ARTICLE IN PRESS

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Field Crops Research xxx (2006) xxx–xxx

Field Crops Research

www.elsevier.com/locate/fcr

Methods to evaluate wheat cultivar testing environments and improve cultivar selection protocols

W.E. Thomason *, S.B. Phillips

Department of Crop and Soil Environmental Sciences, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA

Received 27 October 2005; received in revised form 7 March 2006; accepted 25 March 2006

Abstract

Analysis of cultivar by environment (C × E) interaction can improve efficiency of crop breeding efforts. Variety selection and recommendation based on wheat (*Triticum aestivum* L.) yield testing trials could possibly benefit from this type of analysis as well. The objectives of the present work were to evaluate methods to identify relevant testing environments and improve the predictive value of data from wheat cultivar yield trials in the eastern US. The data come from 32 site years of winter wheat yield trials conducted in Virginia. Biplot analysis revealed that all current testing sites were relevant and that most performed similarly within a year. The degree of relationship or dissimilarity among environments was also evaluated using straight-line distance between observations in variable space measured as the squared Euclidean distance (ED). Analysis using the ED method revealed that all environments contained the centroid and were thus representative testing environments, similar to results from the biplot analysis. Biplots were effective at identifying cultivars and testing locations that were major sources of C × E interaction. Biplots and best linear unbiased predictions (BLUPs) were used to compare cultivar performance across environments. In a separate evaluation, the ED from the centroid to a cultivar mean was used to weight past relative yields for that cultivar and increased the predictability of future yield of a cultivar in three of four seasons. Weighting by ED decreased the number of site years needed to develop confidence in the yield stability of a particular cultivar from six to three. Utilizing the BLUPs for future grain yields, predictive ability of future performance after 1 year was 40% better and overall was 25% better than that achieved by weighting with ED. Overall the BLUPS method of estimating future performance was more accurate and more reliable than weighting with ED.

Keywords: Wheat; Cultivar by environment; Grain yield prediction

1. Introduction

The stability of wheat (*Triticum aestivum* L.) cultivar performance and quality traits is often difficult to accurately estimate due to the interaction of individual cultivar performance with environmental factors. Grain yield stability is affected (Robert, 2002) as are kernel weight and protein content (Groos et al., 2002). Much attention has been devoted to analysing cultivar by environment (C × E) interactions to improve crop breeding efforts but variety selection and recommendation based on wheat yield testing trials could also benefit from this analysis. Such analyses have the potential to reduce the number of site years needed to develop the ‘confidence’ necessary to recommend a particular cultivar across a range of environments, thereby reducing the adoption time for cultivars with superior genetic yield potential.

Stability of cultivar performance for high grain yield across varied environments and broad adaptation are the goals of most wheat breeding and testing programs. Interactions between cultivar and environment often confound the genetic differences that affect yield among wheat cultivars (Brennan and Byth, 1979; Baril, 1992; Yan, 2002). The usual solution to this problem is to evaluate cultivars across a large number of environments or site years to estimate yield potential across randomly occurring cycles of normal and extreme conditions (Rosielle and Hamblin, 1981). However, some previous work suggests that selection for maximum yielding wheat cultivars should be conducted solely in high yielding environments

* Corresponding author. Tel.: +1 540 231 2988; fax: +1 540 231 3075.
E-mail address: wthomaso@vt.edu (W.E. Thomason).

0378-4290/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.fcr.2006.03.007

(Allen et al., 1978). Due to constraints of time and resources, the numbers and locations of testing sites is often less than ideal. This can lead to incorrect selection of inferior lines or promotion of genetic materials that are not truly superior. The solution is to obtain sufficient information while testing across a limited number of environments.

Biplots can be used to approximate a two-way data set in graphical format. This is done using singular value decomposition to break the data matrix into component matrices. The first two principle components (PC1 and PC2) are used to produce a two-dimensional biplot. If a large portion of the variation is explained by these components, a rank-two matrix, represented by a biplot, is appropriate (Yan and Kang, 2003). Using a mixed model analysis may offer superior results when the regression of genotype by environment interaction on environment effect does not explain all the interaction (Piepho, 1997; Yan et al., 2002). This is especially likely to occur when performing analysis over multiple trials with different numbers of replications and when not all genotypes are tested in all trials.

Several methods for determining the optimum type or number of testing environments have been proposed (Eagles et al., 1977; Brennan and Sheppard, 1985) as well as for separating genetic and environmental effects (Byth et al., 1976; Piepho, 1994; Kang and Gauch, 1996). Past research has also proposed methods to weight data from those environments that would provide the most valuable input (Eagles et al., 1977; Vargas et al., 1999).

The goals of this investigation were to evaluate methods to identify relevant testing environments and to assess methods to improve the predictive value of data from wheat cultivar yield trials in the eastern US.

