Paper SAS585-2017

Time-Frequency Analysis Using SAS®

Michael Leonard, Wei Xiao, and Arin Chaudhuri, SAS Institute Inc.

ABSTRACT

Many organizations need to analyze large numbers of time series that have time-varying or frequency-varying properties (or both). The time-varying properties can include time-varying trends, and the frequency-varying properties can include time-varying periodic cycles. Time-frequency analysis simultaneously analyzes both time and frequency; it is particularly useful for monitoring time series that contain several signals of differing frequency. These signals are commonplace in data that are associated with the Internet of Things (IoT). This paper introduces techniques for large-scale time-frequency analysis and uses SAS® Visual Forecasting and SAS/ETS® software to demonstrate these techniques.

INTRODUCTION

For the analysis of data with respect to time, there are many time domain analyses: autocorrelation analysis, cross-correlation analysis, stationarity analysis, seasonal adjustment/decomposition analysis, singular spectrum analysis (principal component analysis for time series), count series analysis, and many others.

For the analysis of data with respect to frequency, there are many frequency domain analyses: Fourier analysis, wavelet analysis, Laplace transform analysis, and many others. Most frequency domain analyses assume that the time series is stationary.

Time-frequency analysis (TFA) *simultaneously* analyzes both time and frequency. TFA is particular useful for time series data that are nonstationary with respect to time, frequency, or both: time series that have time-varying trends, time-varying periodic cycles, irregular cycles, and other time-dependent properties.

TFA is especially useful for monitoring high-frequency time series that have multiple cycles. Some examples follow:

- Forecasting and monitoring hourly demand, such as the demand for electricity, natural gas, water, telecommunications, and other utilities and services.
- Monitoring the Internet of Things (IoT), such as monitoring “edge” components for proper operations, detecting failures, and predictive maintenance.
- Investigating natural processes, such as tides, weather, and their influences as external factors on other systems.

There are many other uses of TFA.

BACKROUND

Many statistical analysis techniques are available for time domain analysis and frequency domain analysis. For time domain analysis, Box, Jenkins, and Reinsel (2015) provide an introductory treatment, and Fuller (1995) provides an extensive treatment. For frequency domain analysis, Bloomfield (2013) provides an introductory treatment. For the combined time-frequency analysis, Coen (1995) provides an introductory treatment. This paper explains some basic principles and demonstrates the SAS® implementation that is based on these texts.

TIME DOMAIN ANALYSIS

A (discrete) time index is represented by $t = 0, \dots, (T-1)$, where T represents the length of the historical time series. A time series value at time index *t* is represented by y_t with $y_t \in \mathbb{R}$. The historical time series vector that you want to analyze is represented by $\{y_t\}_{t=0}^{T-1}$.

There are numerous time domain analyses, many of which are implemented in SAS. For more information about time domain analysis, see the *SAS/ETS User’s Guide*.

FREQUENCY DOMAIN ANALYSIS

The theoretical discrete-time Fourier transform (DFT) is used for frequency domain analysis of a (discrete) time series. The following equations define the transform:

$Y(w) = \mathcal{F}(y_t) = \sum_{t=-\infty}^{\infty} y_t\, e^{-i\omega t}$ or $Y(f) = \mathcal{F}(y_t) = \sum_{t=-\infty}^{\infty} y_t\, e^{-i2\pi f t}$

where i is the imaginary unit ($i^2 = 1$), f is the frequency (in hertz), and $\omega = 2\pi f$ is the radian frequency (in radians per second) whose range is $0 \le \omega < 2\pi$. Note: Using Euler's formula, $e^{i\omega t} = \cos(\omega t) + i \cdot \sin(\omega t)$.

Because an infinite amount of data is impossible, the discrete-time Fourier transform (DTFT) is used in practice,

$$Y_k = \sum_{t=0}^{T-1} y_t\, e^{-i2\pi k t/T}$$

where $\omega = 2\pi k/T$ and $f = k/T$ and where k is the frequency index that ranges between $-(T-1)/2 < k < (T-1)/2$ when T is odd and between $-T/2 < k < (T-1)/2$ when T is even.

When y_t is a real-valued time series, the Fourier transform is symmetric: $Y_{-k} = Y_k$. Therefore, only nonnegative frequencies are sometimes considered. Since $Y_k = a_k + ib_k$ is a complex-number, it is usually visualized in terms of its magnitude, $\|Y_k\| = \sqrt{a_k^2 + b_k^2}$, or spectral power, $\|Y_k\|^2$.

