CSIRO PUBLISHING

Australian Journal of Agricultural Research

Volume 51, 2000

A journal for the publication of original contributions towards the understanding of an agricultural system

www.publish.csiro.au/journals/ajar

All enquiries and manuscripts should be directed to
Australian Journal of Agricultural Research
CSIRO PUBLISHING
PO Box 1139 (150 Oxford St)
Collingwood Telephone: 61 3 9662 7628
Vic. 3066 Facsimile: 61 3 9662 7611
Australia Email: jenny.fegent@publish.csiro.au

Published by **CSIRO** PUBLISHING
for CSIRO Australia and
the Australian Academy of Science

Aust. J. Agric. Res., 2000, **51**, 197–207

Genotype by environment interactions affecting grain sorghum. I. Characteristics that confound interpretation of hybrid yield

S. C. Chapman[AE], *M. Cooper*[B], *D. G. Butler*[C], *and R. G. Henzell*[D]

[A] CSIRO Tropical Agriculture, Long Pocket Laboratories, 120 Meiers Rd, Indooroopilly, Qld 4068, Australia.
[B] School of Land and Food, The University of Queensland, Brisbane, Qld 4072, Australia.
[C] Queensland Department of Primary Industries, PO Box 102, Toowoomba, Qld, 4350, Australia.
[D] Queensland Department of Primary Industries, Hermitage Research Station, Warwick, Qld 4370, Australia.
[E] Corresponding author; email: scott.chapman@tag.csiro.au

Abstract. Past sorghum hybrid trials in north-eastern Australia have detected substantial genotype by environment (G×E) interactions for yield in sampling a variable target population of environments (TPE) that is affected by spatial and seasonal differences in crop water supply. Three datasets, comprising yields of commercial and final stage experimental hybrids and covering 9–17 years (Y) and up to 30 locations (L), were analysed to quantify variance components for trial error, genotypic (σ^2_g), and G×E (σ^2_{gl}, σ^2_{gy}, and σ^2_{gly}) interaction effects.

Whereas trial means varied 2–3-fold across seasons, a greater range was estimated for variance components of trial error (range of 0.05–0.5), G (0–>0.3), and G×L interaction (0.05–>1.0). There was substantial seasonal variation in the ratio of σ^2_g to (σ^2_g+σ^2_{gl}), and in two datasets, 73% of the seasonal σ^2_{gl} was due to poor genetic correlations among locations. This implies that any given set of hybrids in a random set of locations would be ranked differently from season to season. Analysis of locations over years detected 90% of the total G×E interaction as G×L×Y, rather than G×L or G×Y, although this was reduced by accounting for genotype maturity. To achieve repeatabilities of >80%, trials would need to be conducted over at least 5 years and 20 locations per year.

The variable and unpredictable nature of much of the G×E interaction in the region implies that broad adaptation to different water regimes is required, unless prior knowledge of the seasonal weather can be used to choose 'narrowly adapted' cultivars. With current approaches, a large sample of environments is needed to identify such hybrids, and testing across locations and years is equally important. Alternative breeding strategies based on classifying environment types are discussed.

Additional keywords: drought, *Sorghum bicolor*, cultivar, tropical, breeding.

Introduction

During the seasons ending in 1995–1997, Australia produced an average of about 1.3 Mt of rain-fed sorghum at an average yield of just under 2.0 t/ha (FAO 1997). The area planted averaged 650 000 ha, although the majority of the growing region is spread unevenly in a 200–500-km-wide strip extending for over 1000 km from central Queensland (approx. 22°S latitude) to northern New South Wales (32°S). Together with a high season-to-season variability of rainfall (250–800 mm), the extent of the region and the variety of soils generate a great range of patterns of within-season water supply, often causing periods of severe water deficit at almost any stage of crop development. The crop is subject to other stresses such as sorghum midge (*Stenodiplosis sorghicolor* Coquillett) and may occasionally suffer poor nutrition given that nutrients are supplied with the expectation that seasonal rainfall and crop demand for nutrients will be less than optimal.

