

# ESTIMACIÓN DEL RENDIMIENTO DE ORELLANA MEDIANTE MODELOS GAMLSS<sup>a</sup>

## ESTIMATION OF ORELLANA YIELD THROUGH GAMLSS MODELS

GIPSY NARANJO DUEÑAS<sup>b</sup>, EDGAR MONSALVE LUGO<sup>c</sup>, FREDDY HERNÁNDEZ  
BARAJAS<sup>d</sup>

Recibido 20-11-2016, aceptado 08-03-2017, versión final 28-03-2017.

Artículo Investigación

**RESUMEN:** En la literatura científica se pueden encontrar estudios sobre la influencia de variables ambientales en la producción de hongos comestibles, sin embargo, esos estudios han utilizado modelos estadísticos donde se asume que la variable producción se distribuye en forma normal, supuesto que no siempre se cumple. En este artículo se usan los modelos GAMLSS (*Generalized Additive Model for Location Scale and Shape*) para estudiar la influencia de las variables humedad, temperatura, tiempo de aireación, tipo de sustrato y cantidad de sustrato sobre la producción del hongo orellana. Los modelos GAMLSS permiten que el investigador asuma distribuciones estadísticas para la variable respuesta diferentes a la normal y que se puedan modelar todos los parámetros en función de las covariables. Al aplicar GAMLSS a los datos de orellana se encontró que sólo las variables temperatura, tiempo de aireación y tipo de sustrato tienen una influencia significativa en la producción de orellana. En particular, se encontró que a mayor temperatura y tiempo de aireación la producción de orellana aumenta, adicionalmente, se estimó que la producción de orellana se duplica si la siembra se realiza en sustrato de higuera molida. Los resultados de este artículo sirven para cuantificar los efectos de las variables ambientales en la producción industrial de orellana y para determinar combinaciones óptimas de los factores que maximicen la producción.

**PALABRAS CLAVE:** GAMLSS; regresión; modelación; distribución gamma.

**ABSTRACT:** The influence of environmental variables on the production of edible mushrooms has been previously documented in the scientific literature. However, all these studies have in common the use of statistical models where the variable production is distributed in a normal way, although it is well known this is not always true. In this paper, we have used the GAMLSS (*Generalized Additive Model for Location Scale and Shape*) to study the influence of variables such as humidity, temperature, aeration time, type of substrate and amount of substrate on the production of the Orellana mushroom. GAMLSS allows to assume statistical

<sup>a</sup>Naranjo Dueñas, G., Monsalve Lugo, E. & Hernández Barajas, F. (2016). Estimación del rendimiento de Orellana mediante modelos Gamlss. *Revista de la Facultad de Ciencias*, 6 (1), 67–82. DOI: 10.15446/rev.fac.cienc.v6n1.61119

<sup>b</sup>Estudiante de Estadística, Universidad Nacional de Colombia-Sede Medellín. Correo: gtnaranjod@unal.edu.co.

<sup>c</sup>Estudiante de Estadística, Universidad Nacional de Colombia-Sede Medellín. Correo: edmonsalvel@unal.edu.co.

<sup>d</sup>Profesor asistente. Universidad Nacional de Colombia-Sede Medellín. Correo: fhernanb@unal.edu.co.

distributions for the response variable different from normal and enable modeling all parameters according to variables. When applying GAMLSS to the Orellana data, it was found that only the variables temperature, aeration time and type of substrate were influential on the Orellana production. In particular, the analysis revealed that at higher temperatures and aeration times the production of Orellana rises. Additionally, it was estimated that the production of Orellana would double if ground husk were used. The impact of results derived from this investigation can be utilized not just to quantify the effects of environmental variables on the industrial production of Orellana, but also to determine optimal factor combinations to maximize production.

**KEYWORDS:** GAMLSS; regression; modeling; gamma distribution.

## 1. INTRODUCCIÓN

En los últimos años el consumo de hongos ha impulsado la industria de la fungicultura debido a sus propiedades nutricionales. Uno de estos hongos comestible es el *Pleurotus Ostreatus* u orellana que se desarrolla fácilmente en residuos como troncos, ramas, bagazo de caña, entre otros. El cultivo de orellana se realiza en dos etapas. La primera etapa es la incubación que consiste en la preparación del sustrato, siembra e inducción de primordios; la segunda etapa es la fructificación y se realiza entre 7 y 10 días después de la primera etapa. En la Figura 1 se ilustra la forma de cultivo de orellana en una bolsa con sustrato.



Figura 1: Orellana lista para el corte. Tomada de <http://www.aquinohaygeranios.com/>.

La importancia de la orellana en la alimentación es debido a su alto contenido de carbohidratos, alrededor de un 50% en peso, entre 5 y 9% de fibra, bajo contenido de grasa, posee un efecto antioxidante y alto contenido de polisacáridos que ayudan al sistema inmunológico, la orellana es altamente nutritiva y beneficiosa para el consumo humano, (Sierra & Orozco, 2014). Cardona (2011) afirma que el cultivo casero de hongos en países de Centroamérica ha tenido gran aceptación entre las familias ya que éste contribuye a la producción y disponibilidad de alimentos, no requiere altas inversiones de dinero, es muy fácil de manejar, la demanda es creciente y puede ser una fuente de ingresos extra para las familias. Recientemente, Cervantes (2015) planteó un modelo para la producción industrializada de orellana en Colombia al interior de las viviendas de familias

en condición de vulnerabilidad o víctimas de desplazamiento forzado. Este modelo permite que se usen apropiadamente los recursos de cada región para beneficiar a las familias con la producción de orellana, ya sea por el consumo interno o por ventas del producto.

