

MODELOS GAMLSS PARA ANALIZAR EL GRADO DE SECADO DE ACETATO DE CALCIO DIHIDRATADO

BRAHIAN CANO URREGO^{1,a}, EDWIN ALEXANDER CAICEDO CHAMORRO^{1,b},
SANTIAGO TORO ZULUAGA^{1,c}, FREDDY HERNÁNDEZ BARAJAS^{1,d}

¹ESCUELA DE ESTADÍSTICA, FACULTAD DE CIENCIAS, UNIVERSIDAD NACIONAL DE COLOMBIA, SEDE MEDELLÍN,
COLOMBIA

Resumen

En este trabajo se emplearon modelos GAMLSS para reanalizar el problema abordado por los autores Triana et al. (2009); en el cual se creó un modelo para explicar el porcentaje de grado de secado (GS) del acetato de calcio dihidratado mediante un secador de tipo transportador. Con los modelos GAMLSS se encontró una mejor distribución y adecuación para la variable respuesta, llegando así, a un modelo con distribución Weibull con parametrización 3 (WEI3) que cumple con las cualidades de interpolación y modelamiento de los datos, los resultados fueron obtenidos empleando el programa estadístico R, los cuales arrojaron una correlación de 0.92 y un GAIC 71.58.

Palabras clave: GAMLSS, Secador de tornillo transportador, acetato de calcio dihidratado, diseño factorial, modelo de regresión lineal, distribución Weibull.

Abstract

In this work, the GAMLSS models were used to reanalyze the problem addressed by the authors Triana et al (2009); proposing an alternative option that explains the percent of degree of drying (GS) of calcium acetate dihydrate by means of a conveyor-type dryer. Finding with this methodology a better distribution and adaptation for the response variable, arriving at a model with Weibull distribution with parameterization 3 (WEI3) that meets the qualities of interpolation and modeling of the data, the results were obtained using the statistical program R, which showed a correlation of 0.92 and GAIC 71.58 .

Key words: GAMLSS, Conveyor screw dryer, calcium acetate dihydrate, factorial design, linear regression model, Weibull distribution.

1. Introducción

Triana et al. (2009) usaron regresión lineal múltiple para trata de modelar con una distribución normal el porcentaje de grado de secado (GS) del acetato de calcio dihidratado mediante un secador de tipo transportador en función de las covariables: velocidad rotacional del tornillo (N) en dos niveles 3 y 7 rpm, la temperatura (T) del sólido alimentado en tres niveles 45, 53, 61 grados centígrados y el flujo de alimentación ($^{\circ}C$) en tres niveles 6, 8 y 10 L/min.

^aEstudiante de Estadística. E-mail: bcanou@unal.edu.co

^bEstudiante de Estadística. E-mail: eacaicedc@unal.edu.co

^cEstudiante de Estadística. E-mail: satorozu@unal.edu.co

^dProfesor asistente. E-mail: fhernanb@unal.edu.co

El secador de tornillo transportador (SCD) es un secador continuo de calentamiento indirecto, debido a la versatilidad que tiene este tipo de secador recientemente se han realizado estudios relacionados con el desempeño del secador y el comportamiento de las variables hidrodinámicas y térmicas. Nuestro objetivo principal es proponer modelos alternativos que podrían explicar de mejor manera la variable respuesta, utilizando metodología GAMLSS.

2. Caso de estudio

Triana et al. (2009) Realizaron un procedimiento experimental que consta de un cristalizador con chaqueta C-101 donde se controla el porcentaje de humedad del alimento (50-70) y su temperatura 45°, 53°, 61° (T). El sistema de alimentación al tornillo tiene una bomba de doble diafragma B-101 (inyección de aire), una válvula de recirculación V-102 y las válvulas de bloqueo V-101 y V-103 de la línea de descarga. La presión de aire de suministro de la bomba fue de 40 psi. El flujo de alimentación se ajusta con la válvula mariposa V-104 para 6, 8 ó 10 L/min (C). El secador de tornillo se calienta, S-101, con vapor de caldera a través de la válvula V-105. Se acciona el tornillo fijando la velocidad rpm 3 y 7 (N) por medio del variador de frecuencia VG-101. Una vez ajustado el flujo, se purga el condensado remanente de la chaqueta y luego se dispone el flujo de alimentación en la primera boquilla del secador. Cuando el grado de llenado no cambia con el tiempo se mide la velocidad del tornillo.

