

Estimación del porcentaje de carbono orgánico en suelos utilizando Imágenes Satelitales y Modelos Mixtos.

Estimation of the percentage of organic carbon in soils using Satellite Images and Mixed Models.

Ana A. Oroza Hernández ^{1*}, Gladys Linares Fleites ², Hortensia J. Reyes Cervantes ¹, Bulmaro Juárez Hernández³

Resumen La estimación del porcentaje de carbono orgánico en suelos es uno de los problemas actuales que se presentan en los estudios de secuestro de carbono, que es considerado un importante aspecto en la mitigación del cambio climático. La utilización de modelación estadística donde la variable dependiente de una regresión está expresada como una proporción en el intervalo continuo (0,1) es el foco central en estos estudios, por lo que la utilización de los Modelos Lineales Generalizados Mixtos es imprescindible. Estos modelos pueden considerarse una moderna herramienta para integrar información de campo con la obtenida a través de imágenes satelitales. En este artículo se persigue el objetivo de modelar el porcentaje de carbono orgánico en suelos en la Región Terrestre Prioritaria (RTP) 105, Cuetzalan, México. Según la práctica tradicional, inicialmente se consideraron modelos mixtos bajo el supuesto de normalidad de los errores del modelo; posteriormente, bajo el supuesto de distribución gamma de los errores del modelo, se obtuvieron modelos de regresión gamma de efectos mixtos que estiman el porcentaje de carbono orgánico en función de propiedades físico químicas del suelo (efectos fijos) y del Índice de Vegetación Normalizado NDVI (efecto aleatorio), considerándose estos últimos modelos más adecuados que los iniciales. Se utilizó para la modelación el entorno R con la paquetería lme4.

Abstract The estimation of the percentage of organic carbon in soils is one of the current problems that arise in studies of carbon sequestration, which is considered an important aspect in the mitigation of climate change. The use of statistical modeling where the variable dependent on a regression is expressed as a proportion in the continuous interval (0.1) is the central focus in these studies, so the use of Mixed Generalized Linear Models are essential. These models can be considered a modern tool to integrate information of field with that obtained through satellite images. This article pursues the objective of modeling the percentage of organic carbon in soils in the Priority Land Region (RTP) 105, Cuetzalan, Mexico. According to the traditional practice, mixed models were initially considered under the assumption of normality of the errors of the model; subsequently, under the assumption of gamma distribution of the errors of the model, mixed-effect gamma regression models were obtained that estimate the percentage of organic carbon based on physical-chemical properties of the soil (fixed effects) and the Normalized Vegetation Index NDVI (random effect); these latter models being considered more appropriate than the initial ones. The R environment with the lme4 package was used for modeling.

Palabras Clave

Secuestro de Carbono, Modelos Lineales Generalizados Mixtos, Regresión gamma

Keywords

Carbon Sequestration, Generalized Linear Mixed Models, Gamma regression

¹ Facultad de Ciencias Físico Matemáticas, Benemérita Universidad Autónoma de Puebla, aleyda.oroza@alumno.buap.mx, hreyes@fcfm.buap.mx, bjuarez@fcfm.buap.mx

² Departamento de Investigación en Ciencias Agrícolas, Instituto de Ciencias, Benemérita Universidad Autónoma de Puebla, Puebla, México, gladys.linares@correo.buap.mx

*Autor para Correspondencia, Corresponding Author

Introducción

Los modelos de regresión son ampliamente utilizados en una diversidad de áreas de aplicación para describir las asociaciones entre las variables explicativas y la respuesta o variable dependiente. La presencia en estos modelos de una variable dependiente que está expresada como una proporción en el intervalo continuo (0,1) es cada vez más común en estudios y aplicaciones de distintas disciplinas [2]. En la problemática ambiental del cambio climático, que se considera el problema ambiental más importante al que se está enfrentado la humanidad en la actualidad, se presentan estudios de esta índole, ya que en mucho de ellos es de interés tratar de estimar el porcentaje de carbono orgánico en suelos a través de modelos estadísticos, para evaluar el secuestro de carbono en una determinada región.

