The *Journal of Cotton Science 17:279–284 (2013)*
http://journal.cotton.org, © The Cotton Foundation 2013

BREEDING AND GENETICS

Statistical Evaluation of the Cotton Regional Breeders Testing Network (RBTN)

Daryl T. Bowman, Vasu Kuraparthy*, Ted Wallace, and Don C. Jones

ABSTRACT

In the U.S., Ted Wallace coordinates a Regional Breeders Testing Network (RBTN) for cotton that is sponsored by Cotton Incorporated. The objective of this program is to provide testing sites for public breeders and geneticists across a regional basis. This study looked at various statistical aspects of field testing with emphasis on lint yield. Ten years of data (107 environments) were examined. The objectives were to: 1) determine any relationship between error variance and mean lint yield, 2) establish a procedure for rejecting less precise data, and 3) discern the most optimum testing sites in the program. To achieve the first objective the natural log of error variances were regressed on the natural log of mean environmental lint yields and tested for significance. The "b" value of 0.85 was significant indicating that the error variance increased with increasing yield levels. Using a procedure previously published on rejecting less precise data, the second objective was met. Five of the 107 environments were deemed imprecise and should not have been included in across-location tables. A genotypic index regression method was followed to ascertain the most desirable test sites in the program. Twelve of 23 test sites did an acceptable job of discriminating the entries. Thus, by eliminating nearly half of the test sites more reliable data can be produced, less seed would be required, and more efficient use of resources would be achieved. Sites with less than desirable tests might contribute by collecting data on disease tolerance, morphological traits, or insect resistance, etc. and thus might still be valuable in the cotton RBTN.

The cotton (*Gossypium hirsutum* L.) Regional Breeders Testing Network (RBTN) was established as a means of providing a range of testing sites for public breeders and geneticists. Their elite breeding lines are tested in diverse environments in exchange for providing a testing site at their location. Over the years the program has had sites ranging from Virginia to California. The number of sites within a particular year might vary from eight to 15 but the tendency has been to include more sites. This paper reports on a statistical study of the testing program with the objective of finding useful suggestions to improve efficiency. It is generally accepted that certain locations have more desirable resources and are probably more suited to field testing. Various statistical aspects of field testing will be covered in this paper.

It is generally thought that error variance increases with mean yield (Snedecor and Cochran, 1967), but Bowman and Rawlings (1995) showed that for three maturity groups of corn (*Zea mays* L.) and one maturity group of soybean (*Glycine max* L.) there was not a significant relationship. There were significant relationships in barley (*Hordeum vulgare* L.), oat (*Avena sativa* L.), wheat (*Triticum aestivum* L.), and for one maturity group of soybean (Bowman and Rawlings, 1995). Allen et al. (1978) also showed a relationship between error variance and yield. Even though these latter relationships were significantly different from zero, they were also significantly different from 2.0, which is the threshold one needs to use the coefficient of variation (CV) as an indicator of experimental validity (Bowman and Watson, 1997). The relationship between error variance and lint yield in cotton has not been well documented.

Combining data across locations has been a topic of several papers over the years starting with Yates and Cochran in 1938. The issue is combining data from locations with different levels of precision. The RBTN attempts to cover environments representative of the growing area. Even though these trials are nearly identical in size, they are not always similar in precision. Due to soil variability and other factors, results from these trials can be quite variable. The assumptions of the analysis of variance are not always

---

D.T. Bowman and V. Kuraparthy*, Department of Crop Science, North Carolina State University, Raleigh, NC 27695-7620; T. Wallace, Department of Plant & Soil Sciences, Mississippi State University, Starkville, MS 39762; and D. Jones, Cotton Incorporated, 6399 Weston Parkway, Cary, NC 27513
*Corresponding author: vasu_kuraparthy@ncsu.edu

met, such as the assumption that all experimental units within a block are uniform.

Yates and Cochran (1938) stated that less precise locations might need to be discounted but were reluctant to advocate discarding a test altogether. However, Brown et al. (1983) suggested that some less precise trial data might need to be discarded. Various methods of weighting less precise data have been proposed, such as weighting each location by the inverse of the variances of the location. Any genotype-by-location interaction would confound the results when averaged across locations, but that is not the emphasis of this paper. Over the years, several trials have been dropped from across-location tables based mostly on field observations; however, there can be other factors that inflate experimental error. Decisions to drop or include data from various trials should be based on scientific methods if field observations fail to detect variability that leads to inflated error. Bowman and Rawlings (1995) proposed a rejection procedure based on historical error variances. Their method first determined any relationship between error variance and mean productivity level and then established a criterion level based on that relationship (regression) with the pooled error calculated over a large historical data set.

