

Research Article

What Signals Are Removed and Retained by Using an Anomaly Field in Climatic Research?

Chunzai Wang,¹ Zhenya Song,² Fangli Qiao,² and Shenfu Dong³

¹NOAA Atlantic Oceanographic and Meteorological Laboratory, Physical Oceanography Division, 4301 Rickenbacker Causeway, Miami, FL 33149, USA

²The First Institute of Oceanography, State Oceanic Administration, Qingdao 266061, China

³Cooperative Institute for Marine and Atmospheric Studies, University of Miami, Miami, FL 33149, USA

Correspondence should be addressed to Chunzai Wang, chunzai.wang@noaa.gov

Received 22 July 2009; Accepted 22 October 2009

Recommended by Xiao-Hai Yan

Signals in data are often detected by analyzing anomaly field that is calculated by subtracting the mean value over a time length from the data. Here we demonstrate that the anomaly calculation removes signals which satisfy that the ratio between the time length of the mean T and signals' period L is an integer (i.e., $T/L = n$ where n is an integer) and retains other signals if the ratio is not an integer. In climatic and other studies, the time length of the mean is usually chosen as $T = 12$ months from January to December and the mean is called the monthly climatology. Anomaly is calculated by subtracting the monthly climatology from data. This anomaly calculation thus removes the climatic signals with the periods of 12, 6, 4, 3, 2.4, and 2 months which correspond to $(12 \text{ months})/n$ with $n = 1, 2, 3, 4, 5,$ and $6,$ respectively, whereas it retains other signals such as those with the periods of 11, 10, 9, 8, 7, and 5 months. This paper suggests that one should be cautious when an anomaly field is used in research. The conventional notion is that the monthly anomaly calculation removes the annual cycle. However, here we show that the anomaly calculation removes all signals as long as the time length of the mean is an integer multiple of signals' period.

Copyright © 2009 Chunzai Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. Introduction

An important step in scientific research is to analyze data that are either measured in nature/laboratory or obtained by running numerical models. In analyzing data, an anomaly calculation (i.e., a removal of the mean value) is often performed for the purpose of highlighting and emphasizing signals. For example, in climatic and other studies a monthly climatology is usually calculated based on the full record data. Anomaly is then calculated by subtracting the monthly climatology from data. Here we caution that an anomaly calculation will remove some signals and retain others depending on the relationship between the time length of the mean and signals' period. In this paper, we first mathematically and theoretically derive how/why signals are removed by an anomaly calculation and what signals are

removed and retained in an anomaly field. Then, idealized and observed data are used to demonstrate and confirm the mathematical derivation.

2. Mathematical Derivation

Assume a time series x_i ($i = 1, 2, \dots, N$) over the time interval of $[1, N]$. The mean or climatological mean of x_i over the time length T is often defined as

$$\bar{x}_m = \frac{1}{M} \sum_{j=1}^M x_{(j-1)*T+m} \quad (m = 1, 2, \dots, T), \quad (1)$$

where $M \equiv [N/T]$ denotes integer part of N/T (note that N is larger than T). The anomaly of x_i is calculated by subtracting

its mean value \bar{x}_m from x_i (i.e., $x_i - \bar{x}_m$). As an example of $T = 12$ months from January to December, \bar{x}_m ($m = 1, 2, \dots, 12$) is the so-called monthly climatology (or the seasonal cycle) and $x'_i \equiv x_i - \bar{x}_m$ is monthly anomaly field used for detecting climatic signals. These calculations are widely used in studies of climate, oceanography, meteorology, and other disciplines.

Mathematically, the time series x_i ($i = 1, 2, \dots, N$) can be expressed as the compact Fourier series expansions of the following form (e.g., [1]):

$$x_t = a_0 + \sum_k A_k \sin\left(\frac{2\pi kt}{N} + \theta_k\right) \quad (t = 1, 2, \dots, N), \quad (2)$$

where a_0 is constant, A_k is amplitude, θ_k is phase, k is wave number, and its corresponding period is $T_k = N/k$. That is, x_i is represented by a sum of sine functions (waves) of different amplitudes and phases.

