

Long-Term Experiments with cropping systems: Case studies on data analysis

Questa è la versione Post print del seguente articolo:

Original

Long-Term Experiments with cropping systems: Case studies on data analysis / Onofri, A.; Seddaiu, Giovanna; Piepho, H. P.. - In: EUROPEAN JOURNAL OF AGRONOMY. - ISSN 1161-0301. - 77:(2016), pp. 223-235.

Availability:

This version is available at: 11388/84479 since: 2022-05-27T16:23:02Z

Publisher:

Published

DOI:

Terms of use:

Chiunque può accedere liberamente al full text dei lavori resi disponibili come "Open Access".

Publisher copyright

note finali coverpage

(Article begins on next page)

1 Long-term experiments with cropping systems: case studies 2 on data analysis

3
4 ANDREA ONOFRI¹, GIOVANNA SEDDAIU^{2,3}, HANS-PETER PIEPHO⁴

5
6 ¹Dipartimento di Scienze Agrarie, Alimentari e Ambientali, Università degli Studi di Perugia, Italy

7 ²Dipartimento di Agraria, Università di Sassari, Italy

8 ³Nucleo di Ricerca sulla Desertificazione (NRD), Università di Sassari, Italy

9 ⁴Biometrics Unit, University of Hohenheim, Stuttgart, Germany

10
11 Correspondence: Prof. Andrea Onofri, Dipartimento di Scienze Agrarie ed Ambientali, Università
12 degli Studi di Perugia, Borgo XX Giugno 74, 06121, Perugia, Italy. Tel +39-075-5856324; fax:
13 +39-075-5856344; e-mail: andrea.onofri@unipg.it

14 15 **Abstract**

16
17 Data analysis for Long-Term Experiments (LTEs) with cropping systems requires some careful
18 thinking, especially for the most complex designs, characterised by rotations with different
19 durations and/or a different number of test-crops per rotation cycle. This paper takes an
20 example-based approach, built upon a number of datasets, covering the main types of LTEs, with
21 increasing levels of complexity. A procedure is outlined to build statistical models for data
22 analysis that is useful for all LTEs characterised by the simultaneous presence of all rotation
23 phases in all years, together with within-year replication. This procedure is based on the
24 assumption that correct analyses can be performed separately for each year. The use of mixed
25 models and REML estimation is advocated for model fitting with all LTEs, due to the fact that
26 most designs are non-orthogonal, as plots may not produce data for the test-crop under study in
27 all years. Mixed models are also useful to account for the autocorrelation of residuals over time
28 and hints are given for the selection of an appropriate variance-covariance structure. For all our
29 examples, variances were not constant across years and compound symmetry correlation
30 structures with variance heterogeneity of years proved to be the best compromise
31 between parsimony and statistical accuracy. Methods are outlined to test for the need of other
32 more complex correlation structures and examples are also given on how to test for fixed effects,
33 model fertility trends and assess the long-term stability of cropping systems.

34
35 **Key Words:** rotation, repeated measures, mixed model, autocorrelation, variance heterogeneity

36 37 **Highlights**

- 38 • Long-Term Experiments (LTEs) with cropping systems require careful data analyses
- 39 • Five types of LTEs with increasing complexity are considered and analysed
- 40 • A correct model building procedure is outlined and the use of REML estimation is
- 41 advocated
- 42 • The selection of variance-covariance structures and tests for fixed effects are discussed
- 43 • Long-term 'fertility' trends and stability analyses are also considered
- 44

45 **1. Introduction**

46

47 Long-Term Experiments (LTEs) represent an invaluable tool to detect possible slow changes
48 produced by the cropping systems in the long run and reveal possible threats to the
49 environment and to the future fertility of agricultural land. From a methodological point of view,
50 LTEs may be regarded as a particular class of multi-environment experiments, wherein
51 observations are repeated on the same plots for a long period of time, usually from 20 years
52 upwards (Frye and Thomas, 1991). The experimental treatments under comparison may be
53 highly variable (rotations, fertilisations, tillage systems ...), but in most cases they represent
54 several cropping systems, to be investigated by repeated observations of yield or other relevant
55 indicators of productivity, sustainability, efficiency and so on.

56 In general, data analysis for LTEs is not trivial, due to the peculiar characteristics of this
57 class of experiments. In a recent review, Payne (2015) highlighted several special issues, such
58 as: (i) non-uniform correlation structure of within-plot observations; (ii) heterogeneity of
59 variances in different years; (iii) residual effects of agronomic practices, leading to fertility
60 trends over time; (iv) changes in the experimental protocol, due to, e.g., the adoption of new
61 varieties, harvesting methods or weed control methods. Unfortunately, the statistical solutions
62 to these (and other) problems may not be unique and straightforward. Indeed, a survey of
63 literature shows that LTEs employ a wide array of experimental designs with their own
64 peculiarities. For example, rotation experiments may be designed so that all crops in the rotation
65 are grown every year (Cochran, 1939). However, in order to avoid an excessive increase in the
66 total number of plots, LTEs may also be designed without within-year replicates (Patterson,
67 1964), thus requiring a different approach to data analysis. Another example relates to the
68 frequency of repeated measures: in long rotations the test crop may return on the same plot
69 after a relatively long period of time, which may create a totally different correlation structure
70 compared to short rotations.

71 We will not list all the possible peculiarities of LTEs in terms of experimental design, as
72 this has already been accomplished in literature (see, e.g., Frye and Thomas, 1991). It should be
73 clear that each LTE, according to its aims and experimental design, poses peculiar problems in
74 terms of data analysis and may create difficulties in the selection of an appropriate statistical
75 method.

76 In this regard, a survey of literature shows a number of early reviews, discussing
77 ANOVA-like models for LTEs with a wide array of experimental designs (Patterson, 1964; Yates,
78 1954). However, these reviews use traditional methods of data analyses, while the current
79 availability of personal computers and mixed model procedures gives scientists some new
80 challenges and perspectives. On the other hand, more recent reviews about LTEs give examples
81 relating to few experimental designs and do not consider the most complex situations, such as
82 when LTEs include rotations of different lengths and/or with a different number of test crops
83 (see for example Richter and Kroschewski, 2006; Singh and Jones, 2002). Another review about
84 repeated measures includes only an example referring to a LTE comparing rotations of equal
85 lengths (Piepho et al., 2004). A further recent review takes an example-based approach and
86 presents a stepwise procedure for data analysis, though this is based on a single experiment,
87 characterised by no within-year replication (Payne, 2015).

88 Considering the above background, we felt that a further example-based treatment of the
89 statistical analysis of LTE data would be useful. Therefore, we used a number of datasets with
90 contrasting characteristics and elaborated several case studies to give examples on how to: (i)

91 build an appropriate model, (ii) inspect the validity of basic assumptions, (iii) test hypotheses of
92 interest and (iv) obtain reliable measures of productivity, stability and sustainability for the
93 cropping systems under comparison. The code to perform the analyses is made available as
94 supplemental material, with the aim of giving practical directions when dealing with LTEs.

95

96 **2. Five datasets for five types of long-term experiments**

97

98 Focusing on LTEs characterised by the presence of within-year replicates, five main groups can
99 be found in literature, with increasing levels of complexity in terms of data analysis:

100

- 101 1. LTEs with monocultures or perennial crops;
- 102 2. LTEs with different rotations of same length and one test crop per rotation cycle;
- 103 3. LTEs with a fixed rotation (one test crop per rotation cycle) and different treatments;
- 104 4. LTEs with more than one phase per crop and rotation cycle;
- 105 5. LTEs with several rotations of different lengths and/or different numbers of phases per crop
106 and rotation cycle.

107

108 Relating to the terms rotation, phase, cycle, sequence, we will refer to the terminology used in
109 Yates (1954), which is summarised in Table 1.

110

111 [Table 1 about here]

112 We cannot cover all these situations with datasets from real and independent experiments and,
113 therefore, instead of using simulated data, we decided to take a 'hybrid' approach: we started
114 from two real field experiments and formed different subsets of the data to fit the above list. In
115 more detail, we used the datasets from two real long-term field experiments: the first was
116 established around 1974 near Perugia (central Italy), at the Experimental Farm of the
117 Department of Agricultural, Food and Environmental Sciences and the second was established in
118 1994 at the Pasquale Rosati experimental farm in Agugliano (central Italy). Both experiments
119 are still running at present (2015) and a detailed description is given as supplemental material,
120 while other information can be found in Perucci et al. (1997) and elsewhere in this same special
121 issue (Bonciarelli et al., 2016).

122 These two real LTEs were used to extract five exemplary datasets, representing the five
123 types of LTE listed above. It is important to point out that, for the sake of our exercise, these five
124 datasets need not be considered in connection with the two original LTEs and may be taken to
125 represent five 'ideal' and independent LTEs. Readers interested in full-fledged analyses of the
126 original datasets are referred to the aforementioned references.

127 Table 2 summarises the contrasting characteristics of these datasets that will be
128 described in detail afterwards. For each dataset, some records are given as supplemental
129 materials, in order to show the exact structure of data tables.

130

131

[Table 2 about here]

132

133 *2.1 LTEs with monocultures or perennial crops (Dataset 1)*

134

135 This type of LTE does not involve rotations and the treatments under comparison consist of
136 different cropping practices (e.g. fertilisations). In terms of data analysis, these LTEs are similar
137 to experiments with perennial crops, as observations are taken all years on the same plot, for
138 each treatment level and replicate (Eckl and Piepho, 2015; Piepho and Eckl, 2014; Smith et al.,
139 2007).

140

141 *Dataset 1: Continuous wheat cropping.* Wheat is grown in continuous cropping from 1983 to
142 2012, with three fertilisation levels (150, 200 and 250 $kgNha^{-1}$), randomly assigned to three
143 plots in each of three blocks. In all, there are nine plots with yearly sampling, with a total of 270
144 wheat yield observations in 30 years.

145

146 *2.2. LTEs with different rotations of the same length and one test crop per rotation cycle (Dataset*
147 *2)*

148

149 If we have, e.g., two crops in a rotation (maize and wheat), both crops will be grown in different
150 plots in the same year and we will have two possible sequences in time (maize-wheat and
151 wheat-maize). If we consider only one of the two crops, the main difference with respect to the
152 previous situation is that the data obtained in two consecutive years for the same treatment and
153 block are independent, in the sense that they are obtained in different plots. Otherwise, data
154 obtained in a two-year interval (on different rotation cycles) on the same block are correlated,
155 as they originate from the same plot.

156

157 *Dataset 2: Comparing wheat yield for several biennial rotations.* Wheat is grown in five types of
158 two-year rotations, with either pea (*Pisum sativum* L.), grain sorghum (*Sorghum bicolor* (L.)
159 Moench), sugar beet (*Beta vulgaris* L. subsp. *saccharifera*), sunflower (*Helianthus annuus* L.) and
160 faba bean (*Vicia faba* L. subsp. *minor*). For each rotation, there are two possible sequences
161 (wheat in odd years and wheat in even years) and the ten combinations (five rotations by two
162 sequences) are completely randomised to ten plots per each of three blocks (Table 3). Therefore,
163 five wheat plots out of the available ten plots are used from each block and year, for a total of
164 450 observations, from 1983 to 2012.

165

166

[Table 3 about here]

167

168 *2.3 LTEs with a fixed rotation (one test crop per cycle) and different treatments (Dataset 3)*

169

170 This type of LTE is very similar to the previous one, though we will introduce a dataset with a
171 different experimental layout.

172

173 *Dataset 3: Comparing wheat yield in one biennial rotation with different tillage/fertilisation*
174 *systems.* Durum wheat (*Triticum durum* L.) is grown in a two-year rotation with a spring crop and
175 nine treatments, consisting of the factorial combination of three soil tillage methods (T:
176 conventional 40 cm deep ploughing; M: scarification at 25 cm; S: sod seeding with chemical
177 desiccation and chopping) and three N-fertilisation levels (0, 90 and 180 kg N ha⁻¹). The two
178 possible rotation sequences (wheat-spring crop and spring crop-wheat) are arranged in two
179 adjacent fields, which therefore host the two different crops of the rotation in the same year.
180 Within the two fields, there are two independent randomisations, each with two blocks. Tillage
181 levels are randomised to main-plots (1500 m²) and N levels are randomised to sub-plots
182 (500m²), according to a split-plot design with two replicates.

