

Analysis of Diurnal and Seasonal Cycles and Trends in Climatic Records with Arbitrary Observation Times

Konstantin Y. Vinnikov

*Department of Meteorology, University of Maryland, College Park, MD
E-mail: kostya@atmos.umd.edu*

Alan Robock

*Department of Environmental Sciences, Rutgers University, New Brunswick, NJ
E-mail: robock@envsci.rutgers.edu*

Norman C. Grody

*NESDIS/NOAA, Camp Springs, MD
E-mail: Norman.Grody@noaa.gov*

Alan Basist

*Commodity Hedgers, Inc., Asheville, NC
E-mail: abasist@charter.net*

Submitted to *Geophysical Research Letters*

November, 2003

Revised January, 2004

Corresponding author:

Alan Robock
Department of Environmental Sciences
Rutgers University
14 College Farm Road
New Brunswick, NJ 08901-8551
*Phone: 732-932-9478
Fax: 732-932-8644
E-mail: robock@envsci.rutgers.edu*

Abstract

Many long-term records of climate variables have missing data or have had changes in their times of observation. Here we present a technique to analyze such inhomogeneous records. We assume that the underlying climatic processes are nonstationary, where the observations contain a long-term trend superimposed on periodic shorter time seasonal and diurnal cycles. The seasonal and diurnal variations are approximated using a limited number of Fourier harmonics, while the trend is represented by a monotonic function of time whose amplitude can also vary seasonally and diurnally. A least squares method is used to estimate the unknown Fourier coefficients. As an example of the technique, we present an analysis of multi-decadal hourly observations of surface air temperature obtained from several meteorological stations within the United States.

1. Introduction

The amplitude of seasonal and diurnal variations of meteorological variables is generally much larger than weather-related fluctuations. As such, the instruments and analytical techniques used for climatic studies are often different than those used for studying short-term localized weather events. From the very beginning, meteorologists used monthly averages to approximate the seasonal cycle, maximum/minimum temperatures to describe the diurnal cycle of temperature, and other simplifications. Techniques were developed to avoid excessive computations and to fix the calendar-related problems of leap years and unequal length of months.

In addition to diurnal and seasonal variations, much smaller amplitude climatic trends are an important component of meteorological variables, and they can also display diurnal and seasonal cycles. Modern climatologists mainly consider the long-term trend to be linear in time rather than a higher order polynomial. For relatively short time intervals, trend estimates are always contaminated by trend-like and episodic (El Niño events and volcanic eruptions) components of natural climate variability. This is why such trend analysis can only be used as a diagnostic tool and not for extrapolation of climatic data into the future. To address this problem and filter out short-time random variations in climatic trend, trend analysis must use very long time periods.

Although seasonal and diurnal variations in multi-year averages of surface air temperature have been analyzed in the past [*e.g.*, Fassig, 1907], we know of no attempts to determine the trend of such variations. Polyak [1975] investigated different techniques for approximating the seasonal cycle in multi-year averages of meteorological variables. He estimated multi-year averages for five-day periods and then approximated their seasonal variation using Fourier harmonics, polynomials, or smoothing of the pentad averages with

statistically optimal numerical filters. The most recent global analysis of the geographical pattern of amplitude and phase of the first two harmonics of the diurnal cycle in observed temperature for two seasons, winter and summer, revealed the importance of the semidiurnal component of the diurnal cycle [*Dai and Trenberth, 2003*].

The goal of this paper is to introduce a simple technique of approximating both the diurnal and seasonal cycles as well as the climatic trend using a limited numbers of Fourier harmonics. The main advantage of this technique is that it can be applied to data with changing and arbitrary observation times. Changing observation times is well known in the history of meteorological observation in different countries. The same problem also arises with satellite observations of surface and atmospheric variables. Here we use long-term surface air temperature observations at a few regular meteorological stations to illustrate our technique. Simpler versions of this technique can be found in our recent papers [*Vinnikov and Robock, 2002; Vinnikov et al., 2002a, 2002b; Cavalieri et al., 2003*]. This exact technique was used by *Vinnikov and Grody* [2003], but they did not have room to describe the complete technique. It is the purpose of this paper to describe the general technique for others to use.

