An Entomologist Guide to Demystify Pseudoreplication: Data Analysis of Field Studies With Design Constraints

LUIS FERNANDO CHAVES[1,2]

J. Med. Entomol. 47(3): 291–298 (2010); DOI: 10.1603/ME09250

ABSTRACT Lack of independence, or pseudoreplication, in samples from ecological studies of insects reflects the complexity of working with living organisms: the finite and limited input of individuals, their relatedness (ecological and/or genetic), and the need to group organisms into functional experimental units to estimate population parameters (e.g., cohort replicates). Several decades ago, when the issue of pseudoreplication was first recognized, it was highlighted that mainstream statistical tools were unable to account for the lack of independence. For example, the variability as a result of differences across individuals would be confounded with that of the experimental units where they were observed (e.g., pans for mosquito larvae), whereas both sources of variability now can be separated using modern statistical techniques, such as the linear mixed effects model, that explicitly consider the different scales of variability in a dataset (e.g., mosquitoes and pans). However, the perception of pseudoreplication as a problem without solution remains. This study presents concepts to critically appraise pseudoreplication and the linear mixed effects model as a statistical solution for analyzing data with pseudoreplication, by separating the different sources of variability and thereby generating correct inferences from data gathered in studies with constraints in randomization.

KEY WORDS linear mixed effects model, *Culex quinquefasciatus*, *Anopheles nuneztovari*, bootstrap, data analysis

Pseudoreplication is probably one of the most widely cited and misunderstood concepts in the statistical analysis of ecological studies on insects and other organisms. Pseudoreplication is defined as "the use of inferential statistics to test for treatment effects with data from experiments where either treatments are not replicated . . . or replicates are not statistically different . . ." (Hurlbert 1984). This concept has been very influential and pervasive, to the extent that pseudoreplication is widely cited as a major flaw of most field studies (Heffner et al. 1996). Hurlbert's major claim was correct, and he basically showed that mainstream statistical tools at that time (e.g., analysis of variance) were not suitable for the analysis of most experimental designs. However, the uncritical appraisal of his study has been a major barrier for the publication of results and, therefore, the advancement of ecology (Oksanen 2001). Hurlbert's study did not prevent the unsuitable analysis of valuable datasets or the proliferation of unsound experimental practices, such as the movement of sampling units to control for spatial/temporal variability (Alto and Juliano 2001a, 2001b; Reiskind and Wilson 2004). As thoughtfully presented by Oksanen (2001), the goal of ecological studies is not the application of statistical analysis to ecological data per se, but rather its application to the understanding of ongoing ecological phenomena from variation of individual phenotypic traits to the assemblages of organisms in populations, communities, and ecosystems. Unlike physics or chemistry, in which the supply of individual objects of study is practically unlimited, the objects of study for an entomologist (or more generally a naturalist) are finite and constrained. Thus, limitations in randomization will likely arise, and the science of statistics has developed new solutions to correctly analyze the lack of independence in field data since Hurlbert's study (Millar and Anderson 2004). The current forum article presents the following: 1) key concepts of experimental design to critically appraise and demystify the concept of pseudoreplication; 2) linear mixed effects models (LMEMs) as powerful tools to analyze data originated from constrained designs or to produce more general inferences from classical randomized designs (e.g., blocks); and 3) how these tools can be used to further gain insights from the data that can strengthen our understanding of insect ecology. In developing point 2, equations and a guide to interpret them as models used to analyze common entomological data are presented, as well as the implementation of this type of analysis in the open source software R.

[1] Corresponding author: Department of Environmental Studies, Emory University, 400 Dowman Drive, Suite E510, Atlanta GA 30322 (e-mail: lfchave@emory.edu).

[2] Laboratorio de Biología Teórica, Instituto de Zoología y Ecología Tropical, Facultad de Ciencias, Universidad Central de Venezuela, Caracas, Venezuela.

