Considerations on sampling and statistical analysis in grassland ensiling trials

Kroschewski, B.*; Auerbach, H.†; Weiss, K.*
*Humboldt Universität zu Berlin, Berlin, Germany
†International Silage Consultancy (ISC), Halle (Saale), Germany

Key words: grassland ensiling trials; sampling; linear mixed model; block effect; variance homogeneity

Abstract. Critical findings on design, statistical analysis, and interpretation of the results will be addressed based on comparative ensiling trials. For this aim, a lab-scale ensiling trial on biostatistical issues was conducted in 2021. Grass material from a permanent mowing pasture was taken from (i) 10 sampling points, (ii) one sampling point, (iii) a mixture of 10 sampling points. For each sub-trial (based on the sampling design), 3 levels of the fixed treatment factor silage additive were tested with 10 replicates (without additive, chemical silage additive, biological silage additive). The analysis was performed within a linear mixed effects model (LMM) as randomized complete block design (RCBD), accounting for systematic effects of field sampling points (i) and/or time processing (i, ii, iii). In sub-trial (i), variability in trait values was highest and more influenced by treatments (variance heterogeneity), and block effects were most pronounced. In contrast, the block effect was less pronounced in (ii) and (iii), and we could not find a time gradient in the silage trait values. Depending on the nature of the silage trait (distribution, treatment variances), a suitable analysis procedure has to be chosen. The frequently used low number of replications is probably not sufficient.

Introduction

Reliable and reproducible results require both a suitable experimental design and data analysis appropriate to the intended experimental purpose and observed traits. Across all scientific disciplines, only an insufficient number of trials generate reproducible results, which are at the core of the scientific integrity of modern research (Bello and Renter 2018).
In comparative silage trials, the way forage is sampled in the grassland plays a role, which affects the analysis model to be selected, the magnitude of variation in traits, and the scope of inferences related to the results. Based on comparative ensiling trials, critical findings on sampling design and the statistical analysis of selected silage traits will be addressed. Following up on a 2017 ensiling trial with natural grassland (Kroschewski *et al.* 2018), we conducted a second lab-scale ensiling trial in 2021 to focus on biostatistical issues related to this objective, which will be considered here.

Material and Methods

Sampling design and measurements

A 5-row x 5-column grid (= 25 cells) was established on the central area (232.5 m x 500 m) of a permanent mowing pasture (2nd cut). Considering all rows and columns, 10 of these cells were randomly selected and grass material was collected in their center.
Three sub-trials were conducted, using grass material from

(i) 10 different sampling points in the field (P1 … P10): **FIELD**
(ii) one sampling point (P10): **POINT**
(iii) a mixture of all 10 sampling points used (P1 … P10): **MIX**

For each sub-trial, three levels of the fixed treatment factor **silage additive** were tested with 10 replicates (**CON** = without additive, **CHEM**=chemical additive, **LAB** = biological additive). Treatments were independently assigned to each mini-silo as experimental unit. A total of 90 mini-silos were prepared and stored at 22°C for 96 days until opening. A number of traits were measured on the silage after opening. For the measurement of aerobic stability, a portion of the silage removed from the jars was aerated at 21 °C for 22 days. In this paper, we focus only on silage traits, although measurements of the raw material (fresh herbage during ensiling) were also collected.
In order to account for a possible systematic time effect in the analysis model, all processing steps (ensiling, opening of jars, measurements, chemical analyses, etc.) for each sub-trial were organized in a framework of a RCBD (time blocks with random order of treatments per block).

Statistical analysis

The observed values were examined for their distributional characteristics. The normal distribution of the residuals was checked graphically and with the Shapiro-Wilk test. The variance heterogeneity was checked by Levene's test (using the absolute values of the residuals). For each sub-trial, the comparison of treatments was carried out in the framework of an LMM with silage additive as fixed treatment factor in a RCBD. Stroup *et al.* (2018) advise considering block as random factor. For a simple experimental design such as a RCBD with complete blocks, the choice of whether to include blocks as fixed or random factor in the model does not affect the comparison of treatments. Because of convergence problems in the analysis with separate treatment variances, we did both. Data analysis and graphical presentation were performed by Proc MIXED (using REML method) and Proc SGPLOT of SAS 9.4 (SAS Institute Inc., Cary, NC, USA). For all tests, alpha=0.05 was used.

