2002 BELTWIDE COTTON CONFERENCES, ATLANTA, GA – JANUARY 8-12

BIPLOT ANALYSIS OF THE NATIONAL COTTON VARIETY TEST

Gerald O. Myers
LSU Agricultural Center
Dept. of Agronomy
Baton Rouge, LA

Abstract

Plant genotype/variety evaluation is an integral part of plant breeding research. In the advanced stages of this research, multilocation trials become important as a tool to assess variety yield and stability. Such trials do, however, produce a large amount of data that can be difficult to condense and interpret. A recently developed software application, GGEbiplot, allows for the graphical inspection of two-way data in an interactive manner. The objective of this research was to investigate the utility of GGEbiplot by analyzing lint yield data from the 1997-2000 National Cotton Variety Trials. After removing environmental variation, the model (which is based on genotype and genotype by environment interactions) explained 84% of the yield variation. The biplot made it is easy to see which varieties were best adapted to a particular location. The biplot further indicated that testing environments fell into three sector groups based upon vertex varieties. Cluster analysis based upon between location correlations identified five environmental clusters which is also a reduction from the currently utilized seven in the testing program. The calculation of an average tester coordinate indicated that two locations, St. Joseph, LA and Stoneville, MS were most representative of the average environment and were able to discriminate between varieties quite efficiently. The varieties closest to being ideal for both yield and stability across all of the testing locations were Stoneville 474 and Suregrow 125. The ability of GGEbiplot to analyze multi-location data and perform which-won-where queries should prove to be valuable in advancing plant breeding research.

Introduction

The National Cotton Variety Testing Program is a uniform system of reporting data from cotton-yield trials across the US Cotton Belt. This program is coordinated by the USDA Agricultural Research Service and conducted annually at select locations of cooperating State Agricultural Experiment Stations and the Agricultural Research Service. It is composed of a set of National, regional, and interregional standard varieties. Given the large number of testing locations, there are several relevant questions that can be investigated:

1. A producer is interested in selecting varieties that will perform well on his farm. Typically, this will involve looking at yield data from location(s) nearest to theirs. The type of analysis performed here will allow him/her to identify additional, similarly performing locations to aid in the selection of potential varieties for planting.
2. For companies in the business of developing new varieties, the type of analysis performed here would allow them to identify testing locations that are discriminate and to determine locations that are redundant.
3. Investigate the interrelationships among test locations.
4. Examine the mean performance and stability of tested varieties.

The foundation for this type of analysis is founded upon the fact that for a single variety, its yield is a function of:

Yield =Environmental effect + Genotype effect + Genotype x Environment Interaction

While environment is the single largest contributor to yield variation between locations in a multi-environment trial (typically on the order of about 80% with genotype(G) and genotype x environment interaction (GE) each accounting for about 10%), only G and GE are relevant to variety evaluation and mega-environment identification. While numerous methods of analyzing G and GE exist, one method that is finding increasing favor is based upon the use of principal components analysis. The Additive Main effects and Multiplicative Interaction (AMMI) model (Gauch, 1988; Zobel et al., 1988) is historically the most important application of principal components analysis to multi-environment trial (MET) data. Recently, Yan (2000,2001) developed a graphical approach to analyzing MET data called GGEbiplot which is also based upon principal components analysis but appears to have several advantages over AMMI analysis. Presented here are the results of applying GGEbiplot to yield data from the 1997 – 2000 National Cotton Variety Trials. The ability of the program to address questions similar to those posed earlier will be discussed.

Materials and Methods

Yield data for Upland cotton varieties from the 1997 through 2000 National Cotton Variety Trials was extracted from published data sources. Locations are typically subdivided into 7 regions: Eastern, Delta, Central, Blackland, Plains, Western, San Joaquin (Table 1; Figure 1). The total number of testing locations and varieties combined over 1997-2000 was 28 locations and 101 varieties. Since not all varieties were tested in all locations nor in all four years, reduced data sets were developed to facilitate analysis. Additionally, variety mean yields at each location were used rather than data on a per replicate basis.