Field Studies, Experiments, and Statistical Data Analysis

Experiments are one of the major tools for hypothesis testing (Fisher 1935). In general, the idea is to subject individual units of observation to varying degrees of independent and/or controllable factor(s), and to determine how the levels of variation explain a given pattern (Box 1980). Field studies focus on the impacts of natural (or controlled) variation in environmental factors on individual units of observation. Hypothesis testing has been central to the development of modern science, to the point that hypothesis-driven experiments or field studies are one of the most prominent requirements for project support by funding agencies. One of the major reasons for the widespread appeal of hypothesis-driven experiments and field studies has been their close association with tools for data analysis to determine the impact of different independent variables. The best example of a statistical tool guiding experimental design is the use of the linear model (LM). This model assumes that variability across a set of individual units of observation (y_i) is explained by a series of n independent variables (x_1, $x_2, \ldots, x_n$) and by a unique source of unexplained variability, normally referred to as error (ε). These models are linear, because the parameters enter linearly into the equation that relates the independent variables to the outcome (Faraway 2006, Chaves and Pascual 2007). A major constraint of these models is that they assume total independence among the subjects of study, i.e., individual observation units are unrelated at least within strata (i.e., after accounting for the explanatory variables), which is the formal definition of "replication." When there is a lack of independence across objects of study (i.e., pseudoreplication), the use of LMs with a unique source of variability is inappropriate, because the variability is modeled incorrectly and can lead to spurious inferences. For example, if mosquitoes are reared in pans (or kissing bugs in jars) to measure body size of emerging adults from different experimental conditions, the lack of independence that arises from the aggregation into pans (or jars) will inflate the error value (a.k.a. residual variance) of the LM, in some cases leading to incorrect inferences when the LM is compared with a model that explicitly models the lack of independence because of the aggregation into a functional experimental unit (i.e., the pan or jar). More than 20 yr ago, because of the limited statistical toolbox in ecology, this issue was a major problem for the correct analysis of datasets from studies with design constraints (Hurlbert 1984). However, strategies to handle the problem of pseudoreplication were around at the time. For example, in evolutionary ecology, individuals have different degrees of common descent, and this variability by itself is often a subject of study. In the 1980s, it was common to use nested half-sibling designs to estimate the variance of families and individuals belonging to those families (Conner and Hartl 2004). Also, the use of defined designs such as Greco-Roman squares, Roman squares, and fractional factorials was well established in the field of engineering and process control (Montgomery 2005). Some of these balanced designs were even used in studies of medically important insects (Carpenter 1982, Chesson 1984). In fact, sophistication in randomization, when possible, can be very useful to evaluate the impact of strategies to control human-vector contact (Kirby et al. 2008). For example, randomized control trials have been used to demonstrate the importance of mosquito screening in reducing the risk of malaria transmission (Kirby et al. 2009). However, one of the major limitations of designs that handle pseudoreplication is the need for balanced designs, i.e., an equal number of replicates per treatment. The inability to analyze unbalanced designs with unequal number of replicates per treatment has been overcome with the development of maximum likelihood methods, especially the restricted maximum likelihood method and its application to estimate LMEM (Pinheiro and Bates 2000). LMEM has the same fundamental assumptions of the LM; it tries to explain the sources of variability across a set of individual units of observation (y_i) as function of a series of n independent variables ($x_1, x_2, \ldots, x_n$), referred to as "fixed factors," but it can incorporate additional sources of variability (the random factors), besides the error (ε). The random factors can accommodate the lack of independence among the individual units of observation as a result of spatial, temporal, genetic, or any exogenous environmental factor that is not fully randomized. For example, the variance of functional experimental units such as pans for mosquitoes or jars for kissing bugs can be explicitly modeled, thus allowing the proper estimation of the error variance, and thus limiting the chances of committing a type II error, i.e., rejecting the null hypothesis when true. Therefore, LMEMs allow the statistical analysis of pseudoreplicated data. The next section will provide a series of examples illustrating the use of LMEMs and how they compare with similar LMs. The data used in the examples and code to perform the analyses using the open source statistical software R are included as supplementary online material (http://www.envs.emory.edu/research/Chaves_SOM_Pseudoreplication.html).

LMs Versus LMEMs

Factorial Designs. To illustrate the most basic differences between LMs and LMEMs, I reference a field experiment designed to study oviposition by *Culex quinquefasciatus* Say in Atlanta, GA (Chaves et al. 2009). *Cx. quinquefasciatus* larvae are normally absent from lotic systems, such as rivers and creeks. However, several cities have relic sewage treatment systems where runoff water and sewage are combined in the same system, and after large rainfall events the combined sewage effluent can overflow into urban water bodies (Chaves et al. 2009). In this experiment, the effects of combined sewage overflow water and nutrient addition on oviposition site selection by this mosquito species were studied using 10 experimental pools (water containers) at four sites in a forest patch.

Table 1. Analysis of variance for the effects of water quality and nutrient addition on the ln of total number of egg rafts + 1 oviposited over 5 d by *Culex quinquefasciatus* in Atlanta, GA (Chaves et al. 2009)

Factor	df	Sum square	Mean square	*F* Value	Pr(>*F*)
Water (β_1)	1	0.653	0.653	10.1401	0.01111*
Nutrient (β_2)	1	50.634	50.634	786.0312	4.54E-10*
Water × nutrient (β_3)	1	0.013	0.013	0.2055	0.66105
Block (β_4)	3	0.467	0.156	2.4186	0.13339
Error (ϵ)	9	0.58	0.064		

This design was balanced. Original data and R code for analysis are in the supplementary online material (http://www.envs.emory.edu/research/Chaves_SOM_Pseudoreplication.html).

*Statistically significant ($P < 0.05$).

