

## Hierarchical linear mixed models in multi-stage sampling soil studies

Adriana A. Gili · Elke J. Noellemeyer ·  
Mónica Balzarini

Received: 9 May 2011 / Revised: 23 July 2012 / Published online: 15 August 2012  
© Springer Science+Business Media, LLC 2012

**Abstract** The issue of variances of different soil variables prevailing at different sampling scales is addressed. This topic is relevant for soil science, agronomy and landscape ecology. In multi-stage sampling there are randomness components in each stage of sampling which can be taken into account by introducing random effects in analysis through the use of hierarchical linear mixed models (HLMM). Due to the nested sampling scheme, there are several hierarchical sub-models. The selection of the best model can be carried out through likelihood ratio tests (LRTs) or Wald tests, which are asymptotically equivalent under standard conditions. However, when the comparison leads to a restricted hypothesis of variance components, standard conditions are not maintained, which leads to more elaborated versions of LRTs. These versions are not disseminated among environmental scientists. The present study shows the modeling of soil data from a sampling where sites, fields within sites, transects within fields, and sampling points within transects were selected in order to take samples from different vegetation types (open and shade). For soil data, several sub-models were compared using Wald tests, classic LRTs and adjusted LRTs where the distribution of the test statistic under the null hypothesis is the Chi-square mixture of Chi-square

---

A. A. Gili (✉)  
CONICET—Department of Statistics and Experimental Design, Faculty of Agronomy,  
National University of La Pampa,  
La Pampa, Argentina  
e-mail: agili@agro.unlpam.edu.ar

E. J. Noellemeyer  
Department of Soil Studies, Faculty of Agronomy, National University of La Pampa,  
La Pampa, Argentina

M. Balzarini  
CONICET—Department of Statistics and Biometry, Faculty of Agronomy,  
National University of Cordoba, Cordoba, Argentina

distributions. The inclusion of random effects via HLMM and suggested by the latest version of LRT allowed us to detect effects of vegetation type on soil properties that were not detected under a classical ANOVA.

**Keywords** Bulk density · Likelihood ratio test (LRT) · Texture · Total organic carbon · Variance components

## 1 Introduction

In environmental studies of soil and vegetation the fact that usually no fixed experimental design can be set up and the high spatial variability of soil variables make it challenging to meet requirements of traditional statistical methods oriented to comparisons of means. [Lardy et al. \(2002\)](#), despite the high number of samples, found few differences in C and P stocks of soils by applying classic ANOVA models. Similarly, [Feral et al. \(2003\)](#) could only detect statistically significant differences in C contents of soils in the extreme sites of a climatic gradient using traditional lineal models. [Ringrose et al. \(1998\)](#) already showed that vegetation and soil C data had very high variation coefficients (59–67 %), and despite the high intensity of sampling, few differences could be detected. [Wang et al. \(2007\)](#) used one-way ANOVA in paired samples under and outside tree canopy to evaluate the effect of canopy at different soil sites in a rainfall gradient, but the results referred to one site at a time. It is generally acknowledged that relationships between variables may change in sign or magnitude from one scale to another [i.e., scale dependence; ([Wiens 1989](#); [Levin 1992](#))]. Therefore soil scientists have to deal with a set of uncertainties from the design and sampling stage to the analysis of mean differences among conditions of interest. A number of methods of multi-scale analysis have been used to characterize and quantify variability in soil data ([Pelletier et al. 2009a](#)).

The stratification of sampling according to expected variation has shown to improve the ability to detect changes in C stocks ([Post et al. 2001](#)) and other key environmental variables, and [Heim et al. \(2009\)](#) suggest a paired sample approach with a very intensive sample pattern ( $n \geq 25$ ) at each site. Other methods include the classical geostatistical analysis ([Wang et al. 2009](#)) and the use of multivariate methods to assess the multi-scale variability of relationships between variables ([Pelletier et al. 2009b](#)). A disadvantage of conventional geostatistical analysis is that it might fail to identify the spatial dependence correctly when sampling locations are very irregularly distributed within a region ([Lark 2005](#)). Therefore the residual variability between and within sampling sites needed to contrast treatment means may be poorly modeled with classical geostatistical approaches.

A convenient way to compare two treatment conditions taking into account the high variability and correlation in the data is through a linear model taking into account the multi-stage nature of the sampling. In multi-stage sampling there are randomness components in each stage of sampling which can be considered at the modeling stage by introducing random effects through the use of hierarchical linear mixed models (HLMM) which allow modeling the underlying variability for a more precise inference about the treatment effects than those obtained from classical geostatistical techniques

in soil analyses. Even though HLMM constitute an efficient form of statistical analysis for inferring the effects of conditions of interest in multi-stage sampling (Raudenbush and Bryk 2002; Snijders and Bosker 2000), they are rarely used in environmental and ecological studies. In multi-scale studies, the units of analysis are nested within levels of single-factor random samples, and the levels of that factor are nested within the levels of another factor of greater hierarchy, and so on and so forth until the highest level. Level 1 is the level with the lowest hierarchy and it represents the lowest partition of the hierarchy; level  $k$  is the level with the highest hierarchy when there are  $k$  factors that define the sampling stages. For example, in a soil study for which various sample sites are randomly selected, within which even more sample sites are randomly selected, and from which several samples are extracted at a specific depth, each soil sample (unit of analysis) is nested within (or belongs exclusively to) a sample point, and consequently is also nested within the sample site to which that point belongs. In this two-stage sample, the Level 2 model will consider the effect of the site and sample-point factors, while the Level 1 model will only consider the effect of the site factor (the factor with the highest hierarchy); in other words, it will classify the samples only by site, and as a result, samples taken from different points within the same sample site will belong to the same partition.