Estudiar la relación entre los factores de producción para mejorar los rendimientos de orellana es un aspecto importante tanto para el cultivo a escala familiar e industrial. El problema de estimar la producción de hongos en función de variables ha sido abordado en la literatura científica y a continuación se mencionan algunos reportes y las herramientas estadísticas usadas para estudiar esa relación. Hernández *et al.* (2006) usaron pruebas de hipótesis para determinar la influencia del tipo de sustrato (capacho de uchuva, cáscara de arveja y tusa de mazorca) sobre el crecimiento y producción de orellana en Colombia. Bonet *et al.* (2008) usaron modelos lineales mixtos para predecir la producción de tres tipos de hongos silvestres en una región de España en función de la pendiente del terreno, de la altura del nivel mar del terreno y del área de la parcela. Velasco *et al.* (2010) usaron modelos de regresión múltiple para pronosticar la producción de hongos silvestres por parcela en una región de México usando las covariables número de árboles por parcela, diámetro promedio y altura promedio de los árboles en la parcela y cobertura promedio de la parcela. Cañedo (2012) utilizó análisis de varianza para estudiar la influencia del tipo de sustrato (maíz, trigo y frijol) y tres cantidades de sustrato (2, 4 y 6 kg) sobre la producción de orellana en México. Martínez *et al.* (2012) usaron modelos no lineales para predecir la producción anual de tres tipos de hongos silvestres (*Ectomycorrhizal*, *L. group deliciosus* y *B. edulis yield*) en función de las precipitaciones en la región, la altura y edad de los árboles, la altitud, inclinación, densidad y pendiente de la parcela. Tahvanainen *et al.* (2016) usaron modelos mixtos no lineales para predecir la producción de dos tipos de hongos silvestres (*Boletus edulis*, *Lactarius spp*) en Noruega y Finlandia, en función de las precipitaciones, la altura y edad de los árboles, la altitud, inclinación, densidad y pendiente de la parcela.

En este trabajo se abordan estadísticamente los datos de producción de orellana obtenidos de un estudio observacional realizado en la empresa productora de orellana Monterra Productora y Comercializadora Agroindustrial Colombiana S.A.S. ubicada en el municipio de La Palma en Cundinamarca. Los datos se abordan por medio de la metodología GAMLSS (*Generalized Additive Model for Location Scale and Shape*) para cuantificar los efectos de las variables de producción humedad relativa, tiempo de aireación, temperatura, tipo de sustrato y cantidad de sustrato sobre el rendimiento de orellana obtenida de cada bolsa sembrada. Adicionalmente se derivan expresiones matemáticas para la producción media y para la varianza de la producción de orellana.

En la literatura estadística se reporta el uso de modelos GAMLSS en diferentes áreas, algunos de esos reportes son: de Castro *et al.* (2010), usaron la distribución binomial negativa para modelar el tiempo de supervivencia con censura en función de las características de un conjunto de 205 pacien-

tes observados después de la operación para la eliminación de melanoma maligno; Gilchrist *et al.* (2011), por medio de GAMLSS modelaron la probabilidad de padecer una enfermedad respiratoria y la duración en el hospital hasta la recuperación en función de un conjunto de covariables para un grupo personas de la tercera edad; Sandercock *et al.* (2012), construyeron percentiles con GAMLSS con el objetivo de medir la capacidad cardiorespiratoria en pruebas atléticas de 20 metros con niños de 10 a 16 años; Tong *et al.* (2013), aplicaron GAMLSS para modelar las pérdidas en préstamos asumiendo una distribución mixta gamma inflada con ceros (ZAGA) para las pérdidas y usando la información variables relacionadas a los préstamos; Scandroglio *et al.* (2013), usaron GAMLSS para estimar el VaR (*Value-at-Risk*) y ES (*Expected Shortfall*) de los precios del petróleo Brent y WTI y Hernández *et al.* (2015) aplicaron GAMLSS en el área agroindustrial para estudiar el efecto del pH y el tiempo en la obtención de celulosa a partir de desechos de banano.

El presente documento está organizado así: en la sección 2 se hace una descripción breve de los modelos GAMLSS utilizados; en la sección 3 se describe la forma de obtención de los datos, las variables involucradas en el estudio y se presenta el análisis descriptivo; en la sección 4 se muestra la forma como se aplicaron los modelos GAMLSS al problema, los diferentes modelos considerados, los criterios para la elección del mejor modelo y los resultados del mejor modelo; en la sección 5 se presentan las expresiones matemáticas que resumen los efectos y relación entre las variables de producción y el rendimiento de orellana, por último en la sección 6 están las conclusiones del artículo.

## 2. MODELOS GAMLSS

Los modelos GAMLSS (*Generalized Additive Model for Location Scale and Shape*) propuestos por Rigby & Stasinopoulos (2005), son de gran utilidad ya que permiten modelar todos los parámetros de la variable de interés en función de covariables. En la práctica, muy pocas veces la variable respuesta cumple el supuesto de normalidad requerido para el ajuste de modelos lineales, los modelos GAMLSS son útiles porque permiten elegir la distribución para la variable respuesta entre más de 80 distribuciones diferentes. Los modelos GAMLSS asumen que las observaciones  $y_i$  para  $i = 1, \dots, n$  son independientes con función de densidad de probabilidad  $f(y_i|\boldsymbol{\theta}_i)$  donde  $\boldsymbol{\theta}_i = (\mu_i, \sigma_i, \nu_i, \tau_i)^\top$  es el vector de parámetros. Los primeros dos parámetros,  $\mu_i$  y  $\sigma_i$ , son los parámetros de localización y escala mientras que  $\nu_i, \tau_i$  corresponden a los parámetros de forma; los modelos GAMLSS permiten que cada uno de los parámetros  $\mu_i, \sigma_i, \nu_i, \tau_i$  sean explicados en función de covariables. Los modelos GAMLSS permiten incluir en la modelación de los parámetros efectos fijos, términos no paramétricos (funciones suavizadas de las covariables) y efectos aleatorios, la

estructura general de los modelos GAMLSS es la siguiente:

$$g_1(\boldsymbol{\mu}) = \boldsymbol{\eta}_1 = \mathbf{X}_1\boldsymbol{\beta}_1 + \sum_{j=1}^{J_1} \mathbf{Z}_{j1}\boldsymbol{\gamma}_{j1} \quad (1)$$

$$g_2(\boldsymbol{\sigma}) = \boldsymbol{\eta}_2 = \mathbf{X}_2\boldsymbol{\beta}_2 + \sum_{j=1}^{J_2} \mathbf{Z}_{j2}\boldsymbol{\gamma}_{j2} \quad (2)$$

