



Modelos mixtos aplicados a la productividad de hojarasca

Mixed models applied to the litterfall productivity

José F. Zea^a
josezea@usantotomas.edu.co

Miguel A. Murcia^b
miguel.murcia.r@gmail.com

Fidel E. Poveda^c
fierpogo@gmail.com

Resumen

Se propone una metodología para realizar un ajuste de un modelo lineal mixto incorporando una estructura de varianza y de correlación serial adecuada partiendo desde el modelo de regresión lineal clásico siguiendo la estrategia de Zuur et al. (2009). Se utilizan diferentes gráficos de diagnóstico de los residuales y pruebas de razón de verosimilitud a fin de garantizar que la inclusión de determinados efectos fijos y aleatorios esté justificado e igualmente para hallar una estructura de varianzas y covarianzas que permita capturar la heteroscedasticidad y la correlación serial de los residuales del modelo. Se utilizan datos ecológicos tomados de un estudio llevado a cabo por Murcia (2013) en los alrededores de la cuenca del río Pamplonita (Norte de Santander, Colombia), en dicho estudio se tomaron diferentes mediciones en los bosque altoandino y subandino, el peso de caída de hojarasca de 40 colectores ubicados en los dos bosques es cuantificado durante doce meses. Estos datos son analizados utilizando modelos mixtos.

Palabras clave: caída de hojarasca, estructura de varianzas y covarianzas, modelos mixtos, regresión lineal.

Abstract

A methodology is developed to adjust a linear mixed model incorporating a variance-covariance structure starting from a the classical linear regression model following Zuur strategy, Zuur et al. (2009). Different diagnostic plots of residuals and likelihood ratio tests are used to choose fixed and random effects and to find an

^aDocente. Facultad de Estadística. Universidad Santo Tomás. Colombia.

^bDocente. Universidad de Pamplona. Colombia.

^cBiólogo. Universidad Nacional de Colombia. Colombia.

appropriate covariance structure to capture the heteroskedasticity and serial correlation of model residuals. Ecological data from Murcia (2013) research around the river basin Pamplonita (Norte de Santander, Colombia) are used in this paper, different measurements are taken in the high Andean and sub-Andean forests. The weight of litter fallen from 40 collectors located in two forest was quantified for 12 months. This data are analyzed using mixed models.

Keywords: linear regression models, litterfall, mixed model, variance-covariance structure.

1. Introducción

Las medidas repetidas o datos longitudinales corresponden a mediciones que se efectúan en una o más ocasiones en el tiempo para diferentes variables y/o covariables. En los últimos 30 años las metodologías estadísticas para analizar medidas repetidas han tenido un notable desarrollo teórico y una facilidad para su implementación, gracias a los avances de la computación y a un mayor acceso a software especializado, lo cual ha permitido la aplicación de diferentes modelos a ramas tan diversas como las ciencias médicas, biológicas, la industria, la educación, entre otras, (Davis 2002). Las medidas repetidas son recolectadas a través del tiempo y poseen una estructura de correlación entre las mediciones que son tomadas para un mismo individuo, adicionalmente las mediciones repetidas presentan heteroscedasticidad en las observaciones lo cual hace inviable el uso del modelo lineal clásico, ya que este requiere que se cumplan los supuestos de independencia y heteroscedasticidad (Galecky 2013).

Fitzmaurice (2008) menciona que las medidas repetidas se pueden abordar mediante diferentes metodologías estadísticas tales como el análisis univariado (en particular el modelo en dos etapas desarrollado por el Instituto de Salud de los Estados Unidos, conocida como metodología NHI por sus siglas en inglés), el análisis de perfiles utilizando el análisis multivariante de varianza (conocido en la literatura como MANOVA, por sus siglas en inglés de *Multivariate Analysis of Variance*), y finalmente el análisis de medidas repetidas utilizando modelos mixtos. Los dos primeros enfoques tiene facilidades en su implementación, a pesar de esto el uso de modelos mixtos tiene menos restricciones en cuanto a la configuración de los datos y permite la incorporación de efectos fijos y aleatorios, a pesar de esto se requiere mayor capacidad de cómputo y no fue si no hasta años recientes en que los modelos mixtos se popularizaron.

En la metodología NHI se desarrolla un modelo de dos etapas, en la primera etapa se ajustan modelos de regresión (se utiliza el método de mínimos cuadrados o el de máxima verosimilitud) los cuales permiten explicar el comportamiento de una variable de interés en términos del tiempo para cada uno de los individuos del estudio, obteniéndose para cada individuo unos coeficientes de regresión. En la segunda etapa los coeficientes de regresión obtenidos en la primera etapa son analizados utilizando los métodos paramétricos o no paramétricos convencionales,

como por ejemplo las prueba de hipótesis sobre la media, la prueba de diferencia de medias y la prueba de Wilcoxon. Las pruebas anteriores permiten tener un conocimiento aproximado de cómo es el comportamiento de la variable de interés en el tiempo y comparar para diferentes grupos de individuos el efecto del tiempo sobre la variable de interés. Aunque la metodología NIH es ampliamente utilizada presenta algunos inconvenientes, en primer lugar no se tiene en cuenta la autocorrelación de las mediciones utilizadas en la primera etapa y por otra parte las covariables asociadas a los individuos (invariables en el tiempo) solamente permiten modelar las medidas resumen obtenidas en la primera etapa para los individuos, lo cual conlleva una gran pérdida de información, finalmente esta metodología solo está diseñada para incorporar efectos fijos al modelo no pudiéndose ajustar efectos aleatorios asociados a los individuos.

