

Modelos Aditivos Generalizados para optimizar el proceso de hidrofobicidad de la caolinita

Generalized Additive Models to Optimize the Hydrophobicity Process of Kaolinite

Liliana María Usuga Manco¹, Freddy Hernández-Barajas² y Olga Usuga-Manco³

Resumen

El amplio uso industrial de la caolinita requiere que los procesos de extracción sean modelados para determinar las condiciones apropiadas del beneficio. Aunque se han utilizado modelos de regresión lineal clásicos, estos no han sido apropiados debido al incumplimiento de distribución normal para la variable respuesta. Los datos analizados en este estudio corresponden a un proceso de extracción de caolinita mediante fisicoquímica de superficies realizado en La Unión, Antioquia, Colombia. La variable de respuesta fue el potencial zeta y las variables explicativas fueron tipo de solución colectora, concentración y pH. En este artículo se modela la recuperación de caolinita a través de los modelos aditivos generalizados, los cuales permiten elegir la distribución estadística y modelar todos los parámetros en función de variables explicativas. Se seleccionaron cinco distribuciones para la variable respuesta de acuerdo al criterio de información de Akaike (*AIC*). El modelo con distribución generalizada Beta 2 fue el modelo que presentó el mejor desempeño de acuerdo a las métricas utilizadas. A partir de este modelo se encontró que las mejores condiciones de operación obtenidas del análisis de las superficies de respuesta son tipo de colector ácido oleico, concentración 10 unidades y pH de 6.

Palabras clave: caolinita, hidrofobicidad, modelos aditivos, modelos de regresión, potencial zeta.

Abstract

The wide industrial use of kaolinite requires that the extraction processes be modeled to determine the appropriate conditions of the benefit. Although classic linear regression models have been used, these have not been appropriate due to the non-compliance with normal distribution for the response variable. The data analyzed in this study correspond to a kaolinite extraction process by surface physicochemistry carried out in La Unión, Antioquia, Colombia. The response variable was the zeta potential and the explanatory variables were type of collecting solution, concentration, and pH. In this article, the recovery of kaolinite is modeled through generalized additive models, which can choose the statistical distribution and model all the parameters based on explanatory variables. Five distributions were selected for the response variable according to the Akaike information criterion (*AIC*). The model with generalized distribution Beta 2 was the model that presented the best performance according to the metrics used and it was found that the best-operating conditions obtained are the type of oleic acid collector, the concentration of 10 units, and pH 6.

Keywords: additive models, hydrophobicity, kaolinite, regression models, zeta potential.

Recepción: 06-jun-2022

Aceptación: 13-nov-2022

¹Ingeniera Química, Ph.D. en Ingeniería, Ciencia y Tecnología de los Materiales, Universidad de Antioquia, Medellín, Colombia. Correo electrónico: liliana.usuga@udea.edu.co. <https://orcid.org/0000-0003-3811-3839>

²Ingeniero Industrial, Ph.D. en Estadística, Universidad Nacional de Colombia sede Medellín, Medellín, Colombia. Correo electrónico: fhernanb@unal.edu.co. <https://orcid.org/0000-0001-7459-3329>

³Ingeniera Industrial, Ph.D. en Estadística, Universidad de Antioquia Medellín, Colombia. Correo electrónico: olga.usuga@udea.edu.co. <https://orcid.org/0000-0003-3062-1820>

1 Introducción

La caolinita es un mineral industrial con aplicaciones en la producción de papel, cerámica, cosméticos, medicina y porcelana. Actualmente, industrias como la del papel, pinturas y cerámicas exigen caolines más blancos y puros [8]. Debido a que es un material escaso que se encuentra asociado a otros aluminosilicatos, es necesario hacer recuperación usando métodos como flotación inversa, trituración, molienda, clasificación mediante hidrociclones, decantación, cribado en húmedo y secado. Estos métodos conllevan a que se obtenga un producto con medianos niveles de coloración, baja recuperación en peso y un consumo excesivo de agua en los hidrociclones [15].

En la literatura científica se encuentran reportes donde se estudia la influencia de diversos factores en el proceso de recuperación de caolinita. A continuación se presentan algunos de ellos con las herramientas estadísticas utilizadas para analizar los datos experimentales.

Elboughdiri *et al.* [3] prepararon un catalizador de arcilla (montmorillonita y caolinita) el cual se usó para degradar tres compuestos fenólicos obtenidos de la producción de aceite de oliva. Las condiciones de operación de la degradación de estos compuestos fueron optimizadas por la metodología de superficie de respuesta utilizando las covariables pH, tiempo y concentración de H_2O_2 . Zayed *et al.* [17] procesaron caolinita egipcia para sintetizar zeolitas, estos autores estudiaron la influencia de la concentración inicial de manganeso (Mn^{2+}), el tiempo de contacto y la dosis de zeolita en la captación de Manganeso por medio de modelos de regresión de grado tres y superficies de respuesta. En Melo *et al.* [6] se utilizó una metodología secuencial de diseño factorial de experimentos y metodología de superficie de respuesta para evaluar las variables experimentales que influyen en la producción de zeolita sintética tipo FAU. Vadibeler *et al.* [16] exploraron la eficiencia de la goma xantana en la recuperación de partículas finas de caolinita a partir de una suspensión. Los autores usaron optimización estadística por medio del diseño de Box-Behnken para estudiar el efecto del pH, alumbre y goma xantana sobre el porcentaje

de eliminación de turbidez. Ji and Zhang [4] utilizaron modelos de regresión y superficies de respuesta para estudiar el efecto de las variables concentración del ácido (mol/L), temperatura de lixiviación ($^{\circ}C$) y concentración sólida (g/L) sobre el área superficial específica en un proceso de recuperación de partículas finas de caolinita.

