Long-term experiments and strip plot designs

Antonella Plaia*

Department of Scienze Economiche, Aziendali e Statistiche, University Palermo, Viale delle Scienze, 90128 Palermo, Italy

(Received 25 February 2014; accepted 28 April 2015)

In a long-term experiment usually the experimenter needs to know whether the effect of a treatment varies over time. But time usually has both a fixed and a random effects over the output and the difficulty in the analysis depends on the particular design considered and the availability of covariates. Actually, as shown in the paper, the presence of covariates can be very useful to model the random effect of time. In this paper a model to analyze data from a long-term strip plot design with covariates is proposed. Its effectiveness will be tested using both simulated and real data from a crop rotation experiment.

Keywords: experimental designs; repeated measures; covariates

AMS Subject Classification: 62K99; 05B30

1. Introduction

Long-term experiments (LTEs) are commonly used in agronomy, soil science, ecology, biology, medicine and other disciplines to compare the effects of different treatment regimes over an extended length of time, usually years [8]. Treatments are usually assigned to experimental units at the start of the study, and measurements of interest are observed regularly over the course of the trial. Usually the same manipulation is applied one or more times over the course of the trial (annual fertilization, tillage regimes, and so on) and/or manipulation varying in a planned manner over time, such as crop rotation, can be involved. Different experimental designs can be considered, ranging from randomized complete block (RCB) designs to split plot and strip plot designs. Even in real experiments the two things (repeated measures and experimental designs) are usually combined, very often only one out of the two is emphasized during the analysis of the data.

In an LTE usually the purpose is to test and estimate the time × treatment interaction: that is, the experimenter needs to know not only the effect of a treatment but, most of all, if this

*Email: antonella.plaia@unipa.it

© 2015 Taylor & Francis
effect varies over time. In this sense, time should be considered as a fixed effect. Actually, measurements are usually influenced by uncontrollable environmental factors which can be modeled by random year-to-year fluctuations and year \( \times \) treatment random effects. Even if these random effects could be only a nuisance, they need to be taken into account, since variations in the measurements across years are simultaneously due to both fixed and random effects [7].

In this paper, following the approach proposed by Loughin et al. [8], data from a strip plot LTE are analyzed, considering both fixed and random effects of time, eventually including covariates which can be useful to distinguish between the different effects of time.

The paper is arranged as follows: Section 2 describes a suitable approach to analyze data from an LTE, while Section 3 shows the right model and corresponding ANOVA table for a strip plot design. In Section 4 the two approaches will be merged, and a model to analyze data from a long-term strip plot will be proposed. Section 5 describes the use of covariates in the model, while Section 6 shows the results from a real LTE. Finally, some discussion and conclusions can be found in Section 7.

2. Long-term experiment

An easy way to analyze LTE consists in splitting data into series of one-time (1 year) experiments (see [8] for references following this approach), performing separate analyses of variance at each time: in this way no formal study of the eventual cumulative treatment effect is possible. For example, Katsvairo et al. [5] performed separate analysis for each year of the experiment.

As an alternative, statistical methods for longitudinal data are often used in this context but, as shown in [8], most standard approaches to repeated measure analysis may provide incorrect results, since they fail to model both the random environmental effects and their interactions with treatments.

An interesting and useful distinction is the one in [10] between ‘treatment factors’, that can be randomized, and ‘repeated factors’, which cannot: time is a typical example of a ‘repeated factor’. Repeated factors, being unable to be randomized, give raise to serial correlation, that needs to be modeled adequately. This can be done by considering realistically a decay with an increase in the ‘distance’ between observations. An AR(1) model or a power model can be suitable [10].

Singh and Jones [13] focus on modeling the serial correlation between repeated measurements from the same plots, ignoring the random effects associated with annual variations. In particular, they tried to identify the most suitable structure of correlation (covariance) of the plot errors.

Even if the approach followed in [13] considers the effect of time, much emphasis is given to the modeling of correlation between repeated measurements from the same plot, comparing different correlation structures, but almost ignoring random effects associated with annual variations. Although correlation between measurements made on the same plot should be properly considered, Loughin et al. [8] demonstrated that any apparent serial correlation is likely to be masked by random annual variation. They showed (see [8, p. 35]) that correlation between two measurements from the same unit can be very low even if random errors \((\varepsilon_{ikt})\) in model (1) are correlated. This happens when large variations among cycles of time are observed.

For this reason, the approach followed in this paper explicitly separates the fixed from the random effect of time: for this reason we will use the term ‘year’ [8] to mean the random effect representing environmental fluctuation, and the term ‘cycle’ (borrowed from crop rotation literature) for the factor representing the fixed effect of time.

Note that, when an experiment is begun at a single point in time, all measurements are subject to the same \(r\)-year sequence of environmental characteristics and, since measurements in year \(k\) could be influenced both by environmental conditions in year \(k\) and by environmental conditions in years \((1: k - 1)\), unless the variability due to these sequences of environmental conditions can be quantified, inferences should be restricted to sequences of characteristics at least similar to the
observed ones. Standard methods for the analysis of repeated measures data cannot be applied, since one of the fundamental assumptions is that subjects (units) must respond independently from one another. This is not the case of LTE employing standard designs, since all plots are affected simultaneously by the same random effect of year.

In order to quantify and account for random environmental conditions, additional information must be supplied, or some assumptions must be made about the nature of fixed and random effects (‘cycle’ and ‘year’) of time.

In order to separate fixed and random effects of time Loughin et al. [8] suggest to make assumptions on their nature, that is:

- to make an assumption about the structure of the fixed effect of time, for example hypothesizing a particular trend for the mean effect of time;
- to add covariates to the model to account for the random effects of time and time × treatment.

