remedy requires sampling every month and therefore does not allow isolating the time series of a single season or month. Because physical processes can change seasonally, it may be desirable to consider data from a single season. This presents a difficulty that will have to be treated case by case.

5. Summary

Aliasing of relatively high-frequency variations to low frequencies can make determination of slow changes in meteorological time series difficult to quantify. The theoretical and empirical examples presented here suggest that for seasonal or monthly averaged data sampled once per year the fraction of aliased variance exceeds 50% and 80%, respectively, at all frequencies. We propose Fig. 2 as a first-order estimate of aliasing effects for varying averaging lengths and sampling intervals. For any case, the exact nature of aliasing will depend on the true unaliased spectrum of the given time series. In particular, if the low-frequency part of that spectrum has relatively more variance than the AR1 model, then the fraction of aliased variance will be less than Fig. 2 suggests.

Acknowledgments. J. Meehl brought the problem to our attention. D. Shea provided the pressure data, and

E. Rothney typed several versions of the manuscript. An anonymous reviewer's comments helped to improve the manuscript.

REFERENCES

- Bendat, J. S., and A. G. Piersol, 1971: *Random Data: Analysis and Measurement Procedures*. Wiley-Interscience, 407 pp.
- Blackman, R. B., and J. W. Tukey, 1958: *The Measurement of the Power Spectra*. Dover, 190 pp.
- Gray, B. M., and R. A. Madden, 1986: Aliasing in time-averaged tropical pressure data. *Mon. Wea. Rev.*, **114**, 1618–1622.
- Madden, R. A., 1979: A simple approximation for the variance of meteorological time averages. *J. Appl. Meteor.*, **18**, 703–706.
- U.S. Weather Bureau, 1899–1939: *Daily Series Synoptic Weather Maps, Part I. Northern Hemisphere Sea Level and 500 Millibar Charts*. U.S. Dept. of Commerce.
- Wunsch, C., 2000: On sharp spectral lines in the climate record and the millennial peak. *Paleoceanography*, **15**, 417–424.