



UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE CIENCIAS

"MODELOS LINEALES GENERALIZADOS: UNA APLICACION
A LA PREDICCIÓN DE MATERIALES COMBUSTIBLES EN EL
PARQUE NACIONAL LA MALINCHE, PUEBLA-TLAXCALA"

T E S I S
QUE PARA OBTENER EL TITULO DE:
A C T U A R I A
P R E S E N T A :
SOFIA VILLERS GOMEZ



FACULTAD DE CIENCIAS
UNAM

DIRECTOR DE TESIS: DR. CARLOS DIAZ AVALOS

2004





Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.



UNIVERSIDAD NACIONAL
AUTÓNOMA DE
MÉXICO

ACT. MAURICIO AGUILAR GONZÁLEZ
Jefe de la División de Estudios Profesionales de la
Facultad de Ciencias
Presente

Comunicamos a usted que hemos revisado el trabajo escrito:

“Modelos lineales generalizados: Una aplicación a la predicción de materiales combustibles en el parque nacional La Malinche, Puebla- Tlaxcala.”

realizado por Sofia Villers Gómez

con número de cuenta 09956585-7 , quien cubrió los créditos de la carrera de:
Actuaría

Dicho trabajo cuenta con nuestro voto aprobatorio.

A t e n t a m e n t e

Director de Tesis
Propietario

Dr. Carlos Díaz Ávalos

Propietario

M. en A.P. María del Pilar Alonso Reyes

Propietario

Act. Jaime Vázquez Alamilla

Suplente

M. en C. José Antonio Flores Díaz

Suplente

Mat. Margarita Elvira Chávez Cano

Consejo Departamental de Matemáticas

 FACULTAD DE CIENCIAS
 CONSEJO DEPARTAMENTAL DE
 MATEMÁTICAS

A mis padres,
mi hermana,
profesores y amigos.

Índice General

Introducción	iii
1 Modelos Lineales Generalizados	1
1.1 Estructura de los modelos lineales generalizados	5
1.2 Estimación de los parámetros	9
1.3 Análisis del ajuste del modelo	12
1.3.1 Residuos	13
1.3.2 Distancia de Cook	14
2 Metodología	17
2.1 Área de estudio	18
2.2 Selección de sitios y recolección de datos	18
2.3 Definición de variables	22
2.3.1 Variables independientes	22
2.3.2 Variables dependientes	26
2.4 Modelación de datos de combustibles	28
3 Resultados	33
3.1 Análisis exploratorio de los datos	34
3.1.1 Variables independientes	35
3.1.2 Variables dependientes	36

3.2	Modelo para peso del mantillo	38
3.2.1	Análisis del ajuste del modelo e interpretación	41
3.3	Modelo para profundidad del mantillo	45
3.3.1	Análisis del ajuste del modelo e interpretación	50
3.4	Modelo para cantidad de material leñoso por clase diamétrica . .	53
3.4.1	Modelo para cantidad de material leñoso de 0 a 0.5 cm. . .	54
3.4.2	Modelo para cantidad de material leñoso de 0.6 a 2.5 cm. .	57
3.4.3	Modelo para cantidad de material leñoso de 2.6 a 7.5 cm. .	59
3.4.4	Modelo para cantidad de material leñoso mayor a 7.5 cm. .	62
	Conclusiones	65
	Anexo 1	69
	Anexo 2	73
	Anexo 3	75

Introducción

Los incendios forestales son una forma natural de regeneración de flora y fauna, sin embargo la influencia de agentes externos no controlados generan un desequilibrio en el ecosistema provocando en ocasiones su destrucción. Los incendios forestales continuarán sucediendo, y como lo muestran las cifras actuales, posiblemente su presencia y sus efectos en diversas regiones del mundo se intensificarán. Preguntas importantes sobre la ocurrencia de los incendios son: cuándo, dónde y con qué severidad se presentarán ([3]).

Los incendios son una de las principales causas de destrucción forestal en nuestro país y el estudio de las causas y efectos de éstos es aún incipiente. De manera general se considera la existencia de tres tipos de incendios: superficiales, subterráneos y de copa o corona. En México el 90% de los incendios son de tipo superficial ([4]).

En el estudio de los incendios forestales hay dos puntos a considerar: las condiciones en las que se presenta un incendio y la conducta que tiene el incendio. Para la iniciación de un incendio se deben tomar en cuenta como elementos básicos los combustibles, la existencia de calor suficiente para mantener la combustión y la presencia de oxígeno. Estos elementos están relacionados entre sí de tal manera que para detener un incendio es suficiente con eliminar sólo a uno de ellos. Ya generado el incendio, un aspecto interesante es su comportamiento. A fin de predecirlo y controlarlo, se deben considerar tres características fundamentales del terreno: la distribución de los combustibles, la topografía del terreno y el clima.([1])

El clima tiene una fuerte influencia en el comportamiento del fuego, pero a diferencia de los combustibles que se pueden modelar y la topografía que permanece fija, el clima es incontrolable y generalmente impredecible por lo cual es muy difícil su estudio y más aún modelarlo. ([1])

La topografía influye en el comportamiento del fuego a través de características como la pendiente y la orientación de la ladera. Por ejemplo debido a las radiaciones solares, un terreno con orientación al sur tendrá menor humedad que el de orientación norte. Las zonas con mayor pendiente facilitan la dispersión del fuego hacia arriba. Asimismo la topografía del terreno facilitará o dificultará la acumulación de combustibles según sea el caso y esto lleva a considerar la existencia de una relación entre las características topográficas y la distribución de los combustibles. ([1]).

Muchas características de los combustibles como tipo, tamaño, cantidad, y contenido de humedad son importantes para determinar como se incendiará un área boscosa. Los combustibles se clasifican en superficiales o de suelo y aéreos. Los combustibles superficiales comprenden tanto elementos vivos como muertos, y en esta categoría están: raíces, humus o mantillo, hojarasca, ramillas y leña delgada, tocones, troncos y ramas gruesas muertas caídas en el suelo, hierba seca, matorral y regeneración arbórea baja (altura menor a un metro). Entre los combustibles aéreos también hay elementos vivos y elementos muertos como: matorral alto (de uno a dos metros), troncos, ramas y follaje de los árboles y los árboles secos en pie. Existe otra clasificación de acuerdo al tamaño en donde se considera como combustible fino a la hojarasca, las ramillas y leña delgada, la cual es muy inflamable. En período seco, este tipo de material tiene poca humedad, y dispersa fácilmente el fuego. Este tipo de combustible es importante porque la mayoría de los fuegos de superficie comienzan con este combustible ([19]).

La adecuada caracterización de los combustibles, y la determinación de su distribución espacial, son factores críticos para realizar la simulación del comportamiento del fuego. No obstante, la determinación de la distribución espacial de los combustibles forestales ha sido uno de los retos más difíciles a los que se han enfrentado los científicos relacionados con el tema de los incendios forestales ([10]). Según señalan Brown y Bevins (1986), la distribución espacial de los

combustibles forestales sigue un patrón preferentemente discontinuo.

La acumulación de los combustibles y el incendio potencial resultante, depende de la cantidad del material depositado, así como de otras importantes propiedades de ellos, como su compactación y su continuidad vertical y horizontal. Para ser usada como variable importante en la evaluación del comportamiento potencial del fuego, la cantidad de combustible debe ser agrupada por categorías dependientes del tamaño, además de dividirse en componentes vivos y muertos ([3]).

En Estados Unidos, donde el tema de la defensa contra incendios forestales recibe atención desde principios del siglo XX, se cuenta con modelos que ayudan a predecir el comportamiento del fuego en un determinado período de tiempo. FARSITE (Finney 1998), BEHAVE (Andrews 1986) y otros sistemas para la simulación del comportamiento del fuego se basan en un modelo matemático, desarrollado por Rothermel en 1972, que cuantifica la dispersión del fuego en la superficie de los combustibles. Este modelo matemático está compuesto por una serie de cálculos sobre el calor necesario para la iniciación del fuego, el flujo de propagación, intensidad de reacción y el efecto del viento y la pendiente. El clima y los combustibles son variables de entrada para realizar estos cálculos. Un modelo de combustible es una descripción simplificada para el modelo matemático del comportamiento del fuego. Para ser interpretables los modelos de combustibles requieren de un considerable esfuerzo de muestreo, y es difícil desarrollar modelos espaciales con base en estudios de campo. Para simplificar la valoración de los combustibles en E.U., se establecieron 13 modelos estandarizados para combustibles. Debido a que estos modelos no están relacionados con el tipo de vegetación ni con las condiciones en el momento de la predicción, las estimaciones basadas en ellos regularmente resultan con errores significativos ([17]).

Para estudiar el comportamiento del fuego en un incendio forestal es fundamental conocer las características y la distribución de los combustibles. En

este trabajo se consideran solamente los combustibles muertos y las condiciones físicas, biológicas y antropogénicas del terreno en sitios preestablecidos con el objetivo de proporcionar un modelo para los combustibles muertos para un área mayor con características similares. Las condiciones físicas se refieren a los parámetros relacionados con la topografía del terreno, y las biológicas se relacionan con la composición vegetal del terreno. También se consideró una variable referente a la influencia humana en el sitio. Dada la dificultad para estudiar variables climáticas, éstas no se consideraron en este estudio.

El objetivo de esta investigación es determinar a través de los modelos lineales generalizados las características físicas, biológicas y antropogénicas del área que contribuyen a explicar el comportamiento de los combustibles para construir modelos que permitan predecir las variables combustibles con un determinado nivel de significancia.

La información utilizada para realizar este trabajo corresponde a datos de materiales combustible y características biofísicas del terreno. Dichos datos se obtuvieron a través del levantamiento de campo realizado, en dos periodos, el primero de abril 2002 a febrero 2003 (en el que se recolectó la información del material combustible y de vegetación) y el segundo de junio 2003 a noviembre 2003 (se muestrearon las características físicas del terreno), dentro del proyecto CONACyT 38607-V "Comportamiento del fuego y evaluación de riesgos a incendios en áreas forestales de México: Un estudio en el volcán La Malinche." que se lleva a cabo en el Centro de Ciencias de la Atmósfera de la UNAM.

Como resultado de este trabajo se obtuvieron modelos matemáticos que permiten predecir los datos referentes a tres tipos de combustibles, que fueron estudiados en el parque nacional La Malinche. Dichas predicciones tienen un nivel de error el cual se buscó fuera el mínimo. Para la construcción de este modelo se determinó un conjunto de variables con la hipótesis de que explicaban a los datos de combustibles (variables dependientes) a partir de este conjunto se seleccionó el modelo que mejor ajustó a los datos.

La estructura de este trabajo es la siguiente, en el primer capítulo se da una descripción general de los modelos lineales generalizados, explicando cual es la estructura de estos modelos, la forma en que se realizan las estimaciones de los parámetros y los diagnósticos gráficos para analizar el ajuste de los modelos. En el segundo capítulo, se explica la metodología que se siguió para la medición y determinación de las variables, la colecta de los datos y el proceso de ajuste de los modelos de las variables combustibles. Como tercer capítulo se presentan los modelos finales que este trabajo propone para modelar las seis variables de materiales combustibles. Por último se dan las conclusiones y se anexan apéndices que contienen tablas y gráficas, derivadas del análisis exploratorio y de los ajustes de los modelos, que muestran resultados mencionados en el trabajo.

Capítulo 1

Modelos Lineales Generalizados

Un modelo es una abstracción de la realidad, la cual provee una aproximación de algún fenómeno más complejo. Los modelos matemáticos pueden clasificarse de manera general en deterministas y probabilistas. En un modelo determinista, las respuestas están definidas de manera precisa, regularmente por un conjunto de ecuaciones. Este tipo de modelos abundan en las ciencias e ingenierías. En los modelos probabilistas, las respuestas muestran variabilidad, ya que el modelo contiene términos aleatorios que incorporan a la respuesta la influencia de algunos procesos aleatorios.

En los modelos probabilistas la clase más importante es la de los modelos lineales.

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \varepsilon \quad (1.1)$$

donde y es la variable de respuesta, x_1, x_2, \dots, x_k son un conjunto de variables predictoras o regresoras, $\beta_0, \beta_1, \dots, \beta_k$ son un conjunto de parámetros desconocidos, y ε es el término del error, el cual es aleatorio. Las variables regresoras x_1, x_2, \dots, x_k son llamadas covariables o variables independientes. La ecuación (1.1) se llama modelo de regresión lineal. Regularmente se supone que el término de error ε tiene valor esperado igual con cero y que son independientes. En consecuencia el valor esperado de la respuesta y en este modelo de regresión

lineal será

$$E(y) = \beta_o + \beta_1 x_1 + \cdots + \beta_k x_k \quad (1.2)$$

La ecuación (1.1) tiene el nombre de modelo lineal porque la media de la respuesta es una función lineal de los parámetros desconocidos $\beta_o, \beta_1, \dots, \beta_k$. Es decir que un modelo con interacciones por ejemplo con $k = 2$ variables

$$y = \beta_o + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \varepsilon \quad (1.3)$$

o un polinomio de grado dos con $k = 2$ variables

$$y = \beta_o + \beta_1 x_1 + \beta_2 x_2 + \beta_{12} x_1 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2 + \varepsilon \quad (1.4)$$

o hasta un modelo con términos como

$$y = \beta_o + \beta_1 \sin\left(\frac{2\pi x}{12}\right) + \beta_2 \cos\left(\frac{2\pi x}{12}\right) + \varepsilon$$

son todos modelos lineales. El modelo de regresión lineal de la ecuación (1.1) se llama usualmente modelo lineal de primer orden. Cuando este modelo se usa para el análisis de datos en un diseño de experimentos, la ecuación (1.1) recibe el nombre de modelo de efectos principales. El modelo con interacciones (1.3) y el modelo de segundo orden (1.4) también son utilizados en el área del diseño de experimentos.

Los modelos de regresión lineal son ampliamente utilizados por varias razones. En primer lugar, son aproximaciones polinomiales naturales para relaciones funcionales complejas. Es decir, si $E(y) = f(x)$ es la relación real entre la variable de respuesta y la variable predictora x , entonces la aproximación de esta relación por series de Taylor de primer orden en algún punto de interés x_0 es

$$\begin{aligned} E(y) &\cong f(x_0) + \left. \frac{df(x)}{dx} \right|_{x=x_0} (x - x_0) + R \\ &\cong \beta_o + \beta_1 (x - x_0) \end{aligned}$$

en donde, ignorando el residuo R (y aparte del término de error), se tiene un modelo de regresión lineal con una variable. Cuando son consideradas k variables predictoras, la aproximación por series de Taylor de primer orden llevará

directamente a un modelo de regresión lineal con k variables. Del uso de series de Taylor de segundo orden resultará un modelo de segundo orden (1.4), o si las derivadas de segundo orden puras son eliminadas, se tendrá un modelo de interacciones, del cual la ecuación (1.3) es un ejemplo. Ya que los modelos de regresión lineal son usados regularmente (y con éxito) como polinomios que aproximan la relación real entre las variables, en ocasiones se les da el nombre de modelos empíricos.

La segunda razón por la que los modelos de regresión lineal son populares es por la facilidad para estimar los parámetros desconocidos $\beta_0, \beta_1, \dots, \beta_k$. El método de mínimos cuadrados es una técnica para estimar los parámetros que data de principios del siglo diez y nueve. Cuando este método se aplica a un modelo lineal, el analista debe solamente resolver un conjunto de $p = k + 1$ ecuaciones lineales simultáneas.

Finalmente, existe una teoría estadística buena, elegante y bien desarrollada para el modelo lineal. Si se supone que los errores (ε) en el modelo lineal son independientes, con distribución normal y varianza constante, entonces las pruebas estadísticas para los parámetros del modelo, los intervalos de confianza para los parámetros, así como los intervalos de confianza y de predicción para la variable de respuesta se pueden obtener fácilmente. Además, bajo estas condiciones las pruebas de hipótesis son uniformemente más potentes([14])

Usualmente los modelos de regresión lineal se derivan de modelos empíricos de fenómenos más complejos y generalmente desconocidos. Sin embargo, hay situaciones en las que el fenómeno está muy estudiado y es posible describirlo a través de una relación matemática. Por ejemplo la ley de enfriamiento de Newton, la cual establece que la tasa de cambio de la temperatura de un objeto es proporcional a la diferencia entre la temperatura actual del objeto y la temperatura del ambiente que lo rodea. Si T_t es la temperatura actual del objeto y T_A es la temperatura del ambiente, entonces

$$\frac{df}{dt} = -\beta(T_t - T_A) \quad (1.5)$$

donde β es la constante de proporcionalidad. El valor de β depende de la conductividad térmica del objeto y otros factores. Ahora, la temperatura actual del objeto al tiempo t es la solución a la ecuación (1.5) o

$$f(t, \beta) = T_A + (T_I - T_A)e^{\beta t} \quad (1.6)$$

donde T_I es la temperatura inicial del objeto. En la práctica, una persona mide la temperatura al tiempo t con un instrumento, y ambos, la persona y el instrumento son fuentes potenciales de variabilidad no consideradas en la ecuación (1.6). Combinando éstas y todas otras posibles fuentes de variabilidad en un término de error ε , se puede escribir el valor observado de la temperatura actual al tiempo t como

$$\begin{aligned} y &= f(t, \beta) + \varepsilon \\ &= T_A + (T_I - T_A)e^{\beta t} + \varepsilon \end{aligned} \quad (1.7)$$

La ecuación (1.7) es un ejemplo de un modelo de regresión no-lineal, ya que la respuesta no es una función lineal de los parámetros desconocidos β . Los modelos de regresión no-lineales tienen un papel muy importante en las ciencias e ingeniería, y usualmente se derivan del conocimiento de las bases mecánicas que describen el fenómeno. Muchos modelos no-lineales se desarrollan directamente de la solución a ecuaciones diferenciales, como se vio en la ecuación (1.7).

