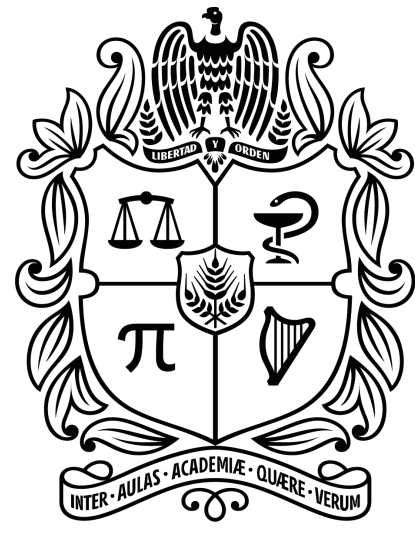


MODELOS GAMLSS PARA ANALIZAR EL GRADO DE SECADO DE ACETATO

DE CALCIO DIHIDRATADO

Brahian Cano Urrego
Edwin Caicedo Chamorro
Santiago Toro Zuluaga
Freddy Hernández Barajas

Universidad Nacional de Colombia sede Medellín



UNIVERSIDAD
NACIONAL
DE COLOMBIA

1. Resumen

En este trabajo se emplearon modelos GAMLSS para re-analizar el problema abordado por los autores Triana et al. (2009); en el cual se creó un modelo de regresión lineal múltiple 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, los resultados fueron obtenidos empleando el programa estadístico R, los cuales arrojaron una correlación de 0.92 y un GAIC 71.58.

2. 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° y el flujo de alimentación (°C) en tres niveles 6, 8 y 10 L/min.

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 modelos GAMLSS.

3. Modelos GAMLSS

El modelo GAMLSS supone observaciones independientes y_i para $i = 1, 2, \dots, n$ con función de probabilidad condicional $f(y_i|\theta^i)$ 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 se requieren $p = 4$ parámetros de distribución. Estos parámetros se denotan 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.

En la mayoría de los estudios para modelar una variable no siempre se cumple el supuesto de normalidad o varianza constante; lo cual brinda una ventaja significativa a la metodología GAMLSS, dado su flexibilidad, ya que ninguno de estos es un requerimiento indispensable para el modelamiento.

4. Caso de estudio

Se presenta la base de datos del experimento de Triana et al. (2009), los datos resaltados fueron eliminados, pues estos causan una desviación adimensional que se atribuye a una propiedad del sólido no controlada.

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

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

Triana et al. (2009) Realizaron un procedimiento experimental que consta de un cristizador con chaqueta C-101

donde se controla el porcentaje de humedad del alimento (50-70) y su temperatura 45°, 53°, 61° (T).