En los estudios reportados anteriormente se observa que los modelos estadísticos usados con más frecuencia fueron los modelos lineales clásicos que asumen que la variable respuesta se distribuye normal. En muchos procesos químicos, las variables de interés toman sólo valores positivos y sesgados lo que hace inadecuados a los modelos de regresión clásicos. Por esta razón, los modelos GAMLSS (*Generalized Additive Model for Location Scale and Shape*) propuestos por Rigby y Stasinopoulos [12] son una alternativa apropiada ya que permiten modelar todos los parámetros de la distribución asumida en función de variables explicativas que se tengan.

En el presente artículo se utilizan los modelos GAMLSS para estudiar el efecto del tipo de solución colectora, la concentración y el pH en la recuperación de caolinita mediante un proceso fisicoquímico de superficies. Estos modelos permiten encontrar de forma precisa la recuperación de caolinita a través de la predicción del potencial zeta promedio.

El artículo está organizado de la siguiente manera: en la Sección 2.1 se presenta una descripción del experimento realizado, en la Sección 2.2 se explican de forma breve los modelos GAMLSS y en la Sección 2.3 se introduce a la distribución estadística GB2 (Generalized Beta 2) que fue la distribución del mejor modelo. Por último en la Sección 3 se presentan los modelos construidos, el mejor modelo según las medidas de desempeño y las ecuaciones y superficie de respuesta para el valor promedio de la variable de interés.

2 Materiales y métodos

En esta sección se presenta una descripción corta del experimento realizado y de las variables de interés en el proceso de hidrofobicidad. Luego se da una explicación breve de los modelos estadísticos utilizados para estudiar el proceso y la última subsección

presenta la distribución estadística GB2 (Generalized Beta 2) que resultó ser la distribución que mejor explica el fenómeno.

2.1 Descripción del experimento

En este estudio se analizan los datos reportados por Usuga et al. [15] sobre un experimento realizado para estudiar el efecto del tipo de solución colectora, la concentración y el pH en la recuperación de caolinita mediante un proceso fisicoquímico de superficies. Las muestras de caolinita empleadas en el experimento fueron tomadas en el municipio de La Unión en Antioquia, Colombia. En la Figura 1 se muestra una fotografía ilustrativa de la caolinita utilizada en el experimento.

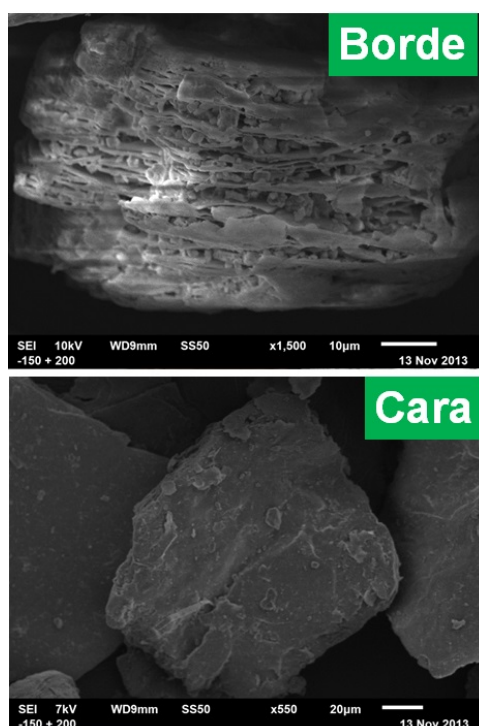


Figura 1. Fotografía ilustrativa de la caolinita (borde y cara) usada en el experimento.

La base de datos obtenida en el experimento consta de 1336 observaciones y 4 variables que se muestran a continuación.

1. Potencial zeta: variable cuantitativa de interés que cuantifica la recuperación del material y es medida en milivoltios [mV].
2. Tipo de solución colectora: se utilizaron tres tipos de reactivos, dodecil sulfato de sodio SDS

($\text{Na C}_{12} \text{H}_{25} \text{SO}_4$, 99,5%), ácido oleico ($\text{C}_{18} \text{H}_{34} \text{O}_2$, 98%) y sin reactivo.

3. Concentración: cantidad de reactivo usado, los valores considerados fueron 0 , 1×10^5 , 1×10^4 , 1×10^3 y 1×10^2 .
4. pH: valor de pH de las soluciones utilizadas para extraer la caolinita, se experimentó con valores diferentes desde 2 hasta 11.

Para obtener la medición del potencial zeta se empleó una fracción de caolinita con un tamaño inferior a $20 \mu\text{m}$, cloruro de sodio (NaCl , 99,7%) 0.01 M como electrolito de soporte y todas las mediciones se hicieron a una temperatura de $23 \text{ }^\circ\text{C}$.

En la Figura 2 se muestra la densidad empírica y el histograma para la variable potencial zeta que presentó valores entre 6.03 y 85.46. De esta figura se observa que la variable sólo asume valores positivos (\mathfrak{R}^+), con una mayor concentración de valores en el intervalo (20,40) y que la distribución tiene un sesgo a la derecha.

