

Accounting for G×E interactions in plant breeding: a probabilistic approach

Reyhaneh Bijari

Iowa State University of Science and Technology: Iowa State University

Sigurdur Olafsson (✉ olafsson@iastate.edu)

Iowa State University of Science and Technology: Iowa State University <https://orcid.org/0000-0002-0088-6898>

Research Article

Keywords: Selection Methodology, Genotype-Environment Interactions, Probability Estimates, Bootstrapping

Posted Date: September 28th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-2052233/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Accounting for $G \times E$ interactions in plant breeding: a probabilistic approach

Reyhaneh Bijari¹ and Sigurdur Olafsson^{1*}

¹Department of Industrial and Manufacturing Systems Engineering, Iowa State University, 3004 Black Engineering Bldg, Ames, 50010, IA, USA.

*Corresponding author(s). E-mail(s): olafsson@iastate.edu;
Contributing authors: rbijari@iastate.edu;

Abstract

In plant breeding trials the mean phenotypic responses of two experimental genotypes are often close; and which genotype appears to perform better depends on the specific environments that are observed. When only one of the genotypes must be selected, the conclusion drawn by comparing means across any set of observed environments may differ from the conclusion that would be drawn if all target environments could have been observed, in which case the wrong selection may be made. This paper proposes a new method of comparing genotypes that aims to select the genotype that is more likely to perform better across a set of environments, rather than the one that has the better mean. The implementation uses bootstrap resampling to estimate the probability that one genotype outperforms another in a set of observed environments, and by doing so accounts for the uncertainty caused by observing limited environments. The results show that due to the different genetic-by-environment ($G \times E$) interaction effects, the genotype that is more likely to be better is sometimes different than the one with the better mean and the probabilistic comparison accounts for both the mean and the interaction effects.

Keywords: Selection Methodology, Genotype-Environment Interactions, Probability Estimates, Bootstrapping

1 Introduction

Selecting the genotype that performs best across a set of target environments is a long-standing problem in plant breeding for which many statistical methods are available [1]. This problem has traditionally been addressed by estimating the mean performance of each genotype and comparing their relative performance [2–9]. However, it is also well known that mean performance, which can be thought of as the main genetic effect of the genotype, or G effect, can be misleading because of the importance of the interaction between genetics and environment, or G×E effects. Thus, the mean information is always complemented in practice with at least one measure of stability or adaptability, which account for some part of the G×E structure [10–13]. Measures that combine the mean and stability have been in popular use for decades [14].

We propose a novel method that shifts the focus from an inherently mean-based comparison to selecting genotypes based on the *probability of performing best across target environments*. We argue that selecting the genotype that is more likely to perform better in a set of environments combines mean information with an account of the entire G×E structure, that is, both aspects of stability and adaptability and the distribution of the quality of the target environments.

The first issue is to determine which probabilities must be estimated. To that end we note that while a plant breeder may ultimately look at a ranking for a set of genotypes, at its core the selection process comes down to making pairwise comparisons between genotypes [7]. Thus, if we can estimate the pairwise probabilities of one genotype being better than another across a set of target environments, those comparisons can be converted into a ranking of all the experimental genotypes. Given that we consider pairwise comparisons, it is natural to ask when the probabilistic pairwise comparisons will result in a different outcome than a simple pairwise comparison of means? To obtain some insights into those differences, we start by presenting a simplified motivating example.

Suppose we have two experimental cultivars of some crop, Cultivar 1 and Cultivar 2, that could be planted in five environments: E_1 , E_2 , E_3 , E_4 , E_5 ; and the phenotype of interest is yield measured in bushels per acre. These five environments represent the entire universe of possible environments (locations and years). The yield of Cultivar 1 is 73.0, 51.0, 61.0, 48.0, and 55.0 in the five environments, respectively; and the yield of Cultivar 2 is 63.5, 53.5, 59.5, 50.5, and 56.5 in the five environments, respectively. By calculating the average we observe that Cultivar 1 has the better mean across all possible environments, or 57.6 versus 57.1. Cultivar 1 also has the better average rank (1.4 versus 1.6) since it has higher yield in more environments. However, a crop will never be planted in all possible environments, not even within a single year, so another quantity of interest would be which variety is *more likely* to have higher yield when planted in some fixed number of environments? For example, we select two out of the five environments at random. There are exactly ten such pairs and it is easy to verify that Cultivar 1 performs better in only four out of those

Table 1 The $G \times E$ effects for two simulated cultivars where Cultivar 2 is more stable.

$G \times E$ effects	E_1	E_2	E_3	E_4	E_5
Cultivar 1	10	0	4	0	1
Cultivar 2	3	3	3	3	3

ten pairs. Specifically, Cultivar 1 does better in every pair that includes the highest-yielding environment E_1 and Cultivar 2 does better in every pair that does not include E_1 . Thus, we can conclude that even though Cultivar 1 has a higher mean yield it will only have higher yield 40% of the time if the two cultivars are planted in two randomly selected environments.

We argue that this small-scale example mirrors what happens in real situations, where there is a large set of potential target environments and each year a crop is only planted in a very small subset of all possible environments. This example therefore motivates the main idea of this paper, which is to propose probabilistic comparison as an alternative to mean-based comparison.