La práctica de la modelación estadística ha estado en constante cambio como resultado del desarrollo de diferentes enfoques metodológicos de la Estadística y el progreso de las Ciencias Computacionales. En los últimos decenios se han alcanzado enormes desarrollos en los resultados analíticos del Modelo Lineal (ML) y del Modelo Lineal Generalizado (MLG) que es una extensión natural del ML [12] y [13]. El MLG ha llegado a suponer “una auténtica revolución estadística” [11], convirtiéndose en una solución especialmente adecuada para modelos de dependencia con datos no métricos. También, bajo el supuesto de normalidad de los errores, pero permitiendo la heteroscedasticidad de la varianza, ha habido considerable trabajo sobre los Modelos Lineales Mixtos (MLM), donde la estructura de la varianza está basada sobre efectos aleatorios [6]. Más recientemente, se han introducido los Modelos Lineales Generalizados Mixtos (MLGM), que constituyen una fusión entre el MLM y el MLG, y representan una herramienta analítica poderosa, ya que permiten considerar diferentes distribuciones de errores, derivadas del ajuste de los modelos a la variable a ser analizada, a la vez que permiten la inclusión y modelación de observaciones no independientes [4].

En el este artículo se persigue el objetivo de modelar el porcentaje de carbono orgánico en suelo en una cierta región del estado de Puebla, México, integrando información obtenida en campo con la obtenida a través de imágenes satelitales. Se utiliza para la modelación el entorno R de programación y análisis estadístico y gráfico, que se ha convertido en una necesidad en los tiempos actuales [15].

El trabajo tiene la siguiente estructura: en la sección 1 se tratan aspectos de los modelos estadísticos utilizados. En la sección 2 se refiere al empleo actual de las imágenes satelitales en estudios sobre la superficie terrestre e introduce el índice de vegetación NDVI utilizado para la modelación del caso bajo estudio. En la sección 3 se desarrolla el procedimiento utilizado para la estimación del porcentaje de carbono orgánico en suelos de la Región Terrestre Prioritaria (RTP) 105,

Cuetzalan, México. Finalmente, se brindan las conclusiones y las referencias.

1. Modelos Lineales Mixtos y Generalizaciones

Los *Modelos Lineales Mixtos* (MLM), permiten tener coeficientes fijos y aleatorios (cuyos niveles son sólo una realización de todos los posibles niveles procedentes de una población) y varios términos de error.

La especificación de este modelo puede plantearse como sigue:

Sea Y_{ki} una variable dependiente continua con realizaciones y_{ki} , con i observaciones en el grupo k , donde $i = 1, \dots, n_k$ y $k = 1, \dots, K$, con

- Componente estocástico: Y_{ki} se distribuye normal con media μ_{ki} y varianza σ^2 .
- Componente sistemático: $\mu_{ki} = X_{ki}\beta + Z_{ki}b_k$.

donde

X_{ki} es el arreglo ($n_k \times p \times K$) de variables explicatorias conocidas (no aleatorias) de efectos fijos,

β es el vector p -dimensional de coeficientes de efectos fijos, Z_{ki} es el arreglo ($n_k \times q \times K$) de variables explicatorias de efectos aleatorios conocidos, y

b_k es el vector q -dimensional de efectos aleatorios, que cumple:

$b_k \sim (0, \Sigma)$, donde Σ es la matriz de varianzas y covarianzas simétrica y semi-definida positiva.

Después de especificar un modelo con cierta parte sistemática y estructura para la parte aleatoria, el próximo paso en el proceso de modelación es estimar los parámetros del modelo. En la literatura sobre MLM se recomiendan dos variantes del método de estimación de máxima verosimilitud: el criterio máximo verosímil y el criterio máximo verosímil restringido, a los que nos referiremos por sus siglas en inglés como ML y REML respectivamente [3]. El método ML puede usarse para estimar tanto los coeficientes β como los parámetros que especifican la estructura de varianza y covarianza residual, sin embargo, los estimadores de la estructura de varianza y covarianza residual son sesgados. El método REML brinda estimadores sólo para los parámetros que especifican la estructura de varianza y covarianza residual; estos estimadores son insesgados en ciertos casos especiales y, generalmente, son menos sesgados que los estimadores ML.

La generalización de los MLM a Modelos Lineales Generalizados Mixtos (MLGM), requiere el establecimiento de una función de enlace conocida $g(\cdot)$, como se establece en el MLG [7]. Los MLG permiten especificar distintos tipos de distribución de errores y son una herramienta muy útil para modelar una proporción continua. Varias propuestas metodológicas [9], se han elaborado para analizar variables dependientes en

el rango continuo (0, 1).