2. Materials and methods

2.1. Yield data

The yield data used in the present study were taken from annual grain yield trials conducted by the Virginia Tech Wheat Variety Testing Program. All experiments used a randomized complete block design with three replications. Twenty-two released cultivars evaluated over all sites from five growing seasons from 1999 to 2003 were chosen for study. The number of test locations in each year varied from five to seven depending on growing conditions (Table 1). Thus, the total number of site years evaluated was 32. Historically, the sites from Holland, Warsaw, and Eastern Shore are classified as the Coastal Plain, those from Orange and Blackstone are in the Piedmont, and Blacksburg and the Shenandoah Valley site are in the Blue Ridge physiographic region of Virginia (Table 1). These classifications are based on similar soil types and climates. The coefficient of variation for grain yield in these trials ranged from a low of 6 to a high of 15, with a mean of 8.1. Mean yield and standard error for all locations is presented in Table 2.

Estimates of variance components for the factors evaluated in this analysis are presented in Table 3. The magnitude of each component measured as a percentage of the total variance is also presented. Analysis was conducted using the MIXED procedure in SAS and since the goal of the analysis was to evaluate testing locations, the effects of variety and associated interactions were considered random effects (Littell et al., 1996; SAS Institute, 2004).

2.2. Biplot analysis

Biplot analysis was performed using the GGEbiplot software program (Yan, 2001) to evaluate the potential for redundant testing environments within the current complete data set. In the current analysis, cultivars were chosen that appeared in all trials and all years thus making the traditional approach to C $\times$ E interaction appropriate. This tool was also used to examine the performance of two representative cultivars, ‘Roane’ and ‘Sisson’, across all environments. This analysis represents stability of the cultivars across environments in terms of principle component analysis. These results tend to be similar to those obtained from stability analysis using regression (Eberhart and Russell, 1966). Biplot analysis was also used to determine which cultivars yielded well in which environment. This information could identify broadly adapted cultivars that offer stable performance across all sites, as well as cultivars that perform well under specific conditions such as high disease or insect pressure.

2.3. Cluster analysis

The degree of relationship or dissimilarity among environments (redundancy) was also evaluated using

Table 1
Wheat variety yield testing locations and site information, 1999–2003

Location	Physiographic region	Taxonomic class	Average yearly rainfall (cm)	Latitude	Longitude	Testing years
Holland	S.E. Virginia, Coastal Plain	Typic Hapludult	125	36.6426	−76.8227	1999–2002
Painter	Eastern Shore, Coastal Plain	Thermic Hapludult	113	37.5891	−75.7995	1999–2003
Warsaw	E. Virginia, Coastal Plain	Typic Hapludult	106	37.9579	−76.7573	1999–2003
Blackstone	S. Virginia, Piedmont	Typic Kanhapludult	122	37.0710	−77.9888	1999–2002
Blacksburg	S.W. Virginia, Blue Ridge	Ultic Hapludalf	94	37.2320	−80.4213	1999–2003
Orange	N. Virginia, Piedmont	Typic Kanhapludult	108	38.2315	−78.0792	1999–2003
Shenandoah Valley	W. Virginia, Blue Ridge	Typic Dystrudepts	90	37.9605	−79.2306	2000, 2001, 2003

Table 2
Grain yield (Mg ha^{-1}) and standard error at testing sites, 1999–2003

Year	Holland		Painter		Warsaw		Blackstone		Blacksburg		Orange		Shen Valley	
	Yield	S.E.[a]	Yield	S.E.	Yield	S.E.	Yield	S.E.	Yield	S.E.	Yield	S.E.	Yield	S.E.
1999	3.70	0.57	6.65	0.67	4.91	0.68	na	0.66	5.98	0.57	5.04	0.74	4.17	0.54
2000	4.30	0.43	5.24	0.74	5.51	0.79	4.10	0.54	5.64	0.52	5.58	0.42	4.84	0.46
2001	4.03	0.79	5.98	0.80	5.58	0.73	4.57	0.55	5.38	0.85	4.84	0.93	5.98	0.73
2002	5.11	0.52	5.24	0.90	6.59	0.60	3.90	1.29	4.64	0.46	5.85	0.64	na	na
2003	na	na	3.96	0.54	4.70	0.71	na	na	4.17	0.67	4.70	0.90	3.63	0.75

[a] Standard error of the mean.

straight-line distance between observations in variable space. The physical distance between points on the graph indicates the degree of relation; the closer two points are, the more similar the results from those testing environments. This physical distance or proximity of environments was determined using the squared Euclidean distance (ED) measure, one of the most common methods used for binary data (Box et al., 1978). Squared Euclidean distance is calculated:

$$\mathrm{ED}\,(x, y) = \sum (x_\mathrm{i} - y_\mathrm{i})^2$$

where ED is the squared Euclidean distance and x and y represent the location of the yield in a multi-dimensional space; the units of which are the same as the input variables. This measurement is powerful because multiple approaches are possible. Cultivar performance can be modeled in multi-dimensional space with each dimension representing performance in an environment. For example, a triangle can be envisioned with the multi-site cultivar mean as one point and the performance of that cultivar in two separate environments as the other two points. The length of the hypotenuse would then represent the distance between those two environments. The ED calculated between environments could then be evaluated for similarity of cultivar performance. Also, environments can be considered in multidimensional space with each dimension performance of a different cultivar. The first method enables evaluation of cultivar performance across environments and the second allows comparison of environments based on how the same cultivars perform in each. The calculation is squared to eliminate the concern over sign and to give progressively greater weight to objects that are further apart or further from a set point. For the purpose of this work, calculations were performed in a spreadsheet program. Statistical software programs such as SAS (SAS Institute, 2004) include functions that perform these calculations.