Otro enfoque muy popular para analizar medidas repetidas es el análisis de perfiles por medio del análisis de varianza multivariado (MANOVA), (Rao 1973). Utilizando este enfoque se puede estimar la estructura de varianzas y covarianzas de los residuales del modelo aunque se exige que estos tengan una distribución normal multivariada, además existen restricciones en la manera como deben estar configurados los datos: la variable independiente se debe disponer en una matriz de tamaño $n \times T$, en donde en la fila i y la columna j se encuentra la medición de la variable independiente para el individuo i en el tiempo j . En el análisis de perfiles, las mediciones de cada uno de los individuos debe ser tomada en los mismos tiempos y no pueden existir datos faltantes de la variable de estudio en ninguno de los periodos. Las covariables que se introducen por otro lado no pueden ser variantes en el tiempo y no se permite introducir efectos aleatorios al modelo (Laird & Ware 1982).

Los modelos mixtos a diferencia de las dos metodologías anteriores permiten analizar datos que disponen de diferente mediciones temporales para los individuos, y se facilita analizar los datos con información faltante, en estos modelos se pueden introducir factores aleatorios que tengan en cuenta las particularidades de cada uno de los individuos sin tener que estimar una gran cantidad de parámetros.

En este artículo se ilustra el uso de los modelos mixtos utilizando un conjunto de dato provenientes de un estudio ecológico llevado a cabo en Norte de Santander, Colombia.

Se analiza la dinámica de la caída de hojarasca (hojas, ramas, troncos, estructuras reproductivas, líquenes y detritus) en dos biomas: bosque altoandino y subandino en la cuenca del Río Pamplonita. En cada uno de estos dos biomas se realizó un muestreo de 20 áreas de 100 metros por 20 metros en cada una de las cuales se instalaron veinte colectores en los cuales se recolecta todo el material desprendido de diferentes especies arbóreas la cual se conoce como hojarasca. Se llevan a cabo mediciones mensuales del peso de hojarasca desprendida en cada uno de los colectores durante un año, obteniéndose un total de 12 mediciones (ver Tabla 1).

Por otro lado se lleva a cabo un análisis del peso acumulada de la hojarasca desprendida en cada uno de los 40 colectores (Murcia 2013). El análisis de datos longitudinales no debe abordarse desde el modelo de regresión clásico ya que los

supuestos de homoscedasticidad y de no autocorrelación de los residuales asociados a este modelo rara vez se verifican.

El uso de modelos mixtos permite incorporar la estructura de autocorrelación de los residuales y posibilita de manera sencilla corregir los problemas de heteroscedasticidad del modelo al incorporarle una estructura adecuada de varianza al modelo, en este trabajo se desarrolla la estrategia de ajuste de un modelo lineal mixto y se muestran algunas de las herramientas de verificación de los supuestos de este modelo. Así mismo, el ajustar un modelos mixtos permitirá realizar pruebas de hipótesis confiables a fin de evaluar la significancia de los parámetros.

Tabla 1: *estructura de los datos. Fuente: elaboración propia.*

Bioma (Comunidad)	Colector	Tiempo
Altoandino	1	1
Altoandino	1	2
⋮	⋮	⋮
Altoandino	1	12
⋮	⋮	⋮
Altoandino	20	1
Altoandino	20	2
⋮	⋮	⋮
Altoandino	20	12
Subandino	1	1
Subandino	1	2
⋮	⋮	⋮
Subandino	1	12
⋮	⋮	⋮
Subandino	20	1
Subandino	20	2
⋮	⋮	⋮
Subandino	20	12

2. Estructura de varianza y covarianza

Zuur et al. (2009) resume las estructuras de varianzas más usuales en la práctica, todas ellas disponibles con el software estadístico R, estas estructuras de la varianza permiten corregir los problemas de heteroscedasticidad y fueron las que se probaron para los datos del artículo.

Tabla 2: estructura de varianzas. Fuente: elaboración propia.

Estructura de varianza	Expresión
Homoscedástica	$V(\epsilon_i) = \sigma^2$
Idéntica por grupo	$V(\epsilon_{ij}) = \sigma_{ij}^2$
Fija	$V(\epsilon_i) = \sigma^2 x_i$
Potencia	$V(\epsilon_i) = \sigma^2 x_i ^{2\delta}$
Potencia por grupo	$V(\epsilon_{ij}) = \sigma^2 x_{ij} ^{2\delta_j}$
Exponencial	$V(\epsilon_i) = \sigma^2 e^{2\delta x_i}$
Exponencial por grupo	$V(\epsilon_{ij}) = \sigma^2 e^{2\delta x_{ij}}$
Constante más una potencia	$V(\epsilon_i) = \sigma^2 (\delta_1 + x_i ^{\delta_2})^2$

Las estructuras de correlación más populares para datos longitudinales se presentan a continuación, se encuentran disponibles en el paquete `nlme` de R, (Pinheiro et al. 2014).

Tabla 3: estructura de correlaciones. Fuente: elaboración propia.

