

# Rank Aggregation in Phenotypic Selection

Shuxiao Chen

## 1 Introduction

In phenotypic selection, we usually want to measure some traits which are difficult for researchers to measure quantitatively, like the taste and the ease of cooking. In this situation, we may adopt “crowdsourcing” methodology to incorporate farmers’ wisdom. The set-up has been described by van Etten in [10], where each farmer receives three varieties and ranks the varieties according to some trait from the best to the worst. This *triadic comparison of technologies* (tricot) is feasible since it only requires farmers to distinguish the best and the worst, so that the task is easy enough to complete and the accuracy can be assured. After receiving all the partial rankings from farmers, we may aggregate them into an overall consensus ranking so that the overall ranking can reflect the level of the traits we ask farmers to measure. Meanwhile, what we have before the actual field trials is the genotype data. Since the additive relationship matrix can be derived from the genotype data and this matrix measures the similarity among all the varieties, we may incorporate the information in this matrix into the rank aggregation process.

## 2 Preliminaries

In this section we introduce some mathematical notations that will be used later. Suppose we have  $m$  varieties we want to measure and  $n$  farmers available. For simplicity varieties are labeled from 1 to  $m$  and farmers are labeled from 1 to  $n$ . We denote the partial ranking given by farmer  $i$  by  $\sigma^{(i)} = \sigma_1^{(i)} \sigma_2^{(i)} \dots \sigma_k^{(i)}$ , where  $\sigma_j^{(i)}$  is the variety ranked in  $j$ -th place by farmer  $i$  and  $k = 3$  in our case. The data we observe is a collection of rankings  $\mathcal{R} = \{\sigma^{(1)}, \dots, \sigma^{(n)}\}$ . Let  $i \succ_\sigma j$  denote “variety  $i$  is ranked higher than  $j$  in the ranking  $\sigma$ ” and let  $r_j^{(i)}$  be the rank of variety  $j$  from farmer  $i$ . Meanwhile we let  $\mathcal{P}(\sigma)$  be the probability distribution that assigns probability on a specific ranking  $\sigma$ .  $\mathcal{P}$  is then referred as the “ranking model”. Lastly we let the  $m \times m$  matrix  $K$  denote the additive relationship matrix derived from the genotype data.

## 3 Models

### 3.1 A Linear Model Approach (LM)

The rank  $r_j^{(i)}$  have two components, namely the actual competence of variety  $j$  and the turbulence from farmer  $i$ . In the overall ranking, we want to extract as much information as possible in the actual competence of varieties. And disregarding the discreteness of ranks

gives rise to the following linear model:

$$\mathbf{r} = \boldsymbol{\mu} + \mathbf{v} + \mathbf{e},$$

where  $\mathbf{r}$  is the ranks,  $\boldsymbol{\mu}$  is the mean vector,  $\mathbf{v}$  is the random effect from varieties and  $\mathbf{e}$  is the random error. In order to incorporate genotypic information, we specify the covariance structure of  $\mathbf{v}$  to be

$$\mathbf{v} \sim \mathcal{N}(\mathbf{0}, \sigma_v^2 K).$$

The overall ranking is then derived by ranking the *best linear unbiased prediction* (BLUP) of  $\mathbf{v}$  in the increasing order.

### 3.2 Thurstone Model (TH)

Thurstone proposed a ranking process in [9] where the ranking  $\sigma^{(k)}$  of  $m$  varieties (full ranking) given by farmer  $k$  is determined by the order of a set of latent values  $\mathbf{y}^{(k)} = (y_1^{(k)}, \dots, y_m^{(k)})$  and

$$y_i^{(k)} \sim \mathcal{N}(s_i, \gamma_i^2)$$

independently, where  $s_i$  is referred as the *score* of variety  $i$  and  $\gamma_i^2$  measures the variance of farmers' measurements on  $i$ . When fitting the model, we let  $\gamma_i^2 = 1$  for all  $i$  in order to let the model be identifiable. Then for each pairwise comparison, we have

$$y_i^{(k)} - y_j^{(k)} \sim \mathcal{N}(s_i - s_j, 2).$$

Let  $\mathbf{s}$  be the vector of scores for all varieties. If we assume independence among all pairwise comparisons, for farmer  $k$  we have

$$\mathcal{P}(\sigma^{(k)} | \mathbf{s}) = \prod_{i \succ_{\sigma} j} \Phi\left(\frac{s_i - s_j}{\sqrt{2}}\right),$$

where  $\Phi$  is the distribution function of the standard normal distribution.

In order to account for the heterogeneity among farmers, we adopt the method of Raman and Joachims in [8], where there is one extra parameter  $\boldsymbol{\eta} = (\eta^{(1)}, \dots, \eta^{(n)})^T$ , which we call the *adherence* parameter. The adherence  $\eta^{(k)}$  measures farmer  $k$ 's adherence to the overall ranking. Then the latent values are distributed as

$$y_i^{(k)} \sim \mathcal{N}(s_i, \gamma_i^2 / \eta^{(k)}),$$

and the ranking model can be derived as

$$\mathcal{P}(\sigma^{(k)} | \mathbf{s}, \boldsymbol{\eta}) = \prod_{i \succ_{\sigma^{(k)}} j} \Phi\left(\frac{\sqrt{\eta^{(k)}}(s_i - s_j)}{\sqrt{2}}\right).$$

Assuming independence among all the farmers, the likelihood function is given by

$$\mathbb{P}(\mathcal{R} | \mathbf{s}, \boldsymbol{\eta}) = \prod_{k=1}^n \prod_{i \succ_{\sigma^{(k)}} j} \Phi\left(\frac{\sqrt{\eta^{(k)}}(s_i - s_j)}{\sqrt{2}}\right).$$

In order to incorporate genotypic information, we use the maximum a priori (MAP) estimator. We assign a normal prior to the score  $\mathbf{s}$ :

$$\mathbf{s} \sim \mathcal{N}(\boldsymbol{\mu}, K),$$

where  $\boldsymbol{\mu} = (1, \dots, 1)^T$  and  $K$  is the additive relationship matrix. Meanwhile we assign a flat, improper prior to the adherence  $\boldsymbol{\eta}$ . Then with some algebra we can derive the log-posterior:

$$\ell(\mathbf{s}, \boldsymbol{\eta} | \mathcal{R}) \propto \text{cons.} + \sum_{k=1}^n \left( -\frac{1}{2n} (\mathbf{s} - \boldsymbol{\mu})^T K^{-1} (\mathbf{s} - \boldsymbol{\mu}) + \sum_{i \succ_{\sigma^{(k)}} j} \log[\Phi(\frac{\sqrt{\eta^{(k)}}(s_i - s_j)}{\sqrt{2}})] \right).$$

Since the log-posterior is convex with respect to  $\mathbf{s}$  according to [8], we can use stochastic gradient descent to optimize it.