The first solution can be considered only if we have an idea of what trend the mean follows, and some graphical representation of data can help in this sense. The second approach need that yearly recorded covariates are available, in order to be included in the model as fixed effects.

Considering a single factor with \( a \) levels, let \( Y_{ikt} \) represent a measurement taken at time \( t = 1, 2, \ldots, T \) on a unit from block \( k = 1, 2, \ldots, K \) receiving treatment \( i = 1, 2, \ldots, a \). The univariate model for repeated measures taken in an RCB design is

\[
Y_{ikt} = \mu + r_k + \alpha_i + (r\alpha)_{ik} + \gamma_t + g_t + (\alpha\gamma)_{it} + (\alpha g)_{it} + \epsilon_{ikt},
\]

where \( \mu \) is a general mean effect, \( r_k \sim N(0, \sigma^2_r) \) is the random effect of block \( k \), \( \alpha_i \) is the fixed main effect of treatment \( i \) (level \( i \) of factor A), \( (r\alpha)_{ik} \sim N(0, \sigma^2_{r\alpha}) \) is the random plot error, that is, a random error effect for factor A, \( \gamma_t \) is the fixed main effect of ‘cycle’ \( t \), \( g_t \sim N(0, \sigma^2_g) \) is the random effect of ‘year’, \( (\alpha\gamma)_{it} \) is the fixed interaction of cycle (time) and treatment, \( (\alpha g)_{it} \sim N(0, \sigma^2_{\alpha g}) \) is the random interaction of time and treatment, \( \epsilon_{ikt} \) is the random error associated with measurement taken at time \( t \) on a unit in a block \( k \) receiving treatment \( i \) and is \( N(0, \sigma^2_e) \). \( g_t \) and \( (\alpha g)_{it} \) are independent of each other, \( \epsilon_{ijk} \) and \( \epsilon_{i'j'k'} \) are independent of each other if \( i \neq i' \) or \( j \neq j' \), but may be correlated with a correlation \( \rho_{kk'} \) if \( i = i' \) and \( j = j' \), and finally \( \epsilon_{ijk} \) are independent of \( (r\alpha)_{ik} \) if \( i, j, k \).

The terms \( g_t \) and \( (\alpha g)_{it} \) account for the random variation associated with years and for possible random variations in treatment effect associated with yearly fluctuation. These two terms can be modeled as functions of covariates, whenever available, as \( g_t = \beta_0 + \beta_1 x_{1t} + \cdots + \beta_q x_{qt} + g'_t \) with \( g'_t \sim N(0, \sigma^2_g) \) and \( (\alpha g)_{it} = \beta_{i0} + \beta_{i1} x_{1t} + \cdots + \beta_{iq} x_{qt} + (\alpha g)_{it}' \) with \( (\alpha g)_{it}' \sim N(0, \sigma^2_{\alpha g}) \).

Let \( \tilde{Y}_{i,t} = K^{-1} \sum_k Y_{ikt}, \tilde{Y}_{i,t} = (KT)^{-1} \sum_k \sum_t Y_{ikt}, \tilde{Y}_{i,t} = (Ka)^{-1} \sum_k \sum_{i, t} Y_{ikt} \), Table 1s (on-line supplemental material) shows the variances of the components for model (1).

Under model (1) with covariates the terms \( \sigma^2_g \) and \( \sigma^2_{\alpha g} \) are substituted by \( \sigma^2_{g'} \) and \( \sigma^2_{\alpha g'} \) respectively, and, if the covariates are effective in explaining the random variation, \( \sigma^2_{g'} < \sigma^2_g \) and \( \sigma^2_{\alpha g'} < \sigma^2_{\alpha g} \).

3. Strip plot experimental design

Strip plot designs are common in plant breeding, animal science, health science experiments, especially when two factors (A and B) requiring large experimental units are to be tested in the same experiment. Lansky [6] shows how strip plot designs can be a powerful tool for identifying subtle effects of factors within biological asseys, while Farewell and Herzberg [3] show their usefulness in studies of the training of medical practitioners.
Figure 1. Lay-out of the crop rotation experiment.

Strip plot designs go also under different names, including strip block design, split block experiment design, two-way whole plot design, sub-treatments in strips across blocks and a criss-cross design (for complete references see [4]). In such a design, experimental field is divided into blocks and each block is divided into strips perpendicular to each other (Figure 1). Factor A is randomly applied to strips in one direction, while Factor B is randomly applied to strips, which are actually a new set of whole plots, orthogonal to the original plots used for factor A. Here, different from a split plot design, characterized by whole plots and subplots, there are two whole plot treatments, A and B: an RCB experiment design is used for factor A treatments, but also the factor B treatments are arranged in an RCB design. The levels of factor B go across all levels of factor A and vice versa in a criss-cross manner. This arrangement, with a different randomization, is repeated in each of the (say $K$) complete blocks.

The usual response model equation is

$$Y_{ijk} = \mu + r_k + \alpha_i + (r\alpha)_{ik} + \beta_j + (r\beta)_{jk} + (\alpha\beta)_{ij} + \varepsilon_{ijk},$$

(2)

with $k = 1, 2, \ldots, K$, $i = 1, 2, \ldots, a$ and $j = 1, 2, \ldots, b$

where $Y_{ijk}$ is the response for the $ijk$th experimental unit; $\mu$ is a general mean effect, $r_k$ is the $k$th block effect and is IID$(0, \sigma_r^2)$, $\alpha_i$ is the effect of the $i$th level of factor A, $(r\alpha)_{ik}$ is a random error effect for factor A and is IID$(0, \sigma_{\alpha}^2)$, $\beta_j$ is the effect of the $j$th level of factor B, $(r\beta)_{jk}$ is a random error effect for factor B and is IID$(0, \sigma_{\beta}^2)$, $(\alpha\beta)_{ij}$ is the $ij$th interaction effect of the two factors A and B, and $\varepsilon_{ijk}$ is a random error effect for the interaction effects and is IID$(0, \sigma_{\varepsilon}^2)$.