Data were analyzed using the Windows based software program GGEbiplot (Yan, 2001). The basic model for a GGE biplot is

$$Y_{ij}-Y_j = \lambda_1\xi_{i1}\eta_{j1} + \lambda_2\xi_{i2}\eta_{j2} + \varepsilon_{ij}$$

where

Y_{ij}	=	average yield of genotype *i* in environment *j*
Y_j	=	average yield over all genotypes in environment *j*
λ_1 and λ_2	=	the singular values for PC1 and PC2, respectively
ξ_{i1} and ξ_{i2}	=	the PC1 and PC2 scores, respectively, for genotype *i*
η_{j1} and η_{j2}	=	the PC1 and PC2 scores, respectively, for environment *j*
ε_{ij}	=	residual for genotype *i* in environment *j*

Results and Discussion

From Figure 2, the model explained 55 + 29% = 84% of the yield variation due to GGE (not to be confused with the total yield variation, which includes environment (E) as well as G and GE). Lines connecting the varieties at the periphery form a multisided polygon. Lines from the origin, drawn perpendicular to the polygon sides divide the polygon into sectors and all test environments fall into the sectors. Stoneville LA887 was the vertex variety in one sector which contains locations TIPTONOK, CHICKOKDRY, ARTNMIRR, and LAMESATXDRY. This means that Stoneville LA887 was the best cultivar in these four environments over the four year period 1997 – 2000. In the rightmost sector containing CSTATIONTX, UPARKNM, BEETX and others, of the three varieties in this sector, Suregrow 747 was the best performing variety followed by Deltapine Nucotton33B. No locations fell into sectors containing the varieties AllTex Atlas, Maxxa, or Paymaster HS26, indicating that these varieties were not the best in any of the locations.

Note that all of the locations fall into three sectors. This compares to the seven regions currently defined by the NCVT program. This implies that only three testing regions may need to be distinguished. Looking at locations historically included in the Plains region, it is apparent that both LUBBOCKTXIRR and CHILLTXDRY do not perform similarly. In fact, by examining Figure 3, it is evident that these locations are highly, negatively correlated since they are at almost a 180° degree angle from the other Plains Region locations. Interestingly, three locations widely separated from each other, DALLASTX, CLARKAR, and AUBURNAL are highly correlated.

To use this information to answer the questions posed earlier:

1. A producer in near CLARKAR may want to see how varieties he is interested in planting perform in DALLASTX and AUBURNAL to help in his decision and pay little attention to how the candidate varieties performed in other Plains Region locations.
2. A company trying to reduce the number of test locations for growing nurseries may not need to test in both CHICKOKDRY and TIPTONOK (or as another example in WSIDECA and LUBBOCKTXIRR) since these locations discriminated among varieties in a similar manner.
3. In Figure 4, the average tester coordinate plot shows the representativeness and discriminating ability of the locations. The vector length (in red) is a measure of discriminating ability: the longer the vector, the more discriminating the location. The greater the distance of a location from this vector the less representative it is. Thus SVILLEMS and SJOELA were highly discriminating and representative. Conversely, TIPTONOK and CHICKOKDRY were neither discriminating or representative. CSTATIONTX while being discriminating was not very representative of the average location. Location rankings based on the discriminating ability and representativenss is presented in Figure 5. The center of the concentric circles is where an *ideal* location should be and since this is on the line it would be absolutely representative of the average location. The closer a location is to this virtual location, the better it is as a test location.

4. Figure 6 ranks the varieties based on both average yield and stability. The center of the concentric circle represents an ideal variety and, since it is on the line, it is absolutely stable. Stoneville 474 was the closest to being an ideal variety since it combined high yield and was the most stable of the eight varieties. SureGrow 747 while being the highest yielding variety was less stable. Maxxa, while stable was very low yielding. It must be noted, however, that it is an Acala type cotton and should not be rigorously compared to the other Upland cotton varieties.