Estructura de varianza	Expresión
Compuesta simétrica	$Corr(Y_{ij}, Y_{ij'}) = \rho, k = 1, 2, \dots$
Autoregresivo de orden 1	$Corr(Y_{ij}, Y_{i,j+k}) = \rho^k, k = 0, 1, 2, \dots$
Autoregresivo continuo de orden 1	$h(s, \rho) = \rho^s, s \geq 0$
No estructurada	$Corr(Y_{ij}, Y_{i,j'}) = \rho^{j \cdot j'}, j < j'$

3. El modelo mixto

El modelo mixto permite extender el modelo de regresión lineal múltiple haciendo posible el análisis de diferentes configuraciones de datos tales como datos jerárquicos, datos longitudinales, entre otros; así mismo ajusta la variable de interés en términos de un componente sistemático y un componente aleatorio. La parte sistemática del modelo contiene lo que se denomina en la literatura estadística efectos fijos, la parte aleatoria del modelo contiene lo que se denomina en la literatura efectos aleatorios. Los modelos mixtos permiten tratar adecuadamente los problemas de heteroscedasticidad y autocorrelación presentes en los datos, (Verbeke & Molenberghs 2000).

Diferentes autores han discutido sobre el significado de lo que es un efecto fijo y un efecto aleatorio. Gelman (2007) desarrolla una extensa discusión sobre estos dos conceptos, en uno de los enfoques se denomina efecto fijo a las variables cuyos niveles son estudiados en su totalidad (tipo de bosque, tiempo) y como efectos aleatorios, las variables cuyos niveles no son estudiados en su totalidad, sino solamente algunos de sus niveles de interés y cuya selección provienen de una muestra aleatoria (en este caso los colectores).

Se dispone de información para n individuos, para los cuales se mide una variable de interés continua (dependiente) y unas variables explicativas (independientes). El modelo mixto para el individuo i , $i = 1, \dots, n$, puede expresarse como sigue:

$$\mathbf{Y}_i = \mathbf{X}_i\beta + \mathbf{Z}_i\mathbf{b}_i + \epsilon_i \quad (1)$$

Donde, Y_i corresponde a la variable de interés, $\mathbf{b}_i \sim N(0, \mathbf{D})$, $\epsilon_i \sim N(0, \Sigma_i)$, $b_1, b_2, \dots, b_N, \epsilon_1, \epsilon_2, \dots, \epsilon_N$, son variables aleatorias independientes. \mathbf{X}_i es la matriz de diseño correspondiente a los efectos fijos de tamaño $p \times 1$, \mathbf{Z}_i es la correspondiente a los efectos aleatorios de tamaño $q \times 1$, β es el vector de efectos fijos, \mathbf{b}_i es el vector de efectos aleatorios, ϵ_i es el vector de errores.

Un caso particular del modelo mixto es el modelo con intercepto aleatorio:

$$Y_i = \alpha + \beta x_i + b_{0i} + \epsilon_i \quad (2)$$

El modelo se puede expresar matricialmente de la siguiente forma:

$$\begin{pmatrix} y_{i1} \\ \vdots \\ y_{i2} \\ \vdots \\ y_{in_i} \end{pmatrix} = \begin{pmatrix} 1 & x_{i1} \\ 1 & x_{i2} \\ \vdots & \vdots \\ 1 & x_{in} \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} b_{0i} + \begin{pmatrix} e_{i1} \\ e_{i1} \\ \vdots \\ e_{in_i} \end{pmatrix} \quad (3)$$

Otro caso particular de un modelo mixto es el modelo con intercepto y pendiente aleatorio el cual se utilizará para analizar la dinámica de la caída de hojarasca:

$$Y_i = \alpha + \beta x_i + b_{0i} + b_{1i}x_i + \epsilon_i \quad (4)$$

Este modelo puede expresarse

$$\begin{pmatrix} y_{i1} \\ \vdots \\ y_{i2} \\ \vdots \\ y_{in_i} \end{pmatrix} = \begin{pmatrix} 1 & x_{i1} \\ 1 & x_{i2} \\ \vdots & \vdots \\ 1 & x_{in} \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \begin{pmatrix} 1 & x_{i1} \\ 1 & x_{i2} \\ \vdots & \vdots \\ 1 & x_{in} \end{pmatrix} \begin{pmatrix} b_{0i} \\ b_{1i} \end{pmatrix} + \begin{pmatrix} e_{i1} \\ e_{i1} \\ \vdots \\ e_{in_i} \end{pmatrix} \quad (5)$$

La única diferencia con la ecuación (2) es la aparición de b_{1i} que corresponde al valor de la pendiente aleatoria para cada individuo donde:

$$\begin{pmatrix} b_{0i} \\ b_{1i} \end{pmatrix} \sim N(0, D) \quad (6)$$

siendo

$$D = \begin{pmatrix} d_{11}^2 & d_{12} \\ d_{12} & d_{22}^2 \end{pmatrix} \quad (7)$$

De la ecuación (1), se puede obtener:

$$y_i = N \sim (\mathbf{X}_i\beta, \mathbf{V}_i) \quad \mathbf{V}_i = \mathbf{Z}_i\mathbf{D}\mathbf{Z}'_i + \Sigma_i \quad (8)$$

3.1. Estrategia para el ajuste de un modelo mixto

Zuur et al. (2009) sugieren la siguiente estrategia para el ajuste de un modelo mixto: se parte del ajuste de un modelo de regresión lineal clásica, tras incorporar la estructura de variabilidad y de autocorrelación de los residuales y los efectos aleatorios, así se obtiene un modelo lineal mixto.