$$g_3(\boldsymbol{\nu}) = \boldsymbol{\eta}_3 = \mathbf{X}_3\boldsymbol{\beta}_3 + \sum_{j=1}^{J_3} \mathbf{Z}_{j3}\boldsymbol{\gamma}_{j3} \quad (3)$$

$$g_4(\boldsymbol{\tau}) = \boldsymbol{\eta}_4 = \mathbf{X}_4\boldsymbol{\beta}_4 + \sum_{j=1}^{J_4} \mathbf{Z}_{j4}\boldsymbol{\gamma}_{j4} \quad (4)$$

donde  $g_k(\cdot)$  es una función de enlace conocida para  $k = 1, \dots, 4$ ;  $\boldsymbol{\mu}$ ,  $\boldsymbol{\sigma}$ ,  $\boldsymbol{\nu}$ ,  $\boldsymbol{\tau}$  y  $\boldsymbol{\eta}_k$  son vectores  $n$ -dimensionales; las  $\mathbf{X}_k$  son matrices de diseño conocidas de orden  $n \times J'_k$  asociadas a los efectos fijos  $\boldsymbol{\beta}_k$  de  $J'_k \times 1$ ; y las  $\mathbf{Z}_{jk}$  son matrices de diseño conocidas de orden  $n \times q_{jk}$  asociadas a los efectos aleatorios  $\boldsymbol{\gamma}_{jk}$  de  $q_{jk} \times 1$  con distribución normal multivariada. La cantidad  $J'_k$  representa el número de covariables usadas en la parte fija del predictor lineal  $\boldsymbol{\eta}_k$ , mientras que  $J_k$  representa el número de efectos aleatorios en  $\boldsymbol{\eta}_k$ . El modelo dado en (1) a (4) puede ser resumido en forma compacta así:

$$g_k(\boldsymbol{\theta}_k) = \boldsymbol{\eta}_k = \mathbf{X}_k\boldsymbol{\beta}_k + \sum_{j=1}^{J_k} \mathbf{Z}_{jk}\boldsymbol{\gamma}_{jk} \quad (5)$$

Los modelos GAMLSS permiten que la distribución de la variable respuesta  $y_i$  sea continua, discreta o mixta, existen varias parametrizaciones para una misma distribución y los detalles de las distribuciones y parametrizaciones usadas en GAMLSS pueden ser consultados en Stasinopoulos, Rigby & Akantziliotou (2008). Los modelos GAMLSS se puede aplicar fácilmente por medio del paquete `gamlss` disponible en R Core Team (2017).

### 3. ANÁLISIS DESCRIPTIVO DE VARIABLES Y DATOS

Los datos analizados corresponden a un estudio observacional con 57 unidades muestrales. Cada unidad muestral corresponde a una bolsa con cáscara de higuera en la cual se sembró orellana. De las 57 bolsas, 37 contenían cáscara de higuera entera y las restantes 20 cáscara de higuera molida. El contenido de cascará de higuera por bolsa fue de 4, 5 o 10 kilogramos; la mayoría de las bolsas estuvieron almacenadas en recintos con temperatura controlada a 19 °C o 24 °C y la humedad relativa del recinto fue también controlada con valores de 58 %, 60 % y 85 %. Todos los días las unidades muestrales fueron sacadas del recinto para recibir una cantidad de luz solar que fue de 4 o 6 horas. Aproximadamente 14 días después de haber iniciado la fructificación, la orellana se retiró de la bolsa y se pesó en una balanza para medir el rendimiento en gramos. La definición y unidades de las variables monitoreadas en el estudio se presentan a continuación:

1. Sustrato: tipo de sustrato o cáscara de higuierilla usada en cada bolsa, los niveles de esta variable son molida o entera.
2. Peso: peso o cantidad (en kg) de sustrato utilizado en cada bolsa.
3. Humedad: humedad relativa (en %) del recinto donde se encontraba la bolsa con orellana.
4. Temperatura: temperatura (en °C) del recinto donde se encontraba la bolsa con orellana.
5. Tiempo: tiempo (en horas) de aireación por día al cual fueron sometidas las bolsas fuera del recinto.
6. Rendimiento: corresponde al rendimiento (en gramos) de orellana cultivada por bolsa.

En la Figura 2 se presenta la densidad para la variable respuesta rendimiento de orellana, de esta figura se observa que el rendimiento tiene una distribución asimétrica con sesgo a la derecha. Las marcas o rayas en el eje horizontal representan cada una de las 57 observaciones, el 80% de los rendimientos centrales se encuentran entre 44.2 y 1150.2 gramos (percentiles 10 y 90).

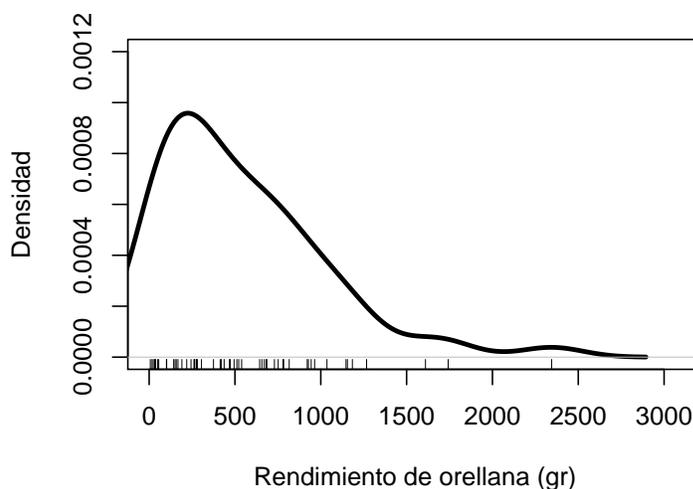


Figura 2: Densidad para el rendimiento de orellana por bolsa. Fuente: Elaboración propia.