Results and discussion

A first impression of the observed values in the 3 sub-trials is given in *Fig. 1*. In FIELD, dry matter in silage *(plot A)* showed a wide range of values between 22.6 and 34.2 % (reflecting the diversity of the sampling points), while the values in POINT (reflecting a special field situation with small values) and MIX (reflecting the average field situation) hardly vary. Treatments did not have an influence. This differentiation of the data between sub-trials was not seen for the other traits presented *(Fig. 1, B-D)*. Instead, treatments differentiated the observations depending on the trait, but the behaviour was similar across the sub-trials for the same trait. Nonetheless, the values appeared to be more scattered for some treatments in FIELD than in the other sub-trials. Also, between the treatments of the same sub-trial the assumption of homogeneous variances seemed to be violated.

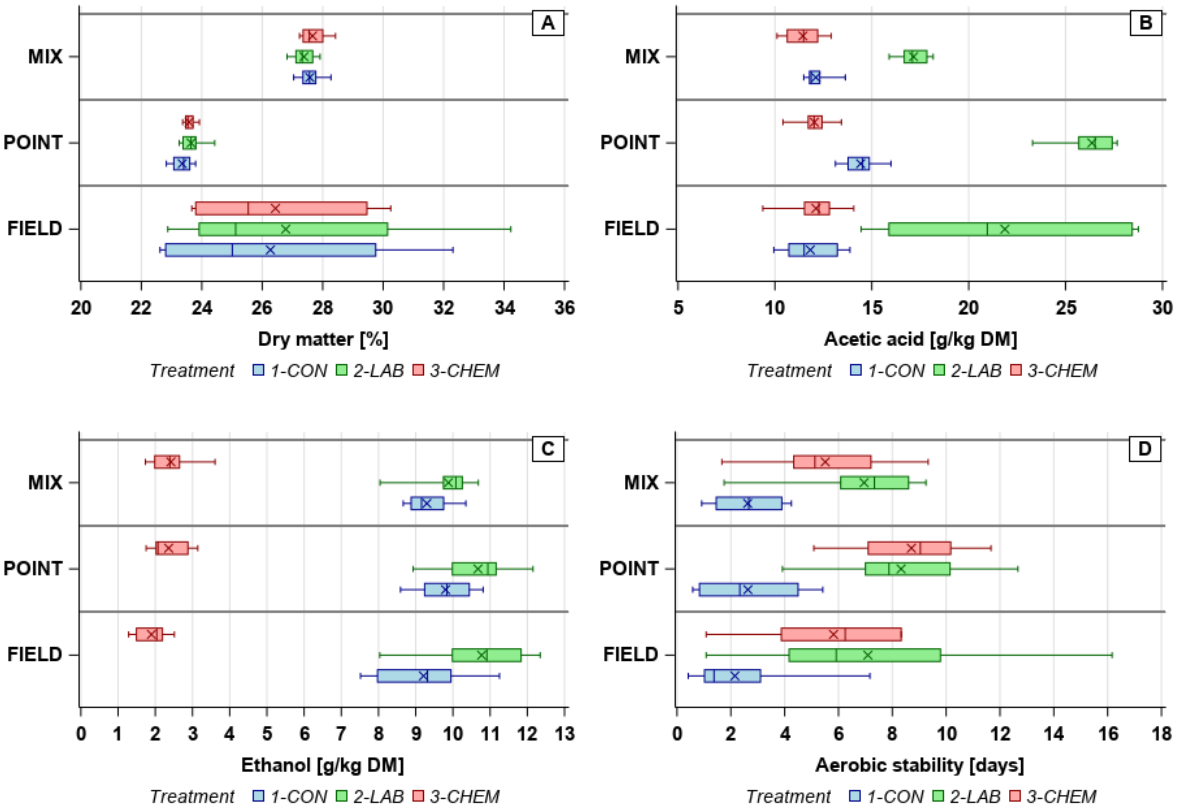

Figure 1. Box-Whisker-Plot for selected silage traits (box indicates 25%, 50%, 75% percentiles, and the whiskers minimum and maximum, mean is marked by cross)