For this example, data will be used from the experiment when egg rafts were removed daily. Data use has been restricted to a randomly extracted subsample from the original data (only four pools per site) to have a dataset similar to that of a balanced design. In this experiment, oviposition (y) was measured by counting the total number of egg rafts oviposited over 5 d. The experiment has three independent variables (i.e., $n = 3$): 1) x_1 = water quality (with two levels: combined sewage overflow water and tap water as control); 2) x_2 = nutrient addition (added or absent as control); and 3) x_3 = sites (four in total). Only x_1 and x_2 are factors (each with two levels), because x_3 is an independent variable considered to test the block effects of forest site on oviposition. Because all treatments were present at each site, this is a randomized block 2 × 2 factorial design, which is randomized because both factors were present in all four sites (blocks in the model), and 2 × 2 factorial because each factor has two levels. The goal of a factorial experiment is to test whether the factors interact, which can be expressed by the following LM:

$$y_{il} = \mu + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \beta_4 x_{3l} + \varepsilon_{il} \quad [1]$$

where μ is the average value of the observations; β_1, β_2, and β_4 quantify the impact of each independent variable; β_3 the interaction of water quality and nutrient addition; and ε is the error, which is assumed to be normally distributed. The subscript l denotes block (site within the forest patch), and i is for individual pools within a block (containers in a forest site). Therefore, y_{il} is the total number of rafts from a given pool and block (i.e., container in a site). Table 1 shows the results of the analysis of variance for the data using the model presented in (1). The natural logarithm transformation of ln ($y + 1$) is done to normalize the data and fulfill model assumptions. Table 1 shows that neither block nor the interaction between water quality and nutrient addition was significant ($P > 0.05$).

The influence of water quality and nutrients on *Cx. quinquefasciatus* oviposition can be reanalyzed using a LMEM. By contrast with the LM analysis, in which inferences are done over blocks, the LMEM assumes that blocks are random samples from a larger population and models block variability as a random factor (Fig. 1). The equivalent to equation 1 for a LMEM is:

$$y_{il} = \mu + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \gamma_l + \varepsilon_{il} \quad [2]$$

where μ, the βs, and y and ε have the same interpretation as in equation 1, and γ quantifies the variability across the blocks, which is assumed to be normally distributed. The significance of factors can be tested using F tests, which work well when designs are balanced (equal number of samples per treatment and block) and parameters can be estimated using maximum likelihood. More generally, significance may be tested using parametric bootstraps, for balanced or unbalanced designs where parameters are estimated by restricted maximum likelihood (Pinheiro and Bates 2000, Faraway 2006). Table 2 shows the results of an analysis of deviance, with parameters of equation 2 estimated using restricted maximum likelihood and inference based on 1000 replications of parametric bootstrap (an analysis in which datasets are simulated and the results of likelihood ratio tests for studied factors are compared with those of the true data to compute the significance of factors). In this example, inference about the impact of water quality and nutrient addition is qualitatively similar using LM or LMEM. However, LMEM provides additional insight; it indicates oviposition is finely grained. The variance at the individual container level is larger than at the block level (Table 2: $\varepsilon = 0.059$ and $\gamma = 0.024$), a pattern also observed in the original study for the full dataset (Chaves et al. 2009). The data can be observed in Fig. 2.

Constrained Designs. One of the major limitations of the LM is that it is not suited to analyze datasets with constraints in randomization. For example, all replicates from a treatment should be present in all blocks. This is a frequent limitation in field studies in which features of a given landscape cannot be altered, an underlying motivation behind split-plot designs, in which some treatments do not vary across blocks. Split-plots are widely used in agriculture (Faraway 2006) and economic entomology (Blumberg et al. 1997, Haile et al. 2000, Oyediran et al. 2007). However, constraints in randomization can also arise as a product of other trade-offs in experimentation or by the nature of the questions asked. For example, the original design of Chaves et al. (2009) was unbalanced in the sense that each block had an unequal number of replicates for each one of the treatments. However, for each block the amount of total nutrients and water quality was constant; one of the questions was to determine the grain of mosquito perception for oviposition choices. The largest variance was among the individual oviposition containers, indicating a finely grained perception, in contrast to a scenario of coarsely grained mosquito perception, in which the largest variance would be expected for the blocks or sites. To illustrate the analysis of constrained designs, the original data of Chaves et al. (2009) are sampled in such a way that all experimental pools with nutrients added and combined sewage overflow water

