Búsqueda de condiciones óptimas de producción en industria química

Víctor Martínez Luaces, Mauricio Ohanian

Núcleo Interdisciplinario de Ingeniería Electroquímica, Espacio Interdisciplinario.
Universidad de la Republica, Montevideo, Uruguay
victorml@fing.edu.uy, mohanian@fing.edu.uy

Palabras clave: diseño experimental, optimización multirespuesta

Resumen

Se trata de optimizar dos características de un producto químico de fabricación industrial. El producto se presenta comercialmente en comprimidos y son características determinantes de su calidad la velocidad de disolución y su dureza superficial. Particularmente es relevante la disolución presentada a la hora y a las 8 horas, lo que puede ser interpretado como una disolución "inicial" y otra "remanente". Simultáneamente, se trata de obtener un buen valor de dureza para los comprimidos, básicamente por un tema de presentación del producto. Por motivos de confidencialidad, no se brindan datos sobre el producto estudiado, ni acerca de las variables consideradas en los experimentos.

Para el objetivo propuesto se preparó un diseño secuencial en 3 etapas distintas:

1. Eliminación de variables de escasa relevancia.
2. Determinación de las variables principales.
3. Optimización multi-respuesta de las variables principales.

Para la primera etapa se utiliza un diseño de Plackett-Burman (P-B) en el cual se consideraron 11 variables, de las cuales 3 resultaron eliminadas. En función de lo anterior, para la segunda etapa, se llevó a la práctica un Diseño Factorial Fraccionario del tipo 2_{IV}^{8-4} , resultando un total de 4 variables principales a optimizar. Finalmente, en la tercera y última etapa, se procedió a la optimización, utilizando para ello la Metodología de Superficies de Respuesta (MSR).

Como resultado final de este proceso, resultan intervalos de trabajo sugeridos para las variables principales y que a su vez son valores experimentalmente adecuados para la optimización basada en criterios de solubilidad y dureza en forma simultanea.

Introducción

El presente trabajo se origina en un asesoramiento para una industria química, que preparaba a escala piloto un nuevo producto, antes que el mismo fuera industrializado y lanzado al mercado. Por razones de confidencialidad, no resulta posible dar mas información del proceso considerado.

Materiales y métodos

Para comenzar esta sección, conviene comentar que previo a cada etapa del diseño, se procedió a al ajuste del Modelo Lineal correspondiente, mediante la utilización de un test estadístico de falta de ajuste (conocido con la sigla LOF, por su nombre en inglés: *Lack of Fit*). La aplicación de esta prueba estadística, permitió concluir que en el rango estudiado, la concentración del producto vs. tiempo no seguía un comportamiento lineal en la variable (i.e., un Modelo Lineal Simple – MLS), por lo que se recurrió, como es habitual en estos casos a un modelo cuadrático en la variable y lineal en los parámetros (i.e., un tipo especial de Modelo Lineal General – MLG) . Finalmente, sobre la base del MLG validado por el test de LOF, se realizaron predicciones puntuales de disolución para tiempos de 1 hora y de 8 horas, respectivamente. Las mencionadas predicciones puntuales, a su vez, fueron los insumos utilizados en las diferentes instancias de tratamiento estadístico de datos correspondientes a las tres etapas del diseño: eliminación de variables de escasa relevancia, determinación de las variables principales y optimización multi-respuesta de las variables principales.

1) Primera etapa: Eliminación de Variables No Significativas.

1.a) Diseño de Plackett-Burman.

Teniendo en cuenta las sugerencias del técnicos de la industria, los recursos disponibles y las propias limitaciones de los diseños aplicables a este fin, se decidió tomar en cuenta un total de 11 variables. En tal sentido, es conveniente mencionar que los diseños de P-B, requieren un número de corridas igual al número de

variables más uno y que a su vez, dicho número de corridas debe ser múltiplo de cuatro [1]. En función de lo anterior, las opciones más razonables eran 7 variables y 8 corridas u 11 variables y 12 corridas. Luego de discutir la viabilidad y conveniencia de ambas opciones, se resolvió utilizar el último de los mencionados de 12 corridas por duplicado. Concretamente, se eligió el Diseño de P-B propuesto en bibliografía [1].

1.b) Comprobación de ajuste y predicción.

Los datos obtenidos en esta primera etapa se analizan a través de un MLS (i.e., se trata de ajustar una recta), pero los resultados en casi todas las corridas llevan al rechazo del modelo propuesto al nivel de significación usual del 0.05.