La práctica más común para modelar proporciones continuas ha sido la aplicación del método de estimación de regresión por mínimos cuadrados ordinarios. Sea o no que se utilice asumiendo los supuestos distribucionales, se aduce con regularidad un argumento asintótico para su aplicación, en el sentido de que tamaños de muestra grandes permiten generar cuantificaciones válidas y confiables. Sin embargo, cuando se analizan la prueba t o la prueba F correspondientes se asume una distribución normal sin importar cuál sea el tamaño de muestra, lo mismo que al emplear algunas pruebas de heteroscedasticidad. Además del supuesto distribucional, el modelo de regresión lineal requiere los supuestos de homocedasticidad y linealidad. En términos conceptuales, este enfoque no es el más adecuado al tratar modelos de regresión donde la variable dependiente es una proporción. Como se ha expresado antes, existen otras alternativas ya que los MLGM permiten especificar distintos tipos de distribución de los errores del modelo.

2. Imágenes satelitales

Las imágenes satelitales son objeto de estudio de la percepción remota o teledetección, ciencia interdisciplinaria y multidisciplinaria que observa desde el espacio objetos físicos a grandes distancias.

Las imágenes procesadas nos permiten detectar cambios en la cobertura terrestre y realizar análisis de usos de suelo, entre otras múltiples aplicaciones [14]. Una definición más formal de esta ciencia la describe como la técnica de adquisición y posterior tratamiento de datos de la superficie terrestre desde sensores instalados en plataformas espaciales, en virtud de la interacción electromagnética existente entre la Tierra y el sensor. Los términos plataformas, sensores y canales deben ser bien comprendidos. Un sensor es el aparato que reúne la tecnología necesaria para captar imágenes a distancia y que es transportado en una plataforma. Puede captar información para diferentes regiones del espectro y cada una de estas regiones se denomina canal o banda.

Las imágenes que proceden del satélite Landsat-5 sensor TM (Thematic Mapper) son imágenes de 8 bits en 7 bandas espectrales; tres de las bandas son visibles y se reconocen como Banda 1, Banda 2 y Banda 3 y cuatro bandas son infrarrojo, conocidas como Banda 4, Banda 5, Banda 6 y Banda 7.

2.1 Índices de vegetación

Los índices de vegetación son medidas cuantitativas que indican el vigor de la vegetación. Su utilidad consiste en que ayudan a la interpretación de la imagen satelital permitiendo la detección de cambios en el uso del suelo, la evaluación de la cubierta vegetal y la densidad de la vegetación. El empleo de índices, calculados a partir de la reflectividad en diferentes bandas, que indican la abundancia y estado de la vegetación, se basan en el comportamiento reflectivo peculiar de la vegetación. La signatura espectral característica de la vegetación

sana muestra un fuerte contraste entre la baja reflectividad en el rojo y la alta reflectividad en el infrarrojo. Esta diferencia es tanto mayor cuanto mayor es la densidad de la vegetación y mejor su estado fitosanitario.

Entre los diversos índices de vegetación definidos [1] se destaca el llamado Índice Normalizado de Vegetación (NDVI, por sus siglas en inglés). Se obtiene utilizando la Banda 3 de la imagen de Landsat TM (reflectividad en el rojo) y la Banda 4 de la imagen de Landsat TM (reflectividad en el infrarrojo cercano). El NDVI es el cociente de la diferencia entre esas bandas y la suma de las mismas. Este índice varía entre -1 y 1, lo que facilita su interpretación. Puede señalarse como umbral crítico para cubiertas vegetales un valor de NDVI de 0.1 y para la vegetación densa de 0.5. Cualquier valor negativo implica ausencia de vegetación por lo que debieran reclasificarse a cero.

3. Estimación del porcentaje de carbono orgánico en suelos de la Región Terrestre Prioritaria (RTP) 105, Cuetzalan, México

3.1 Localización

La zona estudiada, denominada Región Terrestre Prioritaria (RTP) 105, se encuentra ubicada en la porción nor-oriental del estado de Puebla, entre los paralelos $19^{\circ}43'30''$ y $20^{\circ}14'54''$ de latitud norte y los meridianos $97^{\circ}07'42''$ y $97^{\circ}43'30''$ de longitud occidental. Estos suelos cubren una superficie aproximada de 846 Km^2 .