Conclusions

Advances in the stability analysis of variety trial data are coinciding with increased interest by producers and industry in the topic. These programs are providing relatively easy, graphical tools for interpreting what can be a vast amount of data. Biplot analysis of the 1997-2000 National Cotton Variety Test data indicates that producers may wish to consider looking at similarly performing locations that are not necessarily defined by geographic proximity in making their decisions regarding which variety to plant. Those involved in developing new varieties for the marketplace can also be encouraged by the information that biplots can reveal. Such information can be of value to them as they target their research and development resources to the development of high yielding, stable varieties for the future.

References

Gauch, H.G. 1988. Model selection and validation for yield trials with interaction. Biometrics 44:705-715.

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

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.

Zobel, R.W., M.J. Wright, and H.G. Gauch. 1988. Statistical analysis of a yield trial. Agron. J. 80:388-393.

Table 1. Participating Station Codes.

Eastern Region	**Plains Region**
Auburn, AL (AUBURNAL)	Chickasha, OK-irr (CHICKOKIRR)*
Bell Minna AL (BMINNAAL)	Chickasha, OK-dry (CHICKOKDRY)
Florence, SC (FLORSC)	Altus, OK (ALTUSOK)
Delta Region	Tipton, OK (TIPTONOK)
Clarkedale, AR (CLARKAR)	Chilicothe, TX-dry(CHILLTXDRY)
Stoneville, MS (SVILLEMS)	Lubbock, TX-irr (LUBBOCKTXIRR)
St. Joseph, LA (SJOELA)	Lamesa, TX-dry (LAMESATXDRY)
Portageville, MO (PORTAGEMO)	**Western Region**
Central Region	Safford, AZ (SAFFAZ)
Bossier City, LA (BCITYLA)	Artesia, NM-irr (ARTNM)
Weslaco, TX (WESLACOTX)	Las Cruces, NM (UPARKNM)
College Station, TX (CSTATIONTX)	El Paso, TX-irr (ELPASOTXIRR)
Neuces County, TX (BEETX)	Pecos, TX-irr (PECOSTXIRR)
Blackland Region	**San Joaquin Region**
Dallas, TX (DALLASTX)	Five Poins, CA (WSIDECA)
Thrall, TX (THRALLTX)	Shafter, CA (SHAFTERCA)