### 3.3 Bradley-Terry Model (BT)

This model was proposed by Bradley and Terry [1] and has various extensions. This model is also parameterized by the *score* parameter  $\mathbf{s}$  and we have

$$\mathbb{P}(i \succ_{\sigma^{(k)}} j) = \frac{1}{1 + \exp\{-(s_i - s_j)\}}.$$

Assuming independence among all pairwise comparisons and adopting the *adherence* parameter as in [8], the ranking model is given by:

$$\mathcal{P}(\sigma^{(k)} | \mathbf{s}, \boldsymbol{\eta}) = \prod_{i \succ_{\sigma^{(k)}} j} \frac{1}{1 + e^{-\eta^{(k)}(s_i - s_j)}}.$$

Assuming independence among all the farmers, the likelihood is given by:

$$\mathbb{P}(\mathcal{R} | \mathbf{s}, \boldsymbol{\eta}) = \prod_{k=1}^n \prod_{i \succ_{\sigma^{(k)}} j} \frac{1}{1 + e^{-\eta^{(k)}(s_i - s_j)}}.$$

Again we use MAP estimator to incorporate genotypic information. We assign a normal prior to  $\mathbf{s}$ :

$$\mathbf{s} \sim \mathcal{N}(\boldsymbol{\mu}, K)$$

where  $\boldsymbol{\mu} = (1, \dots, 1)^T$  and  $K$  is the additive relationship matrix, and a flat, improper prior to  $\boldsymbol{\eta}$ . With some algebra we can derive the log-posterior:

$$\ell(\mathbf{s}, \boldsymbol{\eta} | \mathcal{R}) \propto \text{cons.} + \sum_{k=1}^n \left( -\frac{1}{2n} (\mathbf{s} - \boldsymbol{\mu})^T K^{-1} (\mathbf{s} - \boldsymbol{\mu}) - \sum_{i \succ_{\sigma^{(k)}} j} \log(1 + e^{-\eta^{(k)}(s_i - s_j)}) \right).$$

Again it has been proved in [8] that the log-posterior is convex with respect to  $\mathbf{s}$ , so we use stochastic gradient descent to optimize it.

### 3.4 Plackett-Luce Model (PL)

This model was proposed independently by Plackett [7] and Luce [5]. It is the generalization of Bradley-Terry model, allowing the comparison of more than 2 varieties concurrently, and it is also parameterized by the score parameter  $\mathbf{s}$ . According to this model, the probability of  $i$  ranked higher than  $j$  by farmer  $k$  is

$$\mathbb{P}(i \succ_{\sigma^{(k)}} j) = \frac{\exp\{s_i\}}{\exp\{s_i\} + \sum_{i \succ_{\sigma^{(k)}} j} \exp\{s_j\}}.$$

Assuming independence among all pairwise comparisons and adopting the *adherence* parameter as in [8], the ranking model is given by:

$$\mathcal{P}(\sigma^{(k)} | \mathbf{s}, \boldsymbol{\eta}) = \prod_{i=1}^m \frac{1}{1 + \sum_{i \succ_{\sigma^{(k)}} j} e^{-\eta^{(k)}(s_i - s_j)}}.$$

Assuming independence among all the farmers, the likelihood is given by:

$$\mathbb{P}(\mathcal{R} | \mathbf{s}, \boldsymbol{\eta}) = \prod_{k=1}^n \prod_{i=1}^m \frac{1}{1 + \sum_{i \succ_{\sigma^{(k)}} j} e^{-\eta^{(k)}(s_i - s_j)}}.$$

Again we use MAP estimator to incorporate genotypic information. We assign a normal prior to  $\mathbf{s}$ :

$$\mathbf{s} \sim \mathcal{N}(\boldsymbol{\mu}, K)$$

where  $\boldsymbol{\mu} = (1, \dots, 1)^T$  and  $K$  is the additive relationship matrix, and a flat, improper prior to  $\boldsymbol{\eta}$ . With some algebra we can derive the log-posterior:

$$\ell(\mathbf{s}, \boldsymbol{\eta} | \mathcal{R}) \propto \text{cons.} + \sum_{k=1}^n \left( -\frac{1}{2n} (\mathbf{s} - \boldsymbol{\mu})^T K^{-1} (\mathbf{s} - \boldsymbol{\mu}) - \sum_{i=1}^m \log[1 + \sum_{i \succ_{\sigma^{(k)}} j} e^{-\eta^{(k)}(s_i - s_j)}] \right).$$

Since the log-posterior is convex with respect to  $\mathbf{s}$  according to [8], we use stochastic gradient descent to optimize it.

## 4 Experimental Design

### 4.1 General Setup

We try out several experimental designs (the allocation of varieties to farmers). Since each farmer can measure 3 varieties,  $\frac{m}{3}$  farmers will be enough to compare all the varieties once. Thus, we first allocate  $m$  varieties to  $\frac{m}{3}$  farmers so that each variety appears exactly once, and we call this allocation one *replicate*. One replicate is demonstrated by a  $\frac{m}{3} \times 3$  matrix, where the entries in each row are the labels of items ranked by one farmer. We then generate  $\text{ceiling}(\frac{n}{m/3})$  replicates and aggregate all the replicates. We choose the first  $n$  rows of the aggregated replicates and the resulting  $n \times m$  matrix represents the experimental design.

### 4.2 Generation of One replicate

#### 4.2.1 Random Allocation (RD)

We assign the label 1 to  $m$  randomly to the  $\frac{m}{3} \times 3$  matrix.