183 Also for this third dataset, the response variable is wheat yield; in terms of data analysis,
184 it is relevant to note that: (i) there are two experimental factors, laid out in a split-plot design;
185 (ii) the two sequences are accommodated in two fields (Figure 1).

186

187 [Figure 1 about here]

188

189 *2.4. LTE with a fixed rotation, different treatments and more than one phase per crop and cycle*
190 *(Dataset 4)*

191

192 The situation becomes more complex if we have, e.g., a three-year rotation with two wheat crops
193 (maize-wheat-wheat). In this case there are three possible sequences and, relating to wheat, we
194 have two distinct phases in each rotation cycle (phase difference: wheat after maize and wheat
195 after wheat).

196

197 *Dataset 4: burial/removal of crop residues on a three-years rotation.* Wheat is grown in a three-
198 year rotation maize-wheat-wheat, under two types of management of crop residues (burial and
199 removal), which are randomised to main plots, while the three possible rotation sequences are
200 randomised to subplots. This experiment has 18 plots (three sequences×two treatment
201 levels×three blocks) and, in every year, 12 of those are cropped with wheat and six with maize.

202 Also in this case, the response variable is wheat yield from 1983 to 2012, i.e. twelve
203 observations per year and 360 observations in total. Data obtained in the same plot in different
204 years belong to two different phases (wheat after maize and wheat after wheat; Table 4).

205

206

[Table 4 about here]

207

208 2.5. LTEs with several rotations of different lengths and different number of phases per crop and
209 rotation cycle (Dataset 5)

210

211 In some cases, it is necessary to compare several rotations with different characteristics (e.g., a
212 different duration and/or a different number of tests crops and /or a different number of phases
213 per crop), which may create a complex design with some degree of non-orthogonality.

214

215 *Dataset 5: comparing five maize-wheat rotations.*Wheat is grown in five maize (M) - wheat (W)
216 rotations of different lengths, i.e., M-W, M-W-W, M-W-W-W, M-W-W-W-W, M-W-W-W-W-W. For
217 all rotations, all phases are simultaneously present in each year, for a total of 20 plots (one for
218 each of the possible sequences, i.e. $2 + 3 + 4 + 5 + 6 = 20$) in each of three blocks. Considering
219 wheat yield as the response variable, we find that only 15 observations are obtained in each year
220 and block, for a total of 1350 records, from 1983 to 2012.

221 Experiments of this type represent a high degree of complexity. Indeed, in contrast to all
222 other examples, after 30 years there are plots with: (i) a different number of observations for the
223 same test crop; (ii) a different number of cycles (in some cases the last cycle is also incomplete);
224 (iii) a different number of phases for wheat.

225

226 3. Steps to model building

227

228 Every good analysis is based on a good model that is biologically relevant and based on realistic
229 assumptions. We will start from the general method and notation proposed by Piepho et al.
230 (2003) that is naturally related to computer implementation by statistical packages for linear
231 models. A similar model building procedure has been devised in Brien and Demetrio (2009),
232 which should yield comparable results for the examples under investigation. With respect to
233 these approaches, we will propose some slight changes, specific for LTEs.

234 In general, ANOVA-like models are based on classification variables, commonly known as
235 factors. In order to build correct models, we need to distinguish three types of factors: (i)
236 *treatment factors*, which are randomly allocated to randomisation units (e.g. rotations,
237 fertilisations, management of crop residues); (ii) *block factors*, which group the experimental
238 units according to some 'innate' (not randomly allocated) physical criterion (e.g. by position),
239 such as the blocks, the locations, the main-plots, the sub-plots and so on. Block factors may
240 represent the randomisation units, to which treatments are randomly allocated; (iii) *repeated*
241 *factors*, which relate to time and thus cannot be randomised (e.g. years, cycles ...).

242 All the above factors are given in capital letters and are combined by using the following
243 operators: (i) the 'dot' operator denotes crossed effects (e.g. $A \cdot B$ means that A and B are
244 crossed factors); (ii) the 'nesting' operator denotes nested effects (e.g. A/B means that B is
245 nested within A and it is equal to $A + A \cdot B$); (iii) the 'crossing' operator denotes the full factorial
246 model for two terms ($A \times B = A + B + A \cdot B$). This brief information may be sufficient for our
247 aims, but other operators and syntactic rules are given in the paper by Piepho et al. (2003) and

248 in a follow-up paper about repeated measures (Piepho et al., 2004), which might be of interest to
249 the readers.

250 The steps to model building may be summarised as follows:

251

- 252 1. Select the repeated factor.
- 253 2. Consider one fixed level of the repeated factor and build a treatment model for the
254 randomized treatment factors.
- 255 3. Consider one fixed level of the repeated factor and build a block model for block factors.
- 256 4. Check whether randomised treatment factors might interact with block effects: if such an
257 interaction is to be expected it should be added to the model.
- 258 5. Include the unrandomised repeated factor into the model.
- 259 6. Combine treatment model and repeated factor model, by crossing or nesting as appropriate.
- 260 7. Consider which effects in the block model reference randomisation units, i.e. those units
261 which receive the levels of a factor or factor combination by a randomisation process.
262 Assign to the corresponding terms a separate random effect, as explicitly recommended in
263 Piepho et al. (2004).
- 264 8. Excluding the terms for randomisation units, nest the repeated factor in all the other terms
265 in the block model.
- 266 9. Combine random effects for randomisation units with the repeated factor, by using the dot
267 operator, in order to derive the correct error terms to accommodate serial correlation
268 structures.

269

270 The key idea for the above approach is that for a properly designed experiment, valid
271 analyses should be possible for the data at each single level of the repeated factor. Such a basic
272 requirement should never be taken for granted, but it should be carefully checked before the
273 beginning of model building process (see later for Dataset 3).

274 Dealing with rotations, steps 1 and 2 are crucial when we are only interested in one of
275 the crops (phases) in the rotation. We will show this with an example.

276

277 *3.1. Analysis of Dataset 2: choosing the repeated factor and treatment model*

278

279 With Dataset 2 (LTE with wheat in five different two-year rotations; Tab. 2) we are looking only
280 at one phase in the rotation (in this case wheat) and we should note that observations are
281 repeated every second year on the same plot (Tab. 3), according to the sequence they belong to.
282 In other words, observations are repeated on each rotation cycle (two years) in the same plot,
283 while there is neither a within-cycle repetition nor a within-cycle phase difference: we have only
284 one observation per plot per cycle. Therefore, it is natural to take the rotation cycle as the
285 repeated factor.

286 As the next step, we should look at what happens in one fixed level of a two-year cycle:
287 what did we randomize to the ten plots? It is clear that, considering only wheat, we randomised
288 each combination of rotation and positioning in the sequence (i.e. wheat as the first crop of the
289 sequence and wheat as the second crop of the sequence; see Table 1). Therefore, sequence and
290 rotation should be included in the model as the treatment factors.

291 This approach is commonly suggested in literature (see Yates, 1954) and it is convenient,
292 mainly because the resulting model is orthogonal and may be fitted by ordinary least squares.
293 Indeed, for Dataset 2 (and similar experiments), there is only one observation for each block,
294 treatment, cycle, sequence and no missing data (in our case: 3 blocks×5 rotations×2 sequences ×
295 15 cycles = 450 observations). The phase should not enter into this model, as we are looking
296 only at one of the two crops (only one phase).

297 However, the drawback is that such an approach cannot be immediately extended to the
298 other more complex examples (e.g. rotations with different lengths and/or with a different
299 number of test-crops). Furthermore, it partitions the year factor into three effects, i.e. 'cycles',
300 'sequences' and 'cycle×sequences', which might make modelling any 'fertility' trends over time
301 less obvious. In this respect, we should note that possible differences between sequences for a
302 given cycle (wheat as the first crop of the sequence and wheat as the second crop of the
303 sequence, i.e. wheat in even years and wheat in odd years) do not carry any meaning that helps
304 understand the behaviour of rotations.

305 Therefore, it may be more convenient to take the year as the repeated factor, considering
306 that, for the case of Dataset 2, this effect is totally confounded with the factorial combination of
307 'cycle' and 'sequence' (15 cycles×2 sequences = 30 years). This gives us a good common
308 modelling platform for all datasets, although it may be argued that the models are no longer
309 orthogonal, as not all plots produce data in all years. It should be recognised, however, that the
310 lack of orthogonality can easily be accommodated within mixed models. If we take the year as
311 the repeated factor, it is more convenient to embrace the general view that in one level of year
312 we randomise the rotations and the phases for each rotation to the plots. Therefore, in the case
313 of Dataset 2, where we are interested only in one of the two phases, the treatment model will
314 only contain the rotation term (see later).

315 The situation is totally different if we look at both the phases of the rotation (e.g. wheat
316 and sunflower): in this case, we have a phase difference within each cycle and, considering one
317 level of the repeated factor year, the treatment model should contain both the rotation and the
318 phase, together with their interaction. When we introduce the year (steps 5 and 6 above), we
319 also introduce the interactions 'year×rotation', 'year×phase' and 'year×rotation×phase', which
320 are all meaningful when studying the behaviour of rotations. We will no longer pursue this
321 aspect with Dataset 2, as we decided to consider only wheat as the test-crop. However, we will
322 again consider the phase difference with Datasets 4 and 5.

323 On the above basis, we will now define the models for all datasets in Table 2, considering
324 the year as the repeated factor. We will also partition the fixed model and the random model, as
325 commonly done in several statistical packages (SAS, ASReml-R...). At first, the random model will
326 include all random terms for randomisation units (terms at steps 7 and 9), while the fixed model
327 will include all the other terms. Later on, we will discuss the possible extensions/changes to this
328 basic approach.

329

330 **4. Building statistical models for all datasets**

331

332 *4.1. Analysis of Dataset 1*

333

334 If we consider Dataset 1 (LTE on continuous wheat at three N-fertilisation levels; Tab. 2), the
335 repeated factor is the year (YEAR). In one year, the treatment factor is nitrogen fertilisation (N)
336 and there are two block factors, i.e. the blocks (BLOCK) and the plots within each block (PLOT).
337 Therefore, the block model is BLOCK + BLOCK · PLOT.

338 We now introduce the repeated factor YEAR and combine it with the treatment model,
339 by including $N \times YEAR = N + YEAR + N \cdot YEAR$. The term BLOCK · PLOT references the
340 randomisation units and receives a random effect. As the year might interact with the block, we
341 add the term BLOCK · YEAR. We also combine the year with the random effect for plots (BLOCK ·
342 PLOT · YEAR), although this residual term is usually automatically fitted and does not need to be
343 explicitly coded when implementing the model. The final model is (the operator ~ means 'is
344 modelled as'):

345

346 $YIELD \sim N + BLOCK + YEAR + N \cdot YEAR + BLOCK \cdot YEAR$

347 $RANDOM: BLOCK \cdot PLOT + BLOCK \cdot PLOT \cdot YEAR$ (Model 1)

348

349 We have already mentioned that one of the problems with LTEs is that the observations
350 taken on the same plot are not independent. Model 1 is able to account for this, owing to the
351 presence of the two random effects BLOCK · PLOT (often called 'plot' error, with variance σ_B^2)
352 and BLOCK · PLOT · YEAR (residual error or within-plot error, with variance σ^2). When the
353 variance for BLOCK · PLOT is higher than 0, it is implied that observations in the same plot 'co-
354 vary', i.e. show a positive covariance equal to σ_B^2 . The variance for one observation (same plot
355 and same year) is equal to the sum $\sigma_B^2 + \sigma^2$ and, therefore, the correlation among observations
356 in the same plot is equal to the ratio $\rho = \sigma_B^2 / (\sigma_B^2 + \sigma^2)$ (intra-class correlation). This correlation
357 is equal for all pairs of observations in each plot, regardless of their distance in time.