Existence of block effects

Here, the block factor accounts for the systematic effect of field sampling locations (only sub-trial FIELD) and the systematic effect of processing time (FIELD, POINT, MIX). For FIELD, both systematic block effects are confounded.

The sub-trials were analyzed as RCBD with both fixed and random block effects. In the case of random block effects, a block variance is estimated in addition to the residual variance. Both variance components are shown in *Table 1*, because of the better interpretability as standard deviations. The percentage represents the relative share of the block variance in the total variance (sum of the two

variance components). For 10 of the 11 traits considered, the residual variance was largest in FIELD and predominantly smallest in MIX.
In FIELD, all block variances were estimated greater than zero; for 5 traits, the analysis as RCBD led to a better model fit compared to CRD (completely randomized design). In POINT and MIX this was true for only 3 traits each. Here, block variances were partially estimated to zero. General, the larger the proportion of block variance to total variance, the more likely model fit was superior as RCBD. For RCBD with fixed block effects, an F-test can be used to evaluate block effects.
Ensiling took about 12 hours from picking the forage material in the field to filling into the jars, so we suspected a temporal gradient in silage trait values that would be due to a possible alteration of the forage during the ensiling day. We found weak correlations between block number and observations for 2 (POINT) and 3 (MIX) traits, but for the same trait the signs were contradictory. Existing block effects do not seem to be due to a time gradient.

Table 1. Assessment of block effects: (a) RCBD with random block effects (variance components for residual and block, better model fit for RCBD versus CRD); and (b) RCBD with fixed block effects (*P* value of F test for blocks)

Trait	FIELD					POINT					MIX				
(g kg^{-1} DM)[&]	Residual	Block		Fit[4]	F test	Residual	Block		Fit	F test	Residual	Block		Fit	F test
	σ_{Error}	σ_{Block}	*(%)*[2]		*P value*	σ_{Error}	σ_{Block}	*(%)*		P value	σ_{Error}	σ_{Block}	*(%)*		*P value*
pH	0.11	0.08	*(34)*	**x**	**0.042**	0.04	0.06	*(64)*	**x**	**<.001**	0.03	0.05	*(73)*	**x**	**<.001**
Lactic acid	10.44	4.49	*(16)*		0.203	5.43	1.52	*(7)*		0.335	4.08	1.91	*(18)*		0.172
Acetic acid	3.37	1.48	*(16)*		0.197	0.95	0.31	*(10)*		0.294	0.54	0.53	*(49)*	**x**	**0.007**
1,2-Propanediol	2.91	0.69	*(5)*		0.159	1.12	0.27	*(5)*		**0.005**	0.27	0.03	*(1)*		0.474
NH_3-N (% total N)	0.50	0.72	*(67)*	**x**	0.369	0.21	0.07	*(11)*		0.368	0.56	0.24	*(16)*		0.449
WSC[5]	4.45	8.42	*(78)*	**x**	**<.001**	1.77	0.02	*(0)*		0.271	1.96	0	*(0)*		0.201
Ethanol	0.94	0.46	*(19)*		**<.001**	0.51	0.52	*(51)*	**x**	0.474	0.62	0	*(0)*		0.509
Ester (mg kg^{-1} DM)	19.0	14.4	*(36)*	**x**	0.034	21.9	0	*(0)*		0.771	15.7	9.4	*(26)*		0.091
ASTA[6] (days)	2.59	2.06	*(39)*	**x**	0.156	2.49	0	*(0)*		**0.014**	1.67	0.92	*(23)*		0.283
TCUM[7] (°C)	209	121	*(25)*		**0.027**	169	44	*(6)*		0.943	149	111	*(36)*	**x**	0.116
DM loss (%)	0.32	0.16	*(19)*		0.099	0.09	0.08	*(44)*	**x**	0.351	0.13	0.04	*(10)*		0.037

