Sommertagung AG Landwirtschaftliches Versuchswesen 29. – 30. Juni 2017

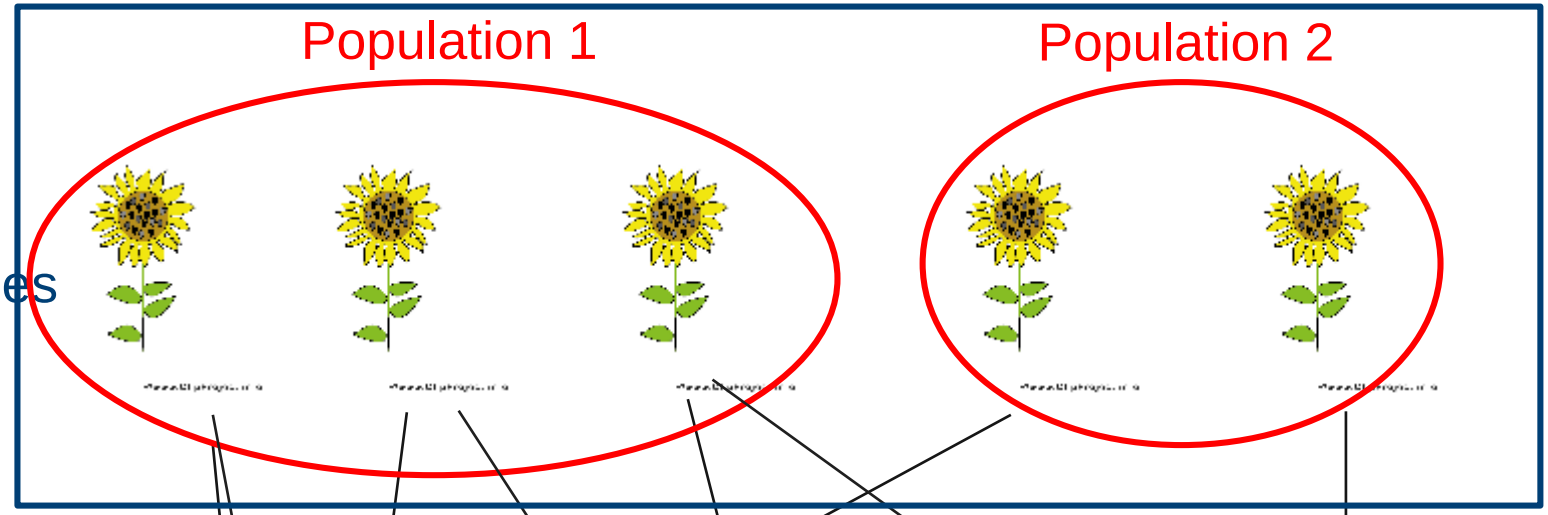
„Der Nutzen von teilwiederholten Versuchen bei Sonnenblumen“

29.06.2017

Christian Sponagel, Hans-Peter Piepho, Jens Möhring
(Uni Hohenheim)

