

APLICACIÓN DE R EN SPACE FILLING DESIGN Y VISUALIZACIÓN DE DATOS EN INGENIERÍA QUÍMICA

Oscar Gonzalez, Facultad de Química, UNAM, Mexico

Elías Martínez, Instituto Mexicano del Petróleo, México

Palabras clave: Latin Sampling Hypercube, DiceDesign Package, Visualización

Abstract: Este trabajo presenta la aplicación de R-Studio en el diseño de experimentos de computadora, así como la visualización de datos para un proceso de separación de biocombustibles en torres de destilación. Este es un problema típico de la ingeniería química. El rellenado de espacios sirve para genera puntos de entrada que se utilizarán para simular el proceso y así obtener una serie de datos de salida para posteriormente generar un modelo de red neuronal artificial. Para generar una serie de datos de entrada se utilizó la paquetería de DiceDesign Package y que contiene el método de Latin Sampling Hypercube, el cual facilita el procedimiento dentro del rango establecido de cada variable. Además, debido a que se cuenta con demasiadas variables de salida se realizó una visualización y análisis de de datos mediante histogramas y gráficas de correlación por medio de R-Studio y así obtener una visión más clara de las variables que afectan el desempeño de las torres de destilación.

Uso de DiceDesign Package en R-Studio

Método de Latin Sampling Hypercube

El DiceDesign Package en R-Studio se describe como un paquete de Diseño de rellenos de espacio (en inglés Space-Filling Designs “SFD”) y criterio de relleno de espacios (basado en la distancia y en la uniformidad), por lo que uno de los métodos que utiliza esta paquetería es la de Latin Sampling Hypercube la cual consiste en ser un método estadístico que genera una muestra de valores aleatorios a partir de distribuciones multivariadas.

Un ejemplo claro de utilizar el método de Latin Sampling Hypercube de DiceDesign Package es en el proceso químico de dos torres de destilación, como se muestra en la figura 1, las cuales van a tener como producto Naphtha, Diesel Verde y bioturbosina.

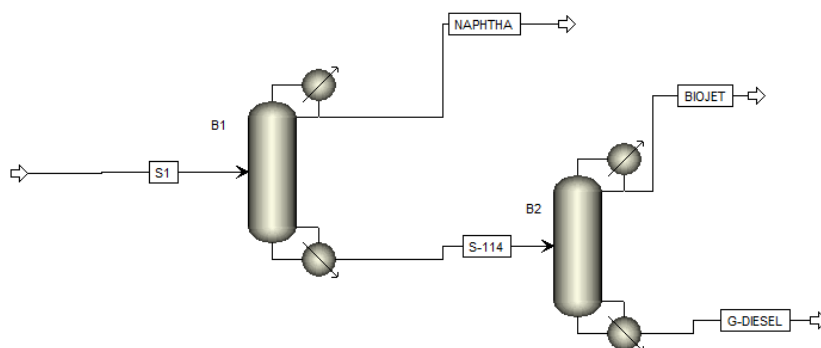


Figura 1. Torres de destilación del proceso.

Debido a que se necesita optimizar este proceso se estudian las variables de entrada y del sistema que pueden afectar a las variables de salida. Se encontró que son en total 7 variables que influyen en los resultados. Se determinaron los rangos en que pueden trabajar las 7 variables sin afectar al proceso para posteriormente utilizar el método de Latin hypercube sampling, el cual consiste generar una serie de datos aleatorios dentro de los rangos de cada variable. Por lo que es una herramienta muy eficaz debido a que se puede generar series de

datos en cuestión de segundos y así poder generar una base de datos para cualquier proceso químico, en este caso dos torres de destilación. Una vez generados los puntos, se llevaron a cabo simulaciones del proceso para generar datos de salida. Entonces se analizó la distribución de cada variable usando visualización de datos.

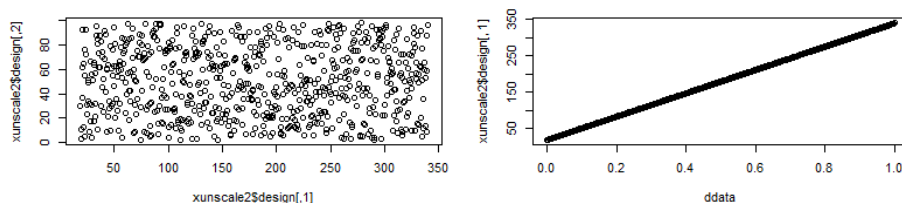


Figura 2. Ejemplo de distribución de datos para una variable del proceso químico en R-Studio

Visualización y análisis de datos

Histogramas y correlaciones de datos

Se necesita visualizar y analizar las variables de salida del proceso químico de las dos torres de destilación, pero al ser demasiadas variables (147 variables de salida), se utilizó un código de programación en R-Studio con las paqueterías de ggplot2, caret, dplyr y tidyr, las cuales tienen como función crear una visualización ordenada y precisa de los datos que uno desea analizar, por lo que al realizar los histogramas y las correlaciones de las variables por medio de estas paqueterías de R-Studio se obtuvo una gran ventaja, la cual fue que se ahorra demasiado tiempo en realizar las gráficas debido a que en una misma corrida del código se puede dar todas las gráficas necesarias, aunque para mejor visualización de las gráficas se recomienda dividir las variables en grupos y realizar varias corridas del código, también otras ventajas fueron que ayudo a poder observar con más claridad los valores fuera de rango o tendencia de cada variable además de cuáles eran las variables con correlación fuerte o débil entre sí y así poder seleccionar las variables necesarias para el modelo que se desea estudiar del proceso químico.

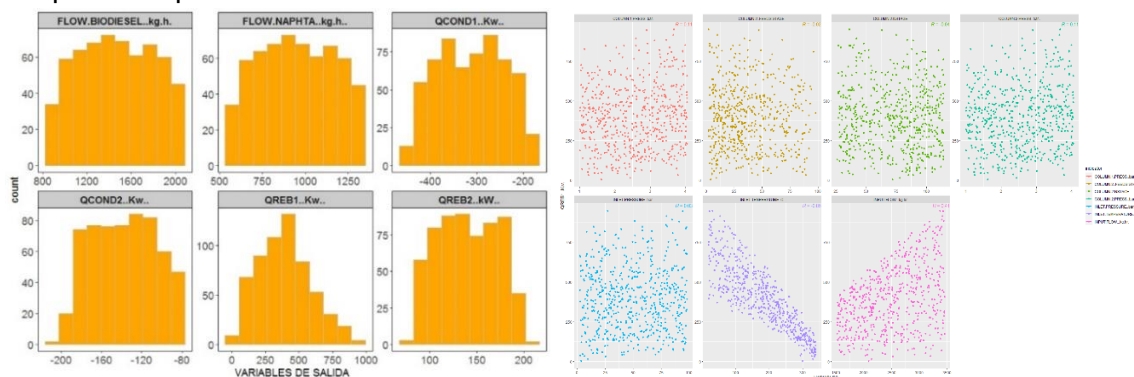


Figura 3. Ejemplo de histograma y correlación que se realizó en R-Studio

Conclusiones

La herramienta de R-Studio tiene grandes ventajas al usarla para aplicaciones en procesos de la industria química, debido a que cuenta con varias paqueterías que pueden ayudar a realizar y visualizar graficas en cuestión de segundos sin necesidad de realizar cada grafica por separado.

Una función de gran importancia de R-Studio es la de Dice Design Package ya que puede generar series de datos sin afectar el rango establecido de cada variable que se desea estudiar.