A couple of important observations need to be made about the motivating example. First, this scenario would not happen if the cultivars had the same $G \times E$ structure. The yield in this example is calculated according to $y_{ij} = 50 + g_i + e_j + h_{ij}$, where $g_i \in \{3, 2.5\}$ is the genetic effect of the two cultivars, $e_j \in \{10, -2, 4, -5, 1\}$ are the environmental effects and the $G \times E$ interaction (h_{ij}) is given in Table 1. We note that $\sum_{j=1}^5 h_{1j} = \sum_{j=1}^5 h_{2j} = 15$ (row sums in the table), and on the average the interaction is therefore the same; but whereas Cultivar 2 is very stable, Cultivar 1 is able to double the environmental effect of good environments but is neutral in poor environments. Cultivar 1 is precisely the type of variety that we expect to appear better with respect to mean performance versus a probabilistic approach.

Second, if the difference in main genetic effect is sufficiently large then the mean-based and probabilistic approach will always reach the same conclusion. For example, if $g_1 = 3$ and $g_2 = 2.4$ but everything else stays the same, that is, the difference in mean increases from 0.5 to 0.6 bushels/acre, it is easy to verify that Cultivar 1 will be selected 50% of the time based on two random environments. And if $g_2 = 1.9$ then Cultivar 1 will be selected 70% of the time. Thus, the proposed probabilistic approach is primarily relevant for comparisons where the difference in genotype effects is relatively small; but we argue that those are also the comparisons that are the most important to plant breeders in practice.

The observation made above regarding differences in $G \times E$ structure suggests that there is a relationship between the proposed approach and stability measures. The connection between the new method and specific stability measures will not be explored in detail in this paper, but it should be pointed out that no stability measure could completely replace the probability-based approach. Just like the mean phenotype is a summary statistic of the probability distribution of phenotype across environments, stability measures provide

another complementary summary statistic. While considering two or more such summary statistics is certainly preferable to a single statistic, no summary statistics can completely replace considering the whole probability distribution. In fact, in some sense *the probabilistic approach accounts for the entire $G \times E$ structure* in whatever is selected as the target environments, rather than the mean and some measure(s) of stability.

2 Method for Probabilistic Comparison

The goal is to understand the performance of n genotypes in m target environments (locations and years). Let y_{ij} denote the phenotype of genotype i in environment j , where $i = 1, 2, \dots, n$; $j = 1, 2, \dots, m$. What is traditionally of interest is the mean of each genotype across all environments, denoted $g_i = \frac{1}{m} \sum_{j=1}^m y_{ij}$, and may be thought of as representing the genotype effect. As noted above, what is specifically of interest is comparing two genotypes i_1, i_2 , which is usually done based on the genotype effect and one approach would be to try to estimate the mean difference $g_{i_1} - g_{i_2}$ with as much precision as possible and use this as the basis of a decision. However, in the method proposed here the analysis is based on the indicator function

$$I(i_1, i_2) = \begin{cases} 1, & \text{if } g_{i_1} > g_{i_2}, \\ 0, & \text{if } g_{i_1} \leq g_{i_2}. \end{cases} \quad (1)$$

In other words, we are only concerned with if one genotype is better than the other, rather than the magnitude of the difference.

Since it is impossible to observe every possible environments, a plant breeder observes some small sample A of m_0 environments, where $m_0 \ll m$, and uses the observed values $\{y_{ij}, j \in A; i = 1, \dots, n$, to obtain estimates $\hat{g}_i = \frac{1}{m_0} \sum_{j \in A} y_{ij}$. From the observed environments A it is straightforward to obtain a point estimate of the indicator of interest

$$\hat{I}_A(i_1, i_2) = I_A(i_1, i_2). \quad (2)$$

This will simply equal either zero or one, depending on the relative performance of the two genotypes in this set A of environments. Based on this estimate, a plant breeder might conclude that genotype i_1 is better than genotype i_2 if $\hat{I}(i_1, i_2) = 1$. This is equivalent to directly estimating the mean difference $\hat{g}_{i_1} - \hat{g}_{i_2}$ and making a decision based on this mean. However, as we argued in the introduction, focusing on the mean can be misleading if the two genotypes have different $G \times E$ structures, in which case it would be more informative to know the distribution of the indicator function (1) rather than just the mean.

Bootstrapping is a well-known technique that can be used to characterize the sampling distribution of a statistic using resampling [15]. To characterize the distribution of $I(i_1, i_2)$, we start by generating B bootstrap samples A_1, A_2, \dots, A_B by sampling the set A of environments with replacement. Due to the $G \times E$ interaction effects, the outcome for the indicator (1), which

indicates which genotype is better, may be different for different samples of environments; and this results in B estimates of the indicator:

$$\hat{I}_{A_1}(i_1, i_2), \hat{I}_{A_2}(i_1, i_2), \dots, \hat{I}_{A_B}(i_1, i_2). \quad (3)$$

Thus, instead of a single estimate, we now have B estimates, capturing the uncertainty due to the set of environments that were observed. Note that if there were no G×E interaction effects or if they were identical for both genotypes then $\hat{I}_{A_j}(i_1, i_2) = \hat{I}_A(i_1, i_2), \forall j$, that is, the conclusion reach for each resampled set A_j of environments would be the same as the original observed environments A .

In practice the magnitude of the G×E interaction effects is usually significant and differs for different genotypes. Using the estimates calculated on the bootstrap samples, the probability of genotype i_1 being better than genotype i_2 in a randomly selected set of environments can then be estimated as

$$\hat{P}(g_{i_1} > g_{i_2}) = \frac{1}{B} \sum_{b=1}^B \hat{I}_{A_b}(i_1, i_2). \quad (4)$$

Given these probability estimates, a plant breeder could now conclude that genotype i_1 is better than genotype i_2 if $\hat{P}(g_{i_1} > g_{i_2}) > \frac{1}{2}$.