358 This would be similar to a split-plot design (split-plot in time) with the important
359 difference that, in this case, the sub-plot factor (year) is not randomised. The corresponding
360 serial correlation structure is known as compound symmetry (CS). It is worth noticing that the
361 same model may be fitted in an alternative way, i.e. by dropping the BLOCK · PLOT random
362 effects and using the residual term BLOCK · PLOT · YEAR to accommodate the CS structure into
363 the model. This may be done very intuitively e.g. in SAS, by using the statement:

364

```
365 repeated year/subject = block*plot type=CS;
```

366

367 This is very useful when a simple CS correlation structure is not satisfactory, because the
368 observations in the same plot are not equally correlated (e.g. observations close in time are
369 more correlated than those distant in time). In this case, we can implement alternative
370 correlation structures by using the residual term BLOCK · PLOT · YEAR, e.g. by appropriately
371 changing the specification of the 'type=' option in the above SAS statement.

372

373 *4.2. Analysis of Dataset 2 (continued)*

374

375 Considering Dataset 2 (LTE with wheat in five different two-year rotations; Tab. 2), we consider
376 the YEAR as the repeated factor. In one year the treatment model is composed only by the
377 rotation (ROT), which is randomised to plots (PLOT), within blocks. The block model for one
378 year is $\text{BLOCK/PLOT} = \text{BLOCK} + \text{BLOCK} \cdot \text{PLOT}$ (the use of the nesting operator gives a more
379 compact definition). We can combine the treatment model with the repeated factor ($\text{ROT} \cdot$
380 YEAR) and add the term $\text{BLOCK} \cdot \text{YEAR}$. The final model is:

381

382 $\text{YIELD} \sim \text{ROT} + \text{BLOCK} + \text{YEAR} + \text{ROT} \cdot \text{YEAR} + \text{BLOCK} \cdot \text{YEAR}$

383 $\text{RANDOM: BLOCK} \cdot \text{PLOT} + \text{BLOCK} \cdot \text{PLOT} \cdot \text{YEAR}$ (Model 2)

384

385 Apart from random effects for randomisation units, this model is similar to the one used
386 for multi-environment genotype experiments and, represents a convenient and clear platform
387 for the analyses of LTE data. We remind the reader that the residual term ($\text{BLOCK} \cdot \text{PLOT} \cdot$
388 YEAR) does not need to be explicitly coded when implementing the model, as it is automatically
389 fitted. This basic model again implies the compound symmetry structure for serial correlation of
390 plot errors; more refined modeling is possible by imposing alternative serial correlation
391 structures.

392

393

394 *4.3. Analysis of Dataset 3*

395

396 Before proceeding to model building with Dataset 3 (LTE for wheat in one biennial rotation with
397 different tillage/fertilisation systems; Tab. 2 and Fig. 1), we need to discuss whether valid
398 analyses are possible at each single level of the repeated factor. Indeed, this is clearly true if we
399 take the year as the repeated factor and consider only one of the two crops in the rotation
400 (wheat, in this case). However, if we intended to consider both crops and compare e.g. their
401 yields, the crop effect would be confounded with the field effect within a single year and,
402 therefore, valid within-year analyses would not be possible. In this case, we should resort to
403 taking the rotation cycle as the repeated factor.

404 Dealing only with wheat, we can therefore take the YEAR as the repeated factor and
405 consider that, in one year, the randomised treatment factors are tillage (T) and nitrogen
406 fertilisation (N) and the treatment model is $\text{T} + \text{N} + \text{T} \cdot \text{N}$.

407 The block factors are the FIELDS, the BLOCKS within fields, the main plots (MAIN) within
408 blocks and the subplots (SUB) within main plots. The block model (for one year) is
409 $\text{FIELD/BLOCK/MAIN/SUB} = \text{FIELD} + \text{FIELD} \cdot \text{BLOCK} + \text{FIELD} \cdot \text{BLOCK} \cdot \text{MAIN} + \text{FIELD} \cdot \text{BLOCK} \cdot$
410 $\text{MAIN} \cdot \text{SUB}$.

411 The treatment and repeated model can be combined as: $(T + N + T \cdot N) \times YEAR = T + N + T$
412 $\cdot N + YEAR + T \cdot YEAR + N \cdot YEAR + T \cdot N \cdot YEAR$.

413 At this stage, the FIELD main effect needs to be removed, as it is totally confounded with
414 the years in this specific example. We assign a random effect to the other randomisation units,
415 i.e. FIELD · BLOCK, FIELD · BLOCK · MAIN and FIELD · BLOCK · MAIN · SUB and combine these
416 random terms with the repeated factor YEAR, by using the dot operator, which leads us to:

417

418 $YIELD \sim T + N + T \cdot N + YEAR + T \cdot YEAR + N \cdot YEAR + T \cdot N \cdot YEAR$

419 RANDOM: FIELD · BLOCK + FIELD · BLOCK · MAIN + FIELD · BLOCK · MAIN · SUB + FIELD ·
420 BLOCK · YEAR + FIELD · BLOCK · MAIN · YEAR + FIELD · BLOCK · MAIN · SUB · YEAR

421 (Model 3)

422

423 As usual, the last term (residual) does not need to be explicitly included when
424 implementing the model, but it can be used, together with the two previous ones (FIELD ·
425 BLOCK · YEAR + FIELD · BLOCK · MAIN · YEAR) to accommodate possible serial correlation
426 structures into the model, by allowing year-specificity of all design effects and the residuals.

427

428 *4.4. Analysis of Dataset 4*

429

430 We can proceed in the same fashion as above with Dataset 4 (LTE with burial and removal of
431 crop residues; Tab. 2), considering that the year is the repeated factor and, in one year, the
432 treatment factors are the management of soil residues (RES, that is randomly allocated to main-
433 plots) and the phases (P; randomly allocated to subplots); the treatment model is indeed: $RES \times$
434 P .

435 In one year, the block model is: $BLOCK/MAIN/SUB = BLOCK + BLOCK \cdot MAIN + BLOCK \cdot$
436 $MAIN \cdot SUB$. Introducing the YEAR as repeated factor, we can combine the treatment model with
437 the repeated model as: $RES + P + RES \cdot P + YEAR + RES \cdot YEAR + P \cdot YEAR + RES \cdot P \cdot YEAR$.

438 The terms BLOCK · MAIN and BLOCK · MAIN · SUB reference randomisation units and
439 should receive random effects. The blocks may interact with the years (BLOCK · YEAR), while
440 the random effects for randomisation units can be made year-specific by adding BLOCK · MAIN ·
441 YEAR and the residual term BLOCK · MAIN · SUB · YEAR. The final model is:

442

443 $YIELD \sim RES + P + RES \cdot P + YEAR + RES \cdot YEAR + P \cdot YEAR + RES \cdot P \cdot YEAR + BLOCK + BLOCK \cdot$
444 $YEAR$

445 RANDOM: BLOCK · MAIN + BLOCK · MAIN · SUB + BLOCK · MAIN · YEAR + BLOCK · MAIN · SUB ·
446 $YEAR$ (Model 4)

447

448

449 *4.5. Analysis of Dataset 5*

450

451 This further case study relates to Dataset 5 (comparison between five maize-wheat rotations of
452 different lengths; Tab. 2), where the repeated factor is again the year. In one year, the treatment
453 factors are the rotation system (ROT) and the rotation phase (P), which are randomly allocated
454 to plots. As there is a different number of phases for each rotation, we nest the phase within the
455 rotation, leading to the following treatment model: $ROT/P = ROT + P \cdot ROT$.

456 In one year, the block model is $BLOCK/PLOT = BLOCK + BLOCK \cdot PLOT$. The repeated
457 factor is included and combined with the treatment model, by introducing $YEAR + ROT \cdot YEAR +$
458 $ROT \cdot P \cdot YEAR$.

459 The term $BLOCK \cdot PLOT$ references randomisation units and needs to receive a random
460 effect, while the term $YEAR \cdot BLOCK$ can be added to the model, together with the residual term
461 $BLOCK \cdot PLOT \cdot YEAR$. The final model is:

462

463 $YIELD \sim ROT + P \cdot ROT + YEAR + ROT \cdot YEAR + ROT \cdot P \cdot YEAR$

464 $RANDOM: BLOCK \cdot PLOT + BLOCK \cdot PLOT \cdot YEAR$ (Model 5)

465

466

467 5. Model implementation

468

469 Models 1 to 5 are fairly similar and they are very closely related to those used for multi-
470 environment experiments. However, apart from Dataset 1, there is always a certain degree of
471 imbalance, as plots do not produce data every year. Therefore, these models are not generally
472 amenable to ordinary least squares estimation and traditional ANOVA and they should be
473 implemented by using the mixed model framework, based on REstricted Maximum Likelihood
474 (REML) or Maximum Likelihood (ML) estimation (see also Payne, 2015).

475 A second aspect that needs to be taken into account is that, in all cases, the above model
476 formulations induce the aforementioned compound symmetry correlation structure ('split-plot
477 in time'), which may not be appropriate in all circumstances. We will discuss alternative
478 correlation structures shortly.

479 Last, but not least, we have already mentioned that, in the above models, only
480 randomisation units have been given a random effect, while all the other effects have been
481 regarded as fixed. Obviously, depending on the aims of the analyses, it might be convenient and
482 appropriate to regard the year factor as random. In this case, all interactions containing this
483 factor are also random. Examples will be given later on.

484 Should we be happy with the above structure (random effects only for randomisation
485 units and compound symmetry), the implementation of models 1 to 5 is straightforward in most
486 statistical packages (R, SAS, GENSTAT . . .). Examples of implementation in R are given as
487 supplemental material (Code S1). Some useful SAS code can be found in Piepho (1999), while
488 GENSTAT, R and SAS code is given in Payne (2015).

489

490 6. Model checking and extensions

491

492 Though the above models are rather easy to implement by using advanced statistical software,
493 we should never forget that they make a number of basic assumptions that should be carefully
494 checked before going ahead with the analyses. In particular, it is assumed that:

495

- 496 1. within group errors are normal and homoscedastic;
- 497 2. random effects for randomisation units are normal and independent from one another (if
498 more than one random effect exists);
- 499 3. subjects in the same group (for example years within plots) are equally correlated, while
500 subjects in different groups are independent.

501

502 This latter assumption (split-plot in time), in particular, seems to be rather questionable
503 and more advanced correlation patterns have been advocated, e.g. autoregressive and
504 unstructured (Richter and Kroschewski, 2006; Singh and Jones, 2002). In brief, the
505 autoregressive approach assumes that observations in the same plot are not equally correlated,
506 but their correlation depends on the time span between them. On the other hand, the
507 unstructured approach is the least restrictive, as it does not assume any particular pattern for
508 within-plot correlation. In principle, this latter approach is rather appealing, though it requires a
509 large number of parameters to be estimated, which may not be supported by the data at hand
510 and may give computational problems.

511 The suggested approach to the selection of variance-covariance structures is based on
512 fitting all possible models and making an 'a posteriori' selection, based on a likelihood ratio test
513 or, preferably, on those statistics which put a penalty on 'complexity', such as the Akaike
514 Information Criterion (AIC: Akaike, 1974), which seeks to find the approximating model
515 (Burnham and Anderson, 2002). Additionally, we suggest a graphical exploration of residuals;
516 indeed, this procedure is very common with traditional linear models and gives a better 'feel' for
517 the data.

518 Accommodating variance and correlation structures in a mixed model is usually done by
519 exploiting the residual term, e.g. BLOCK · PLOT · YEAR and using the YEAR as repeated effect
520 and the non-repeated error term BLOCK · PLOT as subject effect (Piepho et al., 2004). It is
521 important, however, to highlight that, with split-plot designs or in presence of other
522 randomisation structures apart from plots, it is important to allow for year-specificity of all
523 design effects, not only the residual error term. For example, we might use the main plot random
524 effect (e.g. BLOCK · MAIN · YEAR in Dataset 4) as above, to impose a year-specific correlation
525 structure to main-plots.