This type of sampling strategy generates correlated data, since it is assumed that data from a single class are more similar to each other than to data from different classes. This phenomenon of data aggregation due to such partitions is not compatible with the assumption of data independence, and thus it should be taken into account (via the modeling of the corresponding variance and covariance structure) in order to effectively compare the means of the conditions of interest, which are usually represented in a model by fixed effects. This way, models of analysis have both fixed effects (due to the conditions that one is interested in comparing) and random effects which are different from the classic error term (due to the type of sampling). The way in which sampling is done turns out to be a conditioning factor for the analysis of the effects of interest, since the variability of observations among and between the groups, or classes, defined by the sampling strategy can be very different.

When using classical inference based on the sampling strategy, variance structure is examined through the calculation of error estimators obtained according to the sampling strategy used and which are independent of the data (Cochran 1980). In contrast, in the context of HLMM, inference is done through the modeling of these variance components and their covariates based on the identification of the model that best adjusts to the group of data under study (Balzarini 2002). The modeling of the variance structure of the data does not only provide interpretations regarding random variation due to sampling, but also allows for the identification of over-parameterizations or under-parameterizations that could lead to inefficiencies in the inference about fixed effects.

In analyzing data generated by multi-stage sampling, HLMMs allow us to determine whether the variations within each level of the hierarchical structure (different sample stages) have an impact on the dependent variable measured in level 1 of the data. Several models, each one associated with the hierarchical structure up to a specific level, are potential candidates for the analysis of a single group of data. Each model is nested within another, and the model found in the level below a given model

in the hierarchy is said to be the reduced model of the one above it. The group of fixed effects and/or covariance parameters of the reduced model can be obtained through the imposition of restrictions on the model of greatest hierarchy, also known as the reference model. If we assume that the parameters of the fixed effects do not change, models can be compared to each other by assuming that certain variance and/or covariance parameters are equal to zero. In order to select the “best” model, that is, a model that is parsimonious in the number of variance parameters used and also best explains the variability in the dependent variable, tools based on hypothesis proofs of variance components are used (Searle et al. 1992). Among the formal tools used are procedures based on the normal asymptotic theory, such as the Wald test, and procedures based on the Likelihood ratio test (LRT) (Cox and Hinkley 1990).

This paper illustrates the performance of the three statistical tests used to evaluate variance components in the context of an HLMM used to deal with uncertainties due to hierarchical sampling. The motivation for using these models arose from a study based on multi-stage sampling designed to evaluate the effect that vegetation patches, determined by the presence or absence of tree canopy, have on different soil properties in the Caldenal savanna of Central Argentina. The objective is to identify the variance structure of the linear model that best describes the variation among the data, in order to produce precise estimates regarding the effect of vegetation patch on soil texture, bulk density and C contents.

## 2 Materials and methods

### 2.1 Data

The group of data used in the illustration comes from a multi-stage sampling soil study carried out in the Caldenal, in the province of La Pampa, Argentina (between 63° and 66° W longitude and between 35° and 39° S latitude). The vegetation in the area of study is an open forest of *Prosopis caldenia* with gramineous arboreal strata. At present, it is possible to differentiate two well-defined vegetation patches: those without tree canopy and those shaded by tree canopy. In the patches without canopy, forage species predominate and bovine foraging is concentrated here; thus a permanent extraction of plant biomass predominates, which accounts for the scarce deposit of litter and the heterogeneous distribution of material that returns to the ground, which could modify the soil's properties. In contrast, the patches with shaded by canopy, where non-forage species predominate, foraging is scarce, and thus soil exhibits a greater level of homogeneity.

### 2.2 Multi stage sampling scheme

Data come from a multi-stage sampling soil study carried out in the study area. Within the area, six sites were selected according to differences in soil texture, topography and vegetation structure. Within each site, lots were set up, and transects of 100 m each were randomly established in the lots. In each transect, three sampling points were established, with approximately 30 m distance between points. The hierarchy

of multi-stage sampling is defined as follows: site (level 4), lots within sites (level 3), transects within lots (level 2), and sampling points within transects (level 1). In each point, vegetation patches were identified (open and canopy-shade vegetation) and soil samples were taken at two depth levels (1: 0–06 m and 2: 0.06–0.12 m). The following soil variables were evaluated: Texture, Bulk Density (BD) and Total Organic Carbon (TOC).

### 2.3 Soil analyses

Soils core samples were taken at 0–6 and 6–12 cm depth and after drying, BD ( $\text{g cm}^{-3}$ ) was determined on all soil samples by determining dry weight of the soil cores, before grinding them to pass a 2 mm sieve. The soil clay and silt content (Texture, %) of each sample was determined by the hygrometer method of Bouyoucos (Gee and Bauder 1986). Carbon contents of bulk soil (TOC, %) were determined by oxidation with potassium dichromate in acid medium at 120 °C, and colorimetric valuation (Soon and Abboud 1991).