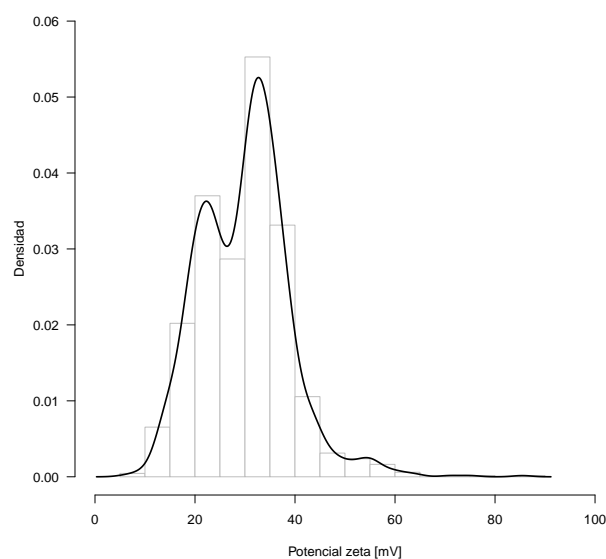


Figura 2. Densidad empírica e histograma para la variable potencial zeta.

En la Figura 3 se muestran tres diagramas para explorar la relación entre el potencial zeta y las variables pH, concentración y tipo de solución. Del panel (a) se observa que a medida que aumenta el pH, el potencial zeta tiende a aumentar su valor promedio y su variabilidad. Del panel (b) se observa un

leve incremento en el potencial zeta a medida que se aumenta la concentración. Finalmente, del panel (c) se observa que el potencial zeta promedio más grande se obtiene con el tipo de solución de ácido oleico y que los menores valores de potencial zeta se obtienen sin solución colectora.

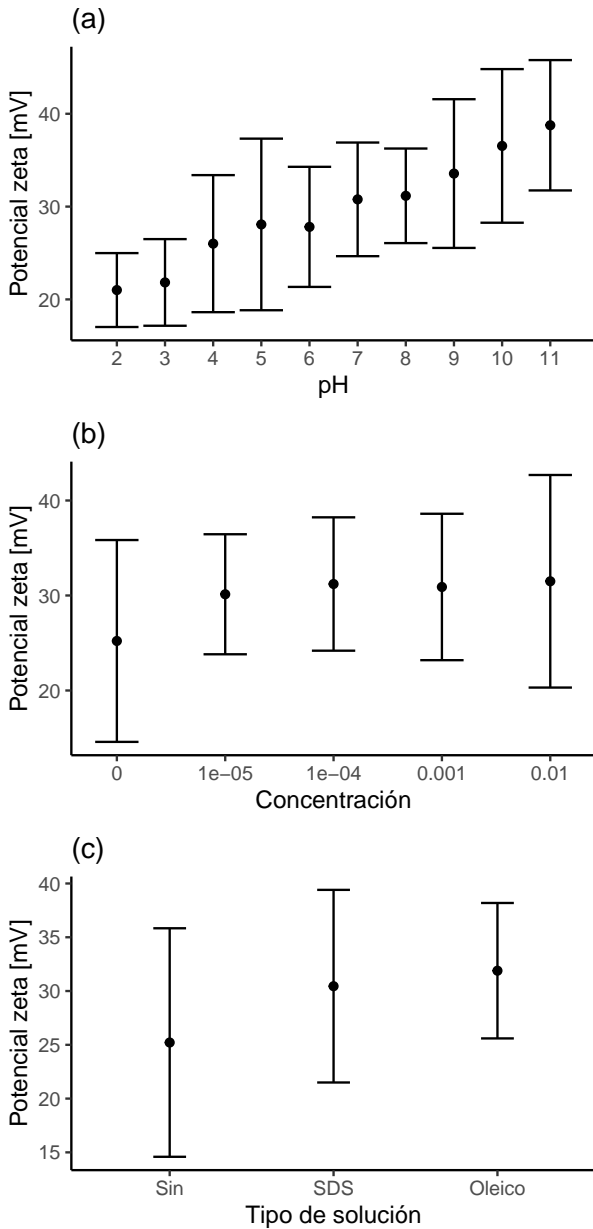


Figura 3. Gráficos del potencial zeta versus pH, concentración y tipo de solución. El punto negro corresponde al potencial zeta promedio (\bar{x}) y las líneas representan un intervalo alrededor de la forma promedio de $\bar{x} \pm s$, siendo s la desviación estándar.

2.2 Modelos GAMLSS

Los modelos GAMLSS (*Generalized Additive Model for Location, Scale and Shape*) propuestos por Rigby y Stasinopoulos [12] son una clase de modelos estadísticos que permiten:

1. modelar todos los parámetros de la distribución estadística de interés,
2. usar variables explicativas para modelar los parámetros de la distribución,
3. elegir la distribución para la variable respuesta entre al menos 100 distribuciones disponibles,
4. incluir efectos fijos, efectos aleatorios y términos aditivos,
5. generalizar varios modelos estadísticos en una sola infraestructura.

En los modelos GAMLSS la variable aleatoria se denota por Y y la estructura de los modelos se puede resumir por medio de las siguientes expresiones:

$$\begin{aligned}
 \mathbf{Y} &\sim D(\boldsymbol{\mu}, \boldsymbol{\sigma}, \boldsymbol{\nu}, \boldsymbol{\tau}), \\
 g_1(\boldsymbol{\mu}) &= \mathbf{X}_1\boldsymbol{\beta}_1 + s_{11}(X_{11}) + \dots + s_{1J_1}(X_{1J_1}), \\
 g_2(\boldsymbol{\sigma}) &= \mathbf{X}_2\boldsymbol{\beta}_2 + s_{21}(X_{21}) + \dots + s_{2J_2}(X_{2J_2}), \\
 g_3(\boldsymbol{\nu}) &= \mathbf{X}_3\boldsymbol{\beta}_3 + s_{31}(X_{31}) + \dots + s_{3J_3}(X_{3J_3}), \\
 g_4(\boldsymbol{\tau}) &= \mathbf{X}_4\boldsymbol{\beta}_4 + s_{41}(X_{41}) + \dots + s_{4J_4}(X_{4J_4}),
 \end{aligned}$$