We refer to decisions based on equation (2) as mean-based selection and decisions based on (4) as probabilistic selection. Thus, decisions are determined by one of the following sets of genotype pairs, which completely describe which genotypes should be selected over others:

$$S_{Mean} = \{(i_1, i_2) \in \{1, 2, \dots, n\}^2 : I_A(i_1, i_2) = 1\}, \quad (5)$$

$$S_{Prob} = \{(i_1, i_2) \in \{1, 2, \dots, n\}^2 : \hat{P}(g_{i_1} > g_{i_2}) > \frac{1}{2}\}. \quad (6)$$

As far as we know, this type of probabilistic selection has not been proposed before in the plant breeding domain, whereas incorporating the mean-based selection is standard practice.

An example of the mean-based and probabilistic selection approaches is illustrated in Fig. 1. In this example the two approaches would reach different conclusions because while one genotype has a better mean across the observed locations, this is not true for the majority of the resampled subsets of locations. The remainder of this paper explores when these two approaches result in different conclusions, that is, $S_{Mean} \neq S_{Prob}$, and the explanations behind those differences.

3 Data

To evaluate the proposed method, we utilize two different data. The first data is simulated and specifically constructed to evaluate the differences between

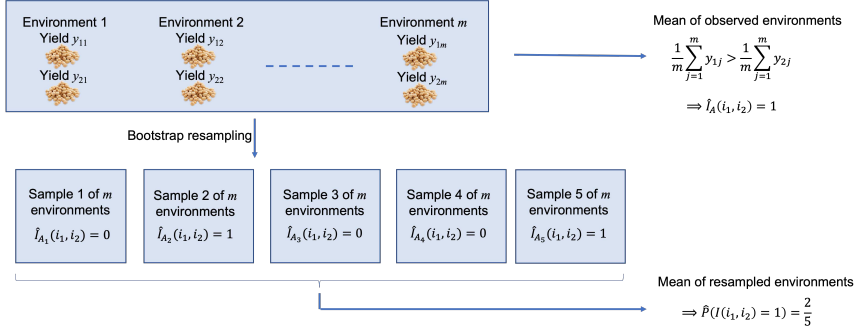
6 Accounting for $G \times E$ interactions

Fig. 1 Mean versus probabilistic comparison of two genotypes. Based on direct observations the first genotype is better, but as it is only better in two out of five resampled environments the probabilistic comparison favors the second genotype.

the new probabilistic approach and the traditional mean-based approach. The second data is a previously analyzed rapeseed data that is used to demonstrate how the method may be applied in practice and how it compares to existing methods that aim to account for at least some of the $G \times E$ interaction effects.

3.1 Simulated Data

To provide insights into if and when probabilistic pairwise comparison differs from mean-based comparison, we generate simulated data that can be considered as a generalized version of the motivational example discussed in the introduction. To generate this data, we assume that the phenotype of interest is plant yield, although any other phenotype could be used, and that yield follows what might be considered a standard linear model involving genetic (G), environmental (E), and genetic-by-environment interaction effects ($G \times E$ effects) [1, 10].

$$\tilde{y}_{ij} = \mu + G_i + E_j + G_i \times E_j + \epsilon_{ij}. \quad (7)$$

As the noise does not provide any insight with respect to the purpose of the new method (we aim to capture the uncertainty due to selection of planting environments, not the uncertainty within each environment), for simplicity we ignore the noise from the above equation.

As will be further described below, the simulation then generates values for each effect according to specific distribution $G_i \sim F_G$, $E_j \sim F_E$, and assumes different structures for the $G \times E$ interactions. This experiment includes three factors that will be set as follows:

1. *Difference in main effect.* These will be determined by the generated G effects for each genotype.
2. *Similarity in interactions.* These are generated as one of three types: a) same in all environments, b) very good in good environments, neutral in others,

Table 2 Thirty simulated genotypes. A set of three genotypes has identical main genetic effect (G), but each of those three has a different $G \times E$ structure.

Name	Main Effect (G)	$G \times E$ Structure
GS1.44	1.44	Stable
GA1.44	1.44	Adaptive
GV1.44	1.44	Variable
GS1.39	1.39	Stable
GA1.39	1.39	Adaptive
GV1.39	1.39	Variable
GS1.13	1.13	Stable
GA1.13	1.13	Adaptive
GV1.13	1.13	Variable
GS0.81	0.81	Stable
GA0.81	0.81	Adaptive
GV0.81	0.81	Variable
GS0.09	0.09	Stable
GA0.09	0.09	Adaptive
GV0.09	0.09	Variable
GS0.39	-0.39	Stable
GA0.39	-0.39	Adaptive
GV0.39	-0.39	Variable
GS0.52	-0.52	Stable
GA0.52	-0.52	Adaptive
GV0.52	-0.52	Variable
GS0.66	-0.66	Stable
GA0.66	-0.66	Adaptive
GV0.66	-0.66	Variable
GS1.23	-1.23	Stable
GA1.23	-1.23	Adaptive
GV1.23	-1.23	Variable
GS1.29	-1.29	Stable
GA1.29	-1.29	Adaptive
GV1.29	-1.29	Variable

c) very good in good environments and very poor in poor environments. The sum of the $G \times E$ effects is kept as constant.

