INVESTIGATING TWO STATISTICAL TECHNIQUES USED IN THE ANALYSIS OF SUGARCANE VARIETY TRIALS

K A REDSHAW, N GOVENDER and M SMIT

South African Sugar Association Experiment Station, P/Bag X02, Mount Edgecombe, 4300, KwaZulu-Natal, South Africa
E-mail: xpbrko@sugar.org.za

Abstract

The main objective of the South African Sugar Association Experiment Station (SASEX) Agronomy variety trials is to compare relative yields amongst sugarcane varieties for making variety recommendations for different agroclimatic zones. Analyses of variety trials, and the interpretation of results, are however often complicated when the yield response to the genotype-by-environment (GxE) interaction is statistically significant. The Residual Maximum Likelihood (REML) and the Additive Main effects and Multiplicative Interaction (AMMI) methods were used to analyse data with GxE interaction in order to:

1) Predict mean estimated recoverable crystal (ERC) yields across different environments.
2) Evaluate the usefulness of AMMI and REML in identifying mega-environments.

Keywords: variety, sugarcane, variety trials, GxE interaction, AMMI, REML

Methods

Data

Six varieties (NCo376, N12, N16, N21, N27 and N29) in 56 dryland environments within KwaZulu-Natal, South Africa were used for this study. Each environment represented a single location-by-year combination with harvesting dates ranging from 1995 to 2001. The unbalanced data set consisted of ERC yields of between four and six replications for each environment. Not all varieties were present at all environments and these gaps were treated as missing data.

AMMI

The AMMI method was chosen because of its acclaimed ability to estimate yields more accurately than the input data, to partition GxE interaction into principle components that have meaning, to enable the user to select winning varieties more accurately and to graph mega-environments in terms of variety performance (Gauch 1992).

AMMI combines ANOVA and Principle Component Analysis (PCA) into a single analysis with both additive and multiplicative parameters. AMMI partitions the residual from the ANOVA into several interaction principle component axes (IPCA) using PCA, resulting in the AMMI model equation (Gauch, 1992):

$$Y_{ger} = \mu + \alpha_g + \beta_e + \Sigma_n \lambda_n \gamma_{gn} \delta_{en} + \rho_{ge} + \varepsilon_{ger} \quad (1)$$

Where:

Y_{ger} = yield of the genotype g in environment e for replicate r
μ = additive parameter, grand mean
α_g = additive parameter, genotype deviation from the grand mean
β_e = additive parameter, environment deviation from the grand mean
λ_n = multiplicative parameter, singular value for PCA axis n, units of yield
n = number of PCA axes retained in the model

γ_{gn} = multiplicative parameter, genotype eigenvector for PCA axis n
δ_{en} = multiplicative parameter, environment eigenvector for PCA axis n
ρ_{ge} = non-additive residual remaining if not all PCA axes are used
ε_{ger} = error term

AMMI analysis was run through MATMODEL (Gauch, 1990). AMMI imputed yields for missing data using the expectation-maximization (EM) algorithm (Gauch and Zobel, 1990). AMMI output parameters were presented in a biplot that provided a graphical representation of the interaction effects.

Further analysis assigned winning varieties to broader groups of environments, referred to as mega-environments (Gauch, 1996). These mega-environments include a group of environments that have similar crop responses in terms of the highest yielding variety at each environment (Gauch, 1992).

REML

REML is a generalised linear mixed model that uses genotypes and environments as fixed effects and replications as random effects (GenStat 5, 2000). The REML model groups the fixed and random terms, by using matrix and vector notation:

$$Y_{ger} = X\alpha + \Sigma^{c}_{i=1} Z_i \beta_i + \varepsilon \qquad (2)$$

Where:
Y_{ger} = vector of data (length n), yield of the genotype g in environment e for replicate r
α = vector of fixed effects (length p)with nxp design matrix X
c = number of random model terms
β_i = vector of random effects (length q) with nxq design matrix Z
ε = vector of random error term

The REML directive in GenStat® estimated the variance components using the Average Information (AI) algorithm (Gilmour *et al*., 1995) to maximise the residual likelihood. These were then used to construct an estimate of variance-covariance matrix. The fixed effects were estimated by generalised least squares and the random effects were predicted using best linear unbiased predictors (BLUPs).

The MVINCLUDE option in the REML directive was used to allow for the inclusion of data with missing values. This allowed the final REML output to be compared with that of AMMI.

Order statistics was used to compare the rankings of varieties by AMMI and REML with the actual mean ERC yields. Kendall's coefficient of concordance was used to measure the association between rankings of three methods of predicting means (actual mean, AMMI and REML). Once the degree of concordance was found to be significant, Spearman's rank correlation coefficient (Sokal and Rohlf, 1981) was used to determine the pairwise degree of association between the rankings of these different methods.

Results and Discussion

Both REML and AMMI methods of analysis indicated that yield variation due to genotype (G), environment (E) and GxE interaction were all statistically significant. The two Interaction Principal Component Axes (IPCA) for the AMMI model were also statistically significant.