Para explorar las distribuciones estadísticas que mejor explican el patrón observado del rendimiento de orellana se utilizó un procedimiento de bondad de ajuste marginal, es decir, sin incluir covariables. En la Figura 3 se presenta el histograma para el rendimiento de orellana acompañado de las cuatro densidades (línea obscura) que mejor se ajustaron y que en su orden fueron exponencial, Weibull parametrización 3, gamma y Pareto parametrización 2. En cada panel de la Figura 3 está

el nombre de la distribución con los parámetros estimados, las expresiones matemáticas para las funciones de densidad de estas distribuciones se pueden consultar en Stasinopoulos, Rigby & Akantziliotou (2008). De esta figura se observa que las cuatro distribuciones logran describir el patrón de variación del rendimiento de orellana, la distribución exponencial lo hace usando un sólo parámetro ( $\mu$ ) mientras que la Weibull 3, gamma y Pareto 2 lo hacen por medio de dos parámetros ( $\mu$  y  $\sigma$ ).

En la Figura 4 se presentan los boxplots que relacionan el rendimiento de orellana con las covariables tiempo de aireación, humedad, temperatura, cantidad de sustrato y tipo de sustrato. De la parte superior de la figura se observa que a mayor tiempo de aireación y a mayor temperatura, el rendimiento de orellana por bolsa aumenta; en la parte superior se observan dos boxplots como líneas debido a que sólo se tiene un registro a humedad de 58 % y un registro a temperatura de 21 °C. En relación a la humedad se nota que los mayores rendimientos sucedieron cuando la humedad fue de 85 %.

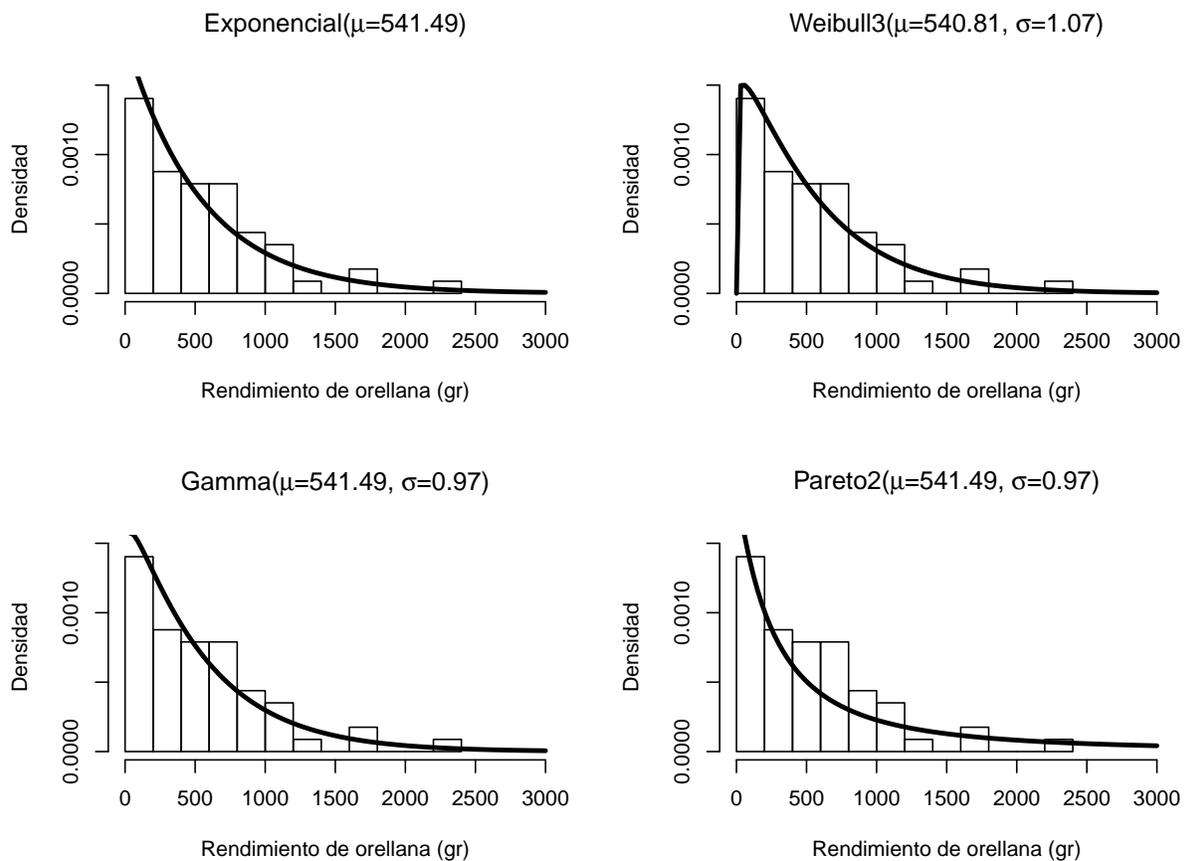


Figura 3: Histograma para el rendimiento de orellana acompañado de las cuatro densidades de probabilidad (líneas oscuras) que mejor se ajustan a la variable respuesta. Fuente: Elaboración propia.