The different random effects $(r\alpha)_{ik}$, $(r\beta)_{jk}$, and $\varepsilon_{ijk}$ are assumed to be independent.

Since the experimental units are different for factors A and B and for their interaction, three different error terms are required in this design.

Tables 1 and 2 show, respectively, the expected values of the various mean squares and the ANOVA table of a standard strip plot design with fixed A and B and random block.

With the assumption of normality for the random error effects in Equation (2), the $F$-tests in Table 2 can be used for fixed effects for both factors. Each of the three $F$-tests requires a
Table 1. Expected mean squares for the ANOVA of a standard strip plot design.

| Source of variation | Degrees of freedom | Expected mean square values |
|---------------------|--------------------|-----------------------------|
| Total               | Kab                | $\sigma^2 + ab\sigma^2$     |
| Block               | $K - 1$            | $\sigma^2 + b\sigma^2_a$   |
| A                   | $a - 1$            | $\sigma^2 + b\sigma^2_a + Kb\sum_{i=1}^a\alpha_i^2/(a - 1)$ |
| Block × A           | $(K - 1)(a - 1)$   | $\sigma^2 + b\sigma^2_a + Kb\sum_{j=1}^b\beta_j^2/(b - 1)$ |
| B                   | $b - 1$            | $\sigma^2 + a\sigma^2_b$   |
| Block × B           | $(K - 1)(b - 1)$   | $\sigma^2 + a\sigma^2_b + Ka\sum_{j=1}^b\beta_j^2/(b - 1)$ |
| A × B               | $(a - 1)(b - 1)$   | $\sigma^2 + K\sum_{i=1}^a\sum_{j=1}^b(\alpha\beta)^2 i/(a - 1)(b - 1))$ |
| Block × A × B       | $(K - 1)(a - 1)(b - 1)$ | $\sigma^2$               |

Table 2. ANOVA table for a standard strip plot design.

| Source of variation | Degrees of freedom | Sum of squares square | Mean | $F$ |
|---------------------|--------------------|-----------------------|------|-----|
| Block               | $K - 1$            | SSBlock               | MSA  | $E_a$|
| A                   | $a - 1$            | SSA                   | MSA  | $E_a$|
| Error(a)            | $(K - 1)(a - 1)$   | SSE$_1$               | MSE$_1$ = $E_a$|
|                     |                    | $ab\sum_{k}^K(\bar{y}_{..,k} - \bar{y}_{..})^2$ |       |     |
| B                   | $b - 1$            | SSB                   | MSB  | $E_b$|
| Error(b)            | $(K - 1)(b - 1)$   | SSE$_2$               | MSE$_2$ = $E_b$|
| A × B               | $(a - 1)(b - 1)$   | SSE(AB)               | MS(AB) | MS(AB)/$E_{ab}$ |
| Error(ab)           | $(K - 1)(a - 1)(b - 1)$ | SSE$_3$   | MSE$_3$ = $E_{ab}$|

3.1 Artificial example

To make ideas clearer, consider a simulated example where data are generated according to model (2), with A (significant) as the row factor and B (significant) as the column factor with three levels each, with field divided into $K = 2$ blocks. Figure 2 shows the interaction plot for the simulated data, while Table 3 shows the corresponding ANOVA table. Both the figure and the tests in the ANOVA table show that the two factors and their interaction are significant. All the analysis is run in R, by functions aov and lmer (library lme4) [2,11].
Table 3. ANOVA table for the simulated data.

| Source of variations | Degrees of freedom | Sum of squares | Mean square | F      | Pr(F)  |
|----------------------|--------------------|----------------|-------------|--------|--------|
| Block                | 1                  | 1.9339         | 1.9339      |        |        |
| Stratum block:A      | 2                  | 2247.96        | 1123.98     | 114.37 | 0.008667 |
| A                    | 2                  | 19.65          | 9.83        |        |        |
| Error (a)            | 2                  | 0.964          | 0.482       |        |        |
| Stratum block:B      | 2                  | 310.173        | 155.087     | 321.61 | 0.0031 |
| B                    | 2                  | 0.964          | 0.482       |        |        |
| Error (b)            | 2                  | 78.773         | 19.6933     | 19.781 | 0.006724 |
| Stratum Within A:B   | 4                  | 3.982          | 0.9956      |        |        |

4. Long-term strip plot

When a standard strip plot designed experiment is conducted at several sites, in several years, or repeated in some other way, the researcher may want to combine the results from the individual experiments, even if results of the individual experiments will need to be obtained and interpreted as well. LTEs and strip plot designs are both schemes followed in agriculture and other field experiments, but, to our knowledge, a complete analysis of data coming from a long-term strip plot design of experiments is absent in the literature. Even if the experiment was conducted according to a long-term strip plot one, data are not analyzed appropriately [5], for example, consider an experiment where treatments were arranged in a strip plot design and observed over six years, but separate analyses for each year are presented. Federer and King [4] describe how to combine results from a strip plot experiment over several sites, but the approach cannot be immediately extended to the case of several years (long-term), though theoretically considered by the authors at the beginning of the book chapter. As a matter of fact, Federer and King [4] do
not distinguish between fixed and random effect of time. Here, we will try to merge the model for
data coming from a strip plot experimental design (2) to the model to analyze data from an RCB
long-term experimental design (1): to do this we will consider ‘time’ as the ‘outmost block’.