GS	14.13	12.08	7.49	5.94	1.17	0.25	7.23	5.80	0.59
T	61	45	61	61	53	45	45	53	53
N	3	3	3	3	3	3	3	3	3
C	6	8	10	8	10	6	10	6	8
GS	3.61	9.58	0.22	1.31	2.09	1.39	3.84	5.35	5.75
T	53	61	53	53	45	45	45	61	61
N	7	7	7	7	7	7	7	7	7
C	6	10	8	10	6	10	8	8	6

Tabla 1: Datos del experimento obtenidos de Triana et al. (2009)

En la tabla 1 se presenta la base de datos del experimento de Triana et al. (2009) en la cual se aprecian los niveles de cada variable además los datos resaltados fueron eliminados por Triana et al. (2009), pues estos datos causan una desviación adimensional que se atribuye a una propiedad del sólido no controlada.

El modelo de regresión lineal múltiple de Triana et al. (2009) es el siguiente:

$$GS = 372.57 - 11.31T - 10.30C - 11.14N + 0.09T^2 + 0.28C^2 + 0.04T * C + 0.10T * N + 0.53C * N$$

Las características de este modelo son; $R^2 = 0.69$ Y su correlación $Corr(Y, \hat{Y})$ con la variable respuesta es de 0.92.

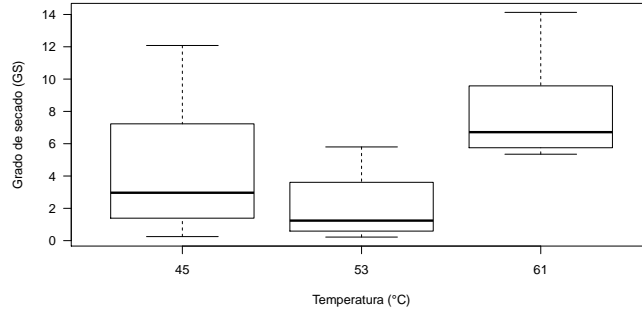


FIGURA 1: Grado de secado discriminado por temperatura

De la figura 1 se muestra el boxplot para el grado de secado diferenciado por el tiempo, de esta figura se puede apreciar una diferencia con respecto a la media de grado de secado (GS) entre los niveles de temperatura de 53° y 61°, siendo mayor para la temperatura del 61°.

3. Modelos GAMLSS

El modelo GAMLSS supone observaciones independientes y_i para $i = 1, 2, \dots, n$ con función de probabilidad (densidad) $f(y_i|\theta^i)$ condicional en θ^i donde $\theta^i = \theta_{i1}, \theta_{i2}, \theta_{i3}, \dots, \theta_{ip}$ es un vector de parámetros p , cada uno de los cuales está relacionado con las variables explicativas. En muchas situaciones prácticas, a lo sumo $p = 4$ se requieren parámetros de distribución. La implementación R denota estos parámetros como $(\mu_i, \sigma_i, v_i, \tau_i)$. Los primeros dos parámetros de población μ_i y σ_i suelen caracterizarse como parámetros de escala y ubicación, mientras que los parámetros restantes, si los hay, se caracterizan como parámetros de forma, aunque el modelo puede ser aplicado de manera más general a los parámetros de cualquier distribución de población. Deje que $y^T = (y_1, y_2, \dots, y_n)$ sea el vector de longitud n de la variable de respuesta. También para $k = 1, 2, 3, 4$, sea $g_k(\cdot)$ se conozcan las funciones de enlace monotónicas que relacionan el k^{th} parámetro θ_k con las variables explicativas por semi-modelos de aditivos paramétricos dados por

$$g_1(\mu) = n_1 = x_1\beta_1 + \sum_{j=1}^{J_1} h_{j1}(x_{j1})$$

$$g_2(\sigma) = n_2 = x_2\beta_2 + \sum_{j=1}^{J_2} h_{j2}(x_{j2})$$

$$g_3(\mu) = n_3 = x_3\beta_3 + \sum_{j=1}^{J_3} h_{j3}(x_{j3})$$

$$g_4(\tau) = n_4 = x_4\beta_4 + \sum_{j=1}^{J_4} h_{j4}(x_{j4})$$

donde μ, σ, v, τ y n_k y x_{jk} , para $j = 1, 2, \dots, J_k$ y $k = 1, 2, 3, 4$, son vectores de longitud n . La función h_{jk} es una función aditiva no paramétrica de la variable explicativa X_{jk} evaluada en x_{jk} . Los vectores explicativos x_{jk} se suponen fijos y conocidos. También X_k por $k = 1, 2, 3, 4$, son matrices de diseño fijo, mientras que β_k son los vectores de parámetros. Tenga en cuenta que en las aplicaciones típicas, un modelo constante u otro modelo simple a menudo es adecuado para cada uno de los dos parámetros de forma (v y τ).