526 The implementation of variance and correlation structures is straightforward with SAS.
527 In the case of R, the implementation may be done by using the commercial package ASReml-R
528 (Butler et al., 2009), while it becomes tricky if a freeware solution is envisaged for designs with
529 several randomisation units. Indeed, it should be considered that the most recent function for
530 mixed models in R, i.e. *lmer()* from the package *lme4* (Bates et al., 2015) does not yet support
531 non-standard variance-covariance structures and it may be therefore necessary to use the
532 function *lme()* from the package *nlme* (Pinheiro and Bates, 2000). However, this latter function

573 In this case, the UN approach gave convergence problems and it was clearly overly
574 complex, with respect to the data at hand (Tab. 5). For the other approaches, however, the log-
575 likelihood always increased when more complex variance-covariance structures were added to
576 the model. Among these, the CSH approach appeared to give the best balance among statistical
577 accuracy and simplicity, based on the AIC statistic (Tab. 5).

578 The code to reproduce the analyses in R is given as supplemental material (Code S3)

579

580 [Table 5 about here]

581

582 7. Testing for fixed effects

583

584 Testing for fixed effects has always been regarded as a fundamental part of data analyses,
585 traditionally based on ANOVA tables and F-tests built upon expected mean squares and 'error
586 strata'. Switching to mixed models, a different approach is required. With ML estimation,
587 likelihood ratio tests (LRTs) would be a straightforward procedure, although the null
588 distribution for these tests is only approximately chi-squared and such an approximation may
589 be rather poor with small sample sizes, leading to 'anti-conservative' p-values. Better results
590 may be achieved by using a parametric bootstrap approach that is shown later.

591 However, REML estimation is usually preferred with mixed models, as variance
592 component estimates are known to have better properties (Searle et al., 1992). In this case,
593 Wald-type F tests can be used as in the usual traditional ANOVA approach, though it is necessary
594 to consider that, when data are unbalanced (which is almost always the case with our models),
595 the denominator degrees of freedom are not easily obtained and some correcting method is
596 mandatory (e.g. the Kenward-Roger approach: Kenward and Roger, 2009; Kenward and Roger,
597 1997).

598

599 7.1. Analysis of Dataset 2 (continued): testing for fixed effects

600

601 For this case study, we were interested in testing the significance of the 'treatment×year'
602 interaction. We fitted Model 2 to Dataset 2 (Tab. 2), regarding all effects as fixed and considering
603 (after preliminary analyses) a simple within-plot compound symmetry (CS) correlation
604 structure, without heteroscedasticity of years. Working with R and the function *lmer()* in the *lme4*
605 package, we performed a conditional F test by using the function *anova()*, which returned a F-
606 value of 1.74, with 116 and 217 unadjusted degrees of freedom ($p = 0.0002$).

607 A Kenward-Roger approximation for degrees of freedom is more reliable and requires
608 two additional packages in R, i.e. *lmerTest* (Kuznetsova et al., 2015) and *pbkrtest* (Halekoh and
609 Højsgaard, 2014). In this case, we obtained a number of denominator degrees of freedom equal
610 to 182.56 ($p = 0.0004$). (See additional material Code S4).

611 The above solution is not general, because the *lmer()* function does not support other
612 variance-covariance structures, apart from CS. A more general solution is to use the commercial

613 package *asreml-R*, which works with all the main variance-covariance structures; an example is
614 given as additional material (Code S4).

615 Alternatively, if a freeware solution is needed, it is necessary to consider that Wald-type
616 F tests with Kenward-Roger adjusted degrees of freedom are not available for the function
617 *lme()* in the *nlme* package. Therefore, the significance of the 'rotation × year' interaction should
618 be tested by using a LRT, which consists of comparing a full model with a reduced model
619 (dropping the 'ROT · YEAR' interaction term) in a fairly similar fashion as with the 'extra-sum-
620 of-squares' test in traditional linear/nonlinear models.

621 In detail, both models (full and reduced) were fitted by full maximum likelihood (ML;
622 REML estimation is not valid, in this particular case) and twice the log-likelihood (LOGLIK) for
623 the full model was compared with twice the log-likelihood of the reduced model:

624

$$625 \quad LRT = 2[LOGLIK_{full} - LOGLIK_{reduced}]$$

626

627 For large samples, the LRT statistic is assumed to have a chi-squared distribution with a
628 number of degrees of freedom equal to the difference in the number of parameters for the two
629 models. In this case we obtained a value of 274.99, with 116 degrees of freedom (the product of
630 the number of years - 1 by the number of rotations - 1, i.e. $29 \times 4 = 116$), which corresponds to a
631 very low p-level ($p = 6.2 \times 10^{-15}$).

632 This very small value leaves us quite confident about rejecting the null hypothesis.
633 However, a parametric bootstrap approach may be used whenever one is in doubt. In this case,
634 the procedure is as follows (Faraway, 2006):

635

- 636 1. simulate a new dataset under the reduced model, using the fitted parameter estimates and
637 assuming normality for the errors and random effects;
- 638 2. fit to this dataset both the full and the reduced model;
- 639 3. compute the LRT statistic;
- 640 4. repeat steps 1 to 3 many times (e.g. 1000);
- 641 5. examine the distribution of the bootstrapped LRTs and compute the proportion of those
642 exceeding 274.99 (empirical p-value)

643

644 An example of calculation in R is given in the supplementary material (Code S4), showing
645 an empirical p-value of 0.000145, still significant, but much higher than the observed one. The
646 bootstrap approach is reliable, but it may be tedious and very computer-intensive.

647

648 **8. 'Fertility' trends**

649

650 In contrast to other multi-environment experiments, LTEs are characterised by the same
651 cropping practices, which are repeated on the same plots for a long period of time. This may
652 bring to a significant 'fertility' trend over time (Berzsenyi et al., 2000) that may be reinforced by
653 temporal changes due to the adoption of innovative cropping practices (e.g. new varieties,
654 harvesting methods, weed control methods), as necessary for experiments lasting for a long
655 period of time (McRae and Ryan, 1996).

656 Long-term 'fertility' trends are not of random nature and relate to the sustainability of
657 each cropping system under investigation: a system with an upward trend should be more
658 sustainable than another one with a downward trend in the long run.

659 The most straightforward way to model fertility trends is by fitting a simple linear
660 regression of yield over time, as shown by Guertal et al. (1994) or Piepho et al. (2014). Nonlinear
661 relationships might also be used, in the form of polynomials (Payne, 2015) or other intrinsically
662 non-linear functions (Yang, 2014), whenever such a complexity is supported by the data at hand.

663

664 *8.1. Analysis of Dataset 4 (continued): fitting a 'fertility' trend*

665

666 For this case study we used Dataset 4 (Tab. 2), relating to the same rotation (M-W-W) with or
667 without burial of crop residues. It might be expected that the continuous incorporation of soil
668 residues creates a build-up of soil fertility over time.

669 Model 4 was used, together with a simple compound symmetry variance-covariance
670 structure. The fixed part of the model was modified by expressing wheat yield as a linear
671 function of a centred continuous covariate for the calendar year from the beginning of
672 experiment (TIME_C), with treatment-dependent intercept and slope (i.e. four different average
673 yield levels, depending on the combined levels of P and RES and two different slopes, depending
674 on RES). All the other effects (YEAR, RES · YEAR, P · YEAR, RES · P · YEAR and all effects
675 containing the BLOCK factor) were included as random, to represent the deviations from the
676 regression lines. Therefore, the model was as follows:

677

678 $\text{YIELD} \sim \text{RES} + \text{P} + \text{RES} \cdot \text{P} + \text{RES}/\text{TIME_C}$

679 $\text{RANDOM: YEAR} + \text{RES} \cdot \text{YEAR} + \text{P} \cdot \text{YEAR} + \text{RES} \cdot \text{P} \cdot \text{YEAR} + \text{BLOCK} + \text{BLOCK} \cdot \text{YEAR} + \text{BLOCK} \cdot$
680 $\text{MAIN} + \text{BLOCK} \cdot \text{MAIN} \cdot \text{SUB} + \text{BLOCK} \cdot \text{MAIN} \cdot \text{YEAR} + \text{BLOCK} \cdot \text{MAIN} \cdot \text{SUB} \cdot \text{YEAR}$
681 $(\text{Model } 6)$

682

683 It is useful to note that the above model partitions the year effect in two components: the
684 fixed part relating to the fertility trend over time and the random remaining part, modelling the
685 year-by-year variation around the trend line, as shown by Loughin et al. (2007).

686 The fitted fertility trends are reported in Figure 4, together with the observed yearly
687 mean yields for the two cropping systems (with and without burial of crop residues) and phases
688 (wheat after maize and wheat after wheat). Within managements, wheat after wheat yielded on
689 average less than wheat after maize. The system was sustainable regardless of the
690 burial/removal of crop residues, thanks to an upward yield trend over time, though, the slope
691 was higher with burial (0.043; SE = 0.021) than with removal (0.031; SE = 0.021). There is,

692 however, a remarkable random variability around these trends, as expected with this type of
693 experiments (Laidig et al., 2014).

694

695 [Figure 4 about here]

696

697 We should point out that we use the term 'fertility' in its widest sense and the above
698 model cannot in general separate the possible reasons for such a trend, which may relate to
699 either possible long-term effects of agronomic practices, or technical innovation, or climate
700 changes (Piepho et al., 2014). The R code to reproduce the analyses is given as supplemental
701 material (Code S5)

702

703 **9. Overall productivity and yield stability of cropping systems**

704

705 Often, there is an interest in assessing long-term average yield levels for the cropping systems
706 under study (broad inference space), together with their ability to minimise random fluctuations
707 relating to the year-to-year variability in weather conditions and other biotic/abiotic factors
708 (e.g. pests). This latter ability has been related to the concept of stability and, for LTEs, it would
709 seem particularly appropriate to assess the 'dynamic' (Type II) stability, i.e. the ability of
710 cropping systems to follow the mean environmental response of all cropping systems in the
711 experiment (Piepho, 1998).

712 Within this frame, it would seem totally appropriate to regard the 'treatment' effect as
713 fixed, together with the possible fertility trend across years, while the remaining part of the year
714 effect and the year-by-treatments interactions should be regarded as random. In this case, the
715 long duration of LTEs makes the estimation of variance components rather reliable.

716 Piepho (1999) has shown that an appropriate specification of the variance-covariance
717 matrix for mixed models may be used to obtain a reliable description of stability. For example,
718 the variance component for the random 'treatment×year' interaction might be assumed to
719 depend on the treatment levels, which would measure their long-term 'stability variance'
720 (Shukla, 1972). We will show this with another example.

721

722 *9.1. Analysis of Dataset 3 (continued): a stability variance model*

723 For this case study we used Dataset 3, where the factorial combination of three tillage methods
724 and three N-fertilisation levels was evaluated on a 19-year LTE, based on a biennial rotation
725 'winter wheat/spring crop' (Tab. 2). Winter wheat yields were analysed to assess the long-term
726 average of the nine cropping systems, together with their estimates of Shukla's stability
727 variance.

728 Preliminary analyses showed that this dataset is characterised by a strong year-to-year
729 heterogeneity of variances, while the existence of within-plot serial correlation is not supported
730 by the data. Therefore, we fitted Model 3, incorporating a fertility trend (as in the previous case
731 study) and allowed for 9 different variance components for the random 'treatment×year'
732 interaction (where treatment is the factorial combination of tillage and N-fertilisation levels).

733

734

[Table 6 about here]

735

736 Results are reported in Table 6 and clearly depict the long-term average behaviour of
737 cropping systems, in terms of yield level, stability and sustainability. The smallest stability
738 variance was observed for wheat grain yield under sod seeding with no N fertilization and under
739 the other two tillage systems with 90 kg N ha⁻¹. Moreover, both the minimum tillage and sod
740 seeding with the highest N fertilization seemed to lead to an upward fertility trend, which makes
741 these cropping systems more sustainable than the other treatments under comparison.

