

Effects of data reduction on trend estimation from German registration trials

04.07.2022

Jens Hartung, Friedrich Laidig, Hans-Peter Piepho



Inhalt

- Cultivar evaluation trials for trend estimation
- Missing not at random
- Simulation approach
- Results
- Interpretation
- A few words about relevance



Selection of breeders

Submission

Post-breeding population

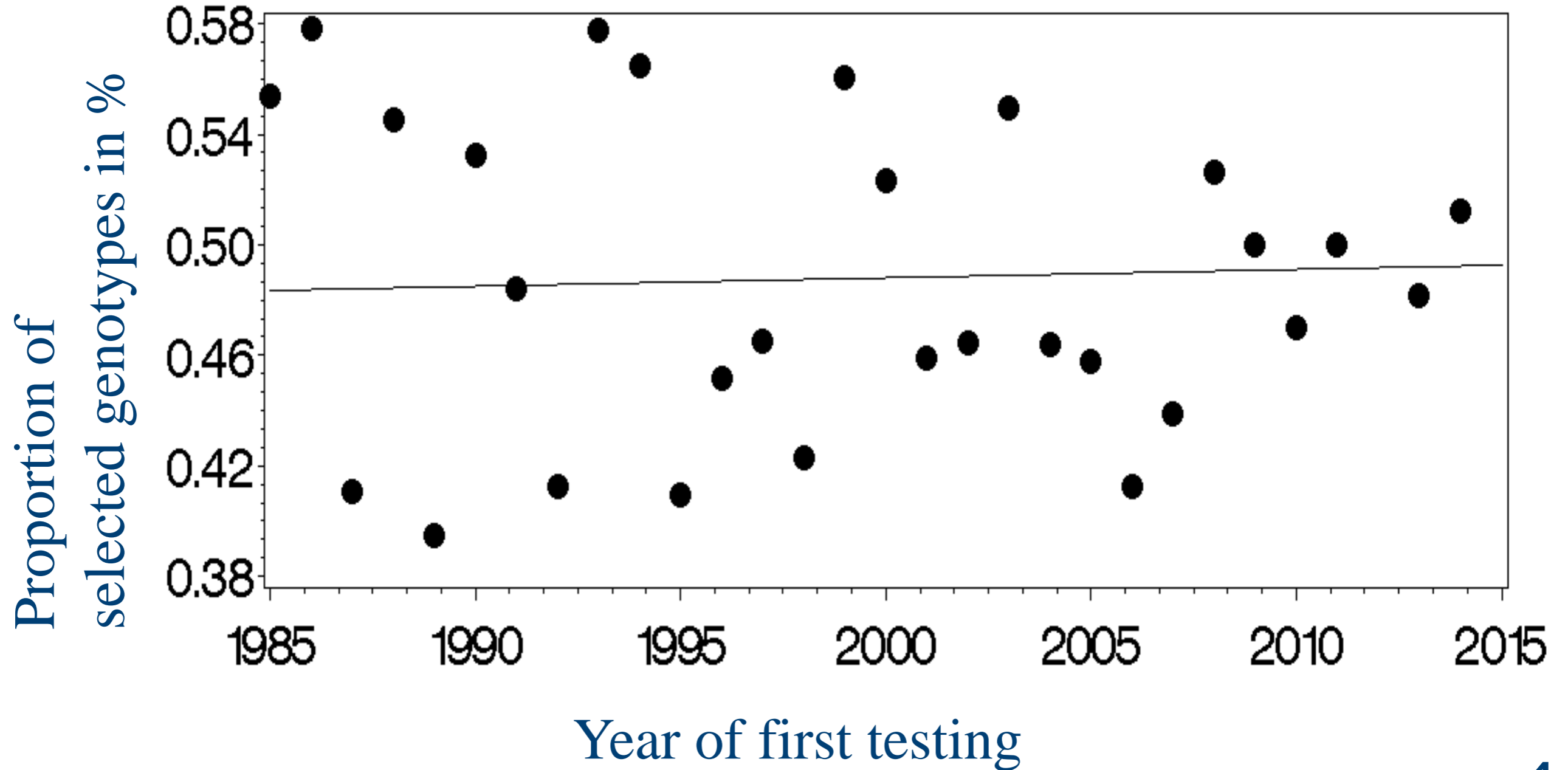
Cultivar evaluation trials (Wertprüfung)