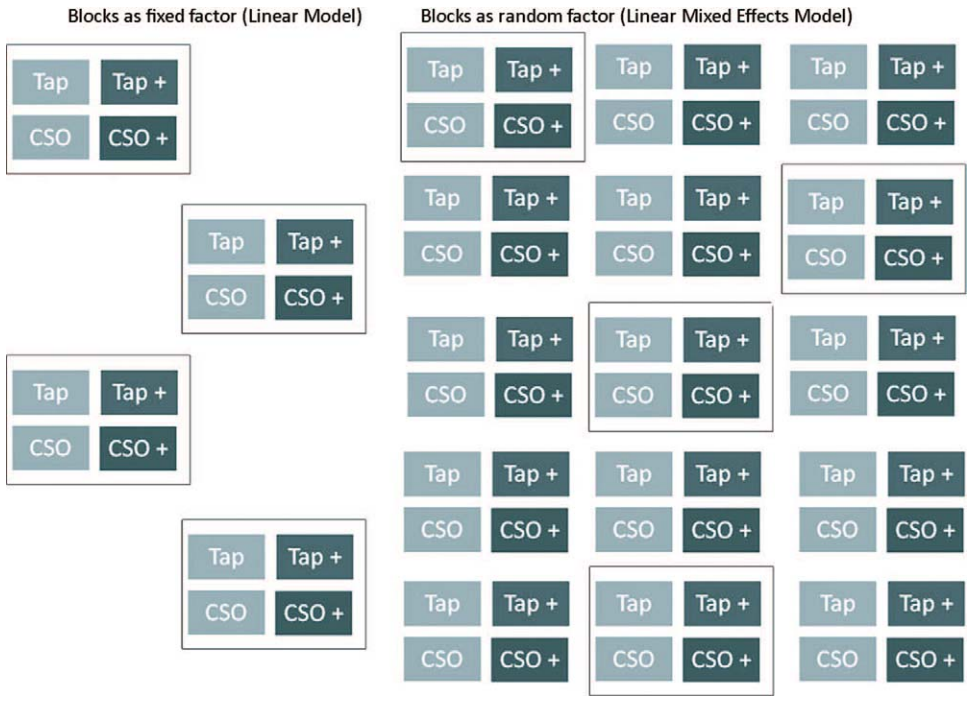

Fig. 1. Blocks: fixed or random? The left panel shows the case for blocks as a fixed factor in LM, in which the assumption is that inferences are exclusive for the observed blocks (within squares). The right panel shows the case for blocks as a random factor in LMEMs, in which the observed blocks (within squares) come from a larger population (i.e., blocks inside and outside squares). (Online figure in color.)

came from the same block (Fig. 2), and therefore the design becomes unbalanced (Fig. 3). Thus, the LM from equation 1 cannot be employed to analyze the data, but the LMEM from equation 2 is suitable for such an analysis. Results are presented in Table 3 and are similar to those presented in Table 2, showing a decreased variability in the blocks and a larger error. In summary, LMEMs can uncover the same variance pattern in data under pseudoreplication, a major advantage over LMs.

Table 2. Analysis of deviance for the effects of water quality and nutrient addition on the ln of total number of egg rafts + 1 oviposited over 5 d by *Culex quinquefasciatus* in Atlanta, GA (Chaves et al. 2009)

Fixed	df	Log likelihood	LRT	$P^{\dagger}$
Water (β_1)	1	−7.241	5.874	0.009*
Nutrient (β_2)	1	−31.089	53.569	0.000*
Water × nutrient (β_3)	1	−4.304	8.612	0.664
Random		**Mean square (variance)**		
Blocks (γ)		0.024		
Error (ϵ)		0.059		

This design was balanced. Original data and R code for analysis are in the supplementary online material (http://www.envs.emory.edu/research/Chaves_SOM_Pseudoreplication.html). LRT, likelihood ratio test.

† Obtained with a parametric bootstrap.

*Statistically significant ($P < 0.05$).

Spatial Variability

Organisms can be clustered in space, for example, the larvae of mosquitoes can be associated with only certain habitats where eggs are oviposited, thus making their abundance autocorrelated in space (Pitcairn et al. 1994). Several statistical tools can accommodate the lack of spatial independence in data from field studies (Fortin and Dale 2005), and they have been widely used with insects of medical importance (Koenraadt et al. 2007, 2008; Vazquez-Prokopec et al. 2008). However, their description is outside the scope of this article. LMEM can also be used to consider spatial variability. LMEM are especially suitable for cases when spatial scales are nested. For example, mosquito larval samples coming from containers in several houses that belong to the same neighborhood are hierarchically nested. Several studies have used this approach in recent studies on medically important insects (Harrington et al. 2008, Chaves et al. 2009, Gurtler et al. 2009). The LMEM fitting procedure is similar to the one used next to consider the lack of temporal independence in longitudinal studies.

Longitudinal Studies

The fact that observations are repeated through time in the same place (or from the same organisms) can lead to data that are not independent and are autocorrelated in time. One approach to this problem

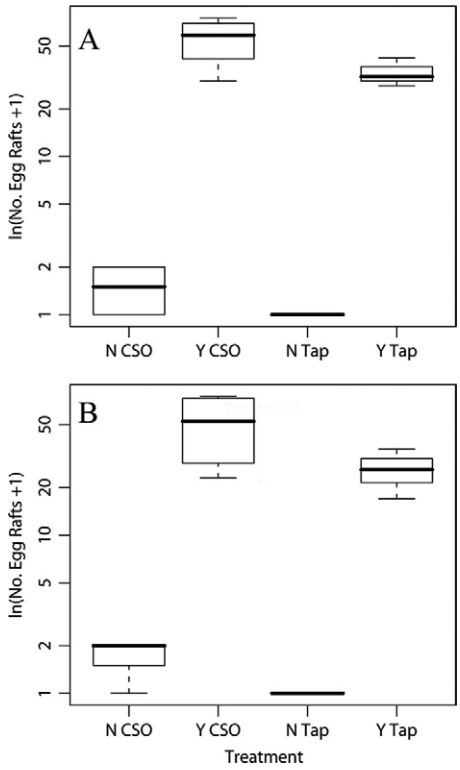

Fig. 2. Boxplots (median and quartiles) for the natural logarithm number of *Cx. quinquefasciatus* egg rafts + 1: (A) in the balanced (Tables 1 and 2) and (B) unbalanced (Table 3) block designs to study the effects of water quality and nutrient enrichment on oviposition. Tap indicates tap water and CSO indicates combined sewage overflow water. Y stands for nutrient addition, and N for no additional nutrients. Data extracted from Chaves et al. (2009). Original data are available in the supplementary online material (http://www.envs.emory.edu/research/Chaves_SOM_Pseudoreplication.html).