3. *Magnitude of interactions.* We consider a set the average magnitude of the $G \times E$ effects as half or double to the average magnitude of the main G effect.

Following the above description, after generating a mean genotype effect, we generate three genotypes with three different $G \times E$ structures. The first is completely stable. i.e., has it has the same $G \times E$ effects over all environments.

Table 3 Planting locations and years.

Year	TX	TGA	GGA	SC	MS	NC	TN	VA	KS	NY	OR	WA	ID	MT
1987	X	X	X	X		X					X	X	X	X
1988	X	X		X	X			X	X		X	X	X	
1988		X	X	X		X	X	X		X		X	X	X

The second might be considered adaptive, i.e., takes advantage of good environments and performs very well in those environments. Finally, the third one is highly variable, with both very good and very poor performance based on the E effect. All three types have a exact same mean phenotype if we observe all of m environments, while their distributions are different. Therefore, with any subset of the m environments, the mean phenotypic response would also be different. Furthermore, for each set of genotypes, we consider uniform distribution of $[-1.5, 1.5]$ for F_G and uniform distribution of $[-10, 10]$ for F_E . As explained, $G \times E$ distributions derive from G and E effects based on their structures. To cover every aspect, we consider different magnitude of $G \times E$ with different fractions of environments to explore how different are the comparisons based on different contributing factors.

The simulated data generated by this procedure consists of 30 different genotypes with three different $G \times E$ interaction effects and 100 environments. For each three set of genotypes, genetic main effects are identical, that is, we have 10 distinct genetic effects. The description of the simulated genotypes is shown in Table 2.

3.2 Rapeseed Data

The intend of the simulated data is to study general features that determine when the proposed probabilistic selection differs from mean-based selection. For a more realistic example of those differences we analyse data obtained from a rapeseed multi-environmental trials that has been described by [16, 17]. We obtained the data via the *agridat* R package [18], which provides a convenient collection of agricultural data for testing new methods.

This data includes 648 observations of six cultivars planted in fourteen locations across three years (1987-89). The locations represent thirteen different states in the USA (TX, GA, SC, MS, NC, TN, VA, KS, NY, OR, WA, ID and MT) with two locations in Georgia (TGA = Tifton, GGA = Griffin). Some locations are planted all three years, whereas others may only have been planted for one or two years for a total of 27 environments (see Table 3). Table 4 summarizes the mean and variance of the yield for each cultivar.

This data was selected to illustrate the new method because based on the original analysis there are intuitive reasons to believe that a probabilistic comparison will yield different results than mean-based comparison. For example, from previous results [16, 17] it is known that the cultivar with the highest mean yield (Bienvenu) also has the worst stability, whereas the cultivar with the second highest mean (Glacier) is also the most stable according to multiple measures. Furthermore, two of the high-yielding cultivars, Bienvenu

Table 4 Yield summary by cultivar in the rapeseed data.

Cultivar	Mean	Variance
Bienvenu	$2.49 \cdot 10^3$	$3.19 \cdot 10^6$
Glacier	$2.48 \cdot 10^3$	$2.31 \cdot 10^6$
Bridger	$2.37 \cdot 10^3$	$1.62 \cdot 10^6$
Cascade	$2.32 \cdot 10^3$	$2.21 \cdot 10^6$
Jet Neuf	$2.28 \cdot 10^3$	$2.38 \cdot 10^6$
Dwarf	$2.27 \cdot 10^3$	$2.77 \cdot 10^6$

and Bridger, have large $G \times E$ effects but in different directions (preferences for different environments). Finally, the stable high-performer (Glacier) performs poorly in the warmest locations, so performance will depend on the target environments. We should expect the probabilistic approach to correctly account for such $G \times E$ structure differences. The expectation is thus not for new insights for this data that has already been studied carefully; but rather to demonstrate that the probabilistic comparison can combine in a single measure what otherwise would require multiple measures. This data also has relatively many environments observed and only 15 pairwise comparisons to be evaluated, implying that it is reasonable to obtain good estimates of the pairwise probabilities.

4 Results

4.1 Simulated Data

We start by looking at the estimated probabilities of one genotype being superior to another genotype. The results are displayed in Fig. 2 and demonstrate how the probabilities of one genotype being better than another depends on the genetic main effect differences and the differences in $G \times E$ structures. All plots in Fig. 2 are showing the comparison among the genotype with better main G main effect with respect to the other genotype in the pair; that is, if the genotype with the better main effect was always selected then all the numbers in the heatmaps in Fig. 2 would be larger than half. Furthermore, the genotypes are ordered according to their main effects with the one with the best mean yield being furthest to the right.

First note that if the $G \times E$ structure does not differ, then the probabilities are all either zero or one, that is there is no ambiguity. The three plots on the left in Figure 2 compares pairs with identical $G \times E$ structure. In those cases, the probability that the genotype with the higher mean is better is always one, even if the difference in the means is very small. On the other hand, when comparing genotypes that differ with respect to their interactions to environments, the likelihood of one genotype being better than the other genotype gets smaller as their genetic effects (yield similarity) get apart from each other. This trend is consistent in all three plots shown on the right hand side of Fig. 2 and