It is understood that locations with higher means have more weight than locations with lower means. Using a percentage of individual entry means has been suggested as a method to avoid this imbalance in weighting data (Yau and Hamblin, 1994). This would entail calculating a percentage for each individual entry compared to the location overall mean. This method has been used by Virginia Tech University in their official crop variety trials; it is particularly helpful when all entries are not tested at all locations. Using relative yield instead of actual yield converts the variance of an entry from biological into an agronomic measure of stability. When using this method with large number of test sites, the impact of any one test with high means is diluted, but the benefit of calculating a useable stability measure might be warranted.

As the number of testing sites in the RBTN expands, the amount of seed required to enter all locations increases. Most breeders have limited supplies of seed in the early stages of testing. A pragmatic approach can be to analyze which locations are more discriminating in separating entries. Thus important questions can be: 1) should we confine the program to the most optimum testing sites and, 2) how do we determine which are the most optimum sites? There are various proposed methods to discern discrimination (Allen et al., 1978; Brown et al., 1983; Yan, 2001). Brown et al. (1983) suggested a genotype index regression method whereby individual location data would be regressed on overall location data. The most optimum sites would have high regression values and high coefficients of determination. Bowman et al. (1993) used this method to analyze a flue-cured tobacco (*Nicotiana tabacum* L.) testing program. Of 13 sites in the program, four were found to be undesirable and needed improved management or discarding altogether.

The focus of this paper is on lint yield, which is the most important but less heritable trait. The objectives of this study were to: 1) determine the relationship between error variance and productivity level in the RBTN, 2) create a rejection procedure for trials with undesirable precision using the procedure of Bowman and Rawlings (1995), and 3) determine the optimum test sites in the program using the genotypic regression method of Brown et al. (1983).

MATERIALS AND METHODS

Data from the RBTN for the years 2002 through 2011 were gleaned from the website www.cottonrbtn.com. Only those locations that were included in the overall location averages were used in this study. Decisions on suitability of the data had been made on some locations that were not included in the overall tables. Twenty-four different locations were used in one year or another during this 10-yr time frame (Table 1). The locations and years used in this study are shown in Table 1. Each location/year combination will be referred to as an environment. There were a total of 107 environments in this study. Management practices used by the various cooperators were common to those locales. Plot size varied with location in terms of length and row width, but most cooperators used two-row plots. All locations in this study used four replicates. Thus error degrees of freedom were the same for all locations within each year.

Table 1. Maximum SD for lint yield in the Regional Breeders Testing Network.

Maximum SD for lint yield			
Metric Units		English Units	
Average Lint Yield (kg/ha)	Maximum SD (kg/ha)	Average Lint Yield (lbs/a)	Maximum SD (lbs/a)
560	132	500	118
1120	225	1000	201
1680	340	1500	271
2240	382	2000	341

To determine a relationship between error variance and lint yield, the natural log of error variances and mean yields at each location were calculated. Then the logs of the individual location (i) error variances ($\ln(s^2)$) were regressed on the logs of the individual location mean yield ($\ln(\hat{Y}_i)$). We then obtained an estimate of $\beta_0 = \ln(\alpha)$ and β, and were able to test the significance of the relationship.

Details of the procedure to follow in establishing a rejection procedure is described by Bowman and Rawlings (1995). Essentially if the relationship is significant, then the expected variance for any particular trial is calculated from the fitted regression using the mean of the trial or

$$\ln(\sigma^2) = \beta_0 + \beta \ln(\hat{Y}_i)$$

The next step is to identify a value “k” such that the error variance of only a few trials exceeds k times the expected variance for that yield, or $s_i^2 > k\,\sigma^2$. We should expect only a few trials having unacceptably low precision as defined as high error variance. This is shown on the log-log graph of error variances and yield where a fitted regression line delineates trials with low precision from those with acceptable precision (Fig. 1). The study by Bowman and Rawlings (1995) used a k factor of 2.0. By using this factor the authors only rejected 22 of 422 trials, approximately 5%. In the current study we also applied a k of 2.0. To apply this criterion to any new trial one would compare the observed error variance of the new trial, s_i^2, to $k\,\sigma^2$, which would be the computed maximum error variance based on the mean yield of that trial.

