Annex VI

Analysis of data of Annex V with SAS/STAT® 15.1.

Cumulative germination through time (data from Sanoubar et al., 2018) are analysed (see Annex V for the dataset used) as an example. It should be noted that germination data were recorded only after some days of incubation in water, thus that the lag time and the timepoints at which germination really began are unknown. Also, some germination curves do not appear to have reached a plateau. Thus, beyond being interval censored as typical for longitudinal germination data, these data are also both left- and right-censored. This is not a full-accomplishing way to record germination data, because some important information was lost. Nonetheless, the time-courses can still be properly analysed for inference. Once relatively complete background information is available for a given seed species and type, in fact, observations at just two carefully chosen timepoints might sometimes be enough for some specific testing applications (Bradford and Still, 2004).

Based on the thorough analysis of Annex IV, the quasi-marginal model with AR(1) covariance structure appears the best candidate to analyse these data. The ANTE(1) structure, indeed, causes convergence problems

```
proc GLIMMIX order=data /*method=Laplace*/ empirical=mbn;
class light_source species plate time;
model germ/n = light_source|species|time / link=probit;
random intercept / subject=plate(light_source*species);
random residual / subject=plate(light_source*species) type=ar(1) group=light_source*species;
nloptions tech=newrap;
lsmeans light_source*species*time / cl ilink plot=meanplot(plotby=light_source sliceby=species join)
slice=time slicediff=time adjust=smm;
covtest zeroG;
covtest diagR;
run;
```

The Laplace approximation can be advisably used to assess the exact overdispersion so to check the completeness of model's specification. To apply it, the 'random residual' statement (invoking R-side modelling) must be omitted, and, thereby, the 'covtest diagR' becomes superfluous. As all the factors described in the experimental setup are already considered, this otherwise recommendable control is not shown here. Note that the sign '|' is utilized as a shorthand for requesting all the interactions among the fixed factors. Two replicate plates are named 'a' and 'b' throughout all levels, and their nesting is therefore specified to make their identities unique. The 'group=' option in the 'random residual' statement requests independent variance/covariance structures across levels of the highest fixed-factor interaction excluding time. The 'lsmeans' statement provides LS-means for all the levels of the highest fixed-factors interaction (including time). The 'nloptions' statement may not be necessary, as the Dual Quasi-Newton optimization technique provides valid estimations if it converges to a fitting solution.

RESULTS (excerpts):

Table 1.

Fit Statistics	
-2 Res Log Pseudo-Likelihood	-53.52
Generalized Chi-Square	30.00
Gener. Chi-Square / DF	1.00

Table 2.

Covariance Parameter Estimates				
Cov Parm	**Subject**	**Group**	**Estimate**	**Standard Error**
Intercept	plate(light_*specie)		0.001183	0.001443
Variance	plate(light_*specie)	light_source*species neon Atriplex hortensis	0.2109	0.1494
AR(1)	plate(light_*specie)	light_source*species neon Atriplex hortensis	0.3653	0.4440
Variance	plate(light_*specie)	light_source*species neon Chenopodium quinoa	0.5222	0.7223
AR(1)	plate(light_*specie)	light_source*species neon Chenopodium quinoa	0.9646	0.05257
Variance	plate(light_*specie)	light_source*species neon Sanguisorba minor	1.2238	1.6258
AR(1)	plate(light_*specie)	light_source*species neon Sanguisorba minor	0.9352	0.09497
Variance	plate(light_*specie)	light_source*species led Atriplex hortensis	0.1483	0.1233
AR(1)	plate(light_*specie)	light_source*species led Atriplex hortensis	0.4381	0.5569
Variance	plate(light_*specie)	light_source*species led Chenopodium quinoa	0.9624	1.1705
AR(1)	plate(light_*specie)	light_source*species led Chenopodium quinoa	0.9247	0.1004
Variance	plate(light_*specie)	light_source*species led Sanguisorba minor	0.04819	0.03720
AR(1)	plate(light_*specie)	light_source*species led Sanguisorba minor	-0.4938	0.4364

Table 3.

Type III Tests of Fixed Effects				
Effect	**Num DF**	**Den DF**	**F Value**	**Pr > F**
light_source	1	6	98.34	<.0001
species	2	6	114.35	<.0001
light_source*species	2	6	5.95	0.0376
time	4	24	97.22	<.0001
light_source*time	4	24	1.77	0.1675
species*time	8	24	8.35	<.0001
light_s*species*time	8	24	1.15	0.3692

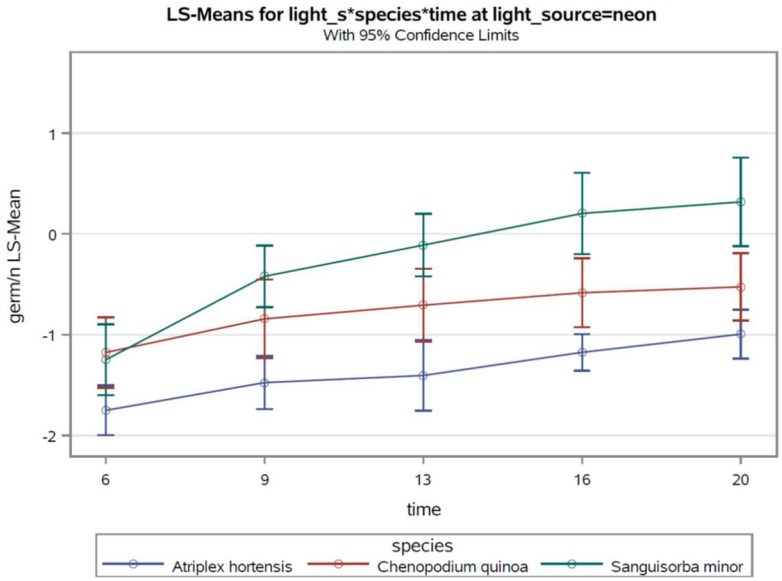

Figure 1.