1. Ajustar un modelo de regresión lineal que incluya los efectos principales y de interacción fijos (componente sistemática del modelo) necesarios para obtener un buen ajuste del modelo.
2. Ajustar una estructura de varianzas al modelo de regresión lineal ajustado en el paso 1 (escoger la estructura más adecuada en términos de criterios de información (AIC, BIC) y que verifique los diagnósticos gráficos de residuales).
3. Ajustar el componente aleatorio del modelo mixto, incluir los efectos aleatorios del modelo y escoger el modelo que tenga los mejores efectos aleatorios en términos de la prueba de razón de verosimilitudes, la cual permite comparar dos modelos anidados con los mismos efectos fijos y estructura de varianzas pero que difieren en un efecto aleatorio.
4. Realizar diagnósticos sobre los residuales del modelo resultante en el paso anterior, si el modelo presenta problemas de heteroscedasticidad o autocorrelación se ajusta una estructura a la matriz de varianzas y covarianzas del modelo anterior. Si es necesario se cambia la especificación de los efectos aleatorios del paso anterior y se prueban diferentes estructuras de varianza y covarianza hasta obtener un modelo en cuyos residuales no se evidencie violación de los supuestos.

4. Resultados

Se observa la caída de hojarasca en cuarenta colectores, en periodos espaciados de aproximadamente de 30 días, para un total de 12 tiempos de observación. Se comenzó el estudio el 8 de Diciembre de 2011 y se finalizó el 7 de Diciembre de 2012.

En este estudio se observó el peso en gramos de la cantidad de hojarasca que cayó en los colectores en cada uno de los 12 meses del estudio. Por otro lado, también se exploró cómo es el comportamiento del peso acumulado de hojarasca obtenida en cada colector, por ejemplo, la hojarasca acumulada hasta el tiempo 4 corresponde a la caída de hojarasca de los tiempos 1,2,3 y 4. Respecto a la acumulación de hojarasca no existen estudios ecológicos previos en los biomas altoandino y subandino, por lo cual se explora en este artículo cómo es la dinámica en la acumulación de hojarasca.

4.1. Análisis Descriptivo

En esta sección se realiza un análisis exploratorio de la caída de hojarasca recolectada al final de cada mes y de la caída de hojarasca acumulada.

4.1.1. Caída de hojarasca

En primer lugar se explora cómo es la relación entre el tiempo y la caída de hojarasca (CHT, medida en gramos) para los bosques altoandino y subandino.

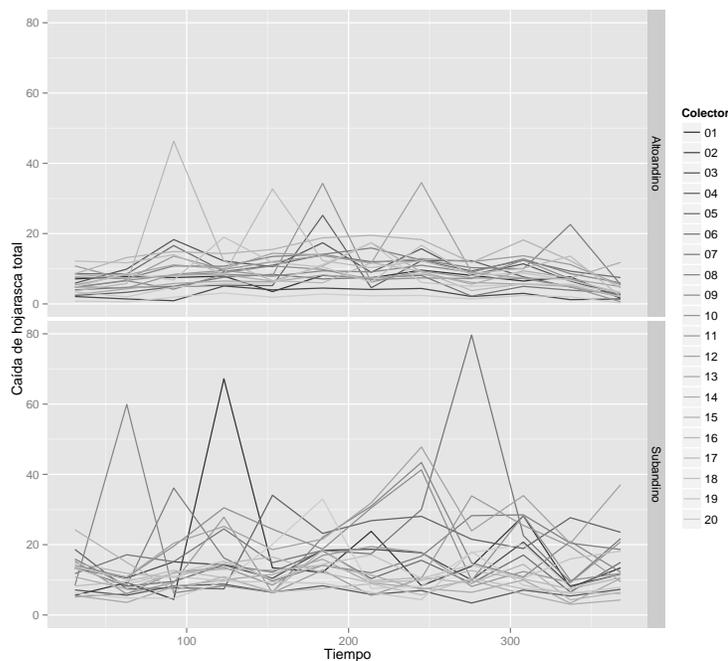


Figura 1: gráfico de perfiles caída de hojarasca en el tiempo según el tipo de bosque. Fuente: elaboración propia.

Como se evidencia en la figura 1, la caída de hojarasca tiene un comportamiento constante a través del tiempo para la mayor parte de los colectores, asimismo se observa que se presentan diferencias considerables en la producción de hojarasca entre los diferentes colectores. Se destaca una mayor producción de hojarasca y variabilidad en los colectores ubicados en el bosque subandino.

Tabla 4: *estadísticas resumen CHT (gramos) por tipo de bosque. Fuente: elaboración propia.*

	Comunidad	min	p25	p50	Media	Desv.est	CVE	p75	max
1	Altoandino	0.30	4.97	7.75	8.61	5.78	67.21	11.10	46.30
2	Subandino	3.03	8.30	11.80	14.60	10.07	68.95	18.20	79.70

En la Tabla 4 se observa que la producción promedio de hojarasca en los colectores ubicados en el bosque subandino es mayor que en aquellos localizados en el bosque altoandino; confirmándose lo observado en la figura anterior. Se presenta una mayor variabilidad en la caída de hojarasca en los colectores del bosque subandino, aunque en términos relativos esta diferencia no es muy relevante (ver coeficiente de variación). Comparando cada uno de los cuartiles, el mínimo y el máximo se confirma que la distribución de la caída de hojarasca es mayor en los colectores localizados en el bosque subandino que en los del bosque altoandino.