### 2.4 HLMM and significance tests

The following HLMMs were adjusted for each independent depth level:

#### A. Level 4 model

$$Y_{ijklmn} = \mu + \tau_i + S_j + L(S)_{k(j)} + T(L)_{l(k)} + P(T)_{m(l)} + \varepsilon_{ijklmn}$$

#### B. Level 3 model

$$Y_{ijklmn} = \mu + \tau_i + S_j + L(S)_{k(j)} + T(L)_{l(k)} + \varepsilon_{ijklmn}$$

#### C. Level 2 model

$$Y_{ijklmn} = \mu + \tau_i + S_j + L(S)_{k(j)} + \varepsilon_{ijklmn}$$

#### D. Level 1 model

$$Y_{ijklmn} = \mu + \tau_i + S_j + \varepsilon_{ijklmn}$$

#### E. Model without random effects

$$Y_{ijklmn} = \mu + \tau_i + \varepsilon_{ijklmn}$$

where,

$Y_{ijklmn}$  is the value of the soil variable in patch  $i$ , site  $j$ , lot  $k$ , and transect  $l$ , and at sample point  $m$ , it is the  $n$ th observation

$\mu$  is the general mean

$\tau_i$  is the effect (considered to be fixed) of the  $i$ th vegetation patch of interest

$S_j$  is the random effect associated with site  $j$  (Level 4)

$$S_j \sim N(0, \sigma_{site}^2)$$

$L(S)_{k(j)}$  is the random effect associated with the lot  $k$  within site  $j$  (Level 3)

$$L(S)_{k(j)} \sim N(0, \sigma_{lot}^2)$$

$T(L)_{l(k)}$  is the random effect associated with transect  $l$  within lot  $k$  (Level 2)

$$T(L)_{l(k)} \sim N(0, \sigma_{transect}^2)$$

$P(T)_{m(l)}$  is the random effect associated with sample point  $m$  within transect  $l$  (Level 1)

$$P(T)_{m(l)} \sim N(0, \sigma_{point}^2)$$

$\varepsilon_{ijklmn}$  is a random error term with normal distribution and mean and variance equal to zero

$$\varepsilon_{ijklmn} \sim iid(0, \sigma^2)$$

The significance of the variance components at level  $k$  was evaluated by comparing the level  $k$  model to the level  $k + 1$  model, and by comparing the model without random effects to the model with only one variance component. Different hypothesis test approximations for variance components were used: Wald test, likelihood ratio test (LRT) with a Chi-square distribution and degrees of freedom equal to the difference in the number of parameters of the models that are compared (Littell et al. 2005), and a LRT but with an asymptotic Chi-square mixture of Chi-square distributions, under the null hypothesis (Molenberghs and Verbeke 2007).

The LRT statistic was calculated by extracting the log-likelihood of the reference model to the log-likelihood of the nested model:

$$-2 \log \left( \frac{L_{nested}}{L_{reference}} \right) = -2 \log(L_{nested}) - (-2 \log(L_{reference})) \quad (1)$$

In Eq. (1),  $L_{nested}$  refers to the value of the likelihood function evaluated at the estimates for restricted maximum likelihood (REML) (Searle et al. 1992) of the parameters in the reduced model (that is to say, that which has a variance component equal to zero), and  $L_{reference}$  is the value of the likelihood function in the reference model (that is to say, the model with the highest hierarchy level). In the classic hypothesis contrast and under the assumption of a normal distribution, the LRT statistic asymptotically follows a Chi-square distribution with degrees of freedom equal to the difference between the number of parameters of the reference model and the nested model to

which it is compared. In the context of the hypothesis regarding variance components (greater than or equal to zero), the Chi-square mixture of Chi-square distributions was used to compare consecutive models from levels 1 to  $k$  with two or more degrees of freedom, taking into account that the models differ in one variance component and in the covariance(s) of that variance component with the other variance components associated with the random effects of the reference model. In comparing models A and B, the significance of the variance associated with Level 4 and the covariances of that level with levels 1, 2 and 3 are evaluated. When a hypothesis is rejected, one concludes that the random effect associated with the sampling done in that stage should be kept in the model.

When comparing models E and D, we used a Chi-square distribution with 0.5 degrees of freedom, because when the comparison involves the model with a random component vs. the model without random effects, it is assumed that the null hypothesis of the LRT is a Chi-square mixture of Chi-square distributions with 0 and 1 degrees of freedom, each with equal weight. In this case, the Chi-square distribution with 0 degrees of freedom is not really a distribution, but a probability mass of 1 for value 0, and therefore the statistic is evaluated with respect to a Chi-square distribution with 0.5 degrees of freedom (Verbeke and Molenberghs 2000). All the tests were done with Proc GLIMMIX, SAS version 9.1 (SAS Institute 2008), with the codes found in the Appendix.