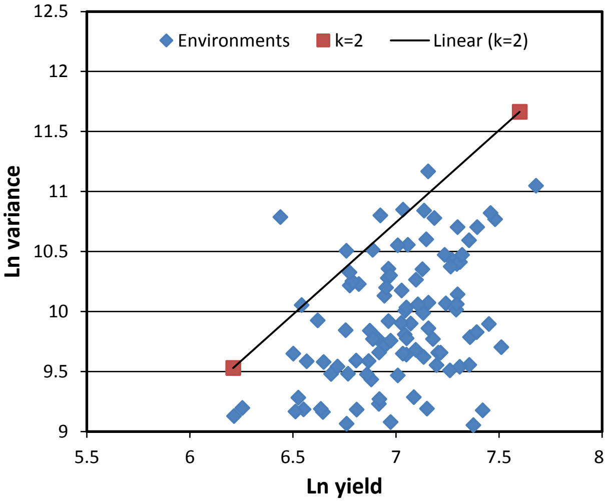

Figure 1. Ln (variance) vs. ln (mean yield) for trials in the cotton Regional Breeders Testing Network for the years 2002 to 2011 along with a line of acceptability.

For determining optimum test sites, we regressed individual location entry means on across-location means using the genotypic index regression method by Brown et al. (1983). The only value used in this study was lint yield, although many other traits were measured. In addition, coefficients of determination were calculated for each location as an average across all years. Using the definition of Brown et al. (1983) an optimum test site would have high regression coefficients (b) and high coefficients of determination (r^2). By using this data set, any bias caused by including less optimum or less precise data would be minimized.

RESULTS AND DISCUSSION

Regressing the natural log of the error variance on the natural log of the location mean yields resulted in a b value of 0.85, which was significantly different from 0 at the 0.01 level. Thus error variances tend to increase as productivity level improves. However the relationship is not strong enough, i.e., it is significantly less than 2.0, to warrant the use of the CV as a test of validity. Thus a maximum acceptable error for trials in this program will need to be adjusted according to environmental means. The pooled error for the entire data set was 2,661,586 or $\ln(s^2) = 10.12$ with 8,400 degrees of freedom.

In establishing a criterion for rejecting trials with low precision, the pooled error was used. Figure 1 shows the log of error variances for each location on the Y axis and the log of mean yields on the X axis. The fitted regression line using k = 2 delineating the trials with acceptable precision from trials with low levels of precision is also shown.

Five of 107 locations/environments or 4.7% had error variances above the fitted line. These locations were: Stoneville (MS)-2007 with a mean yield of 964 kg/ha and a least significant difference (LSD) of 300; Mississippi State-USDA (MS)-2007 with a mean yield of 700 kg/ha and an LSD of 346; Tifton (GA)-2008 with a mean yield of 1138 kg/ha and a LSD of 348; Maricopa (AZ)-2011with a mean yield of 1270 kg/ha and an LSD of 370; and College Station (TX)-2011 with a mean yield of 465 kg/ha and an LSD of 465. All of these environments had LSDs above 300 kg/ha. There might not have been obvious reasons to discard these trials but the error variances were abnormally high for their mean yields. However, having an LSD above 300 kg/ha cannot be used solely as an indicator of low precision. For example, Alexandria (LA)-2009 had an LSD above 300 kg/ha but mean yield = 1820 kg/ha, which meant that its standard deviation (SD) =236 is acceptable. Also, the Mississippi-USDA (MS) site in 2011 had an LSD = 334, but it had a mean yield of 1655 kg/ha and SD = 236, again acceptable. One can simplify the process by calculating the maximum acceptable SD for predetermined yields. Table 2 lists a range of maximum acceptable SDs for various yield levels.

Five environments were deleted from the overall location tables from 2006 through 2011; these were not included in the data set used in this study. Of these five, only one had an LSD above 300 kg/ha. Thus, there could be reasons for discarding data other than level of precision alone. Any factor that can bias the data must be taken in account. For example, in 2011 there were entries that had poor germination at certain environments and those plots were essentially blank throughout the season; thus the adjacent plots were not bordered. Therefore, the adjacent plots should have been discarded or only the bordered row harvested and included in the data set.

Optimum test sites should discriminate among entries. Breeders assume that there are differences among entries, which becomes their null hypothesis. However, the null hypothesis for official variety trials is that there are no significant differences. Regression of location/environment entry means on across-location means provides a measure of discrimination. The higher the b value the larger the separation among entries (Table 1). Those locations with b values equal and above 1.0 are desired.

Table 2. Regression coefficients and coefficients of determination (r^2) for test sites used in the cotton RBTN from 2002 to 2011.