The new model (3), obtained by opportunely merging models (1) and (2), allows to account
simultaneously for the effect of time and the presence of a specific design of experiment (strip
plot):

\[ Y_{ijkt} = \mu + \gamma_t + g_t + (rg)_{kt} + \alpha_i + (\alpha \gamma)_i + \beta_j + (\beta \gamma)_j + (\beta \gamma)_{ij} + \epsilon_{ijkt}, \]

(3)

where \( Y_{ijkt} \) is the response for the \( ijkt \)th experimental unit; \( \mu \) is a general mean effect, \( \alpha_i \) is the
effect of the \( i \)th level of factor A, \( \beta_j \) is the effect of the \( j \)th level of factor B, \( \gamma_t \) is the fixed effect
of time (cycle) \( t \)th, \( g_t \) is the random effect of year and is \( N(0, \sigma^2_g) \), \( (rg)_{kt} \) is the random error
effect for time and is \( N(0, \sigma^2_{rg}) \), \( (\alpha \gamma)_i \) is the random error effect for factor A and is \( N(0, \sigma^2_{\alpha \gamma}) \),
(\( \beta \gamma)_j \) is a random error effect for factor B and is \( N(0, \sigma^2_{\beta \gamma}) \), \( (\alpha \beta)_ij \) is the \( ij \)th interaction
effect of the two factors A and B, \( (\alpha \gamma)_i \) is the fixed interaction of cycle (time) and treatment A, \( (\beta \gamma)_j \)
is the fixed interaction of cycle (time) and treatment B, \( (\alpha \beta \gamma)_{ij} \) is the fixed interaction of cycle (time), treatment A and treatment B, \( \epsilon_{ijkt} \) is a random error effect for the interaction effects and
is \( N(0, \sigma^2_{\epsilon}) \).

Model (3) has been obtained by considering time as the outmost factor and, therefore, starting
from models (1) and (2):

- by substituting \( r_k \) with \( (rg)_{kt} \), the random interaction of time and block effects;
- by substituting \( (r\alpha)_{ik} \) with \( (\alpha \gamma)_{ik} \);
- by substituting \( (r\beta)_{jk} \) with \( (\beta \gamma)_{jk} \).

If we look at model (3) as the extension of model (2) in [8] to the strip plot design, its appropriateness
to model long-term strip plot experiments can be immediately gathered. In model (3) four
different error terms can be distinguished: \( \sigma^2_{rg} \) to test the effect of time, \( \sigma^2_{\alpha \gamma} \) and \( \sigma^2_{\beta \gamma} \) to test the
effect of A and B, respectively, and their interaction with time, while \( \sigma^2_{\epsilon} \) allows to test the A \( \times \) B
and A \( \times \) B \( \times \) time interactions. Table 4 shows the expected mean squares for the ANOVA of a
long-term strip plot design. The corresponding ANOVA table is shown in Table 5. With respect
to Table 2, one more stratum has been added, with sum of squares for Time and the correspond-
ing error term (Error(T)). As it is possible to see from Table 4, there are no proper error terms
for the interactions A \( \times \) Time, B \( \times \) Time and A \( \times \) B \( \times \) Time. The naively F-values computed by
using \( E_{\alpha}, E_{\beta} \) and \( E_{\alpha \beta} \), respectively, as denominator (Table 5) can lead to biased tests for the term
\( \sigma^2_{g} \) (random effect of time) that inflates the numerators.

### 4.1 Artificial example

To make ideas clearer, consider 100 simulated examples where data are generated according to
model (3) (without the second-level interaction \( (\alpha \beta \gamma)_{ij} \)), with A as the row factor and B as the
column factor with three levels each, with field divided into two blocks, the experiment was
conducted for 5 years. Figure 3 shows the interaction plots for a single simulated data set, while
Table 6 shows the true model parameters (according to model (3)) and their mean estimates
over the 100 simulated data sets according to model (3). As expected, in some cases parameter
estimates are not correct: this is especially true for the two variances \( \sigma^2_{g} \) and \( \sigma^2_{rg} \). The problem is
that the fixed and random effects of time are partially confounded. A solution to this problem is
proposed in the next section.
In model (3), as already explained for model (1), the fixed effect of time, $\gamma$, is partially
confounded with its random effect $g$, and the naively $F$ values can lead to biased tests.

As already stated in Section 2, two alternative solutions can be considered to solve this
problem: a model for the fixed effect or, if available, covariates to explain random variation.

The first choice needs to be supported at least by an idea of a possible model for the mean, but
if the chosen model is inadequate the estimated mean results will be misleading.

### Table 4. ANOVA table for a long-term strip plot design.

| Source of variation | Degrees of freedom | Expected mean square values |
|---------------------|--------------------|----------------------------|
| Total               | $Kab$              | $\sigma^2_e + ab\sigma^2_{rg} + Kab\sum_{t=1}^T\gamma^2_t/(T-1)$ |
| Time                | $T - 1$            | $\sigma^2_e + ab\sigma^2_{rg}$ |
| Block × Time        | $T(K-1)$           | $\sigma^2_e + ab\sigma^2_{rg}$ |
| A                   | $a - 1$            | $\sigma^2_e + ab\sigma^2_{rg} + TKb\sum_{a}^a\alpha^2_i/(a - 1)$ |
| A × Time            | $(a - 1)(T - 1)$   | $\sigma^2_e + ab\sigma^2_{rg} + K\sum_{a}^a(a\beta^2_j)/((a - 1)(b - 1))$ |
| Block × A × Time    | $T(K-1)(a - 1)$    | $\sigma^2_e + ab\sigma^2_{rg} + K\sum_{a}^a(a\beta^2_j)/((a - 1)(b - 1))$ |
| B                   | $b - 1$            | $\sigma^2_e + ab\sigma^2_{rg} + TKa\sum_{b}^b\beta^2_j/(b - 1)$ |
| B × Time            | $(b - 1)(T - 1)$   | $\sigma^2_e + ab\sigma^2_{rg} + K\sum_{b}^b\beta^2_j/(b - 1)$ |
| Block × B × Time    | $T(K-1)(b - 1)$    | $\sigma^2_e + ab\sigma^2_{rg}$ |
| A × B               | $(a - 1)(b - 1)$   | $\sigma^2_e + ab\sigma^2_{rg} + K\sum_{a}^a(a\beta^2_j)/(a - 1)(b - 1)$ |
| A × B × Time        | $(T - 1)(a - 1)(b - 1)$ | $\sigma^2_e + ab\sigma^2_{rg} + K\sum_{a}^a(a\beta^2_j)/(a - 1)(b - 1)$ |
| Block × A × B × Time| $T(K-1)(a - 1)(b - 1)$ | $\sigma^2_e$ |