donde el vector $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)^\top$ es el vector columna con las respuestas. Las matrices \mathbf{X}_k con $k = 1, \dots, 4$, son las matrices de diseño conocidas usadas para explicar cada uno de los parámetros $\boldsymbol{\mu}$, $\boldsymbol{\sigma}$, $\boldsymbol{\nu}$ y $\boldsymbol{\tau}$ de la distribución $D(\cdot)$ asumida para la variable Y . Las funciones $g_k(\cdot)$ son funciones de enlace conocidas. Los elementos $\boldsymbol{\mu}$, $\boldsymbol{\sigma}$, $\boldsymbol{\nu}$ y $\boldsymbol{\tau}$ son vectores n -dimensionales. Los elementos s_{kJ_k} son funciones de suavizado no paramétricas y las X_{kJ_k} corresponden a variables explicativas.

El vector de parámetros en los modelos GAMLSS está formado por $\boldsymbol{\theta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \boldsymbol{\beta}_4)^\top$ y la estimación de los elementos de este vector se realiza por medio del método de Máxima Verosimilitud, el cual se realiza por medio del algoritmo Cole y Green [2] o por el algoritmo Rigby y Stasinopoulos [10, 11]. Los modelos GAMLSS se pueden usar fácilmente

por medio del paquete `gamlss` [13] implementado en el lenguaje de programación R [9].

2.3 Distribución GB2

En el presente estudio se consideraron varias distribuciones estadísticas para explicar el potencial zeta en función de la concentración, pH y el tipo de colector. De todas las distribuciones la que mejor explicó el fenómeno fue la distribución GB2 (Generalized Beta 2) y por eso esta subsección presenta los detalles de esa distribución. Para conocer los detalles de las otras distribuciones usadas se recomienda consultar Stasinopoulos, Rigby y Akantziliotou [14].

La distribución GB2 propuesta por McDonald [5] es una distribución con soporte en \mathfrak{R}^+ y cuenta con una gran flexibilidad para modelar diversos fenómenos debido a sus cuatro parámetros μ , σ , ν y τ . Si una variable Y tiene distribución BCPE se denota por $Y \sim \text{GB2}(\mu, \sigma, \nu, \tau)$ y la función de densidad de probabilidad $f_Y(y)$ está dada por

$$f_Y(y) = \frac{|\sigma|y^{\sigma\nu-1}}{\mu^{\sigma\nu}B(\nu, \tau) \left[1 + \left(\frac{y}{\mu}\right)^\sigma\right]^{\nu+\tau}}, \quad (1)$$

donde $B(\cdot)$ corresponde a la función Beta. Los parámetros de la distribución GB2 pueden asumir múltiples valores pero cumpliendo que $\mu, \nu, \tau \in \mathfrak{R}$ y $\sigma > 0$. La distribución GB2 es muy flexible debido a sus cuatro parámetros, y dependiendo de los valores que se asuman, la función de densidad de probabilidad puede tomar diversas formas. En la Figura 4 se muestran cuatro formas diferentes para la distribución GB2.

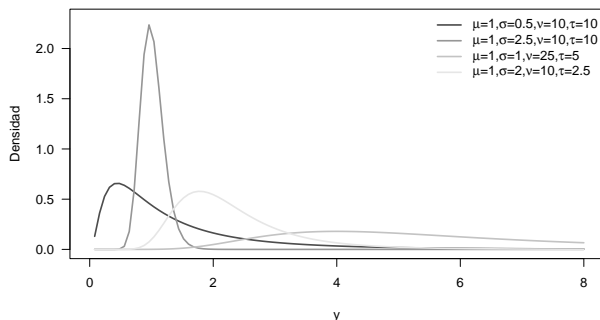


Figura 4. Formas de la densidad GB2 para diferentes valores de los parámetros.

La moda M_o y los momentos $E(Y^h)$ de la distribución GB2 están dados por:

$$M_o = \mu \left(\frac{\sigma\nu - 1}{\sigma\tau + 1} \right)^{\frac{1}{\sigma}}, \quad (2)$$

$$E(Y^h) = \frac{\mu^h B(\nu + h/\sigma, \tau - h/\sigma)}{B(\nu, \tau)}. \quad (3)$$

3 Resultados y discusión

En esta sección se presentan los resultados de los modelos estadísticos construidos para explicar el potencial zeta en función de las covariables pH, concentración y tipo de solución.

3.1 Distribuciones estadísticas para la variable de interés

En este apartado se muestran los resultados de un análisis exploratorio marginal para encontrar las distribuciones estadísticas que mejor explican la variable respuesta potencial zeta. Como la variable de interés sólo puede tomar valores positivos, se consideraron 22 distribuciones candidatas con soporte en \mathfrak{R}^+ para elegir las cinco que mejor explican la variable de interés. Como criterio de selección se utilizó el *Akaike information criterion* propuesto por Akaike [1] y que se define como $AIC = -2 \times \hat{l} + 2 \times df$, donde \hat{l} es el valor de la verosimilitud estimada y df es el número de parámetros estimados; entre menor sea el valor de AIC , mejor es el modelo.