3.2 Variables consideradas en el estudio

Con el propósito de hacer estudios de secuestro de carbono en suelo, especialistas del Departamento de Investigación en Ciencias Agrícolas (DICA) de la Benemérita Universidad Autónoma de Puebla (BUAP) vienen realizando investigaciones desde hace años. En la década de los noventa obtuvieron 51 muestras de suelo. Estas muestras, fueron procesadas en el laboratorio para obtener diferentes propiedades físico y químicas, entre las que se tienen el porcentaje de Carbono orgánico (Corg), el porcentaje de Nitrógeno total (Ntot) y la Densidad Aparente (DA).

Por otra parte, se obtuvieron imágenes satelitales de la zona, lo que permitió, a través del índice de vegetación NDVI, estimar la densidad en la cobertura vegetal del año 1994 (Figura 1). A través de este índice fue posible identificar la presencia de vegetación verde en la superficie y caracterizar su distribución espacial. La actividad fotosintética permitió clasificar la densidad de vegetación de la zona en tres clases: baja, media y alta. Estas clases concuerdan con la cobertura vegetal observada en campo.

3.3 Modelación bajo el supuesto de normalidad de los errores

Para la estimación del secuestro de carbono en la zona bajo estudio, se especificaron varios modelos lineales mixtos.

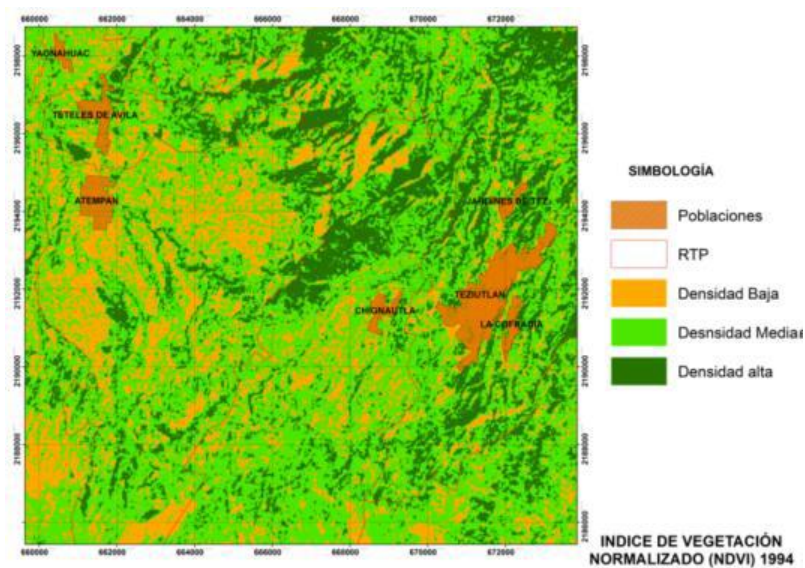


Figura 1. Imágenes con el cálculo de NDVI para el año 1994.

Tabla 1. Métodos de estimación bajo el supuesto de distribución normal de los errores (Modelo 1).

Método REML			Método ML		
Modelo 1: $Corg \sim DA + (1 NDVI)$			Modelo 1: $Corg \sim DA + (1 NDVI)$		
Efectos aleatorios:			Efectos aleatorios:		
Grupos	Nombre	Varianza	Grupos	Nombre	Varianza
NDVI	(Intercepto)	3.695e-05	NDVI	(Intercepto)	1.067e-05
Residuo		8.182e-04	Residuo		8.000e-04
Número de observaciones: 51, Grupos: NDVI, 3			Número de observaciones: 51, Grupos: NDVI, 3		
Efectos fijos:			Efectos fijos:		
	Estimador			Estimador	
(Intercepto)	0.13556		(Intercepto)	0.13465	
DA	-0.12389		DA	-0.12316	
Correlación de efectos fijos:			Correlación de efectos fijos:		
	(Intercepto)			(Intercepto)	
DA	-0.974		DA	-0.982	