As an example, we consider the harmonic component with period of L in (2), which has the following form:

$$x_t^L = A_{N/L} \sin\left[\frac{2\pi(N/L)t}{N} + \theta_{N/L}\right] = A_{N/L} \sin\left(\frac{2\pi t}{L} + \theta_{N/L}\right). \quad (3)$$

Based on (1) and (3), the mean value for this harmonic component over the time length T (where $L \leq T$) is

$$\begin{aligned} \bar{x}_m^L &= \frac{1}{M} \sum_{j=1}^M x_{(j-1)*T+m} \\ &= \frac{1}{M} \sum_{j=1}^M A_{N/L} \sin\left\{\frac{2\pi[(j-1)*T+m]}{L} + \theta_{N/L}\right\} \\ &= \frac{1}{M} \sum_{j=1}^M A_{N/L} \sin\left[\frac{2\pi(j-1)*T}{L} + \frac{2\pi m}{L} + \theta_{N/L}\right]. \end{aligned} \quad (4)$$

If $T/L \equiv n$ is an integer, then (4) becomes

$$\begin{aligned} \bar{x}_m^L &= \frac{1}{M} \sum_{j=1}^M A_{N/L} \sin\left[2\pi(j-1)*n + \frac{2\pi m}{L} + \theta_{N/L}\right] \\ &= A_{N/L} \sin\left(\frac{2\pi m}{L} + \theta_{N/L}\right), \end{aligned} \quad (5)$$

where $m = 1, 2, \dots, T$. Equation (5) is the same as (3), stating that the mean value \bar{x}_m^L contains the original harmonic with period of $L = T/n$ where n is an integer. Thus, the difference between x_t^L and \bar{x}_m^L removes the harmonic component of $L = T/n$. As shown in (2), x_i is represented by a sum of all harmonic components; so the anomaly calculation of $x_i - \bar{x}_m$ removes signals which satisfy that T/L is an integer and retains other signals if T/L is not an integer.

In climatic study, an anomaly field is often calculated by subtracting the monthly climatological mean (i.e., $T = 12$ months from January to December) from the total monthly data. According to the Nyquist folding frequency principle (e.g., [2]), the highest frequency that can be resolved with monthly data is $1/[2 \times (1 \text{ month})] = 0.5 \text{ month}^{-1}$. Thus, the

condition of $L = (12 \text{ months})/n$ tells us that the anomaly calculation removes the signals with the periods of 12, 6, 4, 3, 2.4, and 2 months which correspond to $n = 1, 2, 3, 4, 5$, and 6, respectively (note that $n = 7, 8, \dots, 12$ are associated with the frequencies that cannot be resolved by monthly data based on the Nyquist principle). On the other hand, the anomaly field retains other signals such as those with the periods of 11, 10, 9, 8, 7, and 5 months.

3. Application to Idealized Data

In Section 2, we theoretically derive and approve that the anomaly calculation by subtracting the monthly climatology from a time series removes the climatic signals with the periods of 12, 6, 4, 3, 2.4, and 2 months, whereas it retains the signals such as those with the periods of 11, 10, 9, 8, 7, and 5 months. In this section, we would like to test this result by applying it to idealized time series. We first create a time series by summing a series of sine functions of 2, 4, 6, and 12 months periods (total time record of the created time series $N = 240$ months). These individual sine functions and their summation are shown in Figures 1(a)–1(e). Note that since the result does not depend on amplitude, all amplitudes of the sine functions are chosen as one. With the 240-month time series (Figure 1(e)), we first calculate the climatological mean \bar{x}_m ($m = 1, 2, \dots, 12$). We then compute anomaly field by subtracting \bar{x}_m from the time series. The resultant anomaly is zero, as shown in Figure 1(f). This indicates that the anomaly calculation procedure removes all oscillatory signals of 2, 4, 6, and 12 months from the time series of Figure 1(e). The anomaly calculation also can remove the sine signals with the periods of 2.4 and 3 months (not shown).

We next examine the case in which the anomaly calculation will not remove an oscillatory signal. In this case, we create a time series by summing a series of sine functions of 4, 5, 6, and 12 months periods, as shown in Figures 2(a)–2(e). Again, we calculate the climatological mean from the total summed time series of Figure 2(e) and then subtract the climatological mean from the time series of Figure 2(e). The resultant anomaly field is shown in Figure 2(f). Comparison of Figures 2(b) and 2(f) shows that two time series are identical. This demonstrates that the anomaly calculation retains the oscillatory signal with the period of 5 months, whereas it removes the signals with the periods of 4, 6, and 12 months.