2. Approximation of not stationary component in climatic records

Let us suppose that a climatic variable $y(t)$ has been observed more or less regularly during a long period of time. We can denote the observed value of y at time t as

$$y(t) = Y(t) + y'(t); \quad (1)$$

where $Y(t)$ is the expected value of $y(t)$ and $y'(t)$ is the residual (or anomaly, in the language of meteorologists), which is mostly related to natural weather variability. *Vinnikov et al.* [2002] allowed for a linear trend in the expected value by expressing $Y(t)$ as

$$Y(t) = A(t) + B(t) \cdot t, \quad (2)$$

where $A(t) = A(t+T)$ and $B(t) = B(t+T)$ are periodic functions with a fundamental period of 1 yr ($T = 365.25$ days). That approach was applied to analyze processes without a diurnal cycle. Here we extend the analysis by assuming that $A(t)$ and $B(t)$ consist of short-time diurnal variations with a fundamental period of 1 day ($H = 1$ day) superimposed on the longer time annual cycle. By representing the diurnal and annual cycles by their Fourier series, the amplitudes become a double Fourier series,

$$\begin{aligned} A(t) &= \sum_{k=-K}^K \sum_{n=-N}^N a_{kn} e^{i2\pi t(\frac{n}{T} + \frac{k}{H})}, \\ B(t) &= \sum_{l=-L}^L \sum_{m=-M}^M b_{lm} e^{i2\pi t(\frac{m}{T} + \frac{l}{H})}. \end{aligned} \quad (3)$$

where N and M are the number of harmonics needed to approximate the seasonal variations, while K and L are the number of harmonics needed to approximate the diurnal variations in $A(t)$ and $B(t)$.

The Fourier coefficients a_{00} and b_{00} in (3) are real numbers but all others are complex numbers. These unknown coefficients are found using the least squares condition,

$$\sum_{t=t_1}^{t_p} [y(t) - Y(t)]^2 = F(a_{00}, \dots, a_{KN}, b_{00}, \dots, b_{LM}) = \min, \quad (4)$$

where $t = t_1, t_2, t_3, \dots, t_p$ are the observation times. Although the observations can be irregular, they should cover the different phases of the diurnal and seasonal cycles over different parts of the record. The number of observations should also be much larger than the number of unknown coefficients. Approximations (2) and (3) assume a linear trend, but they can easily be extended to a polynomial trend by adding higher degree terms ($t^2 \cdot C(t)$, $t^3 \cdot D(t)$, ...) in (2). This, however, increases the number of unknown coefficients and should only be used if it is really necessary and makes physical sense.

3. Application of the theory to long-term climatic records of surface air temperature

The technique outlined above is not very sensitive to gaps or irregularities in the observations. As such, it can be applied to observations with arbitrary observation times. Here, however, we apply the technique to a homogeneous 50-60 year period of hourly surface air temperature observations from six reporting stations over the United States (Table 1). As with every temperature record, these data contain diurnal and seasonal variations in both the multi-year averaged temperature and linear trend. These variations require a minimum of two Fourier harmonics in (3) to accurately approximate the seasonal and diurnal cycles, *i.e.*, $K = L = N = M = 2$. The Fourier coefficients in (3) are estimated for each station. They depend on the choice of the beginning time coordinate $t = 0$, but the estimated function $Y(t)$ does not depend on this choice. The functions $Y(t)$, $A(t)$, and $B(t)$ have a leap-year cycle, because the length of a year is not equal to an integer number of days. An analog to a multi-year average temperature (for a linear trend) can be obtained by taking the average of $Y(t)$ over two years which are in the same phase of a leap year cycle. The corresponding trend estimates are equal to $B(t)$ for any year in the same phase of a leap-year cycle.

