Impacto Científico
Revista Arbitrada Venezolana
del Núcleo LUZ-Costa Oriental del Lago
Vol. 11. N°2. Diciembre 2016. pp. 137-154

Determinación de variables en problemas multivariantes. Método: SIMPLISMA

Eddy Rodríguez, Josefina Matera y Maribel Pérez

Universidad del Zulia. Facultad de Ingeniería. Centro de investigación de Matemática Aplicada (CIMA). Maracaibo. Venezuela.
eddyjackeline@yahoo.es

Resumen

En la actualidad los avances instrumentales, la automatización de procesos y el apoyo en sistemas informáticos de alto procesamiento traen como consecuencia la capacidad de generar grandes cantidades de datos, lo cual origina problemas multivariantes de gran escala, haciéndose necesario la utilización de herramientas matemáticas y estadísticas para extraer de estos datos la información relevante en la resolución de los problemas. En este trabajo se desarrolla el método SIMPLISMA como una respuesta a esta necesidad, este método evalúa conjuntos de datos correlacionados con el objetivo de encontrar variables puras dentro de los datos experimentales. La metodología empleada en esta investigación es teórica-práctica donde la revisión teórica parte de la consulta de autores especialistas tales como: Castillo M., Cavanillas S., Mardia K., Kent J.,Bibby J., Tauler R, Maeder M, De Juan A. Los resultados de los ejemplos demuestran estadísticamente que SIMPLISMA es un método eficaz en la obtención de variables puras.

Palabras clave: Método SIMPLISMA; herramientas matemáticas y estadísticas; datos correlacionados; variables puras.

RECIBIDO: 14/6/2016 ACEPTADO: 16/11/2016

Problem determination multivariate variables. Method: SIMPLISMA

Abstract

Today the instrumental advances, process automation and support systems high processing consequently they bring the ability to generate large amounts of data, which results Multivariate large-scale problems, making necessary the use of mathematical and statistical tools these data to extract the relevant information in solving problems. In this work the SIMPLISMA method is developed as a response to this need, this method evaluates data sets correlated with the aim of finding pure variables within experimental data. The methodology used in this research is theoretical and practical where the theoretical review of the consultation of experts authors such as M. Castillo, Cavanillas S., K. Mardia, Kent J., J. Bibby, Tauler R, Maeder M, John A. The results of the examples demonstrate statistically that SIMPLISMA is an effective method for obtaining pure variables.

Key words: SIMPLISMA method; mathematical and statistical tools; data correlated; variables pure.

Introducción

En la actualidad existen varios métodos de análisis multivariado que permiten reducir el número de variables observables en un problema de industria. En su mayoría estos métodos surgieron por la necesidad de solucionar situaciones específicas del momento como lo indican Cavanillas (2014) y Castillo (2007), extendiendo su uso a otras áreas similares.

El estudio de análisis multivariado puede aplicarse a matrices que contienen grandes dimensiones de datos experimentales con multitud de variables, en estos casos la atención puede estar dirigida al análisis de la variación de los datos o servir como paso previo en los procesos donde se discrimina y se selecciona la información antes de construir modelos de calibración o clasificación. En este trabajo se realiza el estudio de un método que permite seleccionar el mínimo de variables necesarias, el cual es: SIMPLISMA.

Al estudiar una matriz de datos, es posible que se encuentre correlaciones altas (en valor absoluto) entre varias variables, el caso más extremo es que una de las variables sea combinación lineal del resto, por esto es adecuado seleccionar un subconjunto de las variables originales o combinaciones lineales de estas. También pueden existir casos donde el número de variables sea tan grande que dificulta su análisis conjunto, y por tanto, es necesario reducirlas a un conjunto de menor dimensión que describa la

matriz de datos.

El objetivo de este trabajo es evidenciar la eficiencia del método SIMPLISMA en la determinación de las variables en problemas multivariantes, para ello se usa una metodología basada en primer lugar en la revisión bibliográfica apoyada en material de páginas web y en segundo lugar en la construcción del algoritmo; con esto se logra una descripción del método SIMPLISMA, mostrando el algoritmo de forma minuciosa e indicando los fundamentos matemáticos y estadísticos que lo sustentan.

Método SIMPLISMA

SIMPLISMA, corresponde a la expresión inglesa (SIMPLe-to-use Interactive Self-modeling Mixture Analysis), éste es un método intuitivo. Fue creado por Willem Winding en el año 1991, bajo un programa de MATLAB, por su naturaleza, pertenece a la familia de las técnicas de resolución de curvas de acuerdo a los autores Tauler y col. (2009) y Enrique (2006). La aplicación de este método permite seleccionar una muestra pequeña de la matriz de datos experimentales, que puede ser utilizada sin comprometer significativamente la información.

SIMPLISMA es usado principalmente en el área de química donde permite hallar variables puras en un conjunto de datos experimentales, la variable pura se define como una variable cuya intensidad se debe principalmente a uno de los componentes de la mezcla en estudio.