Table of contents

1. Introduction



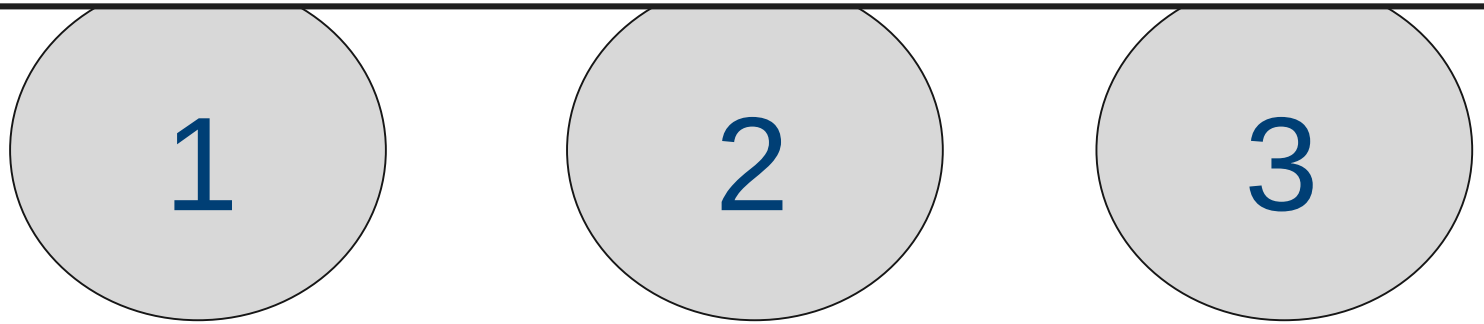
Genotypes

Material-group

→ **Focus is on the distribution of populations and genotypes to the locations**

Locations

With any design for trials



2. Material and methods

The dataset:

- 231 genotypes from 39 populations
- Population size: 1-11 genotypes
- A genotype is being tested in 2-5 locations
- Not every genotype is tested in every locations:

→ partly repeated experiment

- A population is in average tested in 5.59 locations
- The whole experiment extends across 10 locations
- → neither single genotypes, nor populations are being tested in every location

- Each genotype is assigned to one material-group
- Three material-groups are used

Table 2: The material-groups

Number of			Material-group
populations	genotypes	plots	
14	63	261	1
16	107	454	2
9	63	260	3

Table 1: Distribution of populations 1-11 to the locations

populations	locations										sum
	AR	MO	RT01	OD01	TV01	NS01	HFM	MZ01	SC01	VM01	
1		✖			✖			✖	✖	✖	5
2		✖			✖			✖	✖	✖	5
3		✖		✖	✖	✖	✖	✖	✖		7
4		✖				✖	✖	✖	✖	✖	6
5		✖				✖	✖	✖	✖	✖	6
6		✖					✖	✖	✖	✖	5
7	✖	✖	✖	✖	✖	✖		✖	✖	✖	9
8	✖	✖	✖	✖	✖				✖		6
9	✖	✖	✖	✖					✖		5
10		✖	✖	✖				✖	✖	✖	6
11		✖	✖	✖				✖	✖	✖	6

The analysis of the data

$$Genotype_{ij} = P_i + g_{ij}$$

$Genotype_{ij}$ = genetic value of the genotype *ij*

P_i = effect of the population *i*

g_{ij} = effect of the genotype *j* within population *i*

The analysis of the data

Two analyses:

- 1. Fixed genetic effects (2 stage analysis)
- 2. Random genetic effects (3 stage analysis)
- In both analyses weighting is done by the macro %one_big_omega (Damesa et al., 2017)

1. Fixed genetic effects

- : Interaction or nested effects
- **:** :Seperates fixed and random effects
- ____: Residual error

Stage 1 per location:
Block-model: **:EV + W•EV + U• W•EV + P•U•W•EV**
Treatment-model: **Gen(Pop*check)**
Complete model: **Gen(Pop*check) : EV + W•EV+ U• W•EV + P•U•W•EV**

Stage 2 across locations:
Block-model: **:Ort + Obs1**
Treatment-model: **Gen(Pop*check)**
Interactions: **: D1•Pop• Ort + check• Ort + D1•Gen(Pop)• Ort**
Complete model: **Gen(Pop*check): Ort + D1•Pop• Ort + check •Ort + D1•Gen(Pop)• Ort + Obs1**

EV: trial effect
W•EV: replication effect within a trial
U•W•EV: incomplete block effect within a replication
P•U•W•EV: plot effect within a incomplete block
Gen: genetic effect of a genotype
Pop: population effect
Check: effect of a check

Ort: effect of a location
D1: dummy variable with the value 0 for checks and 1 for the other genotypes
Obs1: variable with a level for each observation (residual error)

2. random genetic effects (BLUP‘s)

Stage 1 per location
Stage 2 across locations

Stage 3:

Complete model: Check **:** D1•Pop + D1•Gen(Pop) + Obs2

Gen: genetic effect of a genotype
Pop: population effect
Check: effect of a check
Obs2: variable with a level for each observation

2. random genetic effects (BLUP‘s)

