



Utilización de modelos mixtos lineales generalizados en el análisis de la velocidad del fluido en el interior de un pulverizador agrícola

A. J. Aguirre¹; G. Guevara-Viera²; C. S. Torres-Inga³; R. Guevara-Viera²; J. Badules⁴; A. Boné⁴; M. Vidal⁴ y F. J. García-Ramos⁴

¹ CPIFP Montearagón. Huesca, España; angeljavieraguirre65@gmail.com

² Facultad de Ciencias Agropecuarias de la Universidad de Cuenca. Ecuador.

³ Departamento de Matemáticas y Estadística de la Universidad de Nuevo México. Albuquerque, USA.

⁴ Escuela Politécnica Superior de la Universidad de Zaragoza. Huesca, España.

Resumen: Este trabajo tiene por objetivo formular el mejor modelo mixto lineal generalizado para inferir la velocidad del fluido en el interior de un pulverizador agrícola bajo parámetros operacionales: llenado del depósito, presión interior y número de boquillas abiertas, teniendo en cuenta la geometría del depósito. Se llevó a cabo una formulación de modelos de complejidad creciente, sin transformar la variable, ni eliminar o corregir los numerosos datos extremos y atípicos. El mejor modelo incluyó como efectos fijos: la cantidad de líquido dentro del depósito, la presión interior y el número de boquillas, y como efecto aleatorio la sección de la cuba. La bondad de ajuste de los modelos fue evaluada a través de los menores valores de AICc, BIC y la estima de la varianza residual. La distribución Gamma y la función de enlace logarítmica mejoraron la bondad de ajuste del mejor modelo. El análisis conjunto de los tres efectos fijos puso en evidencia que la mayor influencia sobre la velocidad del fluido la ejercen los litros que contiene el depósito, seguido de la presión del sistema de agitación y, por último, el número de boquillas activas. El modelo sobre estima la velocidad, pero modera los valores elevados.

Palabras clave: AICc; varianza residual; efectos aleatorios; función de enlace; regresión lineal.

1. Introducción

La velocidad del fluido en el interior de un pulverizador agrícola, necesaria para la mezcla homogénea de la materia activa, está determinada por el sistema de agitación interna, de tipo hidráulico en la mayoría de los casos. Esta velocidad varía según (a) algunos parámetros de trabajo: presión, número de boquillas, y nivel de llenado del depósito; y (b) la geometría de la cuba: altura, anchura, tipología de la sección y presencia de obstáculos e irregularidades en el interior del depósito y la ubicación de las boquillas.

Estos dos conjuntos de parámetros dificultan que los datos observados de la velocidad del fluido se ajusten a una distribución normal. Lo que obliga a la utilización de métodos no paramétricos a la hora de evaluar las posibles variaciones de velocidad según parámetros de trabajo y geometría [1]. Esta aproximación analítica permite conocer el efecto de los parámetros por separado, enmascarando las posibles interacciones entre ellos, así como el peso que cada cual ejerce sobre las variaciones de velocidad.

Los modelos lineales generalizados mixtos (GLMM, por sus siglas en inglés) amplían el conjunto de modelos lineales generalizados con la adición de efectos aleatorios al estimador lineal [2], los cuales permiten modelizar observaciones correlacionadas. De igual forma que los modelos

lineales generalizados, los GLMM pueden ser formulados usando una secuenciación en tres partes [3], donde Y_{ij} representa la observación del j -ésimo individuo ($j = 1; 2; \dots; n_i$) dentro del i -ésimo conglomerado ($i = 1; 2; \dots; m$).

Este trabajo tiene como objetivo ajustar y describir un modelo mixto lineal generalizado apropiado para el análisis de la influencia que cada variable (geométricas y de regulación del sistema de agitación) ejerce sobre la velocidad interior del fluido. Teniendo en cuenta que estos modelos permiten que la variable dependiente no tenga una distribución normal, que las observaciones estén correlacionadas y que tenga una relación lineal con las variables independientes y covariables a través de una función de enlace determinada.