Como señalamos anteriormente, los modelos mixtos son usados cuando los datos tienen algún tipo de agrupación, como es el caso que nos ocupa, donde se tienen tres niveles de densidades de vegetación (baja, media y alta), determinados por el índice de vegetación NDVI. Los modelos mixtos permiten tener coeficientes fijos y aleatorios (cuyos niveles son sólo una realización de todos los posibles niveles procedentes de una población) y varios términos de error. Para los efectos fijos se utilizaron, según el modelo especificado, las variables DA y Ntot, elegidas según criterios de especialistas en edafología y estudio exploratorios previos [10]. Se utilizó como variable dependiente o respuesta para los modelos el Corg. El procesamiento estadístico se realizó en lenguaje R [15]. El método de estimación que brinda el paquete estadístico lme4 en el lenguaje R, y que se ha usado aquí, es: primero estimar los parámetros de varianza y covarianza usando REML y,

posteriormente, utilizar estimadores mínimos cuadrados generalizados para los parámetros de la parte sistemática. Debe destacarse que el método ML puro es favorable en situaciones donde se requieran pruebas de hipótesis para comparar modelos, como en este estudio. Se establecieron dos modelos diferentes bajo el supuesto de normalidad de los errores y sus parámetros se estimaron con los métodos REML y ML, haciendo uso de la paquetería lme4 [8]. Los resultados se muestran en las Tablas 1 y 2. También se utilizaron los criterios de información para la comparación de modelos que brinda el paquete estadístico lme4. Los más comúnmente utilizados son el criterio de Información de Akaike (AIC) y el criterio de información bayesiano (BIC). En la literatura consultada para estos y otros criterios de selección de modelos que aparecen en las salidas del programa lme4, se elige como mejor modelo aquel con los valores más peque-

Tabla 2. Métodos de estimación bajo el supuesto de distribución normal de los errores (Modelo 2).

Método REML			Método ML		
Modelo 1: COrg ~ DA +Ntot+ (1 NDVI)			Modelo 1: COrg ~ DA +Ntot +(1 NDVI)		
Efectos aleatorios:			Efectos aleatorios:		
Grupos	Nombre	Varianza	Grupos	<Nombre	Varianza
NDVI	(Intercepto)	5.683e-05	NDVI	(Intercepto)	2.579e-05
Residual		7.531e-04	Residuo		7.210e-04
Número de observaciones: 51, Grupos: NDVI, 3			Número de observaciones: 51, Grupos: NDVI, 3		
Efectos fijos:			Efectos fijos:		
	Estimador			Estimador	
(Intercepto)	0.128136		(Intercepto)	0.12762	
DA	-0.120814		DA	-0.12038	
Ntot	0.014284		Ntot	0.01404	
Correlación de efectos fijos:			Correlación de efectos fijos:		
	(Intercepto)	DA		(Intercepto)	DA
DA	-0.963		DA	-0.982	
Ntot	-0.155	0.054	Ntot	-0.156	0.053

Tabla 3. : Selección de modelos.

	gl	AIC	BIC	logLik	χ^2	$gl - \chi^2$	$\Pr(>\chi^2)$
Modelo1	4	-210.3	-202.6	109.1			
Modelo2	5	-212.8	-203.1	111.4	4.5	1	0.03

ños del criterio. En la Tabla 3, se muestran los resultados del procedimiento de Análisis de Varianza para la comparación de ambos modelos, reajustándolos con ML en vez de REML. El resultado de la prueba χ^2 es significativo al 45 %, lo que apunta a considerar que el modelo 2 es mejor. Los criterios AIC y BIC también señalan que este modelo es ligeramente mejor. El modelo 2 fue ajustado usando el estimador REML, que es recomendado por tener un algoritmo más eficiente, puede ser expresado como

$$y_{ij} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + b_{0i} + \varepsilon_{ij}. \quad (1)$$

donde y_{ij} es el porcentaje de carbono orgánico de la i -ésima muestra en el j -ésimo nivel del índice de vegetación NDVI. Las variables independientes de efectos fijos son: X_1 que representa DA, y X_2 Ntot. La variable de efecto aleatorio es NDVI agrupada en tres niveles. El residuo ε_{ij} se supone que se distribuye normal con varianza σ^2 .

Los estimadores de los parámetros son $\beta_0 = 0.1281364$, $\beta_1 = -0.120814$, $\beta_2 = 0.014284$ y $\sigma^2 = 7.531e-04$. Estos valores destacan, como era de esperarse, la importancia de la densidad aparente para la explicación del porcentaje de carbono orgánico en estos suelos, lo que se manifiesta también al observar la alta correlación negativa con el intercepto. Por otra parte, los residuos estandarizados apuntan a que su distribución no es simétrica ya que el mínimo es aproximadamente -2 mientras que el máximo es aproximadamente 4 .