---

**Algorithm 1** Cluster varieties into 3 clusters using k-means to initialize

---

```
1: procedure KM1( $M, K$ ) ▷  $M$  is the marker matrix
2:    $v$  = varieties not assigned to farmers
3:   Cluster  $v$  into 3 clusters by k-means clustering according to  $M$ 
4:    $x$  = the smallest size of the 3 clusters
5:   Choose top  $x$  varieties from each cluster with the largest competence measure and
      let them be the the varieties in the “chosen clusters”
6:   Delete varieties in “chosen clusters” from  $v$ 
7:   for  $v$  in  $v$  do
8:      $v$  enters the “chosen cluster” with the largest similarity measure (between  $v$  and
      varieties in each “chosen cluster”), unless that cluster has reached the size of  $\frac{m}{3}$ 
```

---

---

**Algorithm 2** Cluster varieties into 3 clusters using a sheer greedy algorithm

---

```
1: procedure GD1( $K$ )
2:    $v$  = varieties not assigned to farmers
3:   Randomly choose a variety from  $v$ , let it enter cluster 1 and delete it from  $v$ 
4:   for  $i = 2, 3$  do ▷ Initialize 3 clusters
5:     for  $v$  in  $v$  do
6:       Compute the similarity measure between  $v$  and all the varieties from cluster
         1 to cluster  $(i - 1)$ 
7:       Choose the variety with smallest similarity measure in  $v$ , let it enter cluster  $i$ 
         and delete it from  $v$ 
8:     for  $v$  in  $v$  do
9:        $v$  enters the cluster with the largest similarity measure (between  $v$  and varieties
         in each cluster), unless that cluster has reached the size of  $\frac{m}{3}$ 
```

---

---

**Algorithm 3** Cluster farmers into  $\frac{m}{3}$  clusters using k-means to initialize

---

```
1: procedure KM2( $M, A$ )
2:    $v$  = varieties not assigned to farmers
3:   Cluster  $v$  into  $\frac{m}{3}$  clusters by k-means clustering according to  $M$ 
4:   for  $i = 1, \dots, \frac{m}{3}$  do
5:     if size of cluster  $i \geq 3$  then
6:       Choose 3 varieties with the largest competence measure to stay in cluster  $i$ ,
         and delete all the other varieties from cluster  $i$ 
7:     Delete varieties already in clusters from  $v$ 
8:     for  $v$  in  $v$  do
9:        $v$  enters the cluster with the largest similarity measure (between  $v$  and varieties
         in each cluster), unless that cluster has reached the size of 3
```

---

---

**Algorithm 4** Cluster farmers into  $\frac{m}{3}$  clusters using a sheer greedy algorithm

---

```
1: procedure GD2( $K$ )
2:    $v$  = varieties not assigned to farmers
3:   Randomly choose a variety from  $v$ , let it enter cluster 1 and delete it from  $v$ 
4:   for  $i = 2, \dots, \frac{m}{3}$  do                                 $\triangleright$  Initialize  $\frac{m}{3}$  clusters
5:     for  $v$  in  $v$  do
6:       Compute the similarity measure between  $v$  and all the varieties from cluster
    1 to cluster  $(i - 1)$ 
7:       Choose the variety with smallest similarity measure in  $v$ , let it enter cluster  $i$ 
        and delete it from  $v$ 
8:     for  $v$  in  $v$  do
9:        $v$  enters the cluster with the largest similarity measure (between  $v$  and varieties
        in each cluster), unless that cluster has reached the size of 3
```

---

---

**Algorithm 5** Cluster farmers into  $\frac{m}{3}$  clusters using a sheer greedy algorithm

---

```
1: procedure GD3( $K$ )
2:    $v$  = varieties not assigned to farmers
3:   Randomly choose a variety from  $v$ , let it enter cluster 1 and delete it from  $v$ 
4:   for  $i = 2, \dots, \frac{m}{3}$  do                                 $\triangleright$  Initialize  $\frac{m}{3}$  clusters
5:     for  $v$  in  $v$  do
6:       Compute the similarity measure between  $v$  and all the varieties from cluster
    1 to cluster  $(i - 1)$ 
7:       Choose the variety with largest similarity measure in  $v$ , let it enter cluster  $i$  and
        delete it from  $v$ 
8:     for  $v$  in  $v$  do
9:        $v$  enters the cluster with the smallest similarity measure (between  $v$  and varieties
        in each cluster), unless that cluster has reached the size of 3
```

---

#### 4.2.2 Clustering algorithms

We use three kinds of clustering algorithms to cluster the varieties so that each farmer can choose accordingly. We first introduce two metrics that are used in the clustering algorithms. The *competence measure* of variety  $i$  in a cluster is calculated by summing up the correlation coefficients (the entry of  $K$ ) between  $i$  and all the other varieties in this cluster. The *similarity measure* between variety  $i$  and a group of varieties is calculated by taking the mean of the correlation coefficients between  $i$  and the other group of varieties.

##### Cluster varieties into three clusters

- **KM1** This algorithm uses k-means clustering based on the marker matrix to initialize and then uses a greedy algorithm based on  $K$  to cluster the varieties into 3 clusters. Then each farmer chooses one variety from each cluster. The algorithm is presented in Algorithm 1.
- **GD1** This algorithm uses a sheer greedy algorithm based on  $K$  to cluster the varieties into 3 clusters. Then each farmer chooses one variety from each cluster. The algorithm is presented in Algorithm 2.

##### Cluster varieties into $\frac{m}{3}$ clusters so that varieties within clusters are most similar

- **KM2** This algorithm uses k-means clustering based on the marker matrix to initialize and uses a greedy algorithm based on  $k$  to cluster the varieties into  $\frac{m}{3}$  clusters, so that varieties within clusters are most similar to each other. Then each farmer chooses one cluster as the varieties they are going to compare. The algorithm is presented in Algorithm 3.
- **GD2** This algorithm uses a sheer greedy algorithm based on  $K$  to cluster the varieties into  $\frac{m}{3}$  clusters, so that varieties within clusters are most similar to each other. Then each farmer chooses one cluster as the varieties they are going to compare. The algorithm is presented in Algorithm 4.

##### Cluster varieties into $\frac{m}{3}$ clusters so that varieties within clusters are most dissimilar

- **GD3** This algorithm uses a sheer greedy algorithm based on  $K$  to cluster the varieties into  $\frac{m}{3}$  clusters, so that varieties within clusters are most *dissimilar* to each other. Then each farmer chooses one cluster as the varieties they are going to compare. The algorithm is presented in Algorithm 5.

#### 4.2.3 Alpha Design (AD)

Alpha design is an experimental design, which is invented by Patterson and Williams [6] and are widely used in breeding trials. Here we use the R package *Agricolae* [2] to generate (0, 1) alpha designs with treatment number =  $m$ , block size = 3 and replications = 3, so that each farmer represents a block and each replication is one replicate.

## 5 Simulation

We then do simulations to compare the performance of the four models and the six experimental designs under various situations. The genotype data we use are wheat marker sequences from Kansas State Genotyping Lab (KSG), available in the Triticeae Toolbox ([https://triticeaetoolbox.org/wheat/display\\_genotype.php?function=typeData&mm=8&mmaf=5&trial\\_code=2014\\_TCAP\\_ABB\\_SRW\\_Mid](https://triticeaetoolbox.org/wheat/display_genotype.php?function=typeData&mm=8&mmaf=5&trial_code=2014_TCAP_ABB_SRW_Mid)). Marker data are filtered by maximum missing data = 8% and minimum MAF = 5%, and we get 368 lines with 9928 markers for each line. The missing data are recovered by imputation method in the rrBLUP package in R [3].