We also apply the above calculations to the signals with the periods of 7, 8, 9, 10, and 11 months (not shown). The anomaly calculation will not remove these oscillatory signals either. However, we would like to notice an error (in spite of smallness) for computing climatological mean for the case with the period of 7 months if the total time record of the time series is still selected as $N = 240$ months. This should be associated with an aliasing error in traditional textbooks (e.g., [3]). To avoid this computational error in the climatological mean, the total time record N should be a multiple of both the climatological month of 12 and the period of 7 (e.g., $N = 252$). This constraint ensures that

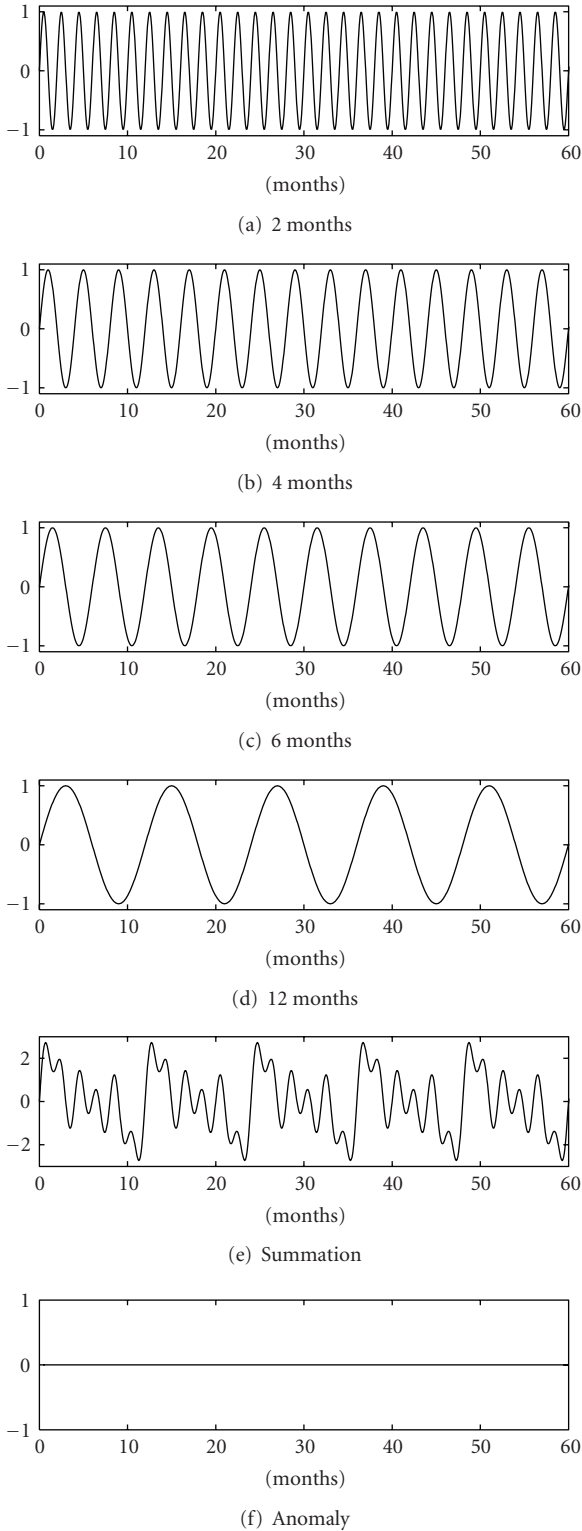


FIGURE 1: The time series of sine functions, their summation, and calculated anomaly. Shown are the sine functions with periods of (a) 2 months, (b) 4 months, (c) 6 months, (d) 12 months, (e) summation of the four time series (a) to (d), and (f) anomaly calculated from (e). The total time record of these time series is $N = 240$ months (the plots are to 60 months). The anomaly calculation removes all signals of 2, 4, 6, and 12 months periods.