Así como en el caso del modelo de regresión lineal, es necesario estimar los parámetros correspondientes al modelo de regresión no-lineal, así como probar hipótesis y construir intervalos de confianza. Existe teoría estadística que soporta la inferencia para el modelo no-lineal. Esta teoría supone que la distribución de los errores es normal, así como que las observaciones ε son independientes y con varianza constante.

De lo anterior es claro que para los modelos de regresión lineal y no-lineal, la distribución normal tiene un papel central. Los procedimientos de inferencia de ambos modelos suponen que la variable de respuesta tiene una distribución

normal. Existen varias situaciones prácticas en las que no será posible satisfacer este supuesto. Por ejemplo, se supone que la variable de respuesta es de tipo discreta, tal como un conteo. En ocasiones se realiza el conteo de los aparatos defectuosos u otros eventos raros, tales como accidentes, pacientes con una enfermedad particular, e incluso la ocurrencia de fenómenos naturales incluyendo temblores y huracanes. Otra posibilidad es que la respuesta sea una variable de tipo binaria. Estudios donde la variable de respuesta es “éxito” o “fracaso” (i.e., 0 o 1) son comunes en todas las áreas científicas. Existen también situaciones donde la variable de respuesta es continua, pero el supuesto de normalidad no es realista. Un ejemplo es el tiempo de falla de un componente electrónico o un sistema. Este tipo de respuesta es no negativa y presenta una distribución de valores asimétrica.

Los modelos lineales generalizados se desarrollaron para poder ajustar modelos de regresión para datos con una variable de respuesta que tenga una distribución que pertenezca a la familia exponencial. La familia exponencial incluye las siguientes distribuciones: normal, binomial, Poisson, geométrica, binomial negativa, exponencial, gamma y normal inversa.([14])

1.1 Estructura de los modelos lineales generalizados

Un concepto relevante en los modelos lineales generalizados es el hecho de que la distribución de la variable de respuesta debe ser parte de la familia exponencial. Las funciones que son miembros de la familia exponencial tienen función de densidad probabilística para una observación y , que puede expresarse de la siguiente forma

$$f(y; \theta, \phi) = \exp \left\{ \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right\} \quad (1.8)$$

donde $a(\cdot)$, $b(\cdot)$ y $c(\cdot)$ son funciones específicas. El parámetro θ es un parámetro natural de localización, y ϕ es usualmente llamado parámetro de dispersión. Si la función de densidad de la variable de respuesta y se puede escribir como la ecuación (1.8) en donde ϕ es conocida, se tiene un modelo con parámetro θ , como es el caso de las distribuciones binomial y Poisson, en donde generalmente $\phi = 1$, excepto en casos de sobredispersión. Si ϕ es desconocido entonces será un miembro de la familia exponencial con dos parámetros, como la normal y la gamma.

Los modelos lineales generalizados tienen la siguiente estructura:

1. Se tiene y_1, y_2, \dots, y_n observaciones de la variable de respuesta, independientes con medias $\mu_1, \mu_2, \dots, \mu_n$ respectivamente.
2. La observación y_i tiene función de densidad que es miembro de la familia exponencial.
3. El modelo se construye alrededor del predictor lineal $\eta = x'\beta = \beta_0 + \sum \beta_i x_i$ el hecho de que se utilice el predictor lineal es lo que sugiere la terminología de modelos lineales generalizados.
4. El modelo se encuentra a través del uso de una función liga

$$\eta_i = g(\mu_i) \quad i = 1, 2, \dots, n$$

el término liga se deriva del hecho de que la función $g(\cdot)$ es la liga entre la media μ y el predictor lineal η . Nótese que el valor esperado de la respuesta es

$$E(y_i) = g^{-1}(\eta_i) = g^{-1}(x'_i\beta)$$

Para el caso particular de los modelos de regresión lineal se tiene que

$$\mu_i = \eta_i = x'_i\beta \quad i = 1, 2, \dots, n$$

lo cual sugiere que la función liga es la función identidad, $g(\mu_i) = \mu_i$.

5. La función liga es una función monótona y diferenciable.
6. La varianza σ_i^2 de $(i = 1, 2, \dots, n)$ es una función de la media μ_i , se llamará función de varianza y su notación será $V(\mu_i)$.

Existen muchas posibilidades para definir la función liga. Se denomina liga canónica si $\eta_i = \theta_i$, donde θ_i es el parámetro de localización de la función de densidad de y_i . Para el caso de la normal la función liga canónica es la identidad ($\eta_i = \mu_i$) y para la Poisson es la función logaritmo ($\eta_i = \ln(\mu_i)$) ([14])

En este trabajo se considera que las variables de respuesta tienen función de densidad normal para las variables continuas (peso del mantillo y profundidad del mantillo) y distribución Poisson para las variables que son conteos (cantidad de material leñoso por cada una de las clases diamétricas). En la tabla 1 se muestran las características de estas dos distribuciones miembros de la familia exponencial.

	<i>NORMAL</i>	<i>POISSON</i>
<i>Notación</i>	$N(\mu, \sigma^2)$	$P(\mu)$
<i>Rango de y</i>	$(-\infty, \infty)$	$\mathbb{N} \cup \{0\}$
<i>Parámetro de dispersión : ϕ</i>	σ^2	1
$b(\theta)$	$\theta^2/2$	$\exp(\theta)$
$c(y; \phi)$	$-\frac{1}{2}(\frac{y^2}{\phi} + \log(2\pi\phi))$	$-\log y!$
$\mu = E(Y; \theta)$	θ	$\exp(\theta)$
<i>Función liga canónica : $g(\mu)$</i>	<i>identidad</i>	log
<i>Función de varianza : $V(\mu)$</i>	1	μ

Tabla 1. Principales características de las distribuciones Normal y Poisson

Para el caso en el que se supone la variable de respuesta y_i tiene distribución normal, el modelo puede verse como un modelo de regresión lineal. Esto implica que la relación entre las variables regresoras x_1, x_2, \dots, x_k se relacionan con $E(y_i)$ a través del siguiente modelo

$$\mu_i = \eta_i = x_i' \beta, \quad i = 1, \dots, n$$

donde $\beta = (\beta_0, \beta_1, \dots, \beta_k)$ es el vector de los parámetros y $x_i' = [1, x_{i1}, x_{i2}, \dots, x_{ik}]$ es el vector de las covariables. Además se cumple que $y_i \sim N(\mu_i, \sigma_i^2)$. Otra forma de ver el modelo es basándose en la variable de respuesta y e introduciendo el error ε , es decir

$$y = x^T \beta + \varepsilon$$

en donde se supuso que $\varepsilon \sim N_n(\underline{0}, \sigma^2 I_n)$. Se puede observar que si el error tiene distribución normal entonces y también tendrá distribución normal. Ésta es la notación utilizada a lo largo del trabajo y en lugar de definir funciones liga, se especifican las transformaciones realizadas a la variable y . En este trabajo para las variables de peso del mantillo y profundidad del mantillo se supone que el error ε tiene distribución normal.

El modelo y las suposiciones que se hacen cuando la distribución de y_i es Poisson son las siguientes: Las respuestas y_1, y_2, \dots, y_n son conteos independientes que tienen distribución Poisson con $E(y_i) = \mu_i$ y $var(y_i) = \mu_i$. El conjunto de variables regresoras x_1, x_2, \dots, x_k describen el valor esperado μ a través del siguiente modelo:

$$\mu_i = e^{\eta_i} = e^{x_i' \beta}, \quad i = 1, \dots, n \quad (1.9)$$

donde $x_i' = [1, x_{i1}, x_{i2}, \dots, x_{ik}]$ y β es el vector de parámetros a estimar. Este modelo exponencial no es el único modelo posible para la distribución Poisson, pero los modelos log-lineales (que se definen por la ecuación 1.9) son modelos que se usan con frecuencia. Una ventaja de estos modelos es que claramente los datos ajustados por el modelo no serán negativos, tal como se necesita sean las

variables de conteo. Las variables de cantidad de material leñoso se ajustaron a modelos log-lineales como la ecuación 1.9.

1.2 Estimación de los parámetros

El método de máxima verosimilitud es uno de los métodos que existen para la estimación de los parámetros de los modelos lineales generalizados. Es posible utilizar este método ya que en estos modelos se supone que la variable de respuesta y_i tiene una distribución que es miembro de la familia exponencial. Sin embargo actualmente el algoritmo basado en iteraciones de las ecuaciones de mínimos cuadrados, llamado por sus siglas en inglés IRLS (iteratively reweighted least squares) se utiliza para resolver los sistemas de ecuaciones de máxima verosimilitud.

El paquete estadístico S-PLUS utilizado para estimar los parámetros de los modelos ajustados en este trabajo, hace uso del algoritmo IRLS para calcular los estimadores máximo verosímiles de los parámetros β . A continuación se describe brevemente el procedimiento para la estimación de los parámetros β .

Se considera el método de máxima verosimilitud aplicado a los modelos lineales generalizados, y se supone que se está usando como función liga a la función canónica. Entonces la función de verosimilitud es

$$\begin{aligned} l(y; \beta) &= \prod_{i=1}^n f(y_i; \theta_i, \phi) = \prod_{i=1}^n \exp \left\{ \frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi) \right\} \\ &= \exp \sum_{i=1}^n \left\{ \frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi) \right\} \\ \text{Sea } L &= \ln(l(y; \beta)) \\ \Rightarrow L &= \sum_{i=1}^n \left\{ \frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi) \right\} \end{aligned}$$

Para la liga canónica se cumple que $\theta_i = \eta_i = g(\mu_i) = x_i' \beta$; esto implica que $\frac{\partial \theta_i}{\partial \beta} = x_i$. Por la regla de la cadena se sabe que $\frac{\partial L}{\partial \beta} = \frac{\partial L}{\partial \theta_i} \frac{\partial \theta_i}{\partial \beta}$ por lo tanto se

cumple que:

$$\frac{\partial L}{\partial \beta} = \sum_{i=1}^n \frac{1}{a(\phi)} \left[y_i - \frac{db(\theta_i)}{d\theta_i} \right] x_i$$

Proposición 1 $\mu = E(y) = \frac{db(\theta)}{d\theta}$ para y con distribución miembro de la familia exponencial.

Demostración. $L = \ln(l(y; \beta)) = \left\{ \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right\} \implies \frac{\partial L}{\partial \theta} = \frac{1}{a(\phi)} \left[y - \frac{db(\theta)}{d\theta} \right]$

Sabemos que $E\left(\frac{\partial L}{\partial \theta}\right) = 0 \implies E\left(y - \frac{db(\theta)}{d\theta}\right) = 0 \implies E(y) = \frac{db(\theta)}{d\theta} = b'(\theta)$ ■

Por la proposición anterior se tiene que

$$\frac{\partial L}{\partial \beta} = \sum_{i=1}^n \frac{1}{a(\phi)} (y_i - \mu_i) x_i \quad (1.10)$$

Se puede observar que el uso de las ligas canónicas simplifica las ecuaciones matemáticas. Con la ecuación (1.10) es posible encontrar los estimadores máximo verosímiles para los parámetros β , resolviendo el siguiente sistema de ecuaciones

$$\sum_{i=1}^n \frac{1}{a(\phi)} (y_i - \mu_i) x_i = 0$$

En la mayoría de los casos la función $a(\phi)$ es una constante lo que reduce el sistema de ecuaciones en $\sum_{i=1}^n (y_i - \mu_i) x_i = 0$. Éste es un sistema con $p = k + 1$ ecuaciones, una para cada parámetro. El sistema de ecuaciones se puede escribir en forma matricial y queda de la siguiente forma:

$$X'(y - \mu) = 0 \quad (1.11)$$

donde $\mu' = [\mu_1, \dots, \mu_p]$. Éstas reciben el nombre de ecuaciones score de máxima verosimilitud. El sistema de ecuaciones score (1.11) se puede resolver a través del algoritmo IRLS.

El estimador de β obtenido a través del método de Newton-Raphson para modelos lineales generalizados con la función canónica como función liga es

$$b = (X'VX)^{-1} X'Vz$$

donde $V = \text{diag}(\text{var}_\mu)$, $\text{var}_\mu = \frac{d\mu}{d\theta}$ y $z' = [z_1, z_2, \dots, z_n]$ con $z_i = \hat{\eta}_i + (y_i - \hat{\mu}_i) \frac{d\eta_i}{d\mu_i}$,

El algoritmo IRLS basado en el método de Newton-Raphson se puede describir en los siguientes pasos:

1. Obtener un estimador inicial (b_0) de β .
2. Se estiman V y μ utilizando b_0 .
3. Sea $\eta_0 = Xb_0$ y definir z_1 basándose en η_0 .
4. Obtener un nuevo estimador b_1 , iterar hasta que se satisfaga algún posible criterio de convergencia.

Sea b el valor final que el algoritmo anterior produce como un estimador de β . Si los supuestos del modelo incluyendo la selección de la función liga son correctos, se puede decir que $E(b) = \beta$ asintóticamente, ya que b es la solución a las ecuaciones score (1.11). La matriz de información de los estimadores $I(b)$ dada por la varianza de la ecuación score es

$$I(b) = \text{var} \left\{ \frac{1}{a(\phi)} [X'(y - \mu)] \right\} = \frac{X'VX}{[a(\phi)]^2}$$

donde $V = \text{diag}\{\text{var}_\mu\}$ y var_μ es la varianza de la respuesta y ; var_μ denota la dependencia de la varianza de la respuesta con su media por lo que es función de μ , depende de la función de densidad que se seleccione. Entonces la matriz de varianzas y covarianzas de b es asintóticamente

$$\text{var}(b) = I^{-1}(b) = [X'VX]^{-1} [a(\phi)]^2 \quad (1.12)$$

Las estimaciones de los errores estándar de los coeficientes se obtienen de la raíz cuadrada de los elementos de la diagonal de la matriz de la ecuación 1.12 con los elementos de V reemplazados por sus estimadores apropiados.([14])

1.3 Análisis del ajuste del modelo

Al ajustar un modelo es necesario verificar que se cumplan tres especificaciones: primero que el modelo proporcione resultados con una adecuada aproximación a los valores reales; en segundo lugar, que ninguno de los supuestos, para el procedimiento de ajuste del modelo sean violados, y por último se debe verificar la existencia de observaciones atípicas que pudieran influir en la determinación del modelo. La verificación de estos tres aspectos se hace a través de gráficas y a este proceso se le denomina análisis del ajuste del modelo. En el presente trabajo se utilizan los valores ajustados por el modelo (\hat{y}_i), los residuos (e_i) y las distancias de Cook de las observaciones para realizar el análisis mencionado. ([14])

Para verificar que el modelo proporcione datos con una aproximación adecuada a los valores reales se realizó la gráfica de los valores observados de y contra los valores ajustados \hat{y} por el modelo. Se diagnosticó como un ajuste aceptable del modelo cuando las observaciones se encuentran relativamente cerca de la recta identidad en la gráfica. Si la gráfica muestra los datos distanciados de la recta, se considera que el modelo propuesto debe utilizarse con cuidado, ya que el error en la predicción puede ser grande.

La devianza es una estadística análoga a la suma de cuadrados de los residuos (e_i) y es utilizada para medir el ajuste del modelo o para comparar modelos. La devianza $D(y; \mu)$ es una función lineal de la función de verosimilitud y se define como: $D(y; \mu) = 2l(\mu^*; y) - 2l(\mu; y)$, donde μ^* maximiza la función de verosimilitud. Para la familia exponencial generalmente se considera $\mu^* = y$. Para el caso particular de la distribución normal la devianza es $D(y; \mu) = \sum_i w_i (y_i - \mu_i)^2$ que es la suma de los cuadrados de los errores.

1.3.1 Residuos

En la regresión lineal estándar, los residuos $(y_i - \hat{y}_i)$ se utilizan para detectar las violaciones a los supuestos como no homogeneidad de la varianza y errores en las especificaciones del modelo. Para los modelos lineales generalizados los residuos ordinarios, calculados a través de la siguiente ecuación $e_i = y_i - \hat{y}_i$, no son los más apropiados, ya que la $var(y_i)$ no es constante.([12])

Para los modelos lineales generalizados existen tres tipos de residuos. El primero es el residuo de Pearson y se define como $r_p = \frac{y-\mu}{\sqrt{V(\mu)}}$ se puede observar que es el residuo ordinario escalado por la estimación de la desviación estándar de y . El segundo residuo lleva el nombre de Anscombe y se deriva de la desventaja del residuo de Pearson en cuanto a que su distribución es asimétrica cuando y tiene distribución distinta a la normal, lo cual afecta las propiedades de los residuos. Anscombe propone definir el residuo utilizando una función $A(y)$ en lugar de y , donde $A(\cdot)$ se escoge de tal manera que la distribución de $A(Y)$ sea lo más normal posible.([12])

El tercer residuo es la devianza. Si la devianza se utiliza como una medida de discrepancia para los modelos lineales generalizados (GLM), entonces cada observación contribuirá con una cantidad d_i a esa medida, i.e. $\sum d_i = D$. Lo anterior implica que si se define el residuo de devianza como

$$r_D = \text{signo}(y - \hat{y}) \sqrt[3]{d_i}$$

el residuo tendrá el mismo signo que $y_i - \hat{y}_i$ y además se cumple que $\sum r_D^2 = D$.