Este método tiene como objetivo seleccionar las variables puras eliminando la colinealidad entre las variables. La matriz, llamada matriz muestra, que se construye con SIMPLISMA se expresa como:

$$X_{(m,n)} = X_{p_{(m,k)}} + E \tag{1}$$

Donde X es la matriz de datos experimentales, X_p es la matriz de las variables puras y E es la matriz de los residuales.

Fundamento algebraico y estadístico del método SIMPLISMA.

El desarrollo del método SIMPLISMA involucra el manejo algebraico de matrices y de conceptos estadísticos. Lo cual conllevan a la escogencia de las variables puras que formaran la matriz muestra. Siguiendo los conceptos algebraicos y estadísticos de Mardia y col. (1979), tenemos la desviación estándar y la media de cada variable de la matriz de datos, ya que estas son usadas para determinar la desviación relativa que permite decidir cuál es la variable pura a escoger:

$$\mu_i = \frac{1}{m}\sum_{i=1}^{m} x_i \tag{2}$$

Donde μ_i es la media de la variable de la columna i

Que puede ser representada en forma matricial como:

$$\mu = \frac{1}{m} X^T 1 \tag{3}$$

Donde X es la matriz de datos de dimensión (m, n) y 1 es una matriz de n columnas de unos.

$$\sigma_i^2 = \frac{1}{m} \sum_{i=1}^{m} (x_i - \mu_i)^2 \tag{4}$$

Donde σ es la desviación estándar de cada columna i

Su representación matricial es:

$$\sigma^2 = \frac{1}{m} \left(X^T X - \frac{1}{m} X^T 1 1^T X \right) \tag{5}$$

Durante la selección de las variables se debe tener en cuenta que tienen que ser independientes unas de otras, para ello se aplica un estudio de correlación entre las variables mediante la siguiente formular:

$$r_{ij} = \frac{\sigma_{ij}}{(\sigma_i \sigma_j)} \tag{6}$$

Donde r_{ij} es el coeficiente de correlación entre las columnas i y j

Su forma matricial es:

$$R = D^{-1} \sigma^2 D^{-1} \tag{7}$$

Donde es la matriz de correlación y $D = diag(\sigma_i^2)$

Obtención de las variables puras

Uno de los parámetros más importantes en el análisis de las mezclas por métodos multivariantes es el número de componentes. Tanto un número insuficiente o excesivo de factores a menudo conduce a resultados incorrectos. En SIMPLISMA la decisión se hace mediante la comparación de los resultados de la etapa actual con los de la anterior y regresar de nuevo si es necesario.

El principal problema es ¿cómo encontrar las variables puras?. Se desarrolla el procedimiento siguiendo a Artyushkova, Fulghum (2001) y Enrique (2006). En este

método se parte de la existencia de una relación entre la desviación estándar y la media de una variable con la pureza de la misma:

$$p_i = \frac{\sigma_i}{\mu_i} \tag{8}$$

Donde p es la desviación estándar relativa, y (σ, μ) la desviación estándar y media de la columna i.

A diferencia de otros métodos que persiguen el mismo objetivo que SIMPLISMA, tales como, el método de análisis de componentes principales o el método de análisis de factores, no se debe centrar los datos, debido a que la escogencia de las componentes involucra al valor de la media en el denominador, como se muestra en la ecuación (8).

Puede aparecer un problema de exceso de alta pureza ocasionado cuando el valor medio tiende a cero (posiblemente por causa de ruido), lo que provoca que el cociente tienda a infinito, esto se soluciona agregando un valor en el denominador. La adición de esta pequeña constante (generalmente definido como un porcentaje de la media) soluciona el problema, suprimiendo el efecto de sobrestimación de la pureza, un valor de desplazamiento de 1% a 5% de la media es recomendable.

Redefiniendo la ecuación (8):

$$p_i = \frac{\sigma_i}{(\mu_i + \alpha)} \tag{9}$$

Una desviación estándar relativa grande indica una alta pureza para esa columna.

El procedimiento consiste primeramente en hallar la columna con la mayor desviación estándar relativa basado en la ecuación (9), y normalizar esta columna.

La selección de las siguientes variables, además de tener la mayor desviación relativa, serán las que tengan mínima correlación con las variables ya seleccionadas, para ello se calcula un factor de ponderación w:

$$w_i = \det\left(\frac{1}{m} Q_i^T Q_i\right) \tag{10}$$

Donde Q_i es la matriz de filas compuesta por las variables puras encontradas y cada i-ésima columna de la matriz de datos que aún no ha sido seleccionada.

El valor del determinante es proporcional a la independencia entre las variables puras halladas y la que está por seleccionarse. De esta manera el valor del determinante será mayor en la medida que las variables no estén correlacionadas, en caso contrario se acercara a cero.

Los elementos de la matriz Q_i se calculan de la siguiente manera:

$$Q_i = \frac{x_i}{\sqrt{\sigma_i^2+(\mu_i+\alpha)^2}} \tag{11}$$

Esta matriz representa la normalización de la matriz de estudio *X*, y tiene por fin dar a todas las variables la misma contribución durante el cálculo de los factores de ponderación. Para escoger la próxima variable pura se modifica la ecuación (9) multiplicándola por el factor de ponderación:

$$p_i = w_i \left(\frac{\sigma_i}{(\mu_i+\alpha)}\right) \tag{12}$$

Con ésta ecuación se selecciona la siguiente variable de mayor pureza, que corresponde al máximo p_i.