En la Figura 5 se muestran las densidades estimadas de las cinco mejores distribuciones con sus respectivos valores de AIC , así como la densidad estimada de la distribución normal que servirá como distribución de referencia. La distribución Generalized Beta 2 fue la distribución con el menor AIC mientras que la distribución normal fue la que presentó el mayor valor de AIC . Las seis distribuciones mostradas en la figura son las distribuciones asumidas para los modelos de regresión que se presentan en la siguiente sección.

Las seis distribuciones mostradas en la Figura 5 tienen diferente número de parámetros. Las distribuciones normal y gamma tienen dos parámetros (μ y σ), la distribución Generalized gamma tiene tres parámetros (μ , σ y ν) mientras que las distribuciones

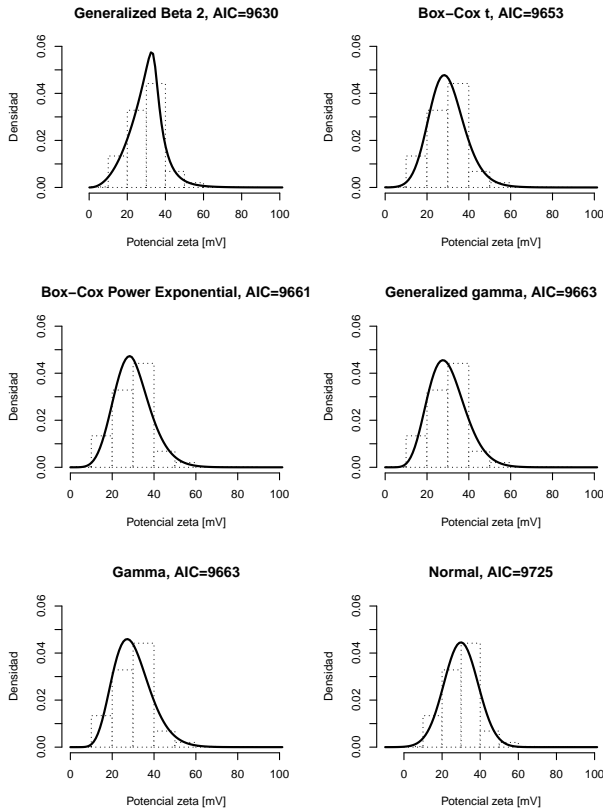


Figura 5. Densidades de las cinco mejores distribuciones que explican la variable potencial zeta. La sexta densidad corresponde a la distribución normal que sirve como distribución de referencia.

Generalized Beta 2, Box-Cox t, y Box-Cox Power Exponential tienen cuatro parámetros (μ , σ , ν y τ).

3.2 Modelos de regresión

En esta sección se presentan los diferentes modelos de regresión contruídos para explicar el potencial zeta (P) en función de las covariables pH (pH), concentración (C) y tipo de solución (S).

El modelo M0 es el modelo de referencia y tiene la siguiente estructura:

$$P_i \sim N(\mu_i, \sigma^2),$$

$$\mu_i = \beta_0 + \beta_1 S_{SDS} + \beta_2 S_{oleico} + \beta_3 pH + \beta_4 C,$$

$$\sigma^2 = constante.$$

En este modelo la distribución de la variable potencial zeta (P) fue la distribución normal, el parámetro μ de la distribución se modeló en función de las covariables pH (pH), concentración (C) y tipo de solución (S), mientras el parámetro σ^2 se asumió

constante. Los coeficientes β 's del modelo corresponden a los efectos de las variables cuantitativas o a los efectos de los niveles de las variables cualitativas, en este modelo no aparece un β asociado al nivel *sin* de la variable solución porque ese nivel es de referencia y está incluido en el β_0 .

Los modelos M1 hasta M6 iniciaron con la variable solución (S) y una estructura polinomial de grado tres para las variables cuantitativas (pH y C). Se decidió usar esta estructura polinomial de grado 3 para capturar posibles relaciones no lineales entre la variable respuesta y las covariables. Un procedimiento de selección de variables ayudará a detectar si el término de grado dos o tres no es importante en los modelos. En la siguiente expresión se resume la estructura usada para los modelos.

$$P_i \sim D(\mu_i, \sigma_i, \nu_i, \tau_i),$$

$$g_1(\mu_i) = \beta_0 + \beta_1 S_{SDS} + \beta_2 S_{oleico} + \text{poly}(pH, 3) + \text{poly}(C, 3),$$

$$g_2(\sigma_i) = \beta_0 + \beta_1 S_{SDS} + \beta_2 S_{oleico} + \text{poly}(pH, 3) + \text{poly}(C, 3),$$

$$g_3(\nu_i) = \beta_0 + \beta_1 S_{SDS} + \beta_2 S_{oleico} + \text{poly}(pH, 3) + \text{poly}(C, 3),$$

$$g_4(\tau_i) = \beta_0 + \beta_1 S_{SDS} + \beta_2 S_{oleico} + \text{poly}(pH, 3) + \text{poly}(C, 3).$$

La letra D de la estructura anterior corresponde a una de las distribuciones mostradas en la Figura 5, y dependiendo de la distribución, se tendrán los parámetros μ , σ , ν y τ , los cuales se modelan en función de las covariables pH, concentración (C) y tipo de solución (S).

Todos los modelos considerados fueron ajustados por medio del paquete `gamlss` [13] y adicionalmente se aplicó un procedimiento de selección de variables para elegir los términos significativos en cada uno de los parámetros de la distribución del modelo. En la tabla 1 se muestran los modelos, la distribución asumida para la variable respuesta, el valor de AIC , el coeficiente de correlación de Pearson ρ entre y_i y $\hat{E}(y_i)$, así como el pseudo R^2 propuesto por Nagelkerke [7].