DM, dry matter; [&]unless stated otherwise; [1]square root of residual variance; [2]square root of block variance, [3]block variance in % of total variance; [4]better model fit (x): AIC (RCBD) < AIC (CRD); [5]water-soluble carbohydrates; [6]aerobic stability; [7]cumulated temperature obtained by summing up the difference in temperature between ambient and silage measured at 2-hours interval

Assumption of normal distribution and homogeneity of treatment variances

Some of the studied traits often had values below the detection limit and/or extremely high dispersion (1,2-propanediol, ester), diagnostic plots (not shown) and Shapiro-Wilk test *(Table 2)* did not support normally distributed residuals. For these, alternative analysis procedures using a different distribution (generalized linear mixed models, GLMM) or ranking procedure (no distribution assumption) should be used for the statistical analysis.

Table 2. Checking the normal distribution of residuals and the assumption of homogeneous or heterogeneous treatment variances in the RCBD model

Trait	FIELD					POINT					MIX				
(g kg^{-1} DM)[&]	CV[1]	S-W-T[2]	Levene[3]	Converg.[4]		CV	S-W-T	Levene	Converg.		CV	S-W-T	Levene	Converg.	
	%	*P value*	*P value*	*fix*	*rand.*	%	*P value*	*P value*	*fix*	*rand.*	%	*P value*	*P value*	*fix*	*rand.*
pH	2.9	**0.033**	**<.001**			1.0	0.543	0.610	x		0.7	0.189	**0.026**		x
Lactic acid	18.7	0.299	**<.001**		x	9.6	0.424	0.651	x	x	6.6	**0.034**	0.085	x	x
Acetic acid	22.1	**0.032**	**<.001**			5.4	**0.037**	0.277	x	x	4.0	0.404	0.555	x	x
1,2-Propanediol	122.0	**<.001**	**0.007**	x		45.5	**<.001**	**0.004**			16.3	0.801	0.455	x	x
NH_3-N (% total N)	7.5	0.580	0.719	x	x	3.1	0.050	0.269	x	x	7.8	**<.001**	0.257	x	x
WSC[5]	38.9	**0.016**	0.562			59.3	**<.001**	**<.001**			25.6	0.753	0.215		x
Ethanol	12.9	0.727	0.335	x		6.7	0.526	0.341	x		8.6	0.189	0.903	x	x
Ester (mg kg^{-1} DM)	36.1	0.180	0.073			43.7	**0.027**	0.627	x	x	24.4	0.065	0.083	x	x
ASTA[6] (days)	51.7	**0.021**	**0.021**			38.0	0.119	0.624	x	x	33.1	0.350	0.714	x	x
TCUM[7] (°C)	46.5	0.253	0.761	x	x	46.6	0.076	0.682	x	x	30.5	0.765	0.051	x	
DM loss (%)	7.0	0.158	**<.001**			1.9	0.507	**0.009**	x		2.9	**<.001**	0.676	x	x

DM, dry matter; [&]unless stated otherwise; [1]residual CV (for RCBD-model with common treatment residual variance); [2]Shapiro-Wilk test, [3]Levene's test; [4]Convergence criteria met for analysis of RCBD-model with specific treatment variances (fix=fixed / rand.=random block effects); [5]water-soluble carbohydrates; [6]aerobic stability; [7]cumulated temperature obtained by summing up the difference in temperature between ambient and silage measured at 2-hours interval

For acetic acid, it can be concluded from *Fig. 1 (B)* that the assumption of equal treatment variances cannot be supported for FIELD. Levene's test *(Table 2)* detected variance heterogeneity, which occurred most frequently for FIELD, less frequently for POINT, and for MIX only for pH (with extremely low dispersion, differences are not surprising). For treatment-specific residual variances, several variance components have to be estimated iteratively, which can lead to convergence problems. If the analysis was performed as RCBD, we found considerable convergence problems, regardless of whether the blocks were included as fix or random effects in the model *(Table 2)*.
In contrast, the analysis as CRD with separate treatment variances almost always led to convergence.