WINDOW FUNCTIONS

A smoothing window function typically tapers values within a chosen range and assigns 0 to values outside this range:

$w(n) > 0$ for $-N \le n < N$ where $L_w = (2N + 1)$ is the smoothing window length

$w(n) = 0$ otherwise

There are many examples of window functions: rectangular, triangular, Parzen, and many others. The window function is useful for smoothing the time series prior to applying the Fourier transform or smoothing the spectra after applying the Fourier transform.

For smoothing in the time domain, you smooth with respect to time, t.

$\tilde{y}_t = \sum_{n=-N}^{N} w(n) y_{t-n}$

For smoothing in the frequency domain, you smooth with respect to frequency, k.

$\tilde{Y}_k = \sum_{n=-N}^{N} w(n) Y_{k-n}$

TIME-FREQUENCY ANALYSIS

Time-frequency analysis (TFA) helps you understand how the frequency domain properties change with time. The theoretical short-time Fourier transform (STFT) is used for frequency domain analysis of a (discrete) time series. The following equations define the transform:

$Y(\tau, \omega) = \mathcal{F}_w(\tau, y_t) = \sum_{t=-\infty}^{\infty} y_t\, s(t-\tau) e^{-i\omega(t-\tau)}$ or $Y(\tau, \omega) = \mathcal{F}_w(\tau, y_t) = \sum_{t=-\infty}^{\infty} y_t\, s(t-\tau) e^{-i2\pi f(t-\tau)}$

where τ is the (discrete) window index and $s(l)$ is a selector window function that is often either zero or one.

A selector window function typically selects the contiguous portion of the time series of concern:

$s(l) = 1$ for $0 \le l < L_s$ where L_s is the window length

$s(l) = 0$ otherwise

Because an infinite amount of data is impossible, the discrete-time STFT is used in practice,

$Y_{\tau,k} = \sum_{t=\tau}^{\tau+L-1} y_t - e^{i2\pi k(t-\tau)/L}$

where τ ranges from $0 \le \tau < (T - L_s + 1)$.

In essence, $Y_{\tau,k}$ represents a "sliding time window" discrete Fourier analysis.

Note that $Y_{\tau,k}$ is a single element of a $\mathrm{T} \times (T - L_s + 1)$ matrix. However, for long time series, many Fourier transforms would be required. So, in addition to the selector window length, L_s, a slide parameter is used to control the amount of overlap between slides.

For example, $Y_{\tau,k}$ can be computed only for $\tau = 0, \vartheta, 2\vartheta, 3\vartheta, \ldots$ where the overlap is half the window length, $\vartheta = L_s/2$.

The resulting matrix is the STFT.

SAS IMPLEMENTATION

The TIMEDATA procedure provides a TFA package for time-frequency analysis. This package contains the following methods for analyzing time series in the frequency domain:

TFA.WINDOW	creates a window of a requested type and length.
TFA.STFT	computes the short-time Fourier transform for a specified real-valued time series.
TFA.FFT	computes the fast Fourier transform for a specified real-valued time series.
TFA.FFTC	computes the fast Fourier transform for a specified complex-valued time series.
TFA.HILBERT	computes the discrete Hilbert transform of a specified real-valued time series.
TFA.PWV	computes the pseudo-Wigner-Ville distribution of a specified real-valued time series.

TIME-FREQUENCY ANALYSIS PLOT

STFT produces a large amount of data that are associated with its computations, so a graphical display is needed for human understanding. Figure 1 illustrates an example of a time frequency analysis (TFA) plot. It is composed of the three graphs: a time series plot, a periodogram plot, and an STFT heat map.

