Tutorials and worked examples for simulation, curve fitting, statistical analysis, and plotting.
http://www.simfit.org.uk

This technique is used when there are two corresponding time series, or in fact any series of signals recorded at a sequence of fixed discrete intervals of time or space etc., and a comparison of the two series is required.

From the main SIMFIT menu choose [Statistics], [Time series], then [Auto- and cross-correlation matrices] and examine the test file `g13dmf.tf1` provided which contains the following data.

X	*Y*
-1.490	7.340
-1.620	6.350
5.200	6.960
6.230	8.540
6.210	6.620
5.860	4.970
4.090	4.550
3.180	4.810
2.620	4.750
1.490	4.760
1.170	10.880
0.850	10.010
-0.350	11.620
0.240	10.360
2.440	6.400
2.580	6.240
2.040	7.930
0.400	4.040
2.260	3.730
3.340	5.600
5.090	5.350
5.000	6.810
4.780	8.270
4.110	7.680
3.450	6.650
1.650	6.080
1.290	10.250
4.090	9.140
6.320	17.750
7.500	13.300
3.890	9.630
1.580	6.800
5.210	4.080
5.250	5.060
4.930	4.940
7.380	6.650
5.870	7.940
5.810	10.760
9.680	11.890
9.070	5.850
7.290	9.010
7.840	7.500
7.550	10.020
7.320	10.380
7.970	8.150
7.760	8.370
7.000	10.730
8.350	12.140

Column 1 contains time series X and column 2 contains the corresponding time series Y. As multivariate time series with more than two variables can be difficult to analyze it is necessary to select any two variables for pairwise analysis using this technique.

The next table illustrates the results from analyzing test file `g13dmf.tf1` for the first ten lags using the SIMFIT cross-correlation matrices time series options.

Auto- and cross-correlation matrices

Sample size $n = 48$
Approximate standard deviation = 0.1443
Mean of $X = 4.37021$
Mean of $Y = 7.86750$
For lag $m = 0$: sample X, Y Correlation coefficient $r = 0.2493$

m = 1	0.7366(***)	0.1743
	0.2114	0.5541(***)
m = 2	0.4562(**)	0.0764
	0.0693	0.2602
m = 3	0.3795(**)	0.0138
	0.0260	-0.0381
m = 4	0.3227(*)	0.1100
	0.0933	-0.2357
m = 5	0.3414(*)	0.2694
	0.0872	-0.2499
m = 6	0.3634(*)	0.3436(*)
	0.1323	-0.2263
m = 7	0.2802	0.4254(**)
	0.2069	-0.1283
m = 8	0.2482	0.5217(***)
	0.1970	-0.0845
m = 9	0.2400	0.2664
	0.2537	0.0745
m = 10	0.1621	-0.0197
	0.2667	0.0047

Indicators: $p < 0.005$(***), $p < 0.01$(**), $p < 0.05$(*)
Maximum off-diagonal, $m = 8$, $|C(1,2)| = 0.5217$

In the above two by two matrices r_{ij} the positions have the following meanings at the lags indicated.

$r(1,1)$: auto-correlation for X
$r(2,2)$: auto-correlation for Y
$r(1,2)$: cross-correlation for X with Y (lags in Y)
$r(2,1)$: cross-correlation for Y with X (lags in X)

The significance levels are indicated if $p \leq 0.05$. These indicate significant auto-correlation for X at lags 1 to 6 but for Y only at lag 1, and also significant cross-correlation for r_{12} at lags 6 to 8.

Theory

It is assumed that the data are from a multivariate time series or similar set of observations of several variables at fixed intervals, and it is wished to make pairwise analysis of such observations.

The data must be supplied as two vectors, say X and Y of length n for instance, with X as column 1 of a n by 2 matrix, and Y as column 2.

The routine first calculates the sample means $\bar{x}$ and $\bar{y}$, the sample variances V_x and V_y, and sample correlation coefficient r. Then, for a selected number of lags $m = 1, 2, \ldots, k$, the auto-correlations and cross-correlations are output as a sequence of 2 by 2 matrices.

Since $1/\sqrt{n}$ is a rough approximation to the standard errors of these estimates, the approximate significance for the sample cross-correlations is indicated as in the table using the following labeling scheme.

$$|r(i,j)| > 3.29/\sqrt{n} : ***$$
$$|r(i,j)| > 2.58/\sqrt{n} : **$$
$$|r(i,j)| > 1.96/\sqrt{n} : *.$$

Finally, the off-diagonal i.e., cross-correlation, coefficient with largest absolute value is indicated. If this value is close to unity it indicates that the series are closely similar, and the value of m at which this occurs indicates the extent to which the series have to be slid past each other to obtain maximum similarity of profiles. Usually, the largest value of m selected for analysis would be for $k \leq n/4$.

Defining the denominator D as follows

$$D = \sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

then the auto-correlations $r(1,1)$ and $r(2,2)$, and the cross-correlations $r(1,2)$ and $r(2,1)$ as functions of m are given by

$$r(1,1) = \frac{1}{D}\sum_{i=1}^{n-m}(x_i - \bar{x})(x_{i+m} - \bar{x})$$
$$r(1,2) = \frac{1}{D}\sum_{i=1}^{n-m}(x_i - \bar{x})(y_{i+m} - \bar{y})$$
$$r(2,1) = \frac{1}{D}\sum_{i=1}^{n-m}(x_{i+m} - \bar{x})(y_i - \bar{y})$$
$$r(2,2) = \frac{1}{D}\sum_{i=1}^{n-m}(y_i - \bar{y})(y_{i+m} - \bar{y})$$

for $m = 1, 2, \ldots, k$.