4.1.2. Caída de hojarasca acumulada en el tiempo

Como se mencionó al principio de esta sección, existe un interés primordial de conocer en detalle como es el comportamiento de la caída de hojarasca acumulada en el tiempo:

En el gráfico 2 se observa que la acumulación en la caída de hojarasca en el tiempo para los diferentes colectores sigue una tendencia lineal, aunque las pendientes varían considerablemente para cada colector, es decir, las velocidades de acumulación de hojarasca en el tiempo varían de colector a colector. Por otra parte, el intercepto con el eje y no parece diferir significativamente en los diferentes colectores. Es interesante observar que La relación lineal entre el tiempo y la caída de hojarasca es más fuerte en el bosque altoandino que en el subandino, además que se presentan mayores fluctuaciones en la caída de hojarasca en el bosque subandino.

Tabla 5: *estadísticas resumen caída de hojarasca acumulada por tipo de bosque. Fuente: elaboración propia.*

	Comunidad	min	p25	p50	Media	Desv.est	CVE	p75	max
1	Altoandino	0.80	20.20	51.00	57.20	41.04	71.77	88.60	170.00
2	Subandino	5.13	42.70	82.10	92.30	64.81	70.22	128.00	305.00

En la Tabla 5 se puede observar que la acumulación promedio de hojarasca es considerablemente mayor en el bosque subandino que en el altoandino.

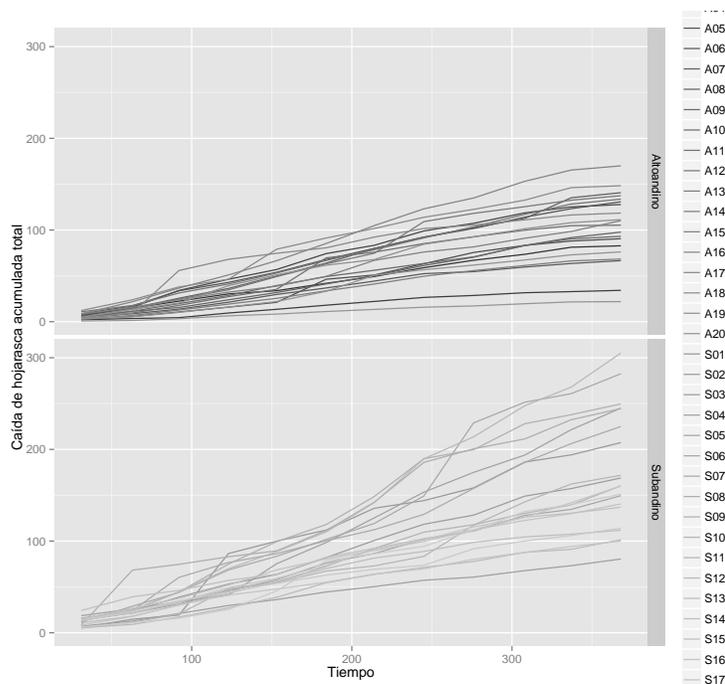


Figura 2: caída de hojarasca acumulada por tiempo y según el tipo de bosque.
Fuente: elaboración propia.

En las siguientes sección se ajusta un modelo mixto para explicar la relación de la caída de hojarasca y la caída de hojarasca acumulada en términos del tiempo y el tipo de bosque.

4.2. Ajuste a través de modelos mixtos

Se lleva a cabo un modelo mixto para la caída de hojarasca en términos del tiempo, el tipo de bosque y considerando como efecto aleatorio el colector. Se utiliza la estrategia definida en Zuur et al. (2009) para el ajuste de un modelo mixto. Inicialmente un modelo sobresaturado de efectos fijos, posteriormente se ajusta la estructura de varianzas para el modelo y finalmente se incorpora el componente aleatorio para así obtener el modelo mixto.

4.2.1. Modelo mixto para los datos de la caída de hojarasca

Usando la estrategia de Zuur se ajusta un modelo lineal (efectos fijos) para la caída de hojarasca en términos del tipo de bosque, el tiempo y la interacción entre estas dos variables. Como se evidencia en la figura 3 los residuales del modelo de

regresión lineal no tienen un comportamiento constante (homoscedástico) ni en el tiempo ni en los dos tipos de bosques (bioma), lo cual nos da indicios de que el modelo requiere de un ajuste en la estructura de varianza. Adicionalmente, el comportamiento de los residuales varía considerablemente de individuo a individuo, lo cual evidencia que es necesario incorporar el efecto de cada uno de los colectores. Dado el número considerable de estos, se ajustará el modelo considerando el colector como un efecto aleatorio y no un efecto fijo.

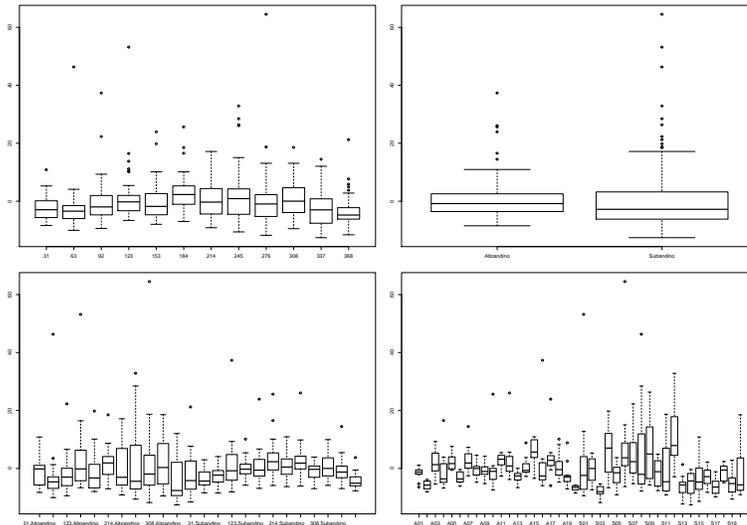


Figura 3: *homogeneidad del Modelo regresión lineal Datos no Acumulados. Fuente: elaboración propia.*