Given the number of varieties  $m$  and the number of farmers  $n$ , we first randomly choose  $m$  varieties out of 368 and we sample 1000 markers out of 9928 and let them be the causal loci. We hence get a  $m \times 1000$  marker matrix  $M$  (with rows being different varieties) coded in  $\{0, 1, 2\}$ . We then sample the effect for each causal locus independently from a normal distribution with mean being a draw from  $Uniform(1, 10)$  and variance being 1. The vector of loci effects is denoted by  $\alpha$ . Then we have the real score vector  $s = M\alpha$ . The ranking data are generated by sampling from the Thurstone model with score being  $s$ . In order to account for farmers' turbulence, we use the heritability measure  $h^2$  which is commonly used in breeding. We have

$$h^2 = \frac{\sigma_s^2}{\sigma_s^2 + \sigma_e^2},$$

where  $\sigma_s^2$  is the variance of  $s$  and  $\sigma_e^2$  is the variance of farmers' choices. Therefore given  $h^2$ , we can compute  $\sigma_e^2$  and we let  $\sigma_e^2$  be the variance parameter in the Thurstone model. Then we sample  $n$  complete rankings from the Thurstone model. The allocation of varieties to farmers are generated according to the experimental designs described in section 4. The additive relationship matrix  $K$  are computed using the remaining 8928 markers by rrBLUP package [3], using shrinkage method introduced by Endelman and Jannink [4].

Now we fit the synthetic ranking data using the four models described above. Then we use two similarity measures for comparison of accuracy among models and experimental designs. The first one is the Kendall's correlation coefficient (referred by "Kendall"). The other one is the proportion of top 20% varieties from the estimated ranking **in** the top 20% varieties from the real ranking (referred by "top20"). Kendall measures the overall similarity between two rankings while top20 measures the specific similarity between two rankings.

### 5.1 Comparison of Models

#### 5.1.1 Models with Adherence Parameters

We first do a simulation to compare the models described in section 3. Namely we include LM, TH, BT and PL with and without genotypic information. All models except LM can accommodate the adherence parameters. We break this simulation into two parts: the pilot simulation and the main simulation. In the pilot simulation, we run the above procedure 100 times with  $h^2 = 0.2, 0.5, 0.8$ , for  $m = 15, 30, 60$  and  $n = 30, 60, 120$ . In the main simulation, we run the above procedure 100 times with  $h^2 = 0.2, 0.5, 0.8$ , for  $m = 120, 240, 360$  and

$n = 180, 240, 300, 360$ . We use the alpha design to allocate varieties to farmers. The main simulation is the situation we may encounter in practice.

### 5.1.2 Models without Adherence Parameters

We then do a simulation to see the performance of models when the adherence parameters are not included. Models without adherence parameters are simply recovered by taking adherences for all farmers to be 1. We include LM, TH, BT and PL with and without genotypic information in this simulation. We run the procedure 100 times with  $h^2 = 0.2, 0.5, 0.8$  with  $m = 240$  and  $n = 180, 240, 300, 360$ . We use the alpha design to allocate varieties to farmers.

### 5.1.3 Inclusion of Outliers

We also do a simulation to see the performance of models when outliers are included. We randomly choose  $10\% \times n$  farmers and the ranking given by those farmers are completely random permutations. We include LM, TH, BT and PL with genotypic information. We run the procedure 100 times with  $h^2 = 0.2, 0.5, 0.8$  with  $m = 240$  and  $n = 180, 240, 300, 360$ . We use the alpha design to allocate varieties to farmers.

## 5.2 Comparison of Experimental Designs

In this simulation, we only include LM and BT with genotypic information. We run the procedure 100 times with  $h^2 = 0.2, 0.5, 0.8$  with  $m = 240$  and  $n = 180, 240, 300, 360$ . We include all 7 experimental designs described in section 4.

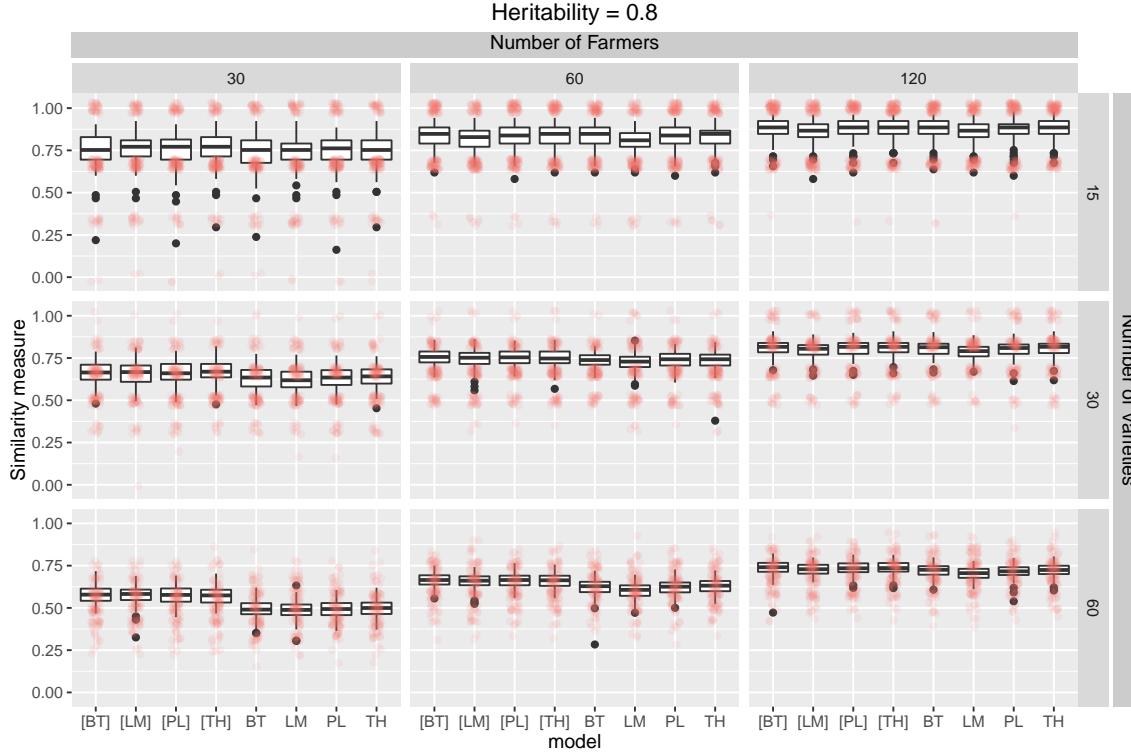
## 6 Results

We presents the simulation results in this section. Models with square brackets ([LM] for example) are models with genotypic information, while models without square brackets (LM for example) are models without genotypic information. The top20 metric is demonstrated by jitter plots, while the Kendall metric is demonstrated by box plots.