Registration

Post-registration population

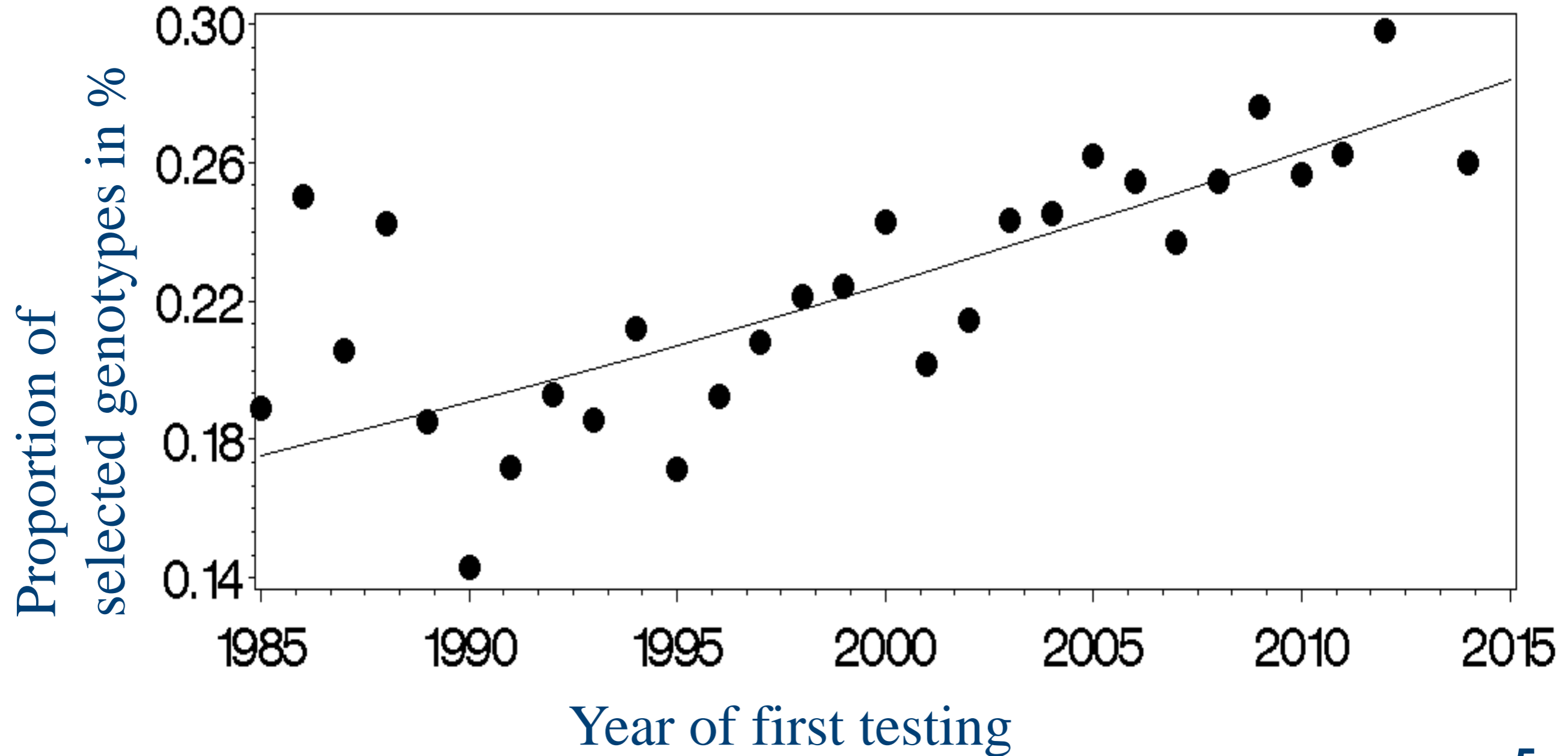


Proportion of genotypes tested in 2nd year





Proportion of genotypes tested in 3rd year





Selection of breeders

Submission

Post-breeding population

Cultivar evaluation trials

Registration

Post-registration population



Selection of breeders

Submission

Post-breeding population

Cultivar evaluation trials

Registration

Post-registration population



Selection of breeders

Submission

Post-breeding population

Cultivar evaluation trials

Registration

Post-registration population



Model

$$\bar{y}_{ijkl} = \mu + \gamma t_k + Z_k + L_j + \beta r_i + H_i + (LY)_{jk} + (LYT)_{jkl} + (GL)_{ij} + (GY)_{ik} + (GLY)_{ijk} + e_{ijkl}.$$