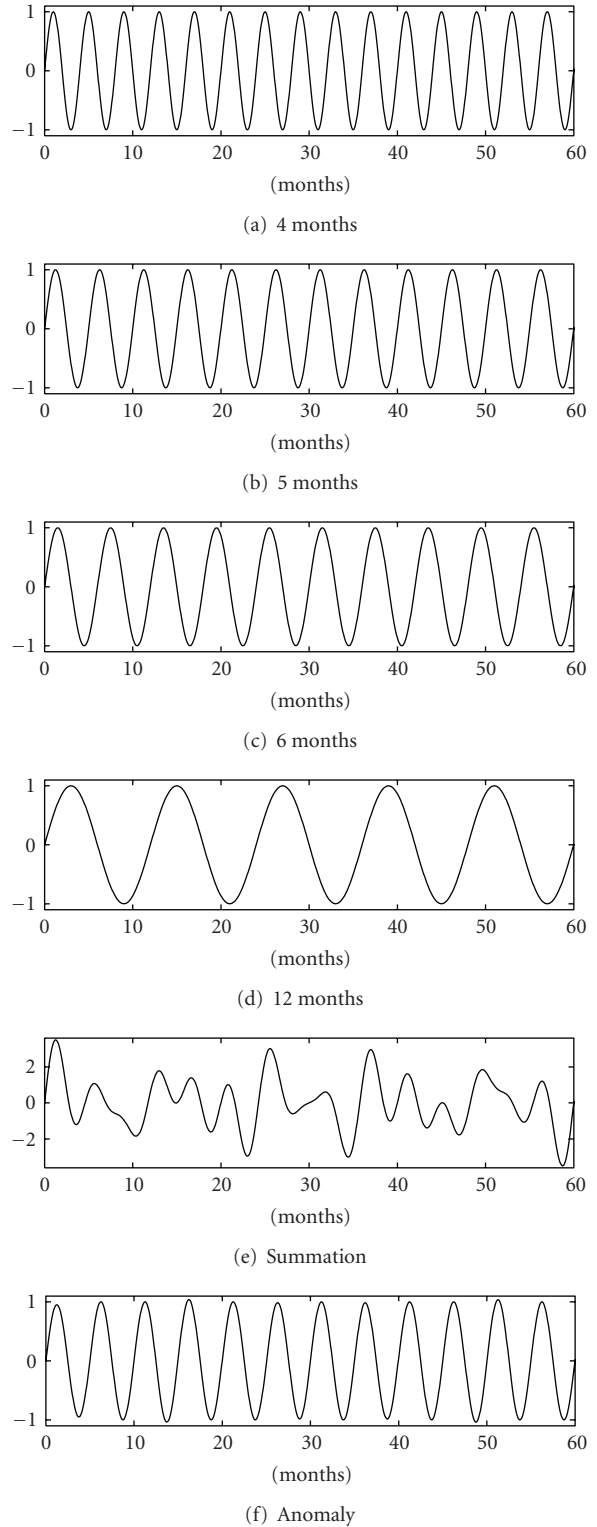


FIGURE 2: The time series of sine functions, their summation, and calculated anomaly. Shown are the sine functions with periods of (a) 4 months, (b) 5 months, (c) 6 months, (d) 12 months, (e) summation of the four time series (a) to (d), and (f) anomaly calculated from (e). The total time record of these time series is $N = 240$ months (the plots are to 60 months). The anomaly calculation removes the signals of 4, 6, and 12 months periods, whereas it retains the signal of 5 months period.