The above factors and other biotic stresses comprise a complex 'target population of environments' (TPE) (Comstock 1977) and complicate breeding programs by creating a substantial genotype × environment interaction (G×E) in any series of variety performance tests. It is difficult to ensure that a multi-environment trial (MET) to test new and current hybrids across a small number of locations and years will adequately sample the existing production TPE. Poor economies of scale compared with, for example, the USA exacerbate the problem. Australia's total sorghum production is only about 8% of that of the USA, yet it is distributed over a similar geographical area. However, the smaller market for hybrid seed cannot support the same geographical density of trials as used in the USA.

Sorghum breeding in Australia is characterised by close relationships between the public and private sectors (Henzell and Hare 1996). The Queensland Department of Primary

10.1071AR99020 0004-9409/00/020207

Industries (QDPI) program has concentrated mainly on the production of midge-resistant parental lines that are licensed to private industry for use directly or as source germplasm. Seed companies also develop parental lines and conduct broad-scale testing of current and experimental hybrids in both research and strip trials, mostly in farmer's fields. Beginning in the 1970s, both QDPI and the NSW Department of Agriculture conducted field variety trials, although the QDPI public trials were terminated in 1992.

Genotype by environment interactions and their analysis

The existence of large G×E interactions is most troublesome for selection when it results in a change in the ranking of genotypes across environments (Haldane 1947; Allard and Bradshaw 1964). The definition of a superior genotype then becomes conditional upon the environment in which the genotypes are tested. Selecting specifically adapted hybrids for each environment (niche breeding) is one response to this problem and has been employed to select barley in water-limited environments (Ceccarelli and Grando 1996). However, this is difficult where the niche environments are not fixed to locations, and seasons are not predictable. Given limited resources, breeding programs in Australia generally aim to produce broadly adapted hybrids, although seed companies typically supply a suite of hybrids differing in maturity and end-use purpose.

Theoretically, broadly adapted hybrids can be selected via large-scale repeated testing ('random sampling') across the production TPE. However, part of the G×E interaction may be 'repeatable' and able to be either controlled or sampled in a more stratified manner. This approach is commonly used when dealing with pests or pathogens. For example, midge-resistance in Australian sorghum has been greatly increased by selection of parent lines in managed nurseries exposed to high levels of midge (Henzell 1992; Henzell and Hare 1996). Allard and Bradshaw (1964) suggested that if G×E interactions are due to differences in soil types, and therefore associated with locations, they should be considered repeatable, cf. the effects of weather, which they considered were unrepeatable. Knowledge of the production factors (e.g. water, midge, soil fertility) that are responsible for repeatable G×E interactions can lead to the definition of target environments and/or management of specific selection nurseries which enable selection for target environments, e.g. irrigated and rain-fed treatments to select for drought tolerance in maize (Fischer *et al*. 1989; Bolaños and Edmeades 1993*a*, 1993*b*; Chapman *et al*. 1996) and wheat (Cooper *et al*. 1995). To sample the TPE with as few trials as possible, our objective should be to ensure that the frequencies of occurrence of different 'types' of environments match those being experienced in the TPE.

The magnitude of variance associated with G×E interactions can be estimated and the form of the interactions evaluated using METs that are routinely conducted as part of plant-breeding programs. An understanding of the relationships between several of the statistical methods (Cooper and DeLacy 1994; Cooper *et al*. 1996; DeLacy *et al*. 1996) has enabled the complementary application of these methods and thus a comprehensive analysis of the magnitude and form of G×E interactions faced by breeding programs, e.g. for sugarcane in Queensland (Mirzawan *et al*. 1993). For this approach, 3 methods are used in combination to examine genotypic variation and G×E in METs: (1) analysis of variance, (2) selection theory, and (3) pattern analysis. The analysis of variance is used to estimate relative sizes and, therefore, importance of variance components for genotypic, environmental, G×E interaction, and error sources of variance. Selection theory can determine the impact of the G×E interactions on selection strategies that are currently used or are being considered for use by breeding programs. Pattern analysis methodology is used to investigate the patterns of performance of genotypes across environments and the relationships among environments in terms of how they influence the relative performance of genotypes. All 3 statistical methodologies used together can identify whether any repeatable G×E interactions are expressed within the TPE and assist in the understanding of their causes.