Algoritmo de SIMPLISMA

Se tiene como matriz de estudio a X_{mxn}, donde es el número de observaciones y *n* es el número de variables.

Procedimiento para la selección de la primera variable pura:

1. Se calculan los vectores media y desviación estándar de *X*:

$$\mu = \frac{1}{m}X^T 1 \qquad \sigma = \sqrt{\frac{1}{m}\left(X^T X - \frac{1}{m}X^T 1 1^T X\right)}$$

2. Se calcula el valor de α:

$$\alpha = 0.01 * \mu_{max}$$

μ_{max} Es el mayor valor de μ

3.Se calcula el vector de la desviación estándar relativa:

$$\mathrm{p} = \frac{\sigma}{(\mu + \alpha)}$$

El máximo valor de *p*, corresponde a la primera variable pura.

Calculo de la matriz de correlación y factor de ponderación:

4. Se calcula la matriz *Q:*

$$Q = \frac{X}{\sqrt{\sigma^2 + (\mu + \alpha)^2}}$$

5. Se calcula la matriz de correlación M_c:

$$M_c = \frac{1}{m} Q^T Q$$

6. Se calcula el factor de peso:

$$w_i = \det\left(M_{c_i}^{\ T} M_{c_i}\right)$$

Donde M_{ci} es la matriz formada por las variables puras encontradas y cada i-ésima columna de la matriz de datos que aún no ha sido seleccionada.

Selección de la segunda variable pura:

7. Se calcula la desviación estándar relativa modificada:

$$p_i = w_i \left(\frac{\sigma_i}{(\mu_i + \alpha)}\right)$$

Donde el subíndice *i*, representa al número de la columna que corresponde a la variable.

La segunda variable a seleccionar será la que tenga mayor *p*.

Para la escogencia de las siguientes variables puras se repiten los pasos 6 y 7 y, se detiene el proceso cuando se ha cumplido con un criterio de parada, dentro de los indicados en la siguiente sección, pre-establecido.

Elección del número de variables

En cada aplicación según Castillo (2007) y Rey (2009), debe ser tomada una decisión acerca de cuantas componentes se deben conservar para resumir eficazmente los datos, esta decisión se basa en el conocimiento técnico de la aplicación acompañado por los siguientes lineamientos que pueden servir de guía:

- Varianza explicada: seleccionar el número de componentes necesarias para explicar una proporción determinada de la varianza por ejemplo un rango de 80% a 90%.

El desafío consiste en la selección de un porcentaje de umbral apropiado. Por ejemplo hay que tener presente que una componente no puede generalizar a la población u otras muestras, una componente está dominada por una sola variable y no representa un resumen compuesto de varias variables.

- Mínimo valor propio: se puede seleccionar las componentes asociadas a valores propios superior a un valor prefijado, por ejemplo la varianza media:

$$varianza\ media = \sum_{i=1}^{k} \frac{\lambda_i}{k}$$

Puede usarse donde hay una gran diferencia entre los dos valores propios que caen a ambos lados de la media, en estos casos los datos pueden ser resumidos con éxito en un número relativamente pequeño de dimensiones.

- Representación gráfica: Para el caso de SIMPLISMA se puede graficar la raíz cuadrada de los errores medios predictivos, si el método forma parte de un proceso de modelado, o se puede graficar las desviaciones medias relativas si el método es usado para seleccionar variables.
- Realizar pruebas sobre la más grande componente, aplicar pruebas de significancia a la componente, asociada al valor propio más grande, o a la variable pura.

Puede ser útil hacer una prueba preliminar de la independencia de las variables, si los resultados indican que las variables son independientes, no hay ninguna necesidad en extraer los componentes, ya que (a excepción de la fluctuación de muestreo) las variables en sí mismas forman los componentes.

Para probar la significancia de los componentes "más grandes", se puede probar la hipótesis de que los últimos valores propios de la población son pequeños e iguales, caso de métodos que calculan valores propios, o se puede probar la hipótesis de que las últimas desviaciones estándar relativas son pequeñas e iguales, caso método SIMPLISMA. La implicación es que los primeros componentes de la muestra capturan todas las dimensiones esenciales, mientras que los últimos componentes reflejan ruido.

La elección de los componentes debe hacerse con cuidado ya que puede darse el caso que los componentes más pequeños pueden llevar información valiosa que no debe ser ignorado rutinariamente.

Ejemplos de estudio

En esta sección se realiza el estudio y análisis de resultado de dos ejemplos reales, tomados de Matlab R2012a. Estos ejemplos son problemas multivariantes, donde con la ayuda del método SIMPLISMA se determinará las variables puras o componentes principales. Para ello se trabaja con las variables de entrada en cada problema. Estos ejemplos fueron seleccionados debido a sus características de colinealidad y diversidad de tamaño, el primer problema obtenido de Woods y col. (1932), corresponde a pocas variables de estudio, tan sólo cuatro y, con un numero de observaciones mayor al número de variables, con alta presencia de correlación.