G=Genotype

L/T=Location/Trial

Y=Year



Model

$$\bar{y}_{ijkl} = \mu + \gamma t_k + Z_k + L_j + \beta r_i + H_i + (LY)_{jk} + (LYT)_{jkl} + (GL)_{ij} + (GY)_{ik} + (GLY)_{ijk} + e_{ijkl}.$$

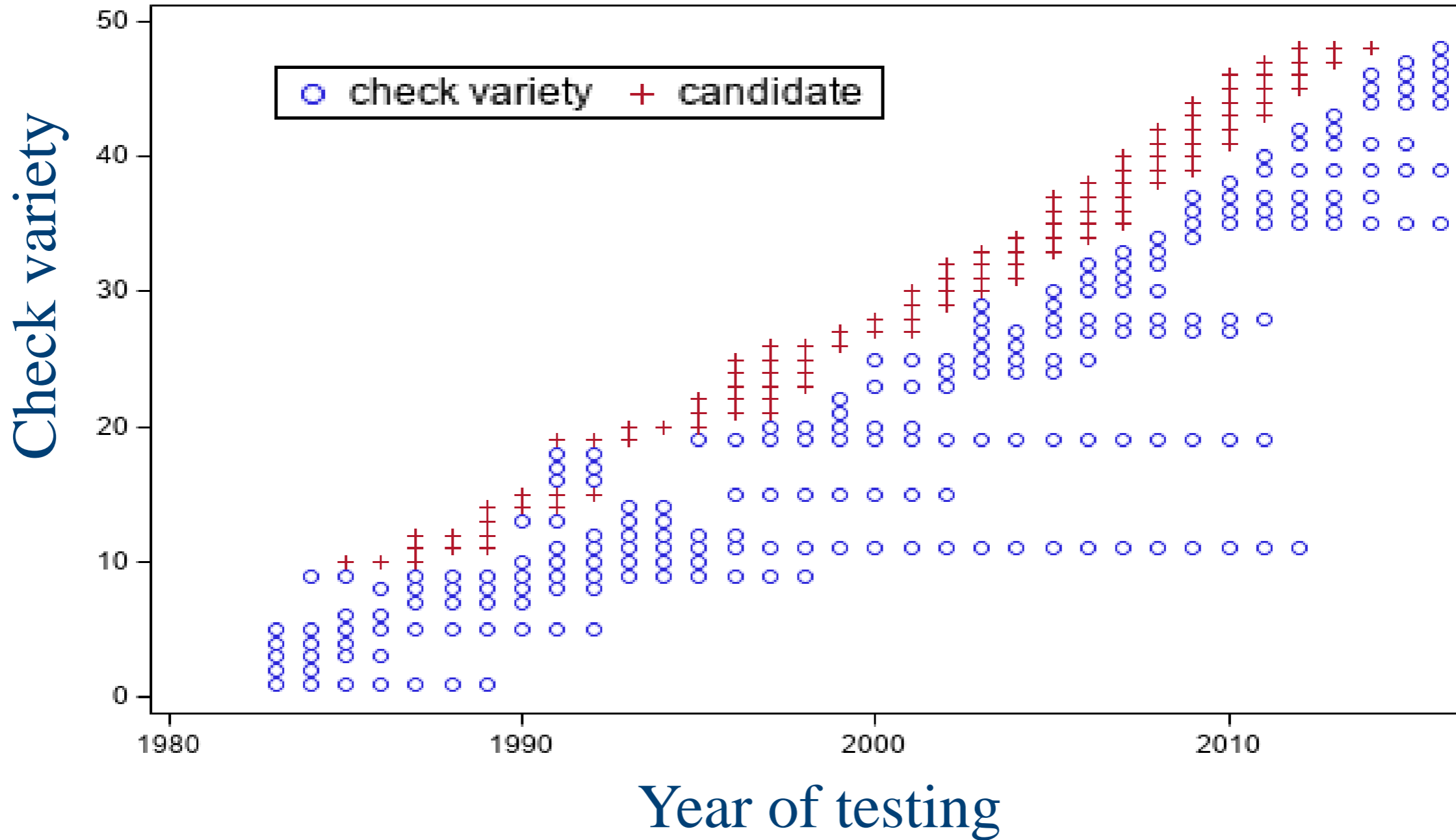
G=Genotype

L/T=Location/Trial

Y=Year

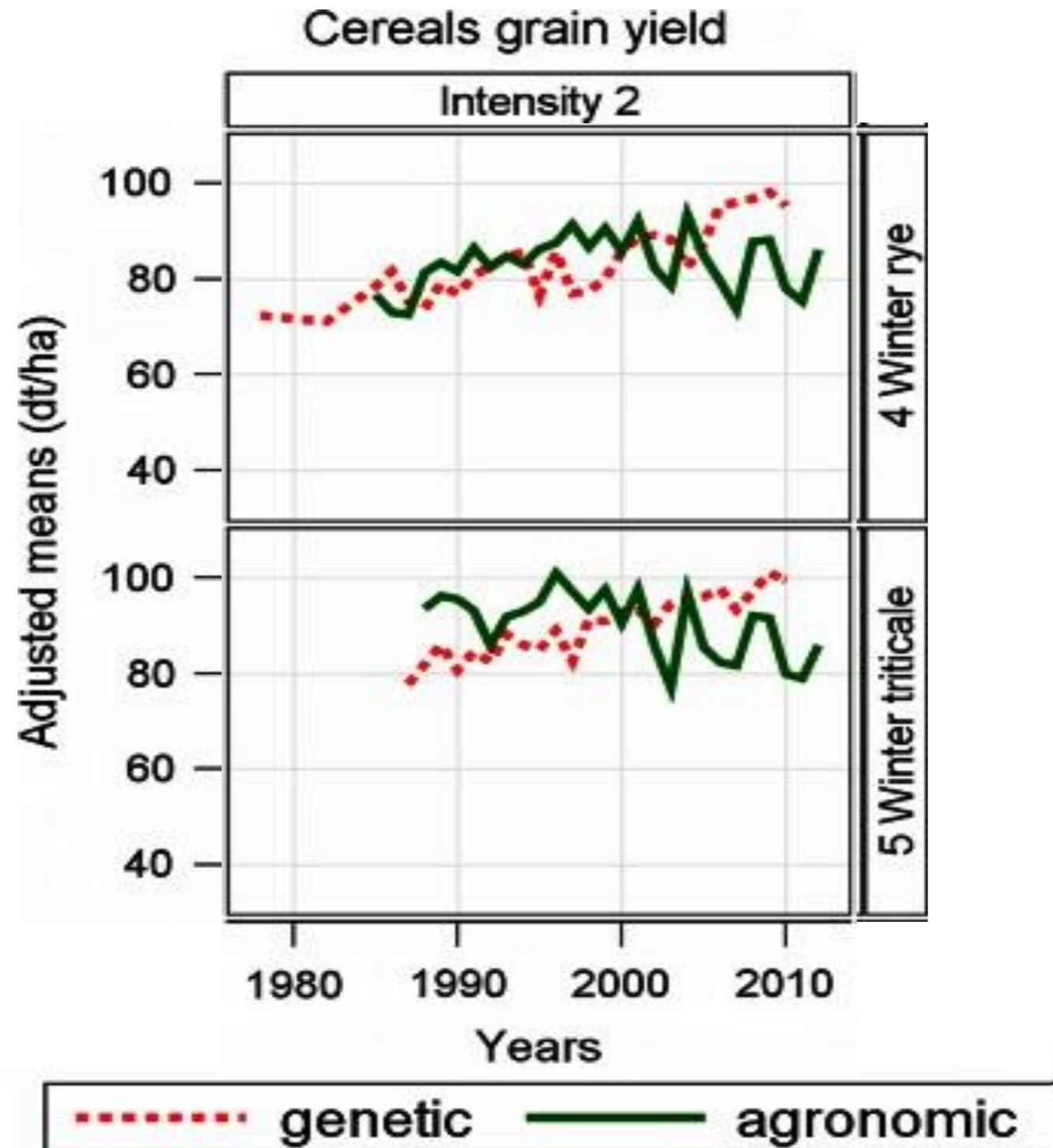


Occurrence of check varieties





Parameter	Estimate for dataset in dt ² ha ⁻²		
	Complete data	Reduced data	Check varieties only
Number of observations	77802	42066	13010
Number of genotypes	2912	702	48
Variance component			
Genotype	13.3	12.1	11.2
Genotype-by-year	3.2	2.9	4.3
Genotype-by-location	2.3	2.2	2.4
Genotype-by-year-by-location	9.6	9.0	9.2
Error	9.4	9.4	9.2
Trend	Estimate (standard error) for dataset in dt ha ⁻¹ year ⁻¹		
Genetic	0.56 (0.03)	0.55 (0.03)	0.56 (0.05)
Non-genetic	0.17 (0.10)	0.13 (0.10)	0.15 (0.10)