In all the cases in which the LRT statistic was used, and in which its value is sufficiently large (larger than the reference value of the distribution of the statistic), there is evidence against the reduced model and in favor of the reference model. But if the likelihood values are very similar in both models, the LRT statistic turns out to be very small, and therefore there is no evidence that favors the reduced model (null hypothesis), because when faced with the lack of statistically significant differences, it is traditionally recommended to select the model with the fewest parameters.

In hierarchical models, it can be observed that the mean and the variance structure are simplified; the  $-2$  log-likelihood REML increases. Nevertheless, in order to select the best model, the model with the smallest  $-2$  log-likelihood REML is preferred since this does not imply an excessive number of unnecessary parameters associated to each stage of sampling.

In this paper, we analyze the significance of the estimated variance components for each depth level, beginning with the Level 1 model and up until the Level 4 model. The model that includes the lowest significant level was selected. Once the best model was identified, we proceeded to study the differences between the means for the fixed effects factor, for each one of the three soil variables in the study.

### 3 Results and discussion

#### 3.1 Model selection

In Tables 1, 2, 3 goodness-of-fit criteria for the each depth level are shown, as are the results of the three statistical tests used to determine the significance of the variance components associated with the different sampling stages. For the case of Texture

**Table 1** Adjustment criteria for the comparison of models for the variable Texture

Model	Depth	Adjustment criteria			Comparison between models		
		-2 log L	Res	Ref	<i>p</i> value Wald <sup>a</sup>	<i>p</i> value LRT <sup>b</sup>	<i>p</i> value LRT <sup>c</sup>
<b>A. Level 4 model</b>		963.207	10.381		<0.001		
B. Level 3 model		985.170	19.876	A	0.141	<0.001	<0.001
C. Level 2 model	0–6 cm	991.550	21.456	B	0.058	0.067	0.094
D. Level 1 model		1,036.951	30.506	C	0.057	<0.001	<0.001
E. Fixed model		1,375.600	33.578	D		<0.001	<0.001
A. Level 4 model		953.543	12.084		0.008		
<b>B. Level 3 model</b>		960.477	16.801	A	0.134	0.104	0.136
C. Level 2 model	6–12 cm	968.015	18.271	B	0.117	0.041	0.058
D. Level 1 model		1,016.408	26.556	C	0.057	<0.001	<0.001
E. Fixed model		1,398.900	38.836	D		<0.001	<0.001

The selected model at a given depth is shown in bold

-2log L: is -2 times the log of the restricted likelihood; Res: is the residual variance; Ref: indicates the reference model used for comparison in each row

<sup>a</sup> Significance of the Wald statistic

<sup>b</sup> Significance of the LRT statistic with an asymptotic Chi-square mixture of  $\chi_k^2$  and  $\chi_{k+1}^2$  distributions, each with an equal weight of 0.5

<sup>c</sup> Significance of the LRT statistic with asymptotic distribution  $\chi_{k+1}^2$

**Table 2** Adjustment criteria for the comparison of models for the variable BD

Model	Depth	Adjustment criteria			Comparison between models		
		-2log L	Res	Ref	<i>p</i> value Wald <sup>a</sup>	<i>p</i> value LRT <sup>b</sup>	<i>p</i> value LRT <sup>c</sup>
A. Level 4 model		-262.077	0.008		0.063		
<b>B. Level 3 model</b>		-259.478	0.009	A	0.232	0.542	0.6268
C. Level 2 model	0–6 cm	-258.229	0.009	B	0.113	<0.001	<0.001
D. Level 1 model		-245.650	0.011	C	0.073	0.7518	0.002
E. Fixed model		-207.500	0.015	D		<0.001	<0.001
A. Level 4 model		-299.705	0.007		0.313		
B. Level 3 model		-299.461	0.008	A	0.417	0.987	0.995
C. Level 2 model	6–12 cm	-299.406	0.008	B	0.297	0.972	0.992
<b>D. Level 1 model</b>		-298.813	0.008	C	0.063	0.590	0.741
E. Fixed model		-208.600	0.015	D		<0.001	<0.001

The selected model at a given depth is shown in bold

-2log L: is -2 times the log of the restricted likelihood; Res: is the residual variance; Ref: indicates the reference model used for comparison in each row

<sup>a</sup> Significance of the Wald statistic

<sup>b</sup> Significance of the LRT statistic with an asymptotic Chi-square mixture of  $\chi_k^2$  and  $\chi_{k+1}^2$  distributions, each with an equal weight of 0.5

<sup>c</sup> Significance of the LRT statistic with asymptotic distribution  $\chi_{k+1}^2$



**Table 3** Adjustment criteria for the comparison of models for the variable TOC

Model	Depth	Adjustment criteria			Comparison between models		
		–2log L	Res	Ref	<i>p</i> value Wald <sup>a</sup>	<i>p</i> value LRT <sup>b</sup>	<i>p</i> value LRT <sup>c</sup>
A. Level 4 model		1,030.198	31.636		–		
B. Level 3 model		1,030.198	31.636	A	0.489	1	1
C. Level 2 model	0–6 cm	1,030.199	31.651	B	–	1	1
<b>D. Level 1 model</b>		1,030.199	31.651	C	0.067	1	1
E. Fixed model		1,088.200	49.826	D		<0.001	<0.001
A. Level 4 model		739.341	4.806		–		
B. Level 3 model		739.341	4.806	A	–	1	1
<b>C. Level 2 model</b>	6–12 cm	739.341	4.806	B	0.119	1	1
D. Level 1 model		756.177	5.606	C	<0.001	<0.001	<0.001
E. Fixed model		877.400	13.341	D		<0.001	<0.001