In this paper, we examine 3 datasets (METs) to determine the magnitude of G and G×E interaction variance components for yield. Our objective was to determine how the relative sizes of genotypic and G×E variance components differ across seasons and locations and to what extent observed G×E interactions were repeatable. Although methodologies now exist to analyse such datasets across all years at once (e.g. Smith *et al*. 1998), we emphasise here the time sequence of variety testing as experienced in breeding programs and its effect on progress. In a further paper, we investigate one of the METs to determine whether different locations produce similar patterns of G×E interaction over time and relate this to the frequency of seasonal patterns of drought as described by crop simulation models (Chapman *et al*. 2000*a*). Finally, we use the simulation model approach to examine temporal trends in the simulated TPE and extend the spatial application to generate a TPE for the entire sorghum-growing region (Chapman *et al*. 2000*b*).

Materials and methods

Datasets and site characteristics

Three datasets were used for the analysis of G×E interactions for yield. All comprised 'Stage 4' (advanced) trials with the entries being current industry hybrids or experimental hybrids that were being considered for release. Two sets of trials were from active breeding programs: the QDPI breeding program and Pacific Seeds breeding program (Table 1). NSW Agriculture provided a further small set of variety testing trials. The data sets differed in dimension, with the QDPI METs testing a larger number of hybrids but in a smaller sample of locations relative to the Pacific dataset, which also included sites in northern NSW. The NSW dataset had small numbers of hybrids and was confined to NSW. Most of the locations in all datasets were on-farm and were planted between September and March, although up to 2 or 3 sites per year were

on research stations managed by QDPI or Pacific. On average, greater experimental replication was present in the QDPI and NSW METs, whereas 69% of Pacific dataset trials were unreplicated strips. All replicated trials were randomised complete block designs. An additional characteristic of the QDPI data was that up to 3 check hybrids were grown in every trial: Pride (early maturing, average anthesis at 61 days after sowing), Texas RS610 (medium maturing, 63 days and 'drought susceptible'), and E57 (late maturing, 68 days and 'drought tolerant'). Throughout this paper, single years in tables or figures (e.g. 87 or 1987) refer to the summer season of that year when the crop would normally be harvested, i.e. 1986–87.

Pacific Seeds defines a location as one of 32 'segments' (aggregated geographical areas), sometimes described by the nearest town. The QDPI and NSW datasets identify locations within a season uniquely by the nearest town and are therefore 'finer in resolution' in the description of a location. Until 1979 in the QDPI dataset, there were often several trials (divided on the basis of maturity) planted on the same date at a single location. This explains the disparity between the total number of trials and the number of unique combinations of location and year in QDPI datasets (Table 1). For the Pacific dataset, the disparity is due to multiple farm trials being assigned the same 'location' name within a season. As for most datasets of this type, the location is a 'loose' spatial reference, which in different seasons may actually be different farms and/or soil types subject to different management regimes.

Initially, the QDPI dataset included 35 different locations. Trials that were known to have had substantial irrigation, or to have suffered notable pest (especially midge) damage, were removed prior to analysis. For the variance component analysis, data were retained only from the 18 locations that remained after the data had been subjected to sequential pattern analysis (Chapman *et al*. 2000*a*).

Data processing and estimation of trial error variance

Hybrids that were tested in fewer than 2 years were dropped from the analysis. The data used were hybrid mean yields (at 15% grain moisture) for each trial as plot data were not available. The total number of means used was 5568, 5082, and 744 for the QDPI, Pacific, and NSW datasets, respectively.