is to use repeated measurements analysis (Faraway 2006) and time series analysis techniques (Shumway and Stoffer 2000, Chaves and Pascual 2007). Time series techniques have been used for longitudinal studies of some vectors (Hayes and Downs 1980, Strickman 1988, Feliciangeli and Rabinovich 1998, Scott et al. 2000, Salomon et al. 2004). Alternatively, LMEM can model the lack of temporal independence as a random factor, which is one of the many methods

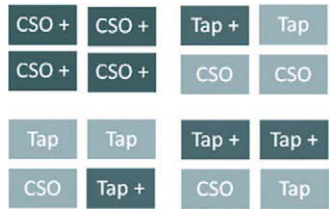

Fig. 3. Unbalanced design. Note that one block contains all samples with added nutrients and combined sewage overflow water, and other treatments have unequal number across other blocks. (Online figure in color.)

Table 3. Analysis of deviance for the effects of water quality and nutrient addition on the ln of total number of egg rafts + 1 oviposited over 5 d by *Culex quinquefasciatus* in Atlanta, GA (Chaves et al. 2009)

Fixed	df	Log likelihood	LRT	$P^{\dagger}$
Water (β_1)	1	−10.45	5.477	0.012*
Nutrient (β_2)	1	−29.03	42.638	0.000*
Water × nutrient (β_3)	1	−7.711	0.0409	0.801
Random	**Mean square (variance)**			
Blocks (γ)	5.0405e-14			
Error (ϵ)	0.125			

This design was unbalanced (because of the sampling from the full dataset). Original data and R code for analysis are in the supplementary online material (http://www.envs.emory.edu/research/Chaves_SOM_Pseudoreplication.html). LRT, likelihood ratio test.

† Obtained with a parametric bootstrap.

*Statistically significant ($P < 0.05$).

for repeated measurements analysis (Faraway 2006). Modeling the lack of temporal independence will be illustrated by examining data from a study on the biting and resting behavior of anophelines using experimental huts in three villages of western Venezuela (Rubio-Palis and Curtis 1992). Mosquitoes were collected during two nights per month and by catching the landing mosquitoes on the legs of two catchers between 1900 and 0700 hours, inside and outside experimental huts. Although several species were found, only data for *Anopheles nuneztovari* Gabaldón from Guaquitas collected between August 1988 and October 1989 will be analyzed in this study (Fig. 4). In this case, the response or dependent variable (y) is the total number of landings for all huts, as presented in the original study (Rubio-Palis and Curtis 1992). The fixed factors are as follows: 1) x_1, the site with two levels, inside and outside the hut; 2) x_2, the landing time with 12 levels corresponding to the hours between 1900 and 0700 hours; and 3) x_3, the rainfall season with two levels: dry (December-May) and wet (June–November). The random factors consider the different scales of temporal variability: 1) γ_l, the year l; 2) δ_{kl}, the month k within a given year l; 3) τ_{jkl}, the sampling day j within a given month k and year l; and 4) ε_{ijkl}, which is the error i (error for an observation belonging to day j within a given month k and year l). All random factors are assumed to be independent and normally distributed. The model equation is as follows:

$$y_{ijkl} = \mu + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_2 x_3 + \gamma_l + \delta_{kl} + \tau_{jkl} + \varepsilon_{ijkl} \quad [3]$$

In this model, μ represents the mean value of all observations; β_1, β_2, and β_3 quantify the impact of each independent variable on the number of landings; and β_4 the interaction of season and landing time. Note that choice of factors is dictated by the study objective: quantification of the seasonal nocturnal biting pattern outside and inside the experimental huts. Such quantification requires landing time, site, and season to be treated as fixed independent variables. The other temporal variables need to be random factors accounting