2. Materiales y métodos

2.1. Origen de los datos

Los datos provienen del ensayo llevado a cabo por [4] con el depósito de un pulverizador hidroneumático en el que midieron la velocidad del fluido (cm/s) con un velocímetro acústico Doppler 3D, en 32 puntos repartidos en varias posiciones de su interior, y variando la presión del sistema de agitación (8, 10, 12 bar), el número de boquillas activas (2, 4), y el nivel de llenado de la cuba (1000, 2000, 3000 L).

2.2. Secuenciación del modelo

La distribución condicional de cada Y_{ij} , dado un vector $q \times 1$ de efectos aleatorios b_i , pertenece a la familia exponencial de distribuciones (binomial, binomial negativa, Poisson, normal, gamma, entre otras). La $\text{Var}(Y_{ij}|b_i) = \phi v(E(Y_{ij}|b_i))$, donde $v(\cdot)$ es una función conocida para la varianza, una función de la media condicional $E(Y_{ij}|b_i)$, y ϕ es un parámetro escalar que puede ser conocido o ser necesario estimarlo. En adición, dado los efectos aleatorios b_i , se asume que los Y_{ij} son independientes entre sí, lo cual supone la asunción de independencia condicional.

La media condicional de Y_{ij} , que depende de los efectos fijos β y los efectos aleatorios b_i , se relaciona con el predictor lineal η_{ij} , a través de la aplicación de una función de enlace conocida, $g(\cdot)$, la cual es monótona y diferenciable [5], de la siguiente manera (Ecuación 1):

$$g\{E(y_{ij}|b_i; x_{ij}, z_{ij})\} = \eta_{ij} = x_{ij}^T \beta + z_{ij}^T b_i \quad (1)$$

donde x_{ij} y z_{ij} son dos vectores de covariables $p + 1$ dimensionales y q dimensionales, respectivamente.

Aunque cualquier función puede ser escogida para $g(\cdot)$, cada distribución perteneciente a la familia exponencial tiene una función de enlace especial llamada función de enlace canónica. Esta función es definida como esa función $g(\cdot)$ tal que (Ecuación 2):

$$g\{E(y_{ij})\} = \theta_i \quad (2)$$

donde θ_i es el parámetro de localización canónico.

En principio, cualquier distribución multivariada puede ser admitida para los b_i . En la práctica, es común asumir que los b_i tienen una distribución normal multivariada, con media cero y matriz de covarianza D de dimensiones $q \times q$. Adicionalmente los efectos aleatorios b_i son asumidos para ser independientes de las covariables X_i , en el caso de haberlas.

En este trabajo se asume como función de enlace canónica la identidad, asumiendo una distribución normal de los errores del modelo y de la variable objetivo en la formulación de los modelos para su comparación. Una vez elegido el mejor modelo y, teniendo en cuenta que la variable objetivo muestra valores de escala positivos que se desvían hacia valores positivos más

grandes, se ensayó hasta obtener la combinación “forma de distribución-función de enlace” que brindó una mayor bondad de ajuste. Se realizó una regresión lineal entre los valores predichos y los observados (Ecuación 3 y Figura 3), valorando su significación a través de un análisis de varianza del modelo de regresión, la constante y la pendiente.