De la tabla 1 se observa que el modelo con las peores medidas de desempeño fue el modelo de referencia M0 ya que presenta el mayor valor de AIC y el menor valor de pseudo R^2 . Por otra parte, el modelo M2 fue el que presentó el menor AIC , el mayor coeficiente de correlación ρ y el mayor pseudo R^2 . Otros modelos a destacar son M3, M4 y M5 porque

éstos presentan unas medidas de AIC , ρ y pseudo R^2 cercanas a las obtenidas por el modelo M2.

Tabla 1. Modelos estadísticos polinomiales para explicar el potencial zeta en función de las variables pH , concentración y tipo de solución.

Modelo	Distribución	AIC	ρ	Pseudo R^2
M0	Normal	8919.32	0.67	0.45
M1	Normal	8425.23	0.68	0.63
M2	G. Beta 2	8064.44	0.73	0.69
M3	Box-Cox t	8159.95	0.70	0.68
M4	Box-Cox PE	8174.17	0.70	0.68
M5	G. gamma	8270.03	0.68	0.65
M6	Gamma	8359.70	0.67	0.63

Se hizo un análisis gráfico de residuales para comparar los modelos M2, M3, M4 y M5. De este análisis se observó que el modelo M2 fue el que obtuvo un mejor comportamiento y en la Figura 6 se muestran los gráficos construidos para el modelo M2. De esta figura no se observan anomalías que indiquen la no normalidad de los residuales del modelo M2.

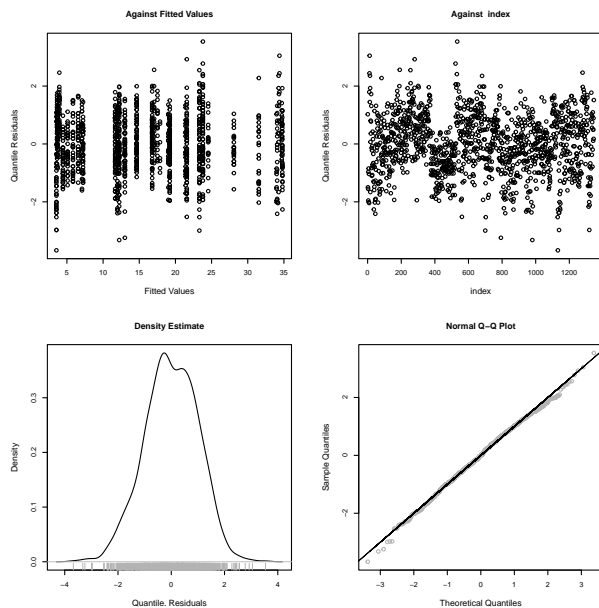


Figura 6. Gráfico de residuales para el modelo M2.

En la tabla 2 se muestran los coeficientes estimados para el modelo M2 con variable respuesta Generalized Beta 2 (GB2). En la primera columna se tienen los cuatro parámetros de la distribución μ , σ , ν y τ . Para cada uno de ellos se tiene la misma función de enlace $\log(\cdot)$ que garantiza que el predictor

lineal genere valores apropiados de los parámetros. La segunda columna de la tabla muestra los términos importantes luego del procedimiento de selección de variables. En las columnas tercera a sexta se muestran los resultados tradicionales al aplicar modelos de regresión.

A partir de los coeficientes estimados reportados en la tabla 2 se pueden escribir expresiones matemáticas para expresar los parámetros estimados μ , σ , ν y τ en función de las covariables S , C y pH . Las siguientes cuatro expresiones se pueden utilizar para calcular el valor estimado de los cuatro parámetros en función de S , C y pH .

$$\begin{aligned} \log(\hat{\mu}) &= 3,272 - 0,379S_{SDS} - 1,570S_{oleico} - 0,312pH + 0,065pH^2 - 0,003pH^3, \\ \log(\hat{\sigma}) &= 3,657 - 2,359S_{SDS} - 3,122S_{oleico} - 82,384C - 0,500pH + 0,080pH^2 - 0,004pH^3, \\ \log(\hat{\nu}) &= -1,677 + 6,330S_{SDS} + 7,978S_{oleico} - 45,530C, \\ \log(\hat{\tau}) &= -1,172 + 4,926S_{SDS} + 6,053S_{oleico}. \end{aligned}$$

Usando la expresión (3) con $h = 1$ se puede escribir el valor esperado de la distribución GB2 en función de $\hat{\mu}$, $\hat{\sigma}$, $\hat{\nu}$ y $\hat{\tau}$ encontradas anteriormente. El valor esperado $\hat{E}(Y)$ del potencial zeta está dado por

$$\hat{E}(Y) = \frac{\hat{\mu} B(\hat{\nu} + 1/\hat{\sigma}, \hat{\tau} - 1/\hat{\sigma})}{B(\hat{\nu}, \hat{\tau})}. \quad (4)$$

En la Figura (7) se muestran las superficies de respuesta para $\hat{E}(Y)$ en función de la concentración, pH y tipo de colector. El eje vertical asociado a $\hat{E}(Y)$ es diferente en los tres casos: cuando no se usa colector el $\hat{E}(Y)$ varía en el intervalo (10, 35), cuando se usa colector SDS el $\hat{E}(Y)$ varía en el intervalo (20, 45) mientras que cuando se usa colector oleico el $\hat{E}(Y)$ varía en el intervalo (20, 120). Esto indica que el uso de colector oleico favorece o aumenta el promedio de potencial zeta que se puede obtener.