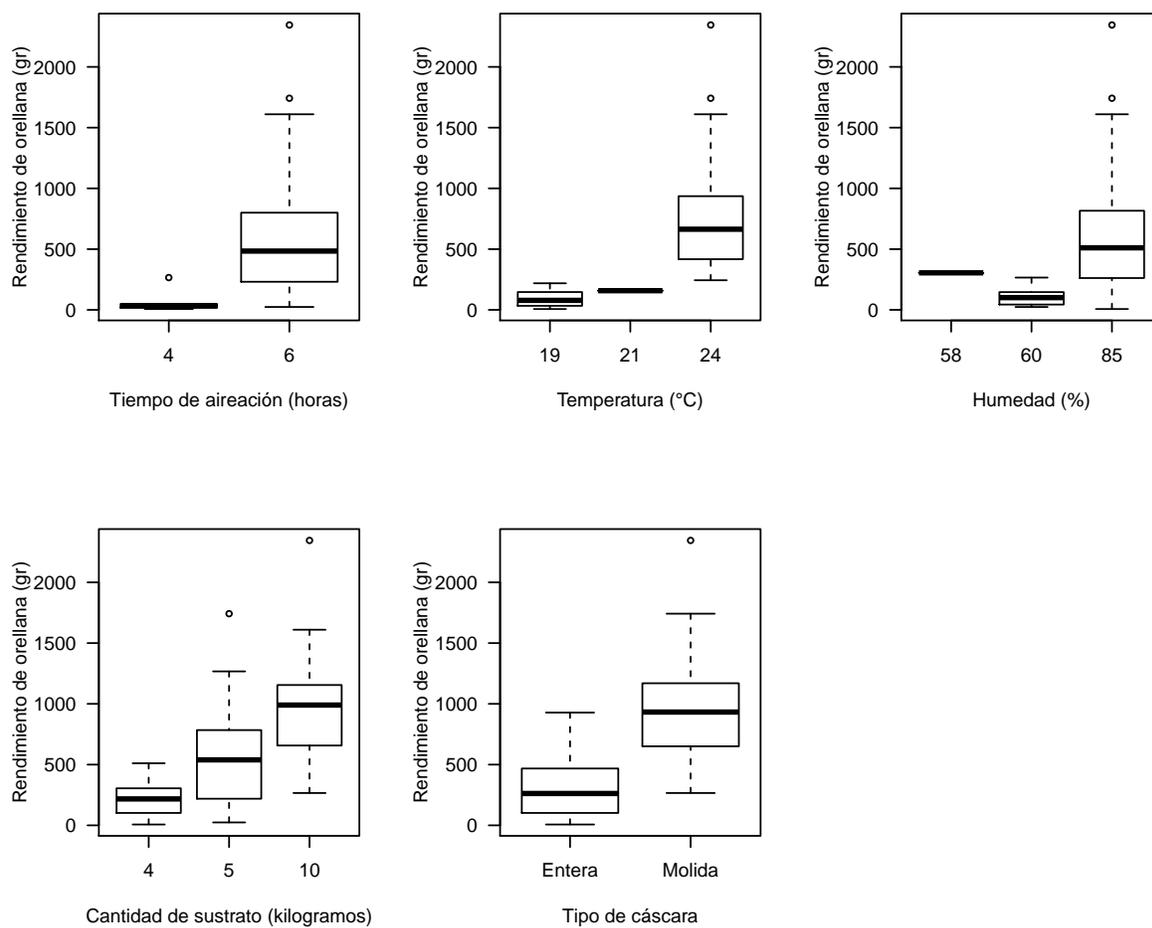


Figura 4: Boxplot para el rendimiento de orellana dado el tiempo de aireación, temperatura, humedad, cantidad de sustrato y el tipo de sustrato o cáscara. Fuente: Elaboración propia.

En la parte inferior de la Figura se muestran los boxplots que relacionan el rendimiento con la cantidad de sustrato y el tipo de sustrato usado en las bolsas donde fueron plantados los primordios. De esta figura se observa que a medida que las bolsas donde se siembran los primordios contienen mayor cantidad de sustrato, el rendimiento obtenido es mayor; se observa también que al usar cáscara molida el rendimiento de orellana aumenta frente al uso de cáscara entera.

## 4. METODOLOGÍA

En esta sección se presenta la aplicación de los modelos GAMLSS para explicar el rendimiento de la orellana en función de las covariables temperatura, humedad relativa, tiempo de aireación, peso de la bolsa y sustrato (variable cualitativa).

Adicional a las cuatro distribuciones (exponencial, Weibull 3, gamma y Pareto 2) que mejor se ajustaban marginalmente y que fueron identificadas en la Sección 3, se consideraron otras siete distribuciones con soporte en  $\mathbb{R}^+$  para modelar el rendimiento de orellana, esas distribuciones fueron: lognormal (LOGNO), inversa gaussiana (IG), gamma generalizada (GG), Box-Cox Cole and Green (BCCGo), inversa gaussiana generalizada (GIG), Box-Cox  $t$  (BCTo), Box-Cox power exponencial (BCPEo). De las distribuciones consideradas sólo la exponencial tiene un parámetro ( $\mu$ ), otras son de dos parámetros ( $\mu, \sigma$ ) y otras tienen hasta 3 parámetros ( $\mu, \sigma, \nu$ ), para conocer los detalles de las funciones de densidad, valor esperado y varianza de las distribuciones usadas se recomienda consultar Stasinopoulos, Rigby & Akantziliotou (2008).

En cada uno de los once modelos (uno por cada distribución considerada) los predictores lineales ( $\eta$ ) incluyeron sólo efectos fijos de la variable cualitativa, términos lineales y cuadráticos de las variables cuantitativas y de las interacciones entre estas mismas variables cuantitativas; también se aplicó un proceso de selección de variables en cada predictor lineal para elegir sólo términos significativos en la modelación.

Para comparar modelos se puede utilizar el *Akaike information criterion* (*AIC*) propuesto por Akaike (1974), que sirve para medir la calidad relativa de los modelos estadísticos, la expresión para calcular el indicador es

$$AIC = -2\hat{l} + 2df,$$

donde  $\hat{l}$  corresponde al valor de log-verosimilitud estimada para el modelo y  $df$  corresponde al número de parámetros estimados del modelo. Modelos anidados pueden ser comparados por medio del *global deviance* (*GD*) dado por la expresión

$$GD = -2\hat{l},$$

y modelos no anidados por medio del *Generalized Akaike information criterion* (*GAIC*) propuesto

por Akaike (1983) y dado por la expresión

$$GAIC = -2\hat{l} + \# df,$$

siendo  $\#$  el valor de penalidad por cada parámetro adicional en el modelo; cuando  $\# = 2$ , el *GAIC* coincide con el *AIC* y el *Schwarz Bayesian criterion (SBC)* propuesto por Schwarz (1978), se da cuando el valor de penalidad es  $\# = \log(n)$ , donde  $n$  es el número de observaciones del modelo; siempre el modelo elegido es aquel modelo con el menor valor de cualquiera de los criterios de información anteriores. En este trabajo se usó el *SBC* para comparar los modelos construidos.

En el Tabla 1 se presentan los valores de *SBC* para los tres mejores modelos de los once considerados, de esta tabla se aprecia que los modelos con variable respuesta gamma generalizada (GG) y gamma (GA) tienen valores de *SBC* muy similares.