Un trabajo realizado por Pierce y Schafer en 1986 sobre los residuos en modelos basados en distribuciones miembros de la familia exponencial, sugiere que los residuos de devianza son semejantes a los residuos generados por la mejor transformación posible para normalizar las observaciones, y recomiendan el uso de los residuos de devianza para construir las gráficas de diagnóstico.([12])

Para el análisis de ajuste de los modelos presentados en este trabajo se utilizaron los residuos de devianza.

La gráfica de los residuos contra los valores ajustados por el modelo se utilizó para diagnosticar el cumplimiento del supuesto de que la varianza del error es constante, si la gráfica muestra los residuos distribuidos aleatoriamente sugiere que la varianza es constante para todos los valores de y , si la varianza depende del valor medio de y la gráfica que se obtiene muestra tendencias, es decir no se ve una distribución aleatoria.

Para los modelos en donde la distribución del error es normal, la verificación de este supuesto se realizó mediante la gráfica de los residuos contra los correspondientes cuantiles de la distribución normal. Si los residuos se aproximan a la línea recta de la identidad en la gráfica entonces se puede diagnosticar que efectivamente el supuesto de normalidad no se viola.

1.3.2 Distancia de Cook

La distancia de Cook se calculó con la siguiente ecuación

$$C_i = \frac{r_i^2 h_{ii}}{p \hat{\phi} (1 - h_{ii})^2} \quad (1.13)$$

en donde h_{ii} son elementos de la diagonal de la matriz H . H es la matriz de proyección de mínimos cuadrados usada en la última iteración del proceso IRLS; r_i son los residuos estandarizados de la última regresión; p es el número de parámetros estimados en el modelo, $\hat{\phi}$ es la estadística ji cuadrada escalar cuyo valor se puede ver en la Tabla 1 en la sección 1.1. La matriz H es de dimensión $n \times n$ y cumple con $\hat{y} = H * y$, por lo que H depende solamente de las variables predictoras, no de y . ([5])

Si el valor de la distancia de Cook para alguna observación resulta ser mayor a uno, entonces se considera esa observación como un posible dato influyente y se trata de buscar la causa y con base en los datos originales confirmar si el dato es influyente. De resultar que el dato realmente difiere significativamente de los datos con características similares (es influyente), se analiza la influencia de la

observación a través de un nuevo ajuste del modelo donde el dato posiblemente influyente es eliminado, y se comparan ambos modelos para ver si el dato provoca cambios significativos en el modelo.

El paquete S-PLUS proporciona las distancias de Cook como resultado de ajustar el modelo a los datos, sólo en los casos en que la función de densidad es normal. En este trabajo se consideraron dos tipos de distribuciones de la familia exponencial, la normal y la Poisson por lo que fue necesario realizar el cálculo de forma manual para obtener las distancias de Cook en el caso de la distribución Poisson. Este cálculo se realizó siguiendo la ecuación 1.13, en donde el valor de $\hat{\phi}$ es uno y los valores h_{ii} y r_i se obtienen de los resultados dados por S-PLUS al ajustar un modelo lineal generalizado.

Capítulo 2

Metodología

Durante varios años se han realizado distintos estudios en el parque nacional La Malinche y la mayoría de éstos están relacionados con el tema de incendios. Uno de ellos es el proyecto de "Comportamiento del fuego y evaluación de riesgo a incendios en áreas forestales de México: Un estudio en el volcán La Malinche". La distribución, el tamaño y los tipos de combustibles son factores importantes en el estudio del comportamiento del fuego, por esta razón para este proyecto es importante el estudio de los materiales combustibles y esto es lo que motiva la realización de este trabajo. Con el objetivo es obtener ecuaciones lineales que modelen las siguientes características de los materiales combustibles: la profundidad del mantillo, el peso del mantillo y la cantidad de material leñoso clasificado por los diámetros, se determinaron las características del terreno y de los combustibles que eran posible medir con un levantamiento de campo.

A partir de los datos recolectados se construyeron las variables con las que se trabajó para ajustar los modelos. En este capítulo se dan algunas características del parque que fue el área de estudio. También se describen cada uno de los pasos que se siguieron desde la selección de los sitios de muestreo hasta el ajuste de los modelos.

2.1 Área de estudio

El Parque Nacional La Malinche se localiza en la zona Centro Oriental de México, en los estados de Puebla y Tlaxcala, entre los 19°05' y 19°22' de latitud norte y los 97°53' y 98°12' de longitud oeste. Abarca poco más de 30 km de diámetro, y se eleva de los 2,300 a los 4,461 metros sobre el nivel del mar (msnm). Ésta es un Área Natural Protegida (ANP) de gran importancia para Tlaxcala.

El parque era una zona ejidal que fue expropiada con el fin de crear una reserva ecológica, razón por la cual existen problemas entre las comunidades cercanas y la administración del parque, además de que está rodeado de zonas agrícolas.

En investigaciones realizadas por Espejel (1996, 1998), se menciona que a lo largo del tiempo el área de la Malinche ha sido fuente de recursos como madera, plantas medicinales, comestibles, ornamentales y religiosas, resinas animales y combustible en forma de leña y carbón. La pérdida de recursos naturales es crítica tanto en el área de influencia, como en la propia Área Natural Protegida ([21])

Con base en diez años de registros de incendios dentro del parque y del análisis de las características físicas en donde éstos se presentan, se ha reportado que los bosques de pino y pastizales han sido los más afectados por los incendios, y que el mayor número se manifiesta entre los 2,600 y 3,200 msnm, donde existe una gran dinámica de uso agropecuario ([20],[23]).

2.2 Selección de sitios y recolección de datos

El número de sitios de muestreo dependió del área total de cada comunidad vegetal identificada previamente, resultado de la fotointerpretación de fotos aéreas del año 2001 escala 1:25,000. Los sitios fueron asignados de acuerdo a los siguientes factores: 1) proporción existente entre la extensión ocupada por cada comu-

nidad vegetal con respecto al área total de la Malinche, 2) distribución dentro del Parque, 3) niveles altitudinales y topografía, 4) estado de conservación ([16]).

Se seleccionó la ubicación de los sitios y considerando que los muestreos cubrieran una superficie total de una ha por sitio, se elaboró para cada una de ellos una red de diez círculos con un área de 0.1 ha y con una distancia de 50 m entre cada uno de los centros de los círculos. El primer círculo se elegía al azar, cuidando de que no se encontrase cerca de los caminos y de las veredas. A partir de él, se estableció la matriz de los círculos restantes (Figura 1).

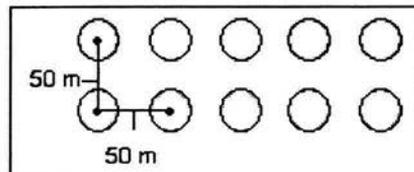


Figura 1. Esquema explica la metodología del muestreo.

Representa un sitio y se compone de 10 círculos de 0.1 ha cada uno. La recolección de datos es por círculo.

Cada círculo tiene un radio de 17.8 m. Tomando como base el punto central, se establecieron tres transectos de 25 m cuya dirección del azimut del primero se determinó al azar y la dirección de los dos restantes se obtuvo sumando 120° y 240° con respecto al azimut del primero (Figura 2)([13]).

Para la colecta y cuantificación del peso de combustible por hectárea, o carga de combustibles, se usó la técnica de intersecciones lineales, siguiendo los lineamientos establecidos por Brown (1974) y McRae et al. (1979). Este método de muestreo es el propuesto por la USFS y Canadá para la recolección de datos de combustibles.

Para medir el mantillo se marcó un área de 400 cm^2 en el extremo de cada transecto y se recogió todo el material vegetal (mantillo) que estuviese en esa

área, además con una regla graduada se midió la profundidad de la capa de mantillo, a cada cinco metros, y a lo largo de cada transecto ([13]). Para contar el material leñoso se consideró solamente el que estuviera sobre la línea marcada por el transecto y se agrupó en cuatro categorías dependiendo del diámetro del leño.

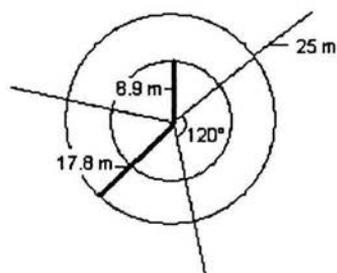


Figura 2. Esquema que representa un círculo de 0.1 ha con los tres transectos y los dos círculos utilizados para el muestreo de los materiales combustibles.

De esta manera el muestreo se realizó en quince sitios con diez círculos cada uno, excepto por un sitio en donde no fue posible hacer los diez círculos por las condiciones del terreno. En ese sitio sólo se recolectó información de seis círculos. Por lo anterior se tuvo un total de 149 círculos, este es el número de observaciones ya que las características físicas, de vegetación y antropogénicas del terreno se midieron por círculo. Las mediciones de combustibles que se hicieron siguiendo la metodología mencionada permitieron que se tuviera para el peso del mantillo tres datos por círculo ya que se tomó una muestra por cada transecto. Y como la profundidad del mantillo se midió cada cinco metros, se generaron 15 datos por círculo. En la figura 3 se observa la distribución de los sitios en el Volcán La Malinche.

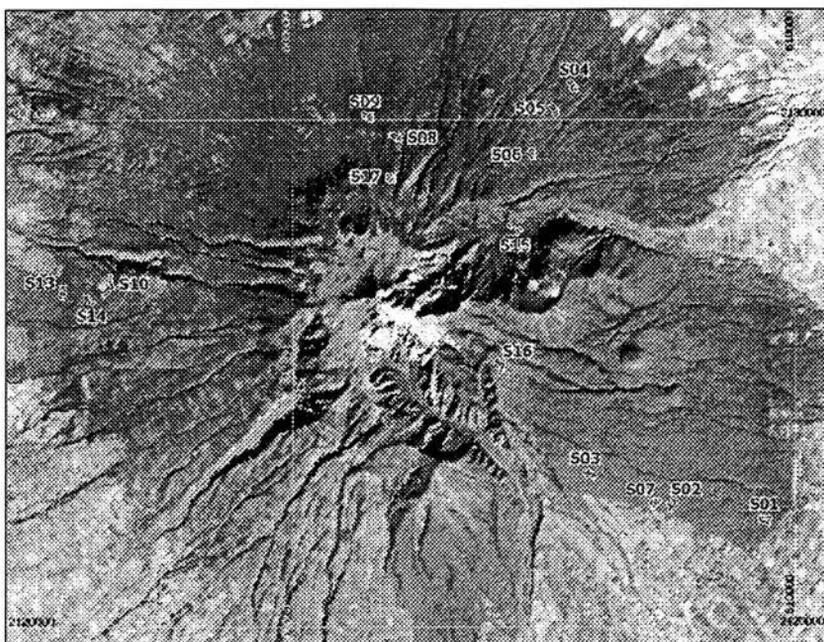


Figura 3. Localización de los quince sitios demuestreo con los círculos correspondientes. Volcán La Malinche, centro de México.

En cada círculo de muestreo se tomaron los datos del género, el DAP (diámetro a la altura del pecho) y la altura total, de cada árbol mayor de 10 cm de diámetro, y en un círculo interior menor (8.9 m de radio), los árboles con diámetros menores a 10 cm. También se realizó un levantamiento de información sobre los parámetros físicos en el que se incluía: altitud, pendiente, orientación, perfil longitudinal, perfil transversal de la ladera; microtopografía del círculo y la distribución de los combustibles en el círculo. Además se hicieron observaciones en los círculos en los que hubiera cauces, depresiones, caminos, brechas cortafuego, señales de tala, ocoteo y áreas incendiadas.

2.3 Definición de variables

Para ajustar un modelo a través de regresión lineal se debe tener datos de dos tipos de variables, las independientes son las que quedan como variables de entrada del modelo, y las dependientes que son a las que se ajusta el modelo, es decir las que se obtienen como resultado del modelo.

A partir de la información recolectada se definieron cuántas y cuales serían cada una de estas variables. A continuación se explica como se midieron cada uno de los parámetros en campo y con base en estos datos como se construyeron las variables.

2.3.1 Variables independientes

En este trabajo se consideraron como variables independientes las derivadas de tres tipos de características del terreno: parámetros físicos, tipo de vegetación y presencia antropogénica.

A continuación se explica como se obtuvieron los datos para cada uno de los parámetros físicos, antropogénicos y de vegetación considerados.

Físicos

Los parámetros físicos que se midieron en cada círculo fueron altitud, pendiente de la ladera, orientación de la ladera, forma del perfil longitudinal de la ladera, forma del perfil transversal de la ladera, microtopografía de la ladera, distribución del combustible, presencia de depresiones, cauces.

Altitud. Se midió en el centro del círculo con un altímetro estándar.

Pendiente de la ladera. Se midió el valor de la pendiente, en grados, mediante el uso de un clisímetro dentro del ámbito de influencia del propio círculo. Siempre se midió en el sentido dominante de la ladera y observando desde la parte más alta del círculo hasta la más baja.

Orientación de la ladera. A partir de la determinación de la dirección dominante de la ladera y observando desde la parte más alta del círculo hasta la más baja, se midió su orientación azimutal usando una brújula tipo Brunton, con graduaciones a cada grado. Esta variable se agrupó en las ocho categorías de orientación más usadas (N, S, E, W, NE, SE, SW y NW).

Forma del perfil longitudinal y transversal de la ladera. A través de la evaluación visual del círculo se selecciona el perfil, las opciones fueron: Plano, Cóncavo y Convexo para los dos casos.

Características microtopográficas. Haciendo una evaluación visual se caracterizó la microtopografía de cada círculo como: Regular, Irregular y con Depresiones.

Distribución de los combustibles. Con una evaluación visual de la distribución de los combustibles dentro del círculo, se seleccionó para cada círculo una de las tres opciones: Regular, Irregular o con Manchones dispersos.

Depresiones y cauces. A través de una observación visual se determinó la existencia de depresiones y/o cauces dentro del círculo.

Algunas de las variables de tipo físico se generaron a partir de una reestructuración de los parámetros medidos, ese fue el caso de la altitud y la pendiente. Para estas dos variables lo que se hizo fue generar cinco categorías tratando de obtener una distribución homogénea de los valores que tomaron esas variables.

El resto de las variables físicas se obtuvieron al considerar las características de orientación y perfiles de la ladera, la microtopografía y la distribución de los combustibles en el círculo y la presencia o ausencia de depresiones y cauces. Estas variables ya tenían sus categorías desde el diseño del muestreo y se conservaron esas categorías.

Vegetación

Se recolectaron los datos del género, el diámetro a la altura del pecho de cada árbol mayor de 10 cm de diámetro que se encontrara dentro del círculo, y para

los árboles con diámetros menores a 10 cm se tomaron los mismos datos pero sólo en un círculo interior menor (8.9 m de radio).

A partir de estos datos fue posible derivar variables referentes a la vegetación del sitio, éstas fueron: composición de la vegetación e índice de dominancia de árboles adultos.

Composición de la vegetación. Considerando los diferentes géneros en el sitio se utilizó el valor de importancia como el determinante de las composiciones. El valor de importancia se construyó de la siguiente forma ([11]):

$$\text{Densidad relativa} = \frac{\text{\# de árboles del género en el sitio}}{\text{\# total de árboles en el sitio}} \times 100$$

$$\text{Dominancia relativa} = \frac{\text{dominancia del género}}{\text{dominancia de todos los géneros}} \times 100$$

$$\text{Frecuencia relativa} = \frac{\text{frecuencia del género}}{\text{frecuencia de todos los géneros}} \times 100$$

Valor de importancia = Densidad relativa + Dominancia relativa + Frecuencia Relativa.

La dominancia es la suma de las áreas basales. La frecuencia del género es el número de círculos en el sitio en los que está presente ese género, la frecuencia de todos los géneros es la suma de todas las frecuencias del género.

El valor de importancia fue calculado para cada género por sitio. Se consideraron siete categorías para esta variable.

Índice de dominancia de árboles adultos. A partir del levantamiento de los datos de los árboles dentro del círculo es posible calcular un número aproximado del total de árboles dentro del círculo. Se consideraron como árboles adultos a aquéllos con diámetro mayor a 10 cm. A partir de estos datos se calculó el índice de dominancia de adultos. La variable fue dividida en cuatro clases (0 a 0.25, 0.25 a 0.5, 0.5 a 0.75 y 0.75 a 1).

Antropogénicos

A través de las observaciones adicionales que se hicieron, usando formatos de campo, se señaló la presencia o ausencia de brechas cortafuegos, caminos, áreas incendiadas, tala y ocoteo y se derivó una variable que mide de manera muy general la perturbación de tipo humano existente en los sitios de muestreo, llamada perturbación.

Perturbación. Esta variable se construyó a partir de las observaciones de presencia o ausencia de brechas cortafuegos, caminos, áreas incendiadas, tala y ocoteo. Esta nueva variable toma valores en tres categorías: 1) No perturbado, cuando en el círculo no observó ninguno de los elementos mencionados. 2) Medianamente perturbado, en el círculo se observó una o dos de las opciones. 3) Perturbado, el círculo tenía tres o más de las observaciones consideradas.

Para el ajuste del modelo se consideraron doce variables independientes, todas de tipo categórico, cuatro ordinales (altitud, pendiente, índice de dominancia de árboles adultos y perturbación) y el resto nominales.

Todas las variables independientes son categóricas por lo que para ajustar un modelo que dependa de estas variables, es necesario reestructurar las variables de manera que se tengan solamente variables dicotómicas, a través del siguiente procedimiento: se toma una variable con k categorías y se generan $k - 1$ nuevas variables dicotómicas, si la observación en esta variable toma el valor de la categoría m , con $m \leq k$, entonces para esa observación la $m - \text{ésima}$ nueva variable tomará el valor uno y el resto tomará el valor de cero.