Dado que en el método de estimación ML se asume la norma-

lidad en la etapa de ajuste del modelo, se requiere de la comprobación de las suposiciones de este modelo. En la Figura 2 se muestran los gráficos de residuos y prueba de normalidad. Puede apreciarse que los datos muestran una estructura no normal y presencia de heteroscedasticidad. Como es conocido, en el pasado, las únicas herramientas disponibles para tratar la ausencia de normalidad eran la transformación de la variable respuesta o la adopción de métodos no paramétricos. Hoy en día, existen otras alternativas, que son los MLGM. Estos modelos permiten especificar distintos tipos de distribución de los errores como puede ser la distribución gamma. Esta distribución es útil con datos que muestran un coeficiente de variación constante, esto es, en donde la varianza aumenta según aumenta la media de la muestra de manera constante, lo que se cumple en este caso [5].

3.4 Modelación bajo el supuesto de distribución gamma de los errores

Bajo el supuesto de distribución gamma de los errores del modelo y, considerando los enlaces “log” y “probit” que brinda la función glmer del paquete lme4 del software R, se obtuvieron otros modelos para explicar Corg en función de la DA y del Ntot como efectos fijos y la densidad de vegetación como efecto aleatorio (NDVIA, NDVIB, NDVIM). Utilizando la notación del paquete lme4 este modelo puede expresarse como el modelo 2 antes considerado:

$$Corg \sim 1 + DA + Ntot + (1 | NDVI). \quad (2)$$

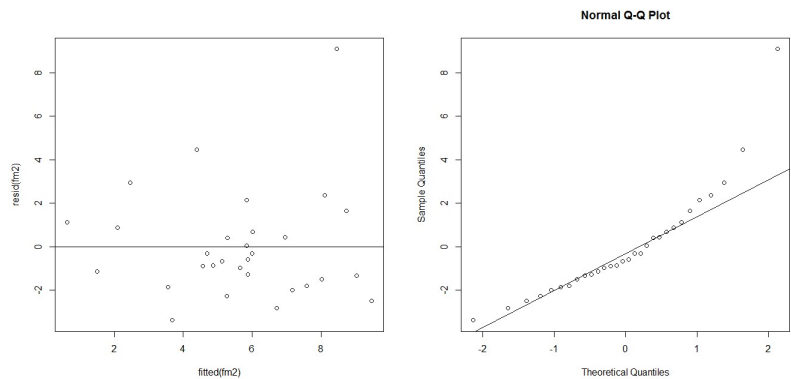


Figura 2. Gráficos de residuos bajo el supuesto de distribución normal.

Tabla 4. Métodos de estimación bajo el supuesto de distribución gamma de los errores.

Familia =Gamma("log")			Familia= Gamma("probit")		
MODELO DE MEDIA			MODELO DE MEDIA		
Estimadores de efectos fijos:			Estimadores de efectos fijos:		
	Estimador	Pr(> t)		Estimador	Pr(> t)
Interc	-1.4442	0.031017 *	Interc	-1.3805	1.75e-05***
DA	-2.7823	0.001657 **	DA	-1.0156	0.00491 **
Ntot	0.6766	0.000645 ***	Ntot	1.1379	2.41e-05 ***
Estimadores de efectos aleatorios:			Estimadores de efectos aleatorios:		
	Estimador	Error Estándar		Estimador	Error Estándar
NDVIA	0.0254	0.1110	NDVIA	.0003	0.0437
NDVIB	0.0455	0.1090	NDVIB	0.0222	0.0428
NDVIM	-0.0709	0.1078	NDVIM	-0.0224	0.0424
MODELO DE DISPERSIÓN			MODELO DE DISPERSIÓN		
Parámetro de dispersión para el modelo de media: 0.6057283			Parámetro de dispersión para el modelo de media: 0.5107561		
Efectos:			Efectos:		
	Estimador	Error Estándar		Estimador	Error Estándar
	-0.5013	0.2053		-0.6719	0.2052
Parámetro de dispersión para el efecto aleatorio: 0.01408			Parámetro de dispersión para el efecto aleatorio: 0.002061		
Efectos:			Efectos:		
. Aleatorio1			. Aleatorio1		
	Estimador	Error Estándar		Estimador	Error Estándar
	-4.2631	1.9066		-6.1846	2.0375