Table 3
REML estimates of variance components and percentage of total variance for each component

Term	Estimate	Percentage of total variance
Cultivar	8707.65	11
Location	25149.26	31
Year	22003.08	27
Cul*Loc	3359.17	4
Cul*Year	2120.61	3
Loc*Year	13631.70	17
Error	6566.19	8
Total	81537.65	

2.4. ED metric analysis

Assessment of environments was performed using the ED metric. Cultivar yields were pooled over environments, and the mean value was identified as the centroid, following terminology common for the Euclidean metric. The environments were also evaluated separately for each year, and the ED between each environment and the centroid for that year was determined. This environment-specific factor was used to compare environments within a year and to search for those that most often produced a value similar to the centroid, indicating performance similar to that for the average of all sites in that year. Results from sites consistently near the centroid should be more relevant for predicting overall cultivar performance.

2.5. BLUPS and yield response

The MIXED procedure in SAS (Littell et al., 1996; SAS Institute, 2004) was used to generate BLUPs for grain yield (Cornelius and Crossa, 1999). All factors were considered random effects and separated by year. Satterthwaite's procedure was used to determine the appropriate degrees of freedom for the BLUPs (SAS Institute, 2004). A comparison of actual yield performance versus estimated yield performance based on BLUPs at each testing location and the average of all locations was made to compare discriminating ability of sites.

2.6. Yield weighting with ED

The method of weighting past cultivar yield is based on a process using ED measure and evaluation parallel to that proposed by Brennan and Sheppard (1985). Weighted cultivar yield will hopefully increase the predictive value of past yields on future results and thus enable researchers to have confidence in cultivar recommendations with fewer site years of field testing. Weighted cultivar yield was determined from the sum of the relative yield of the cultivar

(%) in each environment, calculated as follows:

$$\left\{\left(\frac{Y_c}{Y_{em}}\right)\times 100\times\frac{1}{ED_c}\right\}$$

where Y_c is the cultivar yield, Y_{em} the environment mean yield, and ED_c is the straight line Euclidean distance from the centroid for that environment.

3. Results

3.1. Testing environment evaluation

Biplot analysis showed that year had a greater effect than location on cultivar yields since all sites in a given year tended to have smaller angles between them than the same sites across years (Fig. 1e). Environment vectors that have a