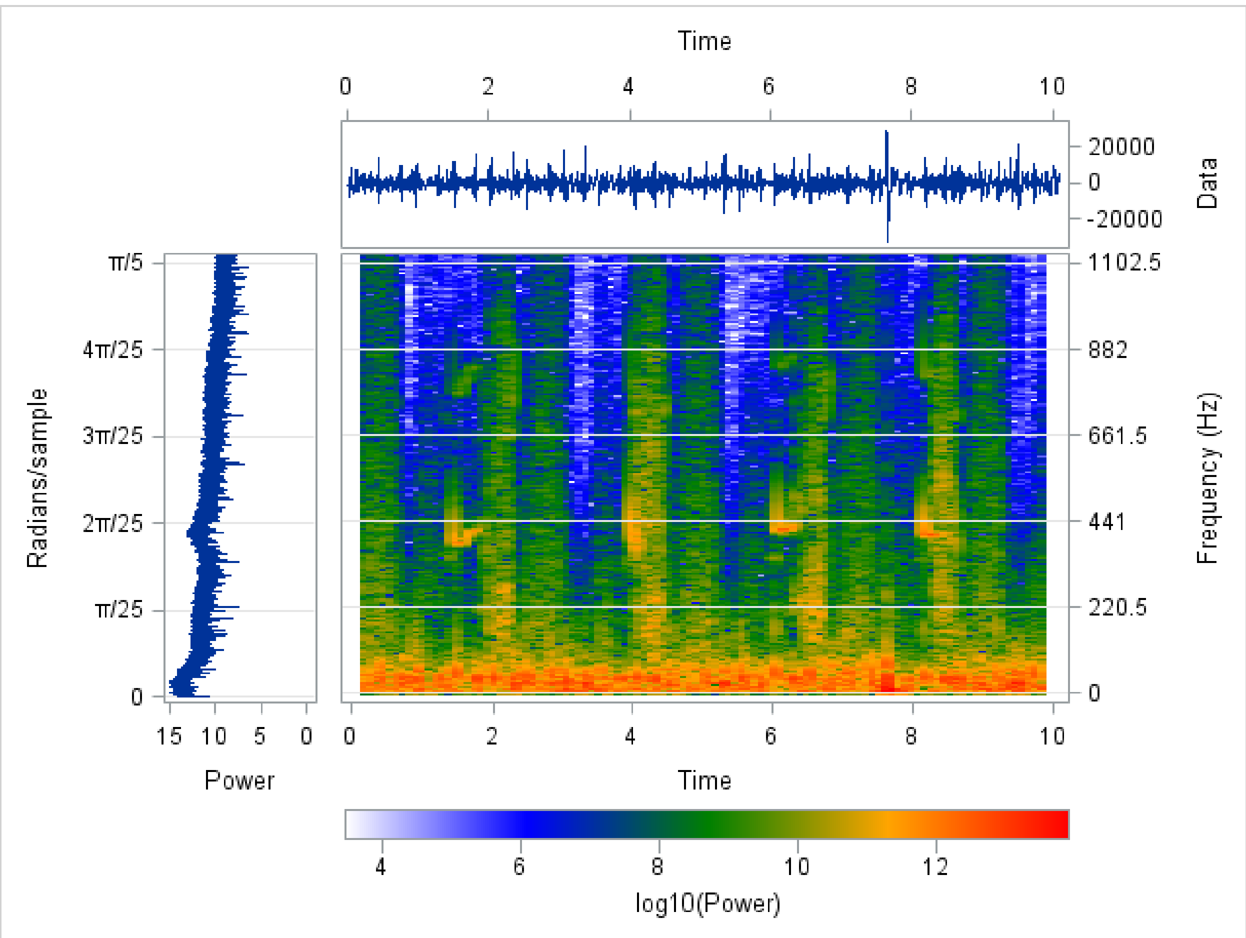


Figure 1: Time-Frequency Analysis Plot

The upper graph is a time series plot. The upper horizontal axis displays the time index: $t = 1, \dots, T$. The left-side vertical axis displays the range of time series values. The blue line displays the time series values: $\{y_t\}_{t=1}^{T}$.

The left-side graph displays the periodogram plot for the *entire* time series history. The left-side vertical axis displays

the radian frequency. The lower horizontal axis displays the power at each frequency. The blue line displays the periodogram values: $\{\|Y_k\|\}_{k=1}^{K}$.

The heat map displays the STFT plot. The lower horizontal axis displays the time index: $t = 1, \ldots, T$. The right-side vertical axis displays frequency (hertz). The colors in the heat map indicate the magnitude of the power, $\|Y_{\tau,k}\|$, as displayed in the legend. Each row of the heat map represents a time series at a particular frequency. Each column of the heat map represents the periodogram that is associated with a particular time.

TIME-FREQUENCY ANALYSIS EXAMPLE

Time-frequency analysis was performed on time series data related to noise from breathing. The goal is to compare breathing patterns between asthma patients. Figure 1 illustrates a patient who is on the verge of an asthma attack.