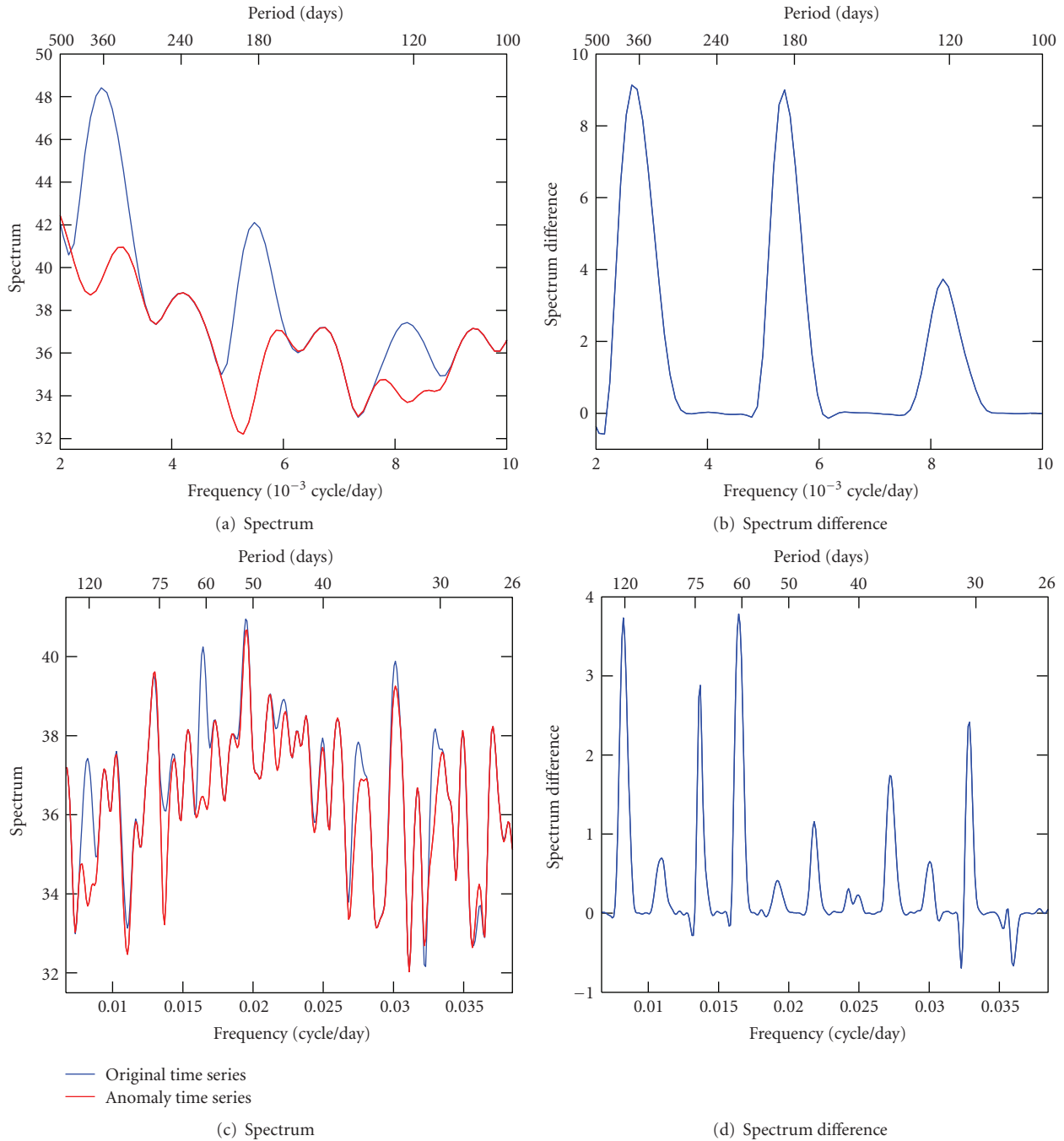


FIGURE 3: Logarithm of power spectrum for the detrended daily OLR time series. Shown are at relatively low frequency from 500 to 100 days of (a) the spectrums for the original (blue) and anomaly (red) OLR time series and (b) their spectrum difference (original minus anomaly) and at relatively high frequency from 150 to 26 days of (c) the spectrums for the original (blue) and anomaly (red) OLR time series and (d) their spectrum difference (original minus anomaly). The daily OLR time series is calculated in the equatorial western Pacific (120°E–150°E, 5°S–5°N).

each climatological month (out of 12) has an equal data point and covers at least a full cycle of the data (the data are from the sine function with the period of 7 months) when the monthly climatology is calculated. Similar situation occurs when the climatological mean is calculated for the cases with the periods of 9 and 11 months.

4. Application to Observed Data

In this section, we use observational data to demonstrate climatic signals observed in nature that are removed and retained by performing the anomaly calculation after the climatological mean is subtracted. Since the equatorial western