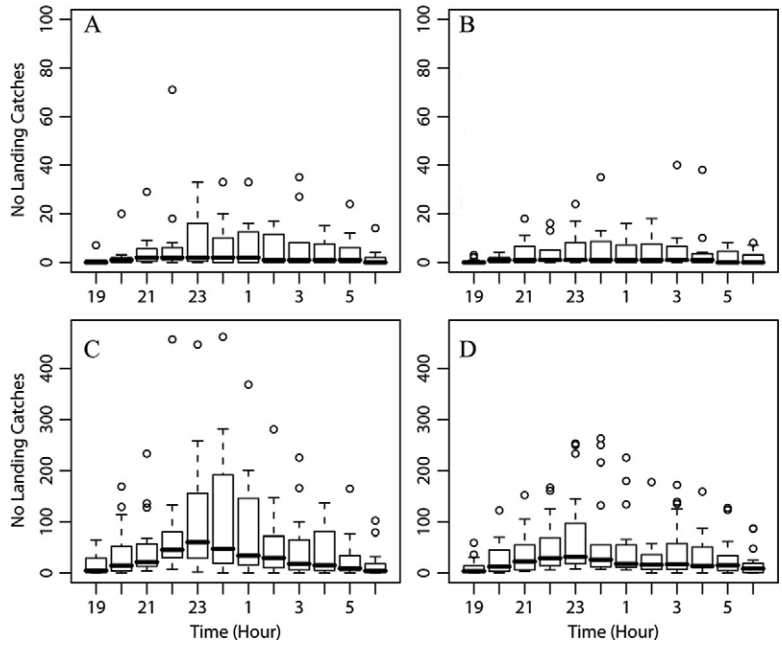

Fig. 4. Boxplots (median and quartiles) for the hourly number of *Anopheles nuneztovari* landings: (A) outside the house, dry season; (B) inside the house, dry season; (C) outside the house, wet season; (D) inside the house, wet season. Data extracted from Rubio-Palis and Curtis (1992). Original data are available in the supplementary online material (http://www.envs.emory.edu/research/Chaves_SOM_Pseudoreplication.html).

for the lack of independence that arises from the repeated measurements through time. There was no variability because of the year of the observation ($\hat{\gamma} = 0$; see supplementary online material http://www.envs.emory.edu/research/Chaves_SOM_Pseudoreplication.html), and a simpler model, without a parameter for the annual variability, was fit, as follows:

$$y_{ijk} = \mu + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_2 x_3 + \delta_k + \tau_{jk} + \varepsilon_{ijk} \quad [4]$$

Results for this model are presented in Table 4. The interaction between season and time and the main effects of site and season are statistically significant ($P < 0.05$). Daily observations for each month were more homogeneous, having a lower variance than those observations across months. For comparison purposes, a LM was also fit, as follows:

$$y_i = \mu + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_2 x_3 + \varepsilon_i \ldots \quad [5]$$

Table 5 shows the results for the analysis with equation 5. All factors are significant with this model ($P < 0.05$). However, when compared with the model with random effects, the main effect for landing time is not significant when the lack of independence in the data is properly modeled with a LMEM (Table 4). These analyses illustrate one of the problems of incorrectly

Table 4. Analysis of deviance for the effects of site, landing time, and season on *Anopheles nuneztovari* abundance in Guaquitas, Venezuela (Rubio-Palis and Curtis 1992)

Factor	df	Log likelihood	LRT	$P^{\dagger}$
Site (β_1)	1	−3550	15.1	0.000*
Landing time (β_2)	11	−3619	154.2	0.073
Season (β_3)	1	−3548	11.2	0.000*
Landing time × season (β_4)	11	−3542	124.4	0.000*
Random	**Mean square (variance)**			
Month (δ)	751.46			
Day (τ)	182.68			
Error (ϵ)	1669.42			

Original data and R code for analysis are in the supplementary online material (http://www.envs.emory.edu/research/Chaves_SOM_Pseudoreplication.html). LRT, likelihood ratio test.

† Obtained with a parametric bootstrap.

*Statistically significant ($P < 0.05$).

Table 5. Analysis of variance for the effects of site, landing time, and season on *Anopheles nuneztovari* abundance in Guaquitas, Venezuela (Rubio-Palis and Curtis 1992)

Factor	df	Sum square	Mean square	*F* Value	Pr(>*F*)
Site (β_1)	1	17,914	17,914	6.727	0.009704*
Landing time (β_2)	11	162,525	14,775	5.5483	1.55E-08*
Season (β_3)	1	286,446	286,446	107.5666	<2.2e-16*
Landing time × season (β_4)	11	77,851	7077	2.6577	0.002432*
Error (ϵ)	671	178,6849	2663		

Original data and R code for analysis are in the supplementary online material (http://www.envs.emory.edu/research/Chaves_SOM_Pseudoreplication.html).

*Statistically significant ($P < 0.05$).

modeling the lack of independence across observations: the LM rejects a null hypothesis that is true (type II error) by saying that landing time by itself is significant (Table 5), when in reality it is only significant when considered in conjunction with the season (Table 4).

Pseudoreplication: an Issue of the Past

As shown in this forum, pseudoreplication no longer is an issue preventing the statistical analysis of experiments and field studies. Current statistical tools such as LMEM can model the lack of independence in field observations. However, pseudoreplication will most likely always be present in any ecological study, because of the complexity of working with living organisms that constrains full randomization or limits the number of replicates. Although other objects of study, like molecules or atoms, are numerous and widespread, samples of living organisms are comparatively few and organisms always are evolutionary and ecologically related at some scale. Although this forum has been focused on demystifying statistical concepts and presents how to use LMEM models to address the lack of independence in datasets, the ingenuity of statisticians is laudable because many other techniques outside the scope of this article have been developed over recent years. A best example includes the extension of LMEM to accommodate non-normal observations in generalized LMEMs (Bolker et al. 2009). Other tools that do not consider the individual variability of observations, but rather the average across all samples, like the generalized estimating equations (Faraway 2006), can address the lack of independence in observations, and have been used in the study of medically important insects (Lindblade et al. 2000, Gurevitz et al. 2009). A third line of new computer-based tools, including neural networks, trees (Olden et al. 2008), and random forests (Ruiz et al. 2010), does not have assumptions on data independence, and has been successfully used to study insects of public health importance (Hu et al. 2006, Ruiz et al. 2010). Thus, pseudoreplication should no longer be considered as a major flaw that impairs the statistical analysis of experiments and field studies. Independence constraints in the manipulation and observation of organisms are adequately handled by many available statistical tools, thus enabling valid inferences from valuable entomological data.