### Table 5. Expected mean squares for the ANOVA of a long-term strip plot design.

| Source of variation | Degrees of freedom | Sum of squares square | Mean | $F$ |
|---------------------|--------------------|-----------------------|------|-----|
| Time                | $T - 1$            | $\sum_{t=1}^T(\bar{y}_{..t} - \bar{y}_{..})^2$ | MST  | $MST/E_T$ |
| Error(T)            | $T(K-1)$           | $\sum_{t=1}^T(\bar{y}_{..t} - \bar{y}_{..})^2$ | MSE1 | $E_T$ |
| A                   | $a - 1$            | $\sum_{a}^a\alpha^2_i/(a - 1)$ | MSA  | $E_a$ |
| A × Time            | $(T - 1)(a - 1)$   | $\sum_{t=1}^T(\bar{y}_{..t} - \bar{y}_{..})^2$ | MS(AT)| $E_a/\sum_{a}^a\alpha^2_i/(a - 1)$ |
| Error(a)            | $T(K-1)(a - 1)$    | $\sum_{a}^a\alpha^2_i/(a - 1)$ | MSE2 | $E_a$ |
| B                   | $b - 1$            | $\sum_{b}^b\beta^2_j/(a - 1)(b - 1)$ | MSB  | $E_b$ |
| B × Time            | $(T - 1)(b - 1)$   | $\sum_{b}^b\beta^2_j/(a - 1)(b - 1)$ | MS(BT)| $E_b/\sum_{b}^b\beta^2_j/(a - 1)(b - 1)$ |
| Error(b)            | $T(K-1)(b - 1)$    | $\sum_{b}^b\beta^2_j/(a - 1)(b - 1)$ | MSE3 | $E_b$ |
| A × B               | $(a - 1)(b - 1)$   | $\sum_{a}^a\alpha^2_i/(a - 1)(b - 1)$ | MS(AB)| $E_ab$ |
| A × B × Time        | $(T - 1)(a - 1)(b - 1)$ | $\sum_{a}^a\alpha^2_i/(a - 1)(b - 1)$ | SS(ABT)| $E_ab/\sum_{a}^a\alpha^2_i/(a - 1)(b - 1)$ |

5. **Long-term strip plot with covariates**

In model (3), as already explained for model (1), the fixed effect of time, $\gamma$, is partially
confounded with its random effect $g$, and the naively $F$ values can lead to biased tests.

As already stated in Section 2, two alternative solutions can be considered to solve this
problem: a model for the fixed effect or, if available, covariates to explain random variation.

The first choice needs to be supported at least by an idea of a possible model for the mean, but
if the chosen model is inadequate the estimated mean results will be misleading.
If covariate data are available, a better choice can be to model $g_t$ as function of covariates: many ‘random’ variations are actually conglomerations of fixed effects [8]. For example, in agronomy, rainfall, solar radiation, etc., or synthetic measures that account for a combined effect of weather variables (like the Water Stress Index, WSI) can be considered, since they are known to affect crop yields and other agronomic measurements. According to this solution, a new model can be introduced by substituting $g_t$ in model (3) with:

$$g_t = \delta x_t + (\alpha x)_{it} + (\beta x)_{jt} + (\alpha \beta x)_{ijt} + g'_{t},$$

where $\delta x_t$ is the effect of the continuous covariate $x$ at time $t$, $(\alpha x)_{it}$ is the interaction of $x$ and treatment A, $(\beta x)_{jt}$ is the interaction of $x$ and treatment B, $(\alpha \beta x)_{ijt}$ is the interaction of $x$, treatment A and treatment B, and $g'_{t}$ is the random effect of time that, if the covariates are effective in explaining the random variations, is $N(0, \sigma^2_{g'})$, with $\sigma^2_{g'} < < \sigma^2_{g}$. As a result, the bias in the $F$ tests in Table 5 is reduced.

The complete model will be:

$$Y_{ijkt} = \mu + \gamma_t + \delta x_t + g'_{t} + (rg)_{tk} + \alpha_i + (\alpha \gamma)_{it} + (\alpha x)_{it} + (\alpha rg)_{itk} + \beta_j + (\beta \gamma)_{jt} + (\beta x)_{jt} + (\beta rg)_{jtk} + (\alpha \beta)_{ij} + (\alpha \beta \gamma)_{ijt} + (\alpha \beta x)_{ijt} + \varepsilon_{ijkt}.$$  \hspace{1cm} (4)

An application of this model, that can help the reader to understand it, is presented in Section 6.

5.1 Artificial example

Again, consider 100 simulated examples where data are generated according to model (4) (without the second-level interaction $(\alpha \beta \gamma)_{ijt}$), with A as the row factor and B as the column factor with 3 levels each, with field divided into 2 blocks, the experiment conducted for 5 years (parameter values are shown in Table 7). A covariate $x$ has now been introduced in the model. Figure 4 shows the interaction plots for a single simulated data set.