3. Resultados y discusión

3.1. Análisis de la distribución de la variable

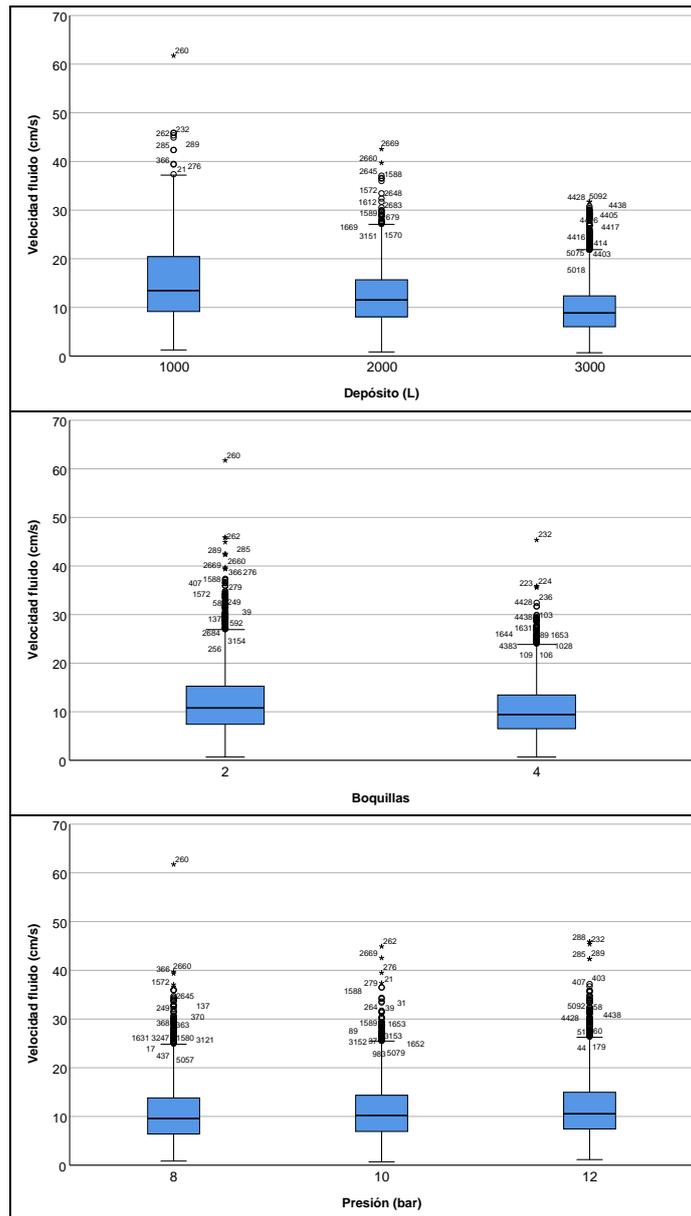


Figura 1. Representación de los diagramas de cajas para la velocidad del fluido (cm/s) observada en el interior del depósito del pulverizador según la cantidad de fluido, número de boquillas abiertas y presión en su interior.

El análisis de la variable velocidad del fluido (cm/s) en su conjunto ($n = 7200$) revela un valor promedio de $11,25 \pm 6,08$ cm/s dentro de un rango de valores desde los 0,69 cm/s hasta los 61,73 cm/s, poniendo en evidencia un desajuste con la distribución normal ($K-S = 0,080$; $p < 0,001$).

La representación de los datos según los parámetros operacionales de interés: llenado del depósito, número de boquillas abiertas y presión dentro del depósito, muestra una gran cantidad de datos extremos y atípicos en la cola derecha de la distribución, Asimetría: 1,242, y Curtosis: 2,680 (Figura 1). Con el objeto de elaborar modelos que describan lo observado en la realidad, ni se procederá a transformar la variable ni se tratarán y/o eliminarán los datos extremos y atípicos [6].

3.2. Comparación de modelos

Se procede a la construcción de modelos con un nivel creciente de complejidad cuya bondad de ajuste se compara (a) con el criterio de información Akaike (AICc) corregido [7], a pesar de tratarse de un elevado número de observaciones en relación con el número de parámetros del modelo (K), de modo que $n/K > 40$; y (b) con el criterio de información Bayesiano (BIC) [8] que modifica la puntuación de los parámetros del modelo respecto del AICc.

Tabla 1. Criterio de información Akaike corregido (AICc) y Bayesiano (BIC) para los modelos planteados y numerados.

Modelo	Nº	Criterio información Akaike corregido (AICc)	Criterio información Bayesiano (BIC)
VF = f(ninguna variable)	1	46.439,097	46.445,978
VF = f(Efectos Fijos: DP, NB, PI)	2	45.440,384	45.447,265
VF = f(Efectos Fijos: DP, NB, PI, SC)	3	44.656,549	44.663,429
VF = f(EF(DP, NB, PI, SC) + EA(AT, LD))	4	44.314,359	44.328,118
VF = f(EF(DP, NB, PI) + EA(SC))	5	43.681,507	43.695,267
VF = f(EF(DP, NB, PI) + EA(SC, AT))	6	43.684,507	43.698,267
VF = f(EF(DP, NB, PI) + EA(SC, AT, LD))	7	43.687,525	43.701,285
VF = f(EF(DP, NB, PI) + EA(AT, LD))	8	45.137,793	45.151,553

1: Sólo el intercepto (modelo nulo).