742

The R code to reproduce the analyses is reported as supplemental material (Code S6)

743

744 **10. Concluding remarks**

745

746 In order to produce reliable results, the process of data analysis for LTEs with cropping systems
747 requires some effort and careful thinking, especially for the most complex designs, characterised
748 by rotations with different durations and a different number of test-crops per rotation cycle.
749 Complex LTEs have been very seldom considered in literature, which motivated the example-
750 based approach taken in this manuscript. We proposed several case-studies based on five
751 exemplary datasets, covering the main types of LTEs, with increasing levels of complexity.
752 Obviously, we could not cover the whole array of techniques that might be successfully used for
753 the analysis of LTEs; for example, further research will be necessary to elucidate whether spatial
754 models might be useful to account for possible spatial trends. For example, spatial correlation
755 between plots errors may be combined with temporal correlation among plot errors in a single
756 model, and there are many options for doing this.

757

Based on the analyses of our datasets, we would like to highlight the following issues.

758

- 759 1. With experiments involving rotations, where all crop phases are grown in every year and
760 within-year replicates are available, statistical models for data analysis can be reliably built
761 using the procedure outlined in Piepho et al. (2004), providing that correct analyses can be
762 performed separately for each year.
- 763 2. In this paper, we propose that the year is included in the model as the repeated factor, by
764 which all the main types of LTEs are put on an equal footing, together with multi-
765 environment (genotype) experiments. However, particularly when we are interested only in
766 a subset of the crop phases for each rotation, the resulting models are rarely orthogonal, as
767 observations for the test crop of interest are not taken all years on all plots. For this reason,
768 traditional analysis of variance is not generally appropriate, while mixed models and REML
769 estimation should be preferred.
- 770 3. One of the main problems with LTEs is that observations are taken repeatedly on the same
771 plot and, therefore, they are correlated. As the first step, we can account for this by
772 including in the model random effects for randomisation units (e.g. plots). Such an approach
773 assumes that within-plot correlation is independent on the time elapsed between two
774 observations (the model is known as compound symmetry, also known as 'split-plot in

775 time'), which is rarely sufficient to appropriately describe the real pattern of variances-
776 covariances of observations. Therefore, it is fundamental that this aspect is considered
777 during data analysis.

778 4. For all our examples, we found that variances were not constant across the repeated factor,
779 which we accounted for by allowing a different variance for each year (compound
780 symmetry with variance heterogeneity of years). This variance-covariance structure proved
781 to be the best compromise among simplicity and statistical accuracy, though this should not
782 be taken as a general rule. Indeed, it is always appropriate to consider the need of other
783 more complex correlation structures.

784 5. For those who are accustomed to traditional ANOVA, testing for fixed effects with mixed
785 models and unbalanced data may require some further attention. Conditional F tests with
786 Kenward-Roger adjusted denominator degrees of freedom may represent a good choice
787 and several simulation studies have shown that they work well in practice (Richter et al.,
788 2015; Spilke et al., 2005). If adjusted denominator degrees of freedom are not available,
789 parametric bootstrap procedures can be used instead of simple likelihood ratio tests, as
790 these latter tend to be anti-conservative and give unreliable p-levels.

791 6. Fertility trends (linear/non-linear) during LTEs may be investigated to measure the long-
792 term sustainability of cropping systems. These trend are always of a fixed nature and may
793 serve the purpose of partitioning the 'year' effect into its random and fixed components, as
794 recommended by Loughin et al. (2007).

795 7. Several statistical methods devised for multi-environment experiments may be successfully
796 used for the analyses of LTEs. For example, in this paper we used Shukla's stability variance
797 model to assess the long-term stability of cropping systems.

798

799 **Bibliography**

800

801 Akaike, H., 1974. A new look at the statistical model identification. *IEEE Transactions on*
802 *Automatic Control* 19, 716–723.

803 Bates, D., Maechler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using
804 *lme4*. *Journal of Statistical Software* 67, 1, 1-48.

805 Berzsenyi, Z., Gyorffy, B., Lap, D., 2000. Effect of crop rotation and fertilisation on maize and
806 wheat yields and yield stability in a long-term experiment. *European Journal of*
807 *Agronomy* 13, 225–244.

808 Bonciarelli, U., Onofri, A., Benincasa P., Farneselli, M., Guiducci, M., Pannacci, E., Tosti, G., Tei, F.,
809 2015. Long-term evaluation of productivity, stability and sustainability for cropping
810 systems in Mediterranean rainfed conditions. *European Journal of Agronomy*, this issue.

811 Brien, C., Demetrio, C., 2009. Formulating mixed models for experiments, including longitudinal
812 experiments. *Journal of Agricultural, Biological and Environmental Statistics* 14, 253–
813 280.

814 Burnham, K.P., Anderson, D.R. 2002. Model selection and multimodel inference. A practical
815 information-theoretic approach. Second edition. Springer, New York.

- 816 Butler, D., Cullis, B., Gilmour, A., Gogel, B., 2009. ASReml-r reference manual: mixed models for S
817 language environments. Queensland Government, Department of Primary Industries and
818 Fisheries, Brisbane, Australia.
- 819 Cochran, W.G., 1939. Long-term agricultural experiments. Supplement to the Journal of the Royal
820 Statistical Society 6, 104–148.
- 821 Eckl, T., Piepho, H.-P., 2015. Analysis of series of variety trials with perennial grasses for
822 subdivided target regions. *Crop Science* 55, 597–609.
- 823 Faraway, J., 2006. Extending the linear model with R. Generalized linear, mixed effects and
824 nonparametric regression models. Chapman & Hall, Boca Raton, Florida.
- 825 Frye, W., Thomas, G., 1991. Management of long-term experiments. *Agronomy Journal* 83, 38–
826 44.
- 827 Gałeczki, A., Burzykowski, T., 2013. Linear mixed-effects models using R: a step-by-step approach,
828 Springer, Berlin.
- 829 Guertal, E., WR Raun, R.W., Boman, R., 1994. Application of stability analysis for single-site, long-
830 term experiments. *Agronomy Journal* 86, 1016–1019.
- 831 Halekoh, U., Højsgaard, S., 2014. A Kenward-Roger Approximation and Parametric Bootstrap
832 Methods for Tests in Linear Mixed Models - The R Package pbkrtest. *Journal of Statistical*
833 *Software*, 59, 1-30.
- 834 Kenward, M.G., Roger, J.H., 1997. Small Sample Inference for Fixed Effects from Restricted
835 Maximum Likelihood, *Biometrics*, 53, 983–997.
- 836 Kenward, M.G., Roger, J.H., 2009. An Improved Approximation to the Precision of Fixed Effects
837 from Restricted Maximum Likelihood. *Computational Statistics and Data Analysis*, 53,
838 2583–2595.
- 839 Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B., 2015. lmerTest: Tests in Linear Mixed
840 Effects Models. R package version 2.0-29. <http://CRAN.R-project.org/package=lmerTest>
- 841 Laidig, F., Piepho, H.-P., Drobek, T., Meyer, U., 2014. Genetic and non-genetic long-term trends of
842 12 different crops in German official variety performance trials and on-farm yield trends.
843 *Theoretical and Applied Genetics* 127, 2599–2617.
- 844 Loughin, T.M., Roediger, M.P., Milliken, G.A., Schmidt, J.P., 2007. On the analysis of long-term
845 experiments. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 170,
846 29–42.
- 847 McRae, K., Ryan, D., 1996. Design and planning of long-term experiments. *Canadian Journal of*
848 *Plant Science* 76, 595–602.
- 849 Patterson, H.D., 1964. Theory of cyclic rotation experiments. *Journal of the Royal Statistical*
850 *Society. Series B (Methodological)* 26, 1–45.
- 851 Payne, R.W., 2015. The design and analysis of long-term rotation experiments. *Agronomy*
852 *Journal* 107, 772–785.
- 853 Perucci, P., Bonciarelli, U., Santilocchi, R., Bianchi, A.A., 1997. Effect of rotation, nitrogen
854 fertilization and management of crop residues on some chemical and biochemical
855 characteristics of soil. *Biology and Fertility of Soils* 24, 311–316.

- 856 Piepho, H.-P., 1998. Methods for comparing the yield stability of cropping systems - a review.
857 Journal of Agronomy and Crop Science 180, 193–213.
- 858 Piepho, H.-P., 1999. Stability analysis using the SAS system. *Agronomy Journal* 91, 154–160.
- 859 Piepho, H.-P., Büchse, A., Emrich, K., 2003. A hitchhiker's guide to mixed models for randomized
860 experiments. *Journal of Agronomy and Crop Science* 189, 310–322.
- 861 Piepho, H.-P., Büchse, A., Richter, C., 2004. A mixed modelling approach for randomized
862 experiments with repeated measures. *Journal of Agronomy and Crop Science* 190, 230–
863 247.
- 864 Piepho, H.-P., Eckl, T., 2014. Analysis of series of variety trials with perennial grasses. *Grass and*
865 *Forage Science* 69, 431–440.
- 866 Piepho, H.-P., Laidig, F., Drobek, T., Meyer, U., 2014. Dissecting genetic and non-genetic sources
867 of long-term yield trend in german official variety trials. *Theoretical and Applied*
868 *Genetics* 127, 1009–1018.
- 869 Pinheiro, J., Bates, D., 2000. *Mixed-Effects models in s and S-Plus*. Springer-Verlag Inc., New York.
- 870 R Core Team, 2014. *R: A language and environment for statistical computing*, ed. R Foundation
871 for Statistical Computing, Vienna, Austria.
- 872 Richter, C., Kroschewski, B., 2006. Analysis of a long-term experiment with repeated-
873 measurement models. *Journal of Agronomy and Crop Science* 192, 55–71.
- 874 Richter, C., Kroschewski, B., Piepho, H.-P., Spilke, J., 2015. Treatment comparisons in agricultural
875 field trials accounting for spatial correlation. *The Journal of Agricultural Science*
876 153,1187-1207.
- 877 Searle, S.R., Casella, G., McCulloch, C.E., 2006. *Variance components*. John Wiley and Sons, Inc.,
878 Hoboken, New Jersey.
- 879 Shukla, G., 1972. Some statistical aspects of partitioning genotype-environmental components of
880 variability. *Heredity* 29, 237–245.
- 881 Singh, M., Jones, M.J., 2002. Modeling yield sustainability for different rotations in long-term
882 barley trials. *Journal of Agricultural, Biological, and Environmental Statistics* 7, 525-535.
- 883 Smith, A., Stringer, J., Wei, X., Cullis, B., 2007. Varietal selection for perennial crops where data
884 relate to multiple harvests from a series of field trials. *Euphytica* 157, 253–266.
- 885 Spilke, J., Piepho, H.-P., Hu, X., 2005. A simulation study on tests of hypotheses and confidence
886 intervals for fixed effects in mixed models for blocked experiments with missing data.
887 *Journal of Agricultural, Biological, and Environmental Statistics* 10, 374–389.
- 888 Yang, R.-C., 2014. Analysis of linear and non-linear genotype × environment interaction.
889 *Frontiers in Genetics* 5, 227.
- 890 Yates, F., 1954. The analysis of experiments containing different crop rotations. *Biometrics* 10,
891 324-346.
- 892

893

TABLES

894

895 Table 1. Glossary of the terms used in this paper (following Yates, 1954).

Term	Meaning
Rotation	A set of crops that are grown in successive years on the same plot (e.g. M-W-W)
Phase	Each crop component in a rotation (e.g. in the rotation M-W-W, M is phase 1, W is phase 2 and W is phase 3). In a LTE we may be interested in all phases, or (more frequently) in a specific phase (test-crop). Rotations may have the same crop in two different phases (phase difference between W1 and W2). The number of phases defines the period (duration) of each rotation
Cycle	Each single repetition of the whole rotation
Sequence	Each of the n possible arrangements for a rotation of n years, having the same crop ordering, but different initial phases (e.g. M-W1-W2, W1-W2-M and W2-M-W1). Each sequences is uniquely identified by its starting phase, which needs to be randomised to each plot at the start of the experiment, according to the basic requirement that all phases of each rotation/system are grown every year. Likewise, if we look at a specific test-crop, the sequence is uniquely identified by its position in the sequence. It is also relevant to note that each sequence shows the test crop in a different position (e.g. the first sequence shows M in first position, the second shows M in third position and the third shows M in second position)

896

897

898 Table 2: Summary of experimental designs for the five datasets

Design	Dataset 1	Dataset 2	Dataset 3	Dataset 4	Dataset 5
Treatments	Three N-fertilisation levels on continuous wheat	Comparison among five two-year rotations based on wheat	Nine combination of soil tillage and N-fertilisation levels on a two-year sunflower-wheat rotation	Two types of management of crop residues on a three-year rotation maize-wheat-wheat	Five wheat (W) – Maize (M) rotations (MW – MWW – MWWW – MWWWW and MWWWWW)
Response	Wheat yield	Wheat yield	Wheat yield	Wheat yield	Wheat yield
Design	RCBD	Split-plot (main-plots: sequences; sub-plots: rotations)	Split-plot (main-plots: tillage; sub-plots: N-levels)	Split-plot (main-plots: crop residues; sub-plots: phases)	RCBD
Years	30	30	18	30	30
Blocks	3	3	2	3	3
Plots	9	30	36	18	60

N. of records per year	9	15	18	12	45
Total data	270	450	360	360	1350

899

900 Table 3. Experimental design for Dataset 2, in one block and four years (two rotation cycles). W:
901 wheat; P: Pea; GS: grain sorghum; SB: sugar beet; SU: sunflower; FB: faba bean.

