Journal of the Air & Waste Management Association

ISSN: 1096-2247 (Print) 2162-2906 (Online) Journal homepage: https://www.tandfonline.com/loi/uawm20

Linear Least Squares Method for Time Series Analysis with an Application to a Methane Time Series

M.A.K. Khalil & F.P. Moraes

To cite this article: M.A.K. Khalil & F.P. Moraes (1995) Linear Least Squares Method for Time Series Analysis with an Application to a Methane Time Series, Journal of the Air & Waste Management Association, 45:1, 62-63, DOI: 10.1080/10473289.1995.10467343

To link to this article: https://doi.org/10.1080/10473289.1995.10467343

Published online: 05 Mar 2012.

ISSN 1047-3289 *J. Air & Waste Manage. Assoc.* **45**: 62-63

Linear Least Squares Method for Time Series Analysis with an Application to a Methane Time Series

M.A.K. Khalil and F.P. Moraes
Global Change Research Center, Oregon Graduate Institute, Beaverton, Oregon

ABSTRACT
A simple method of time series analysis, based upon linear least squares curve fitting, is developed. The method's advantages and disadvantages are discussed, and an example is presented using the Vostok Core methane record.

INTRODUCTION
Time series analysis is one of the most important analytical tools in the experimental sciences. As such, many procedures for finding periodicities in data have been developed. The Linear Least Squares Spectral Analysis (LLSSA) method is an extension of the curve-fitting approach developed in Bloomfield.[1] For environmental science, where signals generally have more than one periodicity but relatively few data points, this method may be the best available. Many procedures exist for determining periodicities in these kinds of data, but this method has several compelling features worth noting:

Unevenly spaced data are analyzed as easily as evenly spaced data.

Any frequency may be considered.

The statistical significance of every peak is easily computed.

In addition to these benefits, all output is easily understood and validated.

These features are not without cost, however. The algorithm is computationally intensive, requiring F·N·(2·M+7) floating point multiplication operations, where F is the number of frequencies searched, N is the number of data points, and M is the number of floating point multiplication operations (typically about 100 operations) necessary for a computation of a sine or cosine. Time series analysis procedures have generally been developed by astronomers and engineers, who both analyze large data sets. For such data sets, this approach may be too slow computationally. In the environmental sciences, however, data sets are generally much smaller, and so this approach to time series analysis is quite reasonable. In addition, due to speed advances in microprocessors, this algorithm is reasonably run on personal computers when the data set is relatively short. For example, it took less than five seconds to run the example presented below on an i80386 class machine.

THEORY
The procedure is straightforward. For each member of a set of frequencies, the best fit sinusoid to the data is found. Usually this set will range from the inverse of the data set length to the Nyquist Frequency. The power for each frequency is then given simply by the square of the amplitude of the fit. The process is discussed in more detail below. For each frequency, ω, we wish to fit the following function to the given time series.

$$y_i - y_{avg} = A\cdot\cos\omega\, t_i + B\cdot\sin\omega\, t_i$$

Using least squares theory (e.g., Bloomfield[1]) we minimize variance by taking its derivative with respect to A and B, setting each equation equal to zero, and solving for A and B. Doing this results in the following functions.

$$A = \frac{1}{\Delta}\left[\Sigma y_i \cos\omega\, t_i \Sigma \sin^2\omega\, t_i - \Sigma y_i \sin\omega\, t_i \Sigma\cos\omega\, t_i \sin\omega\, t_i\right]$$

$$B = \frac{1}{\Delta}\left[\Sigma y_i \sin\omega\, t_i \Sigma \cos^2\omega\, t_i - \Sigma y_i \cos\omega\, t_i \Sigma\cos\omega\, t_i \sin\omega\, t_i\right]$$

$$\Delta = \Sigma \cos^2\omega\, t_i\, \Sigma \sin^2\omega\, t_i - (\Sigma\cos\omega\, t_i \sin\omega\, t_i)^2$$

The power is then given by the sum of the squares of the amplitudes. The error on each of the parameters A and B is given by least squares theory.[2]

$$\sigma_A^2 = \frac{(y-y_{fit})^2\, \Sigma \sin^2\omega\, t_i}{(N-2)\cdot\Delta}$$

$$\sigma_B^2 = \frac{(y-y_{fit})^2\, \Sigma \cos^2\omega\, t_i}{(N-2)\cdot\Delta}$$

The total error on the power is then given simply.

$$\sigma_P^2 = 4\cdot(A^2\sigma_A^2 + B^2\sigma_B^2)$$

The phase of the periodicity and its error are given in the following equations.

$$\varnothing = \begin{cases} \tan^{-1}(-B/A) & A>0 \\ \tan^{-1}(-B/A)-\pi & A<0,\ B>0 \\ \tan^{-1}(-B/A)+\pi & A<0,\ B<0 \end{cases}$$

$$\sigma_\varnothing = \sqrt{\left(\frac{B\cdot\sigma_A}{A^2+B^2}\right)^2 - \left(\frac{A\cdot\sigma_B}{A^2+B^2}\right)^2}$$

For every frequency of interest, the best-fit sinusoid is found. This is different from Fourier series analysis where, in effect, all of the frequencies are simultaneously found, and the frequencies are constrained to be 1/n (n=2,3,4...) of the data set length.

EXAMPLE

The Vostok ice core is a 2,083-meter-long core drilled in East Antarctica. Air bubbles from this core have been analyzed to recreate the temperature, CO_2, and CH_4 levels over the past 160,000 years.[3] We have applied the LLSSA method to the methane data set. The data and the least squares spectrum fit are plotted in Figure 1. The least squares model of the data are composed of the 10 most important periodicities.