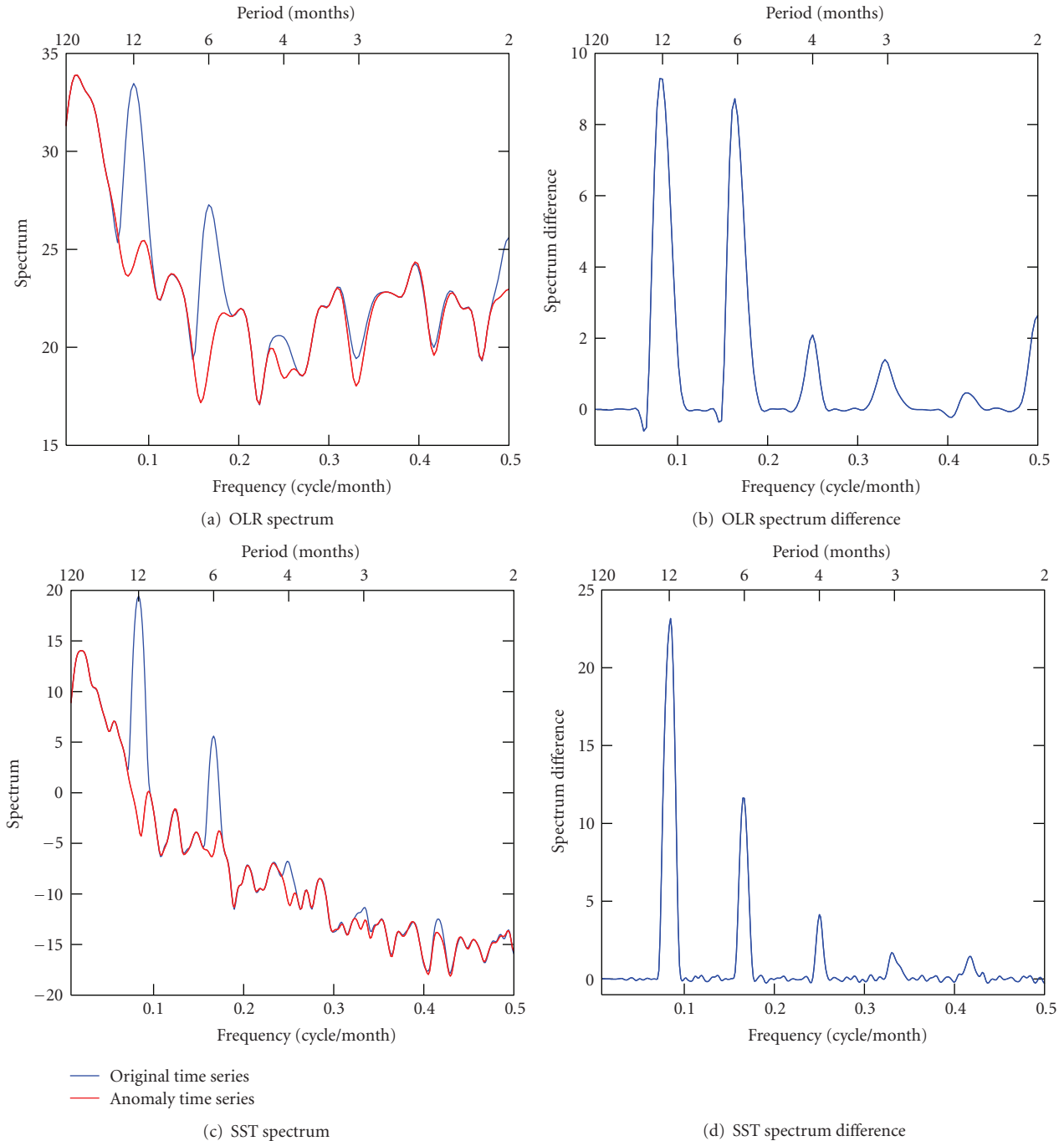


FIGURE 4: Logarithm of power spectrum for detrended monthly data. Shown are (a) the spectrums for the original (blue) and anomaly (red) OLR time series and (b) their spectrum difference (original minus anomaly) and (c) the spectrums for the original (blue) and anomaly (red) SST time series and (d) their spectrum difference (original minus anomaly). The monthly OLR time series is calculated in the equatorial western Pacific (120° E–150° E, 5° S–5° N) and the monthly SST time series is in the Nino3 region (150° W–90° W, 5° S–5° N).

Pacific displays many different climatic signals varying from intraseasonal to interannual timescales, observations in the equatorial western Pacific thus seem to be good candidates for testing signals removed and retained. We obtain the outgoing longwave radiation (OLR) data from the Physical Sciences Division of the National Oceanic and Atmospheric

Administration (NOAA)/Earth System Research Laboratory [4]. Two versions of OLR data are obtained: daily and monthly, both of which are from 1980 to 2007. Based on these daily and monthly OLR data, we create two time series in the equatorial western Pacific region of 120° E–150° E, 5° S–5° N. For the daily OLR time series, we delete the date of