Tabla 1: *SBC* para los tres mejores modelos.

Modelo	Distribución	Grados de libertad	<i>SBC</i>
1	GG	8	765.40
2	GA	7	765.73
3	GIG	8	768.23

En esta aplicación se usaron los residuales *RQR (Randomized Quantile Residuals)* propuestos por Dunn & Smyth (1996) para modelos de regresión donde la variable respuesta no sigue una distribución normal y donde no es posible usar los residuales usuales ( $y_i - \hat{y}_i$ ); los residuales *RQR* se definen como  $r_i = \Phi^{-1}(F(y_i | \hat{\theta}))$  donde  $F$  representa la función de distribución acumulada asumida para la variable respuesta y  $\Phi$  es la función de distribución acumulada para una normal estándar. Este tipo de residuales sigue una distribución normal si los parámetros estimados  $\hat{\theta}$  del modelo fueron bien estimados (Dunn & Smyth, 1996). Para analizar los residuales  $r_i$ , Buuren & Fredriks (2001) propusieron el gráfico de gusano o *Worm plot* que es una modificación del gráfico cuantil cuantil (qqplot), en este nuevo gráfico los valores del eje vertical se reemplazan por la diferencia (*Deviation*) entre la coordenada  $y$  y la coordenada  $x$  del qqplot usual, esta modificación crea un qqplot sin tendencia estocástica y entre más plana sea la secuencia de puntos indica que el modelo está bien ajustado. Buuren (2001) puede ser consultado para conocer el código en R para la creación de este gráfico.

En la Figura 5 se presenta el *Worm plot* para cada uno de los modelos relacionados en el Tabla 1, de esta figura se observa que los residuales para los tres modelos quedan ubicados en la parte central y que no invaden las hipérbolas, esto indica un buen ajuste de los modelos. Como los tres modelos tiene *BIC* muy similares y buen comportamiento de residuales, se decidió utilizar el modelo gamma (GA) debido a que este tiene menor número de parámetros y la idea es obtener un buen modelo y

parsimonioso.

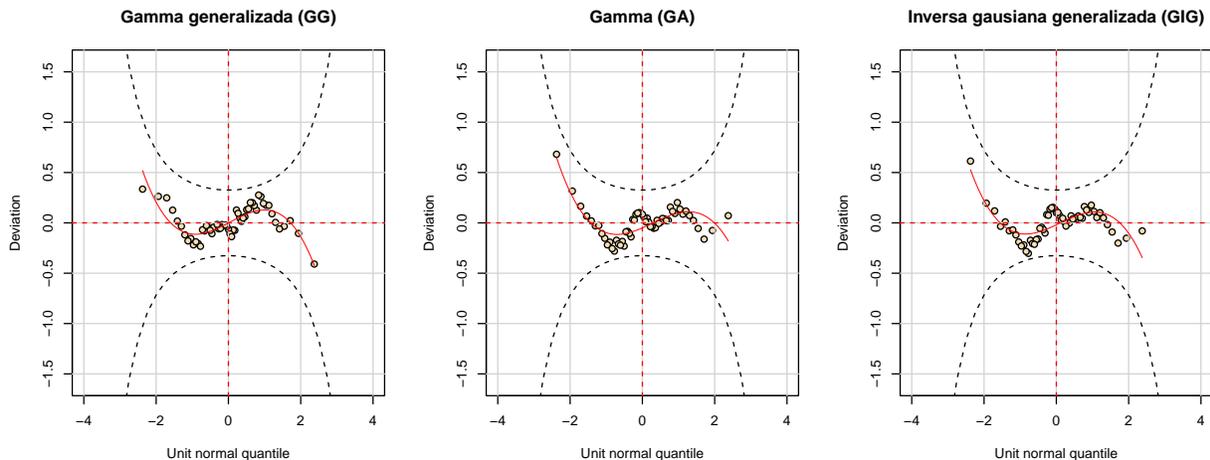


Figura 5: *Worm plot* para cada uno de los modelos ajustados y reportados en el Tabla 1. Fuente: Elaboración propia.

En el Tabla 2 se presentan los parámetros estimados para el modelo en el cual la variable respuesta tiene distribución gamma (GA), de esta tabla se observa que cada una de las variables es significativa a un nivel de 5% para explicar los parámetros  $\mu$  y  $\sigma$ . Las covariables peso de la bolsa y humedad relativa fueron eliminadas en el proceso de selección de variables y por esta razón no aparecen en el Tabla 2.

Tabla 2: Parámetros estimados para el modelo GA.

Modelo para $\log(\mu)$	Estimado	Error estándar	Valor $t$	Valor-P
Intercept	12.430	1.910	6.508	3.54e-08
Temp	-1.355	0.166	-8.185	8.63e-11
Temp <sup>2</sup>	0.038	0.004	9.802	3.14e-13
Tiempo	0.690	0.137	5.023	6.86e-06
Cmolida	0.738	0.133	5.550	1.08e-06
Modelo para $\log(\sigma)$	Estimado	Error estándar	Valor $t$	Valor-P
Intercept	1.106	0.884	1.251	0.217
Temp	-0.082	0.039	-2.105	0.040

A partir de la Tabla 2, y recordando que la función de enlace usada en la modelación fue log, se pueden obtener las expresiones para los parámetros estimados  $\mu$  y  $\sigma$  como se muestra a continuación:

$$\log(\hat{\mu}) = 12.430 - 1.355 \text{ Temp} + 0.038 \text{ Temp}^2 + 0.690 \text{ Tiempo} + 0.738 \text{ Cmolida} \quad (6)$$

$$\log(\hat{\sigma}) = 1.106 - 0.082 \text{ Temp} \quad (7)$$

El código con las instrucciones en R para replicar los análisis aquí presentados se puede consultar en: <https://raw.githubusercontent.com/fhernanb/orellana/master/Orellana.R>

## 5. RESULTADOS

En esta sección se presenta la interpretación de los resultados obtenidos del mejor modelo identificado en la sección anterior.