The selected model at a given depth is shown in bold

–2log L: is –2 times the log of the restricted likelihood; Res: is the residual variance; Ref: indicates the reference model used for comparison in each row

<sup>a</sup> Significance of the Wald statistic

<sup>b</sup> Significance of the LRT statistic with an asymptotic Chi-square mixture of  $\chi_k^2$  and  $\chi_{k+1}^2$  distributions, each with an equal weight of 0.5

<sup>c</sup> Significance of the LRT statistic with asymptotic distribution  $\chi_{k+1}^2$

(Table 1), at depth level 1 (0–0.06 m) the level 4 model was chosen, while at depth level 2 (0.06–0.12 m) the level 3 model was most appropriate. The selection criteria were applied as follows: the significance of the Wald test corresponds to the random factor of the highest level included in the model and when one evaluates whether to keep the same factor in the model, using LRT, the value of the statistic is the difference between the –2 log-likelihood REML of both models ( $985.170 - 963.207 = 21.963$ ). Wald *p* value was <0.001 and therefore significant for the level 4 model. Both LRT tests were significant ( $p < 0.001$ ), thus the model with the highest level of random effects was chosen. A different situation was found when analyzing the models for texture at depth level 2: Using el LRT statistic with asymptotic distribution  $\chi_{k+1}^2$  Level 2 would be selected. On the other hand, using Wald test the highest level would have been included in the analysis. LRT statistic with an asymptotic Chi-square mixture of  $\chi_k^2$  and  $\chi_{k+1}^2$  distributions, each with an equal weight of 0.5 indicated a significant difference ( $p = 0.041$ ) between the –2 log-likelihood REML of the models with 3 and 2 random effects therefore the level 3 model was chosen for further analysis of the data at this depth.

For the variable BD, in both depth levels 1 and 2 and as a function of the Wald statistic, the chosen model is the one with no random factors (Table 2). In contrast, when analysis is based on the LRT statistic, the Level 3 model is chosen for depth level 1, and the Level 1 model is chosen for depth level 2 (Table 2).

For the TOC variable, for depth level 1, when the Wald statistic approximation is used, the chosen model is the one with no random effects. However, when using the

tests based on LRT, one chooses the Level 1 model with depth level 1. For depth level 2, when using the Wald test, the selected model is the Level 1 model, and when using the LRT test, the selected model is the Level 2 model (Table 3).

For all the variables, except for Texture in depth level 2, models with very simplified variance structures were selected when the Wald test was used. The results coincide with the critiques of other authors (Pinheiro and Bates 1998; West et al. 2006) regarding the low sensibility of the Wald statistic, especially in cases such as this one, in which the number of levels for each random factor is low and the true parameter value is within the limit of the parametric space, under the null hypothesis. The identification of the best model to chose for adjusting the experimental data is crucial, since the over-parameterization of the covariance structure can lead to the inefficient estimation and poor evaluation of the standard error for the estimation of differences between means for an effects factor, the under-specification that ignores variance components due to the sampling strategy can lead to incorrect conclusions.

The LRT statistic with an asymptotic Chi-square mixture of Chi-square distributions with  $k$  and  $k + 1$  degrees of freedom, or with 0.5 degrees of freedom when a fixed and mixed model are compared, turned out to be the most potent strategy for identifying the variability due to the sampling strategy. Despite the fact that the  $p$  values associated with the LRT, when dealing with the Chi-square mixture of Chi-square distributions, were always less than the uncorrected LRT  $p$  values, both tests led to the selection of the same model. The fact that there was a reduction in the  $p$  values suggests that in other applications, disregarding this aspect of distributional correction could lead to an oversimplification of the covariance structure.

In the application illustrated in this paper, the models were selected in function of the LRT statistic with asymptotic Chi-square mixture of Chi-square distributions and are in bold in Tables 1, 2, 3. As can be observed, these models include a random factor of the lowest significant level and all the factors of a level higher than that one, regardless of their significance. These factors should remain in the model in order to maintain the correlated structure of the data introduced by the sampling strategy.

It is noteworthy that the selected model for the variance and co-variance structure was not the same at different soil depths. This shows that the variability occurs at different scales at the two depths even for one variable. For Texture and BD we found that at more depth the variation was detected at a higher hierarchical scale, while in surface samples variability was detected even in the finer scale. For TOC, on the contrary, we only detected variability at the highest scale. These findings elucidate the processes that cause variability of the observed soil properties: thus, TOC would be influenced by large scale processes that relate to changes in soil texture, geomorphology and vegetation patches, while BD and Texture at the soil surface would be subject to smaller scale phenomena such as in this particular case, redistribution of soil particles due to wind erosion and the trapping of specific size fractions in vegetation patches in the micro-relief (Pelletier et al. 2009a).