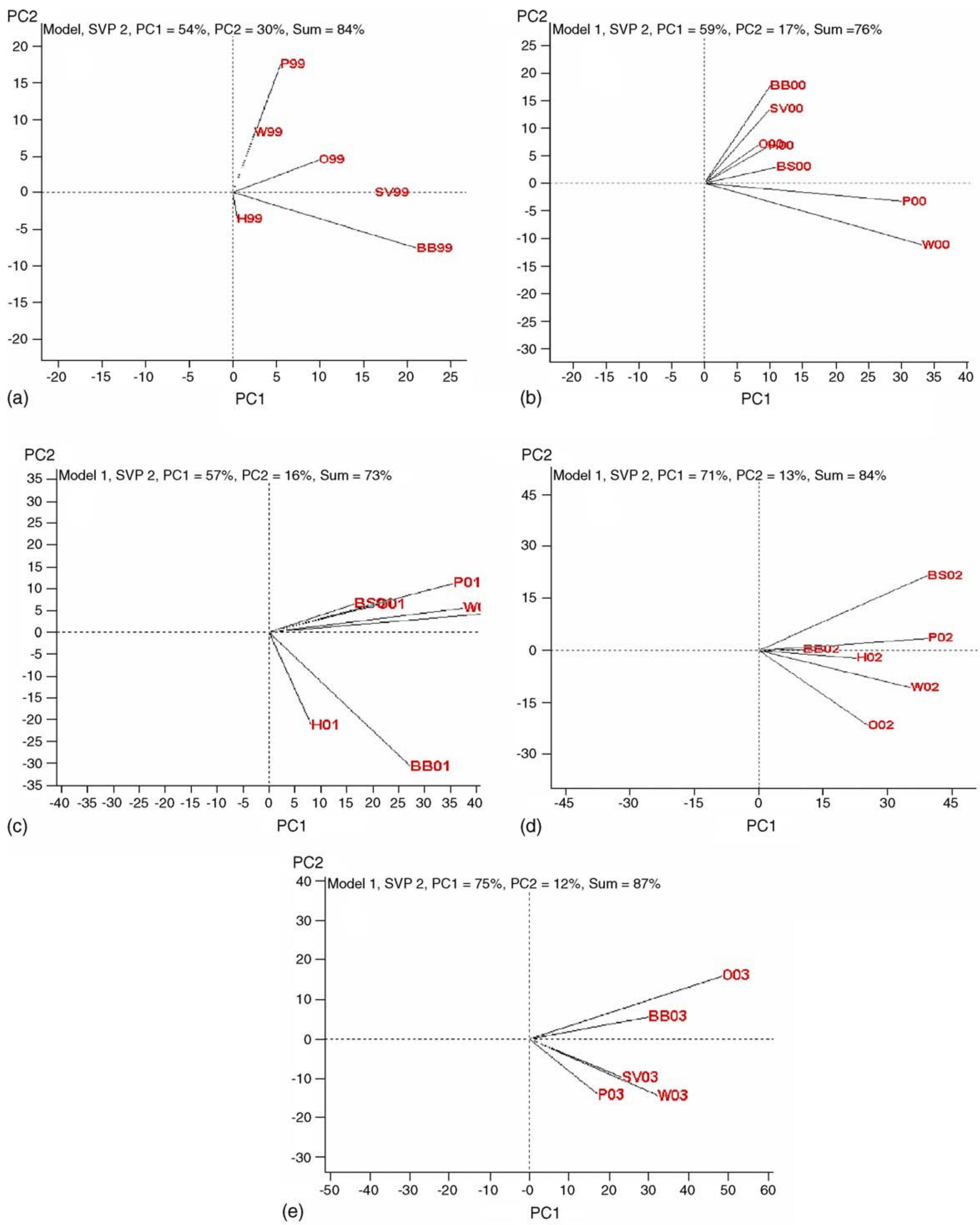

Fig. 1. Biplot and interrelation of all environments (vectors) tested 1999 (a), 2000 (b), 2001 (c), 2002 (d) and 2003 (e). Environments are represented in red and in all capital letters. PC1 and PC2 are first and second principal components, respectively.

small (acute) angle are more similar to each other; therefore it is assumed that conditions in 2003 caused all test sites to perform similarly to one another but differently than the same sites in the other years. Yields in 2003 were, in fact, 0.6 Mg ha^{-1} lower than the average for the other years due to especially wet conditions (Brann et al., 2003). Vector length (physical length of the line leading to a site year) approximates the standard deviation for a particular environment (Yan and Kang, 2003). Since most vectors are relatively short (<30 PC units), excluding those for 2003 (Fig. 1e), most testing environments were visualized to be discriminating and capable of revealing true differences. Aside from 2003, the test environments represented in the analysis were found to be similar, emphasizing the fact that testing across these environments should produce similar results. No unique or outlying locations were noted.

Based on the REML estimates of variance comparing analysis components (Table 3), the largest sources of variation are due to the effects of location and year, 31 and 27% of total variance, respectively. Slightly more than 17% of the total variation was due to the interaction of location and year. Variance components associated with the effect of variety (11%) and interactions of variety with location (4%) and year (3%) are smaller than those associated with location and year. This finding agrees with other studies (Cullis et al., 2000) and implies the same limitations for use of the variety by location interaction as an explanation of how environment will affect varietal performance when not all remaining variance associated with testing environment is explained by the interaction of environment with genotype.

Understanding that year and site location have the largest effect on cultivar yield, it is also noted that researchers have little control over environmental conditions experienced in a particular year. Location, on the other hand, can be changed or modified if sufficient evidence exists to warrant a change. For this reason, cultivar yields from each location were evaluated across years. Vectors with acute angles represent environments where the rank order of cultivar yield was similar. Vectors representing environments that are at right angles are orthogonal to each other and there is little to no correlation in cultivar yield between them. This would be the case for the Holland and Shenandoah Valley sites (Fig. 2). Vectors that are opposite represent crossover interactions. Based on this, cultivar recommendations based on Warsaw data could result in exactly the opposite performance at Blackstone (rank reversal). If this crossover relationship persists then different selection criteria and possibly a different approach in the breeding program would need to be developed for the Southern Piedmont area. Viewing site results across years indicated that sites in the northern Coastal Plain (Painter and Warsaw), northern Piedmont (Orange), and one Blue Ridge site (Blacksburg) were similar (Fig. 2). Blackstone, Holland, and Shenandoah Valley were found to be more similar to one another than to the other sites though they are from the Piedmont, Coastal Plain, and Blue Ridge regions, respectively. Results from the Blackstone site were found to differ most from those of the other testing sites (Fig. 2).

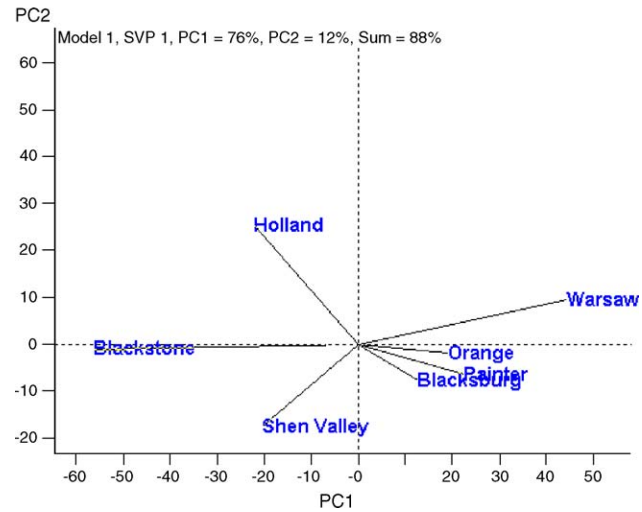

Fig. 2. Biplot of environment mean yields for 21 cultivars averaged across years, 1999–2003. Environments are represented in all capital letters. PC1 and PC2 are first and second principal components, respectively.