Table 7 shows the true model parameters and their mean estimates over 100 simulated data sets by fitting either model (4) and (3). As expected, variances estimated according to model (3)
Table 6. Parameter estimation over 100 simulated samples according to model (3).

| Model parameters | Truevalues | Estimated values (se) |
|------------------|------------|-----------------------|
| Intercept        | 1250       | 1249.8 (0.44)         |
| A2               | 15         | 15.5 (0.27)           |
| A3               | -10        | -8.9 (0.25)           |
| B2               | -5         | -5.0 (0.2)            |
| B3               | -1         | -0.6 (0.25)           |
| time2            | 1          | 1.1 (0.6)             |
| time3            | -2         | -2.5 (0.56)           |
| time4            | -4         | -4.9 (0.67)           |
| time5            | 3          | 3.5 (0.64)            |
| B2:A2            | -3         | -3.3 (0.2)            |
| B2:A3            | -200       | -200.4 (0.22)         |
| B3:A2            | 140        | 140.1 (0.19)          |
| B3:A3            | -5         | -5.1 (0.23)           |
| time2:A2         | 2          | 2.1 (0.35)            |
| time2:A3         | -3         | -4.0 (0.38)           |
| time3:A2         | 3          | 2.9 (0.38)            |
| time3:A3         | 2          | 1.2 (0.35)            |
| time4:A2         | -5         | -5.1 (0.35)           |
| time4:A3         | 2          | 1.0 (0.31)            |
| time5:A2         | 4          | 3.5 (0.4)             |
| time5:A3         | -1         | -2.0 (0.35)           |
| time2:B2         | 5          | 5.0 (0.28)            |
| time2:B3         | 3          | 2.5 (0.25)            |
| time3:B2         | -5         | -4.8 (0.3)            |
| time3:B3         | 2          | 1.5 (0.36)            |
| time4:B2         | 2          | 2.4 (0.24)            |
| time4:B3         | -5         | -5.0 (0.31)           |
| time5:B2         | 6          | 6.2 (0.28)            |
| time5:B3         | -1         | -1.3 (0.29)           |
| $\sigma^2 \gamma$ | 3          | 3.1 (0.1)             |
| $\sigma^2 \eta$  | 2.5        | 2.0 (0.1)             |
| $\sigma^2 \zeta$ | 2          | 1.8 (0.08)            |
| $\sigma^2 \phi$  | 1          | 0.9 (0.08)            |
| $\sigma^2 \epsilon$ | 3          | 3.0 (0.04)            |

(column Model without covariate) are much greater. Different from Table 6, now the variance estimates according to model (4) (column Complete model) are good. Note that, in order to reduce the number of parameters to estimate, Time is now being considered as quantitative (with a linear trend) and not as a factor.

6. A real long-term strip plot trial

Data are here presented from a rotation experiment established in a region in the South of Italy in 1992 [1].

A crop rotation is a cropping system in which different crops are grown in a sequence on the same piece of land, one after the other [13].

Figure 1 shows the layout of the considered experiment. It is a typical strip plot design, with two factors, Tillage and Precession, with three levels each, replicated over two Blocks and observed over 17 years (starting in 1992–1993).

Three crop sequences (Precession) as horizontal treatments, and three soil Tillage systems as vertical treatments have been considered.
Table 7. Parameter estimation over 100 data set generated according to model (4).

| Truemodel parameters | Estimates (s.e.) from complete model | Estimates (s.e.) from model without covariate |
|----------------------|--------------------------------------|-----------------------------------------------|
| Intercept            | 1250                                 | 1249.7 (0.76)                                 |
| A2                   | 15                                   | 15.4 (0.62)                                   |
| A3                   | −10                                  | −10.1 (0.61)                                  |
| B2                   | −5                                   | −5.2 (0.5)                                    |
| B3                   | −1                                   | −0.4 (0.61)                                   |
| time                 | 1                                    | 0.98 (0.12)                                   |
| x                    | 5                                    | 5.0 (0.02)                                    |
| A2:B2                | −3                                   | −3.4 (0.21)                                   |
| A3:B2                | −200                                 | −200.1 (0.2)                                  |
| A2:B3                | 140                                  | 140.1 (0.18)                                  |
| A3:B3                | −5                                   | −5.2 (0.18)                                   |
| A2:time              | 2                                    | 2.0 (0.09)                                    |
| A3:time              | −3                                   | −3.0 (0.09)                                   |
| B2:time              | 3                                    | 3.1 (0.09)                                    |
| B3:time              | 2                                    | 2.0 (0.09)                                    |
| A2:x                 | −5                                   | −5.0 (0.01)                                   |
| A3:x                 | 2                                    | 2.0 (0.01)                                    |
| B2:x                 | 12                                   | 12.0 (0.01)                                   |
| B3:x                 | −3                                   | −3.0 (0.01)                                   |
| $\sigma^2_{\epsilon}$ | 1.5                                 | 1.4 (0.13)                                   |
| $\sigma^2_{\epsilon}$ | 1                                   | 1.1 (0.09)                                   |
| $\sigma^2_{\epsilon}$ | 1.2                                 | 1.1 (0.07)                                   |
| $\sigma^2_{\epsilon}$ | 1.5                                 | 1.3 (0.07)                                   |
| $\sigma^2_{\epsilon}$ | 3                                   | 2.9 (0.04)                                   |

Figure 4. Artificial example of data generated according to model (4).