Ranking of varieties
AMMI and REML picked the same winning variety as the actual means in 61% and 91% of the environments, respectively. AMMI picked the same winner as REML in 59% of the environments. AMMI output indicated that a yield benefit could be achieved were the AMMI predictions to be used rather than the actual means. The yield benefit varied from 0 tERC/ha to 1.57 tERC/ha, depending on the environment, with an average yield gain of 0.22 tERC/ha across all environments. As an example, the AMMI output for Ottawa harvested in 2000 indicated that if variety N27 was selected rather than N16, which was the variety that would have been selected were the actual means used, then there would be a yield benefit of 1.01 tERC/ha. This could be of great economic benefit to a grower at this specific location.

Kendall's coefficient of concordance was 0.733, indicating that the degree of concordance between actual means, AMMI and REML predicted means was significant. The Spearman's rank correlation coefficient between actual and AMMI ranking was 0.513, between actual and REML was 0.813 and between AMMI and REML was 0.538.

REML predicted the same ranking of varieties as that of the actual means, except for those environments where there were missing values. This explained the high correlation coefficient for rankings obtained by REML and the actual means. This indicated that there would be no benefit in using REML for this particular data set if varieties were selected based on their rank order.

It must be cautioned that the differences in imputed values for the two methods may affect the reliability of the final output of AMMI and REML.

Identification of mega-environments

The AMMI2 biplot identified four mega-environments in terms of variety superiority (Figure 1).

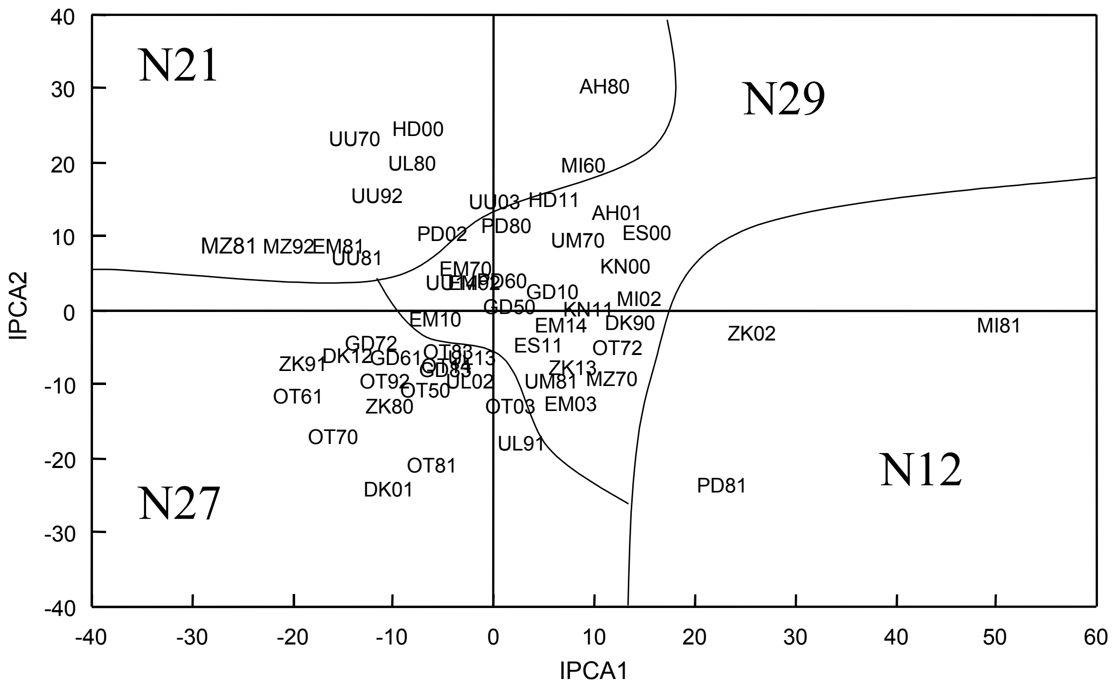

Figure 1. Defining mega-environments with an AMMI2 biplot. The environment code presented in the figure can be explained as follows: The first two letters indicate the location of the trial, eg. OT = Ottawa (see Table 1), the third letter represents the year of harvest (0 = 2000, 9 = 1999, etc) and the last letter indicates the ratoon (0 = plant crop, 1 = first ratoon, etc).

These mega-environments represent homogeneous areas based on the winning variety's own unique way of integrating environmental factors (Gauch, 1996). Each environment was assigned to one of the four mega-environments based on its winning variety. Future research will now attempt to identify the principle environmental characteristics of these mega-environments. The combination of the AMMI2 mapping and environmental characteristics may then be used to identify suitable locations and reduce duplication of homogeneous locations. In this study there were only two locations that had the same winning variety across all years of harvest. Kearsney (KN00 and KN11) and Eshowe (ES00 and ES11) were identified as homogeneous locations in terms of variety response.

While REML could not be used to group winning varieties into mega-environments, it could be used to predict winning varieties at each location (this data included all the years of harvest at each site) (Table 1).

Table 1. REML predicted mean ERC yields (t/ha) over years for 17 different locations.