4. Aplicación de modelos GAMLSS

En primera instancia se buscó marginalmente las distribuciones que pudieran explicar el patrón de variación de la variable respuesta, en nuestro caso el porcentaje grado de secado; la función *fitDist* del paquete *gamlss* permite hacerlo. Encontramos las mejores cuatro distribución para modelar el porcentaje de grado de secado (GS); de esto pudimos obtener las siguiente figura:

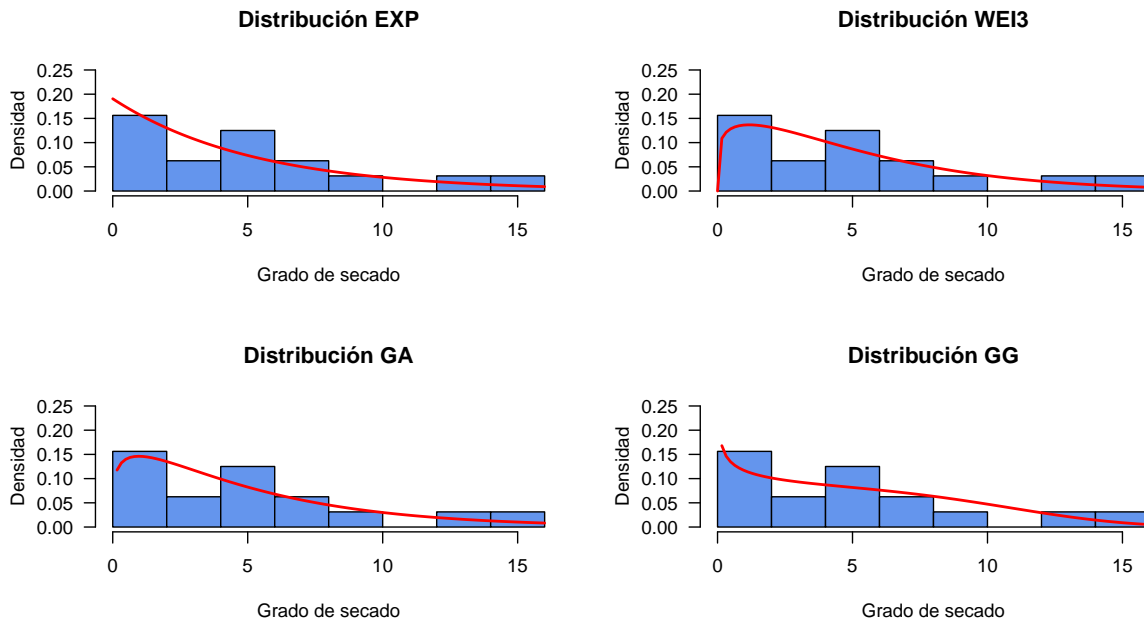


FIGURA 2: Histograma para la variable grado de secado y densidades Exponencial, Weibull3, Gamma y Gamma generalizada

Dentro del desarrollo del trabajo, se propusieron diferentes tipos de modelos, entre ellos lm, no paramétricos, transformaciones y GAMLSS. Para la selección de los mejores modelos se utilizaron los criterios: Normalidad de los errores, GAIC, Correlación con la variable respuesta (capacidad de interpolar). En busca del mejor modelo posible para cada distribución, se realizó un proceso de selección de variables con la función *StepAICALLA()* para los mejores modelos considerados, arrojando como resultado final los siguientes modelos:

Modelos	Distribución	V.Respuesta	Cor	GAIC	R_{adj}^2
modgs	Normal	GS	0.95	79.02	0.69
mod1	Normal	\sqrt{GS}	0.91	28.39	0.76
mod2	Exponencial	GS	0.76	85.36	0.45
mod3	Gamma	GS	0.75	49.61	0.97
mod4	Gamma G	GS	0.92	31.81	0.99
mod5	Weibull 3	GS	0.92	71.58	0.83

Tabla 2: Tabla comparativa de las medidas de ajuste

Después de analizar los modelos con los criterios GAIC, correlación y normalidad de los errores, descartamos los modelos que tienen violaciones graves de los supuestos normalidad, en segunda instancia filtramos por los de menor GAIC y por ultimo elegimos el modelo con la mejor capacidad de explicar la variable respuesta en términos de correlación y ajuste.