En la parametrización de la distribución gamma con parámetros  $\mu$  y  $\sigma$  usada en los modelos GAMLSS, el valor esperado está dado por  $E(Y) = \mu$  y la varianza por  $Var(Y) = \sigma^2 \mu^2$ . De esta forma, el valor esperado y varianza estimada para el rendimiento de la orellana se puede escribir como:

$$\widehat{E}(Y) = \hat{\mu} = e^{12.430 - 1.355 \text{ Temp} + 0.038 \text{ Temp}^2 + 0.690 \text{ Tiempo} + 0.738 \text{ Cmolida}} \quad (8)$$

$$\widehat{Var}(Y) = \hat{\mu}^2 \hat{\sigma}^2 = e^{27.072 - 2.874 \text{ Temp} + 0.076 \text{ Temp}^2 + 1.380 \text{ Tiempo} + 1.476 \text{ Cmolida}} \quad (9)$$

De la expresión 8 se observa que el promedio de orellana obtenido de cada bolsa es una función de la temperatura y del tiempo de aireación, a mayor temperatura y mayor tiempo se esperan mayores rendimientos de orellana, estos resultados coinciden con lo observado en los dos primeros paneles de la Figura 4. Para una temperatura constante se espera que por cada hora adicional de aireación, la producción media de orellana casi se duplique (obtenido de  $\exp(0.69)=1.994$ ). De esta misma expresión se observa que para temperatura y tiempos fijos, los rendimientos para bolsas con cascarilla molida se duplican frente a los rendimientos de bolsas con cascarilla entera (obtenido de  $\exp(0.738)=2.092$ ), esto significa que plantar en cascarilla molida favorece la producción de orellana. En la Figura 6 se presentan los gráficos de contornos para el valor esperado del rendimiento de orellana, de esta figura se observa que a mayor temperatura y a mayor tiempo de aireación el valor esperado aumenta. En la Figura 7 se presentan los gráficos de contornos para la varianza del rendimiento de orellana dada en la expresión 9, de esta figura se observa que la varianza crece a medida que la temperatura y el tiempo aumentan, se observa también que la varianza para el rendimiento de orellana cuando se usa cascarilla molida es aproximadamente 4.38 veces mayor frente a la cascarilla entera (obtenido de  $\exp(1.476)=4.375$ ).

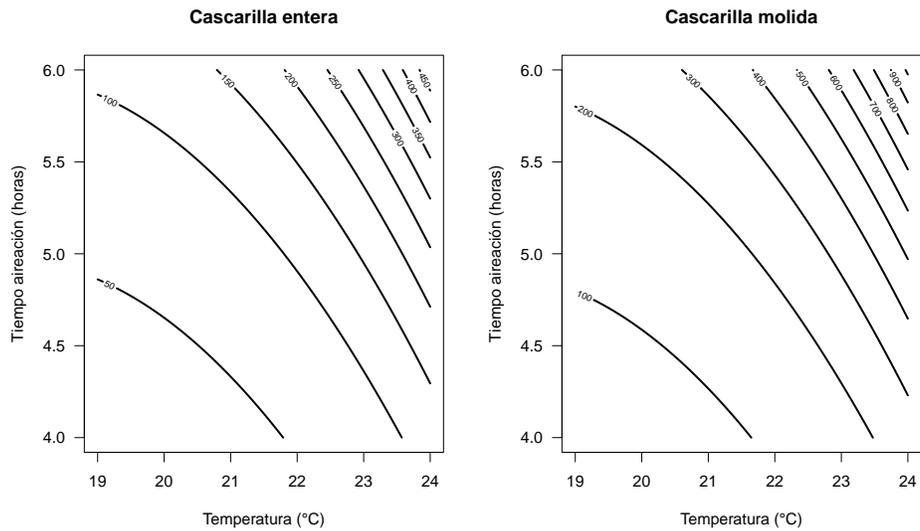


Figura 6: Gráfico de contornos para el valor esperado  $\widehat{E}(Y)$  del rendimiento de orellana en función de la temperatura, tiempo y el tipo de cascarilla. Fuente: elaboración propia.

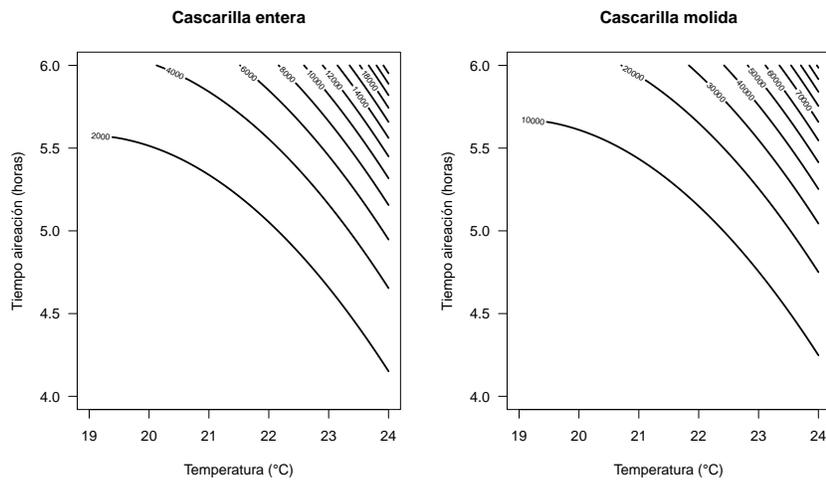


Figura 7: Gráfico de contornos para la varianza esperada  $\widehat{Var}(Y)$  del rendimiento de orellana en función de la temperatura, tiempo y el tipo de cascarilla. Fuente: elaboración propia.