Teniendo en cuenta lo anterior, se pasa a considerar un modelo cuadrático (i.e., se ajusta una parábola, utilizando la metodología del MLG) y en este último caso el ajuste del modelo resultó satisfactorio, excepto en unas pocas corridas que fueron repetidas hasta lograr el ajuste deseado. Como ya se comentó, dichos ajustes fueron probados utilizando el test de LOF [2], a nivel de significación de 0.05 como es habitual en estos casos.

Una vez comprobado el ajuste de los modelos anteriormente mencionados (es decir, los 24 MLG cuadráticos con el test de LOF), se hicieron predicciones puntuales a la hora y a las 8 horas de disolución.

1.c) Tratamiento de datos.

Utilizando los valores interpolados en la sección 1.b, se procede a calcular los valores críticos de la distribución *t* [3] a fin de determinar qué variables resultaban poco significativas para el estudio posterior. El resultado fue que se eliminaron las tres variables, las cuales no son tomadas en cuenta para los sucesivos estudios.

2) <u>Segunda etapa: Determinación de variables principales</u>.

2.a) Elección del Diseño Factorial Fraccionario utilizado.

En esta etapa se tiene en cuenta las 8 variables remanentes codificadas con las letras *A, B, C, D, E, F, G* y *H* y los recursos disponibles, procurándose no utilizar más de un 50 % de los mismos en esta instancia. Por este motivo, se decide utilizar un diseño del tipo 2_{IV}^{8-4}, el cual permite con solamente 16 corridas (y otras tantas en la réplica) estudiar las variables en dos niveles con sus correspondientes interacciones [1]. En este tipo de diseño, existen patrones de confusión que podrían dificultar la identificación de algunas interacciones de variables. Sin embargo, poseen la ventaja de que no introducen confusión entre variables principales e interacciones de segundo orden, como surge de la bibliografía [4].

2.b) Comprobación de ajuste y predicción de los modelos.

Teniendo en cuenta lo sucedido en la etapa 1.b, los datos en esta oportunidad no fueron analizados a través de un MLS y se pasó a considerar directamente el modelo cuadrático, el que resulta satisfactorio, excepto en unas pocas corridas, que fueron repetidas hasta lograr el ajuste deseado, ya que en esta etapa se busca un modelo confiable de ajuste. Los mismos fueron analizados utilizando el test de LOF al nivel de significación del 0.05.

Una vez validados los modelos anteriormente mencionados (es decir, los 32 MLG cuadráticos), se hicieron predicciones puntuales a la hora y a las 8 horas de la disolución en cada uno de los 32 casos, los que a su vez resultaron utilizados como insumos informativos para el tratamiento de datos de la etapa.

2.c) Tratamiento de datos.

El tratamiento estadístico se realiza a través de varios cuadros de Análisis de Varianza (ANOVA [1]), correspondientes a original y réplica, para la hora y para las 8 horas. Del análisis de dichos cuadros de ANOVA resultan significativas las variables *B, D* y *H*.

Como complemento de lo anterior, se realiza un cuadro de ANOVA, en planilla Excel®, utilizando como variable de respuesta el promedio de disolución. Esto se debe a que los programas comerciales estadísticos acostumbran a incluir las interacciones dentro del error, sin desglosarlas, como correspondería a este tipo de diseño en que las interacciones no se confunden con los efectos principales.

De este segundo análisis, resulta significativa la interacción *CD*, lo que justifica agregar la variable C a la lista de variables principales, con las que se construye el diseño de la etapa siguiente.

3) Tercera etapa: Optimización de Variables Principales.

3.a) Diseño Experimental utilizando MSR

Tal como se expresó anteriormente, en esta etapa se trabajó únicamente con las variables *B, C, D* y *H*. Dichas variables son codificadas y renombradas como X_1, X_2, X_3 y X_4, las cuales se optimizan recurriendo a un diseño octogonal con punto central replicado. Se elige este diseño, ya que el mismo permite estudiar efectos de curvatura y de interacción entre las variables, con un número razonablemente pequeño de corridas. En efecto, este tipo de diseño plantea utilizar 5 niveles distintos de las variables (codificados como –2, –1, 0, +1 y +2), combinados de modo tal que posibilita la obtención de una superficie correspondiente a una función polinómica de grado 2 en cuatro variables. Para ello, el polinomio utilizado debe tener un término independiente, cuatro monomios lineales, cuatro monomios cuadráticos de una sola variable (uno para cada variable) y seis monomios cuadráticos de interacción (correspondientes a las interacciones *BC, BD, BH, CD, CH* y *DH*).