Once the variance structure was modeled for each variable, we studied the effect of the vegetation patches (fixed factor) contained in the mean structure of the same model. This unified approximation for the analyses is a substantial practical difference between the use of HLMM compared to classical geostatistical techniques which attend the multi-scale variability in a different analytical stage than that used for

studying average trends, and in this sense they are more efficient in the use of the available information.

For the variable Texture, differences were found at depth level 1 between soils covered by different vegetation patches ( $p < 0.001$ ), but these differences were not statistically significant at depth level 2 ( $p = 0.4562$ ) (Table 4). This suggests a differential effect of the type of soil cover on the texture of the first soil stratum with respect to the one with the greatest depth level. With regards to the variables BD and TOC, significant differences between vegetation patches were observed ( $p < 0.001$ ) for two depth levels (Table 4). Since the random effect model showed smaller standard errors, the amplitude of the confidence interval for the differences in means is always lower for the selected model than for the model without random effects (Table 4). This implies that the inference regarding the fixed effects is more precise when the random components associated with the sampling strategy are included. For the variable Texture, failing to model the variance structure results in finding no differences between the vegetation patches in depth level 1, while difference are indeed found when this structure is taken into account. In contrast, for the variable BD, marked changes regarding the length of the confidence interval are not observed in any of the depth levels, with or without modeling the covariance structure (Table 4). With respect to the variable TOC, in depth level 1, it does not have much influence on the inference regarding the fixed effects of modeling the covariance structure, but it does in depth level 2, because if the random effects induced by the sampling strategy are not included, the hypothesis regarding the difference in means between vegetation patches fails to be rejected, and the reverse occurs in the case that these are indeed included in the model.

These results coincide with the concepts of (Glasscock et al. 2005) who defined soil texture as the key determinant in defining plant community sub models for simulating vegetation dynamics in Texan rangelands. The textural control on herbaceous species distribution can be attributed to the soil's hydraulic dynamics that are determined by texture (Quiroga et al. 1999, 1998). The variation of soil texture at a very small scale in surface samples could be explained by the relevance of wind erosion in the region (Buschiazzo et al. 1999). Areas not covered by tree canopy are preferably browsed by cattle and frequently present higher proportion of bare soil than those under tree

**Table 4** Comparison of means and error standard of means comparison at depth level 1 and depth level 2

Depth	Texture <sup>1</sup> Clay + Silt cont. (%)		BD <sup>1</sup> (Mg m <sup>-3</sup> )		TOC <sup>1</sup> (g kg <sup>-1</sup> )	
	Fixed model	Selected model	Fixed model	Selected model	Fixed model	Selected model
0–6 cm						
Open	42.7 ± 2.7a	43.7 ± 0.5a	1.16 ± 0.02b	1.16 ± 0.01b	15.3 ± 1.1b	15.3 ± 0.9b
Shade	40.0 ± 2.7a	41.0 ± 0.5b	0.95 ± 0.02a	0.95 ± 0.01a	21.9 ± 1.1a	21.9 ± 0.9a
6–12 cm						
Open	42.9 ± 2.9a	43.9 ± 0.5a	1.25 ± 0.02b	1.25 ± 0.01b	11.2 ± 0.6a	11.2 ± 0.3b
Shade	42.4 ± 2.9a	43.4 ± 0.5a	1.18 ± 0.02a	1.18 ± 0.01a	12.2 ± 0.6a	12.2 ± 0.3a

<sup>1</sup> Different letters in the same column indicate statistically significant differences ( $p \leq 0.05$ )

canopy. Wind erosion will transport fine particles at great distances while sand-sized particles will be trapped by the near-by canopy shaded vegetation, thus producing an increase in coarser particles in those areas. This process obviously only affects the soil surface while at greater depth the variability of soil texture will be controlled at a larger scale, more related to differences in parent material and landscape relief. The differences in scale, detected through the statistical procedure for BD, with a smaller scale in surface samples, reflect the effect of litter deposition and tree canopy and, perhaps less compaction due to cattle grazing in the canopy-shaded areas, while at greater depth this parameter is not affected by small-scale differences due to vegetation patches, but more associated to landscape scale differences in texture and TOC contents. These findings are consistent with those of [Feral et al. \(2003\)](#) and [Wang et al. \(2007\)](#) who also found higher TOC contents under tree canopies in the Kalahari desert, with less differences detected between canopy and non-canopy areas than in our study.

## 4 Conclusions

The proposal to consider an application of HLMM in multiple stage sampling soil studies is presented as an analytical tool to improve treatment comparison in environmental and ecological studies. Variance components related at different scale of samplings are estimated to consider induced correlation in the data at the same level of the hierarchical study. To test the significance of variance components, adjustments of the classical LRT are recommended. The comparison of soil variable means between vegetation conditions gives a better discrimination of vegetation patch effects when random factors related to the sampling scheme were included in the linear model.

**Acknowledgments** We thank Dr. Raúl Machiavelli, professor at the University of Puerto Rico, for his comments on this approach. This work was supported by the Faculty of Agronomy, National University of La Pampa, Argentina and the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina.