These results suggest that cultivar recommendations for locations in most of the state should rely more heavily on data from the sites at Warsaw, Painter, Orange, and Blacksburg. Recommendations for the southern and southeastern regions of the state should probably place more emphasis on the results from Holland and Blackstone. Recommendations for the northern Blue Ridge region should consider the Shenandoah Valley data more heavily.

3.2. *Cultivar stability and comparison across environments*

Yields of 'Roane' and 'Sisson' were compared to one another across all tested environments in Fig. 3. 'Sisson' was consistently higher yielding across all environments

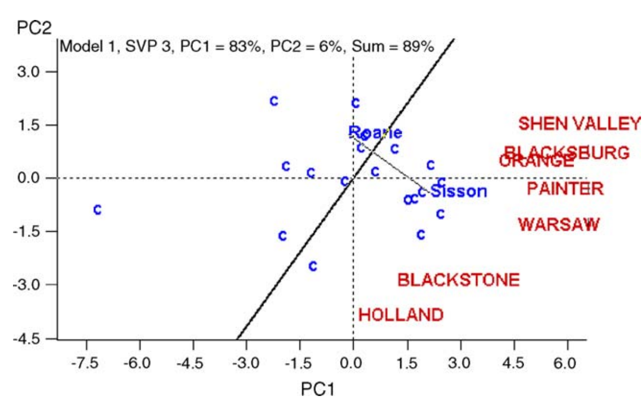

Fig. 3. Biplot comparing the performance of two successful Virginia Tech released cultivars, Roane and Sisson, across all tested environments (1999–2003). Environments are represented in all capital letters. Lowercase 'c' represents other tested cultivars. PC1 and PC2 are first and second principal components, respectively.

compared to 'Roane' as evidenced by the fact that the symbols for all environments are on the same side of the figure with 'Sisson'. Mean yield of 'Roane' (4.90 Mg ha^{-1}) was near the mean yield for all tests while 'Sisson' yields were well above the mean (5.44 Mg ha^{-1}). 'Sisson' had a BLUP value across sites of 0.756 ($P > 0.001$) while the BLUP for 'Roane' was 0.078 ($P = 0.563$) indicating that expected performance for 'Roane' would be near the mean of the tests, while 'Sisson' would be expected to yield higher than the mean. This was proven accurate by further testing. This type of visualization can assist in analysis between any two cultivars across environments. It especially facilitates comparison with a standard or control cultivar.

The evaluation of the best cultivar or group of cultivars for a particular environment is represented by Fig. 4. An outer line or hull is drawn such that cultivars are all contained within that outer line. Perpendicular lines are drawn from the origin to each side of the hull to divide the plot into sections. The analysis places cultivars that excel in a particular environment near the entry for that environment. For example, some cultivars, like Massey, did not appear near any environment entries and were consistently lower yielding than the average at all sites, 'SS535' was near the average, represented by appearance near the dotted line at zero, and 'Sisson', 'McCormick', 'Tribute', and 'SS520' were typically higher yielding than average. The cultivar 'Tribute' yielded especially well at Blackstone, Warsaw, Shenandoah Valley, and Orange while 'SS520' outperformed the others at the remaining sites. This type of data presentation allows identification of cultivars with stable performance across environments as well as an understanding of where, and under what conditions, narrowly adapted cultivars excel. Air temperature post-heading and spring rainfall had the greatest effect on the magnitude of the C × E interaction across sites. Interaction was greatest at sites with the warmest temperatures post heading.

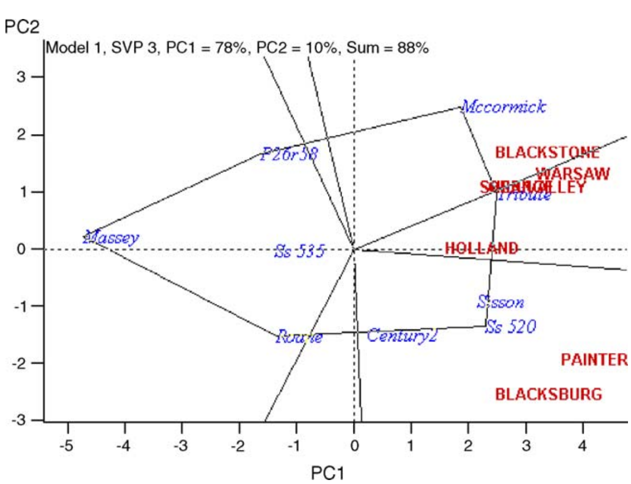

Fig. 4. Biplot of selected cultivar means and environments showing the highest yield cultivar in each environment (1999–2003). Environments are represented in all capital letters. PC1 and PC2 are first and second principal components, respectively.

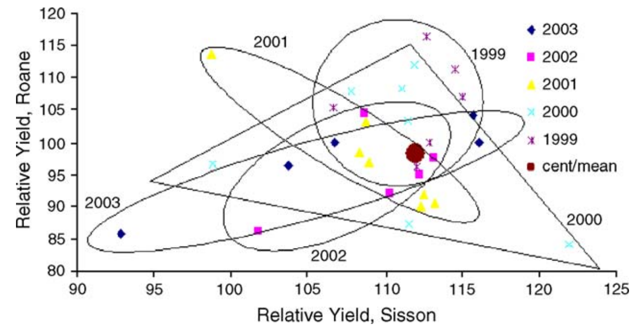

Fig. 5. Comparison of the relative yield, centroid and yearly environment grouping of two successful Virginia Tech released cultivars, Roane and Sisson, across all tested environments (1999–2003).