- In order to use the covariance between genotypes of the same population, a Compound-Symmetry (CS) - structure is used for the variance-covariance-matrix in stage 3

$$var\,(Genotype_{ij}) = \sigma^2_{pop} + \sigma_{gen}{}^2 \qquad var\begin{pmatrix} Genotype_{ij} \\ Genotype_{ij'} \end{pmatrix} = \begin{pmatrix} \sigma^2_{pop} + \sigma^2_{gen_j} & \sigma^2_{pop} \\ \sigma^2_{pop} & \sigma^2_{pop} + \sigma^2_{gen_{j'}} \end{pmatrix}$$

$cov\,(Genotype_{ij}\,;\,Genotype_{ij'}) = \sigma^2_{pop}$

i = Index of the population i

j = Index of the genotype j

Evaluation criteria

- The experimental designs are evaluated by the average standard error over all pairwise differences ($\overline{s.e.d.}$)

- In the analysis with fixed genetic effects this is done by the option pdiff in SAS for the lsmeans of the genotypes

- In the analysis with random genetic effects this is done by using the inverse of the mixed-model-equations (MMEQ)

Calculation of the average sed from the MMEQ

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + D^{-1} \end{bmatrix}^{-1} = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}.$$

X: design matrix for fixed effects
R: variance-covariance of the residual error
Z: designmatrix for random effects
D: variance-covariance matrix of random effects

C_{22} : variance-covariance matrix of the residual error for the estimations of the random effects

- The matrix C_{22} can be get by the option MmeqSol in proc mixed in SAS

Calculation of the average s.e.d. from the MMEQ

- To calculate the average s.e.d. from the matrix C_{22} a macro (Ould Estaghvirou et. al, 2013) is used

$$\bar{v}_D = \frac{2}{n(n-1)}\left[n \times trace(V_M) - 1_n V_M 1_n^T\right]$$

$\bar{v}_D$ = average variance of all pairwise differences

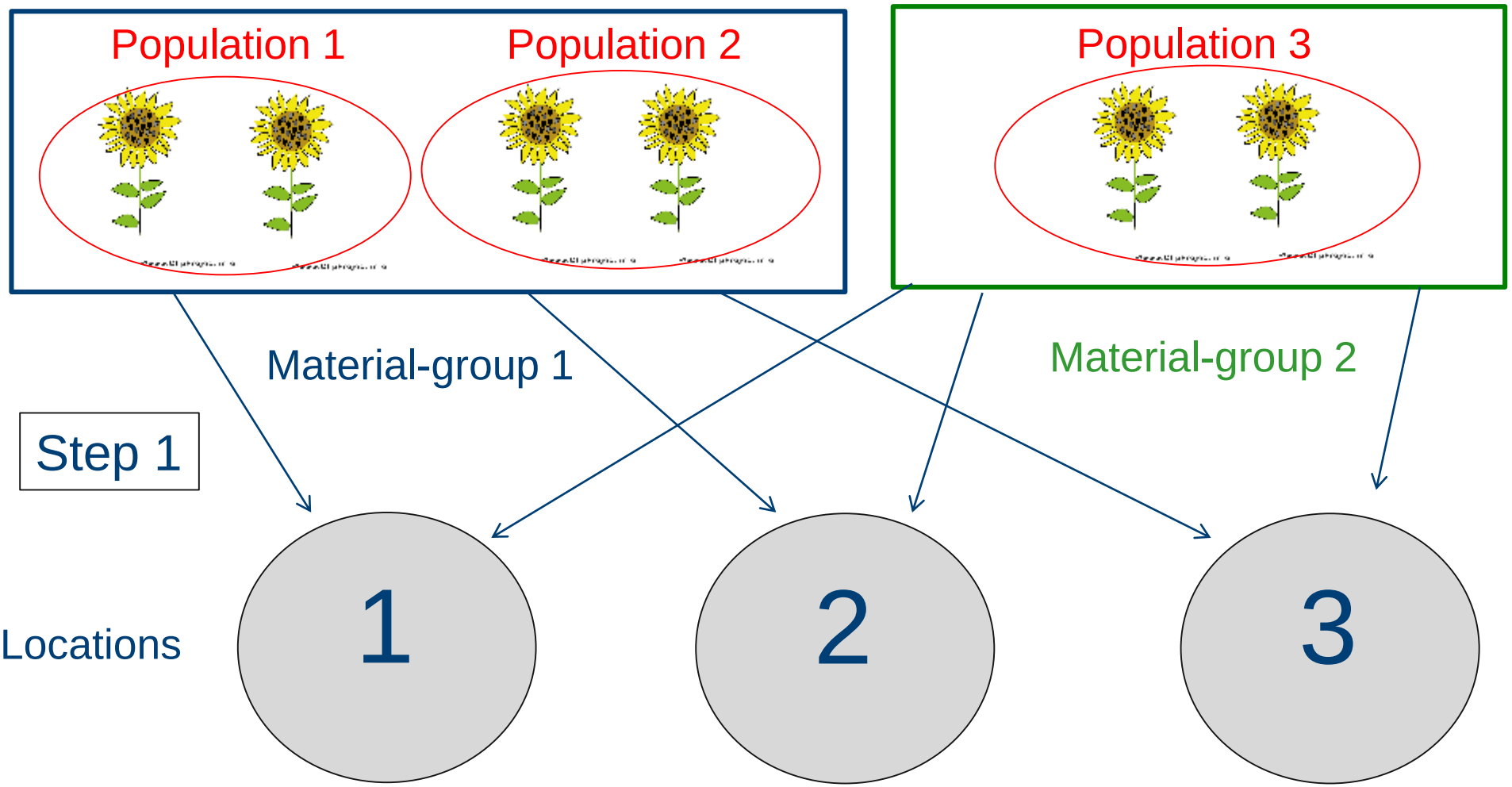
n = number of parameters that are compared