2: Efectos fijos (EF): llenado deposito (DP), número de boquillas (NB) y Presión interior (PI).

3: EF: DP, NB, PI y la Sección (SC).

4: EF: DP, NB, PI, SC, más la Altura (AT) y el Lado (LD) como efectos aleatorios (EA).

5: EF: DP, NB, PI, más SC como efecto aleatorio (EA).

6: EF: DP, NB, PI, más la SC y la Altura (AT) como efectos aleatorios.

7: EF: DP, NB, PI, más la SC, AT y el Lado (LD) como efectos aleatorios.

8: EF: DP, NB, PI, más la AT y LD como efectos aleatorios.

Como se observa en la Tabla 1 el mejor modelo, es decir, aquel que mejor representa la realidad mostrada por los datos, es el número 5, el que incluye los parámetros operacionales: llenado del depósito, número de boquillas abiertas y presión en el interior del depósito, como efectos fijos, y la sección de la forma geométrica del depósito como efecto aleatorio.

El modelo 5, siendo el que muestra un menor coeficiente de los criterios de información AIC y BIC, además muestra el mayor peso relativo Akaike (w_i ; Tabla 2), seguido de los modelos 6 y 7 respectivamente, con una plausibilidad menor y decreciente en lo relativo a su ajuste con los datos observados.

Tabla 2. Incrementos del criterio de información Akaike corregido de cada modelo respecto del mejor modelo (mínimo) y la verosimilitud relativa de cada modelo expresada a través de su peso relativo Akaike (w_i).

Modelo	$\Delta_i = AIC_i - AIC_{\min}$	$w_i = \exp(-0,5 \cdot \Delta_i) / \sum(\exp(-0,5 \cdot \Delta_i))$
1	2.757,59	0,00000E+00
2	1.758,88	0,00000E+00
3	975,04	1,47120E-212
4	632,85	2,97363E-138
5	0,00	7,85872E-01
6	3,00	1,75352E-01
7	6,02	3,87757E-02
8	1.456,29	0,00000E+00

El modelo 5, junto a los modelos 6 y 7, son los que realizan una estima de la varianza residual menor dentro del conjunto de modelos (Tabla 3).

Tabla 3. Valores de la estima de la varianza residual de cada modelo, su error estándar, el valor del estadístico, el p-valor de la prueba de significación y los valores de los límites inferior y superior del intervalo de confianza al 95% de la estima de la varianza residual.

Modelo	Estima Varianza Residual ($\hat{\sigma}_r$)	Error estándar	Z	P	Límite Inferior	Límite Superior
1	37,017	0,617	59,996	< 0,001	35,827	38,246
2	32,202	0,537	59,975	< 0,001	31,167	33,272
3	28,868	0,481	59,962	< 0,001	27,940	29,827
4	27,405	0,457	59,933	< 0,001	26,523	28,316
5, 6, 7	24,761	0,414	59,845	< 0,001	23,964	25,586
8	30,743	0,513	59,945	< 0,001	29,754	31,765

Aquellos modelos en los que no se ha incluido la Sección (SC), ni en los efectos fijos ni en los aleatorios (modelos 1, 2 y 8), han mostrado los menores pesos relativos (w_i) en el conjunto, así como las mayores estimas de la varianza residual. Teniendo en cuenta que los valores más pequeños de varianza residual corresponden a los modelos que mejor se ajustan a lo descrito por las observaciones [9], se pone en evidencia la importancia de la SC, como descriptor de la geometría del depósito del pulverizador, a la hora de evaluar la velocidad del fluido en su interior, y más aún, la necesidad de su inclusión como efecto aleatorio y no como efecto fijo.