Simulation

- Mimic structure of cultivar evaluation trials
- 1622 trials, 2912 genotypes
- 77802 observations (genotype-by-trial means)
- No trends
- Complete data
- Reduced data







Results

Approach	Genotype variance in $dt^2 \text{ ha}^{-2} \text{ a}^{-2}$		Trend in $dt \text{ ha}^{-1} \text{ a}^{-1}$		
	simulated	estimated	genetic	non-genetic	
Complete data	13.3	13.3	0.0004	-0.0002	






Results

Approach	Genotype variance in $dt^2 \text{ ha}^{-2} \text{ a}^{-2}$		Trend in $dt \text{ ha}^{-1} \text{ a}^{-1}$		
	simulated	estimated	genetic	non-genetic	
Complete data	13.3	13.3	0.0004	-0.0002	
Reduced data	13.3	5.9	0.1153	-0.1057	







Results

Approach	Genotype variance in $dt^2 \text{ ha}^{-2} \text{ a}^{-2}$		Trend in $dt \text{ ha}^{-1} \text{ a}^{-1}$		
	simulated	estimated	genetic	non-genetic	
Complete data	13.3	13.3	0.0004	-0.0002	
Reduced data	13.3	5.9	0.1153	-0.1057	
Reduced data, random selection	13.3	13.3	0.0003	0.0001	



Results

Approach	Genotype variance in $dt^2 \text{ ha}^{-2} \text{ a}^{-2}$		Trend in $dt \text{ ha}^{-1} \text{ a}^{-1}$		
	simulated	estimated	genetic	non-genetic	
Complete data	13.3	13.3	0.0004	-0.0002	
Reduced data	13.3	5.9	0.1153	-0.1057	
Reduced data, random selection	13.3	13.3	0.0003	0.0001	
Reduced data, large G×J variance	13.3	5.4	0.6403	-0.6674	



Relevance

Genetic trend in real VCU data:	$0.5532 \text{ dt ha}^{-1} \text{ a}^{-1}$
Bias found in simulation	$0.1153 \text{ dt ha}^{-1} \text{ a}^{-1}$
Genotype variance from complete data:	$13.3 \text{ dt}^2 \text{ ha}^{-2} \text{ a}^{-2}$
Genotype variance from reduced data:	
• Real VCU data:	$12.1 \text{ dt}^2 \text{ ha}^{-2} \text{ a}^{-2}$
• Simulation:	$5.9 \text{ dt}^2 \text{ ha}^{-2} \text{ a}^{-2}$



Conclusion (1/2)

Post-breeding Population

Cultivar evaluation trials

Post-registration Population



Conclusion (2/2)

- Use of complete data for estimation of post-breeding trends
- Reduced data resulted in small bias of post-breeding trends
- Use of post-registration trials (in Germany LSV)



References:

Laidig F, Piepho H-P, Drobek T, Meyer U (2014) Genetic and non-genetic long-term trends of 12 different crops in German official variety performance trials and on-farm yield trends. *Theor Appl Genet* 127:2599–2617. doi:101007/ s00122-014-2402-z

Piepho H-P, Möhring J (2006) Selection in cultivar trials: Is it ignorable? *Crop Sci* 46:192–201. doi:102135/cropsci200504-0038

Hartung J, Laidig F, Piepho H-P (2022) Effects of systematic data reduction on trend estimation from German registration trials. In preparation.



Parameter	Estimate for dataset in dt ² ha ⁻²		
	Complete data	Reduced data	Check varieties only
Variance component			
Year	25.1314	23.8879	22.9941
Location	53.1506	53.4990	54.0750
Year-by-location	74.1434	73.9777	72.8957
Year-by-location-by-trial	8.8293	9.0251	9.5473
Genotype	13.2716	12.0976	11.1719
Genotype-by-year	3.1598	2.9439	4.2967
Genotype-by-location	2.2527	2.1570	2.3762
Genotype-by-year-by-location	9.6216	9.0116	9.1895
Error	9.3552	9.4141	9.1666
Trend	Estimate (standard error) for dataset in dt ha ⁻¹ year ⁻¹		
Genetic	0.5588 (0.02912)	0.5532 (0.03068)	0.5551 (0.05029)
Non-genetic	0.1743 (0.09960)	0.1258 (0.09762)	0.1528 (0.09888)