Trial residual variances on a plot basis (s_ε^2) were available as error mean squares or were computed from a coefficient of variation (CV) or standard error (s.e.) of a mean:

$$s_\varepsilon^2 = (\mathrm{CV}/100 \times \mu)^2 = \mathrm{s.e.}^2 \times r$$

where μ is the mean of a single trial and r is the number of replicates in the trial.

To accommodate variation in data quality and trial-to-trial variation in σ_ε^2, and to estimate σ_ε^2 for trials with no experimental estimates (e.g. strip trials), we decided to model the structure of the observed error variances, rather than to use the actual values. Cullis *et al*. (1996) used an error variance model for this purpose:

$$\sigma_i^2 = \exp(1 + \alpha_i' \delta)$$

where σ_i^2 is the expected error variance for trial i, δ is a vector of unknown parameters, and α_i' is a transposed vector of explanatory variables. Variables included in α were the log of the trial mean [log(μ)], location (l), and year (y). As the Pacific data had no σ_ε^2 for years 1984 and 1985, the log-linear model was fitted to these data without the year effect. Using S-Plus (MathSoft Inc., Seattle, WA), the model was fitted as a generalised linear model with gamma errors, a log link, and weights of $v_i/2$ where v_i are the residual degrees of freedom associated with σ_i^2 from each trial.

The weight for trial i (w_i) was calculated as:

$$w_i = r_i \bullet \bar{s}^2/\hat{s}_i^2$$

where r_i is the number of replicates in trial i, $\hat{s}_i^2$ is the fitted error variance (actual for NSW dataset) for trial i, and $\bar{s}^2$ is the average error variance across all trials.

Variance component analysis and selection theory

The hybrid means were processed by a residual maximum likelihood method (Patterson and Thompson 1975) using the ASREML software (Gilmour *et al*. 1995, 1998) to estimate the genotypic components of variance, assuming genotypes (g_i) in any season to be a random sample of the currently available hybrids. Locations (l_j) and years (y_k) were assumed to be fixed. The model for individual years was:

$$x_{ijl} = u + g_i + l_j + (gl)_{ij} + \varepsilon_{ijl}$$

and for environments over years was:

$$x_{ijkl} = u + g_i + l_j + y_k + (gl)_{ij} + (gy)_{ik} + (gly)_{ijk} + \varepsilon_{ijkl}$$

where x is the lth observation of genotype i in location j (and year k in the latter case), u is the expected mean over all genotypes and locations (and years), and the remaining terms are the relevant main effects and associated interactions. The data were unbalanced across years (and often within years) for the inclusion of genotypes and locations. Although genotypes were presumably improving with time, these trials all contained the same 3 checks of differing maturity and adaptation.

Within years, hybrid means were processed by ASREML using the individual-years model and the average error variance for the year, with each trial weighted by its number of replicates. The datasets were also processed using the trial weights (w_i) described above. These analyses provided estimates of the variance components for genotypes (σ_g^2) and for interactions between genotype and locations (σ_{gl}^2) within each year of testing. The multiple-years model was applied to entire datasets to estimate interactions between genotype and years (σ_{gy}^2) and between genotypes, locations, and years (σ_{gly}^2). Where variance components were less than zero, that component was assumed to be zero and was dropped from the model prior to determining the final estimates.