Figure 1 shows the result derived for each station in the form of contour plots. The plots in the left column display the diurnal and seasonal variation of the multi-year average temperature, estimated as $Y(t)$, averaged for the two leap years, 1948 and 2000. Similarly, the middle column displays the trend estimates for one of the leap years, $B(t)$. The diurnal and seasonal variations of the average temperature show reasonable patterns, with the amplitude and phase of these variations depending on climatic conditions. The four northernmost stations show a tendency for winter warming. Three of them show a diurnal asymmetry in the trends with more warming at night and less warming or even a small cooling trend in the daytime. Such asymmetry has been found at many continental meteorological stations using maximum and

minimum temperatures [Karl *et al.*, 1993], but our technique displays the pattern in much more detail. The diurnal asymmetry is mostly a summertime phenomenon, as previously shown by Vinnikov *et al.* [2003b]. An opposite diurnal asymmetry of temperature trend, with maximum warming in the daytime, is found in Honolulu, Hawaii. The diurnal and seasonal cycles of the standard deviation of surface air temperature are shown in the right column of Figure 1 and display details of the expected winter and daytime maximums in temperature variability. They were estimated by applying the same technique (but with $B(t) = 0$) to the time series of squared detrended anomalies $(y')^2$ of surface temperature. We analyzed seasonal and diurnal cycles in variance, but displayed standard deviations for convenience.

4. Discussion and conclusions

We have developed a model of the expected value of a variable that combines three climatic variations (diurnal, seasonal and trend) into a single function. Vinnikov *et al.* [2002b] previously treated the diurnal and seasonal scales separately. The seasonal and diurnal variations are approximated using a limited number of Fourier harmonics, while the trend is represented by a monotonic function of time whose amplitude can also vary seasonally and diurnally. A least squares method is used to estimate the unknown Fourier coefficients. This new approach to data analysis improves the description of diurnal and seasonal cycles in the observed data and provides a more efficient means of analyzing data with changing observation times. As an example of the technique, an analysis is presented of the hourly observations of surface air temperature obtained from several meteorological stations within the United States. Both the multi-year averaged temperature and linear trend estimates show detailed diurnal and seasonal patterns that were difficult to analyze previously.

Application of this technique has a few obvious limitations. The main question is whether seasonal and diurnal variations in a climatic record can be approximated by only a small

number of Fourier harmonics. What is quite acceptable for surface air temperature may not be good at all for precipitation or other climatic variables. If we need too many harmonics to approximate the variations we would have to estimate too many unknown parameters from observed data. Available climatic records are often not long enough, and the length of a climatic record limits the number of harmonics in approximation (2, 3). If the record is too short, the errors of approximation of the expected value $Y(t)$ will be included in the residuals $y'(t)$ and misinterpreted as meteorological anomalies. To solve this problem we can use other non-harmonic periodic functions for the approximation of $A(t)$ and $B(t)$ in (2) that require only a small number of parameters. This will be a subject of one of our future publications. In the example here, we set the number of Fourier harmonics to 2 (*i.e.*, $K = L = N = M = 2$) in (1-3) since these harmonic components provide a number of statistically significant amplitudes and result in a qualitatively good approximation of the seasonal and diurnal cycles in multi-year averages and trends. To test of this assertion we assume that the autocorrelation function of the meteorological anomalies can be approximated as $\overline{y'(t)y'(t+\tau)} = \sigma^2 e^{-\tau/\lambda}$, where σ^2 is the variance, τ is time lag, and λ is the autocorrelation scale. This model ignores seasonal and diurnal variations in the variance and lag correlation. To evaluate the effect of such an autocorrelation on a number of independent data in the time series of observed hourly temperatures we used an empirically estimated value of $\lambda \approx 40$ hours. By increasing the number of harmonics up to $K = L = N = M = 4$, we find that we do not add statistically significant amplitudes and that we decrease the variance of residuals by not more than 0.6% of its value (decrease of standard deviation of $\sim 0.015^\circ\text{C}$). Even for $K = L = N = M = 2$, approximately half of the coefficients in the approximation (2-4) are not significant compared to their root mean squared errors estimated for independent observations.

Another problem appears when we are trying to analyze short climatic records that include very large natural climatic anomalies, for example those related to very strong El Niños or volcanic eruptions. If such a phenomenon occurs near the beginning or the end of the record, the trend estimates may be significantly biased. There is no better solution of this problem than to extend the record into the past or to update it if possible. If trend estimates change when a few years of observations are added to the record, they should not be simply interpreted as climatic trends.