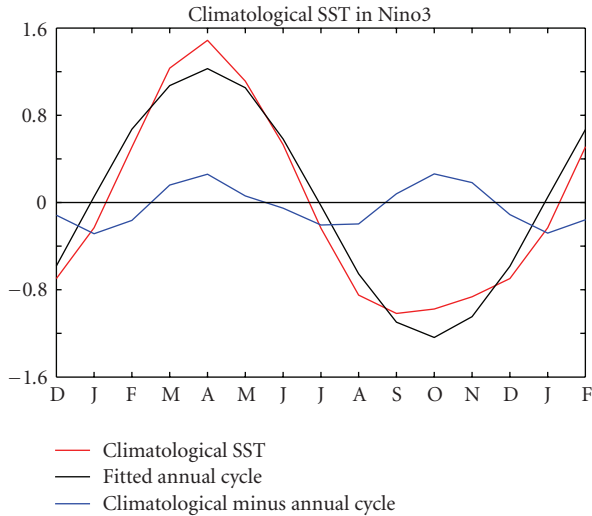


FIGURE 5: Climatological SST (minus 25.8°C) in the Nino3 region (150°W–90°W, 5°S–5°N). The climatological mean SST (red) is calculated from the detrended Nino3 SST time series from 1950 to 2007. The climatological mean SST is fitted to an annual cycle with the form of $A \sin(2\pi t/T_A + \theta)$ where $T_A = 1$ year (black). A semiannual cycle is shown in the difference between the climatological mean SST and the fitted annual cycle (blue).

February 29 and make sure that each year has 365 days, and thus the total time record is $N = 10220$ days. The total time record of the monthly OLR time series is $N = 336$ months. To avoid a distortion of the spectrum at lower frequency band [1], two OLR time series are detrended prior to spectral analysis; that is, a linear trend removal is applied to two OLR time series.

With the detrended original daily OLR time series, we first calculate the mean value \bar{x}_m ($m = 1, 2, \dots, 365$ days). The daily OLR anomaly is then computed by subtracting \bar{x}_m from the original daily OLR time series. We next perform the power spectral analyses to both the original and anomaly OLR time series. Their corresponding spectrums are shown in Figure 3. For demonstration purpose we separate the spectrums into two parts, with the relatively low-frequency signals of periods from 500 to 100 days shown in top two panels and the relatively high-frequency signals of periods from 150 to 26 days shown in bottom two panels. The spectrum differences between the original and anomaly OLR time series occur mainly at frequency bands peaked around 365, 182.5, 121.67, 91.25, 73, 60.83, 52.14, 45.63, 40.56, 36.5, 33.18, and 30.42 days. These signals removed are consistent with the above theoretical result that satisfies the relationship of $L = (365 \text{ days})/n$. These twelve removed signals correspond to $n = 1, 2, \dots, 12$, respectively. At other frequency bands, two spectrum curves are almost the same. It is well known that climatic variability in the equatorial western Pacific includes intraseasonal variability associated with the Madden-Julian Oscillation, the semiannual cycle, and the annual cycle. Our spectrum analyses show that these climate variations are removed or much reduced if anomaly field is used to study

them. Thus, it is not reliable to use anomaly field to study climate variations in the equatorial western Pacific.

We also perform a similar calculation on the monthly OLR time series. For this case, the mean value is the so-called monthly climatology \bar{x}_m ($m = 1, 2, \dots, 12$ months; i.e., from January to December) that is widely used in climatic research. The power spectrums calculated from the original and anomaly monthly OLR time series and their spectrum difference are shown in Figures 4(a) and 4(b). The removed or reduced climatic signals occur mainly around the periods of 12, 6, 4, 3, 2.4, and 2 months, whereas the power spectrums are relatively unchanged at other frequency bands. The removed or reduced climatic signals are exactly what our theory predicts, which satisfy that $L = (12 \text{ months})/n$ with $n = 1, 2, 3, 4, 5$, and 6, respectively. In particular, the large spectrum peaks at the annual and semiannual cycles disappear in the anomaly OLR time series. Figures 4(a) and 4(b) also show that two spectrum curves at lower frequency (lower than annual timescale) are almost identical, indicating that the anomaly calculation does not affect climate variability for frequencies of interannual timescales.

To further confirm the result, we also examine climate variability in the equatorial eastern Pacific. We use the NOAA monthly sea surface temperature (SST) data from 1950 to 2007 [5]. With this SST data, we produce a detrended original SST time series in the Nino3 region of 150°W–90°W, 5°S–5°N. Based on this monthly SST time series from 1950 to 2007, we perform a similar calculation to that for the monthly OLR time series. The power spectrums of the original and anomaly SST time series and their spectrum difference are shown in Figures 4(c) and 4(d). For this case, the removed or reduced climatic signals are around 12, 6, 4, 3, and 2.4 months. These removed or reduced signals satisfy that $L = (12 \text{ months})/n$ with $n = 1, 2, 3, 4$, and 5, respectively. Since the SST time series does not have a peak signal at 2 months, it does not apply to $n = 6$. Again, the interannual signal of El Niño-Southern Oscillation (ENSO) is unchanged.