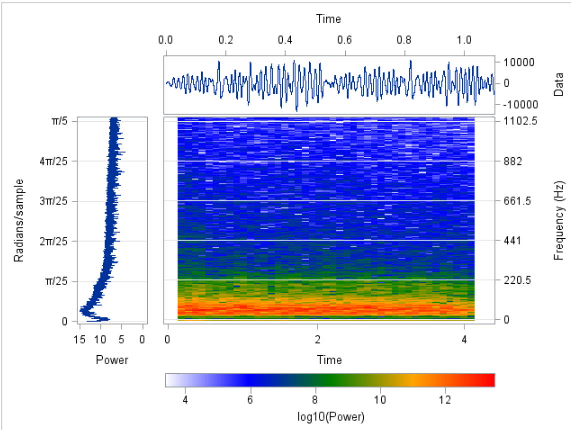
The following code uses the TIMEDATA procedure to perform FFT and STFT analyses. The DATA= option specifies the input data set, Breath. The VAR statement requests that the X variable be read and that an X array be created for subsequent programming statements. The REQUIRE statement requests that the TFA package be loaded. The DECLARE statement initializes the TFA object for use in subsequent programming statements. The three ARRAY statements create arrays to store the results of the subsequent analysis.

```
proc timedata data=breath out=_null_;
   var x;
   require tfa;
   declare object tfa(tfa);
   array fft[1]/nosymbols;
   array window[1]/nosymbols;
   array stft[1,1]/nosymbols;

  /*--- perform FFT on input data ---*/
   rc = TFA.FFT(x, 'forward', fft); if rc then stop;
   rc = append_array('fft', fft, 'coef_re', 'coef_im'); if rc then stop;

   /*--- create a HANNING window of size 4096 ---*/
   rc = TFA.WINDOW('hanning',4096,,window); if rc then stop;

   /*--- perform STFT on input data, with an overlap of 3072 ---*/
   rc = TFA.STFT(x,window,3072,4096,0,1,stft); if rc then stop;
   rc = append_array('stft', stft, 'frequency', 'time', 'power',
                     'amplitude', 'phase', 'coef_re', 'coef_im');
   if rc then stop;
run;
```

The rest of the code preforms the computations.

1. The TFA.FFT method performs the fast Fourier transform and outputs the result as a data set named FFT by using the APPEND_ARRAY function.
2. The TFA.WINDOW method creates a Hanning window of length 4,096.
3. The TFA.STFT method performs the short time Fourier transform by using the Hanning window and an overlap of 4,096 and outputs the result as a data set named STFT by using the APPEND_ARRAY function.

Similar analyses were performed on other patients. Figure 2 illustrates a patient who has normal breathing. Figure 3 illustrates a patient who has asthmatic breathing. If you compare the time series plots alone, it is difficult to discern the differences (mainly due to scaling). If you compare the periodogram plots alone, it is difficult to discern the differences. However, by comparing the two TFA analyses you can readily identify the differences.



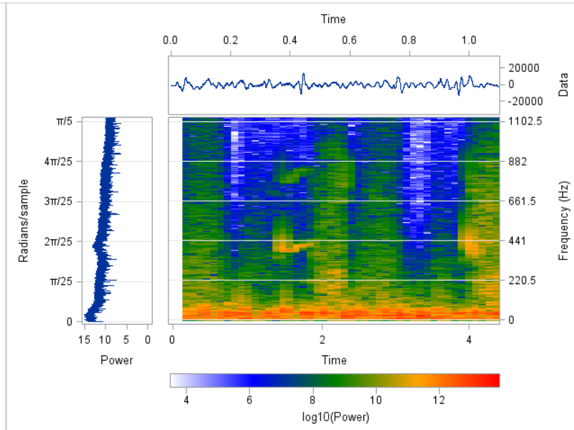


Figure 2: TFA of Normal breathing

Figure 3: TFA of Asthmatic Breathing

CONCLUSION

Time-frequency analysis (TFA) is particularly useful for time series data that are nonstationary with respect to time, frequency, or both. Such time series have time-varying properties: trends, periodic cycles, irregular cycles, and other properties. TFA is useful for differentiating between two time series with differing frequency patterns. This paper uses SAS/ETS and SAS Visual Forecasting software to demonstrate TFA.

REFERENCES

Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. P. 2015. *Time Series Analysis: Forecasting and Control.* 5th ed. New York: Wiley.

Bloomfield, P. 2013. *Fourier Analysis of Time Series: An Introduction, 2*nd ed., New York: Wiley.

Coen, L. 1995. *Time Frequency Analysis: Theory and Applications*. Upper Saddle River, New Jersey: Prentice Hall PTR.

Fuller, W. A. 1995. *Introduction to Statistical Time Series.* 2nd ed. New York: Wiley.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Michael Leonard
SAS Campus Drive
Cary, NC 27513
919-531-6967
Michael.Leonard@sas.com

Arin Chaudhuri
SAS Campus Drive
Cary, NC 27513
919-531-1282
Arin.Chaudhuri@sas.com

Wei Xiao
SAS Campus Drive
Cary, NC 27513
919-531-3294
Wei.Xiao@sas.com