Por ejemplo para el caso de la variable pendiente de la ladera se tienen las siguientes cinco categorías: -75° a -26° , -26° a -17° , -17° a -11° , -11° a -7° y -7° a -3° , entonces se reestructuran generando cuatro variables correspondiendo a cada una de las categorías PEND1, PEND2, PEND3 y PEND4, para una observación cuya pendiente de la ladera tomará el valor de -26° a -17° , se tendría que $\text{PEND1}=0$, $\text{PEND2}=1$, $\text{PEND3}=0$ y $\text{PEND4}=0$, otra observación con pendiente -7° a -3° tendría $\text{PEND1}=0$, $\text{PEND2}=0$, $\text{PEND3}=0$ y $\text{PEND4}=0$ y en

este caso se puede observar que son $k - 1$ variables porque con ellas es suficiente para considerar las k categorías. Estas nuevas variables son las que se utilizaron en este trabajo para ajustar los modelos de los materiales combustibles, por lo que a continuación se presenta una relación de las variables utilizadas con las correspondientes categorías de las cuales fueron derivadas.

Variable	Categoría de:	Variable	Categoría de:
	Presencia de Depresiones		Orientación
DEPRES	NO	ORIENT1	NE
	Presencia de Cauces	ORIENT2	Este
CAUCE	NO	ORIENT3	SE
	Altitud	ORIENT4	Sur
ALT11	2820-3030 msnm	ORIENT5	SO
ALT12	3030-3100 msnm	ORIENT6	Oeste
ALT13	3100-3220 msnm	ORIENT7	NO
ALT14	3220-3428 msnm		Perfil longitudinal de la ladera
	Pendiente	PERFLONG1	concavo
PEND1	-75° a -26°	PERFLONG2	convexo
PEND2	-26° a -17°		Perfil transversal de la ladera
PEND3	-17° a -11°	PERTRANS1	concavo
PEND4	-11° a -7°	PERTRANS2	convexo
	Índice de dominancia de árboles adultos		Microtopografía de la ladera
INDICE1	0 a 0.25	MICROTOPO1	depresiones
INDICE2	0.25 a 0.5	MICROTOPO2	irregular
INDICE3	0.5 a 0.75		Distribución del material combustible
	Vegetación	DISTCOMB1	irregular
VEG1	aile-pino	DISTCOMB2	manchones dispersos
VEG2	oyamel		Perturbación
VEG3	oyamel-desconocido	PERTURBA1	No perturbado
VEG4	pino	PERTURBA2	Medianamente perturbado
VEG5	pino-aile		
VEG6	pino-aile-desconocido		

Tabla 2. Descripción de las variables independientes utilizadas en los ajustes de los modelos

2.3.2 Variables dependientes

En este trabajo se quiere encontrar un modelo lineal para los datos combustibles, entonces las variables dependientes serán aquéllas que se refieran al material combustible.

Las variables dependientes o respuesta son las que se derivan de la colecta y cuantificación que se hizo de la carga de combustibles en cada círculo.

A continuación se describe como se recolectaron los datos de los combustibles en el círculo y como fue que se agruparon para determinar las variables dependientes.

Peso del mantillo. Al final del transecto se colocó un cuadrado de 20 x 20 cm, con el cual se recolectó todo el mantillo que estuviese dentro, este material fue secado y pesado. Se obtuvieron tres datos por círculo. El dato con el que se trabajó fue el peso y este se consideró como una sola variable de tipo continua. La unidad de medida con la que se trabajo esta variable fue kilogramos

Profundidad del mantillo. Sobre el transecto se hicieron cinco mediciones (a partir del centro y a cada cinco metros) de la profundidad del matillo. Con lo que se tienen cinco medidas por transecto, siendo tres transectos dan un total de quince datos de profundidad del mantillo por círculo. A ésta también se le consideró como una sola variable continua. La unidad de medida con la que se trabajo esta variable fue centímetros.

Cantidad de material leñoso. Todo el material leñoso que se encontró en el trayecto de los tres transectos fue seleccionado y medido, agrupándose en cuatro categorías de acuerdo con su diámetro (0 a 0.5 cm, 0.6 a 2.5, 2.6 a 7.5, y mayor a 7.5 cm). Se midió la frecuencia de todo el material leñoso que coincidió con cualquiera de las categorías. En este trabajo se generó una variable de combustible para cada clase diamétrica y como los datos se refieren a la frecuencia, las variables son consideradas de conteo (sólo toma valores en los enteros positivos).

En este estudio se consideraron seis variables dependientes (combustibles), dos de ellas de tipo continuo y cuatro de conteo. Para cada una de ellas se seleccionó el modelo que mejor ajustara a los datos recolectados.

2.4 Modelación de datos de combustibles

En esta sección se explica el procedimiento general que se siguió para hacer el ajuste de los datos y obtener el modelo para cada una de las variables dependientes. S-PLUS fue el paquete estadístico que se utilizó para realizar los análisis estadísticos necesarios para este trabajo. La metodología utilizada para obtener un modelo que ajuste a los datos de combustibles se basa en la teoría de los modelos lineales generalizados descrita en el capítulo anterior. Con un diagrama de flujo se explica brevemente el procedimiento y posteriormente se detalla cada uno de los procesos. (Figura 4)

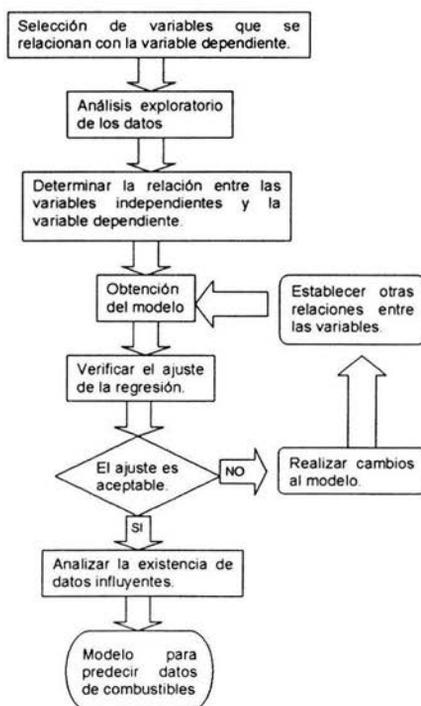


Figura 4. Diagrama de la metodología para ajustar un modelo lineal a las variables de respuesta.

A continuación se describen cada uno de los pasos realizados para los ajustes de los modelos.

Las variables independientes que se consideraron en este trabajo son de tres tipos: físicas, de vegetación y antropogénicas y cada una de ellas fue seleccionada de distinta manera.

Para la selección de variables independientes de tipo físicas se consideraron los parámetros físicos que fuera posible medir en un levantamiento de campo, la determinación de estas variables fue previo a su muestreo.

Con base en observaciones, relacionadas con la influencia humana en los sitios, realizadas en el muestreo de los parámetros físicos, se generó una variable que mide el nivel de presencia antropogénica en el sitio, llamada perturbación y ésta es la única variable antropogénica considerada en el trabajo.

También se consideraron variables de vegetación que se derivaron de la información referente a los árboles de cada uno de los sitios, recolectada en un muestreo previo. Es decir, al momento de empezar este trabajo ya se contaba con los datos referentes a la vegetación.

El análisis exploratorio de los datos tuvo como objetivo conocer de manera general el comportamiento de cada una de las variables, es decir, saber cuáles son los valores más frecuentes, en qué intervalo están los valores de la variable y si tiene alguna distribución conocida, para ello se hizo uso de estadísticas descriptivas y gráficas. La estadística que se utilizó para las variables independientes son la moda y como complemento al análisis, la distribución en porcentajes de las categorías para cada variable. Para el caso de las variables de combustibles se presentan, la media y la mediana, como medidas de tendencia central; la varianza, la desviación estándar, el valor máximo y el valor mínimo como medidas de dispersión.

Para establecer la relación entre las variables por primera vez se siguió el criterio de encontrar el modelo más simple. No se consideró la existencia de alguna relación "a priori", basada en estudios anteriores. Con base en lo anterior

se determinó que la relación entre las variables fuese lineal.

En el procedimiento de la obtención del modelo, se utilizó el método stepwise, el cual partiendo de un modelo inicial, a través de iteraciones agrega o elimina variables, dando como resultado el modelo con mejor ajuste. Con este método es posible también generar interacciones entre las variables, a las cuales les aplica el mismo procedimiento que a los efectos principales. Para este trabajo se consideraron interacciones de primer nivel, es decir entre solamente dos variables, no era posible realizar más porque el número de parámetros a estimar resultaba mayor al número de observaciones. Como resultado de la regresión se obtuvo un modelo con efectos principales e interacciones entre dos variables. El método stepwise busca explicar con el menor número de variables independientes el mayor porcentaje de la variabilidad de la variable dependientes, es decir el criterio que utiliza este método para determinar cual es el mejor modelo en cada uno de los pasos se basa en escoger el modelo que mejor explique la variabilidad y el proceso iterativo se detiene cuando ya no es posible mejorar el porcentaje de variabilidad explicado por el modelo.

La gráfica de los valores ajustados contra los residuos, la gráfica de los valores ajustados con los valores reales fueron los elementos con los que se verificó el ajuste de la regresión. Para el caso de la distribución normal también se usó el porcentaje de variabilidad explicado por el modelo (R^2). Dichas gráficas y valores se obtuvieron como resultados que el programa S-PLUS da al ajustar el modelo. Con base en ellos se decidió si se aceptaba o no el modelo.

Para el caso en el que el modelo no fue aceptado, fue necesario plantear otra relación entre las variables y repetir el proceso. En el trabajo la razón por la que el modelo no se aceptó fue por que no cumplió con la suposición de la variabilidad constante y lo que se hizo fue aplicar una transformación a la variable dependiente.

En caso de que el modelo fuese aceptado, con las gráficas anteriores, con las distancias de Cook, se determinó la posible existencia de datos influyentes,

después de ser identificados estos datos se analizaron las causas por las que eran significativamente distintos al resto de los datos basándose en las observaciones originales y después se ajustó el modelo eliminado el dato influyente y se analizaron las diferencias entre ambos del modelos.

Como resultado del procedimiento anterior se obtuvo una tabla con las variables que están en el modelo resultante y para cada una de ellas el valor estimado del parámetro y su correspondiente nivel de significancia, con base en estos datos, se eliminaron del modelo las variables cuya estimación del parámetro fuera cero o su nivel de significancia fuese mayor a 0.05. A partir de esta selección se obtuvo el modelo para predecir datos de combustibles que mejor ajustaban a los datos y con el mejor nivel de explicación posible.

Capítulo 3

Resultados

En este capítulo se presenta un análisis exploratorio realizado a todas las variables consideradas en el estudio así como los modelos que mejor ajustaron a cada una de las seis variables de materiales combustibles que se estudiaron. Para el caso de las variables de peso y profundidad del mantillo siendo del tipo continuo, en los modelos propuestos se supuso que el error tenía una distribución normal o posiblemente gamma. Las variables de cantidad de material leñoso siendo discretas (son conteos), la distribución del error se supuso Poisson.

Para cada una de las variables primero se describió el procedimiento seguido para encontrar un modelo con un ajuste estadísticamente aceptable para los datos observados. El proceso para encontrar el modelo, como se explicó en el capítulo anterior, empieza por suponer una relación entre las variables independientes y la variable dependiente (que es la que se está modelando), a esta relación recibe el nombre de modelo. Ya que se determinó la forma del modelo, se selecciona un subconjunto de variables independientes, a través del método stepwise, que proporciona los valores de los estimadores de los parámetros para todas las variables de este subconjunto. El siguiente paso es verificar que los supuestos, realizados para poder ajustar el modelo, se cumplan. De existir una violación a los supuestos es necesario hacer modificaciones al modelo. En este

trabajo las modificaciones realizadas fueron transformaciones a los datos observados o un cambio en la distribución de la variable dependiente y o el error ε ; en ambos casos se redefinió la relación entre las variables y se volvió a aplicar el método de stepwise, para encontrar el mejor modelo posible con base en esa relación de variables. La relación entre las variables se redefinen cuantas veces sea necesario para asegurar que el modelo obtenido cumpla todos los supuestos. En el proceso de encontrar un modelo aceptable para los datos se realizó como último paso un análisis de datos influyentes con el cálculo de la distancia de Cook y se analizó su influencia en el ajuste del modelo. El análisis de influencia consistió en eliminar el dato y realizar de nuevo el ajuste, si el nuevo modelo ajustado era mejor el dato era influyente y quedaba fuera del estudio, esto sólo sucedió para una observación en el modelo para el peso del mantillo.

Por último se presenta el análisis de ajuste del modelo final y su interpretación para cada una de las variables de combustibles.

3.1 Análisis exploratorio de los datos

Con la intención de conocer de manera general el comportamiento de cada una de las variables, es decir, saber cuáles son los valores más frecuentes, en qué intervalo están los valores de la variable y si tiene alguna distribución conocida, se realizó un análisis exploratorio a todos los datos con los que se trabajaron para los ajustes de los modelos. Para ello se utilizaron estadísticas descriptivas y gráficas, dependiendo del tipo de variable.

La estadística que se utilizó para las variables independientes, que son de tipo categóricas, fue la moda y como complemento al análisis, la distribución en porcentajes de las clases para cada variable. Para el caso de las variables independientes se calcularon: la media, la mediana y la moda, como medidas de tendencia central; la varianza, la desviación estándar, el valor máximo y el valor mínimo como medidas de dispersión.

3.1.1 Variables independientes

Para las variables independientes que se relacionan con las características físicas, de vegetación y antropogénicas de los círculos, se recolectaron un total de 146 datos, correspondientes a 14 sitios de diez círculos cada uno y un sitio de solamente seis círculos. Para estas variables no hubo omisión de información.

Los datos muestran que en 82% de los círculos no hay presencia de depresiones y que en el 93% de los casos no hay presencia de cauces.

Para las variables de altitud y pendiente los datos se distribuyen en forma uniforme en las diferentes categorías ya que de esa forma se construyeron las variables.

En el 55% de los círculos la composición de la vegetación es de pino o pino-aile. Para el caso del índice de dominancia de árboles adultos el 40% de los círculos está en el intervalo de 75% a 100% de adultos.

No existe círculo que tenga orientación norte de la ladera, la orientación de la ladera se distribuye de manera más o menos uniforme en el resto de las orientaciones.

El 85% de los círculos presentan perfil longitudinal de la ladera plano; mientras que el 69% tiene perfil transversal plano.

En la mayoría de los círculos la microtopografía de la ladera es irregular o regular. La distribución de los combustibles es regular en un 82% de los casos.

Para la selección de los sitios de muestreo se buscó que fueran lugares con poca perturbación humana, lo cual se refleja en las estadísticas de la variable de perturbación, en donde se observa que el 88% de los círculos son no perturbado o medianamente perturbado.

3.1.2 Variables dependientes

Por omisión en el levantamiento de la información, la variable peso del mantillo presentó seis casos sin información y para la variable de profundidad del mantillo quince de las observaciones no tenían información, por lo que el número de observaciones se redujo ligeramente.

Para los ajustes de los modelos se utilizaron 432 observaciones (son tres muestras por círculo) para el caso de peso del mantillo; 2175 datos para la variable profundidad del mantillo (cinco mediciones por círculo) y 438 datos para cada una de las cuatro variables de cantidad de material leñoso.

Los resultados obtenidos en el análisis exploratorio aplicado a las variables dependientes fueron los siguientes:

El peso del mantillo toma valores de 0 a 439.4 gramos (g), el valor medio es de 70 g y la dispersión es 55.48 g. Para los sitios con la misma composición de vegetación, los valores promedio del peso del mantillo resultaron similares entre los diferentes tipos de vegetación en estudio, los cuales toman valores en el rango de 30 a 141 gramos. (Figura 5)

La profundidad del mantillo tiene valores en el intervalo 0 a 32 cm. Su media es 4.4 cm. y la dispersión es 4.1 cm. La cantidad de material leñoso varía significativamente entre los sitios y más de la mitad del material está clasificado en las categorías con diámetros más pequeños (de 0 a 0.5 cm y de 0.6 a 2.5 cm) para todos los sitios (Figura 6). Las tablas con los resultados del análisis exploratorio se presentan en el anexo 1

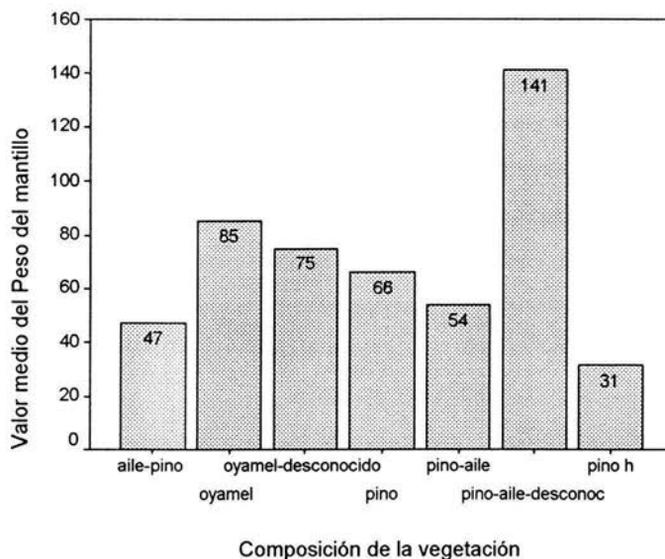


Figura 5. De los datos recolectados del muestreo en el parque La Malinche se obtiene el valor medio del peso del mantillo por cada una de las composiciones de la vegetación en estudio.