Plot No.	Rotation	Sequence	Year			
			1983	1984	1985	1986
4	3	1	W	SB	W	SB
6	1	1	W	P	W	P
12	5	1	W	FB	W	FB
13	2	1	W	GS	W	GS
14	4	1	W	SU	W	SU
20	5	2	FB	W	FB	W
21	2	2	GS	W	GS	W
22	4	2	SU	W	SU	W
27	3	2	SB	W	SB	W
29	1	2	P	W	P	W

902

903

904

905 Table 4. Experimental design for Dataset 4, in one block and
906 three years (one rotation cycles). Three phases can be
907 identified: W1 (wheat after maize), W2(wheat after wheat) and
908 M (maize).

Plot No.	Sequence	Crop Residues	Year		
			1983	1984	1985
1	3	Buried	W2	M	W1
2	1	Buried	M	W1	W2
3	2	Buried	W1	W2	M
4	1	Removed	M	W1	W2
5	2	Removed	W1	W2	M
6	3	Removed	W2	M	W1

909

910

911
 912
 913
 914
 915
 916
 917
 918
 919

Table 5. Selection among several patterns of within-plot variance-covariance structures, following mixed model fits to Dataset 1 using REML. NC: no within-plot correlation; CS: compound symmetry; CSH: compound symmetry with heteroscedastic errors by year; AR(1): autoregressive of order 1; ARH(1): autoregressive of order 1, with heteroscedastic errors by year.

Model	df	AIC ¹	log-likelihood
NC	151	619.97	-158.98
CS	152	612.39	-154.19
AR(1)	152	619.52	-157.76
CSH	181	604.56	-121.28
ARH(1)	181	606.37	-122.18

920
 921
 922

¹: Akaike Information Criterion (Akaike, 1974)

923
 924
 925
 926

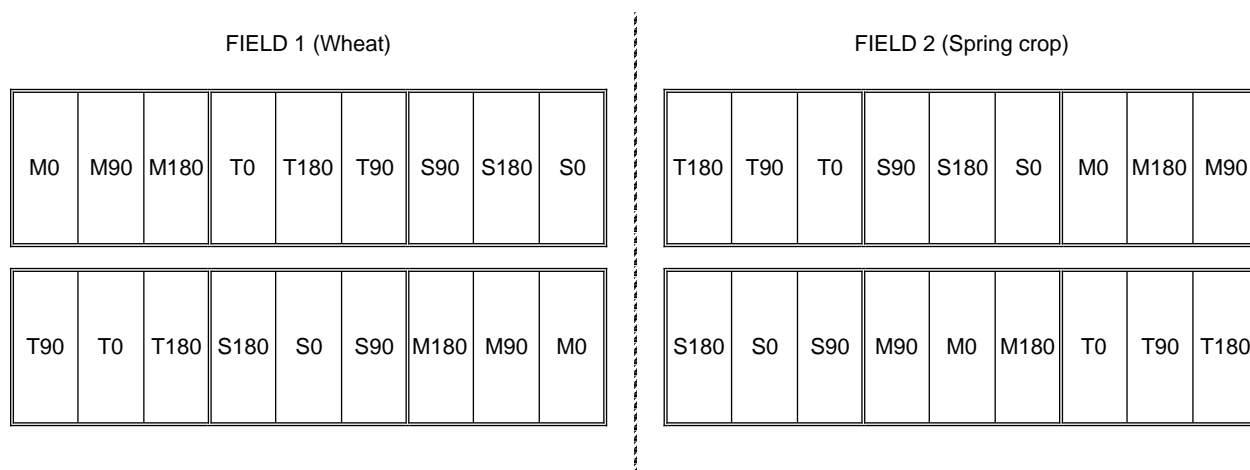
Table 6. Mean yield level, stability standard deviation (SD) and fertility trend for wheat, grown with nine different combinations of tillage and N-fertilisation levels (T: conventional ploughing; M: scarification; S: sod seeding), at the end of a 19-year LTE (Dataset 3).

Tillage	kg N ha ⁻¹	Overall mean (kg ha ⁻¹)	SEM	Stability SD (kg ha ⁻¹)	Slope	SE
M	0	1488	213.4	402.0	12.0	37.95
M	90	3022	189.1	30.3	6.6	32.90
M	180	3807	214.8	414.2	42.5	38.23
S	0	1471	189.0	24.4	13.7	32.89
S	90	2924	211.9	387.4	0.3	37.63
S	180	3554	223.7	489.3	43.7	40.06
T	0	1409	218.3	445.1	-14.7	38.96
T	90	3082	193.9	169.3	-38.9	33.91
T	180	4111	278.3	851.1	16.0	51.14

927
 928

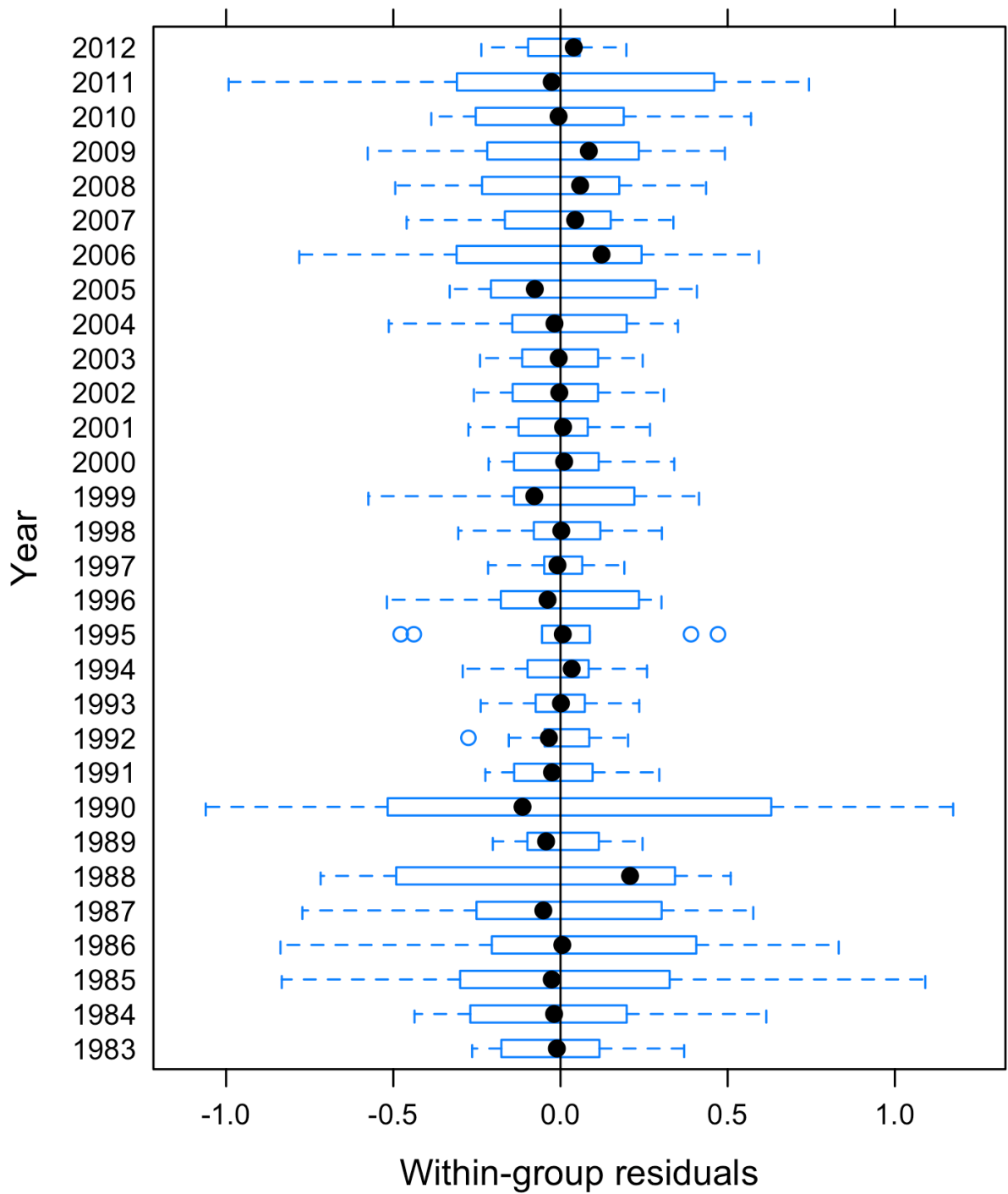
929
930
931

FIGURES



932
933
934
935
936
937

Figure 1. Experimental design for Dataset 3 in one year. The position of wheat and spring crop is exchanged in the following year (T: conventional ploughing; M: scarification; S: sod seeding; 0, 90 and 180 kg N ha⁻¹)