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

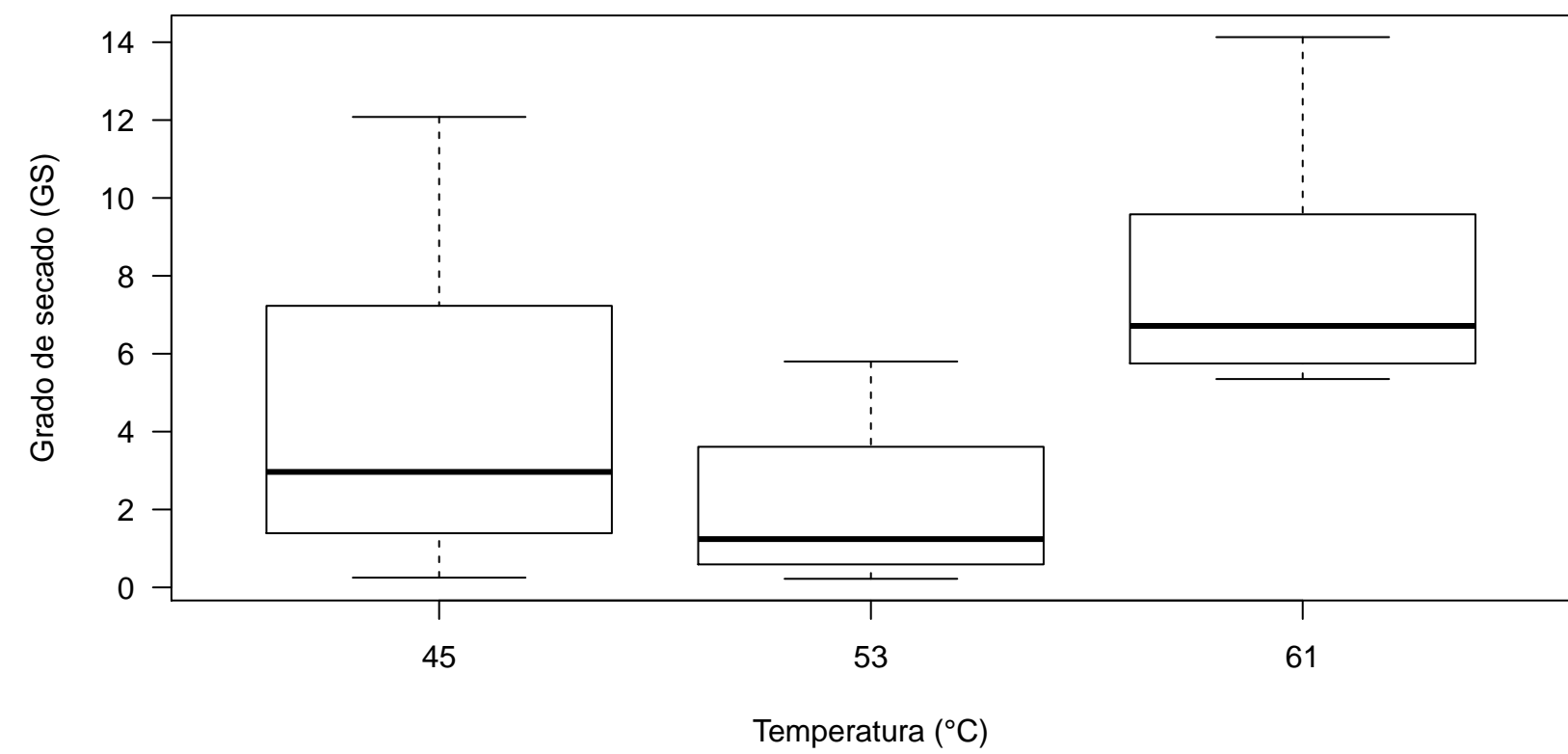


Figure 1: Grado de secado discriminado por temperatura

5. 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 la siguiente figura:

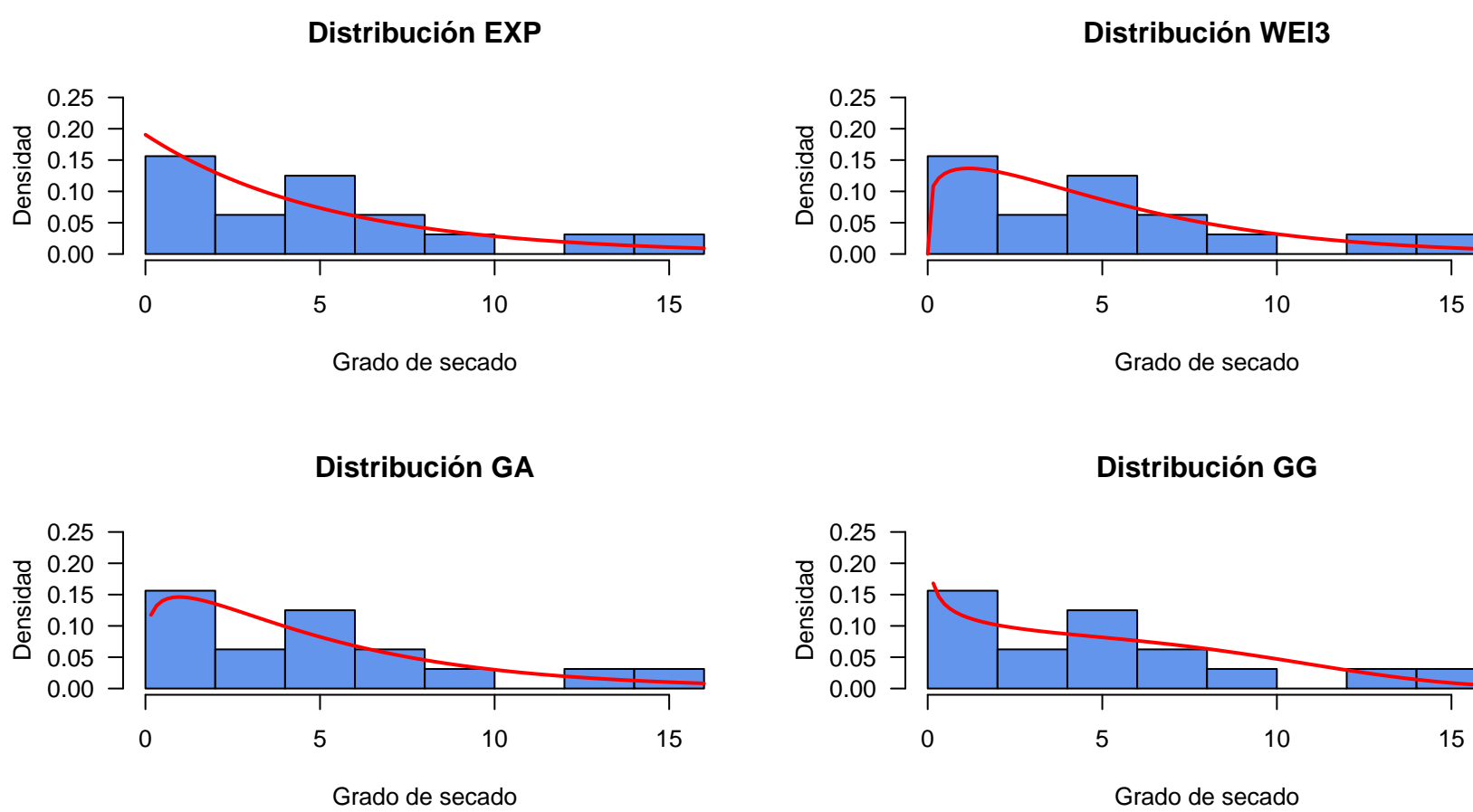


Figure 2: Posibles distribuciones del grado de secado.

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 *StepAICall.A* para los mejores modelos considerados, arrojando como resultado final los siguientes modelos:

Modelos	Distribución	V.Respuesta	Cor	GAIC	R^2_{adj}
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.