* DRY = dryland; IRR = irrigated

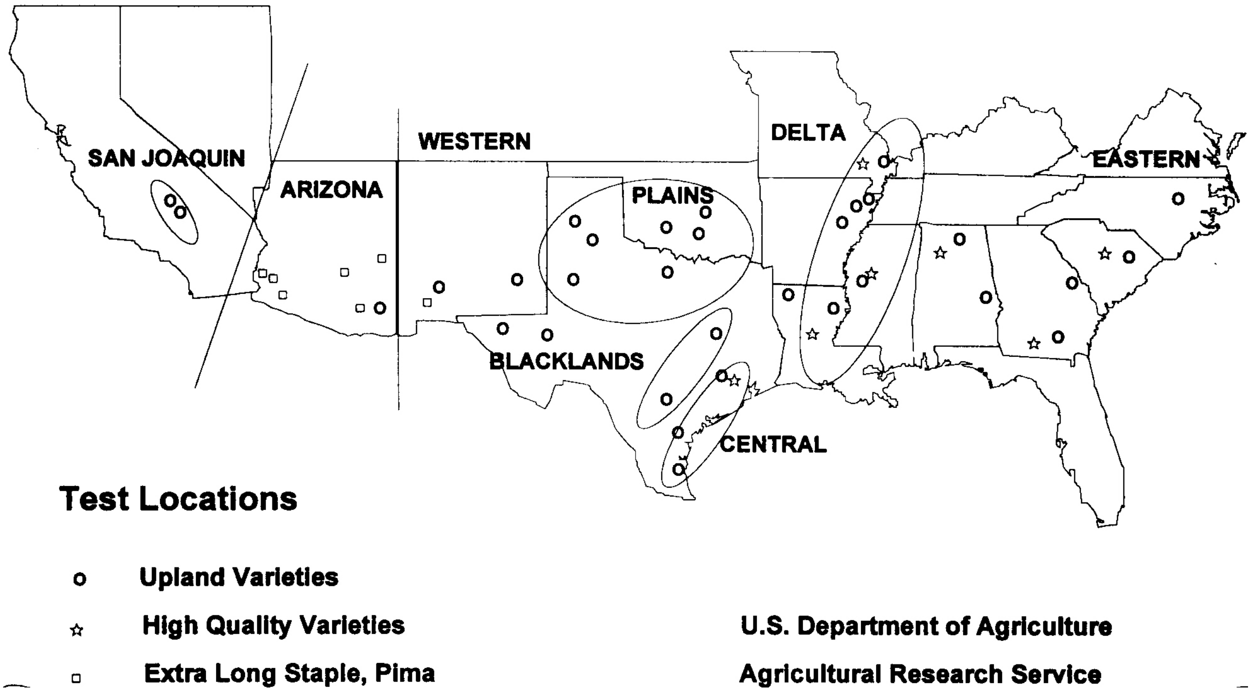
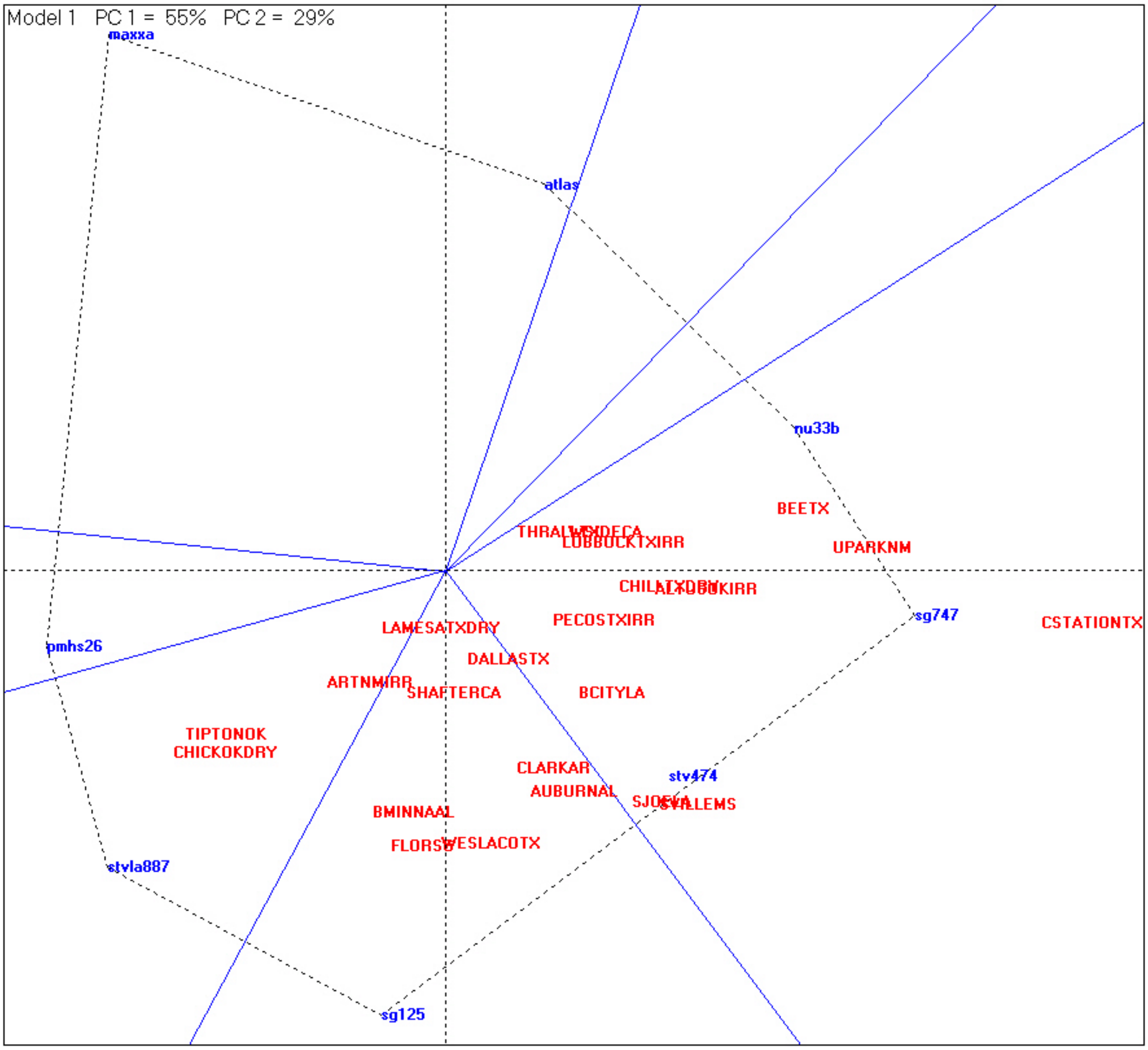