Location	Regression coefficient (b)											
	Year										Avg.	r^2
	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011		
Tallassee, AL		**0.8**	**0.91**	**1.3**	**0.5**	**0.65**	**1.25**	**0.69**		**0.58**	**0.84**	**0.3**
Keiser, AR	**0.54**	**0.46**	**0.78**	**1.29**	**0.89**	**0.67**	**0.77**	**0.83**	**1.02**	**0.64**	**0.79**	**0.44**
Maricopa, AZ					**1.23**	**1.37**			**1.1**	**0.37**	**1.02**	**0.26**
Shafter, CA									**0.6**	**1.03**	**0.82**	**0.29**
Albany, GA	**0.87**	**1.2**							**1.24**		**1.10**	**0.3**
Tifton, GA						**1.4**	**0.62**	**1.39**	**0.66**	**0.76**	**0.97**	**0.42**
Alexandria, LA	**0.93**		**0.05**	**0.75**	**0.54**	**1.07**	**0.73**	**0.94**		**1.44**	**0.81**	**0.39**
Bossier City, LA	**0.82**	**0.86**	**0.79**	**1.16**	**1.04**	**0.9**					**0.93**	**0.35**
St. Joseph, LA									**1.04**	**1.26**	**1.15**	**0.62**
Miss. State, MS	**1.75**	**0.98**	**0.46**	**1.35**	**1.41**	**1.19**	**0.86**				**1.14**	**0.45**
Stoneville, MS	**1.75**	**1.24**	**1.22**	**0.99**	**1.94**	**1.49**	**1.26**	**1.52**	**0.71**	**1.31**	**1.34**	**0.5**
Miss. State USDA, MS		**0.82**	**0.74**			**0.74**			**1.5**	**1.35**	**1.03**	**0.43**
Winterville, MS	**1.47**										**1.47**	**0.59**
Rocky Mt., NC	**0.66**			**0.94**	**1.18**				**0.78**		**0.89**	**0.43**
Las Cruces, NM							**0.9**	**0.53**	**1.08**	**0.58**	**0.77**	**0.32**
Hartsville, SC	**0.26**	**1.64**	**0.74**								**0.88**	**0.26**
Florence, SC				**0.27**		**0.53**	**1.18**	**1.47**		**0.66**	**0.82**	**0.42**
Jackson, TN									**1.67**	**1.07**	**1.37**	**0.55**
College Station, TX			**1.06**	**1.19**	**1.13**				**1.33**	**1.14**	**1.17**	**0.45**
Idalou, TX	**1.8**		**1.75**								**1.78**	**0.53**
Lubbock, TX	**0.06**	**0.06**	**1.35**	**0.76**	**0.14**		**1.22**	**0.51**	**0.53**		**0.58**	**0.22**
Weslaco, TX								**1.12**			**1.12**	**0.69**
Suffolk, VA									**0.74**		**0.74**	**0.55**

Tallassee (AL) had below average b and r^2 values with b values above 1.0 in only two of eight years (Table 1). Keiser (AR) also had below average b values, average r^2 values, and b values above 1.0 in only two of 10 years. Maricopa (AZ) had an average b value of 1.02; if the data from 2011 as mentioned above were discarded, then the average b value would be 1.24 making it a desirable test site. Its average r^2 value is below average suggesting rank changes or a genotype-by-environment interaction, thus the data might not be pertinent to the rest of the testing region. Shafter (CA) had b values below 1.0, but an average r^2 value; it had only one year where b = 1.0. Albany (GA) had b values average above 1.0 but below average r^2 values, which might indicate ranking interaction; this site had b > 1.0 in two of three years. Tifton (GA) had an average b value near 1.0 and an average r^2 value. If the 2008 data were discarded as mentioned previously, then the average b value would be 1.06; with b > 1.0 in two of five years and an average b >1.0, this site would be acceptable.

Alexandria (LA) had below average b values but an average r^2 with only two of seven years having a b > 1.0 (Table 1). Bossier City (LA) also had below average b values, below average r^2 value, and only two b values > 1.0 in six years. St. Joseph (LA) was only in the program two years but in both years the b values were > 1.0 with above average r^2 values making it a desirable test site.

Mississippi State-MAFES (MS) had high b values and a good r^2 values; five of seven years this site had a b value >1.0 making it a desirable test site (Table 1). The Stoneville (MS) site also had good b values and above average r^2 values. It had b values above 1.0 in nine of 10 years; if you remove the 2007 data, then the b value would average 1.32, making this an excellent test site. Mississippi State-USDA (MS) had below desirable b values and below averager^2 values; but if the undesirable data in 2007 is removed, b = 1.1 making it an acceptable test site. The Winterville (MS) site was used one year and it had above average values making it a highly desirable site if available.