6. Resultados

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*

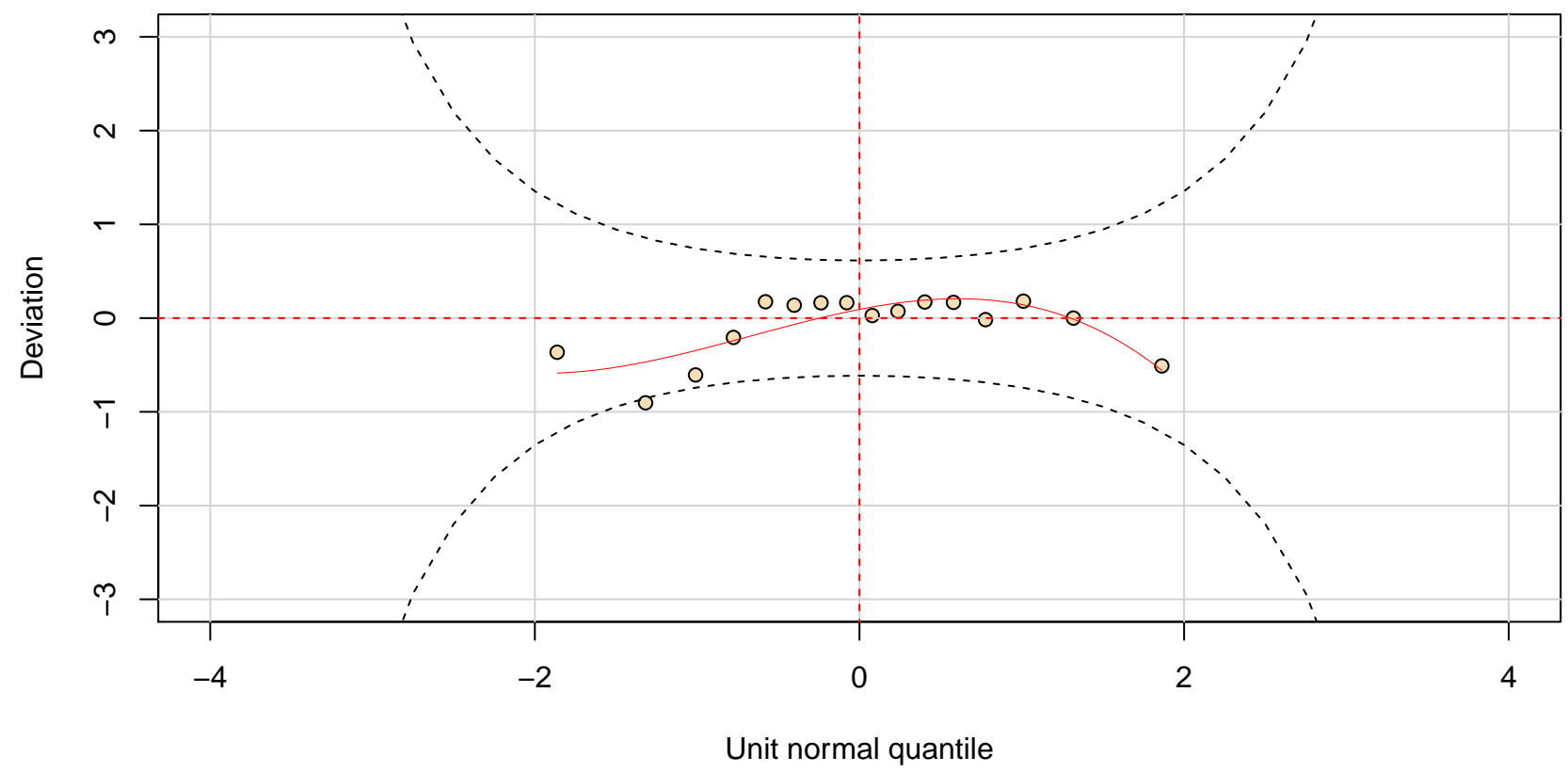


Figure 3: Wormplot del modelo WEI3.

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^2_{adj} de 0.83 las siguientes ecuaciones describen el modelo $GS \sim WEI3(\mu, \sigma)$ con parámetros:

$$\hat{E}(GS) = \exp(78.60 - 1.94N + 0.02T^2 - 2.66T - 0.59C + 0.02N * T + 0.10N * C)$$

$$\widehat{Var}(GS) = \exp(1.11)$$

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 rpm, la cual es la optima, para poder obtener una idea visual del fenómeno.

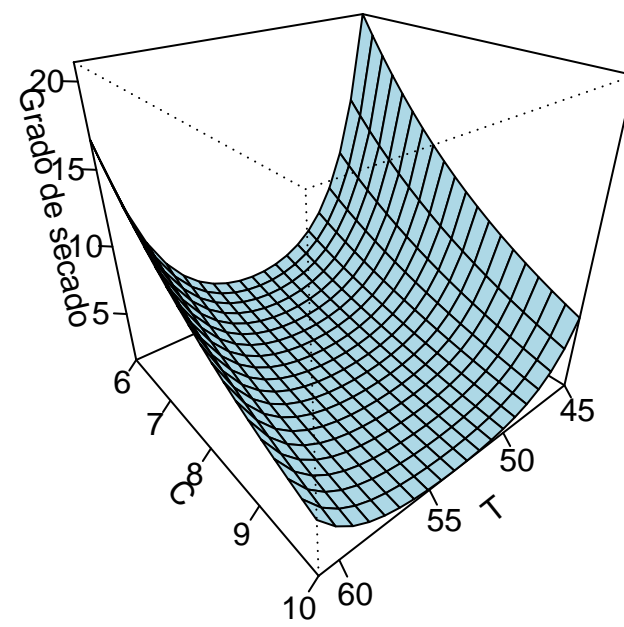


Figure 4: Superficie de respuesta, grado de secado (GS) a una velocidad rotacional de 3 rpm.

7. 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/>.