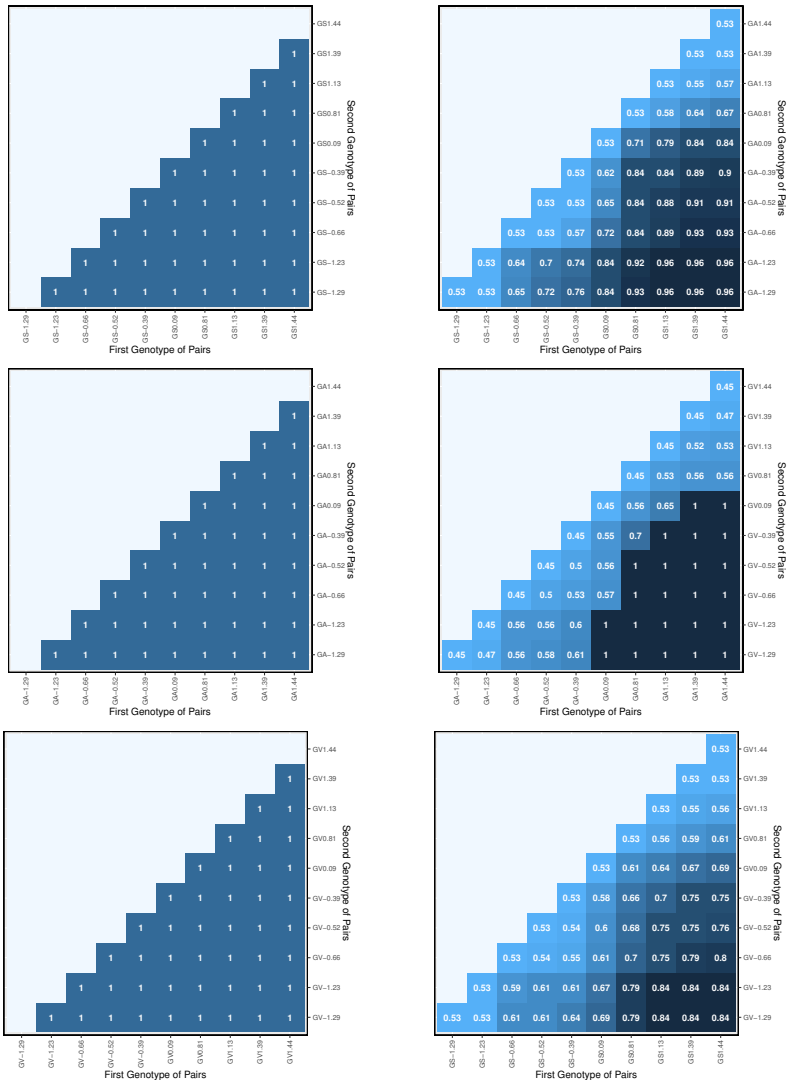


Fig. 2 Heat maps showing probabilistic comparison of genotypes pairs where the G×E structure is stable vs stable (top-left), stable vs adaptive (top-right), adaptive vs adaptive (middle-left), adaptive vs variable (middle-right), variable vs variable (bottom-left), and variable vs stable (bottom-right). Genotypes are ordered according to main effects.

the bigger the genetic difference, the more certain the comparison gets, that is, the probabilities converge to one as the difference in main effect becomes larger. A more subtle observation is that absolute certainty (probability equal one) is observed when comparing adaptive genotypes to stable and variable genotypes, while it is not observed in comparison of two other structures. The reason may be that adaptive genotypes never exhibit very poor performance

as they take advantage of good environments; therefore, when the difference is large enough, the comparison becomes certain at a fast rate.

It can be further observed from Fig. 2 that perfectly stable genotype is always selected over the adaptive and variable genotypes when the G main effect is bigger, even if they are very close; that is, for the plots in the top and bottom rows the probabilities are always greater than $\frac{1}{2}$ for all pair-wise comparisons. On the other hand, the comparison among highly variable genotypes and adaptive ones is more complicated. Even though adaptive genotypes are more stable than the corresponding highly variable genotype, it does not automatically imply that it will be preferred. The reasons for this may be complex and depend both on the $G \times E$ structure and the distribution of environments in the target environments. The highly variable genotype will be worse in poor environments and better in very good environments, but if the distribution of environments is such that very good environments are likely and very poor environments are unlikely then the highly variable genotype may be better than the adaptive one even though it is less stable. This is reflected in the results reported in Figure 2. This illustrates the complexity of the the process and how the probabilistic comparison incorporates both the $G \times E$ structure and the distribution of environments.

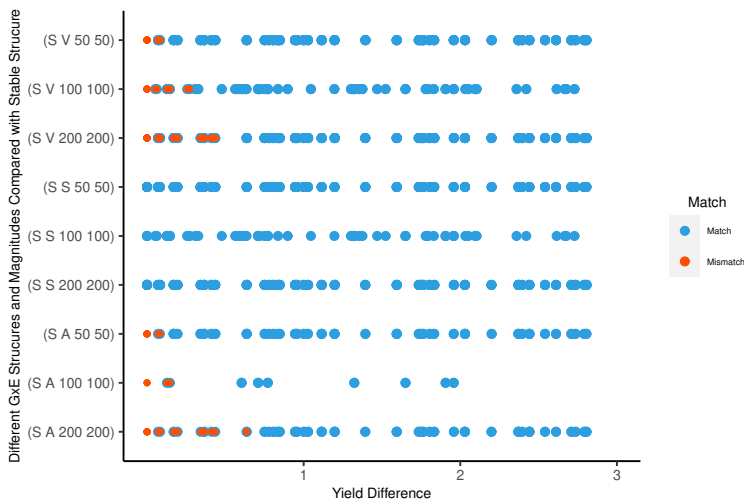


Fig. 3 Mean and probabilistic comparison match/mismatch for stable genotypes and three levels of the magnitude of $G \times E$ interactions (50, 100 and 200). The red dots indicate pairs where the genotype with the better mean is *not* the genotype that is *more likely* to perform better.

