**Appendix**

Assuming *n* samples with all variables (*m*) for each sample, the matrix is:

(1)

where *xij*≥0 (*i* =1, 2, …, *m*; *j* = 1, 2, …, *n*), with at least one nonzero point in each row and column. According to:

(2)

*xi* and *xj* can be expressed as:

(3)



(4)



From Equation (2), we obtain Equation (5):

(Grunsky, 2010) (5)



Based on the eigenvalues of the variance-covariance matrix (λ*1*≥ λ*2*≥ ... ≥ λ*m*), the corresponding unit eigenvectors (*u1*, *u2*, ..., *up*) are calculated using the first P eigenvalues (λ*1*, λ*2*, ..., λ*p*), whose cumulative percentage of eigenvalues is ≥80%, to obtain the R-type factor loading matrix as follows:

(6)

Scatter data analysis based on the matrix is the so-called R-type factor analysis, which indicates the degree of correlation between variables.

The contribution rate of the selected elements was calculated according to the R-type factor loading matrix of the variable analysis. The elements that had a significant impact on the sample analysis were screened at the same time. Q-type factor analysis shows the relationship between the samples according to the scatter plot of the Q-type factor load matrix on the factor plane. Sample classification was based on the correlation coefficient (*Cij*) from the Q factor analysis. The correlation coefficient is the angle cosine obtained after data normalization:

(7)

Clustering effectiveness can be assessed using the similarity coefficient. The similarity coefficient between the clustered data *yi* and the original signal *sj*is defined by equation (8):

(Everittc and Hothorn, 2011) (8)

If ξ*ij* = 1, it indicates that the separated signal *yi* differs from the original signal *sj* only in amplitude, meaning the separated signal is very pure. If ξ*ij* = 0, it indicates that the separated signal *yi* is completely different from the original signal *sj*, and the signal has not been separated. Therefore, the closer the value of ξ*ij* is to 1, the better the separation effect. To ensure the validity of the separation results, this paper classifies the data using ξ*ij* = 0.80 as the threshold.