Fig. 5 uses cluster analysis to compare the relative yields of two successful Virginia Tech cultivars, 'Roane' and 'Sisson', over the 5 years from 1999 to 2003 with each point representing a testing location. The proliferation of points within ten units of the centroid indicates that the most frequent response was near the mean. This suggests that these cultivars performed predictably over most years and locations. However, results in 2000 and 2003 indicated large differences in relative performance, suggesting that further evaluation of those growing seasons is warranted. No consistent pattern of both cultivars increasing or decreasing exists across years, indicating that the two cultivars do not necessarily respond similarly to the same environment, but 'Sisson' is typically higher yielding than 'Roane'. This yield advantage may be due to the fact that 'Sisson' matures earlier than 'Roane' enabling it to take advantage of cooler temperatures during grain fill.

The same data were examined through by-year grouping by outlining the area encompassing the cultivar yield points at each site (Fig. 5). This allows visualization of cultivar performance in different years and identification of unique past years that do not encircle the centroid. An example of how this information can be used to search for explanations for differences would be the large distances between means (variable yields) encountered in the data for the year 2000. This is most likely a result of barley yellow dwarf virus infection, especially with 'Roane', and delayed harvest due to heavy rainfall at some sites. Identification of a normal response zone or centroid allows researchers to easily observe abnormal conditions, such as those in the 2000 testing year, and to estimate future yields with more confidence since outlying data are easily spotted and managed. Identification of abnormal conditions through comparison with a yearly mean yield can also help identify cultivars that perform well under those specific conditions. This can lead to development of narrowly adapted cultivars particularly suited to localized areas or specific conditions such as day length sensitive lines that delay reproductive growth even under warm spring temperatures, enabling them to avoid late season freeze damage. Lines encompassing the yield points for each year show that all the years contain the centroid, again indicating that data from all years and locations is relevant for evaluation. Four of the 5 years evaluated produced similar oval response diagrams (Fig. 5) but in all years, particular sites could be found that gave the more frequently observed yield response, i.e. those found near the centroid based on ED. Nearness is relative to a particular data set and must be evaluated as such. The identification of environments that are more predictive of the overall cultivar mean is important because they can be weighted more heavily during future evaluation than those consistently farther from the centroid. In other words, some sites will probably be better predictors of overall mean yields than others.

Usefulness of an environment for prediction of relative overall yield is indicated by a small ED value between the location mean and the centroid for that year. Those locations testing consistently high, like Blackstone, are not likely to be good overall predictors. Those with consistently low values, such as Blacksburg and Warsaw, should be good predictors of performance (Table 4). Estimation of performance over the entire Commonwealth for the period examined would benefit from putting more weight on data from Blacksburg and Warsaw and less emphasis on data from Blackstone. The other use of these data may be to show that results from Blackstone may be unique to that area and narrowly adapted traits may be of importance, similar to what was noted with biplot analysis. Some past research has demonstrated a precedent for discarding data from environments with ED values >150% relative yield (Brennan and Sheppard, 1985). However, all environments were included in this analysis with the understanding that ED values >150% often represent peculiar or unique results.

3.3. Improving yield predictions

Squared Euclidean distance based on past cultivar performance was also evaluated for the potential to add predictive value to previous yield results through weighting values from each environment. The relative yield of a cultivar weighted (multiplied) by the reciprocal of the ED was compared to the centroid for all cultivars in the environment to determine a weighted relative yield. The

Table 4
Euclidean distance as calculated from the variance of the observations from the centroid, or overall mean, for each site year

Location	Year				
	2003	2002	2001	2000	1999
Holland		39.26	361.45	144.12	484.92
Painter	25.37	44.11	196.72	56.94	729.65
Warsaw	49.82	625.82	49.88	54.68	64.24
Blackstone		108.23	100.57	265.47	256.94
Blacksburg	21.29	64.54	61.28	101.79	80.62
Orange	49.88	144.46	105.63	81.74	196.74
Shenandoah Valley	121.96		196.54	95.68	
Average	64	171	153	114	302

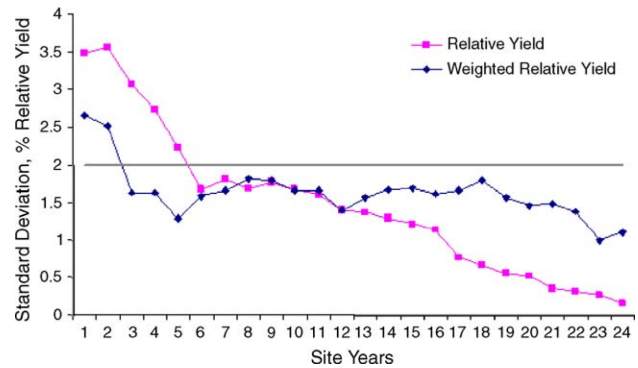

Fig. 6. Standard deviation from the mean for relative cultivar yield and relative cultivar yield weighted by ED for 15 cultivars across 25 site years.