De este proceso se obtuvo como mejor modelo el WEI3, con $E(GS) = \mu$ y $\widehat{Var}(GS) = \hat{\mu}^2 \frac{\Gamma(\frac{1}{\sigma}+1)}{\Gamma(\frac{1}{\sigma}+1)^2-1}$

Observemos la normalidad de los residuales que se proponen en el paquete *gamlss* usando la función: *wp*

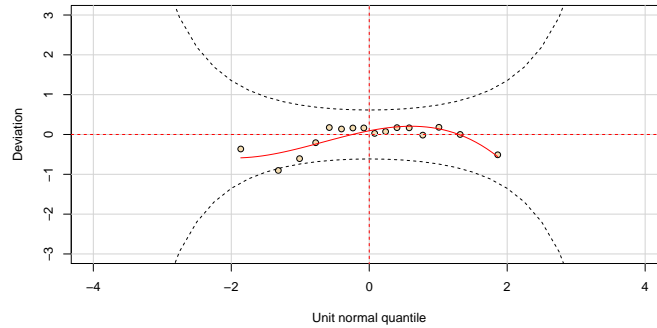


FIGURA 3: Wormplot del modelo WEI3

No podemos observar violaciones graves del supuesto de normalidad de los errores.

Este modelo tiene un GAIC= 71.58 , una Correlación con la variable respuesta de 0.92 ; un R_{adj}^2 de 0.83 las siguientes ecuaciones describen el modelo $GS \sim WEI3(\mu, \sigma)$ con parámetros:

$$\begin{aligned}\widehat{E}(GS) &= \exp(109.58 - 2.05N - 3.30T - 4C + 0.03T^2 + 0.14C^2 + 0.02N \cdot T + 0.07N \cdot C + 0.02T \cdot C) \\ \widehat{Var}(GS) &= \exp(152.47 + 0.05T^2 - 5.07T - 0.92N)\end{aligned}$$

Con el modelo resultante calculamos los valores para los cuales se optimiza el porcentaje de grado de secado, y además generamos una superficie respuesta, fijando la velocidad rotacional del tornillo en 3 rmp, la cual es la optima, para poder obtener una idea visual del fenómeno.

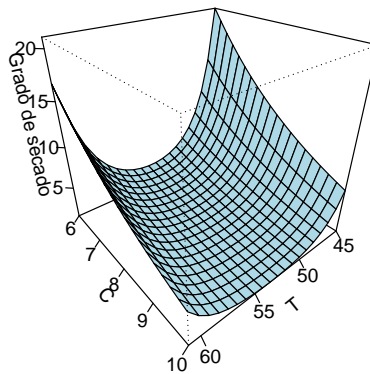


FIGURA 4: Superficie de respuesta de grado de secado (GS) a una velocidad rotacional de 3 rpm

5. Conclusiones

Se obtiene como resultado un mejor modelo con respecto al propuesto en el artículo en términos de interpolación y adecuación del fenómeno evidenciado por la correlación y el GAIC, con este resultado se evidencia la utilidad que tiene dar un uso adecuado de la metodología GAMLSS y no limitarse exclusivamente a modelar la variable respuesta con una distribución normal.

Se observó que el punto óptimo del porcentaje de grado de secado (GS) es 21 , el cual se obtiene con unos valores de $TE=45^{\circ}C$, $N=3$, $C=6$ esto se contradice la creencia inicial de la parte descriptiva que nos indicaba que la temperatura que podría optimizar el grado de secado era de 61° , esto puede deberse a que no se tomaba en cuenta la interacción entre las covariables o que solo tomaba en cuenta el termino lineal.

Referencias

- Triana, M, A., Ayala, M, A., De La Puente, F & Camargo , R (2009). *Grado de secado y eficiencia térmica de un secador de tornillo transportador drying DYNA*, **78(165)**, 196–206.
- Rigby, B.; Stasinopoulos, M (2005). *Generalized additive models for location scale and shape. Applied Statistics*, **54**, 507–554.
- R Core Team (2018). *R: A Language and Environmental for Statistical Computing, R Foundation for Statistical Computing*. Vienna, Austria. <http://www.R-project.org/>.