Al modelo de regresión lineal se le incorpora una estructura de varianza en donde se considera para cada uno de los tiempos y los tipos de bosques una varianza diferente, se obtienen 12 estimaciones de varianza (once para los tiempos y una para los bosques). El tiempo 31 y el bosque subandino se toman como categorías de referencia para las estimaciones de la varianza. En la figura 4 se observa cómo se estabiliza la varianza de los residuales para los diferentes tiempos y bosques al incorporar la estructura de varianza. Sin embargo, aún incorporando la estructura de varianza la variabilidad de la caída de hojarasca sigue siendo considerable en cada uno de los colectores, lo que nos da indicios de la necesidad de incorporar los colectores en el modelo. En el paso siguiente se estima un modelo con la estructura de varianzas estimado en el paso anterior y el colector como efecto aleatorio. Se ajusta un modelo con intercepto aleatorio y con estructura de varianza diferente en cada comunidad y tiempo.

```
## Linear mixed-effects model fit by REML
## Data: CHAcum
## AIC BIC logLik
```

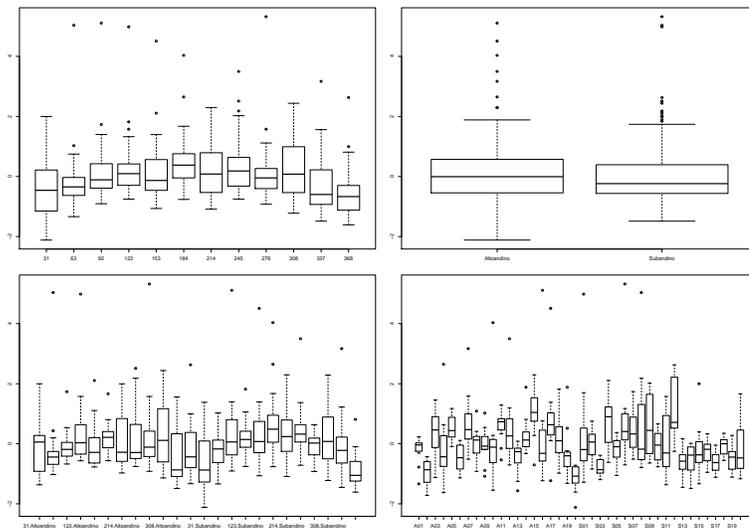


Figura 4: *homogeneidad del Modelo con Estructura de Varianza en Datos no Acumulados. Fuente: elaboración propia.*

```
## 3249 3323 -1606
##
## Random effects:
## Formula: ~1 | Colector
## (Intercept) Residual
## StdDev: 3.488 2.8
##
## Combination of variance functions:
## Structure: Different standard deviations per stratum
## Formula: ~1 | Comunidad
## Parameter estimates:
## Altoandino Subandino
## 1.000 1.893
## Structure: Different standard deviations per stratum
## Formula: ~1 | Tiempo
## Parameter estimates:
## 31 63 92 123 153 184 214 245 276 308 337 368
## 1.000 1.608 2.530 1.951 1.579 2.222 1.354 2.236 2.131 1.116 1.204 1.574
## Fixed effects: Hojarasca ~ Tiempo * Comunidad
## Value Std.Error DF t-value p-value
## (Intercept) 7.566 0.9436 438 8.019 0.0000
## Tiempo 0.002 0.0023 438 0.811 0.4177
## ComunidadSubandino 5.138 1.5843 38 3.243 0.0025
## Tiempo:ComunidadSubandino 0.005 0.0049 438 1.036 0.3009
```

```
## Correlation:
##                (Intr) Tiempo CmdddS
## Tiempo          -0.483
## ComunidadSubandino -0.596  0.288
## Tiempo:ComunidadSubandino 0.226 -0.467 -0.616
##
## Standardized Within-Group Residuals:
##      Min      Q1      Med      Q3      Max
## -2.28399 -0.48442 -0.05811  0.33659  5.52529
##
## Number of Observations: 480
## Number of Groups: 40
```

Con el modelo mixto de intercepto aleatorio se confirma que la caída de hojarasca es mayor en la comunidad subandina (en más de 5 gramos), y no se observa un efecto significativo del tiempo en la caída de hojarasca, manteniéndose constante en el tiempo.

En el modelo anterior, el intercepto aleatorio de cada colector traslada la estimación fija del intercepto, obteniéndose diferentes ajustes lineales para cada uno de los colectores. La varianza asociada al intercepto aleatorio $\sigma_{b_0}^2$ es de 3.488, por otro lado, la varianza de los residuales, σ_e^2 es de 2.8. Con el modelo mixto, se controla una parte de la variabilidad de los datos al considerar el colector como efecto aleatorio, por lo tanto se logra controlar indirectamente algunas de las condiciones ambientales y del suelo asociadas al colector.