Figure 1. National Cotton Variety Test Locations.



***Accurate positions of cultivars and environments are at the beginning of the labels.**

Figure 2. GGEbiplot of 1997-2000 NCVT*.

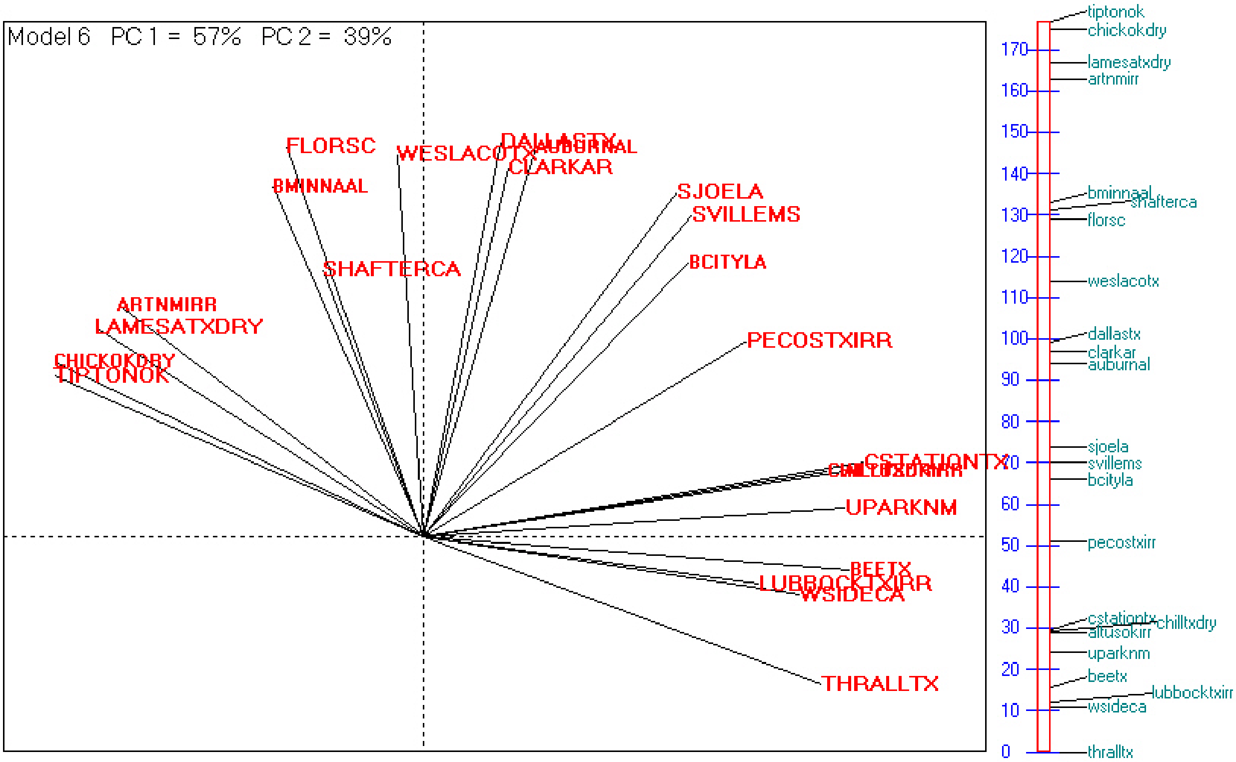


Figure 3. Cluster analysis using r matrix 1997-2000 NCVT.