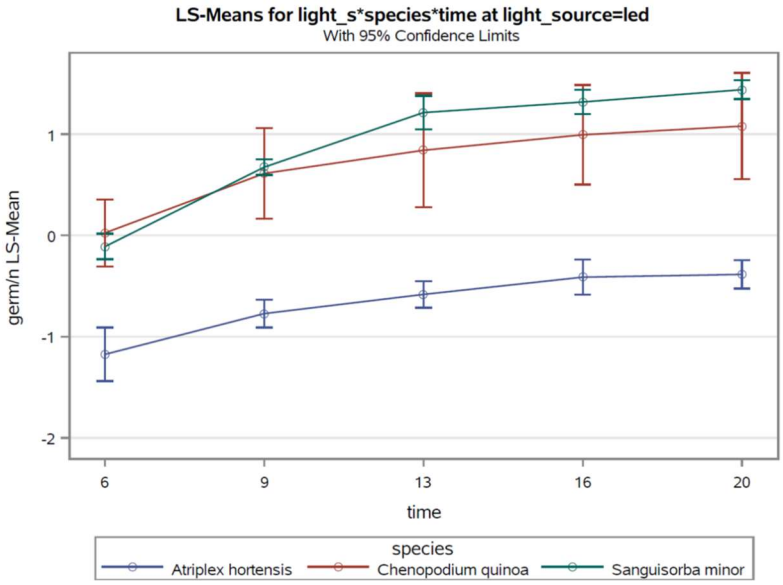

Figure 2.

Table 4.

Tests of Effect Slices for light_s*species*time Sliced By time				
time	Num DF	Den DF	F Value	Pr > F
6	5	24	43.58	<.0001
9	5	24	119.18	<.0001
13	5	24	81.88	<.0001
16	5	24	139.73	<.0001
20	5	24	158.01	<.0001

Table 5.

Tests of Covariance Parameters Based on the Residual Pseudo-Likelihood					
Label	**DF**	**-2 Res Log P-Like**	**ChiSq**	**Pr > ChiSq**	**Note**
No G-side effects	1	-52.1602	1.36	0.1215	MI
Conditional Independence	6	-43.5487	9.97	0.1257	DF

DF: P-value based on a chi-square with DF degrees of freedom.
MI: P-value based on a mixture of chi-squares.

In presence of a random factor and in the absence of an integral approximation, like the Laplace approximation, pseudo-likelihood is applied, and fixed effects are profiled by linear fitting on the linked scale. No overdispersion is apparent (Table 1), and indeed, if the Laplace approximation were used, the exact overdispersion parameter would be 0.20 (not shown), indicating underdispersion, as typical for binomial data. As usual (at least in absence of relevant sources of heterogeneity), a very small variance is found for the random plate effect (corresponding to a maximum standard deviation of 1.4 % around 50 % on the percentile scale), even smaller than its standard error (Table 2). Error variances, as well as autocorrelations, seem to vary considerably (Table 2) among levels of the highest interaction excluding time, here 'light_source*species', as frequently observed. All the factors are significant ($P \leq 0.05$), as are the 'light_source*species' and 'species*time' interactions (Table 3). Figures 1 and 2 display the estimated LS-means on the linked scale. Heterogeneous variances and an overall upward shift when LED light instead of fluorescent light ('neon') is used are apparent. Table 4 shows that significant differences between LS-means were found at every timepoint. Finally, tests of covariance parameters (Table 5) suggest poor significance of both the random effect and the variance/covariance structure. These tests are however based on pseudo-likelihood ratios and are therefore merely indicative, at least as regards variance structures. The 'covtest' for conditional independence ('diagR') is used to evaluate the effect of covariances alone by reducing the R-side covariance structure to the diagonal form. Thereby, the null model for testing covariance parameters is fitted to the final pseudo-data of the converged optimization, and this therefore is an exact test. Statements modelling the random effect and the variance/covariance structure could be removed without substantially affecting the inferences (not shown), but this is not always the case. If discrepant inferences were obtained, the model that properly considers the random effect(s) and the variance/covariance structure ought to be anyway preferred, because it better reflects the experimental design. So, these tests are mostly useful as suggestive of an improper modelling of the random factor(s) and/or the variance/covariance structure when they result grossly non-significant. In such instances, the model should be double-checked and reconsidered.

It might be noted that even though germination of these herb seeds was slow and incomplete, at least within the (too short) time of observation used in this experiment, the random plate effect was very small, and smooth average curves were observed even though only two replicate plates were used. This indicates that the number of seeds per plate was large enough, and the seed samples were uniform, apart from physiological inter-seed variability. Although poor and sluggish germination can be associated with seed lot heterogeneity due to genetic, health or seed management inconveniences, slow and partial germination may well be a species trait, or it can be due to dormancy or suboptimal germination conditions. In these cases, slow germination is linked to a continuous variability of the physiological status of the seeds, characteristic of every seed population (Bradford and Still, 2004), and is not linked to heterogeneity of the seed batch, which causes discrete anomalies in the germination time-course.

References

Bradford K.J., Still D.W. (2004). Applications of hydrotime analysis in seed testing. *Seed Technology* 26:75-85.

Sanoubar R., Calone R., Noli E. and Barbanti L. (2018). Data on seed germination using LED versus fluorescent light under growth chamber conditions. *Data in Brief* 19:594-600.