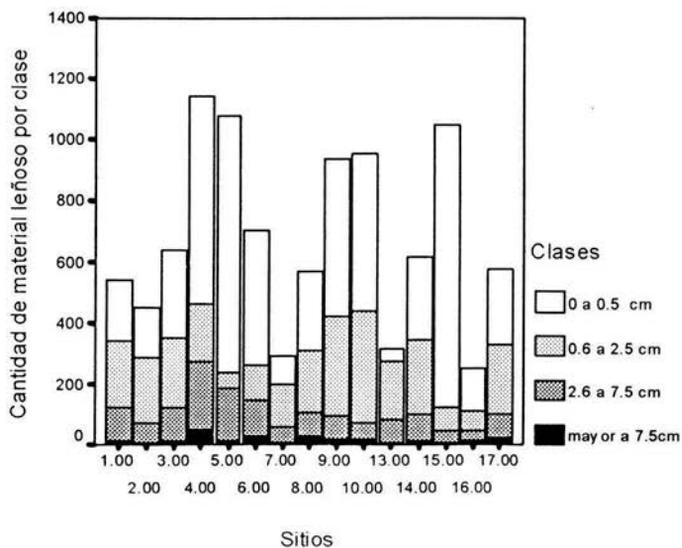


Figura 6. Cantidad de material leñoso agrupado en las cuatro clases diamétricas para cada uno de los sitios de muestreo del parque La Malinche.

3.2 Modelo para peso del mantillo

Con base en la metodología se consideraron todas las variables independientes que se muestrearon y se construyó un modelo lineal

$$y = x^T \beta + \varepsilon$$

A partir de este modelo se aplicó el método stepwise, con el paquete estadístico S-PLUS, para obtener el mejor modelo que ajuste a los datos, aceptando hasta interacciones de primer grado. El modelo resultante no cumplió con el supuesto de la varianza constante, ya que como puede observarse en la figura 7 existe una tendencia a que la varianza de los residuos e_i aumente cuando los valores estimados por el modelo (\hat{y}_i) aumentan. En la tabla 1 del anexo 2 se observan las variables que quedaron como resultado de buscar el mejor modelo que ajustara a los datos.

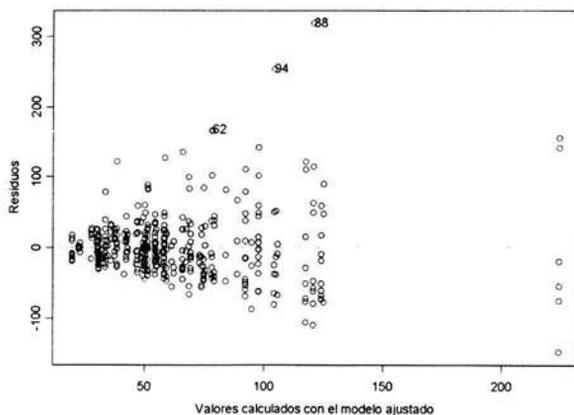


Figura 7. Residuos contra los valores ajustados por el modelo para la variable de peso del mantillo, con datos del parque nacional La Malinche.

Para corregir el problema de varianza se aplicó la transformación logaritmo a la variable dependiente y se ajustaron los datos a este segundo modelo

$$\log(y) = x^T\beta + \varepsilon \quad (3.1)$$

para este modelo no hubo indicios de violación a los supuestos. En la tabla 2 del anexo 2 se presentan las variables seleccionadas por el proceso de stepwise para explicar la variabilidad del peso del mantillo.

Después de obtenidos los parámetros del modelo y verificar que se cumplieran los supuestos, se hizo un análisis de datos influyentes (calculando las distancias de Cook para cada uno de los valores ajustados) (figura 8). Se encontró que una de las observaciones tiene distancia de Cook de 1.4, dado que es mayor a uno, se puede decir que es un posible dato influyente. Se identificó la observación en la base de datos y resultó que este dato tenía valor 40.80 g. y revisando los datos resultó no ser un error de medición ya que los datos de los círculos contiguos fueron entre 30 y 40 g., pero los datos con las mismas características es decir del mismo círculo, fueron cero. Coincidió en este caso, que los datos de peso del mantillo en este círculo difieren mucho y sea esa la causa de que el dato sea influyente.

Se eliminó el dato influyente de la base y se volvió a realizar el ajuste de un tercer modelo para determinar si el dato era influyente. El modelo resultante tuvo una valor de R^2 mayor, el modelo anterior explicaba un 48% de la variabilidad y este modelo explica el 58%. Las variables que quedaron dentro de los modelos fueron las mismas, esto se puede observar en las tablas 2 y 3 del anexo 2.

El modelo final para la variable de peso del mantillo esta formado por las siguientes variables:

Variables	Valor del parámetro	Error Estándar	Valor t
Intersección	5.34	0.19	27.63
CAUCE	-1.57	0.16	-10.01
INDICE2	0.22	0.10	2.27
ORIENT2	3.45	0.99	3.49
ORIENT3	2.55	0.26	9.98
ORIENT5	-1.21	0.32	-3.72
ORIENT6	-1.06	0.26	-4.13
PERFLONG2	-0.36	0.18	-1.99
CAUCEPERFTRANS1	1.70	0.20	8.44
CAUCEPERFTRANS2	-3.11	0.27	-11.55
ORIENT1PERFTRANS1	-2.58	0.61	-4.21
ORIENT3PERFTRANS1	2.82	0.28	10.02
ORIENT4PERFTRANS1	-2.43	0.50	-4.89
ORIENT5PERFTRANS1	-0.93	0.32	-2.93
ORIENT6PERFTRANS1	-1.36	0.32	-4.31
ORIENT1PERFTRANS2	2.43	0.41	5.94
ORIENT4PERFTRANS2	2.39	0.31	7.83
CAUCEPERFLONG1	-0.37	0.18	-2.13
CAUCEPERFLONG2	0.51	0.17	2.93

Tabla 3. Variables y parámetros estimados del modelo para peso del mantillo en el parque nacional La Malinche.

Los valores de β_i representan el cambio esperado en y con un cambio de unidad en x_i , cuando las demás variables son controladas. Se puede decir que la presencia de cauces en el círculo reduce el valor promedio del peso del mantillo. Si el círculo tiene orientación NE aumenta el valor esperado del peso del mantillo, mientras que si el círculo tiene orientación Sur, el valor disminuye.

En la tabla 3 del anexo 2 se puede observar las variables que el procedimiento de stepwise determinó fueran las consideradas en el tercer modelo, ya que en el modelo final sólo se seleccionaron aquellas variables con nivel de significancia menor a 0.05.

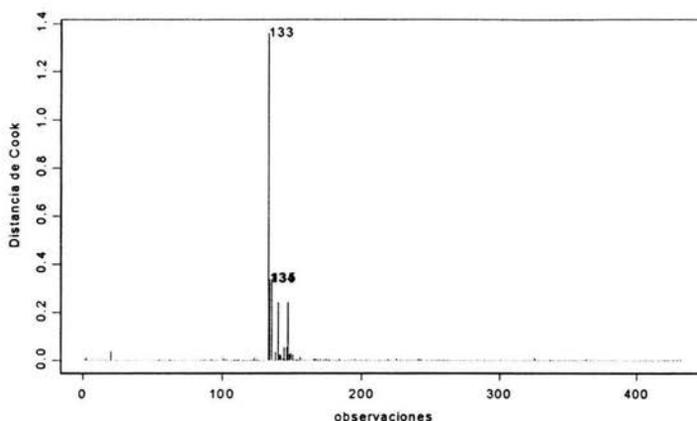


Figura 8. Valores de la distancia de Cook, calculadas para hacer el análisis de datos influyentes a las observaciones de peso del mantillo realizadas en La Malinche.

3.2.1 Análisis del ajuste del modelo e interpretación

El modelo final para la variable de peso del mantillo está dado por la ecuación 3.1 y las variables independientes en este modelo X son las variables mencionadas en la tabla 3. Interesara saber como es el ajuste de este modelo a los datos observados en el muestreo, ya que con esta información se sabrá la precisión que se tiene al predecir datos del peso del mantillo. También es importante tratar de explicar porqué es que el peso del mantillo está determinado por las covariables resultantes de este modelo, es importante mencionar que los expertos en el tema no tenían información documentada que estableciera alguna relación específica entre el peso del mantillo y las variables independientes consideradas en el estudio.

En la gráfica de la figura 9, los residuos e_i contra los valores estimados por el modelo \hat{y}_i , se observa la existencia de tres datos atípicos (139, 145 y 146) revisando la base de datos dos de ellos resultaron ser observaciones con valor del peso del mantillo cero, y por el uso de la transformación logaritmo son observaciones atípicas. La observación 145 tiene valor de 28.8 gr. para el peso del mantillo pero las dos observaciones de este mismo círculo son cero y 2.20, la diferencia si es significativa. Es importante mencionar que estas observaciones y la que resultó influyente son todas del mismo sitio (el sitio 16 con vegetación de *Pino hartwegii*) y si existía la posibilidad de que los muestreos cayeran en zonas donde no había mantillo, no son errores en los datos. A partir de esto se puede concluir que el modelo lineal propuesto es adecuado y que el término de error ε tiene varianza constante.

Para verificar el supuesto de normalidad de los errores se graficaron los residuos contra sus correspondientes cuantiles de la normal estándar (figura 10), se observó que la mayoría de las observaciones estaban sobre la recta, también se observó que las colas de la distribución son pesadas, es decir la probabilidad de encontrar datos en los extremos de la distribución es mayor de lo que se esperaba lo cual no contradice el comportamiento normal de las observaciones, con lo que se concluyó que los datos no daban evidencia de violación al supuesto de normalidad.

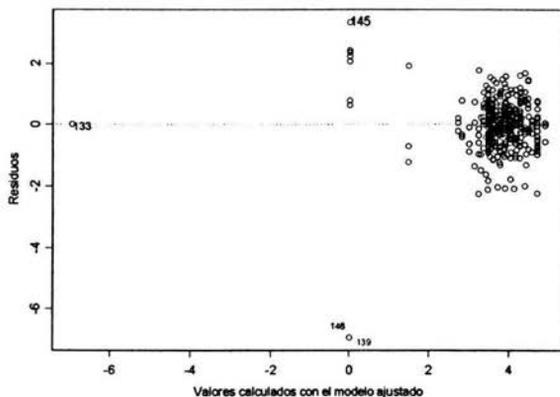


Figura 9. Residuos contra los valores del modelo ajustado a la variable de peso del mantillo, para probar el supuesto de varianza constante.

Con datos del parque nacional La Malinche

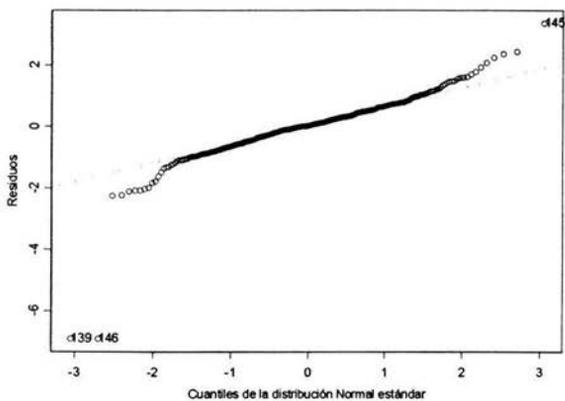


Figura 10. Residuos del modelo para la variable de peso del mantillo contra los cuantiles de la distribución normal estándar, para datos del parque La Malinche.

En la figura 11 se aprecia el ajuste del modelo, se grafican los valores reales obtenidos del muestreo contra los valores obtenidos con el modelo propuesto, si los puntos se encuentran cercanos a la recta se concluye que el ajuste del modelo es bueno. En este caso los puntos se caen en una franja cercana a la recta con lo que se concluye que el ajuste es aceptable.

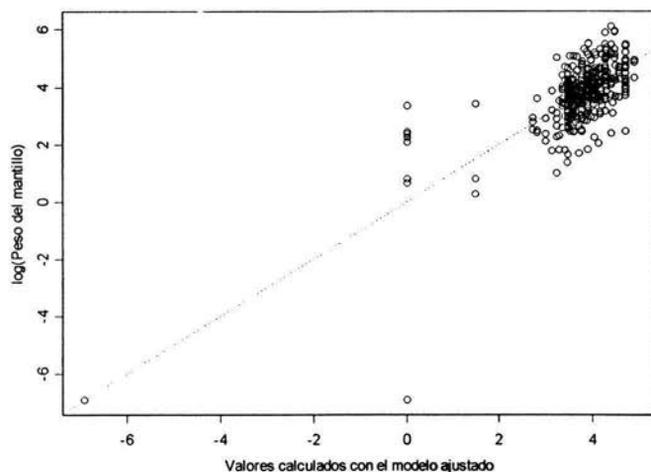


Figura 11. Valores de la transformación logaritmo aplicada a la variable peso del mantillo contra los valores obtenidos por el modelo ajustado a esta misma variable. Para datos del parque La Malinche.

La exactitud del modelo propuesto, la proporción de la variación del peso del mantillo explicada por las variables incluidas en el modelo, está reflejada por R^2 (coeficiente de determinación) que en este caso tiene un valor de 57.6%, lo cual sugiere la existencia de otras variables o interacciones entre las variables no consideradas en el estudio que podrían también explicar la variabilidad del peso del mantillo.

De una superficie de 400 cm² se recolecta todo el material sobre el suelo, el peso en seco de este material es la variable de peso del mantillo. Parte de la variabilidad de esta variable es explicada por la segunda categoría del índice de

dominancia de árboles adultos, lo cual da evidencias de que la variable de índice de dominancia se relaciona con la cantidad de mantillo que se acumule en el suelo. Existe una relación entre el tamaño del árbol y la cantidad de material que se desprenderá de éste depositándose en el suelo, esto justifica la relación mencionada.

El perfil transversal de la ladera, que puede ser plano, cóncavo o convexo afecta a la estimación del peso del mantillo siempre y cuando en el círculo no haya cauces, esto se debe a que ambas variables favorecerán o no la acumulación del mantillo en esas áreas, dependiendo de las categorías que tomen dichas variables; el mismo caso es para el perfil longitudinal de la ladera condicionado con la variable de presencia de cauces en el círculo.

Por el modelo obtenido es posible decir que un porcentaje de la variabilidad del peso del mantillo se explica con la variable de perfil transversal de la ladera condicionado a la orientación que tenga la ladera. La interacción entre estas dos características topográficas del terreno afecta el valor medio del peso del mantillo de manera positiva o negativa dependiendo de las categorías.

3.3 Modelo para profundidad del mantillo

Considerando a todas las variables independientes, a través del proceso de step-wise realizado en el paquete S-PLUS, se ajustó a los datos el siguiente modelo:

$$y = x^T \beta + \varepsilon$$

en donde se supone que $\varepsilon \sim N_n(\mathbf{0}, \sigma^2 I_n)$.

Al analizar los supuestos del modelo se encontró que los residuos e_i crecen al aumentar los valores ajustados por el modelo \hat{y}_i , como lo muestra la figura 12. Esto indica que el supuesto de que la función de varianza sea constante no se cumple, hay una relación entre la varianza del error y el valor medio de la variable de respuesta calculados por el modelo, por esta razón se dice que el

modelo propuesto no es el apropiado para modelar estos datos.

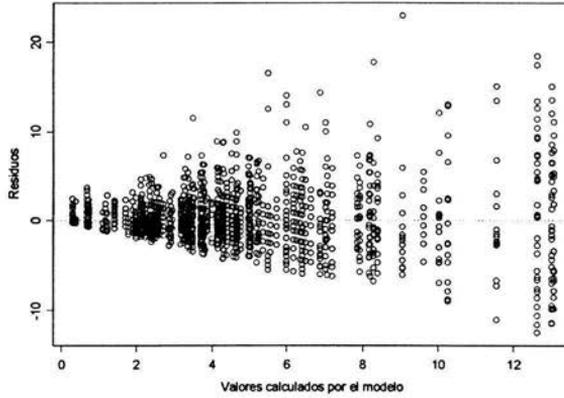


Figura 12. Residuos contra los valores ajustados por el modelo para la variable de profundidad del mantillo del parque La Malinche.

Para analizar el supuesto de varianza constante, el cual no se cumple.

Considerando que la variable de profundidad del mantillo toma valores en los reales positivos, con la intención de mejorar el problema de violación de los supuestos se ajustaron los datos a un nuevo modelo.

$$y = x^T \beta + \varepsilon$$

en donde y tiene distribución gamma, es decir la función liga que establece la relación entre $\mu = E(y)$ y los predictores lineales $x^T \beta$ es

$$g(\mu) = -\frac{1}{\mu}$$

y la función de varianza, que determina como la variabilidad de y depende de la media μ es

$$V(\mu) = \mu^2$$

con este modelo no se observaron violaciones a los supuestos. Pero después de seleccionar las variables que mejor explicaban a la variable de profundidad del mantillo resultó que el número de parámetros dentro del modelo era de 92, razón por la cual se buscó otro modelo que mejorara la parsimonia

El nuevo modelo propuesto fue a partir del primero quedando:

$$\sqrt[2]{y} = x^T \beta + \varepsilon$$

con $\varepsilon \sim N_n(\underline{0}, \sigma^2 I_n)$. Se aplicó también la transformación logarítmica pero al analizar la varianza se observó la existencia de tendencias en la gráfica lo cual contradecía el supuesto de variabilidad constante.

Con este modelo la gráfica de los residuos e_i contra los valores ajustados por el modelo \hat{y}_i tuvieron una distribución aleatoria con lo que es posible asegurar que la variabilidad de los errores cumple con el supuesto de ser constante, esto se puede observar en la figura 13.