El segundo problema proveniente de Kalivas (1997), está formado por un gran número de variables, 401 en total, y con pocas observaciones en comparación con el número de variables, en este ejemplo, las variables presentan diferentes porcentajes de correlación. Con estos ejemplos se evidencia la efectividad del método para diferentes estilos de problemas.

- Datos del ejemplo 1:

Cemento Portland de datos
Ingredientes (%):
columna1: 3CaO.Al2O3 (tricalcium aluminate)
columna2: 3CaO.SiO2 (tricalcium silicate)
columna3: 4CaO.Al2O3.Fe2O3 (tetracalcium aluminoferrite)
columna4: 2CaO.SiO2 (beta-dicalcium silicate)
Calor (cal / g): el calor de endurecimiento después de 180 días.

En estos datos la variable de salida o respuesta está formada por el vector Calor de 13 muestras, las variables de entrada corresponde a la matriz ingredientes, que es una matriz de 13 muestras y 4 variables. Se aplicara el método SIMPLISMA a la matriz ingredientes para seleccionar las variables puras.

- Datos del ejemplo 2:

Los datos espectrales y octanaje de la gasolina:
Espectros NIR y los números de octano de 60 muestras de la gasolina
NIR: espectros NIR, medido en intervalos de 2 nm de 900 nm a 1700 nm
Números de octano: octanos

En estos datos la variable de salida o respuesta es el vector números de octano, que contiene 60 muestras y las variables de entrada corresponden a la matriz NIR, que está compuesta de 401 variables y 60 muestras. A esta matriz NIR, se le aplicara el método SIMPLISMA para seleccionar las variables puras.

Análisis de los resultados

El estudio de los resultados se centra en diferentes pruebas que permitirán seleccionar el número óptimo de variables, las cuales son:

- Estimación del número de variables por valor propio mínimo.
- Análisis comparativo de las desviaciones estándar relativas, presentadas mediante gráficas.
- Análisis de la correlación de las variables en cada paso graficando el factor de ponderación, con lo que se indica si es necesario la entrada de una nueva variable.
- Estudio del porcentaje de varianza explicada, mediante pruebas de hipótesis.

Con la primera se tiene un estimado del posible número de variables puras, con la segunda y la tercera se determinan las variables puras y con la última se comprueba si las variables seleccionas son suficientes.

Ejemplo 1:

Se presenta primero la matriz de correlación de la data original:

	V1	V2	V3	V4
V1	1.0000	0.2286	-0.8241	-0.2454
V2	0.2286	1.0000	-0.1392	-0.9730
V3	-0.8241	-0.1392	1.0000	0.0295
V4	-0.2454	-0.9730	0.0295	1.0000

En esta matriz se observa que la primera variable tiene alta correlación con la tercera variable (-0.8241) y, la segunda variable con la cuarta variable (-0.9730), con lo cual se afirma que no todas las variables del problema son puras.

Se determina los valores propios de la matriz de covarianza:

Valores propios = [0.2372 12.4054 67.4964 517.7969]

Se obtiene el promedio de la varianza, que es: 149.4840, siguiendo lo indicado en la sección 2, se puede estimar un número de variables a seleccionar de uno.

Al comparar las desviaciones estándar relativas (p), ver figuras 1 y 2, se observa que la primera variable a seleccionar es la uno con una desviación estándar relativa de 0.6, seguida de la variable cuatro con una desviación estándar relativa de 0.18.

Con un análisis de las figuras 3 y 4, donde se grafica cada variable versus el factor de ponderación, se observa por ejemplo, en la figura 3, que el valor de para las variables que no han sido seleccionadas es superior a 0.2, de manera que es factible considerar que existen variables puras aun no seleccionadas. Con respecto a la figura 4, los valores de no llegan a 0.1 lo que permite confirmar junto con lo expuesto en la matriz de correlación, que las variables uno y cuatro ya seleccionadas forman las variables puras del ejemplo en estudio.

Además, de acuerdo a lo establecido en la sección 2, indica la independencia entre las variables puras halladas y las que están por seleccionarse y, cuando este se acerca a cero significa que hay correlación entre las variables.

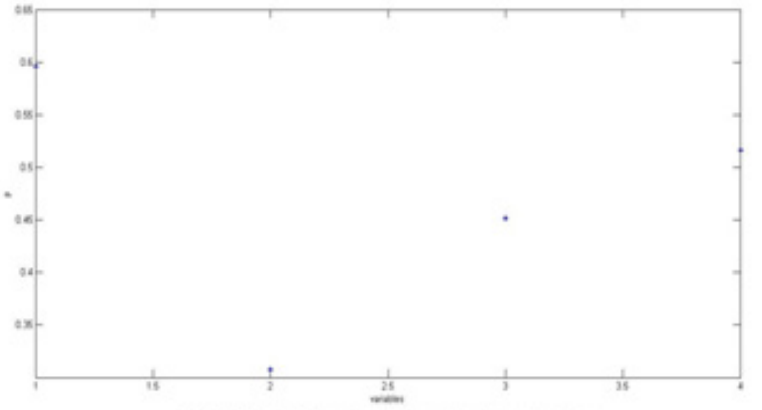

Figura 1. Ejemplo 1. Valores de la desviación estandar relativa para seleccionar la primera variable más pura.