En la Tabla 4 se muestran los resultados obtenidos. Con el enlace “log” se obtuvo la convergencia en 7 iteraciones y las pruebas t se basaron en 47 grados de libertad (gl), mientras que el enlace “probit” convergió en 8 iteraciones y las pruebas se basaron en 48 gl. Todos los estimadores de los efectos fijos fueron significativos a diferentes niveles de significación en ambos enlaces, lo que corrobora la importancia de la DA y del Ntot en la explicación del Corg. Estos modelos pueden considerarse más adecuados que los obtenidos bajo el supuesto de normalidad de los errores del modelo, ya que no se presentan incumplimientos de las suposiciones.

Conclusiones

Los Modelos Lineales Generalizados Mixtos (MLGM), pueden considerarse una moderna herramienta para integrar información obtenida en campo y procesada en laboratorios con la información obtenida a través de imágenes satelitales, en estudios de secuestro de carbono en suelos. El análisis de índices de vegetación en la Región Terrestre Prioritaria 105: Cuetzalan, Puebla, México, mediante el uso de sensores remotos, permitió visualizar el contraste entre tres clases de densidad de vegetación, baja, media y alta, mostrando que las observaciones tienen un comportamiento de agrupación y,

dado que la variable respuesta es una proporción, los MLGM permitieron una modelación más adecuada que la práctica tradicional de considerar normalidad en los errores. Se obtuvieron modelos de regresión gamma de efectos mixtos que estiman el porcentaje de carbono orgánico en función de propiedades fisicoquímicas del suelo (efectos fijos) y del Índice de Vegetación Normalizado (NDVI) considerado como efecto aleatorio, con los enlaces “log” y “probit”.

Referencias

- [1] A. Bannari, D. Morin, F. Bonn, and A. R. Huete. A review of vegetation indices. *Remote Sensing Reviews*, 13(1-2):95–120, 1995.
- [2] W. Bonat, Ribeiro Jr. Paulo, and Zeviani Walmes. Regression models with responses on the unity interval: specification, estimation and comparison. *Revista Brasileira de Biometria*, 30, 01 2012.
- [3] T. Burzykowski and A. Galecki. *Linear Mixed Effects Models Using R. A Step by Approach*. Springer, 2013.
- [4] L. Cayuela. *Modelos lineales mixtos (LMM) y modelos lineales generalizados mixtos (GLMM) en R*. Universidad Rey Juan Carlos, Madrid, España, 2012.
- [5] L. Cayuela. *Modelos lineales generalizados (GLM)*. Universidad Rey Juan Carlos, Madrid, España, 2014.
- [6] E. Demidenko. *Mixed Models. Theory and Applications with R*. John Wiley & Sons, 2013.
- [7] A. J. Dobson and A. Barnett. *An Introduction to Generalized Linear Models*. Chapman and Hall and CRC, 2008.
- [8] B. Douglas, Martin Mächler, Ben Bolker, and Steve Walker. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48, 2015.
- [9] R. Kieschnick and B D McCullough. Regression analysis of variates observed on $(0, 1)$: percentages, proportions and fractions. *Statistical Modelling*, 2003.
- [10] G. F. Linares, M.G. A. Tenorio, E. T. Trejo, and A. A. H. Oroza. Estimación del carbono orgánico en suelos por teledetección y modelos de regresión. *Revista Latinoamericana el Ambiente y las Ciencias*, 8(18):26–40, 2017.
- [11] J. M. Losilla, J. B. Navarro, A. Palmer, M. F. Rodrigo, and M. Ato. *Modelo Lineal Generalizado*. Girona: Documenta Universitaria, 2005.
- [12] P. McCullagh and J. A. Nelder FRS. *Generalized Linear Models*. Chapman and Hall, 1989.
- [13] P. McCullagh and J. Nelder. *Generalized Linear Models*. Chapman and Hall, 2001.
- [14] A. Ramos and A. Domínguez. *Universidad Tecnológica de Puebla, México*. El mundo de los Satélites: Fundamentos y Aplicaciones, 2002.
- [15] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2016.