The spectrum of the SST in the Nino3 region shows a peak at the semiannual timescale (Figure 4(c)). This seems to be a little surprising since the conventional notion is that the equatorial eastern Pacific has only a strong annual cycle because of a strong equatorial ocean upwelling there. To examine how and where the semiannual cycle comes from, we plot the climatological SST in the Nino3 region (Figure 5). We fit the climatological SST to an annual cycle with the form of $A \sin(2\pi t/T_A + \theta)$ where $T_A = 1$ year and then we subtract the fitted SST annual cycle from the climatological SST. The residual shows a semiannual cycle, with two peaks in April and October and two valleys in January and July, respectively. A closed examination of Figure 5 shows that asymmetry of the climatological SST corresponds to or results in the semiannual SST cycle. Comparison of the climatological SST and fitted SST annual cycle shows that the warm season in the spring is warmer and shorter (i.e., the climatological SST in the spring is warmer and lasts shorter than the fitted SST annual cycle) and the cold season in the fall is warmer and longer (i.e., the climatological SST in the fall is warmer and lasts longer than

the fitted SST annual cycle). Physically, this semiannual cycle is consistent with the fact that sun passes the equator twice every year. It is the solar radiation that results in the SST semiannual cycle in the equatorial eastern Pacific (although its amplitude is much smaller than that of the annual cycle).

5. Summary

Given a time series that includes signals with various timescales, how to analyze and detect these signals is important in research. One of analysis methods is to focus on anomaly field that is calculated by subtracting the mean value from the time series. This paper mathematically and theoretically demonstrates that the mean value over the time length T contains the signal with the period of $L = T/n$ where n is an integer. Since anomaly is calculated as the difference between the time series and the mean, the anomaly calculation removes the signal with the period of L from the time series. In other words, the anomaly calculation removes all signals which satisfy that T/L is an integer and retains signals if T/L is not an integer.

In studies of climate, oceanography, meteorology, and other disciplines, an anomaly calculation is widely used. The most common case is to first calculate the monthly or daily climatological mean and then to subtract the climatological mean from data. As an example of $T = 12$ months from January to December, the climatological mean is the so-called monthly climatology. If the monthly climatology is subtracted from data, the resultant anomaly field does not include the signals with the periods of 12, 6, 4, 3, 2.4, and 2 months ((12 months)/ n with $n = 1, 2, 3, 4, 5$, and 6, resp.), whereas it retains other signals such as with the periods of 11, 10, 9, 8, 7, and 5 months. This paper suggests that one should be cautious when an anomaly field is used in research. The conventional notion is that the anomaly calculation removes the annual cycle. However, here we show that the anomaly calculation removes all signals as long as the time length of the mean is an integer multiple of signals' period. For example, a monthly anomaly field also removes the semiannual cycle in addition to the annual cycle. Another example is that intraseasonal variability in the equatorial western Pacific is removed or much reduced if one uses an anomaly field. Thus, it is not a good idea to use an anomaly field for studying intraseasonal variability.

Acknowledgments

The work was initiated when C. Wang visited the First Institute of Oceanography in Qingdao, China, during September of 2008. This work was supported by the National Oceanic and Atmospheric Administration (NOAA) Climate Program Office, the base funding of NOAA Atlantic Oceanographic and Meteorological Laboratory (AOML), and National Natural Science Foundation of China (no. 40476017). The findings and conclusions in this report are those of the author(s) and do not necessarily represent the views of the funding agency.

References

- [1] W. J. Emery and R. E. Thomson, *Data Analysis Methods on Physical Oceanography*, Elsevier, Amsterdam, The Netherlands, 2004.
- [2] H. von Storch and F. W. Zwiers, *Statistical Analysis in Climate Research*, Cambridge University Press, Cambridge, UK, 2003.
- [3] J. S. Bendat and A. G. Piersol, *Random Data: Analysis and Measurement Procedures*, Wiley-Interscience, New York, NY, USA, 2nd edition, 1986.
- [4] B. Liebmann and C. A. Smith, "Description of a complete (interpolated) outgoing longwave radiation dataset," *Bulletin of the American Meteorological Society*, vol. 77, pp. 1275–1277, 1996.
- [5] T. M. Smith and R. W. Reynolds, "Improved extended reconstruction of SST (1854–1997)," *Journal of Climate*, vol. 17, pp. 2466–2477, 2004.