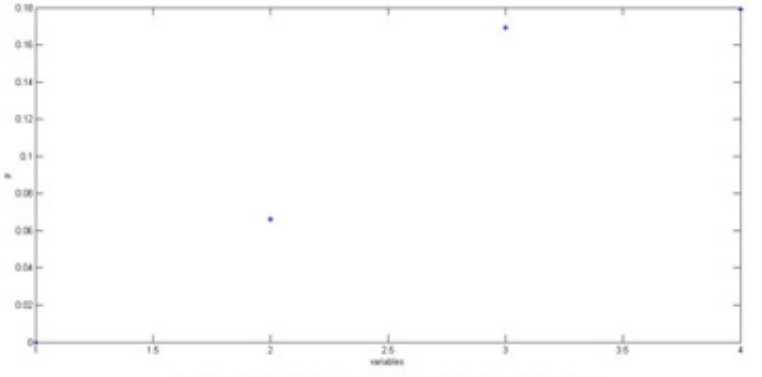

Figura 2. Ejemplo 1. Valores de la desviación estándar relativa para seleccionar la segunda variable más pura.

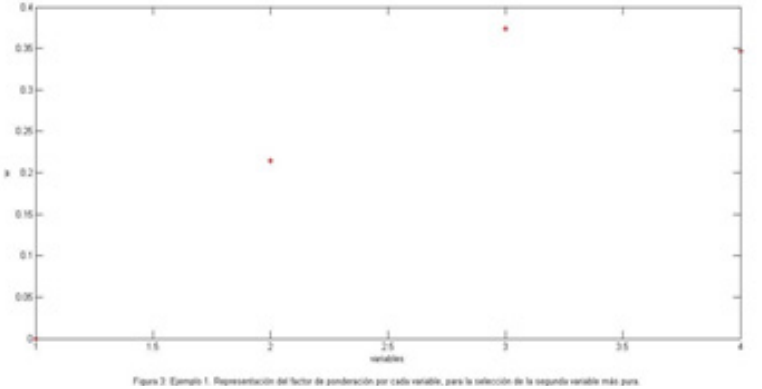

Figura 3. Ejemplo 1. Representación del factor de ponderación por cada variable, para la selección de la segunda variable más pura.

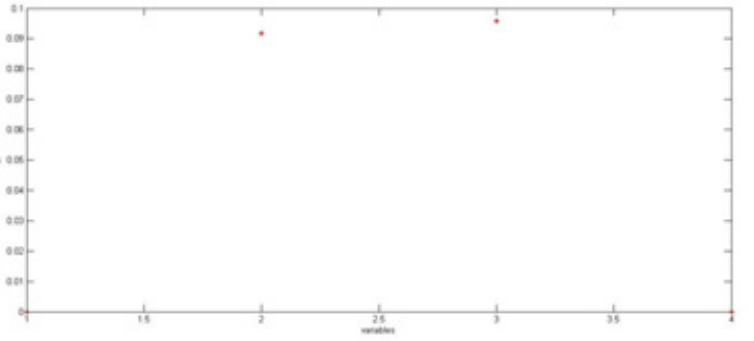

Figura 4. Ejemplo 1. Representación del factor de ponderación por cada variable, para la selección de la tercera variable más pura.

Aplicando prueba de hipótesis como en Mardia y col. (1979) se puede establecer si la varianza aportada por el número de variables puras seleccionadas explica una proporción significativa de la varianza total, o si por el contrario no son suficientes las variables seleccionadas. Para ello, considérese:

$$\theta = \frac{\lambda_1 + \lambda_2}{\lambda_1 + \cdots + \lambda_4}$$

Donde λ_i representa la varianza aportada por cada variable y, θ es la varianza explicada con las variables puras seleccionadas.

Se analiza la hipótesis nula de que la varianza explicada sea igual a 0.9, contra la hipótesis alterna de que la varianza explicada sea menos que 0.9.

El valor calculado de θ = 0.9789

Con una desviación estándar de σ = 0.0112

Con un nivel de significancia de α = 0.05 y un $z_{0.05}$ = -1.645

Da un $z_{Calculado}$ = 7.0366

Como $z_{Calculado}$ = 7.0366 > $z_{0.05}$= -1.645, entonces, no se puede rechazar la hipótesis nula.

Con un intervalo de confianza de 95% se puede afirmar que las variables puras uno y cuatro explican entre el 95.69% y el 100% de la varianza de la población.

Ejemplo 2:

Con un análisis previo de la matriz de correlación se determina que existen variables que tienen 99% de correlación, de esta manera se puede afirmar que no todas las variables son puras.

El promedio de la varianza es: 1.5175e-04.