In each dataset, an estimate of phenotypic variance (σ_p^2) was calculated using the variance components from the weighted analyses for individual years:

$$\sigma_p^2 = \sigma_g^2 + \sigma_{gl}^2/l + \sigma_\varepsilon^2/(l.r)$$

or across years:

$$\sigma_p^2 = \sigma_g^2 + \sigma_{gl}^2/l + \sigma_{gy}^2/y + \sigma_{gly}^2/(l.y) + \sigma_\varepsilon^2/(l.y.r)$$

where l, y, and r were any given number of locations, years, and replicates, respectively. Several sets of values were used for l, y, and r to compare different testing strategies. For each of these combinations, repeatability (h^2), computed in the form of a heritability (broad sense) and defined and used as a measure of the chance of detecting the same differences among hybrids in future experiments (Fehr 1987, p. 97), was estimated as the ratio:

$$h^2 = \sigma_g^2/\sigma_p^2$$

Cooper and DeLacy (1994) gave a theoretical development of relationships between direct selection and indirect selection theory as applied to the measure of genotype performance in more than one environment. This includes an extended discussion of an expression given by Cockerham (1963) that partitioned the G×E interaction component of variance (σ_{ge}^2) into components due to heterogeneity of genotypic variance (V) and lack of genetic correlation among environments (L_c). The latter is of particular relevance, as this is the part of σ_{ge}^2 that can result in changes in the ranking of genotypes across environments and therefore complicate selection of superior hybrids. Cooper and DeLacy

(1994) review expressions from Cockerham (1963) and Dickerson (1962) to estimate the magnitudes of V and L:

$$V = \frac{\sum_{j<j'j'} (\sigma_{g(j)} - \sigma_{g(j')})^2}{e(e-1)}$$

$$L_c = \sigma^2_{ge} - V$$

for comparisons among environments j to j' and where e is the number of environments. Values of V and L_c were estimated for the QDPI and NSW datasets for each year and across years. These relationships can be developed further to estimate pooled genetic correlations and review the relative importance of h^2 when L_c is a significant part of the G×E interaction (Cooper *et al*. 1996). At present, this work is being extended to unbalanced datasets and is still under development (B. R. Cullis, pers. comm.).

In about 80% of the 144 QDPI trials, estimated anthesis dates were available for each hybrid. Using ASREML, we repeated the across-years analysis of yield for the QDPI data and applied anthesis date as a co-variate in the main effects and interactions (location and year). This provided estimates of genetic and genotype interaction variance components that had been adjusted for maturity effects.

Results

Mean trial yields across all years were similar for each dataset (Table 1), but the range of yields across and within years (Fig. 1) was larger in the QDPI and Pacific datasets than for NSW. Estimated average trial error variance ($\bar{s}^2$) was substantially lower in the QDPI and Pacific datasets than in the NSW data (Table 1). There was a linear relationship between log (trial mean yield) and log (s^2_ε) in all 3 datasets (Fig. 2*a*), although several highly variable trials from NSW lie between log mean yields of about 1 and 2. The boxplots in Fig. 2*b* and *c* illustrate that the distribution of variability in σ^2_ε among the 156 QDPI trials varied with both location and year. For all 3 datasets, the log (trial mean), and for the QDPI and NSW datasets, the location and year, had a significant ($P < 0.01$) influence on the magnitude of error variance (Table 2). Over each of the seasons of the QDPI trials, estimates of σ^2_g using the modelled error variances were about 15% greater than if observed error variances were used (data not presented), although the methods were highly correlated ($r = 0.98$).

For all 3 datasets, estimates of σ^2_g (Fig. 3*a*) and σ^2_{gl} (Fig. 3*b*) varied 3–5-fold across years. The average values of these components (shown in the figure legends) were greatest in the Pacific dataset, followed by those in the QDPI and then NSW datasets. In the QDPI dataset, σ^2_g was estimated to be zero in 1978, and this also occurred in both the Pacific and NSW datasets in 1985. For the 8 seasons that were in common across the 3 datasets (1984–91), the values of σ^2_g were significantly correlated between the QDPI and Pacific datasets ($r = 0.83$, $P < 0.05$), but not between the NSW dataset and other datasets. There was no correlation between the datasets for the estimates of σ^2_{gl} across years.