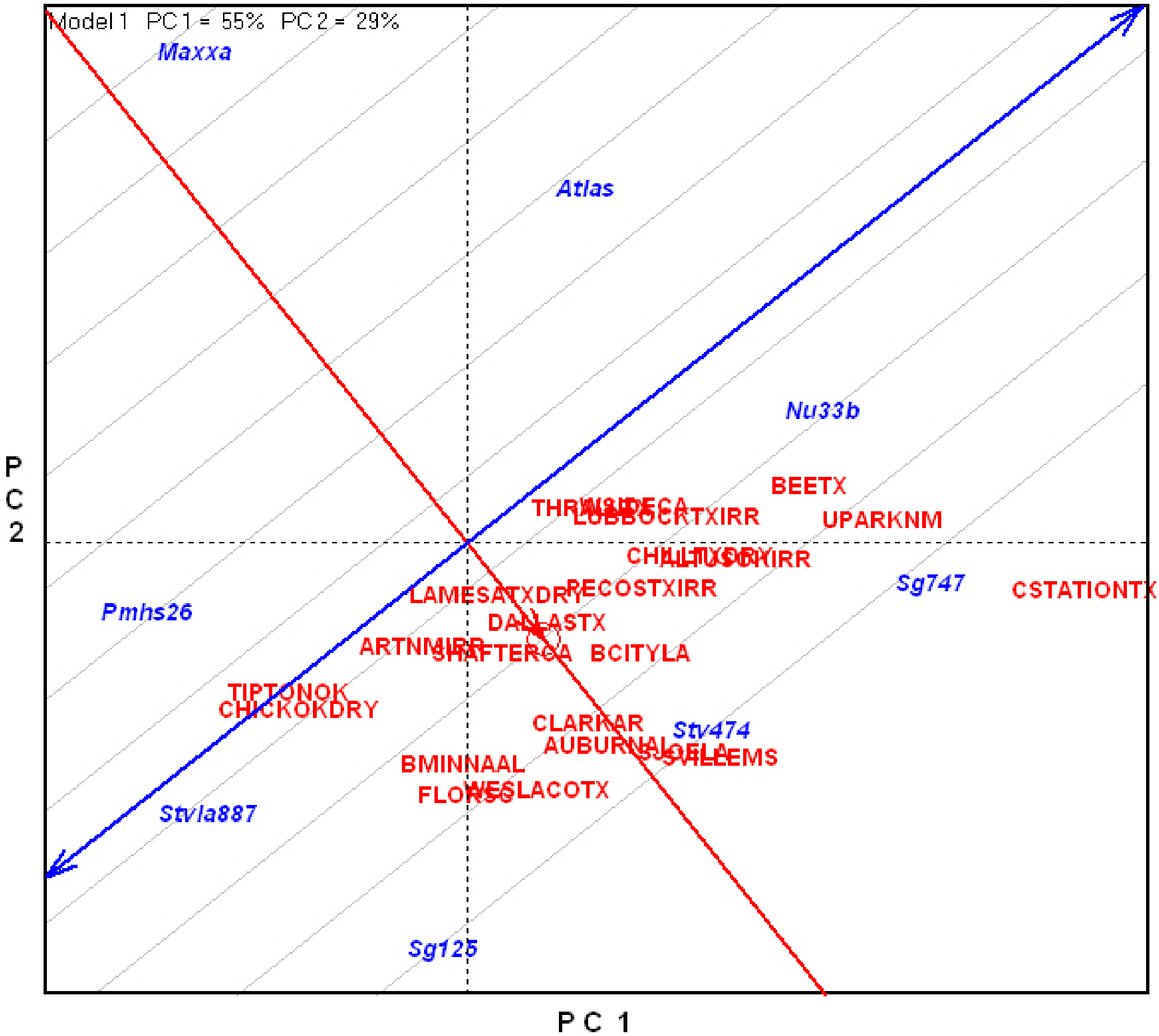


Figure 4. Average Tester Coordinate biplot.