En esta clase de modelos generalizados de efectos mixtos se ha supuesto la distribución paramétrica normal [10] para los efectos aleatorios, fundamentalmente por razones convencionales [11], entendiendo que, a pesar de haber elegido cuatro secciones transversales para la localización de los puntos de control dentro del depósito del pulverizador, la sección podría tomar cualquier valor desde un extremo hasta el otro.

No se aporta más información acerca de la forma de la distribución conjunta de los efectos aleatorios (SC, AT, LD), de modo que el supuesto de la distribución normal para estos efectos aleatorios en conjunto no ha sido evaluada directamente en los modelos 6, 7 y 8 [12]. Su no elección soslaya en parte la preocupación natural por la especificación incorrecta de los efectos aleatorios en los modelos GLMM [13], y su posible influencia en la inferencia de los efectos fijos [14].

3.3. Descripción del mejor modelo

Los datos se estructuran a partir de los puntos de control, como unidades de observación. Son independientes entre sí. En ellos se han llevado a cabo las mediciones según los distintos niveles de los efectos fijos: llenado del depósito, presión y abertura de pares de boquillas; en cuatro secciones y a tres alturas y distancias del perímetro (lados).

La variable objetivo (dependiente) es la velocidad del fluido en el interior del depósito. Como ya se advirtió anteriormente, atendiendo a la forma de la distribución de la variable objetivo y, fruto de los diversos ensayos realizados, la combinación que mejor se ajusta a lo observado resulta de una distribución tipo Gamma [15] con una función de enlace logarítmica ($f(x) = \log(x)$). Al tener un elevado número de observaciones y una estructura de covarianza sencilla como la identidad escalada (por tener sólo un efecto aleatorio), se eligió el método residual (fijo para todas las pruebas) para el cálculo de los grados de libertad. En el cálculo de la matriz de covarianzas de las estimaciones de los parámetros se seleccionó la estimación robusta, a la vista de los datos extremos y atípicos cuyo efecto podría vulnerar la premisa de homoscedasticidad del modelo [2]. De este modo, los indicadores de bondad de ajuste son los siguientes: (a) AICc = 8.371,421; (b) BIC = 8.385,182; y (c) $\hat{S}_r = 0,183$; $p < 0,001$.

Tabla 4. Estadístico de contraste y p-valor para el Modelo completo y cada uno de los efectos fijos (EF). Medias estimadas para los distintos niveles de los efectos fijos. Letras distintas ponen en evidencia diferencias significativas ($p < 0,05$).

Origen	F	p	Medias estimadas Velocidad del Fluido (cm/s)							
			Niveles de los efectos fijos							
Modelo	130,423	< 0,001	1000 L	8 bar	2	2000 L	10 bar	4	3000 L	12 bar
EF DP	38,865	< 0,001	15,048c			12,698b			9,296a	
EF PI	28,052	< 0,001		11,511a			12,073a			12,782b
EF NB	6,403	0,011			12,944b			11,331a		

Como se observa en la Tabla 4 el modelo resulta significativo, así como los tres EF, siendo el llenado del depósito (DP) el que más influencia tiene sobre la velocidad del fluido, seguido de la presión interior (PI) del mismo y, muy por detrás, pero también significativamente, el número de boquillas abiertas (NB). Cabe señalar que son significativamente distintas las velocidades del fluido en los tres niveles estudiados, disminuyendo según se va llenando el depósito; en lo relativo a la presión interior, la velocidad del fluido no es significativamente distinta a 8 bar o 10 bar, pero sí a estas dos presiones es significativamente menor que con una presión de 12 bar; y, por último, la velocidad del fluido disminuye significativamente cuando se duplica el número de boquillas abiertas (Figura 2) debido en este caso a la ubicación geométrica de las mismas, enfrentadas entre sí.