The Rocky Mount (NC) site had data in four years and only had a b value above 1.0 one time; thus it had a below desirable average b value but average r^2 values making it an undesirable site (Table 1). Las Cruces (NM) also had four years of data and below average values making it a less than desirable test site; it only had one year with acceptable values. Hartsville (SC) had three years of data and only had one year of discriminating data making it a less than desirable test site. Florence (SC) had five years of data and only two of those years were discriminating; however, its r^2 values were acceptable. Jackson (TN) was an acceptable site even though only two years of data were collected; it had good values in both years.

In Texas there have been four locations involved in the program (Table 1). College Station (TX) had above average b values and average r^2 values; all five years of data had good b values. If the 2011 data were discarded, then the average b = 1.18, making it an excellent test site. Idalou (TX) was used two years and produced good data both years making it a desirable test site. Lubbock (TX) was used eight times and had below average b and r^2 values with only two years of data producing b > 1.0; this would be an undesirable test site. Only one year of data is available from Weslaco (TX) and the data appeared acceptable. Finally, the Suffolk (VA) site was included one year and had undesirable b values.

In conclusion, of the 23 test sites in this data set 12 would be deemed desirable. They are: Maricopa (AZ), Albany (GA), Tifton (GA), St. Joseph (LA), Mississippi State-MAFES (MS), Mississippi State-USDA (MS), Stoneville (MS), Winterville (MS), Jackson (TN), College Station (TX), Idalou (TX), and Weslaco (TX). These test sites represent a range of environments from the humid Southeast (Georgia), to the Midsouth (Louisiana, Mississippi, and Tennessee), to the Southwest (Texas and Arizona). Wallace and Jones (2008) also identified Maricopa (AZ), Stoneville (MS), and Mississippi State as having discriminating sites for yield in 2006. Blanche et al. (2008) identified via GGE biplot that Bossier City (LA) was a poor discriminating site, whereas St. Joseph (LA), College Station (TX), and Mississippi State were better discriminating test sites. Some of these sites might not be available or were only available at the time. By limiting the number of yield test sites the breeder can gain in two ways. First, limited seed would be wisely allocated to the best testing sites. Secondly, limiting test sites to the more discriminating locations makes it much easier to interpret data and draw proper inferences. Some sites could still be used for collecting other data such as disease data in California and Texas, and insect resistance data in Arkansas and Mississippi State-USDA (MS). Thus cooperators could contribute and still participate.

Criteria should be established that would take some of the guess work out of deciding when to include data if field observations fail to detect abnormal variability and when to exclude data based on a scientific approach or method. Historical data indicate that the error variances increase as yield increases (b = 0.85) so any criteria must account for this correlation/relationship.

ACKNOWLEDGEMENTS

We thank the participants of the RBTN for making available the lint yield data used in the current study.

REFERENCES

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

Blanche, S.B., G.O. Myers, W.D. Caldwell, and T.P. Wallace. 2008. Determining selection gains and discriminating environments via GGE biplots. J. Crop Improvement 21:13–25.

Bowman, D.T., C.A. Wilkinson, J.M. Moore, M.G. Stephenson, and R.E. Currin, III. 1993. Optimum farm test sites for flue-cured tobacco. Tobacco Sci. 37:1–3.

Bowman, D.T., and J. O. Rawlings. 1995. Establishing a rejection procedure for crop performance data. Agron. J. 87:147–151.

Bowman, D.T., and C.E. Watson. 1997. Measures of validity in cultivar performance trials. Agron. J. 89:860–866.

Brown, K.D., M.E. Sorrells, and W.R. Coffman. 1983. A method for classification and evaluation of testing environments. Crop Sci. 23:889–893.

Snedecor, G.N., and W.G. Cochran. 1967. Statistical methods. 6th ed. Iowa State Univ. Press, Ames, IA.

Wallace, T.P., and D.C. Jones. 2008. Cotton regional breeders testing network—GGE biplot analysis. p. 889–893 *In* Proc. 2008 Beltwide Cotton Conf. Nashville TN. 8-11 Jan. 2008. Natl. Cotton Counc. Am., Memphis, TN.

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

Yates, F., and W.G. Cochran. 1938. The analysis of groups of experiments. J. Agric. Sci. 28:556–580.

Yau, S.K., and J. Hamblin. 1994. Relative yield as a measure of entry performance in variable environments. Crop Sci. 34:813–817.