$trace(\boldsymbol{V_M})$ = trace of the varianz-covariance-matrix $\boldsymbol{V_M}$

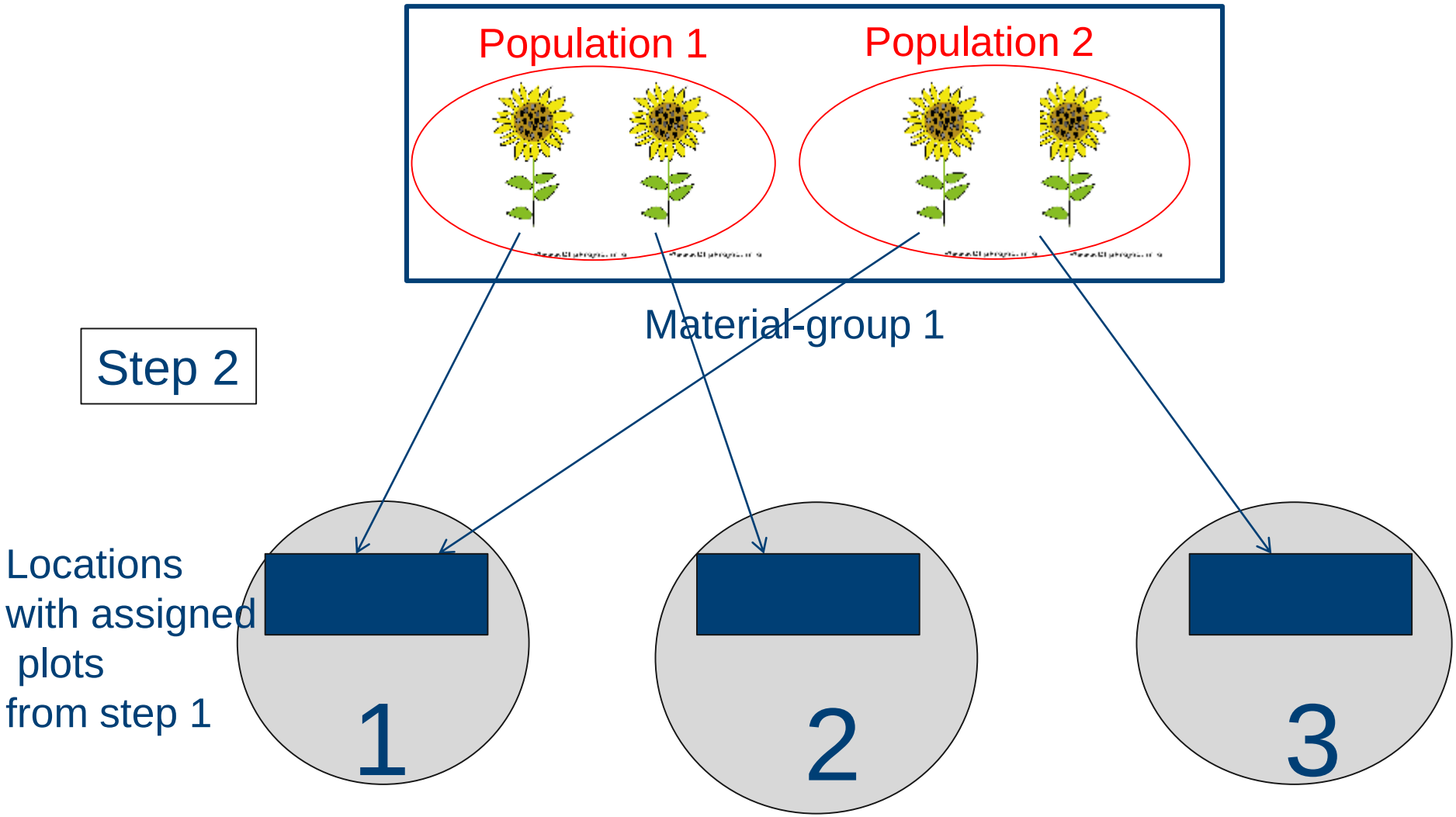
$1_n \boldsymbol{V_M} 1_n^T$ = sum of all elements in the matrix $\boldsymbol{V_M}$