Comparisons of treatments (n=10) and for sub-samples (n=3, n=6) for selected traits
For three traits shown in *Fig. 1*, the analysis was performed as RCBD and Tukey's test procedure for pairwise comparisons *(Table 3)*. For FIELD, separate treatment variances would have to be estimated for acetic acid and aerobic stability *(Table 2)*. Due to convergence problems, this was not an option, so we always used a common residual variance.

Table 3. Pairwise comparisons of treatments for total sample (n=10) and subsets by Tukey's test procedure

Trait	sub-trial	LS-Means for total sample, n=10			SED	HSD (α=0.05)	No. of significant treatment differences from: 210 subsets with n=6			120 subsets with n=3		
		1-CON	2-LAB	3-CHEM			1 - 2	1 - 3	2 - 3	1 - 2	1 - 3	2 - 3
Acetic acid	FIELD	11.8 a[§]	21.9 b	12.1 a	1.51	3.8	210	0	210	67	0	63
(g kg^{-1} DM)	POINT	14.4 b	26.4 c	12.0 a	0.43	1.1	210	210	210	120	69	120
	MIX	12.1 b	17.2 c	11.4 a	0.24	0.6	210	60	210	120	13	120
Ethanol	FIELD	9.2 b	10.8 c	1.9 a	0.42	1.07	106	210	210	25	120	120
(g kg^{-1} DM)	POINT	9.8 b	10.7 c	2.4 a	0.23	0.58	110	210	210	24	120	120
	MIX	9.3 b	9.9 b	2.4 a	0.28	0.70	35	210	210	13	120	120
Aerobic	FIELD	2.15 a	7.09 b	5.82 b	1.16	2.96	200	72	0	24	31	0
stability	POINT	2.63 a	8.33 b	8.71 b	1.11	2.84	210	210	0	82	76	0
(days)	MIX	2.63 a	6.95 b	5.51 b	0.74	1.90	210	143	22	67	20	6

[§]treatment means per sub-trial with no letter in common are significant different

The sub-trials showed largely the same differentiation of treatments in the traits examined with n=10. Usually silage trials are performed with a smaller number of repetitions. In order to show the influence of the sample size on the test decision for the present real experimental data, samples with smaller sizes (n=6, n=3) were extracted from the total sample (n=10).
For acetic acid and ethanol, *Fig. 2* showed a clear differentiation between treatments, but not so clear for aerobic stability. The smaller the sample size and the smaller the treatment effects, the more inconsistent the test decisions in the subsamples. However, it is critical to consider the biological relevance of the effects when assessing significance.

Conclusions
Forage sampled at several locations represent the field. The replicates can correspond to the sampling points and their variability reflects arbitrary field locations, or a composite sample is used and the replicates reflect the average field situation. The latter led to smaller variability of observations and less frequent variance heterogeneity. Systematic time effects can be accounted for by blocking. We could not find a time gradient in the silage trait values.
Depending on the nature of the silage trait, a suitable analysis procedure has to be chosen (linear model assuming normally distributed residuals, generalized linear model assuming a different distribution, rank procedure without distribution assumption). The frequently used n= 3 is rather insufficient, especially if GLMM or rank procedures are to be applied.

References

Bello, N. M., and Renter, D. G. 2018. Invited review: Reproducible research from noisy data: Revisiting key statistical principles for the animal sciences. *Journal of Dairy Science*., 101: 1-23.

Kroschewski, B., Auerbach, H., Weiss, K. 2018. Statistics and experimental design in silage research: Some comments on design and analysis of comparative silage experiments. *Proceedings of the XVIIIth International Silage Conference*., 554-560.

Stroup W. W., Milliken G. A., Claassen E. A., Wolfinger, R. D. 2018. *SAS for Mixed Models: Introduction and Basic Applications.* SAS Institute.