standard deviation was calculated for both relative yield and relative yield weighted by ED for 15 representative cultivars with 25 site years of data over the past 4 years (Fig. 6). Evaluation began in 2000 with each succeeding site year incorporated into a rolling average and the standard deviation re-evaluated based on the additional data from another environment. Standard deviations from relative yield alone decreased to acceptable levels (<2.0) when relative yields from six site years were averaged. This indicates that at least six site years were needed to obtain reliable, repeatable results. Weighting the relative yields with ED increased the reliability of results as indicated by a lower standard deviation and decreased the number of site years necessary to reach a standard deviation <2.0 to 3. Weighting relative yield for future yield prediction was also compared with regression of overall average yield on 1 year, the average of 2 years, the average of 3 years, and the average of 4 years (Table 5). Comparing the relative yield (regression) of 1 year to the next was the best non-weighted method. Weighting using ED increased the degree of prediction in all but 1 year and resulted in R^2 values greater than 0.60 in two comparisons. Surprisingly, even when averaged across the preceding 3 and 4 years, future yields were not predicted as accurately as when the estimates were based on weighted relative yields from the previous year. None of these methods did an excellent job of predicting future yields for a given cultivar. Based on these simple evaluation methods, future yield estimates based on historical yields are not necessarily improved with multiple years of input. Mathematical weighting of relative yield for released cultivars may decrease the number of site years necessary to develop a high level of confidence in their performance and decrease the amount of time necessary to begin recommending cultivars with improved genetic yield potential.

A much more statistically rigorous methodology was also applied to the data to generate the BLUP for cultivar grain yield at each location. A comparison of actual yield performance of the 22 tested cultivars versus estimated yield performance based on BLUP for cultivars at each testing location and the average of all locations is presented in

Table 5
Cultivar yield prediction estimates based on the regression of average relative yield on 1 year, the average of 2 years, the average of 3 years, and the average of 4 years, 1999–2003 (R^2)

	Year				
	1998–1999	1999–2000	2000–2001	2001–2002	2002–2003
Relative yield	1.000	0.3754	0.7715	0.5645	0.4784
Relative yield, 2-year average			0.6190	0.2441	0.3263
Relative yield, 3-year average				0.1794	0.2925
Relative yield, 4-year average					0.2217
Relative yield weighted with ED		0.3071	0.8205	0.6560	0.4990

Table 6. Values represent the regression of predicted yields on actual yields. Higher values indicate a closer relationship between observed and predicted yields. As years of testing increased, the predictive value increased at all locations. The overall accuracy of prediction did not increase when results from all sites were combined to form a statewide value probably because sites such as Holland and Blackstone produced more variable results than the others. Likewise, grouping testing locations by physiographic region did not improve prediction of future yield performance (Table 6).

4. Discussion

This data set comes exclusively from Virginia, and it is not the intent of this paper to attempt to encompass a large land area or region, but rather to illustrate some methods to improve and speed the cultivar recommendation process using local data. These methods are conceptually simple and could be easily applied to data from other regions or from larger areas.

Using BLUPs, biplot, and cluster evaluation, all seven of the Virginia Tech Wheat Variety testing sites evaluated were found to produce relevant results. Visual examination of biplot analysis identified cultivars and environments that exhibited major sources of C $\times$ E interaction as well as those that were stable, similar to results reported by Yan et al. (2000). Cluster analysis provided information on groups of sites where cultivars performed similarly and can potentially be used to identify sites that provide the most broadly applicable data. Some locations, such as Blacksburg, Warsaw, and Shenandoah Valley were found to more accurately reflect the average yield across all sites than others, and, similar to the recommendation of Brennan and Sheppard (1985), it is proposed that future results from these locations be more heavily weighted than results from sites that tend to produce yields that do not mirror the average of all testing sites. For example, if all seven Virginia testing locations were equally weighted, approximately 14.3% of the statewide average would be attributable to any one site. Based on the responses observed, it would be appropriate for the results from Blacksburg and Warsaw to be weighted to represent 20% of the final results with each of the other sites contributing 12%.

Weighting of relative cultivar yield differences based on ED for that environment was used as a method to predict cultivar yield. Weighting cultivar relative yield using ED resulted in better predictions of yield than past relative yield alone. None of these methods for future yield estimates were highly accurate (average $R^2 = 0.56$). Brennan and Sheppard (1985) also reported that weighting increase predictive accuracy with regression coefficients ranging from 0.375 to 0.555. Weighting with ED decreased the number of site years necessary to achieve acceptable standard deviations and be confident in stability of cultivar performance from six to three.

Employing mixed models and the BLUP for future grain yields, predictive ability of future performance after 1 year was 40% better than that achieved by weighting with ED and 25% more reliable over 4 years of continuous evaluation. Overall, this method of estimating future cultivar performance was more accurate and more reliable than weighting with ED.

As more wheat cultivars are developed and marketed by commercial companies, cultivars are introduced and aggressively marketed before adequate local testing has occurred to develop confidence in local performance. Applying statistical techniques that provide more interpretative value from the same collected data will allow cultivar recommendations with more confidence from fewer site years of field testing. In an era of increasing costs and decreasing resources for cultivar yield trials the value of a more rigorous statistical approach to yield trial results can also be realized in terms better use of limited resources.