Con la prueba de razón de verosimilitudes se puede comparar los tres modelos ajustados: el modelo de regresión lineal, el modelo de regresión lineal con ajuste de la estructura de varianza y el modelo mixto con estructura de la varianza:

```
anova(modelo1, modelo2, modelo3)
##      Model df  AIC  BIC logLik  Test L.Ratio p-value
## modelo1   1  5 3405 3426 -1698
## modelo2   2 17 3316 3387 -1641 1 vs 2 113.14 <.0001
## modelo3   3 18 3249 3323 -1606 2 vs 3  69.79 <.0001
```

De la salida anterior se puede concluir que el modelo con intercepto aleatorio es considerablemente mejor que los otros dos modelos, por otra parte, el modelo lineal tiene un desempeño muy pobre frente a los otros dos modelos

4.2.2. Modelo mixto de los datos de la caída de hojarasca acumulada

Como se mencionó anteriormente, es de interés conocer la dinámica de la caída de hojarasca acumulada. Inicialmente se ajusta un modelo lineal para la caída de hojarasca acumulada en términos del tiempo y el tipo de bosque. Los residuales de este modelo presentan problemas de heteroscedasticidad, en la medida que

aumenta el tiempo se presenta una mayor dispersión en los residuales, asimismo los residuales presentan una mayor dispersión en el bosque subandino que en el bosque altoandino, ver figura 5.

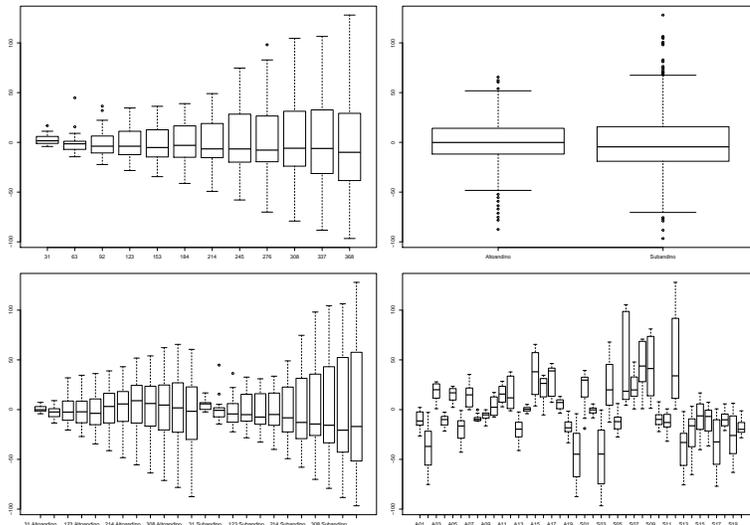


Figura 5: homogeneidad del modelo lineal para datos acumulados. Fuente: elaboración propia.

Al modelo lineal ajustado anteriormente se le ajusta un modelo con estructura de varianza diferenciada para cada tiempo y tipo de bosque, ver figura 6.

Al incorporar la estructura de varianzas se obtiene un modelo estadísticamente mejor que el modelo lineal ajustado inicialmente, como se refleja en la prueba de razón de verosimilitudes mostrada a continuación:

```
anova(modelo_acum1, modelo_acum2)
##           Model df  AIC  BIC logLik  Test L.Ratio p-value
## modelo_acum1     1  5 4683 4704  -2336
## modelo_acum2     2 17 4360 4431  -2163 1 vs 2   346.6 <.0001
```

En el gráfico 6 se presenta el comportamiento de los residuales a través de las covariables del modelo (tiempo y tipo de bosque) y para cada uno de los colectores. El modelo con ajuste de estructura de varianzas, permite corregir la heteroscedasticidad que se presentaba en las covariables, sin embargo se evidencia la presencia de grandes fluctuaciones de los residuales al interior de cada colector, lo cual es un indicio de la necesidad de introducir como efecto aleatorio el colector.

Lo anterior permite estabilizar la varianza de los residuales del modelo y obtener una estimación más precisa de los efectos fijos (tiempo y tipo de bosque).

Para ajustar un efecto aleatorio asociado al colector existen diferentes alternativas,

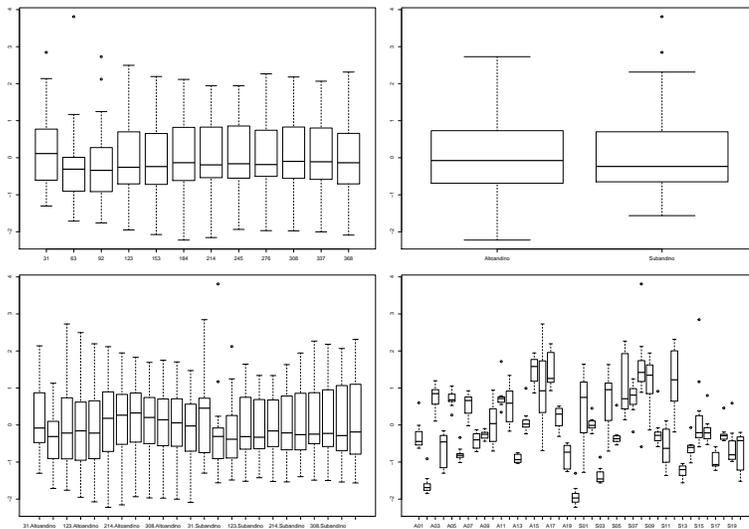


Figura 6: *homogeneidad del Modelo para Datos Acumulados con Estructura de Varianza. Fuente: elaboración propia.*

puede ajustarse un modelo con intercepto aleatorio, pendiente aleatoria y pendiente e intercepto aleatorio, se ajustan los tres modelos; sin embargo únicamente el modelo de pendiente aleatorio es satisfactorio en términos de convergencia. Lo anterior puede explicarse debido a que el punto de partida de la caída de hojarasca para los diferentes colectores es cero, por lo tanto las líneas ajustadas para los diferentes colectores pasan muy cerca al origen. 2