$$\overline{sed} = \sqrt{\overline{var}}$$

Creating new experimental designs (scenario 1-3)



Creating new experimental designs (scenario 1-3)



Creating new experimental designs (scenario 1-3)

- The optex procedure in SAS is used:
- Genotypes and populations have to be distributed optimally to the locations, i.e. testing the populations in as many locations as possible

- Step 1: Distributing the material-groups to the locations
- → each material-group gets a capacity of plots in each location, according the number of populations and total number of plots that the group includes

Creating new experimental designs (scenario 1-3)

Step 2:

- Now the genotypes within each material-group are distributed to the assigned capacity of plots in each location
- But also the populations have to be distributed optimally
- So treatments are both genotype and population
- Blocks are the locations
- In the blocks-statement also sub-structures are possible and with the prior option a optimal distribution of the treatments to the blocks and “sub-blocks“ is possible

Creating new experimental designs (scenario 1-3)

- But in the block-statement there are no substructures, there is just the location
- In the treatment-statement there is no prior option available, but there are substructures: Populations and genotypes within a population

→ So treatments and blocks are changed (John & Williams, 1995)

Treatments: locations
Blocks: populations
Subblocks: genotypes within a population

- Now the belonging of genotypes to populations is taken into account
- Populations are now tested in as many locations as possible

Creating new experimental designs (scenario 1-3)

- Also in step 1 the treatments and blocks are changed:

Treatments: locations

Blocks: material-group

Subblock: Population

- For the selection of the experimental design the optex-procedure uses the D- and A-efficiency
- The D- and A-efficiency can also be calculated for a existing design (option init=chain and iter=0 in SAS)

Fixing of the variance-components

- In all stages of the analysis of the dataset there are variance-components for the random effects
- These variance-components are fixed and used to calculate the average s.e.d. in the scenarios 0-3

Table 3: Comparison of the scenarios

	Number of plots in the whole experiment	Number of Replicates of each genotype	Number of locations	Size of locations	Distribution of genotypes and populations
Scenario 0			10		
Scenario 1			10		
Scenario 2			20	½ Size	
Scenario 3			10	Equal size	

Scenario 0 (real dataset)