Los valores propios de la matriz de covarianza que son mayores a 1.5175e-04, son:

Valores propios = [0.0002 0.0002 0.0003 0.0006 0.0008 0.0028 0.0042 0.0069 0.0442]

Con lo cual se estima que el número de variables puras es de nueve.

Analizando las figuras 5, 6, 7, 8 y 9, y la tabla 1, se observa que la desviación estándar relativa tiene su mayor valor en la variable ochenta y uno, en la primera selección, seguida de las variables diecinueve, ciento cinco, cincuenta y ochenta, siendo el valor de p=0.0002 para la última variable seleccionada muy próxima a cero, lo que nos indica que el proceso puede detenerse con este número de variables puras seleccionadas.

A través de las figuras 10, 11, 12 y 13 se observan las variables versus el factor de ponderación, donde se evidencia que a partir de la quinta variable pura, es cercano a cero, de manera que la correlación entre las variables seleccionadas y las que faltan por entrar es significativo y como tal, las no seleccionadas no son variables puras.

Con un estudio de prueba de hipótesis se puede verificar si las variables puras seleccionadas explican satisfactoriamente la varianza del problema. Para esto considérese el mismo parámetro θ estudiado en el ejemplo anterior y como hipótesis nula que el valor de la varianza explicada por las variables puras seleccionadas sea de 0.9, contra la hipótesis alternativa de que éste valor sea menor a 0.9.

El valor calculado de θ=0.9670

Con una desviación estándar de σ = 0.0019

Con un nivel de significancia de α = 0.05 y un $z_{0.05}$ = -1.645

Da un $z_{Calculado}$ = 34.7623

Como $z_{Calculado}$ = 34.7623 > $z_{0.05}$= -1.645, no se puede rechazar la hipótesis nula.

Con un intervalo de confianza de 95% se puede afirmar que las variables puras ochenta y uno, diecinueve, ciento cinco, cincuenta y ochenta explican entre el 96.32% y el 97.08% de la varianza de la población.

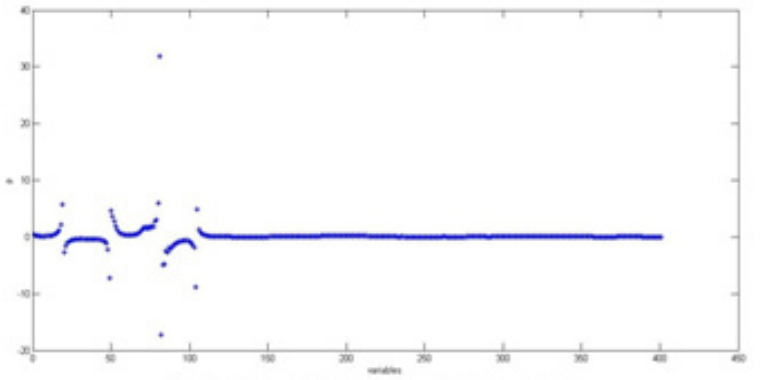

Figura 5. Ejemplo 2. Valores de la desviación estándar relativa para seleccionar la primera variable más pura.

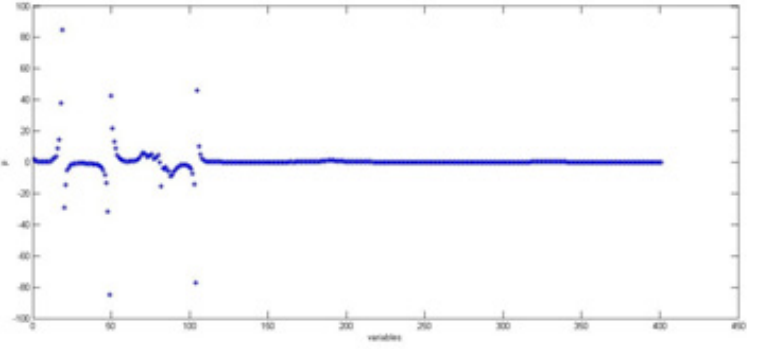

Figura 6. Ejemplo 2. Valores de la desviación estándar relativa para seleccionar la segunda variable más pura.

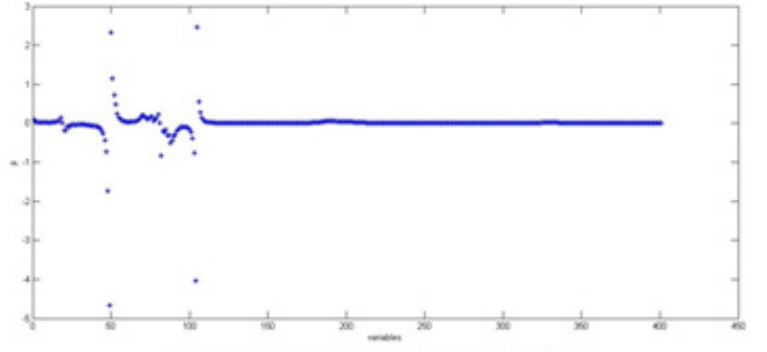

Figura 7. Ejemplo 2. Valores de la desviación estándar relativa para seleccionar la tercera variable más pura.