3.b) Comprobación del ajuste y Predicción a la hora y las 8 horas.

Una vez más se recurrió a la metodología de MLG cuadrático en variables y lineal en coeficientes para modelar la disolución vs. tiempo. También se utiliza nuevamente el Test de LOF para comprobar el ajuste de los modelos, antes de efectuar las predicciones puntuales como en las etapas anteriores. En esta ocasión no hubo necesidad de efectuar repeticiones de corridas, ya que todas presentaron un buen ajuste sin mayores inconvenientes.

Nuevamente, los resultados de este análisis constituyen los insumos informativos del tratamiento de datos de esta tercera y última etapa. Los mismos se presentan como predicciones puntuales para la hora y para las 8 horas.

3.c) Tratamiento de datos.

Esta etapa es la más complicada desde el punto de vista del tratamiento estadístico de los datos. En efecto, los programas disponibles trabajan con modelos polinómicos (y otros de uso habitual) para una sola variable y/o con modelos lineales de primer grado en varias variables. Es decir que, para un caso como el aquí estudiado, en el cual se trabaja con 4 variables (X_1, X_2, X_3 y X_4), sólo es posible ajustar un hiperplano, lo que no permite determinar curvaturas ni interacciones; para ser más explícitos, sólo es posible ajustar un polinomio con término independiente y cuatro monomios lineales, sin poder agregar otros elementos al modelo.

En función de lo anterior, se prepara un programa *ad hoc*, en el que se realiza la regresión polinómica antedicha con 15 coeficientes: uno independiente, 4 para los monomios lineales, 4 para los cuadráticos uni-variados y 6 para las interacciones de dos variables. El resultado de la mencionada regresión polinómica tiene la forma de 15 coeficientes para la primera regresión (a la hora) y otros 15 para la regresión correspondiente a las 8 horas.

Utilizando dichos modelos se predicen valores puntuales para todos los nodos de una grilla construida con las 4 variables, separadas por distancias de 0.4 entre –2 y +2 (en los valores codificados). En otras palabras, para cada variable se dispone de 11 valores codificados distintos –2.0, -1.6, -1.2, -0.8, -0.4, 0.0, 0.4, 0.8, 1.2, 1.6 y 2.0 y estos a su vez son combinados con los valores de las otras 3 variables. Esto produce un total de 11x11x11x11=14641 datos para la hora y otros tantos para las 8 horas, i.e., casi 30.000 datos de estimación puntual de disolución según las distintas combinaciones posibles.

Para el análisis de estos resultados se tomaron en cuenta los siguientes criterios, en cuanto a la primer variable de respuesta (disolución):

a) Que la disolución fuera menor o igual a 15 % para la hora.
b) Que la disolución fuera mayor o igual a 57 % para las 8 horas.
c) Que la segunda variable (*C*), en lo posible, tomara un único valor por razones experimentales y a las demás variables les correspondiera el mayor rango posible.

Para los resultados encontrados, se realiza una intersección con los conjuntos de nivel correspondientes a los mayores valores de dureza, encontrados de manera similar (pero con nuevas experiencias, teniendo en cuenta las variables relevantes encontradas) a la descripta.

Resultados

Aplicando los criterios ya mencionados (literales *a*, *b* y *c*) con respecto a la disolución, se obtiene una región representable como un prisma en el espacio de variables *B*, *D*, *H*, mientras que la variable *C* se logró mantener fija en un valor adecuado, con las consiguientes ventajas operativas.

Finalmente, se realiza la intersección del mencionado prisma con los conjuntos de nivel correspondientes a los mayores valores de dureza alcanzables en el experimento. De este proceso surgen condiciones de trabajo óptimas, dadas en forma de intervalos para 3 de las variables ya mencionadas (*B*, *D* y *H*) y un valor fijo para la restante. Esto permite obtener condiciones de trabajo convenientes para las variables que admiten ciertos márgenes de variación y al mismo tiempo se logra fijar la restante variable en un valor determinado, lo que estaría en fase con las posibilidades reales que se disponen en planta.

Referencias

[1] Douglas C. Montgomery, Design and Analysis of Experiments, Wiley & Sons Inc., 2005, MA

[2] Norman Draper and Harry Smith, Applied Regression Analysis, Wiley & Sons Inc., 1998, MA

[3] Martínez Luaces, V.; CD: Notas del curso de Diseño Experimental, AQFU, 2003, Montevideo

[4] George E.P. Box, J. Stuart Hunter and William G. Hunter, Statistics for Experimenters: Design, Innovation, and Discovery, Wiley & Sons Inc. 2005, MA