## Appendix

### Appendix A: SAS code to fit HLMM by sample depth

```
/*----- Level 4 Model -----*/
proc glimmix data=datos plots=diffplot plots=studentpanel(type=blup);
by depth;
class site lot transect sample_point vegetation;
model Texture= vegetation /ddfm=kr;
random intercept/ subject= site;
random intercept/ subject= lot(site);
random intercept/ subject= transect(lot);
random intercept/ subject= sample_point(transect);
lsmeans vegetation / diff cl ;
run;
```

```

/*-----Level 3 Model -----*/
proc glimmix data=datos plots=diffplot plots=studentpanel(type=blup);
by depth;
class site lot transect sample_point vegetation;
model Texture= vegetation /ddfm=kr;
random intercept/ subject= site;
random intercept/ subject= lot(site);
random intercept/ subject= transect(lot);
lsmeans vegetation / diff cl ;
run;

/*-----Level 2 Model -----*/
proc glimmix data=datos plots=diffplot plots=studentpanel(type=blup);
by depth;
class site lot transect sample_point vegetation;
model Texture= vegetation /ddfm=kr;
random intercept/ subject= site;
random intercept/ subject= lot(site);
lsmeans vegetation / diff cl ;
run;

/*-----Level 1 Model -----*/
proc glimmix data=datos plots=diffplot plots=studentpanel(type=blup);
by depth;
class site lot transect sample_point vegetation;
model Texture= vegetation /ddfm=kr;
random intercept/ subject= site;
run;

/*-----Model without random effects -----*/
proc glimmix data=datos plots=diffplot plots=studentpanel(type=blup);
by depth;
class site lot transect sample_point vegetation;
model Texture= vegetation /ddfm=kr;
run;

```

## Appendix B: SAS codes for obtaining LRT with asymptotic distribution $\chi^2_{k+1}$

```

/*----- Comparison of Level 4 model and Level 3 model-----*/
data _null_;
lrtstat= 985.170-963.207;
pvalue=1-probchi(lrtstat,4);
format pvalue 10.8;
put lrtstat=pvalue;
run;

```

```

/*----- Comparison of Level 3 model and Level 2 model-----*/
data _null_;
lrtstat= 991.550-985.170;
pvalue=1-probchi(lrtstat,3);
format pvalue 10.8;
put lrtstat=pvalue=;
run;

/*----- Comparison of Level 2 model and Level 1 model-----*/
data _null_;
lrtstat= 1036.951-991.550;
pvalue=1-probchi(lrtstat,2);
format pvalue 10.8;
put lrtstat=pvalue=;
run;

/*----- Comparison of Level 1 model and the model with fixed effects-----*/
data _null_;
lrtstat= 1375.600-1036.951;
pvalue=1-probchi(lrtstat,1);
format pvalue 10.8;
put lrtstat=pvalue=;
run;

```

### Appendix C: SAS codes for obtaining LRT with asymptotic Chi-square mixture of $\chi_k^2$ and $\chi_{k+1}^2$

```

/*----- Comparison of Level 4 model and Level 3 model -----*/
data _null_;
lrtstat= 985.170-963.207;
pvalue=0.5*(1-probchi(lrtstat,4))+0.5*(1-probchi(lrtstat,3));
format pvalue 10.8;
put lrtstat=pvalue=;
run;

/*----- Comparison of Level 3 model and Level 2 model-----*/
data _null_;
lrtstat= 991.550-985.170;
pvalue=0.5*(1-probchi(lrtstat,3))+0.5*(1-probchi(lrtstat,2));
format pvalue 10.8;
put lrtstat=pvalue=;
run;

```

```

/*----- Comparison of Level 4 model and Level 3 model -----*/
data _null_;
lrtstat= 985.170-963.207;
pvalue=0.5*(1-probchi(lrtstat,4))+0.5*(1-probchi(lrtstat,3));
format pvalue 10.8;
put lrtstat=pvalue=;
run;

/*----- Comparison of Level 3 model and Level 2 model-----*/
data _null_;
lrtstat= 991.550-985.170;
pvalue=0.5*(1-probchi(lrtstat,3))+0.5*(1-probchi(lrtstat,2));
format pvalue 10.8;
put lrtstat=pvalue=;
run;