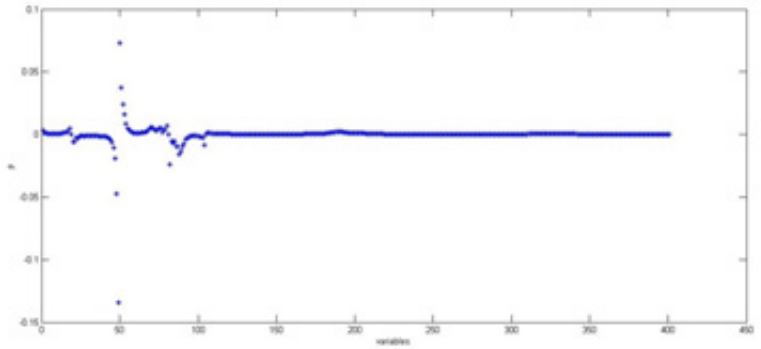

Figura 8. Ejemplo 2. Valores de la desviación estándar relativa para seleccionar la cuarta variable más pura.

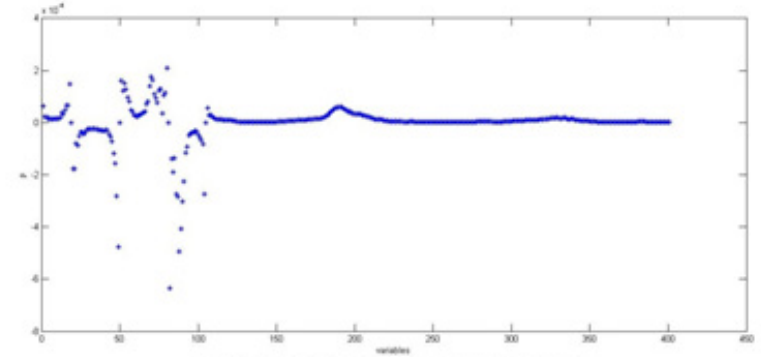

Figura 9. Ejemplo 2. Valores de la desviación estándar relativa para seleccionar la quinta variable más pura.

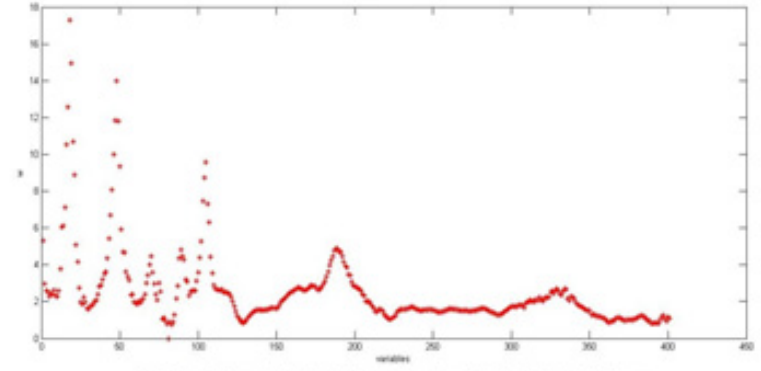

Figura 10. Ejemplo 2. Representación del factor de ponderación por cada variable, para la selección de la segunda variable más pura.

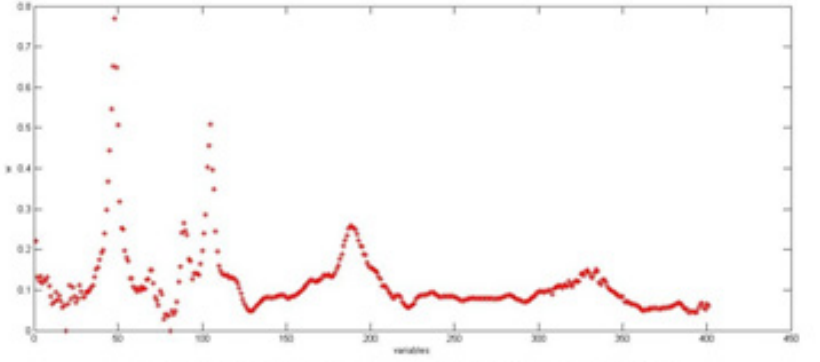

Figura 11: Ejemplo 2. Representación del factor de ponderación por cada variable, para la selección de la tercera variable más pura.

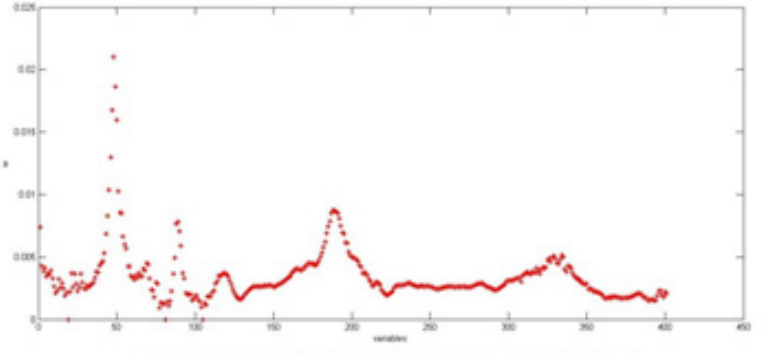

Figura 12: Ejemplo 2. Representación del factor de ponderación por cada variable, para la selección de la cuarta variable más pura.