Un modelo con pendiente aleatoria (sin intercepto aleatorio) es ajustada para la caída de hojarasca acumulada y se modela la estructura de varianzas de la misma forma a como se ajustó en el modelo presentado anteriormente ($\sigma_{hj}^2 = \sigma_h^2 \sigma_j^2$).

```
## Linear mixed-effects model fit by REML
## Data: CHAcum
##   AIC   BIC logLik
## 3526 3601 -1745
##
## Random effects:
## Formula: ~(Tiempo - 1) | Colector
##           Tiempo Residual
## StdDev: 0.1431      2.9
##
## Combination of variance functions:
## Structure: Different standard deviations per stratum
## Formula: ~1 | Comunidad
## Parameter estimates:
```

```

## Altoandino Subandino
##      1.000      2.003
## Structure: Different standard deviations per stratum
## Formula: ~1 | Tiempo
## Parameter estimates:
##      31      63      92      123      153      184      214      245      276      308
## 1.0000 1.8852 2.5427 2.5883 2.6124 2.2899 2.2042 2.0683 0.9277 0.7223
##      337      368
## 1.1630 2.5175
## Fixed effects: Hojarasca_acum ~ Tiempo * Comunidad
##
##              Value Std.Error  DF t-value p-value
## (Intercept)   -4.423   0.5900 438  -7.496 0.0000
## Tiempo         0.312   0.0321 438   9.735 0.0000
## ComunidadSubandino -1.291  1.3207  38  -0.978 0.3343
## Tiempo:ComunidadSubandino 0.189  0.0456 438   4.150 0.0000
## Correlation:
##              (Intr) Tiempo CmnddS
## Tiempo         -0.065
## ComunidadSubandino -0.447  0.029
## Tiempo:ComunidadSubandino 0.046 -0.704 -0.102
##
## Standardized Within-Group Residuals:
##      Min      Q1      Med      Q3      Max
## -3.66749 -0.68236 -0.05029  0.41870  3.40042
##
## Number of Observations: 480
## Number of Groups: 40

```

La varianza del efecto aleatorio correspondiente a la pendiente es de 0.1431, la varianza de los residuales por otro lado es de 2.9. El ajuste de los colectores mejora considerablemente con esta pendiente, las estimaciones del parámetro para los bosques altoandino 1.000 y para el bosque subandino 2.003; es decir la velocidad de acumulación en la caída de hojarasca en el bosque subandino es cerca del doble a la del bosque altoandino.

```

anova(modelo_acum1, modelo_acum2, modelo_acum4)
##           Model df  AIC  BIC logLik  Test L.Ratio p-value
## modelo_acum1    1  5 4683 4704  -2336
## modelo_acum2    2 17 4360 4431  -2163 1 vs 2   346.6 <.0001
## modelo_acum4    3 18 3526 3601  -1745 2 vs 3   836.5 <.0001
# modelo_acum4$coefficients

```

Entre los tres modelos ajustados, el modelo con pendiente aleatorio presenta el mejor ajuste en términos de mayor verosimilitud y menores criterios de información (Akaike, BIC).

5. Conclusiones

En el presente artículo se ajustó un modelo lineal mixto partiendo del modelo de regresión lineal clásico, se incorporaron efectos fijos y aleatorios al modelo así como una estructura de varianzas para los residuales. El modelo ajustado permitió cuantificar los efectos del tiempo y del tipo de bosque en la dinámica de la caída de la hojarasca e igualmente en la acumulación del peso de la hojarasca, para ambos modelos se introdujeron como efectos aleatorios los colectores.

Ni el tiempo ni la interacción del tiempo con el tipo de bosque son significativos para modelar el peso de la hojarasca caída mensualmente; las diferencias en el peso de la hojarasca son explicadas principalmente por el tipo de bosque. En contraste, para la caída de hojarasca acumulada tanto el tiempo como la interacción entre el tiempo y el tipo de bosque son variables significativas, siendo la relación entre la caída de hojarasca acumulada y el tiempo de tipo lineal.

Recibido: 01 de abril de 2014

Aceptado: 13 de agosto de 2014

Referencias

- Davis, C. (2002), *Statistical Methods for the Analysis of Repeated Measurements*, Springer.
- Fitzmaurice, G. (2008), *Longitudinal Data Analysis*, Chapman and Hill.
- Gelman, A. (2007), *Data Analysis Using Regression and Multilevel/Hierarchical Models*, Cambridge University Press, New York.
- Laird, N. M. & Ware, J. H. (1982), 'Random effects models for longitudinal data', *Biometrics* **38**(4), 963–974.
- Murcia, A. (2013), 'Respiración del suelo en una comunidad sucesional de pastizal del bosque altoandino en la cuenca del río pamplonita', *Caldasia* **30**(2), 337–353.
- Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D. & R Core Team (2014), *nlme: Linear and Nonlinear Mixed Effects Models*, <http://cran.r-project.org/package=nlme> edn.
- Rao, C. R. (1973), *Linear statistical inference and its applications*, Wiley, New York.
- Verbeke, G. & Molenberghs, G. (2000), *Linear Mixed Model for Longitudinal Data*, Springer, New York.
- Zuur, A., Ieno, E., Walker, N., Saveliev, A. & Smith, G. (2009), *Mixed Effects Models and Extensions in Ecology with R*, Springer, New York.