We have observed that the genotype with the better mean is not always the genotype that is more likely to perform better, given the different $G \times E$ structure of the pair of genotypes, as well as the distribution of the environmental effects in the target environments. We now further explore whether decisions

made using the new probabilistic approach characterized by Equation (4) are different than decisions made using the traditional mean approach characterized by Equation (2), that is, is $S_{Prob} = S_{Mean}$? And if yes, what circumstances lead to such differences? Thus, we look at the fraction of pair-wise comparisons that are different as a function of both the mean yield difference and difference in $G \times E$ structure. For further insights, we calculate the results of probabilistic comparisons when the average $G \times E$ magnitude is either reduced by half or doubled. Fig. 3 illustrates the cases where the results of two methods differ or are the same with respect to their difference in yield for different $G \times E$ structures and magnitudes. As it can be seen, when comparing two perfectly stable genotypes, the comparison between two stable genotypes is always be the same independent of the magnitude of the $G \times E$ interaction effects. On the other hand, when the interaction structures differ the two methods may result in different results. This is especially true when the genotypes are close and the fraction of disagreements increases when the $G \times E$ magnitude increases.

The results reported in this section demonstrates that comparing genotypes using means does not always equate to selecting the genotype that is most likely to perform better. When the difference in main effects is large or the two genotypes have the same $G \times E$ structure then the genotype with the better mean is also the one that is more likely to perform better across the target environments. However, in cases where the means are closer and the $G \times E$ structure differs, this may not be the case and indeed in such cases the simulated data experiments show that in some cases the majority of pair-wise comparisons differ in these two criteria. These differences also depend on the magnitude of the $G \times E$ effects and the specific subsets of environments that are observed. The proposed probabilistic comparison incorporates all of those factors in order to identify the genotype that is more likely to perform better across the environments.

4.2 Rapeseed Data

The probabilistic comparison of the six rapeseed cultivars across all locations and years is shown in Fig. 4. We first observe that Glacier is expected to win against all of the other cultivars, and would hence rank highest according to a probabilistic ranking. In particular, it has an estimated 59% win probability against both of the other high performers (Bridger and Bienvenu). Previous reported results from an AMMI model show that both Bridger and Bienvenu have strong interaction effects with the environment whereas the Glacier cultivar had the least interaction effects; and other stability measures show that Glacier is the most stable cultivar, whereas Bridger and Bienvenu are the least stable [16, 17]. This appears to be a likely explanation for why the probabilistic ranking of Glacier is better than its mean ranking. In other words, by simultaneously accounting for the the mean performance, the $G \times E$ interaction effects and the type of environments, the probabilistic approach concludes that Glacier is the preferred cultivar across this set of test environments.

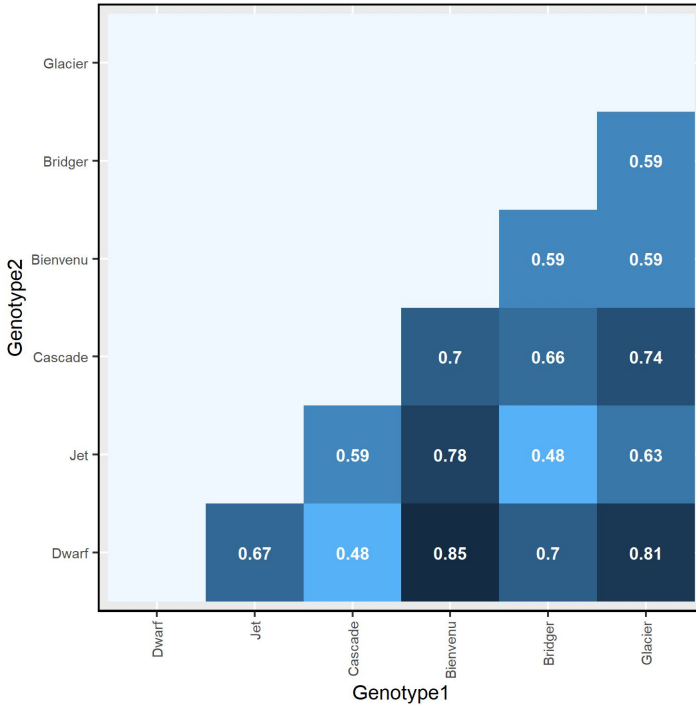


Fig. 4 Heat maps showing probabilistic comparison of rapeseed cultivar pairs across all environments. Cultivars are ordered according to main effects. The figure shows the winning probability of the cultivar on the x-axis over the cultivar on the y-axis.

Table 5 Rapeseeds' ranks based on probabilistic and some traditional ranking methods.

