THE ART OF CONTEMPORARY TIME SERIES ANALYSIS & FORECASTING

D. A. HEGER, PhD
CEO/Owner DHTechnologies
& Data Nubes

DHTechnologies & Data Nubes, Austin, USA

In General, a time series reflects a collection of quantitative observations that are evenly spaced in time and are measured successively. Examples of time series datasets comprise the uninterrupted monitoring of a person's heart rate, the hourly readings of air temperature values, the daily closing price of a company's stock price, some monthly rainfall data at location x, or some yearly sales figures.

Time series data is analyzed in order to understand the underlying structure and function embedded into the dataset that actually produces the observations. Understanding the means of a time series allows developing a mathematical model that explains the data so that forecasting, monitoring, or control functions can be applied.

Observations made over time can be either discrete or continuous. Both types of observations can be equally spaced, unequally spaced, and/or may have missing data points. Discrete measurements can be recorded at any time interval (are most often taken at evenly spaced intervals). Continuous measurements can be spaced randomly in time (such as measuring the occurrence of an earthquake) as the recording instruments are constantly collecting data. Constant measurement scenarios normally refer to a natural phenomenon such as air temperature values or a process such as velocity assessment of an airplane.

Time series data is considered as being rather complex as each observation is somewhat dependent upon the previous observation and each data point is frequently influenced by more than 1 previous observations. The stochastic error behavior is also important and has to be studied from 1 observation to another. These influences are labeled *autocorrelation* and describe dependent relationships among successive observations of the same variable. One of the challenges in time series analysis is to extract the autocorrelation elements of the data set to either understand the trend itself or to model the underlying mechanisms. To reiterate, time series data reflect the stochastic nature of most measurements over time. Ergo, the data may be skewed, the mean and the variation may not be constant, the data set may not be normally distributed, not randomly sampled or may not reflect independent data points. Further, the data set may not be evenly spaced out in time due to either instrument or sensor failure scenarios or simply due to the fluctuation of the actual number of days in a month. Obviously, the introduction of Big Data and its current focus on real-time predictive analytics has profoundly changed the landscape of time series analysis as vast amounts of data may today have to be forecasted in a (near) real-time fashion.

This presentation focuses on discussing the pros and cons of time series algorithms such as ARIMA, Bayesian Networks, and Artificial Neural Networks (ANN) for time series forecasting. Further, an approach where the time series data is moved into the frequency domain (via Fast Fourier Transform) and forecasted that way via ANN's is presented. All algorithms are introduced by means of actual demos during the talk.