A este modelo se le aplicó un análisis de datos influyentes, calculando las distancias de Cook para cada una de las observaciones en el ajuste. El resultado fue que los valores de las distancias de Cook para todas las observaciones fueron menores a uno lo cual indica que ninguno de los datos fue una posible observación influyente esto se puede observar en la figura 14.

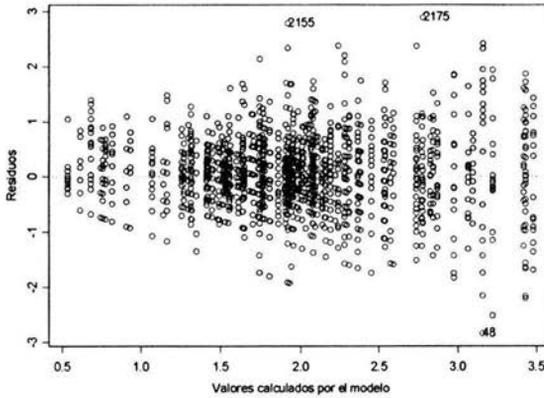


Figura 13. Residuos contra los valores ajustados por el modelo para la variable de profundidad del mantillo. Para analizar el supuesto de varianza constante, el cual se cumple

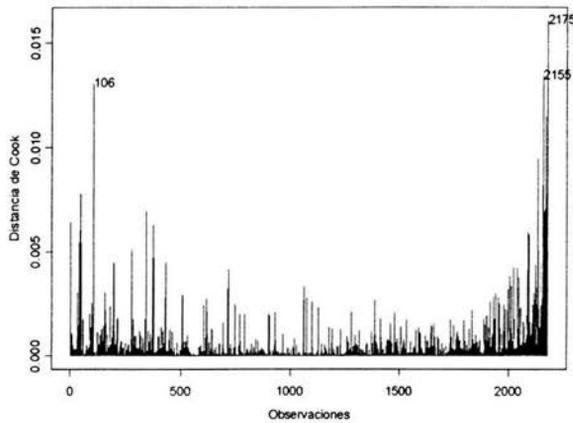


Figura 14. Distancia de Cook, calculadas para hacer el análisis de datos influyentes a las observaciones para la variable de profundidad del mantillo.

El modelo final para la variable de profundidad del mantillo fue:

$$\sqrt[3]{y} = x^T \beta + \varepsilon \quad (3.2)$$

en donde y tiene distribución normal. Las variables x y sus respectivos parámetros β son los siguientes

Variabes	Valor del parámetro	Error Estándar	Valor t
Intersección	1.59	0.81	1.96
CAUCE	0.65	0.18	3.67
ALTI4	-0.91	0.15	-5.97
PEND4	-0.54	0.25	-2.20
INDICE2	0.64	0.28	2.25
VEG1	0.85	0.30	2.84
VEG2	1.60	0.50	3.19
VEG3	-2.29	0.31	-7.48
VEG5	0.35	0.17	2.09
PERTRANS1	0.29	0.05	5.76
PERTRANS2	-0.13	0.05	-2.54
MICROTOPO1	0.30	0.12	2.41
DISTCOMB1	-0.42	0.17	-2.40
PERTURBA1	0.61	0.18	3.32
PERTURBA2	-0.20	0.09	-2.17
ALTI3ORIENT1	1.39	0.56	2.46
ALTI3ORIENT3	1.78	0.53	3.34
ALTI3ORIENT4	-1.87	0.38	-4.94
ALTI3ORIENT5	-1.32	0.29	-4.60
INDICE2ORIENT3	0.71	0.26	2.68
INDICE3ORIENT3	-0.85	0.26	-3.25
INDICE2ORIENT5	-1.12	0.21	-5.33
PEND1INDICE2	0.50	0.25	2.01
CAUCEINDICE2	-0.30	0.08	-3.58
DEPRESVEG4	-0.86	0.23	-3.74
DEPRESPEND3	0.72	0.24	3.01
DISTCOMB1PERTURBA1	0.22	0.09	2.30
DISTCOMB1PERTURBA2	-0.27	0.09	-3.07
CAUCEDISTCOMB1	0.57	0.17	3.44
CAUCEPERTURBA1	-0.27	0.10	-2.69

Tabla 4. Variabes y parámetros estimados del modelo para profundidad del mantillo en el parque nacional La Malinche

Al igual que en el caso del modelo para la variable de peso del mantillo, para el modelo final de la variable de profundidad del mantillo solamente se consideraron aquellas variables que resultaron ser significativas en el último ajuste del modelo, es decir que tuvieran un nivel de significancia menor a 0.05.

3.3.1 Análisis del ajuste del modelo e interpretación

El modelo final para la variable combustible de profundidad del mantillo está dado por la ecuación 3.2 y las variables independientes consideradas en este modelo X son las mencionadas en la tabla 4. Al igual que para el modelo del peso del mantillo, en esta sección se presenta el análisis gráfico realizado para evaluar el ajuste del modelo propuesto así como una interpretación del mismo modelo.

Para el modelo que ajustamos a los datos de profundidad del mantillo se supuso que el error ε tenía distribución normal. La función liga, que define la relación entre la media $\mu = E(\sqrt[3]{y})$ y los predictores lineales $x^T\beta$, es $g(\mu) = \mu$. La función de varianza es $var(y) = \phi$, donde ϕ es una constante.

Por lo anterior se deseaba probar que efectivamente los errores tuvieran distribución normal y que su varianza fuese constante. La figura 13 muestra que la gráfica de los residuos contra los valores ajustados por el modelo se distribuyen en forma aleatoria en una banda, lo cual permitió decir que los errores tenían varianza constante. En la figura 15 se observa que al graficar los residuos con sus correspondientes cuantiles de la distribución normal los datos tienden a concentrarse muy cerca de la recta identidad, a pesar de que el comportamiento de las colas no es el que se esperaba ya que las observaciones tienen mayor probabilidad en las colas, no se contradice el supuesto de que los errores tienen distribución normal.

Para verificar el ajuste del modelo a los datos se realizó la gráfica de los valores calculados con el modelo ajustado contra la raíz cuadrada de los valores de profundidad del mantillo obtenidos del muestreo, como se ve en la figura 16 el ajuste lineal propuesto para los datos es correcto, el modelo resultante es aceptable.

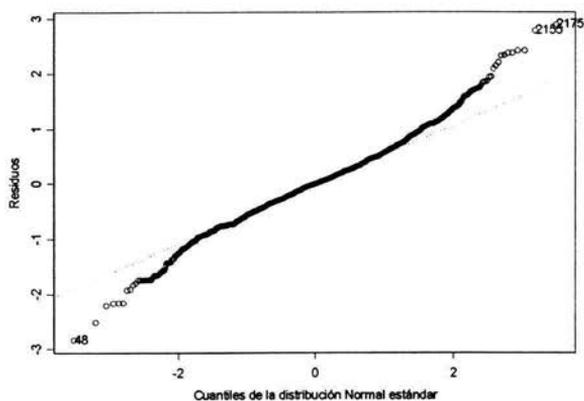


Figura 15. Residuos del modelo contra los cuantiles de la distribución normal estándar, para probar el supuesto de normalidad de los errores. Modelo para la variable de profundidad del mantillo en el parque nacional La Malinche.

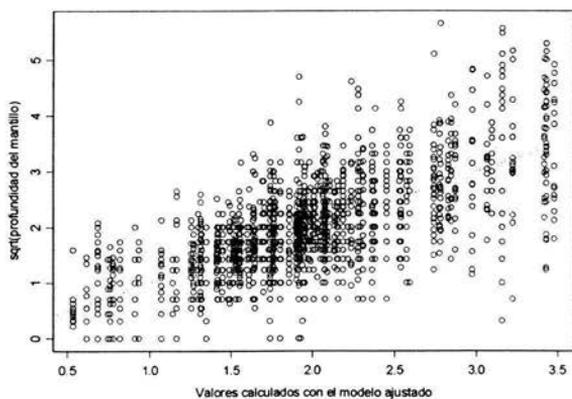


Figura 16. Valores de la transformación raíz cuadrada aplicada a profundidad del mantillo contra los valores obtenidos por el modelo ajustado a esa misma variable, con datos tomados en el parque La Malinche.

La exactitud del modelo propuesto que está reflejada por R^2 (coeficiente de determinación) en este caso tiene un valor de 47.3%, lo cual sugiere la clara existencia de otras variables o interacciones entre las variables que no fueron consideradas en el estudio que podrían explicar la otra parte de la variabilidad de la profundidad del mantillo.

La profundidad del mantillo es la distancia entre el suelo y la parte superior del material que se deposita de manera natural sobre el suelo. Con base en el modelo puede decirse que la variable de profundidad del mantillo depende del tipo vegetación, esto se explica con el hecho de que dependiendo del tipo de árboles será el tipo y la cantidad de material que se acumule en el suelo, por ejemplo los árboles con hojas latifoliadas (aile) son mucho más grandes que las de los pinos, generando que el mantillo acumulado sea diferente. El perfil transversal de la ladera ayudará a que el material se concentre o disperse según el tipo de perfil (cóncavo, convexo o plano) del círculo, el mismo caso es para la variable de microtopografía la cual dependiendo de la característica que tome ayudará a que la profundidad del mantillo sea mayor o no.

La orientación de la ladera es una variable que influye en la estimación de la profundidad del mantillo si y sólo si la altitud del terreno esta en la categoría de 3100-3220 msnm. Condicionado a que la distribución de los combustibles sea irregular, la variable de perturbación estará relacionada con la profundidad del mantillo.

La presencia de cauces en el círculo facilita el acceso de las personas a estas zonas, por esta razón condicionado con que no haya cauces en el terreno tres variables afectan el valor resultante para la profundidad del mantillo. Una es la variable de distribución de combustibles de forma irregular, la cual afecta positivamente, las otras dos se relacionan negativamente y son el índice de dominancia de árboles adultos en la categoría de 0.25 a 0.5 y la variable de perturbación en su categoría de no perturbado.

El índice de dominancia de árboles adultos afectará a la profundidad del mantillo, ya que los árboles grandes generaran mayor cantidad de material que se acumulará sobre el suelo, por esta razón condicionado a los valores que tome este índice, la orientación de la ladera será un factor que influirá en la estimación de la profundidad del mantillo con este modelo. Es en las depresiones en donde con mayor facilidad se acumula el material que forma el mantillo, esto explica que condicionado con que no haya depresiones en el círculo el hecho de que la vegetación sea pino disminuirá la profundidad del mantillo estimada.

3.4 Modelo para cantidad de material leñoso por clase diamétrica

La cantidad de material leñoso está dividido en cuatro clases, las cuales se basan en el diámetro del leño, estas clases son importantes ya que con base a ellas se determina la cantidad de humedad que retiene la madera, factor para el cálculo de la flamabilidad del leño. Por esta razón se consideró respetar las clases y a partir de los datos, ajustar un modelo para cada una de ellas.

Dado que el procedimiento para ajustar el modelo fue el mismo para las cuatro variables de cantidad de material leñoso, primero se describe el proceso general seguido para determinar estos modelos, posteriormente se especifican cada uno de los modelos finales resultantes, se analiza el ajuste de los modelos y se da una interpretación individual para los modelos propuestos en este trabajo.

La variable de cantidad de material leñoso es una variable de conteo, lo cual implica directamente que se debe considerar que el valor esperado de la variable de respuesta y tiene distribución Poisson. Por esta razón el primer modelo propuesto para estas variables fue:

$$E(y) = \mu = e^{x'\beta}$$

El paquete estadístico S-PLUS a través del proceso de stepwise aplicado a este modelo, seleccionó de las variables independientes aquellas que mejor explicaran la variabilidad de μ . Al verificar el cumplimiento del supuesto de que la varianza dependiera de la media, no se observó ninguna violación, por lo que por esta razón no fue necesario realizar alguna transformación. Al analizar la presencia de datos influyentes, los valores de las distancias de Cook fueron menores a uno, ninguna observación presentó indicios de ser influyente.

El modelo final propuesto para la cantidad de material leñoso fue:

$$E(y) = \mu = e^{x'\beta}$$

Las variables X y sus respectivos parámetros β varían dependiendo de la clase diamétrica. Se puede observar que el número de variables independientes en el modelo disminuye conforme aumenta el tamaño del diámetro del material leñoso esto se debe principalmente a que al aumentar la clase diamétrica el número de observaciones distintas de cero son cada vez menores, es decir en el muestreo se observó que la presencia de material leñoso con diámetro mayor a 7.5 cm. es muy pequeña, esto puede verse en la figura 6.

Las gráficas utilizadas para verificar el cumplimiento de los supuestos y las de las distancias de Cook, para cada una de las clases diamétricas se pueden ver en el anexo 3

3.4.1 Modelo para cantidad de material leñoso de 0 a 0.5 cm.

Para esta clase diamétrica las variables independientes de este modelo están en la tabla 5, también en esta tabla se pueden observar los valores estimados para los parámetros del modelo. El análisis gráfico del ajuste del modelo que se presenta en la figura 17, muestra que el modelo propuesto para la cantidad de material leñoso de 0 a 0.5 cm. es aceptable, ya que los puntos no distan significativamente de la recta identidad.

Variables	Valor del parámetro	Error Estándar	Valor t
DEPRES	-0.56	0.23	-2.39
CAUCE	0.23	0.08	3.01
PEND1	-4.42	2.00	-2.21
PERFLONG1	1.30	0.31	4.17
PERFLONG2	-0.54	0.25	-2.19
PERFTRANS1	-0.34	0.12	-2.76
MICROTOPO2	-1.00	0.50	-2.00
DISTCOMB1	0.59	0.11	5.50
DISTCOMB2	-0.91	0.11	-8.32
PERTURBA1	0.47	0.15	3.23
PERTURBA2	0.33	0.09	3.77
PEND1INDICE2	3.36	1.26	2.67
ALTI3ORIENT5	-3.28	0.55	-5.96
VEG3MICROTOPO1	-5.24	1.19	-4.41
ALTI1PEND1	-11.92	5.11	-2.33
ALTI2PEND1	4.63	2.26	2.04
ALTI3PEND1	7.28	2.22	3.29
ALTI4PEND1	6.78	3.13	2.17
ALTI1PEND2	6.49	2.00	3.25
ALTI3PEND2	1.45	0.73	1.97
ALTI1PEND3	6.05	1.49	4.07
ORIENT2MICROTOPO1	2.43	1.05	2.32
INDICE1PERFTRANS1	1.25	0.22	5.67
INDICE2PERFTRANS1	-1.53	0.30	-5.27
INDICE2PERFTRANS2	0.62	0.25	2.52
INDICE3PERFTRANS2	-0.84	0.28	-3.02
INDICE1ORIENT5	-2.36	0.52	-4.51
PEND1ORIENT1	2.10	0.75	2.79

Tabla 5. Variables y parámetros estimados del modelo para cantidad de material leñoso de 0 a 0.5 cm. en el parque nacional La Malinche.

Los leños de esta clase diamétrica, son material combustible ligero y por ello la presencia de depresiones y cauces en los círculos así como la distribución de los combustibles influyen en la cantidad de material leñoso que se pueda acumular. La presencia de brechas cortafuego, caminos o áreas incendiadas, características medidas por la variable de perturbación, pueden hacer que la cantidad de material leñoso aumente, en el caso de las brechas, en donde se facilita el depósito del material, o disminuya en el caso de los caminos y áreas incendiadas. El perfil longitudinal de la ladera que puede ser cóncavo, convexo

o plano ayudará a que la concentración de este material aumente o disminuya según sea el perfil.

La variable que mide la pendiente de la ladera también influye en el comportamiento del valor promedio de la cantidad de material leñoso de 0 a 0.5 cm. de diámetro pero esta condicionado con la altitud que tenga el terreno. La microtopografía con depresiones influirá en el valor promedio de dos formas diferentes: lo disminuye si y sólo si la vegetación del sitio es del tipo oyamel-desconocido, y lo aumenta si y sólo si la orientación es Este. El perfil transversal de la ladera variará el valor esperado de la cantidad de material leñoso de esta clase diamétrica condicionado con la variable del índice de dominancia de árboles adultos. Si la ladera tiene la mayor pendiente, de las consideradas en el trabajo, el hecho de que la orientación de la ladera sea NE afectará positivamente el valor esperado.

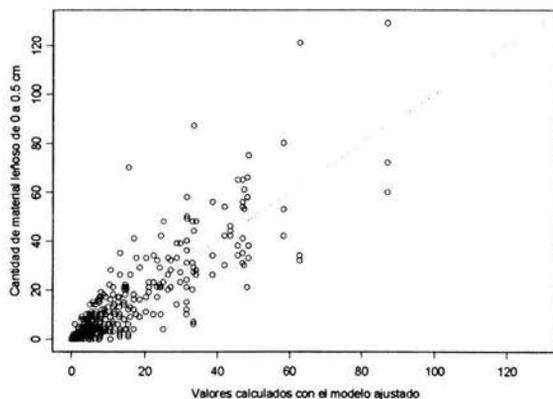


Figura 17. Valores de la cantidad de material leñoso de 0 a 0.5 cm. contra los valores obtenidos por el modelo ajustado a esta variable, con observaciones realizadas en el parque La Malinche.

3.4.2 Modelo para cantidad de material leñoso de 0.6 a 2.5 cm.

El modelo ajustado para la cantidad de material leñoso en esta clase diamétrica está determinado por las variables y sus respectivas estimaciones de los parámetros que se muestran en la tabla 6. Para analizar el ajuste del modelo, se graficaron los valores observados contra los valores calculados por el modelo propuesto como se puede ver en la figura 18, los puntos de la gráfica se encuentran lo suficientemente cerca de la recta identidad como para concluir que este modelo tiene un ajuste aceptable.

