



Universität Hohenheim (340), D-70599 Stuttgart

Fruwirthstr. 23  
70599 Stuttgart  
Telefon: (0711) 459 – 22386  
Telefax: (0711) 459 – 24345  
e-mail: piepho@uni-hohenheim.de  
URL: www.uni-hohenheim.de

Hohenheim, September 2020

## **Master thesis: Method comparison for analysis of multi-year field trial data for selection decisions**

### **Background**

In a plant breeding program, field trials are the main tool for germplasm enhancement. Selection candidates are evaluated across multiple locations in several consecutive years to estimate their breeding values. This requires elaborated (mixed) linear models to correct for macro and micro environmental effects. The gold standard is a one-stage analysis where all available data is analyzed in one step. However, this leads to high computational burdens. At present, computational capacities and algorithms commonly available to breeders don't allow for one-stage analysis of multi-year breeding data in a timely manner. Often approximate solutions like two-stage or three-stage analyses are used instead, where adjusted means for genotypes computed in one stage are carried forward to the next stage. Stage-wise analyses may be performed with or without weighting based on error variances or use of covariances among genotypes and environments. In two-stage analyses, data is analyzed per year in the first stage and then pooled across years in a second stage. Two-stage analyses are still relatively complex and computationally demanding. Therefore, different variants of three-stage analysis are in common usage, for instance, by analyzing data across trials and testers per location in the first stage, then analyzing across locations in the second stage, and finally pooled across years in the third stage. Without weighting, in two- and three-stage analyses the information on the number of data points per selection candidate and the precision of adjusted means from previous stages is lost, leading to a potential loss in accuracy for selection decisions.

With the advent of cost-effective genotyping platforms, obtaining marker profiles and predicting genomic estimated breeding values (GEBV) are an additional selection criterion. GEBVs are an attractive framework to incorporate data from related selection candidates from multi-year field trials. However, the integration into the statistical model introduces additional challenges. The marker information can be modelled either in a one-stage analysis (on a plot level) or can be incorporated in the second stage of a two-stage analysis (on a genotype x environment level) or can be used in a third stage (on a per genotype level). There is as yet little evidence to judge how these alternatives compare and which option should be preferred for routine use in breeding programs.

### **Project**

Within this project, experimental data from five years (2016 to 2020) from both the seed and pollen-parent gene pool of the KWS rye breeding program are available. Per year, the data are

composed of yield trials from 3 selection stages for general combining ability (GCA1 – GCA3) in combination with two to four testers evaluated in 7 to 14 locations in two replicates per location. The goal is to compare the breeding values of different models for phenotypic data analysis (two- and three-stage weighted, two- and three-stage unweighted, two- and three-stage with use of genotype and environment covariances) with respect to the resulting one-stage breeding values. The relative merit of different modelling approaches can be assessed by means of the correlation and the standard error of estimates with respect to the one-stage analysis as well as an empirical validation with breeding values estimates from a future year. An important output is a workflow suggestion for a routine evaluation of multi-year field trial data.

In addition, the integration of marker information to obtain GEBVs shall be investigated and potential different models compared. For this work package, the comparison criterion is the predictive ability to predict future years. To evaluate this, marker data for 10,000 SNP markers and around 6500 genotypes (rye lines) per gene pool from 5 years is available.

### **Partners**

The thesis will be jointly supervised by Prof. Dr. Piepho (Institute of Crop Science, Biostatistics 340 c, Univ. Hohenheim, hans-peter.piepho@uni-hohenheim.de) and KWS (Angela Bernal, Cereals Breeding Technology, angela.bernal@kws.com).

### **Time frame**

The data for this Master Thesis can be made available from the 1<sup>st</sup> of November 2020. Envisioned completion of the Master Thesis is summer 2021.

### **Reading**

Bernal-Vasquez, A.M., Gordillo, A., Schmidt, M., Piepho, H.P. (2017): Genomic prediction in early selection stages using multi-year data in a hybrid rye breeding program. *BMC Genetics* **18**, 51.  
Damesa, T.M., Hartung, J., Gowda, M., Beyene, Y., Das, B., Semagn, K., Piepho, H.P. (2019): Comparison of weighted and unweighted stage-wise analysis for genome-wide association studies and genomic selection. *Crop Science* **59**, 2572-2584.