The crop sequences:

W–W: continuous wheat
W–FB: wheat–faba bean
W–BC: wheat–berseem clover
The tillage systems:

DT: Double Tillage
CT: Conventional Tillage
NT: No Tillage

Different aspects of crop productivity have been considered in the experiment: grain yield will be here analyzed.

Each phase of each rotation is present each year since, within each replication, there are two groups of plots (established in the first year) – one group of plots receives all the Tillage–Precession combination according to a strip plot design in their grain phase, while the second group of plots was planted to the second phase of the rotations.

Thus, although the year-wise observations (grain yield) are available for combinations of Tillage, Precession, and replication, they actually come (1) from one of the two groups of plots and (2) from the same plot only in alternate years.

As an example, the following schemes show the year-wise yield of grain \( Y \) from Plots 1 and 2 under the same crop-sequence/tillage combination and replication. We assume that in the W–FB rotation, Plot 1 is in the wheat (W) phase and Plot 2 in the faba-bean (FB) phase in the first year.

| Year | 1 | 2 | 3 | 4 | 5 | 6 | ... | T (= 17) |
|------|---|---|---|---|---|---|-----|---------|
| Plot1 | W | FB | W | FB | W | FB | ... | W |
| Plot2 | FB | W | FB | W | FB | W | ... | FB |
| Yield | \( Y_{11} \) | \( Y_{21} \) | \( Y_{12} \) | \( Y_{22} \) | \( Y_{13} \) | \( Y_{23} \) | ... | \( Y_{19} \) |

The series of grain yields \( Y_{11}Y_{12} \ldots \) arise from Plot 1 (in the grain phase (W) in odd years) while the grain yields \( Y_{21}Y_{22} \ldots \) from Plot 2 (beginning with the clover phase in Year 1 and receiving the grain phase in even years). The observations over years within each of the plots can be correlated, but independent (or with ignorable correlation) when they come from different plots.

Since data come from the same plot in alternate year, and emphasis is given to random environmental effects and their interaction with treatments, serial correlation will not be accounted for: this is justified by the autocorrelation plot within each of the 18 \( = 3 \times 3 \times 2 \) plot units shown in Figure 1 which shows the autocorrelation among measures \( y \) coming from the same plot at different lags (3 levels of Crop Sequence \( \times \) 3 levels of Tillage \times 2 blocks). As Payne [9] suggests, one of the issues to consider is the possible different amounts of random variation in different years. For this reason, a test for homogeneity of variances has been carried out resulting to be not significant.

For the same years meteorological data are available, which, for sake of simplicity, have been transformed into a crop WSI according to [12].

Since the experiment was conducted for several years, we are interested in combining the results from the single-year experiments. The availability of a covariate (WSI) allows to adapt model (4) to the data.

The ANOVA table for the whole data set is reported in Table 8, and the results can be verified looking at the figures in the online supplemental material.

In particular:

1. the first row (from the top) in Figure 2s shows the effect of Time (left, not significant) and WSI (right, significant), and corresponds to the tests in the Time:Rep stratum in Table 8,
Table 8. ANOVA table for the complete data.

| Source of variation | Degrees of freedom | Sum of squares | Mean square | F value | Pr(F) |
|---------------------|--------------------|---------------|-------------|---------|-------|
| **Time:Rep**        |                    |               |             |         |       |
| WSI                 | 1                  | 82,822,770    | 82,822,770  | 5.019   | 0.032 |
| Time                | 1                  | 16,595,463    | 16,595,463  | 1.006   | 0.323 |
| Residuals           | 31                 | 5.12E+08      | 16,503,301  |         |       |
| **Time:Rep:Tillage**|                    |               |             |         |       |
| Tillage             | 2                  | 779,373       | 389,686     | 1.002   | 0.373 |
| WSI:Tillage         | 2                  | 10,315,640    | 5,157,820   | 13.263  | 0.000 |
| Time:Tillage        | 2                  | 1,879,839     | 939,920     | 2.417   | 0.097 |
| Residuals           | 62                 | 24,111,912    | 388,902     |         |       |
| **Time:Rep:Prec**   |                    |               |             |         |       |
| Prec                | 2                  | 1.17E+08      | 58,397,984  | 32.846  | 1.88E-10 |
| WSI:Prec            | 2                  | 32713720      | 16,356,860  | 9.200   | 0.000317 |
| Time:Prec           | 2                  | 1976862       | 988,431     | 0.556   | 0.576366 |
| Residuals           | 62                 | 1.1E+08       | 1,777,948   |         |       |
| **Within**          |                    |               |             |         |       |
| Tillage:Prec        | 4                  | 3,245,229     | 811,307     | 4.543   | 0.00186 |
| WSI:Tillage:Prec    | 4                  | 212,431       | 53,108      | 0.297   | 0.8792 |
| Time:Tillage:Prec   | 4                  | 3,933,655     | 983,414     | 5.506   | 0.00041 |
| Residuals           | 124                | 22,145,961    | 178,596     |         |       |

(2) the second row in Figure 2 shows the interaction between Tillage and Time (not significant) and Tillage and WSI (significant), and corresponds to the tests in the Time:Rep:Tillage stratum in Table 8,
(3) the third row in Figure 2 shows the interaction between Crop-Sequence and Time (not significant) and Crop-Sequence and WSI (significant), and corresponds to the tests in the Time:Rep:Prec stratum in Table 8,
(4) Figure 3 shows the interaction between Tillage, Prec and Time (significant), while Figure 4 shows the interaction Tillage, Prec and WSI (not significant), and correspond to the tests in the ‘Within’ stratum in Table 8.