Variables	Valor del parámetro	Error Estándar	Valor t
DEPRES	0.57	0.18	3.17
CAUCE	0.39	0.10	3.77
ALTI1	-0.85	0.16	-5.36
ALTI2	-1.42	0.16	-8.83
ALTI4	0.68	0.17	4.13
VEG4	-7.19	3.42	-2.10
MICROTOPO1	1.30	0.47	2.75
DISTCOMB1	-3.39	1.55	-2.19
PERTURBA2	-1.48	0.39	-3.83
INDICE3ORIENT3	2.84	0.84	3.36
VEG1MICROTOPO1	1.49	0.34	4.33
VEG3MICROTOPO1	-1.72	0.63	-2.72
VEG4MICROTOPO1	0.56	0.19	2.99
VEG5MICROTOPO1	0.86	0.22	3.86
VEG6MICROTOPO1	1.48	0.27	5.49
VEG1MICROTOPO2	-1.00	0.26	-3.84
INDICE1VEG2	-9.64	2.14	-4.51
INDICE1VEG3	8.63	1.19	7.26
INDICE1VEG4	-1.54	0.37	-4.22
ORIENT2PERTURBA1	8.84	3.26	2.71
ORIENT3PERTURBA1	-2.16	0.77	-2.81
ORIENT4PERTURBA1	1.81	0.70	2.58
ORIENT3PERTURBA2	1.39	0.40	3.49
PERFTRANS1MICROTOPO1	0.95	0.47	2.03

Tabla 6. Variables y parámetros estimados del modelo para cantidad de material leñoso de 0.6 a 2.5 cm. en el parque nacional La Malinche.

De acuerdo con el modelo, el valor promedio de la cantidad de material leñoso de 0.6 a 2.5 cm. aumentará cuando en los círculo no haya presencia de cauces o depresiones, este valor es afectado por la altitud del terreno, tanto positiva como negativamente. Si la distribución de los combustibles en el círculo es irregular (DISTCOMB1) entonces el valor esperado de la cantidad de material leñoso de esta clase diamétrica disminuye.

El tipo de vegetación influye en el tipo y la cantidad de combustible que se acumule en el suelo, por esta razón condicionado con que el índice de dominancia de árboles adultos en el círculo este entre 0 y 0.25 la variable de vegetación influirá en el valor esperado de la variable ajustada con este modelo. Este valor esperado también esta relacionado con la microtopografía del terreno condicionado con el tipo de vegetación.

Si la orientación de la ladera es Este o Sur entonces el no existir perturbación en el terreno aumentará el valor promedio de la cantidad de material leñoso de 0.6 a 2.5 cm., mientras que la condición de que la orientación sea SE afecta negativamente y positivamente, si el terreno no esta perturbado o esta medianamente perturbado respectivamente.

Condicionado con que el perfil transversal de la ladera sea cóncavo, si la microtopografía de la ladera es irregular habrá un aumento en el valor esperado de la cantidad de material leñoso para esta clase diamétrica. La variable de orientación en su categoría SE influirá en el cálculo de la cantidad del material leñoso si y sólo si el índice de dominancia de árboles adultos esta entre 0.5 y 0.75.

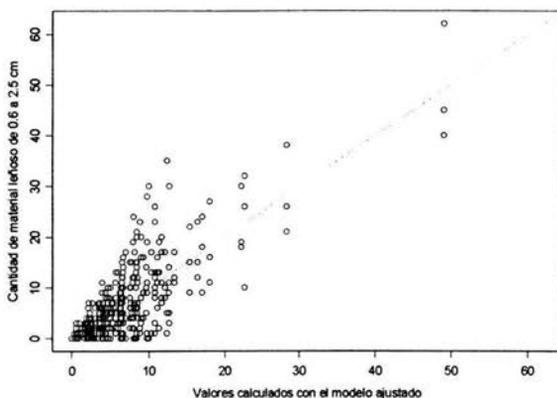


Figura 18. Valores de la cantidad de material leñoso de 0.6 a 2.5 cm. contra los valores obtenidos por el modelo ajustado a esta misma variable, con datos recolectados en el parque La Malinche.

3.4.3 Modelo para cantidad de material leñoso de 2.6 a 7.5 cm.

El modelo ajustado para esta clase diamétrica consta de las variables mencionadas en la tabla 7, en esta misma tabla están las estimaciones de los parámetros del modelo, como ya se mencionó el proceso que se siguió para determinar este modelo, asegura que éste es el mejor ajustable a los datos observados con ciertas características especificadas al principio de esta sección. En la figura 19 se hizo la gráfica de los valores reales observados contra los ajustados por el modelo y es posible observar unas líneas horizontales muy marcadas, esto se debe a que la variable de respuesta son conteos y las observaciones toman valores en un intervalo reducido, amplificando esta gráfica en relación a las de clases diamétricas

más pequeñas. De la figura 19, considerando la escala de la gráfica se puede decir que el modelo ajusta adecuadamente a los datos observados de la cantidad material leñoso de 2.6 a 7.5 cm.

Variables	Valor del parámetro	Error Estándar	Valor t
CAUCE	-0.55	0.22	-2.50
ALTI3	1.59	0.34	4.68
ALTI4	0.89	0.28	3.22
VEG2	2.23	0.91	2.44
VEG3	-3.85	0.95	-4.07
VEG5	-0.69	0.30	-2.25
ORIENT6	-1.19	0.34	-3.44
PERTURBA2	-0.57	0.15	-3.93
PEND1DISTCOMB1	-0.52	0.22	-2.36
PEND2DISTCOMB1	0.38	0.18	2.18
PEND3DISTCOMB1	-1.54	0.54	-2.88
PEND1DISTCOMB2	-0.86	0.35	-2.47
PEND3DISTCOMB2	2.10	0.57	3.71
INDICE1ORIENT6	-1.15	0.35	-3.25
CAUCEPERFTRANS1	1.30	0.41	3.20
CAUCEPERFTRANS2	-1.91	0.50	-3.84
CAUCEINDICE2	-1.17	0.42	-2.77

Tabla 7. Variables y parámetros estimados del modelo para cantidad de material leñoso de 2.6 a 7.5 cm. en el parque nacional La Malinche.

El tipo de vegetación determina en gran medida la clase diamétrica del material leñoso que se genere en esa área, por ejemplo si en el círculo la vegetación es pino los leños tendrán un diámetro menor a los que se encuentren en círculo donde haya ailes, ya que estos últimos tienen más ramas y de mayor diámetro. La altitud es una variable que se relaciona con la vegetación de la zona, el crecimiento de la mayoría de las plantas está limitado por la altitud. Por esta razón la cantidad de material leñoso con diámetro entre 2.6 y 7.5 cm. está relacionado con la altitud y el tipo de vegetación del área.

Los leños cuyo diámetro esté en esta clase son lo suficientemente grandes para ser usados como combustibles para fogatas, en zonas en donde la perturbación humana está presente, la cantidad de material leñoso de esta clase diamétrica deberá ser pequeña.

Condicionando a que el terreno tenga pendiente entre -75° y -11° , la distribución de los combustibles en el terreno (regular, irregular o manchones dispersos) será un factor que afectará el valor promedio de la cantidad de material leñoso de esta clase diamétrica. Si el terreno tiene un índice de dominancia de árboles adultos entre 0 y 0.25 y además la orientación del terreno es Oeste el valor esperado disminuirá. Las variables de perfil transversal de la ladera modificarán el valor esperado de la cantidad de material leñoso de 2.6 a 7.5 cm. de diámetro si y sólo si en el terreno no hay presencia de cauces.

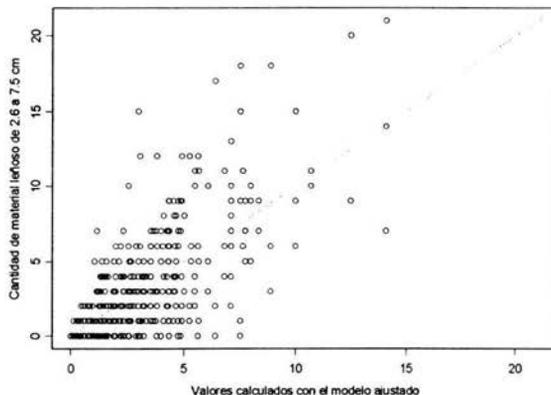


Figura 19. Valores de la cantidad de material leñoso de 2.6 a 7.5 cm., recolectados del parque nacional La Malinche, contra los valores obtenidos por el modelo ajustado a esta variable.

3.4.4 Modelo para cantidad de material leñoso mayor a 7.5 cm.

Para el ajuste de este modelo es importante mencionar que las observaciones mostraron que la presencia de material leñoso de esta clase diamétrica en el parque es muy reducida, esto a consecuencia de la influencia humana que tiene la zona. Las variables que forman parte de este modelo son las de la tabla 8. Como se puede ver en la figura 20 la variabilidad de la variable de respuesta explicada por este modelo es pequeña debido principalmente a que las observaciones distintas de cero son pocas. De lo anterior se puede decir que si se hacen predicciones para esta variable con este modelo el error que se cometa podría ser significativo.

VARIABLES	Valor del parámetro	Error Estándar	Valor t
DEPRES	-2.36	0.65	-3.63
ALTI2	-1.54	0.71	-2.17
ALTI4	1.56	0.69	2.27
PERFLONG1	2.33	0.77	3.02
PERFLONG2	-1.41	0.65	-2.17
PERFTRANS2	-0.76	0.31	-2.43
MICROTOPO1	-3.42	0.94	-3.64
MICROTOPO2	1.23	0.52	2.39
DISTCOMB1PERTURBA1	-1.59	0.65	-2.45
DISTCOMB1PERTURBA2	-1.18	0.55	-2.14

Tabla 8. Variables y parámetros estimados del modelo para cantidad de material leñoso mayor a 7.5 cm. en el parque nacional La Malinche.

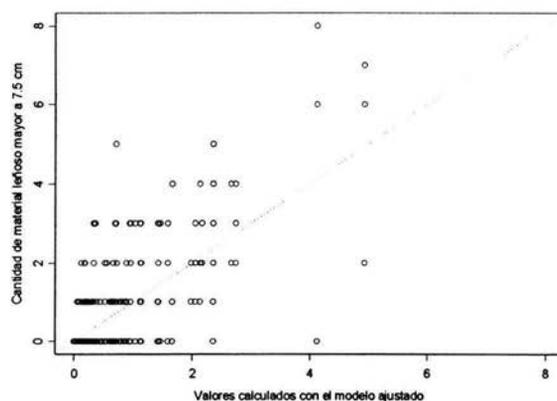


Figura 20. Valores de la cantidad de material leñoso mayor a 7.5 cm. contra los valores obtenidos por el modelo ajustado a esta variable. Las observaciones utilizadas son del parque La Malinche.

Conclusiones

Los incendios forestales son considerados como una de las causas principales de la destrucción forestal en nuestro país, por esta razón actualmente existe mucho interés en el estudio del comportamiento del fuego.

Con la finalidad de predecir y controlar los incendios, el estudio del comportamiento del fuego se basa fundamentalmente en el conocimiento de las características y la distribución de los combustibles. Para la realización de este trabajo las variables que se utilizaron para medir las características de los combustibles fueron: peso del mantillo, profundidad del mantillo y cantidad de material leñoso para cuatro clases diamétricas.

Las variables independientes, son las que describen las características físicas, el tipo vegetación y la influencia humana (antropogénicas) sobre el terreno, con estas observaciones se ajustaron modelos lineales para las variables dependientes de peso del mantillo, profundidad del mantillo y cantidad de material leñoso para cuatro clases diamétricas.

El modelo resultante de este trabajo para la variable de peso del mantillo esta formado por las siguientes variables: índice de dominancia de árboles adultos en la categoría de 0.25 a 0.5, perfil transversal de la ladera condicionado a ciertas categorías de la orientación de la ladera y perfil longitudinal y transversal solamente si no hay presencia de cauces en el círculo

Con base en el modelo la estimación de la variable de profundidad del mantillo depende del tipo vegetación, del perfil transversal de la ladera, de la microtopografía, de la orientación de la ladera condicionado a que la altitud sea 3100-3220 msnm o con la condición de que el índice de dominancia de árboles adultos este entre 0.25 y 0.75, de la perturbación condicionado con la variable de distribución de combustibles, del índice de dominancia de árboles adultos, de la distribución de combustibles y de la perturbación las tres condicionadas a la no presencia de cauces.

A partir del modelo obtenido para la variable de cantidad de material leñoso de 0 a 0.5 cm. de diámetro, esta variable esta relacionada con la presencia de depresiones y cauces, el perfil longitudinal de la ladera, la distribución de los combustibles, la perturbación, la pendiente condicionada a la altitud, de la microtopografía de la ladera si y sólo si la orientación es Este y del perfil transversal condicionado con la variable de índice de dominancia de árboles adultos.

Con el modelo resultante podemos decir que el valor promedio de la cantidad de material leñoso de 0.6 a 2.5 cm. depende de la presencia de cauces y depresiones en el círculo, la altitud, la distribución de los combustibles, la microtopografía condicionado a la vegetación, de la vegetación si y sólo si el índice de dominancia de árboles adultos esta entre 0 y 0.25 y de la perturbación condicionado con la orientación de la ladera

Los leños con diámetro mayor a 2.6 cm. son lo suficientemente grandes para ser usados como combustibles para fogatas. La mayoría de los sitios del muestreo son zonas en donde la perturbación humana está presente, aunque se buscó que fuera lo menor posible, los resultados de este trabajo muestran que la cantidad de material leñoso de las clases diamétricas de 2.6 y 7.5 cm. y mayores a 7.5 cm. es pequeña; y como consecuencia se tuvo que los modelos propuestos tienen un ajuste pobre.

Pese a la variabilidad de las variables combustibles (dependientes) si fue posible determinar un conjunto de variables independientes a través de las cuales se explica el comportamiento de los materiales combustibles. Los modelos propuestos, proporcionan un ajuste aceptable y satisfacen todos los supuestos de la teoría de los modelos lineales generalizados, por lo anterior a partir de estos modelos es posible realizar las predicciones de estos combustibles en otras zonas del parque nacional La Malinche con condiciones similares a las de los muestreos.

En conclusión, de los resultados obtenidos en este trabajo se puede decir que la técnica de modelos lineales generalizados es útil para construir modelos predictivos para las características de los materiales combustibles : la profundidad y peso del mantillo y las cantidades de material leñoso por clase diamétrica.

Anexo 1

Las siguientes tablas muestran los resultados obtenidos del análisis exploratorio realizado a las variables independientes. Se presenta la variable con sus categorías, el número de casos y su correspondiente porcentaje.

Presencia de Depresiones	No. de casos	Porcentaje
NO	120	82.19
SI	26	17.81

Presencia de Cauces	No. de casos	Porcentaje
NO	136	93.15
SI	10	6.85

Alitud	No. de casos	Porcentaje
2820-3030 msnm	30	20.55
3030-3100 msnm	30	20.55
3100-3220 msnm	28	19.18
3220-3428 msnm	29	19.86
3428-3850 msnm	29	19.86

Pendiente	No. de casos	Porcentaje
-75° a -26°	36	24.66
-26° a -17°	31	21.23
-17° a -11°	27	18.49
-11° a -7°	21	14.38
-7° a -3°	31	21.23

Índice de dominancia de árboles adultos	No. de casos	Porcentaje
0 a 0.25	34	23.29
0.25 a 0.5	27	18.49
0.5 a 0.75	24	16.44
0.75 a 1	61	41.78

Vegetación	No. de casos	Porcentaje
aile-pino	20	13.70
oyamel	6	4.11
oyamel-desconocido	10	6.85
pino	40	27.40
pino-aile	40	27.40
pino-aile-desconocido	10	6.85
pino h	20	13.70

Orientación	No. de casos	Porcentaje
Norte	0	0.00
NE	13	8.90
Este	19	13.01
SE	30	20.55
Sur	23	15.75
SO	32	21.92
Oeste	13	8.90
NO	16	10.96

Perfil longitudinal de la ladera	No. de casos	Porcentaje
concavo	8	5.48
convexo	13	8.90
plano	125	85.62

Perfil transversal de la ladera	No. de casos	Porcentaje
concavo	21	14.38
convexo	24	16.44
plano	101	69.18

Microtopografía de la ladera	No. de casos	Porcentaje
depresiones	11	7.53
irregular	82	56.16
regular	53	36.30

Distribución del material combustible	No. de casos	Porcentaje
irregular	18	12.33
manchones dispersos	8	5.48
regular	120	82.19

Perturbación	No. de casos	Porcentaje
No perturbado	59	40.41
Medianamente perturbado	70	47.95
Perturbado	17	11.64

Las siguientes tablas muestran los resultados obtenidos del análisis exploratorio realizado a las variables dependientes. Se presenta para cada una de las variables los valores mínimo y máximo, la media, la mediana, la varianza y la desviación estándar.