As a point of reference, the power spectrum (the error bars are one sigma) is presented in Figure 2 along with the power spectrum of the Fast Fourier Transform (FFT). The raw Vostok ice core data were used with the LLSSA method to determine the spectrum, but because the FFT requires evenly spaced data, it was necessary to linearly interpolate the data in order to determine the spectrum using this technique. To be fair, the LLSSA method searched only those frequencies that the FFT did (1/n times the data set length). But it is important to remember that any frequency less than the Nyquist Frequency could have been searched—and this is one of the biggest strengths of the LLSSA method in addition to its ability to use unevenly spaced data as inputs.

CONCLUSIONS

The LLSSA method of time series analysis provides an effective and robust method for the determination of periodicities in noisy data. It is of particular value when comparing coupled processes where phase and error information are important.

The algorithm has been fully developed in a computer program.[4] The program allows users to search through a set of evenly spaced frequencies or periods. It outputs frequency (or period), power, and the error on the power by default, but may be changed by the user to output the A and B coefficients and their errors as well. The executable version of this program (for MS-DOS, OS/2, and Linux only) or C source code is available upon request from the authors.

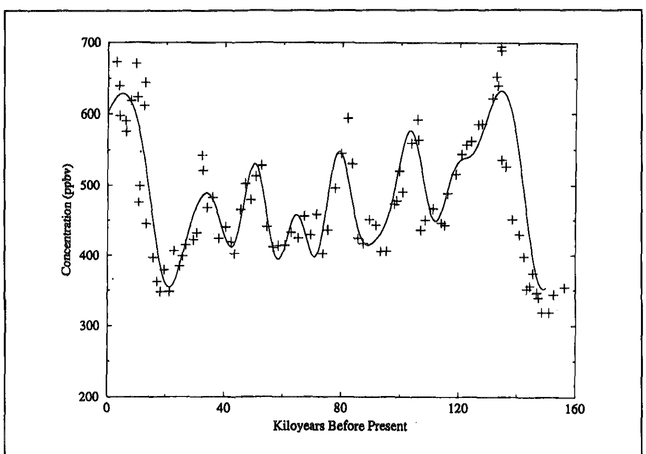

Figure 1. The Vostok ice core methane database (plus signs) with the top 10 frequency fit based upon analysis with the LLSSA method (solid line). The frequencies are 0.001, 0.008, 0.016, 0.024, 0.032, 0.040, 0.047, 0.055, 0.062, and 0.071 (all in inverse kiloyears).

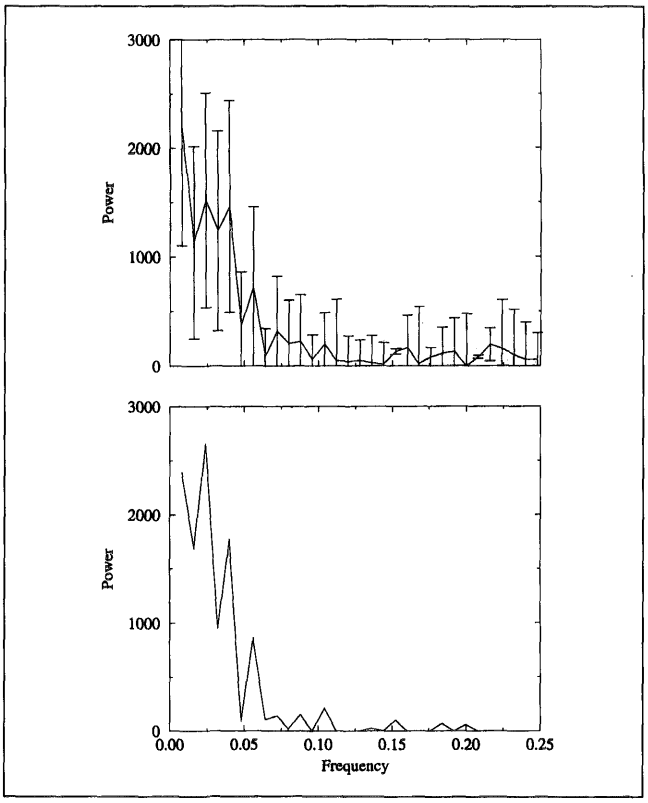

Figure 2. The power spectrum of the Vostok ice core methane database using (top) the LLSSA method with one-sigma error bars as given by Equation 7 in the text, and the FFT (bottom).

REFERENCES

1. Bloomfield, P. *Fourier Analysis of Time Series: An Introduction*, John Wiley & Sons, 1976.
2. Snedecor, G.W.; Cochran, W.G. *Statistical Methods*, 7th ed.; The Iowa State University Press, 1980.
3. Chappellaz, J.; Barnola, J.M.; Raynaud, D.; Korotkevich, Y.S.; Lorius, C. "Ice-core record of atmospheric methane over the past 160,000 years," *Nature* **1990**, *345*, 127.
4. Moraes, F.P.; Khalil, M.A.K. "LLSSA: A computer program for spectral analysis of unevenly spaced data," Technical report, Oregon Graduate Institute, 1993.

About The Authors

M.A.K. Khalil and Francis Moraes are with the Global Change Research Center, Oregon Graduate Institute, 20000 N.W. Walker Rd., Beaverton, OR 97006. Dr. Khalil is director of the Global Change Research Center and a professor at the Department of Environmental Science and Engineering. Francis Moraes is a research associate in the Global Change Research Center, specializing in modeling of remote atmospheric chemistry.