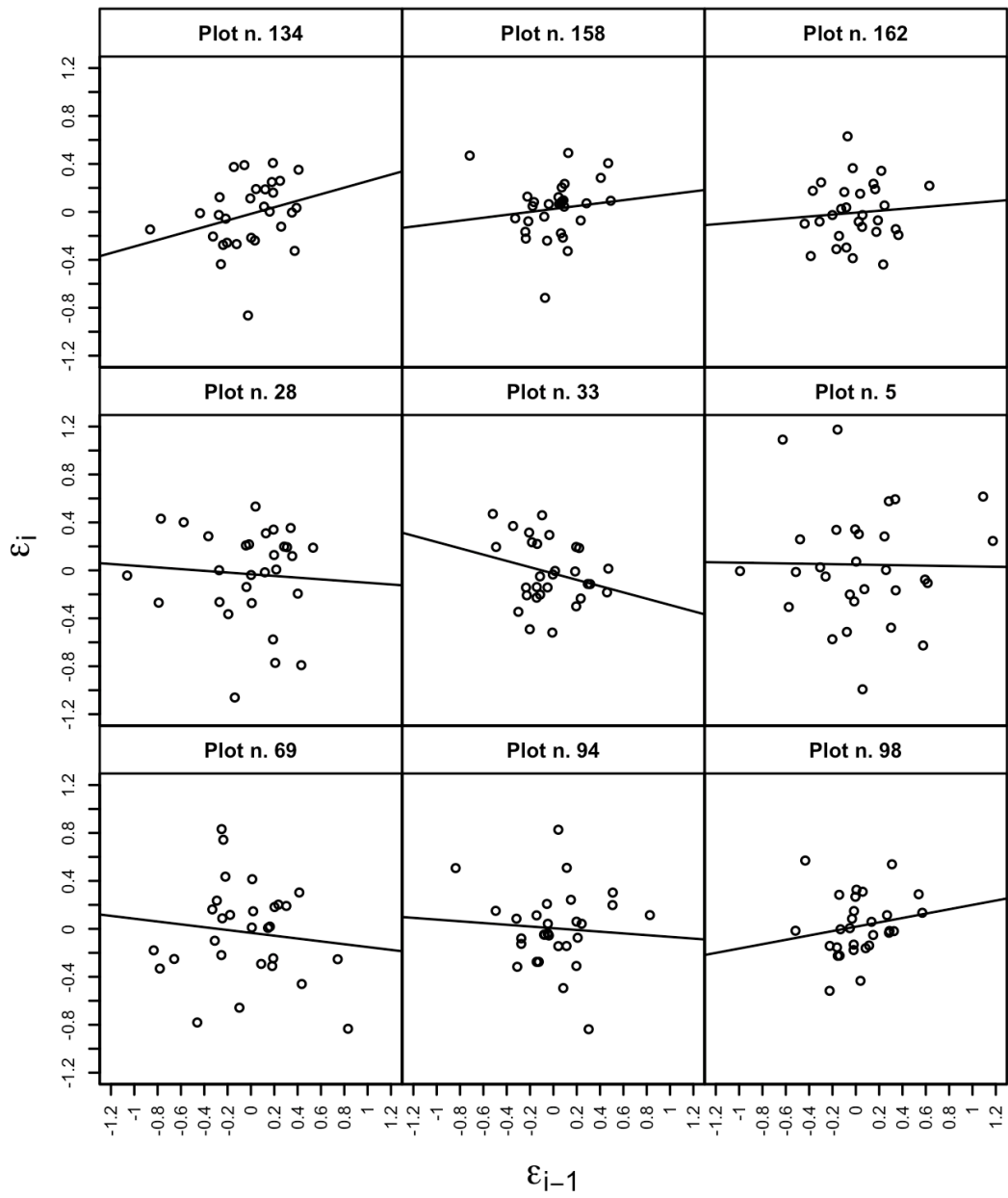
939

940

Figure 2. Boxplot of within-group residuals by year, after a mixed model fit to Dataset 1

941

942

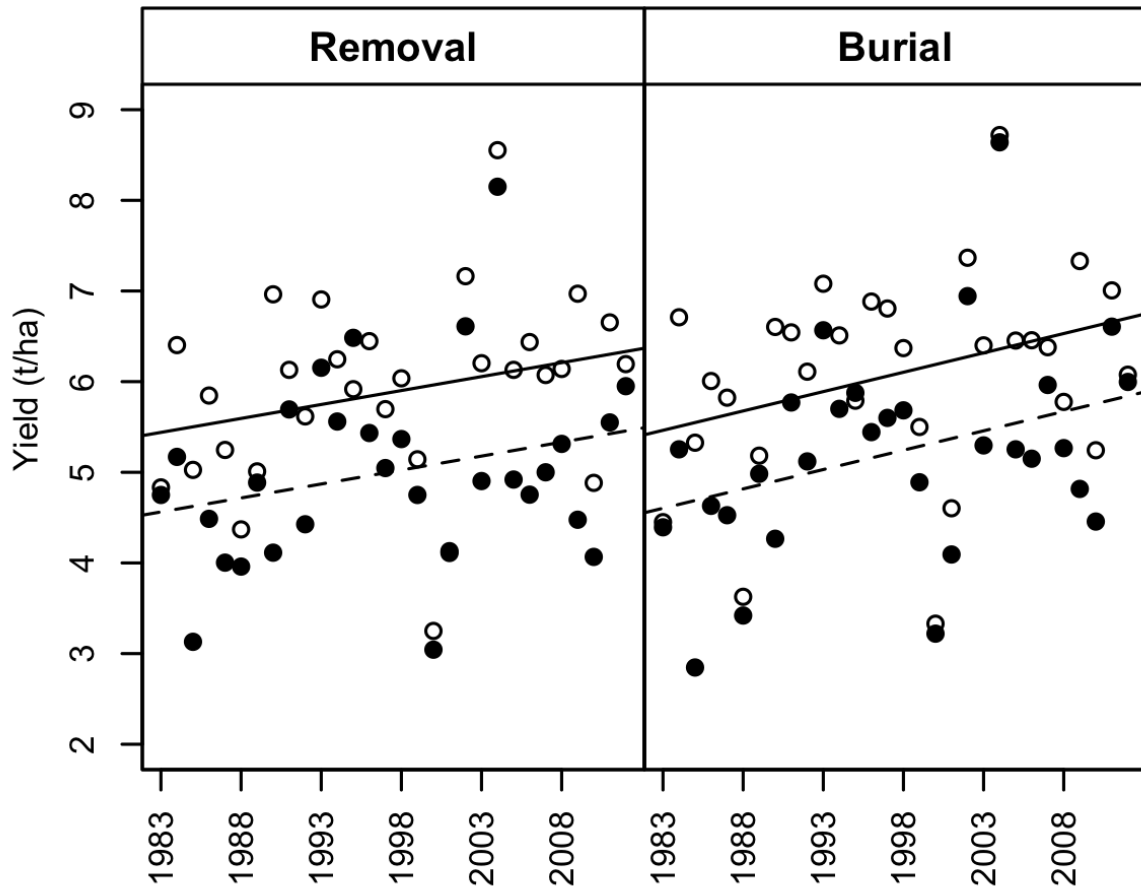


944

945

Figure 3. Lag-one residual plot, after a mixed model fit to Dataset 1.

946



948

949

950

951

952

953

954

955

Figure 4. Fertility trends for wheat in a three-year rotation (Maize-Wheat-Wheat; Dataset 4) with either removal or burial of crop residues. For both management types, closed circles and dotted lines represent the wheat following wheat (phase 3) and open circles and solid lines represent wheat after maize (phase 2). Symbols show observed data, while lines show fitted trends.

Supplemental material

956

957

958 1. The two real LTEs

959

960 Two long-term field experiments are used. The first one (Exp. 1) was established around 1974 at
961 Perugia (central Italy), at the Experimental Farm of the Department of Agricultural, Food and
962 Environmental Sciences (Papiano, 42.9569N, 12.3767E, 196 m a.s.l.) and it is still running at
963 present (2015). The aim is to compare different cropping systems based on winter wheat
964 (*Triticum aestivum* L.) and resulting from the factorial combination of two different types of
965 management of crop residues and 13 types of crop rotation and mineral N fertilisation rates. In
966 particular, crop residues are either (i) removed, or (ii) incorporated into the soil at ploughing,
967 while the 13 rotations/fertilisations include (i - iii) three continuous winter wheat (W) systems,
968 differing in N fertilization rates, i.e. 150 (WN150, 'standard' rate), 200 (WN200) and 250
969 (WN250) kg ha⁻¹, split into 50% at tillering and 50% at shooting; (iv - viii) five types of two-year
970 rotations, wherein wheat is rotated with either pea (*Pisum sativum* L.), grain sorghum (*Sorghum*
971 *bicolor* (L.) Moench), sugar beet (*Beta vulgaris* L. subsp. *saccharifera*), sunflower (*Helianthus*
972 *annuus* L.) or faba bean (*Vicia faba* L. subsp. *minor*). (ix - xiii) five maize (M) - wheat (W)
973 rotations of different lengths, i.e. MW, MWW, MWWW, MWWWW, MWWWWW. For all rotations,
974 all phases are contemporarily present in each year, for a total of 66 plots (33 plots for each of 33
975 possible crop sequences by two residue management levels) in each of three blocks (198 plots in
976 total), arranged according to a split-plot design, with crop residue management on the main
977 plots and crop rotations/fertilisation levels on the subplots (24 m² each). Several measurements
978 are taken on variable intervals, though we will only consider yearly yield measurements on each
979 plot. Further details about this experiment can be found elsewhere (Bonciarelli et al., 2016;
980 Perucci et al., 1997).

981 The second experiment (Exp. 2) was established in 1994 at the Pasquale Rosati
982 experimental farm in Agugliano, Italy (43.32N, 13.22E, 100 m a.s.l.), located on a 10% slope. It is
983 based on a two-year rotation with durum wheat (*Triticum durum* L.) followed either by
984 sunflower (until 2001) or maize (from 2002 onwards). The aim was to compare several
985 cropping systems consisting of a factorial combination of different soil tillage and N-fertilisation
986 practices. The tillage treatments were: (T: conventional 40 cm deep ploughing; M: scarification
987 at 25 cm; S: sod seeding with chemical desiccation and chopping). The three N fertilizer
988 treatments were 0, 90 and 180 kg N ha⁻¹, distributed in two rates for wheat and at seeding for
989 maize/sunflower. The crop rotation was duplicated in two adjacent fields to allow for all crops
990 to be present each year. Specifically, the two fields receive the two sequences, so that they have
991 the two different crops of the rotation in the same year. Within fields, there are two independent
992 randomisations, with two blocks, tillage levels randomised to main-plots (1500 m²) and N levels
993 randomised to sub-plots (500 m²), according to a split-plot design with two replicates.

994

995 2. Structure of datasets

996

997 Datasets are composed by a number of records as shown in Table 2 and columns are named
 998 according to the definition of models 1 to 5. For illustration of the general structure, the first six
 999 observations are listed here for each dataset.

1000

1001 *Dataset 1*

	Plot	N	Year	Block	Yield
1002	1	33	fn150	1983	1 4.54
1003	2	98	fn150	1983	3 4.18
1004	3	162	fn150	1983	2 3.70
1005	4	33	fn150	1984	1 4.57
1006	5	98	fn150	1984	3 5.04
1007	6	162	fn150	1984	2 5.06
1008				
1009				
1010				

1011

1012 *Dataset 2*

	Block	Plot	Rot	Year	Yield
1013	1	1	4 SBW	1983	5.10
1014	2	3	70 SBW	1983	4.50
1015	3	2	135 SBW	1983	4.53
1016	4	1	27 SBW	1984	5.83
1017	5	3	95 SBW	1984	6.26
1018	6	2	160 SBW	1984	6.22
1019				
1020				
1021				

1022

1023 *Dataset 3*

	Year	Field	Cycle	Main	Sub	Block	T	N	Yield
1024	1	1995	1	1	1_1_M	1	1 M	0	855
1025	2	1995	1	1	1_1_M	2	1 M	90	2256
1026	3	1995	1	1	1_1_M	3	1 M	180	2513
1027	4	1995	1	1	1_1_S	4	1 S	90	2302
1028	5	1995	1	1	1_1_S	5	1 S	0	1006
1029	6	1995	1	1	1_1_S	6	1 S	180	3125
1030								
1031								
1032								

1033

1034 *Dataset 4*

	Block	Sub	Res	Year	P	Yield
1035	1	1	15 asp	1983	2	5.05
1036	2	1	23 asp	1983	1	5.08
1037	3	2	43 int	1983	2	4.00
1038						

```

1039 4      2  57 int 1983 1  4.64
1040 5      3  81 asp 1983 2  4.79
1041 6      3  89 asp 1983 1  5.00
1042 .....
1043 .....

```

1044

1045 *Dataset 5*

```

1046   Block Plot Rot Year P Yield
1047 1      1  16 f1m 1983 1  5.36
1048 2      3  82 f1m 1983 1  5.20
1049 3      2 148 f1m 1983 1  4.99
1050 4      1   8 f1m 1984 1  6.46
1051 5      3  75 f1m 1984 1  6.23
1052 6      2 147 f1m 1984 1  6.35
1053 .....
1054 .....

```

1055

1056 3. R Code for all case studies

1057

1058 The following code has been used to perform the analysis with the R software (R Core Team,
1059 2014) on the above mentioned datasets.