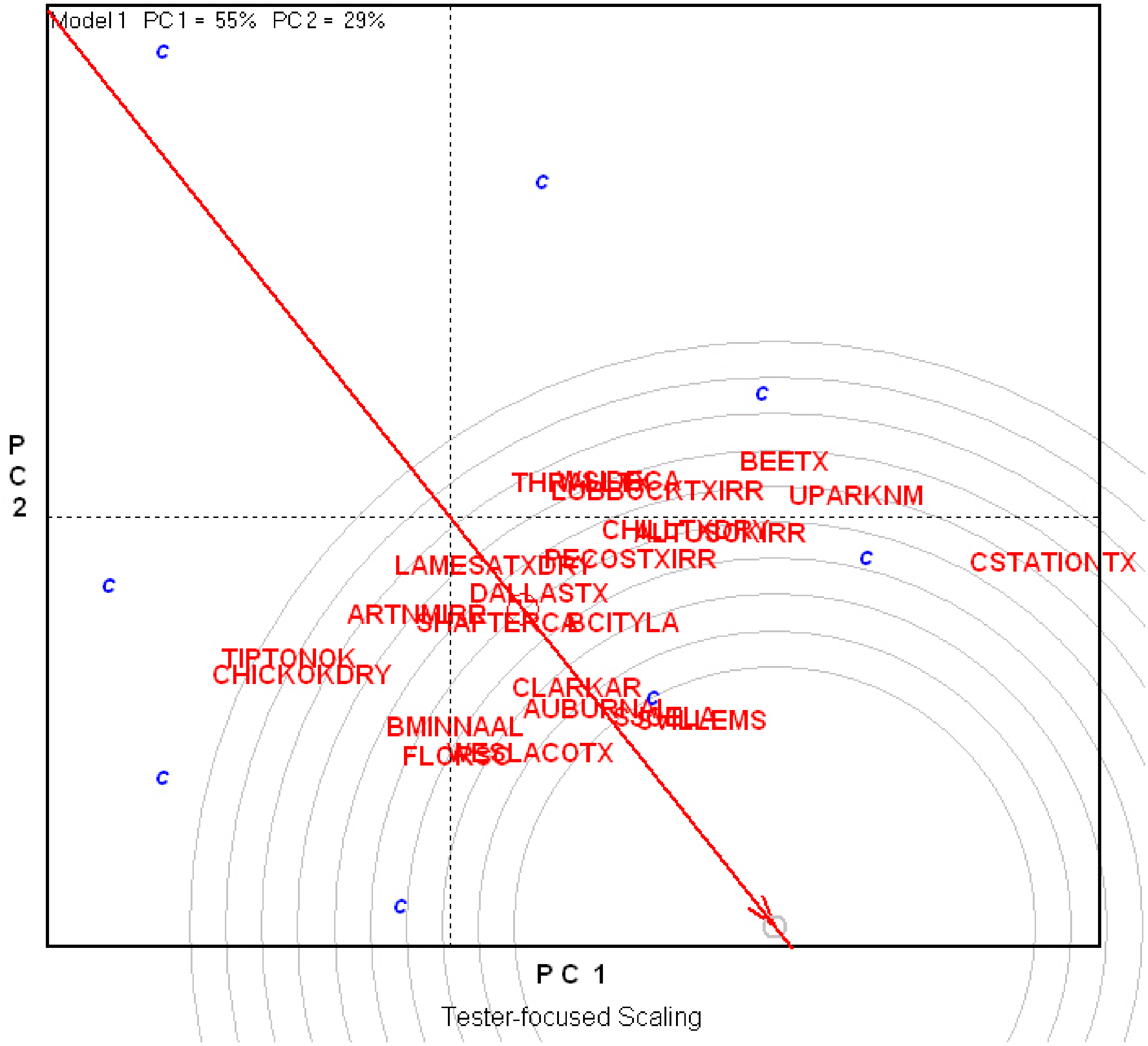


Figure 5. Environment ranking based on discriminating ability and representativeness 1997-2000 NCVT.