Tabla 2. Valores estimados para el modelo M2 con variable respuesta Generalized Beta 2.

Parámetro	Término	Estimación	Error estándar	Valor t	$P(> t)$
$\log(\mu)$	Intercepto	3.272	0.064	51.234	0.000
	S_{SDS}	-0.379	0.081	-4.703	0.000
	S_{oleico}	-1.570	0.131	-11.950	0.000
	pH	-0.312	0.020	-15.931	0.000
	pH^2	0.065	0.002	39.576	0.000
	pH^3	-0.003	0.000	-1155.300	0.000
$\log(\sigma)$	Intercepto	3.657	0.580	6.302	0.000
	S_{SDS}	-2.359	0.606	-3.890	0.000
	S_{oleico}	-3.122	0.596	-5.238	0.000
	C	-82.384	5.744	-14.342	0.000
	pH	-0.500	0.018	-27.201	0.000
	pH^2	0.080	0.002	50.921	0.000
	pH^3	-0.004	0.000	-2753.114	0.000
$\log(\nu)$	Intercepto	-1.677	0.635	-2.641	0.008
	S_{SDS}	6.330	0.720	8.796	0.000
	S_{oleico}	7.978	0.640	12.474	0.000
	C	-45.530	3.992	-11.405	0.000
$\log(\tau)$	Intercepto	-1.172	0.653	-1.795	0.073
	S_{SDS}	4.926	0.762	6.466	0.000
	S_{oleico}	6.053	0.700	8.642	0.000

Al analizar con detalle la primer superficie de respuesta de la Figura (7) para $\hat{E}(Y)$ sin colector se observa una forma de “S”. De esta figura se nota que al aumentar el pH el valor esperado del potencial zeta tiende a aumentar, adicionalmente, al aumentar la concentración se disminuye el valor esperado del potencial zeta. En la segunda superficie de respuesta de la Figura (7) para $\hat{E}(Y)$ con colector SDS se observa una forma de “S” más suave. De esta figura se nota que al aumentar el pH o la concentración el valor esperado del potencial zeta tiende a aumentar. De la tercera superficie de respuesta para $\hat{E}(Y)$ con colector oleico no se observa un patrón claro como en los casos anteriores.

Otro resultado interesante que se observa de la Figura (7) es que cuando no se usa colector, el mayor valor de potencial zeta esperado ocurre cuando el pH es 10 y la concentración es cero. En el caso de colector SDS se espera que el mayor valor de potencial zeta se dé para la combinación pH de 10 y concentración de 10. Por último, para el caso de colector oleico, se espera que cuando pH es 6 y la concentración es 10, ocurra el mayor valor esperado de potencial zeta.

4 Conclusiones

En este artículo se utilizaron los modelos GAMLSS para modelar el potencial zeta en la extracción de caolinita como función de las variables explicativas tipo de solución colectora, concentración y pH. El modelo estadístico que mejor predice la variable respuesta es el modelo M2 que obtuvo un coeficiente de correlación de Pearson de 0.73 y un pseudo R^2 de 0.69, siendo estos los valores más altos comparados con los otros modelos construidos para modelar los datos del potencial zeta. Adicional a ello, se muestra de manera gráfica la normalidad de los residuales de este modelo M2, con lo cual se comprueba que el modelo es apropiado para explicar el potencial zeta.

El valor esperado del potencial zeta para el modelo M2 puede obtenerse a partir de los coeficientes estimados reportados en la Tabla 2 o mediante la Ecuación (4) substituyendo los valores estimados $\hat{\mu}$, $\hat{\sigma}$, $\hat{\nu}$ y $\hat{\tau}$ para ciertos valores de las variables experimentales tipo de colector, concentración y pH. Adicionalmente, en este artículo se presenta la superficie de respuesta en la Figura 7 que es una herramienta gráfica útil para representar el comportamiento de $\hat{E}(Y)$ en función de la concentración, pH y tipo de colector.

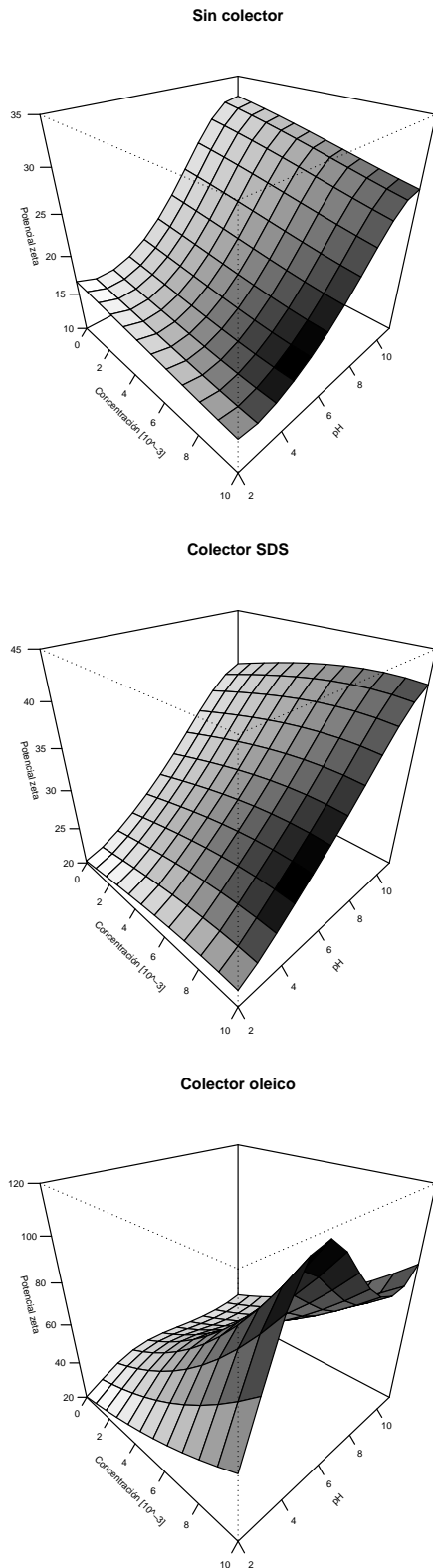


Figura 7. Superficie de respuesta para $\hat{E}(Y)$ en función de la concentración, pH y tipo de colector.