Cultivar	Ranking Method						
	Probabilistic Rank	Mean Yield	Shukla's Variance	Kang's Rank-Sum	Superiority Index		
					All	Fav.	Unfav.
Glacier	1	2	1	1	2	2	3
Bridger	2	3	5	3	5	6	1
Bienvenu	3	1	6	2	1	1	4
Cascade	4	4	4	3	3	4	2
Jet	5	5	3	3	6	5	5
Dwarf	6	6	2	3	4	3	6

Of course there are other approaches that combine mean and stability, and a classic approach is Kang's rank-sum [14] that adds the mean-rank and the rank according to Shukla's variance [19]. Table 5 compares the probabilistic rank, mean rank and Kang's rank-sum. We observe that Kang's rank-sum also captures Glacier as the preferred cultivar but it ranks Bienvenue second and then has a four-way tie for third. This tie reflects the lack of granularity that occurs when combining rank versus the original metrics. Table 5 also includes a

comparison with the superiority index of Lin and Binns [20], which is interesting for comparison as it can be calculated for all environments, only favorable environments and only unfavorable environments. The superiority index ranks either Bridger or Bienvenu as first. However, it is interesting that only Glacier is in the top three for all three variants; whereas Bridger is worst for favorable environments and Bienvenu is poor for unfavorable environments. The more consistent ranking of Glacier appears to be reflected in its probabilistic rank.

As noted in the simulated data experiments, typically difference between the mean-based and probabilistic comparison occur when the differences in main effects are close. Table 4 shows those mean differences for the rape-seed cultivars. We note as before that the Bienvenu cultivar ranks the highest based on mean yield, followed closely by the Glacier cultivar (2487.95 and 2481.78 kg/ha). Given those close mean yield values and the superior stability of Glacier, it is not surprising given the results for the simulated data that the probabilistic analysis ranks Glacier ahead of Bienvenu. However, differences can also occur when the mean differences are larger. The cultivar that is third according to mean yield rank (Bridger) also ranks above Bienvenu according to the probabilistic analysis. The mean yield of Bridger is almost 118 kilograms smaller than Bienvenu and only 101 kilograms more than the Dwarf cultivar that has the smallest yield. Furthermore, both cultivars have large G×E effects [17]. Nonetheless, the probabilistic comparison shows that Bridger is more likely to perform better than Bienvenu across these environments, demonstrating that the explanation for the mean versus probabilistic comparison differences may be complex. This conclusion also disagrees with Kang's rank-sum, which is heavily influenced by the mean comparison and hence has Bienvenu ranked second and Bridger in a four-way tie for third. A hint for why Bridger is more likely to perform better may be found by noting it is ranked first according to the superiority index in unfavorable environments, but the probabilistic comparison reached this conclusion without the need to look at multiple measures and weighting their importance.

To illustrate how the probabilistic approach incorporates the specific target environments of interest, Fig. 5 shows the pairwise comparison probabilities for two subsets of environments: a) locations in the south (TX, GA, SC, MS, NC and TN) for all years, and b) locations in the northwest (OR, WA, ID and MT)) for all years. We observed that as the target locations change as do the pairwise probabilities and any resulting probabilistic ranking. As expected the probabilistic comparison reflects the set of environments, both in terms of the distribution of the target environments and the interaction of cultivars with those environments.

5 Conclusions

Experimental genotypes are frequently compared according to mean phenotypic response, and such comparison is then used as the basis of further decision-making. This paper introduces a new method to compare genotypes,

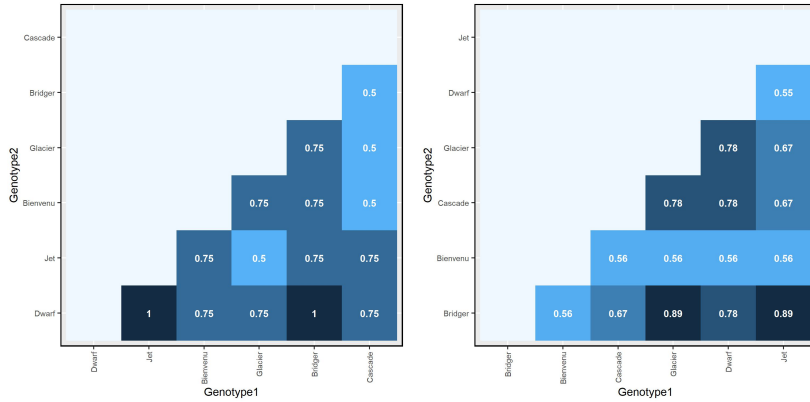


Fig. 5 Heat maps showing probabilistic comparison of rapeseed cultivar pairs across two subsets of environments - southern locations (left) and northwestern locations (right). Cultivars are ordered according to main effects. The figure shows the winning probability of the cultivar on the x-axis over the cultivar on the y-axis.

namely to estimate and select genotypes based on which genotype is most likely to perform better across a set of environments. We further evaluate how often this differs from mean-based selection, and evaluate the underlying reasons for why the genotype with the better mean is not always most likely to perform best. The probabilistic approach accounts for both mean and $G \times E$ interactions and thus incorporates both main effects and in some sense the stability and adaptability of the genotypes. However, the probability estimates account for not only the uncertainty that stems from the selection of planting locations but also the distribution of locations in which they have been planted.

Results from simulated data demonstrate that when the difference in main effects is large or the two genotypes have the same $G \times E$ structure then the genotype with the better mean is also the one that is more likely to perform better across the target environments. However, these cases are likely to be considered straightforward in practice as any selection approach is likely to result in the same decision. It is in the more difficult cases where the results show that the probabilistic and mean-based approaches differ. Specifically, when the means are close, the $G \times E$ structure differs, and the magnitude of the $G \times E$ effects is large relative to the main effects, then the majority of pairwise comparisons may differ.

Results from previously analyzed rapeseed data indicate that the probabilistic method effectively captures comparative information about the cultivars, which may be only partially captured using multiple measures using prior approaches. By incorporating the main effects, the $G \times E$ effects, and the distribution of the environmental effects for the target environments into a single probability, the probabilistic method provides a new way to identify the cultivar that is more likely to perform better across each set of target environments.