Acknowledgments

I am thankful to Yasmin Rubio-Palis for sharing her original data on *Anopheles nuneztovari* from Guaquitas, Venezuela. This work was funded by a Gorgas Research Award from the American Society of Tropical Medicine and Hygiene and Emory University. This work also benefited from comments by the editor, anonymous reviewers, Jorge Rabinovich, Nicole Gottdenker, and Greg Decker, and helpful discussions from a National Institutes of Health-Research and Policy on Infectious Disease Dynamics (NIH-RAPIDD) study group on mosquito-borne diseases.

References Cited

Alto, B. W., and S. A. Juliano. 2001a. Precipitation and temperature effects on populations of *Aedes albopictus* (Diptera: Culicidae): implications for range expansion. J. Med. Entomol. 38: 646–656.

Alto, B. W., and S. A. Juliano. 2001b. Temperature effects on the dynamics of *Aedes albopictus* (Diptera: Culicidae) populations in the laboratory. J. Med. Entomol. 38: 548–556.

Blumberg, A.J.Y., P. F. Hendrix, and D. A. Crossley. 1997. Effects of nitrogen source on arthropod biomass in notillage and conventional tillage grain sorghum agroecosystems. Environ. Entomol. 26: 31–37.

Bolker, B. M., M. E. Brooks, C. J. Clark, S. W. Geange, J. R. Poulsen, M. H. Stevens, and J. S. White. 2009. Generalized linear mixed models: a practical guide for ecology and evolution. Trends Ecol. Evol. 24: 127–135.

Box, J. F. 1980. R. A. Fisher and the design of experiments, 1922–1926. Am. Stat. 34: 1–7.

Carpenter, S. R. 1982. Stemflow chemistry: effects on population dynamics of detritivorous mosquitoes in tree-hole ecosystems. Oecologia 53: 1–6.

Chaves, L. F., and M. Pascual. 2007. Comparing models for early warning systems of neglected tropical diseases. PLoS Negl. Trop. Dis. 1: e33.

Chaves, L. F., C. L. Keogh, G. M. Vazquez-Prokopec, and U. D. Kitron. 2009. Combined sewage overflow enhances oviposition of *Culex quinquefasciatus* (Diptera: Culicidae) in urban areas. J. Med. Entomol. 46: 220–226.

Chesson, J. 1984. Effect of notonectids (Hemiptera, Notonectidae) on mosquitos (Diptera, Culicidae): predation or selective oviposition. Environ. Entomol. 13: 531–538.

Conner, J. K., and D. L. Hartl. 2004. A primer of ecological genetics. Sinauer, Sunderland, MA.

Faraway, J. J. 2006. Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models. CRC, Boca Raton, FL.

Feliciangeli, M. D., and J. Rabinovich. 1998. Abundance of *Lutzomyia ovallesi* but not *Lu-gomezi* (Diptera: Psychodidae) correlated with cutaneous leishmaniasis incidence in north-central Venezuela. Med. Vet. Entomol. 12: 121–131.

Fisher, R. A. 1935. The design of experiments. Oliver & Boyd, Edinburgh, United Kingdom.

Fortin, M. J., and M.R.T. Dale. 2005. Spatial analysis: a guide for ecologists. Cambridge University Press, Cambridge, United Kingdom.

Gurevitz, J. M., U. Kitron, and R. E. Gurtler. 2009. Temporal dynamics of flight muscle development in *Triatoma infestans* (Hemiptera: Reduviidae). J. Med. Entomol. 46: 1021–1024.

Gurtler, R. E., F. M. Garelli, and H. D. Coto. 2009. Effects of a five-year citywide intervention program to control *Aedes aegypti* and prevent dengue outbreaks in northern Argentina. PLoS Negl. Trop. Dis. 3: e427.

Haile, F. J., D. L. Kerns, J. M. Richardson, and L. G. Higley. 2000. Impact of insecticides and surfactant on lettuce physiology and yield. J. Econ. Entomol. 93: 788–794.

Harrington, L. C., A. Ponlawat, J. D. Edman, T. W. Scott, and F. Vermeylen. 2008. Influence of container size, location, and time of day on oviposition patterns of the dengue vector, *Aedes aegypti*, in Thailand. Vector Borne Zoonotic Dis. 8: 415–423.

Hayes, J., and T. D. Downs. 1980. Seasonal changes in an isolated population of *Culex pipiens quinquefasciatus* (Dipter: Culicidae): a time series analysis. J. Med. Entomol. 17: 63–69.

Heffner, R. A., M. J. Butler, and C. K. Reilly. 1996. Pseudoreplication revisited. Ecology 77: 2558–2562.

Hu, W., S. Tong, K. Mengersen, B. Oldenburg, and P. Dale. 2006. Mosquito species (Diptera: Culicidae) and the transmission of Ross River virus in Brisbane, Australia. J. Med. Entomol. 43: 375–381.