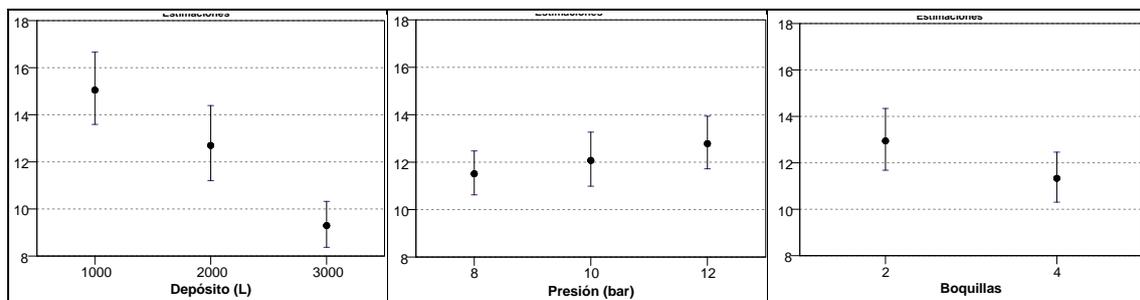


Figura 2. Medias estimadas por el modelo para la velocidad del fluido (cm/s) e intervalos de confianza al 95% según el nivel de llenado del depósito, la presión interior y el número de boquillas abiertas.

En el modelo elegido el residuo tipificado (Pearson) se distribuye $f \sim (0,00; 0,995)$, con una asimetría = 0,545 y curtosis = 0,417. Su distribución es significativamente distinta de una normal (K-S = 0,034; $p < 0,001$). Los valores predichos reproducen los mayores valores en el sentido positivo de la variable objetivo (VF), sobre estimando las predicciones desde los 2 cm/s hasta los 35 cm/s, pero reduciendo la velocidad máxima predicha al menos en 15 cm/s (Figura 3).

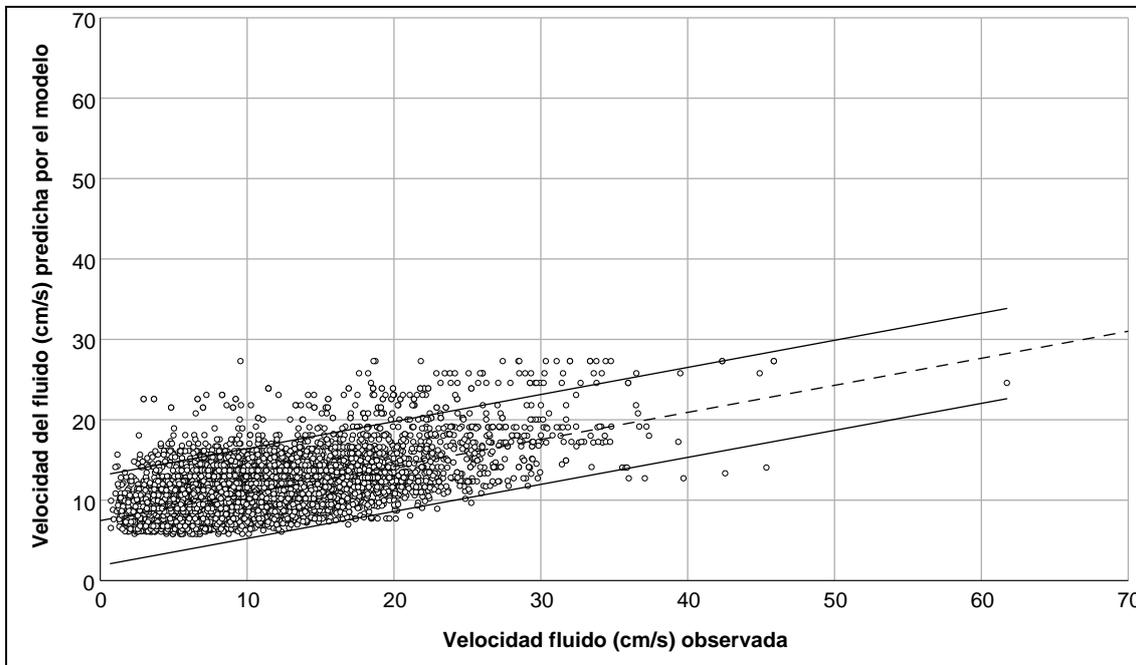


Figura 3. Regresión lineal entre la velocidad del fluido (cm/s) predicha por el modelo y la observada durante el ensayo.