References

- [1] van Eeuwijk, F.A., Bustos-Korts, D.V., Malosetti, M.: What should students in plant breeding know about the statistical aspects of genotype × environment interactions? *Crop Science* **56**, 2119–2140 (2016). <https://doi.org/10.2135/cropsci2015.06.0375>
- [2] Abberton, M.: *Molecular plant breeding*. by y. xu. wallingford, uk: Cabi (2012), pp. 752, £59.95. isbn 9781845939823. *Experimental Agriculture* **48**(3), 461–461 (2012). <https://doi.org/10.1017/S0014479712000178>
- [3] Cooper, M., DeLacy, I.H.: Relationships among analytical methods used to study genotypic variation and genotype-by-environment interaction in plant breeding multi-environment experiments. *Theoretical and Applied Genetics* **88**, 561–572 (1994). <https://doi.org/10.1007/BF01240919>
- [4] Gauch, H.G.: A Simple Protocol for AMMI Analysis of Yield Trials. *Crop Science* **53**(5), 1860–1869 (2013). <https://doi.org/10.2135/cropsci2013.04.0241>
- [5] Piepho, H.P.: Best Linear Unbiased Prediction (BLUP) for regional yield trials: a comparison to additive main effects and multiplicative interaction (AMMI) analysis. *Theoretical and Applied Genetics* **89**(5) (1994). <https://doi.org/10.1007/BF00222462>
- [6] Piepho, H.P., Möhring, J., Melchinger, A.E., Büchse, A.: BLUP for phenotypic selection in plant breeding and variety testing. *Euphytica* **161**(1-2) (2008). <https://doi.org/10.1007/s10681-007-9449-8>
- [7] Vargas, M., Combs, E., Alvarado, G., Atlin, G., Mathews, K., Crossa, J.: Meta: A suite of sas programs to analyze multienvironment breeding trials. *Agronomy Journal* **105**, 11–19 (2013). <https://doi.org/10.2134/agronj2012.0016>
- [8] Malosetti, M., Ribaut, J.M., van Eeuwijk, F.A.: The statistical analysis of multi-environment data: Modeling genotype-by-environment interaction and its genetic basis. *Frontiers in Physiology* **4** MAR (2013). <https://doi.org/10.3389/fphys.2013.00044>
- [9] Yang, R.-C.: Mixed-Model Analysis of Crossover Genotype-Environment Interactions. *Crop Science* **47**(3) (2007). <https://doi.org/10.2135/cropsci2006.09.0611>
- [10] Becker, H.C., Léon, J.: Stability Analysis in Plant Breeding. *Plant Breeding* **101**(1), 1–23 (1988). <https://doi.org/10.1111/j.1439-0523.1988.tb00261.x>

- [11] Eberhart, S.A., Russell, W.A.: Stability parameters for comparing varieties. *Crop Science* **6**, 36–40 (1966)
- [12] Happ, M.M., Graef, G.L., Wang, H., Howard, R., Posadas, L., Hyten, D.L.: Comparing a Mixed Model Approach to Traditional Stability Estimators for Mapping Genotype by Environment Interactions and Yield Stability in Soybean [*Glycine max* (L.) Merr.]. *Frontiers in Plant Science* **12** (2021). <https://doi.org/10.3389/fpls.2021.630175>
- [13] Olivoto, T., Lúcio, A.D.C., da Silva, J.A.G., Marchioro, V.S., de Souza, V.Q., Jost, E.: Mean Performance and Stability in Multi-Environment Trials I: Combining Features of AMMI and BLUP Techniques. *Agronomy Journal* **111**(6), 2949–2960 (2019). <https://doi.org/10.2134/agronj2019.03.0220>
- [14] Kang, M.S.: A rank-sum method for selecting high-yielding, stable corn genotypes. *Cereal Research Communications* **16**(1/2), 113–115 (1988). Accessed 2022-06-10
- [15] Efron, B.: Better Bootstrap Confidence Intervals. *Journal of the American Statistical Association* **82**(397), 171–185 (1987). <https://doi.org/10.2307/2289144>
- [16] Shafii, B., Mahler, K.A., Price, W.J., Auld, D.L. *Crop Science* **32** (1992). <https://doi.org/10.2135/cropsci1992.0011183x003200040017x>
- [17] Shafii, B., Price, W.J.: Analysis of genotype-by-environment interaction using the additive main effects and multiplicative interaction model and stability estimates. *Journal of Agricultural, Biological, and Environmental Statistics* **3** (1998). <https://doi.org/10.2307/1400587>
- [18] Wright, K.: *Agridat: Agricultural Datasets*. (2021). R package version 1.19. <https://kwstat.github.io/agridat/>
- [19] Shukla, G.K.: Some statistical aspects of partitioning genotype-environmental components of variability. *Heredity* **29**, 237–245 (1972). <https://doi.org/10.1038/hdy.1972.87>
- [20] Lin, C.S., Binns, M.R.: A superiority measure of cultivar performance for cultivar \times location data. *Canadian Journal of Plant Science* **68**(1), 193–198 (1988) <https://doi.org/10.4141/cjps88-018>. <https://doi.org/10.4141/cjps88-018>

Statements and Declarations

This research is partially funded by Syngenta Seed and the authors are grateful for their support. The authors have no relevant financial or non-financial competing interests to disclose.