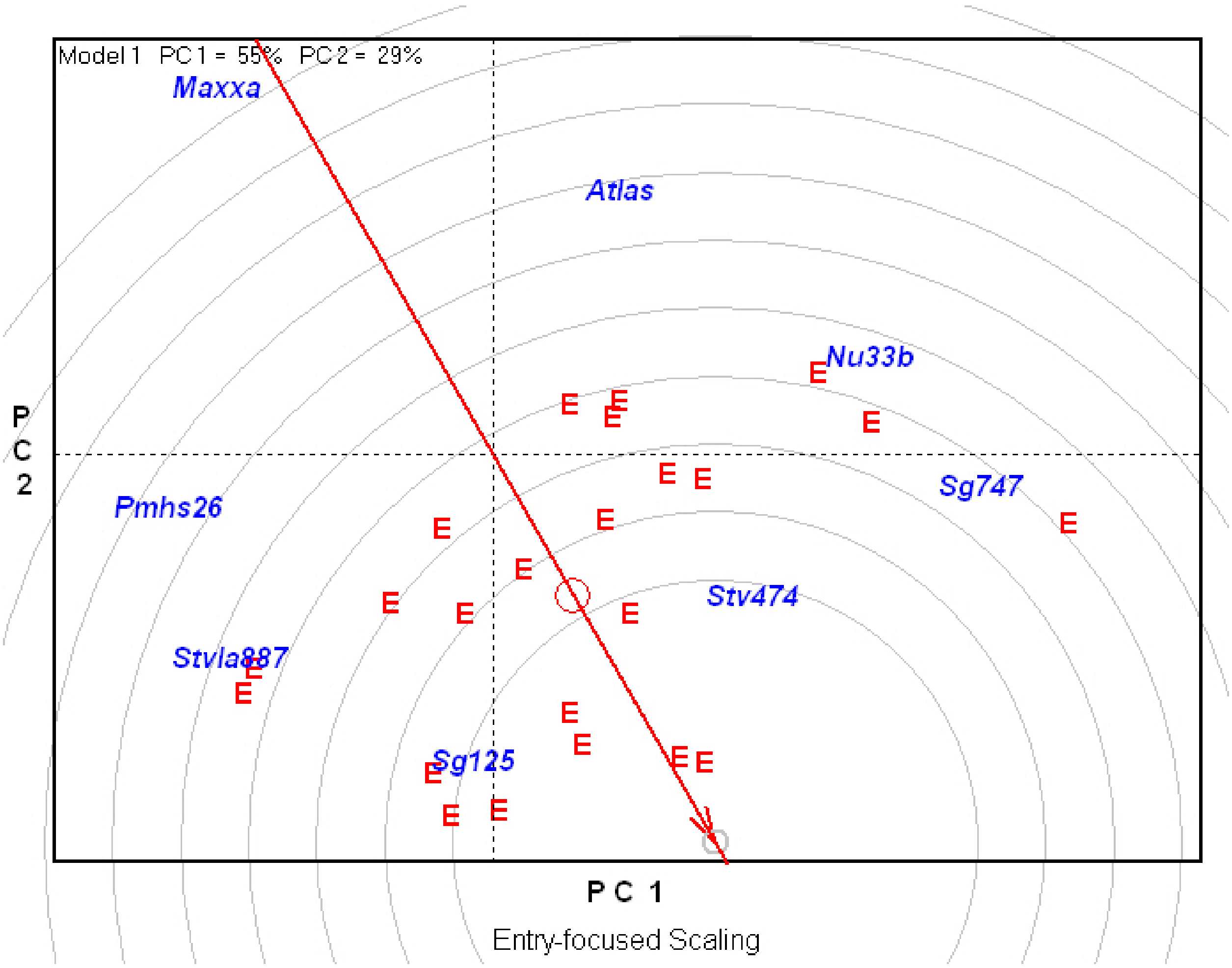


Figure 6. Variety ranking based on both average yield and stability.