Evaluation of statistical significance of the estimates is always a problem. Here we used the ordinary least squares technique for independent observations which are really not independent at all. The errors of the model (2, 3) coefficients obtained in such a simplification are underestimated. The number of hourly observations for 50 years is very large (438,300) compared to the number of unknown parameters. But the number of independent data is approximately 4.3 times less than the number of hourly observations. More realistic models for lag-correlation of residuals have to be studied and taken into account to estimate the actual confidence intervals of the unknown parameters. Nevertheless, our experience shows that taking into account the autocorrelation of the residuals affects the estimated $Y(t)$ only a little, but improves the statistics of the errors quite a bit. We would first have to estimate the non-stationary expected value and then to estimate the non-stationary variance and autocorrelation. We would then use them to improve the non-stationary expected value estimates and iterate.

Despite the above limitations, this is a powerful technique that can be used for traditional and satellite derived climatic records. The technique presented in this paper was used by *Vinnikov and Grody* [2003] in their recent analysis of satellite observed trend in globally averaged mean tropospheric temperature, and we look forward to its application for many other climatological time series.

Acknowledgements. We thank Abram Kagan, Semyon Grodsky, and Anandu Vernekar for useful discussions, and an anonymous reviewer for valuable suggestions. Meteorological data have been selected and kindly provided by NCDC/NOAA scientists. This work supported by NOAA grants NAO6GPO403 and NA17EC1483.

References

- Anderson, T. W., *The Statistical Analysis of Time Series*, Wiley, New York, 755 pp., 1971.
- Cavalieri, D. J., C. L. Parkinson, and K. Y. Vinnikov, 30-Year satellite record reveals contrasting Arctic and Antarctic decadal sea ice variability, *Geophys. Res. Lett.*, *30* (18), 1970, doi:10.1029/2003GL018031, 2003.
- Dai, A., and K. E. Trenberth, The diurnal cycle and its depiction in the Community Climate System Model, *J. Climate*, in press, 2003.
- Fassig, O. L., The climate and weather of Baltimore, in *Maryland Weather Service*, vol. 2, Johns Hopkins Press, Baltimore, 27-511, 1907.
- Polyak, I. I., *Numerical Methods in Analysis of Observations*, Gidrometeoizdat, Leningrad, 211 pp., 1975. (In Russian)
- Vinnikov, K. Y., and A. Robock, Trends in moments of climatic indices, *Geophys. Res. Lett.*, *29* (2), doi:10.1029/2001GL014025, 2002.
- Vinnikov, K. Y., A. Robock, D. J. Cavalieri, and C. L. Parkinson, Analysis of seasonal cycles in climatic trends with application to satellite observations of sea ice extent, *Geophys. Res. Lett.*, *29* (9), doi:10.1029/2001GL014481, 2002a.
- Vinnikov, K. Y., A. Robock, and A. M. Basist, Diurnal and seasonal cycles of trends of surface air temperature, *J. Geophys. Res.*, *107* (D22), 4641, doi:10.1029/2001JD002007, 2002b.
- Vinnikov, K. Y., and N. C. Grody, Global warming trend of mean tropospheric temperature observed by satellites, *Science*, *302*, 269-272, 2003.

Table 1. List of stations.

Station Name	State	Station ID	Lat. (N)	Lon. (W)	Elevation	Start	End
Bellingham International Airport	WA	BLI	48°48'	122°32'	45.4 m	1948	2001
Bellevue Offutt Air Force Base	IL	OFF	41°07'	95°55'	319.1 m	1948	2001
Belleville Scott Air Force Base	NE	BLV	38°33'	89°51'	135.0 m	1938	2001
Andrews Air Force Base	MD	ADW	38°49'	76°52'	86.0 m	1944	2001
Maxwell Air Force Base	AL	MXF	32°23'	86°21'	53.0 m	1936	2001
Honolulu International Airport	HI	HNL	21°19'	157°56'	2.1 m	1939	2001