The results of pilot simulation is shown in Figure 1, 2 and 3. We can see the improvement of both similarity measures when number of farmers and heritability increases. Meanwhile, the models with genotypic information achieves a larger similarity measure. This difference is especially evident when the number of varieties are large. When  $h^2 = 0.8$  nad 0.5, we cannot tell the difference among the four models with genotypic information. However, [LM] performs better than the other three models when  $h^2 = 0.2$ .

The results of main simulation is shown in Figure 4, 5, and 6. The main simulation is the simuation we may encounter in practice. We can see [LM] achieves a larger similarity measure in all simuations. Also, the performance of [LM] improves as the number of farmers increases. However, the performance of the other three models doesn't improve much as the number of farmers increases.

Figure 1: Results of pilot simulation - part I



The influence of the exclusion of adherence parameters is shown in Figure 7. We see a decrease of similarity measure when we get rid of adherence parameters of [TH], [BT] and [PL]. And [LM] still substantially outperforms all the other models.

When we include 10% outliers, we can still see in Figure 8 that [LM] substantially performs better than the other three models. But the similarity measure decreases compared to the results in main simulation, where there are no outliers.

The performance of different experimental designs is shown in Figure 9 and 10. Compared to the random allocation, the alpha design achieves a larger similarity measure in all situations. However, the performance of clustering algorithms isn't promising. GD1 and KM1, which both cluster the varieties into 3 clusters so that each farmer can choose one variety from each cluster, performs similarly to RD. GD3, which clusters the varieties into  $\frac{m}{3}$  clusters and varieties within clusters are most dissimilar, also achieves a very similar similarity measure to RD. Meanwhile, GD2 and KM2, which both cluster the varieties into  $\frac{m}{3}$  clusters and varieties within clusters are most similar, perform substantially poorly compared to RD.

## 7 Discussion

When deriving the MAP estimators for Thurstone model, Bradley-Terry model and Plackett-Luce model, we assume the independence among pairwise comparisons and the indepen-