- For the real dataset the average s.e.d. is calculated again with the fixed variance-components
- All observations are set to the value 1 (any value can be used)
- Now also observations with missing values are included
- Scenario 0 is used to compare the existing experimental-design with the designs of the scenarios 1-3

Table 4: Comparison of the scenarios

Experimental-design	Average number of	
	Locations a genotype is being tested in	Genotypes from one population in one location
Scenario 0	**5.59**	**4.47**
Scenario 1	**9.21**	**2.62**
Scenario 2	**15.74**	**1.47**
Scenario 3	**9.62**	**2.54**

Table 5: Variance-components in stage 3

effect	variance in $(kg/plot)^2$
Pop	0.000530
Gen(Pop)	0.007971

→ population-variance is rather small

Table 6: Average s.e.d. in the scenarios

Random genetic effects			Fixed genetic effects		
	Average sed	Change to scenario 0		Average sed	Change to scenario 0
Scenario 0	0.082	+/- 0 %	Scenario 0	0.228	+/- 0 %
Scenario 1	0.070	- 14 %	Scenario 1	0.224	- 1.5 %
Scenario 2	0.065	- 20 %	Scenario 2	0.227	- 0.4 %
Scenario 3	0.069	- 16 %	Scenario 3	0.225	- 1.5 %

- Greater impact of the new experimental desings when genetic effects are random
- Smallest s.e.d. in scenario 2 with random genetic effects
- Fixed genetic effects: smallest s.e.d. in scenario 1

- Greater impact on the s.e.d. when genetic effects are random
- Population-effect (covariance between genotypes of the same population) is not used when genetic effects are fixed

→ More precise estimations of the population-effects have an impact on the precision of the genetic value of a genotype

Random genetic effects			Fixed genetic effects		
	Average sed	Change to scenario 0		Average sed	Change to scenario 0
Scenario 0	0.082	+/- 0 %	Scenario 0	0.228	+/- 0 %
Scenario 1	0.070	- 14 %	Scenario 1	0.224	- 1.5 %
Scenario 2	0.065	- 20 %	Scenario 2	0.227	- 0.4 %
Scenario 3	0.069	- 16 %	Scenario 3	0.225	- 1.5 %

- The entire number of plots is equal in all scenarios, also with 20 locations
- It remains to be seen whether the use of 20 locations is also economic
- Scenario 1 is probably the most realistic scenario

- Same number of plots for each genotype would probably be beneficial

- To compare two genotypes the s.e.d. is important,
  so the s.e.d. was used to evaluate the scenarios

4. Conclusion

- It is beneficial to test the populations in more locations, instead of testing genotypes of the same population together in a location

→ Partly repeated experiments make sense !

Damesa, T., Möhring J., Piepho H.-P., 2017, One Step at a Time: Stage-Wise Analysis of Series of Experiments in: Agronomy Journal (in revision)

Piepho, H.-P., 2015, Biometrie, Vorlesungsskript, Hohenheim

Williams, E.R. & John, J.A. , 1995, Cyclic and Computer Generated Designs, Chapman & Hall, London (Erstauflage 1987)

www.ClipProject.info, http://www.clipartsfree.de/images/joomgallery/originals/bauernhof_4/sonnenblume_bild_clipart_kostenlos_20130706_1744935743.jpg (Abruf am 26.01.2017)

https://bmcgenomics.biomedcentral.com/articles/10.1186/1471-2164-14-860 (Abruf am 07.12.2016)
Ould Estaghvirou S.B., O Ogutu J., Schulz-Streeck T., Knaak C., Ouzunova M., Gordillo A., Piepho H.-P., 2013, Evaluation of approaches for estimating the accuracy of genomic prediction in plant breeding, in: BMC Genomics 14:860 (2013), licensee BioMed Central Ltd.

https://support.sas.com/documentation/onlinedoc/qc/132/optex.pdf (Abruf am 01.12.2016),
SAS Institute Inc. 2014. SAS/QC® 13.2 User´s Guide. Cary, NC: SAS Institute Inc.

Vielen Dank für Ihre Aufmerksamkeit!