Usando los resultados del modelo se encontró que las mejores condiciones para optimizar el potencial

zeta son usar colector oleico, concentración 10 y pH 6, lo cual coincide con los resultados experimentales reportados.

Referencias

- [1] H. Akaike. A new look at the statistical model identification. *IEEE transactions on automatic control*, vol. 19, no. 6, pp. 716-723, 1974.
- [2] T. J. Cole and P. J. Green. Smoothing reference centile curves: The LMS method and penalized likelihood. *Statistics in Medicine*, vol. 11, no. 10, pp. 1305-1319, 1992. ISSN 02776715. <https://onlinelibrary.wiley.com/doi/10.1002/sim.4780111005>.
- [3] N. Elboughdiri, A. Mahjoubi, A. Shawabkeh, H. Khasawneh, and B. Jamoussi. Optimization of the degradation of hydroquinone, resorcinol and catechol using response surface methodology. *Advances in Chemical Engineering and Science*, vol. 5, pp. 111-120, 2015. ISSN 0169-1317.
- [4] B. Ji and W. Zhang. Rare earth elements (rees) recovery and porous silica preparation from kaolinite. *Powder Technology*, vol. 391, pp. 522-531, 2021. ISSN 0032-5910. <https://www.sciencedirect.com/science/article/pii/S0032591021005520>.
- [5] J. B. McDonald. Some generalized functions for the size distribution of income. *Econometrica*, vol. 52, no. 3, pp. 647-663, 1984. ISSN 00129682, 14680262. <http://www.jstor.org/stable/1913469>.
- [6] C. C. A. Melo, B. L. S. Melo, R. S. Angélica, and S. P. A. Paz. Gibbsite-kaolinite waste from bauxite beneficiation to obtain fau zeolite: Synthesis optimization using a factorial design of experiments and response surface methodology. *Applied Clay Science*, vol. 170, pp. 125-134, 2019. ISSN 0169-1317. <https://www.sciencedirect.com/science/article/pii/S0169131719300158>.
- [7] N. J. Nagelkerke et al. A note on a general definition of the coefficient of determination. *Biometrika*, vol. 78, no. 3, pp. 691-692, 1991.

- [8] X. Qiu, X. Lei, A. Alshameri, H. Wang, and C. Yan. Comparison of the physicochemical properties and mineralogy of chinese (beihai) and brazilian kaolin. *Ceramics International*, vol. 40, no. 4, pp. 5397-5405, 2014. ISSN 0272-8842. <https://www.sciencedirect.com/science/article/pii/S027288421301393X>
- [9] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2022 <https://www.R-project.org/>
- [10] R. A. Rigby and D. M. Stasinopoulos. Mean and dispersion additive models. In W. Härdle and M. G. Schimek, editors, *Statistical Theory and Computational Aspects of Smoothing*, pp. 215-230. Physica-Verlag Heidelberg, 1 edition, 1996. ISBN 978-3-7908-0930-5.
- [11] R. A. Rigby and D. M. Stasinopoulos. A semi-parametric additive model for variance heterogeneity. *Statistics and Computing*, vol. 6, no. 1, pp. 57-65, 1996. ISSN 0960-3174. <http://link.springer.com/10.1007/BF00161574>.
- [12] R. A. Rigby and D. M. Stasinopoulos. Generalized Additive Models for Location, Scale and Shape. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 54, no. 3, pp. 507-554, 2005. <http://www.jstor.org/stable/3592732>.
- [13] R. A. Rigby and D. M. Stasinopoulos. Generalized additive models for location, scale and shape,(with discussion). *Applied Statistics*, vol. 54, pp. 507-554, 2005.
- [14] M. Stasinopoulos, B. Rigby, and C. Akantzi-liotou. Instructions on how to use the gamlss package in R second edition, 2008.
- [15] L. Usuga-Manco, L.-V. A., and B.-R. M. Estudio de la hidrofobicidad de la caolinita de la Unión, Antioquia. *Tecnológicas*, vol. 18, pp. 71-81, 2015.
- [16] D. Vadibeler, E. C. Ugwu, N. Martínez-Villegas, and B. Sen Gupta. Statistical analysis and optimisation of coagulation-flocculation process for recovery of kaolinite and calcium carbonate from suspensions using xanthan gum. *Journal of Food, Agriculture and Environment*, vol. 18, no. 2, pp. 103-109, 2020. ISSN 1459-0263. <https://www.wflpublisher.com/Journal>.
- [17] A. M. Zayed, A. Q. Selim, E. A. Mohamed, M. S. Abdel Wahed, M. K. Seliem, and M. Sillanpa. Adsorption characteristics of na-a zeolites synthesized from egyptian kaolinite for manganese in aqueous solutions: Response surface modeling and optimization. *Applied Clay Science*, vol. 140, pp. 17-24, 2017. ISSN 0169-1317. <https://www.sciencedirect.com/science/article/pii/S0169131717300455>.