```

## References

- Balzarini M (2002) Applications of mixed models in plant breeding. In: Kang MS (ed) Quantitative genetics, genomics, and plant breeding. CABI Publishing, UK, pp 353–365
- Buschiazzo DE, Aimar S, Zobeck T (1999) Wind erosion in soils of the semiarid Argentinian Pampas. *Soil Sci* 164:133–138
- Cochran W (1980) Técnicas de muestreo. CECSA, México
- Cox DR, Hinkley DV (1990) Theoretical statistics. Chapman & Hall, London
- Feral CJW, Epstein HE, Otter L, Aranibar JN, Shugart HH, Macko SA, Ramontsho J (2003) Carbon and nitrogen in the soil-plant system along rainfall and land-use gradients in southern Africa. *J Arid Environ* 54:327–343
- Glasscock SN, Grant WE, Drawe DL (2005) Simulation of vegetation dynamics and management strategies on south Texas semi-arid rangeland. *J Environ Manag* 75:379–397
- Gee GW, Bauder JW (1986) In: Klute A (ed) Methods of soil analysis. Part I, chap 15. ASA Publications, Madison, WI, USA
- Heim A, Wehrli L, Eugster W, Schmidt MWI (2009) Effects of sampling design on the probability to detect soil carbon stock change at the Swiss CarbEurope site Lägeren. *Geoderma* 149:347–354
- Lardy LC, Brossard M, Assad MLL, Laurent JY (2002) Carbon and phosphorus stocks of clayey Ferralsols in Cerrado native and agroecosystems. *Brazil Agr Ecosyst Environ* 92(2–3):147–158
- Lark RM (2005) Exploring scale-dependent correlation of soil properties by nested sampling. *Eur J Soil Sci* 56(3):307–317
- Levin SA (1992) The problem of pattern and scale in ecology. *Ecology* 73(6):1943–1967
- Littell RC, Milliken GA, Stroup WW, Wolfinger RD, Schabenberger O (2005) SAS<sup>®</sup> system for mixed models, 2nd edn. SAS Institute Inc., Cary, NC
- Molenberghs G, Verbeke G (2007) Likelihood ratio, score, and Wald test in a constrained parameter space. *Am Stat* 61(1):22–27
- Pelletier B, Dutilleul P, Larocque G, Fyles JW (2009a) Coregionalization analysis with a drift for multi-scale assessment of spatial relationships between ecological variables 1. Estimation of drift and random components. *Environ Ecol Stat* 16:439–466
- Pelletier B, Dutilleul P, Larocque G, Fyles JW (2009b) Coregionalization analysis with a drift for multi-scale assessment of spatial relationships between ecological variables 2. Estimation of correlations and coefficients of determination. *Environ Ecol Stat* 16:467–494
- Pinheiro JC, Bates DM (1998) lme: linear mixed-effects methods and classes for S and S-plus, Version 3.0
- Post WM, Izaurralde RC, Mann LK, Bliss N (2001) Monitoring and verifying changes of organic carbon in soils. *Clim Chang* 51:73–99

- Quiroga A, Buschiazzo D, Peinemann N (1999) Soil compaction is related to management practices in semiarid Argentine pampas. *Soil Tillage Res* 52:21–28
- Quiroga A, Buschiazzo D, Peinemann N (1998) Management discriminant properties in semiarid soils. *Soil Sci* 163:591–597
- Raudenbush SW, Bryk AS (2002) Hierarchical linear models: applications and data analysis methods, 2nd edn. Sage Publications, Thousand Oaks, CA (Advanced quantitative techniques in the social sciences series no. 1)
- Ringrose S, Matheson W, Vanderpost C (1998) Analysis of soil organic carbon and vegetation cover trends along the Botswana Kalahari Transect. *J Arid Environ* 38(3):379–396
- Searle SR, Casella G, McCulloch CE (1992) Variance components. Wiley, New York
- SAS Institute Inc (2008) SAS/STAT software: version 9.1. SAS Institute Inc, Cary, NC
- Snijders T, Bosker R (2000) Multilevel analysis: an introduction to basic and advanced multilevel modeling. Sage, London
- Soon Y, Abboud S (1991) Comparison of some methods for soil organic carbon determination. *Commun Soil Sci Plant Anal* 22:943–954
- Verbeke G, Molenberghs G (2000) Linear mixed models for longitudinal data. Springer, New York
- Wang L, Okin GS, Caylor KK, Macko SA (2009) Spatial heterogeneity and sources of soil carbon in southern African savannas. *Geoderma* 149:402–408
- Wang L, D'Odorico P, Ringrose S, Coetzee SH, Macko SA (2007) Biogeochemistry of Kalahari sands. *J Arid Environ* 71(3):259–279
- West B, Welch KB, Galecki AT (2006) Linear mixed models: a practical guide using statistical software. Chapman & Hall/CRC, Boca Raton, FL
- Wiens JA (1989) Spatial scaling in ecology. *Funct Ecol* 3:385–397

## Author Biographies

**Adriana A. Gili** is a Ph.D. candidate in biostatistics at the graduate school of the Facultad de Ciencias Agrarias of the Universidad Nacional de Córdoba, and her basic formation was as an agronomist at the Universidad Nacional de La Pampa. Currently, she is Professor in the Department of Statistics at the Facultad de Agronomía, Universidad Nacional de La Pampa and her main interests are statistics, ecology, landscape ecology, precision agriculture and geostatistics.

**Elke J. Noellemeyer** is Professor of Soil Science at the Facultad de Agronomía, Universidad Nacional de La Pampa. She was trained as a soil scientist at the College of Agriculture of the University of Saskatchewan, Canada, and has been working on soil management, with a special interest in organic matter and soil physical properties. She has been involved in several international research networks and actively participates in the science-policy interface regarding environmental and agronomic topics.

**Mónica Balzarini** main background is in Statistics, and her research interests are in applications of statistics in agricultural and environmental sciences. She is an agronomist from the Universidad Nacional de Córdoba and a Ph.D. from Louisiana State University-USA. She is Professor in the Department of Biometrics, Facultad de Ciencias Agrarias, Universidad Nacional de Córdoba and an independent researcher in the Instituto Nacional de Ciencias y Tecnología of Argentina.