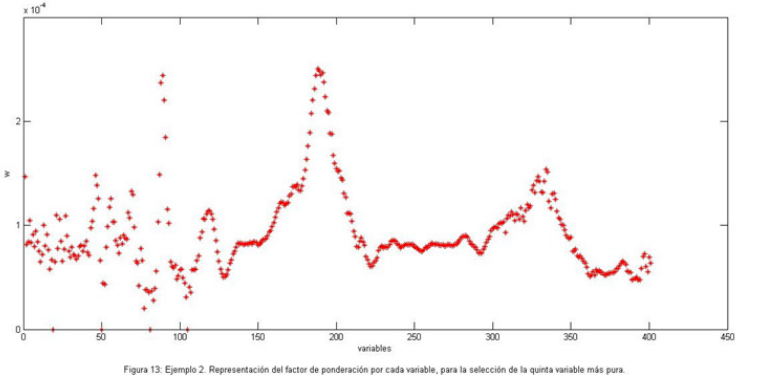

Figura 13: Ejemplo 2. Representación del factor de ponderación por cada variable, para la selección de la quinta variable más pura.

Conclusiones

- Los métodos explicados en esta investigación permiten que la matriz reducida conserven las características de variabilidad e independencia de la matriz de datos experimentales.
- Es viable la implementación del algoritmo de SIMPLISMA en computadora, debido a que el método se fundamenta en estructuras algebraicas de matrices.
- La elección del número de variables debe hacerse tomando en consideración el riesgo de la perdida de información posible luego de la reducción de la matriz.
- A través de los ejemplos se constató la eficiencia del método SIMPLISMA para la determinación de las variables puras, concluyéndose a partir del análisis de los resultados que con la selección de las variables puras se explica más del 90% de la varianza de los datos de los problemas.

Referencias bibliográficas

Artyushkova K. Fulghum J (2001). Identification of chemical components in XPS spectra and images using multivariate statistical analysis methods. Journal of Electron Spectroscopy and Related Phenomena. Vol. 121, 33-55. (Documento en línea). Disponible en: http://www.unm.edu/~kartyush/articles/8jesrp8_OK.pdf

Castillo M. (2007) Aplicación de la Espectroscopia NIR al Control Analítico de Procesos de la Industria Química. Tesis Doctoral. Universidad Autónoma de Barcelona. En: http://grupsderecerca.uab.cat/chemometrics/sites/grupsderecerca.uab.cat.chemometrics/files/Castillo_thesis.pdf

Cavanillas S. (2014). Desarrollo de metodologías y herramientas quimiométricas para el tratamiento de datos electroquímicos no lineales. Aplicación a sistemas de interés biológico y medioambiental. Tesis Doctoral. Universitat de Barcelona. p.p. 282. En: http://www.tesisenxarxa.net/handle/10803/285264

Enrique M. (2006). Seguimiento Cuantitativo de Reacciones de Resinas Epoxi mediante Espectroscopia de Infrarrojo Cercano y Métodos de Resolución de Curvas. Tesis Doctoral. Universitat Rovira I Virgili. p.p. 309. (Documento en línea). Disponible en: http://tesisenred.net/handle/10803/9009

Kalivas, John H., "Two Data Sets of Near Infrared Spectra," Chemometrics and Intelligent Laboratory Systems, v.37 (1997) p.p. 255–259. Fuente: Matlab R2012a.

Mardia K., Kent J.,Bibby J. (1979). **Multivariate analysis.** Academic Press. p.p. 551

Rey E. (2009). Estudio de Mezclas por Resonancia Magnética Nuclear. Tesis para de Magister en Ciencias Química. Universidad Nacional de Colombia. En: http://www.yo-que.ch/nmrlab/mediawiki-1.15.1/images/d/do/Tesis-ERRC.pdf

Tauler R, Maeder M, De Juan A. (2009). Multiset Data Analysis. Extended Multivariate Curve Resolution. Elsevier. B. V. All rights reserved. (Documento en linea). En: http://www.iasbs.ac.ir/chemistry/chemometrics/history/11th/multiset%20data%20analysis.pdf

Woods,H., H. Steinour, H. Starke, "Effect of Composition of Portland Cement on Heat Evolved during Hardening," Industrial and Engineering Chemistry, v.24 no.11 (1932), p.p.1207-1214. Fuente: Matlab R2012a.

UNIVERSIDAD DEL ZULIA

Impacto Científico

Revista Arbitrada Venezolana
del Núcleo LUZ-Costa Oriental del Lago

Vol. 11. N°2_____

Esta revista fue editada en formato digital y publicada en diciembre de 2016, por el ***Fondo Editorial Serbiluz, Universidad del Zulia. Maracaibo-Venezuela***

www.luz.edu.ve
www.serbi.luz.edu.ve
produccioncientifica.luz.edu.ve