Figure 2: Results of pilot simulation - part II

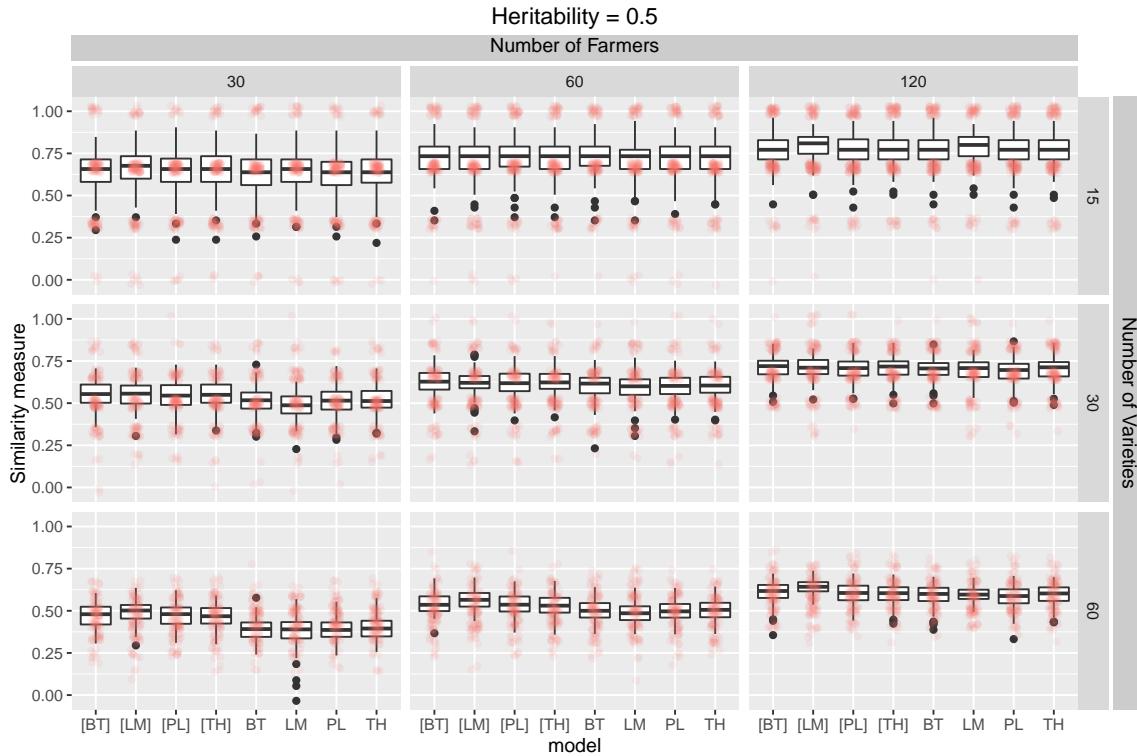


Figure 3: Results of pilot simulation - part III

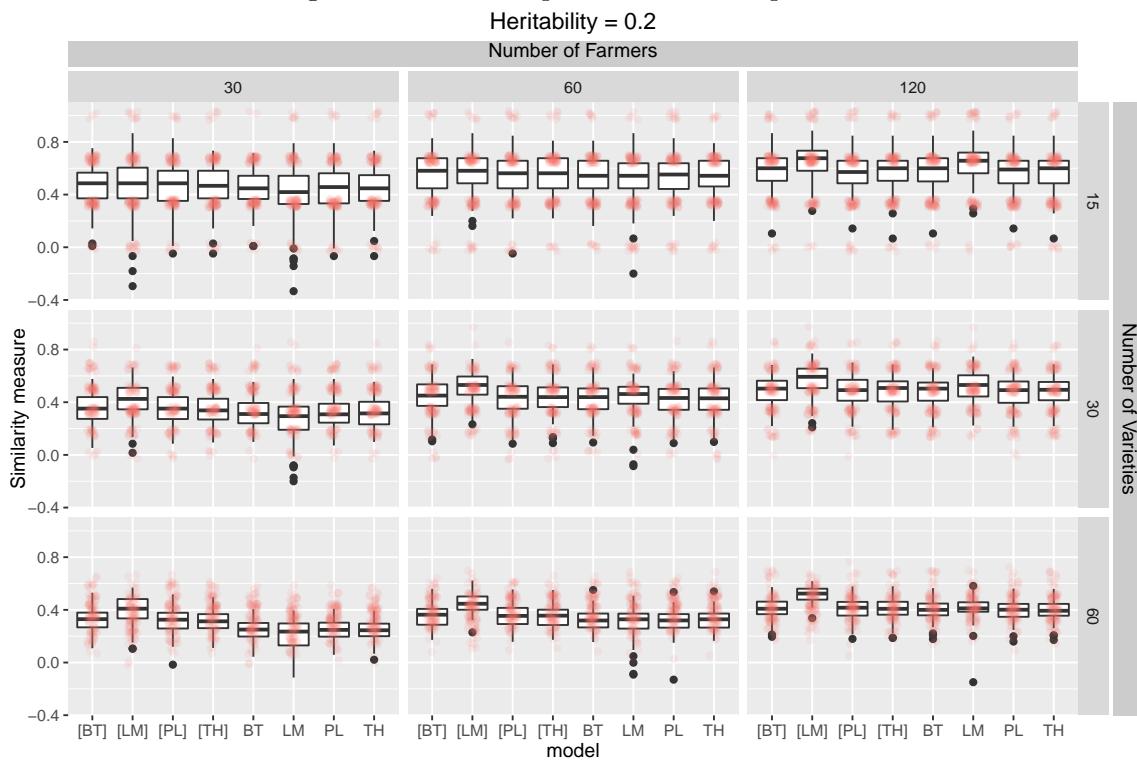


Figure 4: Results of main simulation - part I

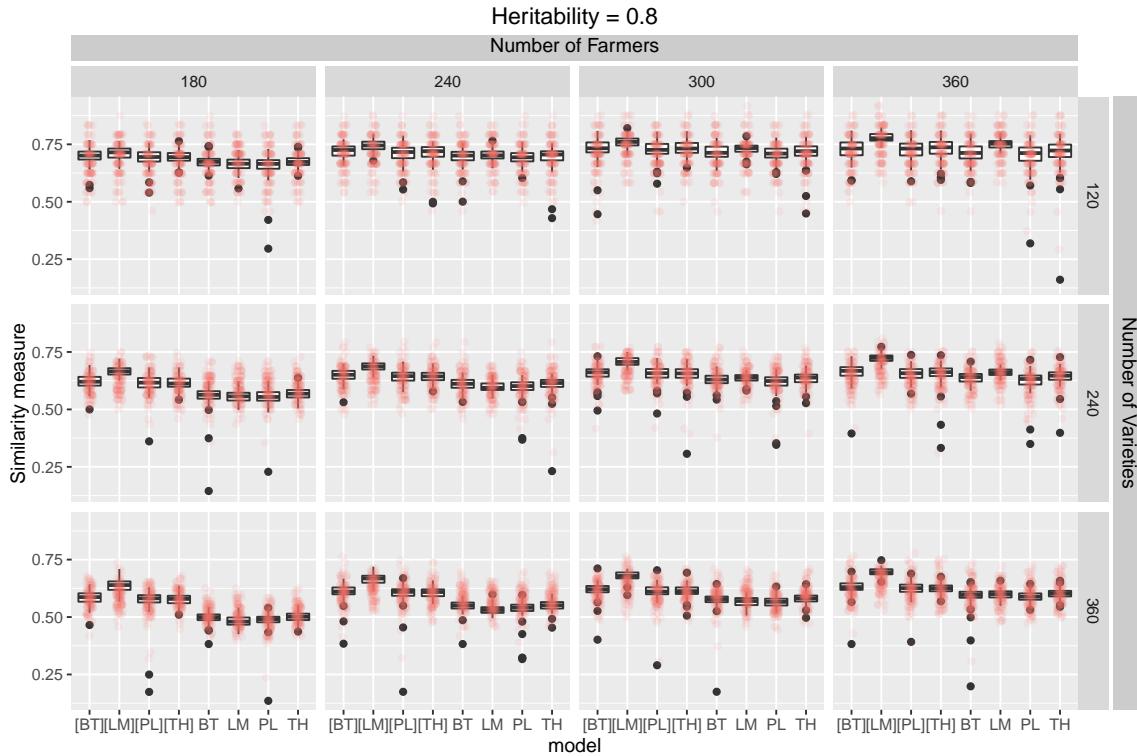


Figure 5: Results of main simulation - part II

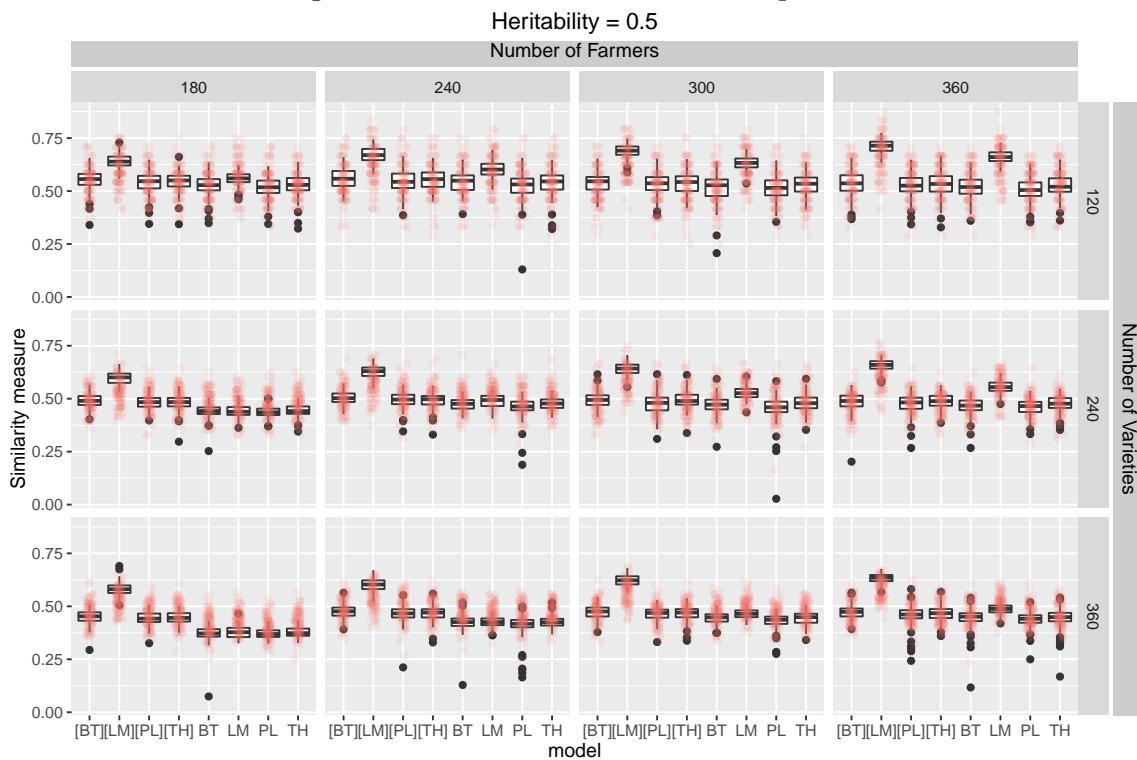


Figure 6: Results of main simulation - part III

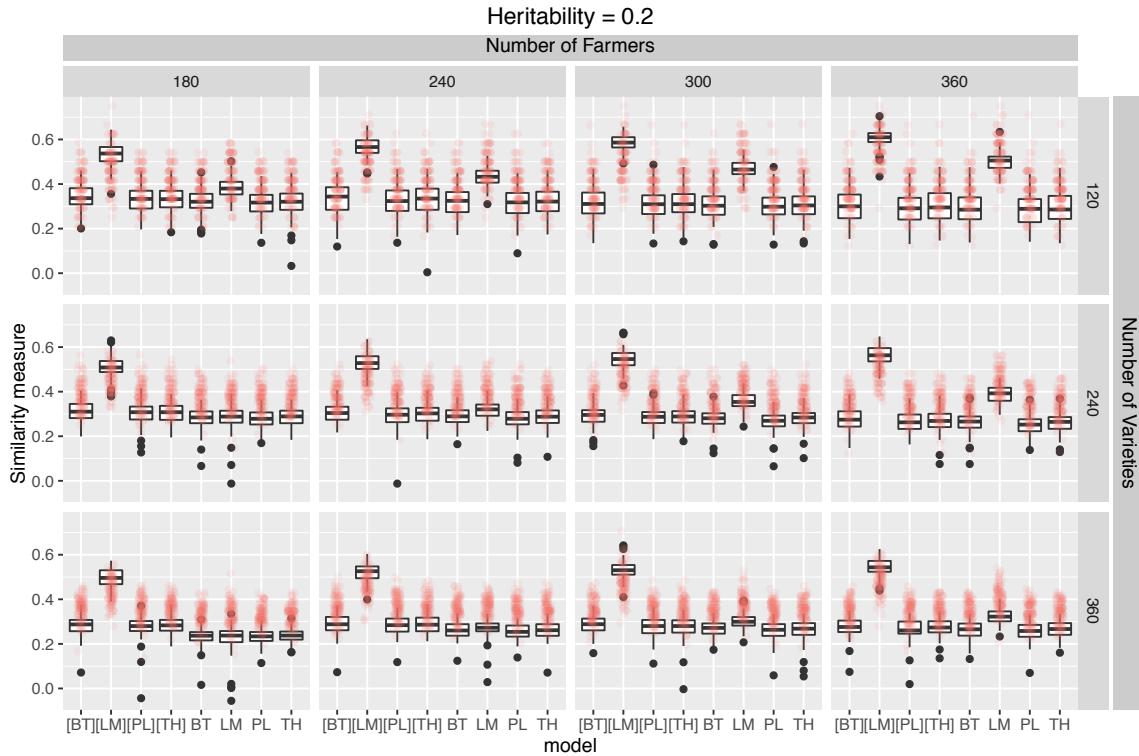


Figure 7: Results of simulation when adherence parameters are not included

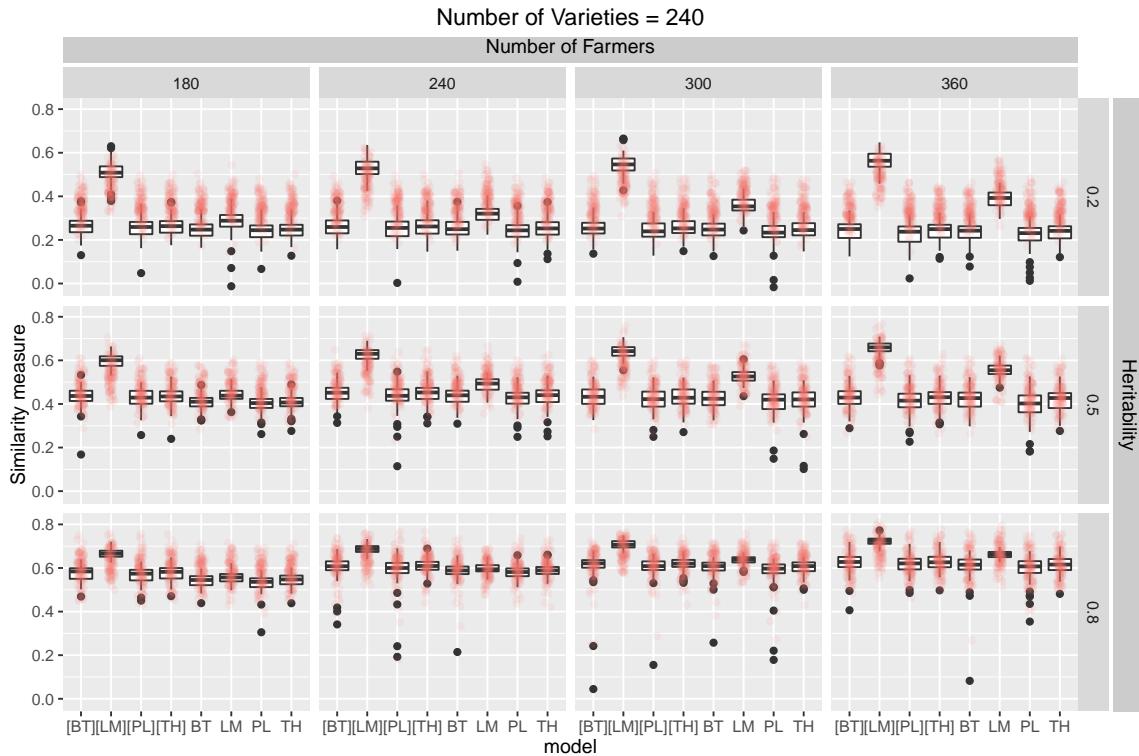


Figure 8: Results of simulation when 10% of the farmers are outliers

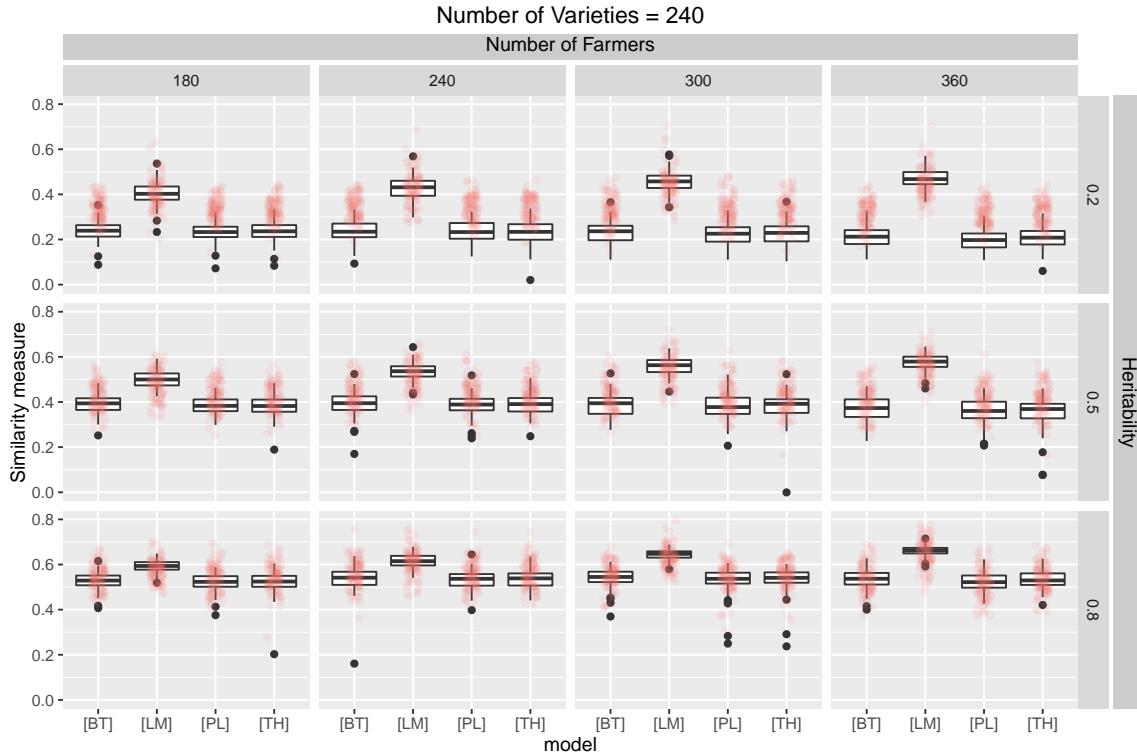
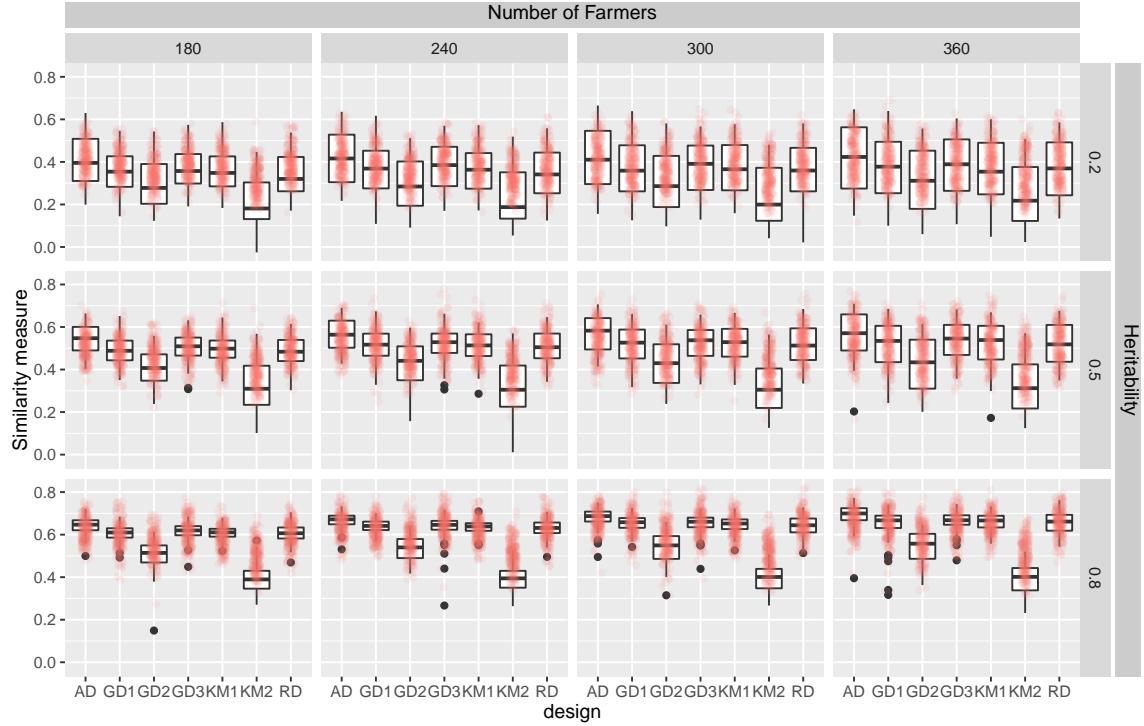


Figure 9: Comparison of experimental designs - part I



Figure 10: Comparison of experimental designs - part II  
 Number of Varieties = 240, Model = [BT]