Peso del mantillo (gramos)	
Estadística	Valor
Valor mínimo	0
Media	61.96
Mediana	49.5
Valor máximo	439.4
Varianza	3078.84
Desviación estándar	55.48

Profundidad del mantillo (cm)	
Estadística	Valor
Valor mínimo	0
Media	4.44
Mediana	3.3
Valor máximo	32
Varianza	16.86
Desviación estándar	4.1

Cantidad de material leñoso mayor a 7.5 cm	
Estadística	Valor
Valor mínimo	0
Media	0.54
Mediana	0
Valor máximo	8
Varianza	1.23
Desviación estándar	1.11

Cantidad de material leñoso de 2.6 a 7.5 cm	
Estadística	Valor
Valor mínimo	0
Media	3.17
Mediana	2
Valor máximo	21
Varianza	12.9
Desviación estándar	3.59

Cantidad de material leñoso de 0.6 a 2.5 cm	
Estadística	Valor
Valor mínimo	0
Media	6.53
Mediana	4
Valor máximo	62
Varianza	54.52
Desviación estándar	7.38

Cantidad de material leñoso de 0 a 0.5 cm	
Estadística	Valor
Valor mínimo	0
Media	12.82
Mediana	6
Valor máximo	129
Varianza	298.71
Desviación estándar	17.28

Anexo 2

Resultados del ajuste del modelo para la variable de peso del mantillo.

Tabla 1. Variables del primer modelo.

Variables	Valores	Error Estándar	Valor t	Pr(> t)
Intersección	79.51		5.34	0.00
DEPRES	-5.94	4.04	-1.47	0.14
INDICE1	2.22	4.69	0.47	0.64
INDICE2	5.59	4.94	1.13	0.26
INDICE3	6.54	5.12	1.28	0.20
ORIENT1	49.02	18.98	2.58	0.01
ORIENT2	37.62	9.58	3.93	0.00
ORIENT3	-28.88	17.85	-1.62	0.11
ORIENT4	-13.95	18.52	-0.75	0.45
ORIENT5	-17.77	17.82	-1.00	0.32
ORIENT6	-8.60	17.40	-0.49	0.62
DISTCOMB1	-12.66	6.17	-2.05	0.04
DISTCOMB2	5.89	7.67	0.77	0.44
PERTURBA1	7.01	16.39	0.43	0.67
PERTURBA2	-3.30	16.50	-0.20	0.84
ORIENT1PERTURBA1	-50.52	19.83	-2.55	0.01
ORIENT2PERTURBA1	9.11	13.71	0.66	0.51
ORIENT3PERTURBA1	36.19	18.67	1.94	0.05
ORIENT4PERTURBA1	31.75	20.23	1.57	0.12
ORIENT5PERTURBA1	-0.71	22.54	-0.03	0.98
ORIENT6PERTURBA1	-22.94	19.83	-1.16	0.25
ORIENT1PERTURBA2	88.15	23.06	3.82	0.00
ORIENT3PERTURBA2	-7.37	19.18	-0.38	0.70
ORIENT4PERTURBA2	-19.16	20.00	-0.96	0.34
ORIENT5PERTURBA2	-10.48	19.23	-0.55	0.59
DEPRESPERTURBA1	-22.12	5.94	-3.72	0.00
DEPRESPERTURBA2	3.94	4.88	0.81	0.42

Tabla 2. Variables del segundo modelo.

Variables	Valores	Error Estándar	Valor t	Pr(> t)
Intersección	5.25	0.21	24.55	0.00
CAUCE	-0.98	0.16	-6.12	0.00
INDICE1	-0.09	0.10	-0.87	0.38
INDICE2	0.22	0.11	2.05	0.04
INDICE3	0.08	0.11	0.75	0.46
ORIENT1	-0.37	0.33	-1.10	0.27
ORIENT2	5.30	1.07	4.93	0.00
ORIENT3	2.05	0.28	7.38	0.00
ORIENT4	-0.71	0.28	-2.57	0.01
ORIENT5	-1.71	0.36	-4.82	0.00
ORIENT6	-1.56	0.28	-5.62	0.00
PERFLONG1	0.12	0.19	0.63	0.53
PERFLONG2	-0.36	0.20	-1.79	0.07
PERFTRANS1	1.88	0.49	3.80	0.00
PERFTRANS2	-0.65	0.37	-1.77	0.08
CAUCEPERFTRANS1	1.11	0.21	5.22	0.00
CAUCEPERFTRANS2	-1.93	0.27	-7.25	0.00
ORIENT1PERFTRANS1	-3.08	0.68	-4.55	0.00
ORIENT2PERFTRANS1	7.66	2.02	3.78	0.00
ORIENT3PERFTRANS1	2.32	0.31	7.55	0.00
ORIENT4PERFTRANS1	-2.94	0.55	-5.36	0.00
ORIENT5PERFTRANS1	-1.44	0.35	-4.14	0.00
ORIENT6PERFTRANS1	-1.86	0.34	-5.41	0.00
ORIENT1PERFTRANS2	2.43	0.45	5.36	0.00
ORIENT2PERFTRANS2	-2.87	0.97	-2.95	0.00
ORIENT4PERFTRANS2	2.39	0.34	7.06	0.00
CAUCEPERFLONG1	-0.37	0.19	-1.92	0.06
CAUCEPERFLONG2	0.51	0.19	2.64	0.01

Tabla 3. Variables del tercer modelo.

Variables	Valores	Error Estándar	Valor t	Pr(> t)
Intersección	5.34	0.19	27.63	0.00
CAUCE	-1.57	0.16	-10.01	0.00
INDICE1	-0.09	0.09	-0.97	0.33
INDICE2	0.22	0.10	2.27	0.02
INDICE3	0.08	0.10	0.83	0.41
ORIENT1	0.14	0.30	0.46	0.65
ORIENT2	3.45	0.99	3.49	0.00
ORIENT3	2.55	0.26	9.98	0.00
ORIENT4	-0.21	0.26	-0.81	0.42
ORIENT5	-1.21	0.32	-3.72	0.00
ORIENT6	-1.06	0.26	-4.13	0.00
PERFLONG1	0.12	0.17	0.70	0.48
PERFLONG2	-0.36	0.18	-1.99	0.05
PERFTRANS1	0.78	0.46	1.70	0.09
PERFTRANS2	0.53	0.35	1.49	0.14
CAUCEPERFTRANS1	1.70	0.20	8.44	0.00
CAUCEPERFTRANS2	-3.11	0.27	-11.55	0.00
ORIENT1PERFTRANS1	-2.58	0.61	-4.21	0.00
ORIENT2PERFTRANS1	3.45	1.88	1.84	0.07
ORIENT3PERFTRANS1	2.82	0.28	10.02	0.00
ORIENT4PERFTRANS1	-2.43	0.50	-4.89	0.00
ORIENT5PERFTRANS1	-0.93	0.32	-2.93	0.00
ORIENT6PERFTRANS1	-1.36	0.32	-4.31	0.00
ORIENT1PERFTRANS2	2.43	0.41	5.94	0.00
ORIENT2PERFTRANS2	-0.51	0.91	-0.56	0.57
ORIENT4PERFTRANS2	2.39	0.31	7.83	0.00
CAUCEPERFLONG1	-0.37	0.18	-2.13	0.03
CAUCEPERFLONG2	0.51	0.17	2.93	0.00

Anexo 3

Las gráficas de los residuos contra los valores ajustados por el modelo para verificar los supuestos de varianza y las distancias de Cook para todas las observaciones para identificar posibles observaciones atípicas, para cada uno de los modelos ajustados a los datos de cantidad de material leñoso por clase diamétrica.

Cantidad de material leñoso de 0 a 0.5 cm

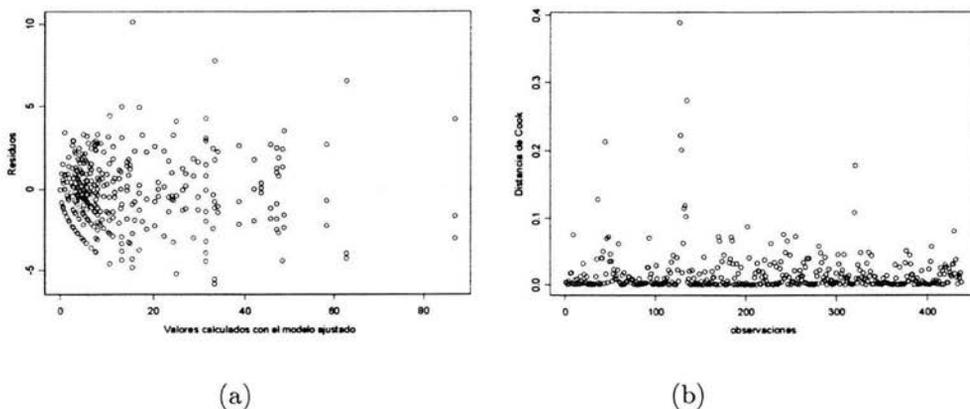


Figura 1 (a) residuos contra los valores ajustados por el modelo (b) Distancia de Cook para cada una de las observaciones

Cantidad de material leñoso de 0.6 a 2.5 cm

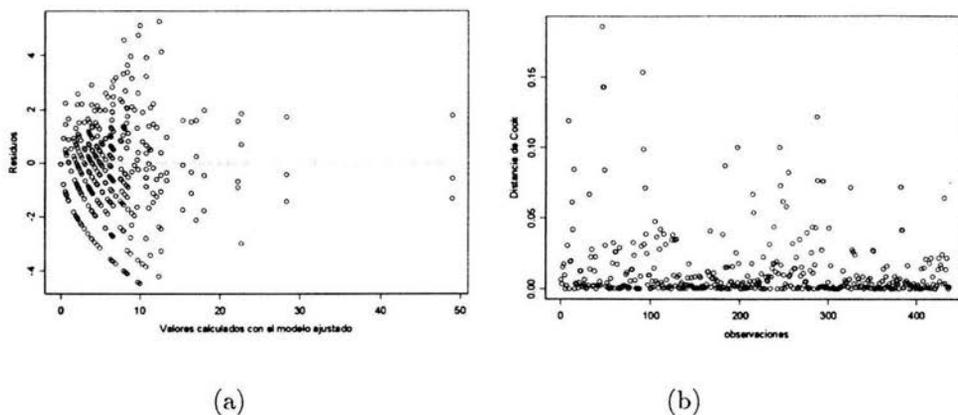


Figura 2 (a) residuos contra los valores ajustados por el modelo

(b) Distancia de Cook para cada una de las observaciones

Cantidad de material leñoso de 2.6 a 7.5 cm

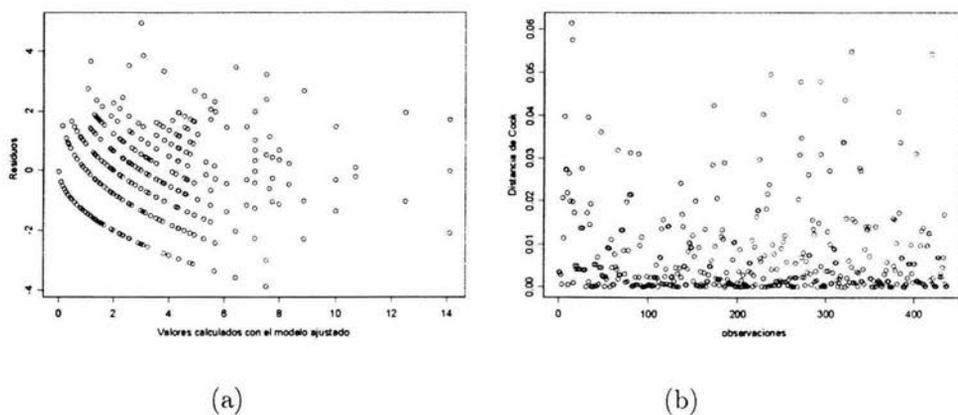


Figura 3 (a) residuos contra los valores ajustados por el modelo

(b) Distancia de Cook para cada una de las observaciones

Cantidad de material leñoso mayor a 7.5 cm

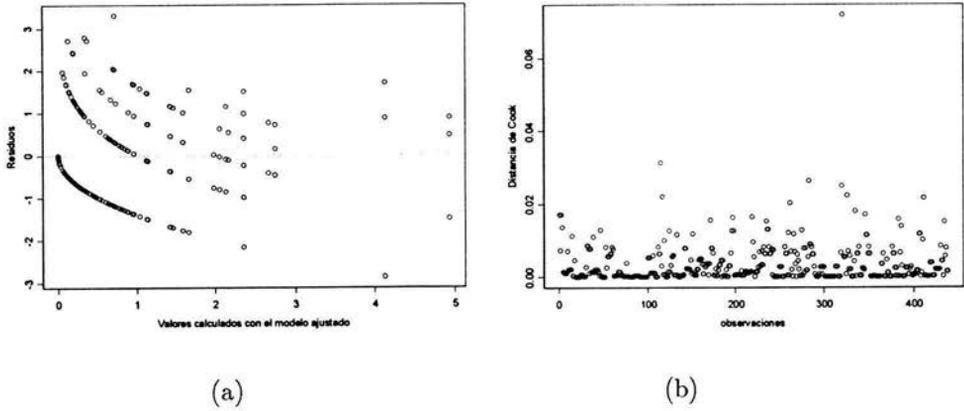


Figura 4 (a) residuos contra los valores ajustados por el modelo

(b) Distancia de Cook para cada una de las observaciones

ESTA TESIS NO SALE
DE LA BIBLIOTECA

Bibliografía

- [1] Arno, Stephen F., Allison-Bunnell, Steven (2002). "Flames in Our Forest. Disaster or Renewal?". Island Press. pp. 37-45.
- [2] Brown, J.K. (1974) "Handbook for inventorying downed woody material". USDA Forest Service, Intermountain Forest Range Experimental Station, Ogden, Utah. Gen. Tech. Rep. INT-16. 24 pp.
- [3] Brown, J.K. (2000) "Ecological principles, shifting fire regimes and management considerations". In: Wildland Fire in Ecosystems. Effects of fire on flora. USDA Forest Service Gen. Tech. Rep. RMRS-GTR.42 vol. 2. pp. 185-203.
- [4] Cedeño, S.O. (1999) "Incendios Forestales en México en 1998: Magnitud, Extensión, Combate y Control". En: Incendios forestales y agropecuarios: prevención e impacto y restauración de los ecosistemas UNAM/PUMA-IPN, pp. 1-18.
- [5] Chambers, J., Hastie, T. (1993) "Statistical Models in S". Chapman and Hall. 608 pp.
- [6] Dobson, Annette J. (1990) "An Introduction to Generalized Linear Models". Chapman and Hall 174 pp.
- [7] Espejel, A (1996) "La Malinche: Una versión retrospectiva de su deterioro y conservación". Gaceta Ecológica. No.41 119 pp.

- [8] Espejel, A. (1998) "La importancia y el deterioro del recurso bosque en La Malinche, estado de Tlaxcala". Tesis de Maestría en Análisis Regional. Universidad Autónoma de Tlaxcala. 194 pp.
- [9] Fahrmeir, L. Tutz, G. (1994) "Multivariate Statistical Modelling Based on Generalised Linear Models" Springer. 425pp.
- [10] Flores, G. J.G. (2001) "Modeling the spatial variability of forest fuel arrays". Ph.D. Dissertation, Department of Forest Sciences, Colorado State University, Fort Collins, Colorado. 201 pp.
- [11] Kent, Martin and Coker, Paddy (1992) "Vegetation Description and Analysis. A Practical Approach" John Wiley & Sons. 363 pp.
- [12] McCullagh, P., Nelder, J.A. (1989) "Generalized Linear Models". Chapman and Hall. 511 pp.
- [13] McRae, D.J.; Alexander, M.E.; and Stocks B.J. (1979) "Measurement and description of fuels and fire behavior on prescribed burns". A handbook. Dep. Environ., Can. For. Serv., Sault Ste. Marie, Ontario Report 0-X-287. 44 pp.
- [14] Myers, R., Montgomery, D., Vining, G. (2002) "Generalized Linear Models with Applications in Engineering and the Sciences" Wiley Interscience. 342 pp.
- [15] Neter, John; Wasserman, William; Kutner, and Michael, H. (1989) "Applied Linear Regression Models". Irwin. 667 pp.
- [16] Peña del Valle, A. (2003) "Captura de Carbono: Un estudio en el Parque Nacional La Malinche, Tlaxcala-Puebla". Tesis de Maestría en Geografía, Facultad de Filosofía y Letras, UNAM. 85 pp.

- [17] Reich, R., Lundquist, J., Bravo, V. (2004) "Spatial models for estimating fuel loads in the Black Hills, South Dakota, USA". *International Journal of Wildland Fire*. No 13 pp.119-129.
- [18] Schmid-Haas, P.; Baumann, E.; Wernwe, J. (1993) "Forest inventories by unmarked permanent samples plots: instructions". Birmensdorf, Swiss Federal Institute for Forest, Snow and Landscape Research. 135 pp.
- [19] Vélez, Ricardo (2000). "La Defensa Contra Incendios Forestales. Fundamentos y Experiencias". McGraw-Hill.
- [20] Villers, L., Arellano, A. y Peña del Valle A. (2001) "Los incendios forestales en el volcán La Malinche y su relación con ciertos aspectos físicos y humanos en el área". XVI Congreso Nacional de Geografía, Mérida 6 al 8 de Junio.
- [21] Villers, L (2001) "Programa de reforestación y áreas afectadas por los incendios forestales en el volcán La Malinche". Manuscrito Informe Técnico Final, rendido al Fondo Mexicano para la Conservación de la Naturaleza.
- [22] Villers, L (2001) "Comportamiento del fuego y evaluación de riesgo a incendios en áreas forestales de México: Un estudio en el volcán La Malinche". Manuscrito Solicitud de apoyo a proyecto de investigación del CONACYT.
- [23] Villers, L., Peña del Valle A. y Arellano, A. (2002) "Recurrencia de los incendios forestales en el volcán La Malinche y la presencia del fenómeno de El Niño 1998". En Sánchez Crispín (Editor) México en su unidad y diversidad territorial, tomo I:162-175 pp.