1060

```

1061 #####
1062 #Code S1. Examples of implementation in R, by way of mixed models and REML
1063   estimation
1064
1065 #Load data
1066 rm(list=ls())
1067 dataset <- read.csv(file="Dataset2.csv", header=T)
1068 head(dataset)
1069
1070 #Make all numeric variables as factors
1071 dataset$Block <- factor(dataset$Block)
1072 dataset$Plot <- factor(dataset$Plot)
1073 dataset$YearF <- factor(dataset$Year)
1074 dataset$Rot <- factor(dataset$Rot)
1075
1076 #Implementation with R and lmer()
1077 #Random effect for main-plots was disregarded
1078 library(lme4)
1079 mod1 <- lmer(Yield ~ Block + Rot*YearF + Block:YearF +
1080             (1|Plot),
1081             data=dataset)
1082
1083 #Implementation with R and lme()
1084 library(nlme)
1085 mod2 <- lme(Yield ~ Block + Rot*YearF + Block:YearF,
1086            random = ~ 1|Plot,

```

```

1087         data=dataset)
1088
1089 #Implementation with R and gls()
1090 mod3 <- gls(Yield ~ Block + Rot*YearF + Block:YearF,
1091            correlation=corCompSymm(form=~1|Plot),
1092            data=dataset)
1093
1094 #Implementation with ASRemL-R
1095 library(asreml)
1096 mod4 <- asreml(fixed = Yield ~ Block + Rot*YearF + Block:YearF,
1097              random = ~ Plot, data = dataset)
1098
1099
1100 #####
1101 ### Code S2: Checking models
1102
1103 #Load data
1104 rm(list=ls())
1105 dataset <- read.csv(file="Dataset1.csv", header=T)
1106 head(dataset)
1107
1108 #Make all numeric variables as factors
1109 dataset$Block <- factor(dataset$Block)
1110 dataset$Plot <- factor(dataset$Plot)
1111 dataset$YearF <- factor(dataset$Year)
1112 dataset$N <- factor(dataset$N)
1113
1114 #Cumulative analyses with lme()
1115 library(nlme)
1116 mod <- lme(Yield ~ Block + N +
1117           YearF + Block:YearF + YearF:N,
1118           random=~1|Plot,
1119           data=dataset)
1120
1121 plot(mod) #Plot of residuals vs fitted values
1122 qqnorm(mod, ~ resid(.)) #QQ plot of residuals
1123
1124 plot(mod, Plot ~ resid(.), abline=0) #Residuals by plot
1125 #Residuals by year
1126 plot(mod, YearF ~ resid(.), abline=0, xlab="Within-group residuals", ylab="Year")
1127
1128 #Other diagnostic plots
1129 plot(mod, resid(.) ~ fitted(.)|Plot, abline=0)
1130 plot(modMix, N ~ resid(.), abline=0)
1131 plot(modMix, Block ~ resid(.), abline=0)
1132
1133 #Lag-plot 1
1134 library(reshape)
1135 res <- data.frame(plot=names(residuals(mod)), Year = dataset$YearF, residui=residuals(mod))
1136 res <- res[order(res$Year), ]
1137 mat <- cast(Year ~ plot, data = res, value = "residui", fun=mean)
1138 mat2 <- rbind(mat[2:30,], rep(NA,30))
1139 par(las=3, omi=c(1,0.7,0.5,0.5), mfrow=c(3,3))
1140 for(i in 1:9){
1141   par(mai=c(0,0,0.4,0))
1142   lab <- paste("Plot n.", dimnames(mat)[[2]][i+1], sep=" ")
1143   plot(0, type="n", xlim=c(-1.2,1.2), ylim=c(-1.2,1.2), axes=F, main=lab)
1144   points(mat[,i+1] ~ mat2[,i+1])

```

```

1146     abline(lm(mat[,i+1] ~ mat2[,i+1]))
1147     box(which = "plot")
1148     box(which = "figure")
1149     if(i>6) axis(1, at=seq(-1.2,1.2,by=0.2), labels=round(seq(-1.2,1.2,by=0.2), 2))
1150     if(i==1 | i==4 | i == 7) axis(2,at=seq(-1.2, 1.2, by=0.2), labels=round(seq(-1.
1151 2,1.2,by=0.2), 2))
1152 }
1153 expression(varepsilon [i])
1154 mtext(expression(epsilon [i - 1]), las=1, side=1, outer=T, at=c(0.5), padj=3.5, ce
1155 x=1.0)
1156 mtext(expression(epsilon [i]), outer=T, side=2, las=0, at=c(0.5), padj=-2.3, cex=
1157 1.0)
1158
1159 #####
1160 ### Code S3: Implementing within-plot correlation structures by using gls()
1161 #Load data
1162 rm(list=ls())
1163 dataset <- read.csv(file="Dataset1.csv", header=T)
1164 head(dataset)
1165
1166 #Make all numeric variables as factors
1167 dataset$Block <- factor(dataset$Block)
1168 dataset$Plot <- factor(dataset$Plot)
1169 dataset$YearF <- factor(dataset$Year)
1170 dataset$N <- factor(dataset$N)
1171
1172 #CS: compound symmetry
1173 modCS <- gls(Yield ~ Block + N +
1174             YearF + Block:YearF + YearF:N,
1175             correlation=corCompSymm(form=~1|Plot),
1176             data=dataset)
1177
1178 #CSH: Compound symmetry with heteroscedastic errors by year
1179 modCSH <- gls(Yield ~ Block + N +
1180             YearF + Block:YearF + YearF:N,
1181             correlation=corCompSymm(form=~1|Plot),
1182             weights=varIdent(form=~1|YearF),
1183             data=dataset)
1184
1185 #AR: autoregressive
1186 modAR <- gls(Yield ~ Block + N +
1187             YearF + Block:YearF + YearF:N,
1188             correlation=corAR1(form=~1|Plot),
1189             data=dataset)
1190
1191 #ARH: autoregressive with heteroscedastic errors by year
1192 modARH <- gls(Yield ~ Block + N +
1193             YearF + Block:YearF + YearF:N,
1194             correlation=corAR1(form=~1|Plot),
1195             weights=varIdent(form=~1|YearF),
1196             control=list(opt="optim", numIter=3000),
1197             data=dataset)
1198
1199 #UN: unstructured (does not converge!)
1200 modMix4Bis <- gls(Yield ~ Block + N +
1201             YearF + Block:YearF + YearF:N,
1202             correlation=corSymm(form=~1|Plot),
1203             weights=varIdent(form=~1|YearF),
1204             control=list(opt="optim", numIter=3000),

```

```

1205         data=dataset)
1206
1207 AIC(modCS, modCSH, modAR, modARH)
1208
1209 #####
1210 #Code S4. Testing for fixed effects
1211
1212 rm(list=ls())
1213 dataset <- read.csv(file="Dataset2.csv", header=T)
1214
1215 #Making all numerical variables as factors
1216 dataset$Block <- factor(dataset$Block)
1217 dataset$Plot <- factor(dataset$Plot)
1218 dataset$YearF <- factor(dataset$Year)
1219 dataset$Rot <- factor(dataset$Rot)
1220
1221 #Kenward-Roger approximation (by using the lmerTest)
1222 library(lmerTest)
1223 library(pbkrtest)
1224 modlF2 <- lmer(Yield ~ Block + Rot*YearF + Block:YearF +
1225               (1|Plot),
1226               data=dataset)
1227 anova(modlF2, ddf = "lme4")
1228 anova(modlF2, ddf = "Kenward-Roger")
1229
1230 #Kenward-Roger approximation (by using the asreml-R package)
1231 library(asreml)
1232 modFA <- asreml(fixed = Yield ~ Block + Rot*YearF + Block:YearF,
1233                random = ~ Plot, data = dataset)
1234 wald(modFA, denDF="default", ssType="conditional")
1235
1236
1237 #Fitting a full and reduced mixed model by lme()
1238 #Models are fit by maximum likelihood and a LRT test is performed
1239 #by the function anova()
1240 #Random effect for main-plots was disregarded
1241 library(nlme)
1242 modlF <- lme(Yield ~ Block + Rot*YearF + Block:YearF,
1243             random=~1|Plot,
1244             data=dataset, method="ML")
1245 modlR <- lme(Yield ~ Block + Rot + YearF + Block:YearF,
1246             random=~1|Plot,

```

```

1247         data=dataset, method="ML")
1248 anova(modlF, modlR)
1249
1250 #Parametric bootstrap (may be extremely slow!!!!!!!)
1251 #This reproduces the procedure outlined in the paper, although
1252 #the nsim argument may be used to avoid the 'for' cycle
1253 nsim <- 100000
1254 lrtSim <- numeric(nsim)
1255 for (i in 1:nsim){
1256     y <- simulate(modlR, nsim=1, m2=modlF, method="ML")
1257     lrtSim <- as.numeric(2*(y$alt$ML[,2] - y>null$ML[,2]))
1258     lrtSim
1259     print(i)
1260 }
1261 mean(lrtSim > 274.99)
1262
1263
1264 #####
1265 #Code S5. Fertility trends
1266
1267 rm(list=ls())
1268 dataset <- read.csv(file="Dataset4.csv", header=T)
1269 head(dataset)
1270
1271 #Making all numerical variables as factors
1272 dataset$Block <- factor(dataset$Block)
1273 dataset$Main <- factor(dataset$Block:dataset$Res)
1274 dataset$Sub <- factor(dataset$Sub)
1275 dataset$YearF <- factor(dataset$Year)
1276 dataset$P <- factor(dataset$P)
1277
1278 #Fitting a fertility trend
1279 library(lme4)
1280 dataset$YearC <- scale(dataset$Year, scale=F)
1281 options(contrasts=c("contr.sum", "contr.poly"))
1282 modLST <- lmer(Yield ~ I(Res:P) + Res/YearC +
1283               (1|YearF) + (1|YearF:Res) + (1|YearF:P) + (1|YearF:Res:P) +
1284               (1|YearF:Block) +
1285               (1|Block/Main/Sub),
1286               data=dataset)
1287 summary(modLST)
1288
1289
1290 #####
1291 #Code S6. Stability variance model
1292
1293 rm(list=ls())
1294 dataset <- read.csv(file="Dataset3.csv", header=T)

```

```

1295 head(dataset)
1296
1297 #Making all numeric variables as factors
1298 dataset$Field <- factor(dataset$Field)
1299 dataset$Block <- factor(dataset$Block)
1300 dataset$Sub <- factor(dataset$Sub)
1301 dataset$FieldBlock <- factor(I(dataset$Block:dataset$Field))
1302 dataset$FBM <- factor(dataset$FieldBlock:dataset$Main)
1303 dataset$FBMS <- factor(dataset$FBM:dataset$Sub)
1304 dataset$Till <- factor(dataset$T)
1305 dataset$N <- factor(dataset$N)
1306 dataset$Treat <- factor(dataset$Till:dataset$N)
1307 dataset$YearF <- factor(dataset$Year)
1308 dataset <- within(dataset, one1 <- one2 <- one3 <- one4 <- one5 <- 1L)
1309 dataset$YearC <- scale(dataset$Year, scale=F)
1310
1311 options(contrasts=c("contr.sum", "contr.poly"))
1312 mod <- lme(Yield ~ Treat - 1 + Treat/YearC,
1313           random = list(one1 = pdIdent(~Block - 1),
1314                         one3 = pdIdent(~FieldBlock - 1),
1315                         one4 = pdIdent(~FBM - 1),
1316                         one5 = pdIdent(~FBMS - 1),
1317                         YearF = pdBlocked(list(pdIdent(~1),
1318                                               pdDiag(~Treat - 1))))),
1319           weights=varIdent(form=~1|YearF),
1320           control=list(opt="optim", numIter=3000),
1321           data=dataset)
1322
1323 tTable <- summary(mod)$tTable
1324 shuk <- as.numeric(VarCorr(mod)[61:69,2])
1325 cbind(tTable[1:9,1:2], "shukla"=shuk, tTable[10:18,1:2])

```