Location	Code	Lat (ºS)	Long (ºE)	Alt (m)	MAP (mm)	N12	N16	N21	N27	N29	NCo 376
Ahrens	AH	29.05	30.75	1236	721	8.50	8.51	8.10	5.94	6.69	7.14
Doornkop	DK	29.22	31.23	442	1056	11.57	10.90	11.39	12.89	13.30	11.42
Empangeni – Paddison*	EM	28.80	31.92	74	1334	12.57	12.17	13.06	13.26	13.85	12.70
Empangeni – SASEX*	EM	28.80	31.92	74	1334	3.23	3.58	3.77	4.28	3.57	2.97
Eshowe	ES	28.85	31.45	536	1160	13.00	11.12	11.89	11.78	13.52	11.38
Gingindlovu – SASEX#	GD	29.02	31.57	24	1109	4.26	5.16	4.78	4.73	6.76	6.20
Gingindlovu – Etalmage#	GD	29.08	31.58	45	1125	9.40	9.70	9.31	11.10	10.44	8.70
Hibberdene	HD	30.57	30.48	90	917	11.74	13.00	13.96	13.57	14.66	13.48
Kearsney – SASEX	KN	29.28	31.27	277	1014	8.20	8.16	7.61	7.72	8.98	8.39
Mid Illovo	MI	29.92	30.52	732	851	16.18	15.45	15.14	14.34	16.34	16.63
Mtunzini – SASEX	MZ	28.93	31.73	36	1257	8.39	9.62	10.69	10.68	9.40	9.80
Ottawa	OT	29.67	31.03	75	927	7.90	8.85	7.33	9.41	8.65	7.32
Paddock	PD	30.75	30.25	503	856	13.76	13.07	13.41	14.07	15.33	14.65
ULOA – Monzi	UL	28.45	32.30	15	1079	13.08	13.43	13.16	13.74	13.98	13.35
Umdhloti	UM	29.65	31.13	100	1040	6.50	8.06	5.82	6.60	8.57	7.01
Upper Umhlali	UU	29.47	31.20	50	991	11.73	13.74	13.76	12.56	13.48	11.65
Zinkwazi	ZK	29.27	31.38	127	1005	9.87	9.91	9.08	10.83	10.31	9.59
Standard Error of Difference (5%) = 1.360											

Where: * and # = these locations were very similar in terms of soil type and environment and were allocated the same area code

These results showed some similarity with the AMMI output, e.g. AMMI picked N29 as the winner in 5 out of 6 years at Empangeni and REML picked N29 as the overall winner at this location, AMMI picked N27 as the winner in 8 out of 9 years at Ottawa and REML picked the same overall winner. REML could therefore be used for making more location-specific recommendations.

Conclusions

It is the view of the authors that this initial investigation of AMMI and REML provided useful and thought-provoking information that warrant their future use as statistical tools for the analysis of data from sugarcane variety trials across a range of environments.

An advantage of AMMI was its use of Principle Component Analysis to interpret the GxE interaction and to identify mega-environments. AMMI output allowed for results to be presented graphically and detailed explanations of some of the analyses were included in the final output. This assisted with the interpretation of the results. AMMI can be used to identify specific and broadly adapted varieties across a range of environments. The identification of homogeneous locations can assist with the positioning of new trials as well as avoiding duplication of similar locations. REML was useful for predicting yields for varieties at specific locations over a number of years.

AMMI, which was run through the MATMODEL software package, required data to be set up in a very specific, time-consuming way. REML may be processed through GenStat®, which is a menu-driven Windows-based system. An AMMI directive for balanced data has been added to the GenStat® library ([1]Smith, personal communication) and this will be investigated further.

Finally, validation of AMMI and REML methods still needs to be conducted.

Acknowledgements

The authors would like to thank Abraham Singels, Head of Agronomy Department, SASEX, for his invaluable comments on this paper. The contribution of the SASEX Agronomy technicians in the collection of the data from all the field trials is also gratefully acknowledged.

REFERENCES

Gauch, HG (1990). *MATMODEL Version 2.0: AMMI and related analyses for two-way data matrices.* Microcomputer Power, Ithaca, New York.

Gauch, HG (1992). *Statistical analysis of regional yield trials: AMMI analysis of factorial designs.* Elsevier, Amsterdam, London, New York, Tokyo.

Gauch, HG (1996). *AMMIWINS*. Microcomputer Power, Ithaca, New York.

Gauch, HG and Zobel, RW (1990). Imputing missing yield trial data. *Theoretical and Applied Genetics* 79: 753-761.

GenStat 5 (2000). *The guide to GenStat® Release 4.2, Part 2: Statistics*. Oxford Science Publications, pp 413-503.

Gilmour, AR, Thompson, R and Cullis, BR (1995). Average information REML, an efficient algorithm for variance parameter estimation in linear mixed models. *Biometrics* 51: 1440-1450.

Sokal, RR and Rohlf, FJ (1981). *Biometry. The principles and practice of statistics in biological research.* WH Freeman and Company, San Francisco, pp 601-616.

[1] Smith, M (2002), Agricultural Research Centre, Biomety Department, Pretoria