## 6. CONCLUSIONES

Los modelos GAMLSS son una herramienta muy útil para modelar la relación entre una variable respuesta y un conjunto de covariables ya que permiten que la distribución para la variable de interés sea continua, discreta o mixta. En este artículo se mostró la aplicación de GAMLSS para estudiar los efectos de la temperatura, tiempo de aireación, humedad, cantidad de sustrato y tipo de sustrato sobre la producción industrial de orellana. De los resultados se encontró que la temperatura, el tiempo de aireación y el tipo de cascarilla son las variables que ayudan a explicar de forma significativa la producción de orellana; en particular, a mayor temperatura y a mayor tiempo de aireación la producción de orellana aumenta, se encontró también que la cascarilla molida favorece la producción de orellana duplicándola, frente al uso de cascarilla entera. Teniendo como referencia los rangos de las covariables en el estudio, se encontró que la producción de orellana se maximiza cuando la explotación se dá a temperatura de 24 °C, tiempo de aireación de 6 horas diarias y con cascarilla molida.

Los resultados aquí obtenidos pueden ser de mucha utilidad para las familias, cooperativas y empresas del sector fungicultor, se pueden utilizar para definir la mejor combinación de factores con el objetivo de maximizar la producción de orellana. Adicionalmente, la metodología usada en este artículo puede servir en un futuro como base para analizar datos experimentales de un estudio a mayor escala con covariables adicionales y con mayor número de observaciones.

## Referencias

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19 (6), 716–723.
- Akaike, H. (1983). Information measures and model selection. *Bulletin of the International Statistical Institute*, 50, 277-290.
- Bonet, J., Pukkala, T., Fisher, C., Palahí, M., Martínez de Aragón, J. & Colinas, C. (2008). Empirical models for predicting the production of wild mushrooms in Scots pine (*Pinus sylvestris* L.) forests in the Central Pyrenees. *Annals of Forest Science, Springer Verlag/EDP Sciences*.
- Buuren, S. (2001). An S Plus Program for drawing the Worm plot. [Consultada en noviembre de 2016]. Disponible en: <http://www.stefvanbuuren.nl/wormplot/wp.SSC.txt>.
- Buuren, S. & Fredriks, M. (2001). Worm plot: a simple diagnostic device for modelling growth reference curves. *Statistics in Medicine*, 20, 1259-1277.

- Cañedo, J. (2012). Cultivo de *Pleurotus ostreatus* en el valle de el Fuerte, Sinaloa: una alternativa de aprovechamiento de esquilmos agrícolas. (Tesis de doctorado). Universidad Autónoma Indígena de México, Sinaloa, México.
- Cardona, A. (2011). Colección Buenas Prácticas, Hongos tipo ostra. *FAO, Organización de las Naciones Unidas para la Alimentación y la Agricultura*.
- Cervantes, N. (2015). Cultivando setas de la especie *pleurotus pulmonarius* en el interior de las viviendas como un modelo empresarial que le permitirá a las víctimas del desplazamiento forzado la superación de su vulnerabilidad y propiciar su reinclusión social. *Revista de la Asociación Colombiana de Ciencias Biológicas*, 27, 80–87.
- de Castro, M.; Cancho, V. & Rodrigues, J. (2010). A hands-on approach for fitting long-term survival models under the GAMLSS framework. *Computer Methods and Programs in Biomedicine*, 97 (2), 168–177.
- Dunn, P. K. & Smyth, G. K. (1996). Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, 5 (3), 236-244.
- Gilchrist, R.; Stasinopoulos, D.; Rigby, R.; Sedgwick, J. & Voudouris, V. (2011). Forecasting Film Revenues Using GAMLSS. *Proceedings of the 26th International Workshop on Statistical Modelling*. Disponible en: <https://ssrn.com/abstract=1782783>.
- Hernández, R.; López, C. & Suárez, C. (2006). Evaluación de crecimiento y producción de *pleurotus ostreatus* sobre diferentes residuos agroindustriales del departamento de Cundinamarca (Tesis de pregrado). Pontificia Universidad Javeriana, Bogotá, Colombia.
- Hernández, F.; Torres, M.; Arteaga, L. & Castro, C. (2015). GAMLSS models applied in the treatment of agro-industrial waste. *Comunicaciones en Estadística*, 8 (2), 245–254.
- Martínez, F., de Miguel, S., Pukkala, T., Bonet, J., Ortega, P., Aldea, J. & Martínez, J. (2012). Yield models for ectomycorrhizal mushrooms in *Pinus sylvestris* forests with special focus on *Boletus edulis* and *Lactarius group deliciosus*. *Forest Ecology and Management*, 282, 63-69.
- R Core Team (2017). *R: A Language and Environmental for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>.
- Rigby, B. & Stasinopoulos, M. (2005). Generalized additive models for location scale and shape. *Applied Statistics*, 54 (3), 507–554.
- Sandercock G., Voss C.; Cohen D., Taylor M. & Stasinopoulos D. M. (2012). Centile curves and normative values for the twenty metre shuttle-run test in English schoolchildren. *Journal of Sports Sciences*, 30 (7), 679–687.

- Scandroglio, G., Gori, A., Vaccaro, E. & Voudouris, V. (2013). Estimating VaR and ES of the spot price of oil using futures-varying centiles. *International Journal of Financial Engineering and Risk Management*, 1 (1), 6–19.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6 (2), 461-464.
- Sierra, F. & Orozo, J. (2014). Sistema de cultivo para la gestión de hongos comestibles como complemento alimenticio dirigido a los habitantes de la Comuna 18 de Cali para mejorar la seguridad alimentaria causada por la inaccesibilidad (Tesis de pregrado). Universidad Icesi, Santiago de Cali, Colombia.
- Stasinopoulos, M., Rigby, B. & Akantziliotou, A. (2008). Instructions on how to use the gamlss package in R. [Consultado en noviembre de 2016].
- Tahvanainen, V., Miina, J., Kurttila, M. & Salo, K. (2016). Modelling the yields of marketed mushrooms in *Picea abies* stands in eastern Finland. *Forest Ecology and Management*, 362, 79–88.
- Tong, E. N., Mues, C. & Thomas, L. C. (2013). A zero-adjusted gamma model for mortgage loan loss given default. *International Journal of Forecasting*, 29, 548–562.
- Velasco, E., Zamora, M., Nieto, C., Martínez, J. & Montoya, A. (2010). Modelos predictivos de la producción de hongos silvestres comestibles en bosques de coníferas, Tlaxcala, México. *Revista Mexicana de Ciencias Forestales*, 1 (1), 95–104.