Table 6
Accuracy of BLUP predicted to actual yield response as a percentage of the site maximum and region maximum for each testing site for 22 cultivars across multiple sites and multiple years of testing (R^2)

Test location	1 year	2 years	3 years	4 years
Holland	0.57	0.65	0.71	0.78
Painter	0.73	0.82	0.86	0.90
Warsaw	0.86	0.89	0.91	0.92
Blackstone	0.73	0.77	0.81	0.84
Blacksburg	0.80	0.84	0.87	0.91
Orange	0.73	0.79	0.85	0.89
Shenandoah Valley	0.81	0.88	0.91	0.94
Statewide	0.74	0.80	0.85	0.88
Test region				
Coastal plain	0.72	0.79	0.83	0.87
Piedmont	0.73	0.78	0.83	0.87
Blue ridge	0.81	0.86	0.89	0.93

References

Allen, F.L., Comstock, R.E., Rasmussen, D.C., 1978. Optimal environments for yield testing. Crop Sci. 18, 747–751.

Baril, C.P., 1992. Factor regression for interpreting genotype-environment interactions in bread wheat trials. Theor. Appl. Genet. 83, 1022–1026.

Box, G.E.P., Hunter, W.G., Hunter, J.S., 1978. Statistics for Experimenters. John Wiley and Sons, Inc., New York.

Brann, D.E., Griffey, C., Behl, H., Rucker, E. Pridgen, T., 2003. Small Grains in 2003. Va. Coop. Extn. Pub. 424-001. Blacksburg, VA: Virginia Polytechnic Institute and State University.

Brennan, P.S., Byth, D.E., 1979. Genotype $\times$ environmental interactions for wheat yields and selection for widely adapted wheat genotypes. Aust. J. Agric. Res. 30, 221–232.

Brennan, P.S., Sheppard, J.A., 1985. Retrospective assessment of environments in the determination of an objective strategy for the evaluation of the relative yield of wheat cultivars. Euphytica 34, 397–408.

Byth, D.E., Eisemann, R.L., DeLacy, I.H., 1976. Two-way pattern analysis of a large data set to evaluate genotypic adaptation. Heredity 37, 215–230.

Cornelius, P.L., Crossa, J., 1999. Prediction assessment of shrinkage estimators of multiplicative models for multi-environment cultivar trials. Crop Sci. 39, 998–1009.

Cullis, B.R., Smith, A., Hunt, C., Gilmour, A., 2000. An examination of the efficiency of the Australian crop variety evaluation programmes. J. Agric. Sci. 135, 213–222.

Eagles, H.A., Hinz, P.N., Frey, K.J., 1977. Selection of superior cultivars of oats by using regression coefficients. Crop Sci. 17, 101–105.

Eberhart, S.A., Russell, W.A., 1966. Stability parameters for comparing varieties. Crop Sci. 6, 36–40.

Groos, C., Robert, N., Bervas, E., Charmet, G., 2002. Genetic analysis of grain protein-content, grain yield, and thousand-kernel weight in bread wheat. Theor. Appl. Genet. 106, 1032–1040.

Kang, M.L., Gauch, Jr., H.G. (Eds.), 1996. Genotype-by-Environment Interaction. CRC Press, Boca Raton, FL.

Littell, R.C., Milliken, G.A., Stroup, W.W., Wolfinger, R.D., 1996. SAS System for Mixed Models. SAS Institute, Inc., Cary, NC.

Piepho, H.P., 1994. Best linear prediction (BLUP) for regional yield trials: a comparison to additive main effects and multiplicative interaction (AMMI) analysis. Theor. Appl. Genet. 89, 647–654.

Piepho, H.P., 1997. Analyzing genotype-environment data by mixed models with multiplicative terms. Biometrics 53, 761–766.

Robert, N., 2002. Comparison of stability statistics for yield and quality traits in bread wheat. Euphytica 128, 333–341.

Rosielle, A.A., Hamblin, J., 1981. Theoretical aspects of selection for yield in stress and non-stress environments. Crop Sci. 21, 943–946.

SAS Institute, 2004. SAS/STAT 9.1 Users Guide. SAS Institute, Inc., Cary, NC.

Vargas, M., Crossa, J., van Eeuwijk, F.A., Ramirez, M.E., Sayre, K., 1999. Using partial least squares regression, factorial regression, and AMMI models for interpreting genotype $\times$ environment interaction. Crop Sci. 39, 955–967.

Yan, W., 2001. GGEbiplot—A windows application for graphical analysis of multienvironment trial data and other types of two-way data. Agron. J. 93, 1111–1118.

Yan, W., 2002. Singular-value partitioning in biplot analysis of multi-environment trial data. Agron. J. 94, 990–996.

Yan, W., Kang, M.J., 2003. GGE Biplot Analysis. CRC Press, Boca Raton, FL.

Yan, W., Hunt, L.A., Sheng, Q., Szlavnics, Z., 2000. Cultivar evaluation and mega-environment investigation based on the GGE biplot. Crop Sci. 40, 597–605.

Yan, W., Hunt, L.A., Johnson, P., Stewart, G., Lu, X., 2002. On-farm strip trials vs. replicated performance trials for cultivar evaluation. Crop Sci. 42, 385–392.