Hurlbert, S. H. 1984. Pseudoreplication and the design of ecological field experiments. Ecol. Monogr. 54: 187–211.

Kirby, M., P. Milligan, D. Conway, and S. Lindsay. 2008. Study protocol for a three-armed randomized controlled trial to assess whether house screening can reduce exposure to malaria vectors and reduce malaria transmission in The Gambia. Trials 9: 33.

Kirby, M. J., D. Ameh, C. Bottomley, C. Green, M. Jawara, P. J. Milligan, P. C. Snell, D. J. Conway, and S. W. Lindsay. 2009. Effect of two different house screening interventions on exposure to malaria vectors and on anemia in children in The Gambia: a randomized controlled trial. Lancet 374: 998–1009.

Koenraadt, C.J.M., J. Aldstadt, U. Kijchalao, A. Kengluecha, J. W. Jones, and T. W. Scott. 2007. Spatial and temporal patterns in the recovery of *Aedes aegypti* (Diptera: Culicidae) populations after insecticide treatment. J. Med. Entomol. 44: 65–71.

Koenraadt, C.J.M., J. Aldstadt, U. Kijchalao, R. Sithiprasasna, A. Getis, J. W. Jones, and T. W. Scott. 2008. Spatial and temporal patterns in pupal and adult production of the dengue vector *Aedes aegypti* in Kamphaeng Phet, Thailand. Am. J. Trop. Med. Hyg. 79: 230–238.

Lindblade, K. A., E. D. Walker, A. W. Onapa, J. Katungu, and M. L. Wilson. 2000. Land use change alters malaria transmission parameters by modifying temperature in a highland area of Uganda. Trop. Med. Int. Health 5: 263–274.

Millar, R. B., and M. J. Anderson. 2004. Remedies for pseudoreplication. Fisheries Res. 70: 397–407.

Montgomery, D. C. 2005. Design and analysis of experiments. Wiley, New York, NY.

Oksanen, L. 2001. Logic of experiments in ecology: is pseudoreplication a pseudoissue? Oikos 94: 27–38.

Olden, J. D., J. J. Lawler, and N. LeRoy-Poff. 2008. Machine learning methods without tears: a primer for ecologists. Q. Rev. Biol. 83: 171–193.

Oyediran, I. O., M. L. Higdon, T. L. Clark, and B. E. Hibbard. 2007. Interactions of alternate hosts, postemergence grass control, and rootworm-resistant transgenic corn on western corn rootworm (Coleoptera: Chrysomelidae) damage and adult emergence. J. Econ. Entomol. 100: 557–565.

Pinheiro, J. C., and D. M. Bates. 2000. Mixed effects models in S and S-plus. Springer, New York, NY.

Pitcairn, M. J., L. T. Wilson, R. K. Washino, and E. Rejmankova. 1994. Spatial patterns of *Anopheles freeborni* and *Culex tarsalis* (Diptera: Culicidae) larvae in California rice fields. J. Med. Entomol. 31: 545–53.

Reiskind, M. H., and M. L. Wilson. 2004. *Culex restuans* (Diptera: Culicidae) oviposition behavior determined by larval habitat quality and quantity in southeastern Michigan. J. Med. Entomol. 41: 179–86.

Rubio-Palis, Y., and C. F. Curtis. 1992. Biting and resting behavior of anophelines in western Venezuela and implications for control of malaria transmission. Med. Vet. Entomol. 6: 325–334.

Ruiz, M. O., L. F. Chaves, G. L. Hamer, T. Sun, W. M. Brown, E. D. Walker, L. Haramis, T. L. Goldberg, and U. D. Kitron. 2010. Local impact of temperature and precipitation on West Nile virus in *Culex* species mosquitoes in northeast Illinois, U.S.A. Parasit. Vectors 3: 19.

Salomon, O. D., M. L. Wilson, L. E. Munstermann, and B. L. Travi. 2004. Spatial and temporal patterns of phlebotomine sand flies (Diptera: Psychodidae) in a cutaneous leishmaniasis focus in northern Argentina. J. Med. Entomol. 41: 33–39.

Scott, T. W., A. C. Morrison, L. H. Lorenz, G. G. Clark, D. Strickman, P. Kittayapong, H. Zhou, and J. D. Edman. 2000. Longitudinal studies of *Aedes aegypti* (Diptera: Culicidae) in Thailand and Puerto Rico: population dynamics. J. Med. Entomol. 37: 77–88.

Shumway, R. H., and D. S. Stoffer. 2000. Time series analysis and its applications. Springer, New York, NY.

Strickman, D. 1988. Rate of oviposition by *Culex quinquefasciatus* in San Antonio, Texas, during three years. J. Am. Mosq. Control Assoc. 4: 339–344.

Vazquez-Prokopec, G. M., M. C. Cecere, U. Kitron, and R. E. Gurtler. 2008. Environmental and demographic factors determining the spatial distribution of *Triatoma guasayana* in peridomestic and semi-sylvatic habitats of rural northwestern Argentina. Med. Vet. Entomol. 22: 273–282.

Received 12 October 2009; accepted 20 January 2010.