La regresión entre los valores predichos y los observados (Figura 3) permite formular una relación lineal significativa ($p < 0,001$) con un coeficiente de determinación ajustado $R^2_{aj} = 0,340$ (Ecuación 3).

$$\widehat{VF} \text{ (cm/s)} = 7,459 + (0,336 \times VF \text{ (cm/s)}) \quad (3)$$

4. Conclusiones

El mejor modelo de los ensayados incluyó como efectos fijos el nivel de líquido en el interior del depósito, la presión interior y el número de boquillas abiertas, y como efecto aleatorio, únicamente la sección de la cuba, resultando irrelevantes la altura y el lado de los puntos de control.

A través de este modelo mixto lineal generalizado se han podido evaluar conjuntamente efectos fijos y aleatorios, de modo que el efecto fijo que más importancia ejerce sobre la velocidad del fluido es el llenado del depósito, seguido de la presión interior, para terminar con el número de boquillas abiertas; parámetros instrumentales para el operador.

Este modelo suaviza y modera los datos extremos y atípicos observados en el ensayo, pero no elimina totalmente su influencia, de modo que sobre estima las predicciones, pero limita los valores más elevados.

Referencias

1. Armenante P.M., Luo C., Chou C., Fort I., Medek J. Velocity profiles in a closed unbaffled vessel: comparison between experimental LDV data and numerical CFD predictions. *Chemical Engineering Science*, 1997, 52(20), 3483-3492.
2. McCulloch C.E., Searle, S.R. *Generalized, Linear, and Mixed Models*. 2001. New York, USA: John Wiley and Sons, Inc.
3. Fitzmaurice G.M., Laird N.M., Ware J.H. *Applied Longitudinal Analysis*. 2012. Boston, Massachusetts, USA: John Wiley and Sons.
4. García-Ramos F.J., Badules J., Boné A., Gil E., Aguirre A.J., Vidal M. Application of an Acoustic Doppler Velocimeter to Analyse the Performance of the Hydraulic Agitation System of an Agricultural Sprayer. *Sensors*, 2018, 18(11), 3715.
5. Gad A.M., El Kholly R.B. Generalized Linear Mixed Models for Longitudinal Data. *International Journal of Probability and Statistics*, 2012, 1(3), 41-47.
6. Krishnaiah, P.R. (Ed.). *Handbook of Statistics 1: Analysis of Variance*. 1980. Amsterdam, Netherlands: North-Holland Publishing Company.
7. Akaike H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 1974, 19(6), 716-723.
8. Burnham K.P., Anderson D.R. Multimodel inference: understanding AIC and BIC in Model Selection. *Sociological Methods and Research*, 2004, 33, 261-304.
9. Litière S., Alonso A., Molenberghs G. The impact of a misspecified random-effects distribution on the estimation and the performance of inferential procedures in generalized linear mixed models. *Statistics in Medicine*, 2008, 27(16), 3125-3144.
10. Abad A.A., Litière S., Molenberghs G. Testing for misspecification in generalized linear mixed models. *Biostatistics*, 2010, 11(4), 771-786.
11. Tsonaka R., Rizopoulos D., Verbeke G., Lesaffre, E. Nonignorable Models for Intermittently Missing Categorical Longitudinal Responses. *Biometrics*, 2010, 66(3), 834-844.
12. Xiang L., Yau K.K.W., Lee A.H. The robust estimation method for a finite mixture of Poisson mixed-effect models. *Computational Statistics and Data Analysis*, 2012, 56(6), 1994-2005.
13. Huang X. Diagnosis of Random-Effect Model Misspecification in Generalized Linear Mixed Models for Binary Response. *Biometrics*, 2009, 65(2), 361-368.
14. Komárek A., Lesaffre E. Generalized linear mixed model with a penalized Gaussian mixture as a random effects distribution. *Computational Statistics and Data Analysis*, 2008, 52(7), 3441-3458.
15. Montgomery D.C., Runger G.C. *Applied Statistics and Probability for Engineers*, (3rd ed.). 2003. New York, USA: John Wiley & Sons, Inc.