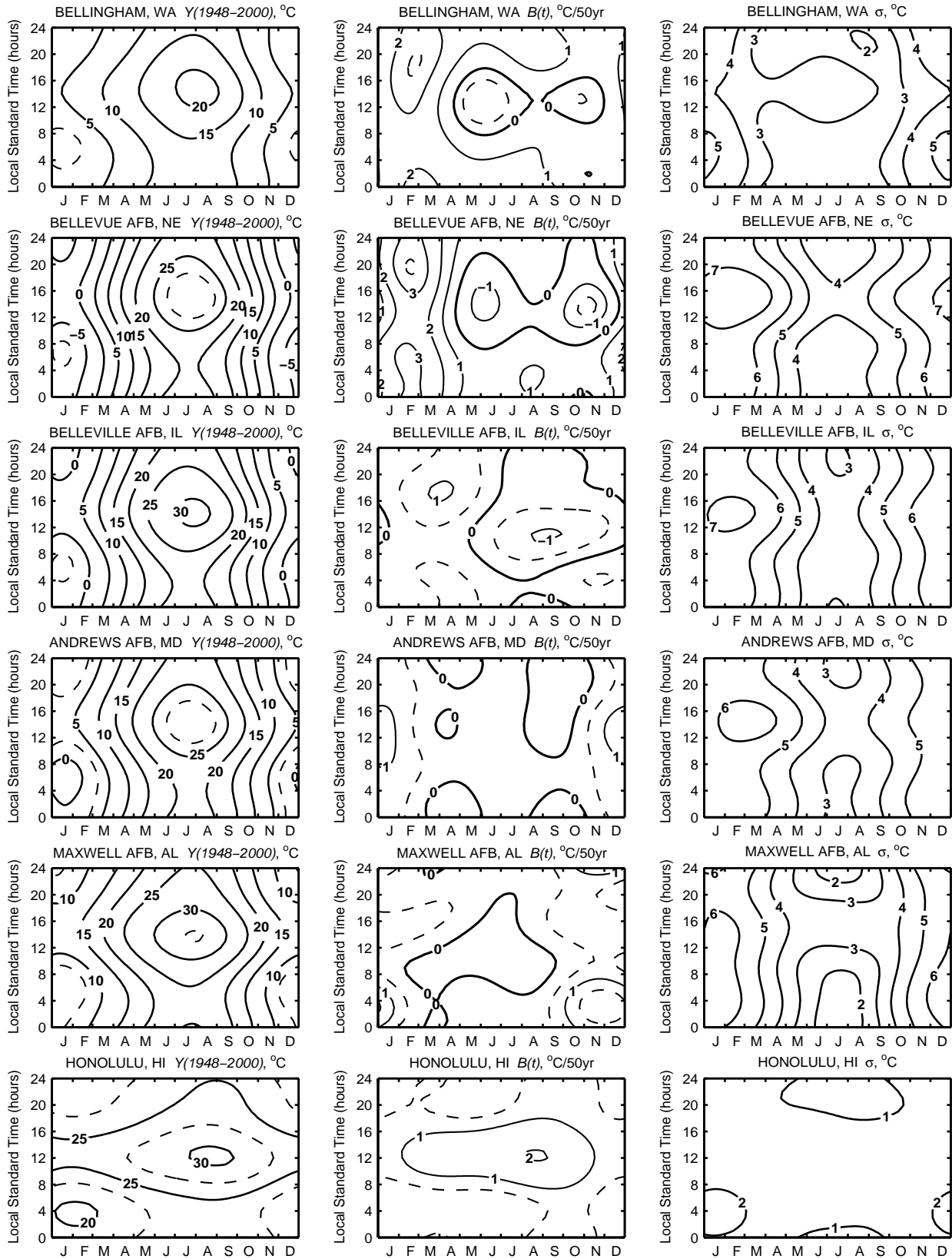


Figure 1. Diurnal and seasonal variations in averages $Y(t)$ for years 1948 and 2000, linear trends $B(t)$, and standard deviations $\sigma(t)$ of surface air temperature at stations listed in Table 1. The x-axis shows time (months) from the beginning of the year while the y-axis shows the hours within a day.