dence among all farmers. While the latter is a resonable assuption and can be controlled during field experiments, the former might not be a good assuption to make, since each farmer compares the three varieties concurrently and the transitive law of ranking (e.g.  $A \succ B, B \succ C$ , then we must have  $A \succ C$ ) may break the independence. That is one possible reason why those models cannot compete with the linear model.

## References

- [1] Ralph Allan Bradley and Milton E Terry. Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika*, 39(3/4):324–345, 1952.
- [2] Felipe De Mendiburu. Agricolae: statistical procedures for agricultural research. *R package version*, 1:1–6, 2014.
- [3] Jeffrey B Endelman. Ridge regression and other kernels for genomic selection with r package rrblup. *The Plant Genome*, 4(3):250–255, 2011.
- [4] Jeffrey B Endelman and Jean-Luc Jannink. Shrinkage estimation of the realized relationship matrix. *G3: Genes—Genomes—Genetics*, 2(11):1405–1413, 2012.
- [5] R.. Dukan Luce. *Individual Choice Behavior a Theoretical Analysis*. John Wiley and sons, 1959.
- [6] HD Patterson and ER Williams. A new class of resolvable incomplete block designs. *Biometrika*, 63(1):83–92, 1976.
- [7] Robin L Plackett. The analysis of permutations. *Applied Statistics*, pages 193–202, 1975.
- [8] Karthik Raman and Thorsten Joachims. Methods for ordinal peer grading. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1037–1046. ACM, 2014.
- [9] Louis L Thurstone. A law of comparative judgment. *Psychological review*, 34(4):273, 1927.
- [10] Jacob van Etten. Crowdsourcing crop improvement in sub-saharan africa: a proposal for a scalable and inclusive approach to food security. *IDS bulletin*, 42(4):102–110, 2011.