The effects of tillage varied greatly by climate (the WSI–Tillage interaction). When the WSI was high, grain yield was greater with conservative tillage techniques (especially NT) than with CT, whereas the opposite was true when the WSI was low; in growing seasons of medium water stress, no differences in grain yield were observed among the three tillage techniques. The effects of Tillage varied greatly by crop sequence (Tillage–Prec interaction was significant). On average, grain yield was higher with NT than CT when wheat was grown after FB or BC and lower with NT than CT in continuous wheat (W level). The latter effect grew stronger with time (Time–Tillage–Prec interaction was significant). The effects of the Tillage technique on grain yield were stable over time when wheat was grown after FB or BC, whereas RT and especially NT had a detrimental effect on continuous wheat (W) that increased with time. The effects of crop sequence were stable over time (the Time–Prec interaction was not significant) but varied greatly by WSI (the WSI–Prec interaction is significant).

7. Discussion and conclusions

In an LTE usually the purpose is to test and estimate the time × treatment interaction: that is, the experimenter needs to know whether the effect of the treatment varies over time. In this sense time should be considered as a fixed effect. On the other hand, measurements are usually
influenced by uncontrollable environmental factors which can be modeled by random year-to-year fluctuations and year \times treatment random effects. Even if these random effects could be only a nuisance, they need to be taken into account, since variations in the measurements across years are simultaneously due to both fixed and random effects [7]. As a result, time can have both a fixed and a random effect over the output.

Moreover, the analysis can be more difficult due to the particular design followed in the experiment, not necessarily a standard RCB one, and to the presence of covariates that, opportunely introduced in the model, can be very useful to model the random effect of time.

For example, a strip plot is a quite common experimental design and, at the same time, long-term experiments are frequently considered to account for the fixed and/or random effect of time. But the true potentiality of a particular experimental design established for a long time is fully exploited only if a model that properly accounts, at the same time, for the particular design and the presence of replications in time is formulated.

Even if LTE and strip plot designs are both schemes followed in Agriculture and other field experiments, up to our knowledge a complete analysis of data coming from a long-term strip plot design of experiments is absent in the literature. Even if the experiment was conducted according to a long-term strip plot one, data are not analyzed appropriately.

In this paper a model to analyze data from a long-term strip plot design with covariates is proposed. The model allows the fixed and random effects of time to be separated, introducing a covariate that can account for the random effect. Even if only a linear effect of covariates has been proposed, this model can be easily extended to a nonlinear effect of this covariate.

Both simulated and real data are used to show the adequacy of the proposed model, verified also by appropriate graphical representations which highlight the significant/not significant effects of main factors, time and covariate, together with their two- and three-way interactions.

Acknowledgements
The author is grateful to the associate editor and the anonymous referees for their very thoughtful and helpful comments and suggestions that doubtless improved this paper. The author thanks Prof. Gaetano Amato and Dario Giambalvo (Department of Scienze Agrarie e Forestali – University of Palermo) for the data analyzed in Section 6 recorded at Pietranera farm (Sicily, Italy).

Disclosure statement
No potential conflict of interest was reported by the authors.

Supplemental data and research materials
Supplemental data for this article can be accessed at 10.1080/02664763.2015.1046821. The code to simulate the data sets and analyze them will be available on request. Figures 1s–4s and Table 1s referenced in the article are available as online supplements to the article.

References
[1] G. Amato, P. Ruisi, A.S. Frenda, G. Di Miceli, S. Saia, A. Plaia, and D. Giambalvo, Long-term tillage and crop sequence effects on wheat grain yield and quality, Agron. J. 105 (2013), pp. 1317–1327.
[2] D. Bates, M. Maechler, B. Bolker, and S. Walker, lme4: Linear Mixed-effects Models using S4 Classes. R Package Version 1.0-6, 2014. Available at http://CRAN.R-project.org/package=lme4.
[3] V.T. Farewell and A.M. Herzberg, Plaid designs for the evaluation of training for medical practitioners, J. Appl. Stat. 30 (2003), pp. 957–965.
[4] W.T. Federer and F. King, Variations on Split Plot and Split Block Experiment Designs, John Wiley & Sons, Hoboken, NJ, 2007.
[5] T. Katsvairo, D. Wright, J. Marois, J. Rich, and P. Wiatrak, *Plant growth and development in two cotton rotations under irrigated and non-irrigated conditions*, Crop Sci. 49 (2006), pp. 2233–2245.

[6] D. Lansky, *Strip-plot designs, mixed models, and comparisons between linear and non-linear models for microtitre plate bioassays*, Dev. Biol. 107 (2002), pp. 11–23.

[7] T.M. Loughin, *Improved experimental design and analysis for long-term experiments*, Crop Sci. 46 (2006), pp. 2492–2502.

[8] T.M. Loughin, M.P. Roediger, G.A. Milliken, and J.P. Schmidt, *On the analysis of long-term experiments*, J. R. Stat. Soc. Ser. A 170 (2007), pp. 29–42.

[9] R. Payne, *The design and analysis of long-term rotation experiments*, Agron. J. (2012). doi:10.2134/agronj2012.0411.

[10] H.P. Piepho, A. Büchse, and C. Richter, *A mixed modelling approach for randomized experiments with repeated measures*, J. Agron. Crop Sci. 190 (2004), pp. 230–247.

[11] R Development Core Team, *R Foundation for Statistical Computing*, Vienna, Austria, 2014. Available at http://www.R-project.org/, ISBN 3-900051-07-0.

[12] F. Rizza, F.W. Badeck, L. Cattivelli, O. Lidestri, N. Di Fonzo, and A.M. Stanca, *Use of a water stress index to identify barley genotypes adapted to rainfed and irrigated conditions*, Crop Sci. 44 (2006), pp. 2127–2137.

[13] M. Singh and M.J. Jones, *Modeling yield sustainability for different rotations in long-term barley trials*, J. Agric. Biol. Environ. Stat. 7 (2002), pp. 525–535.