The ratio of σ^2_g to ($\sigma^2_g + \sigma^2_{gl}$) also varied greatly with season (Fig. 3*c*). On all except 3 occasions, σ^2_g was less than σ^2_{gl} (i.e. ratio <0.5). Large σ^2_{gl} particularly dominated the Pacific dataset in all years (average ratio of 0.08) except 1989.

Given a 'standard' test of 10 locations with experiments of 3 replicates in the years for which trials were actually conducted, the repeatability of a MET was generally greater for the QDPI and NSW datasets than for the Pacific dataset (Fig. 3*d*). For the years in which trials overlapped (1984–91), the average h^2 was 0.65, 0.35, and 0.50 for the QDPI, Pacific, and NSW datasets, respectively. Notably, in all except one year, the h^2 in the QDPI dataset was >0.5, whereas in the 1985 and 1986 seasons, the h^2 in the Pacific and NSW datasets was <0.1.

When analysed over all years in a dataset, σ^2_g averaged about 0.03 in the QDPI and NSW datasets, but was 3 times larger in the Pacific dataset (Table 3). In the Pacific and NSW datasets, the variance components σ^2_{gl} and σ^2_{gy} were estimated to be negligible or negative and these terms were removed and the analysis re-run. Standard errors of the variance component estimates were lowest in the QDPI dataset. The σ^2_{gly} variance component was 3.5–6 times the size of σ^2_g so that ratios of σ^2_g to the sum of σ^2_g and interaction components were <0.3 in all datasets.

When the QDPI data were adjusted for anthesis date, the σ^2_g increased by 40%, whereas both σ^2_{gl} and σ^2_{gly} were decreased by at least 25% (Table 3). Together with a 10% decrease in σ^2_{gl}, this increased by 50% the ratio of σ^2_g to all sources of genotypic variance.

Given the same standard test described previously, but using the variance components derived from the 'all years' analysis in Table 3, the h^2 estimates for a 2-year testing program were 0.58, 0.70, and 0.68 for the QDPI (without anthesis date adjustment), Pacific, and NSW datasets, respectively. For more extensive testing over 3 years the corresponding estimates were 0.66, 0.78, 0.76 or over 5 years were 0.75, 0.86, 0.84. For up to 3 replicates, 20 locations, and 5 years, the h^2 of the QDPI dataset was 82–92% of the average of the h^2 of the Pacific and NSW datasets.

Fig. 4*a* and *b* shows the σ^2_{gl} of the QDPI and NSW datasets, respectively, partitioned into interaction due to heterogeneity of variance (V), and that due to lack of genetic correlation among locations (L_c). Seasons 1984 and 1990 are missing from the NSW dataset where σ^2_{gl} was negligible. In both datasets, the G×L interaction due to L_c dominated the V effects such that, averaged across years, L_c represented 70 and 73% of σ^2_{gl} in the QDPI and NSW datasets, respectively; i.e. even within a single season, the ranking of genotypes across locations can change greatly. Fig. 5*a* and *b* illustrates this crossover effect for 2 consecutive years (1987–88) of QDPI data for 3 hybrids grown in the same 7 locations. Substantial seasonal effects on mean location yield are apparent (e.g. BW cf. GD), as are differences in hybrid rankings at the same location in different seasons, or at different locations within the same season. It is, therefore, difficult to identify hybrids that have broad adaptation to all of the environments.

Table 1. Summary of characteristics of sorghum hybrid trial datasets from QDPI, Pacific Seeds, and NSW Agriculture variety testing programs

Values in parentheses refer to total number of possible levels of a factor, and the sequences of 3 numbers are the minimum, mean, and maximum values for factors (trial, year, location, genotype, replicate) or trial attributes [grain yield, coefficient of variation (CV), or error variance (s^2_ε)]. For locations, the minimum, mean, and maximum values refer to values within years, whereas for the other columns, they refer to values within trials. Trials refer to the number of unique combinations of years and locations, whereas the number in parentheses is the total number of trials grown, including those instances where more than one trial was grown at a location within a year. In QDPI and Pacific datasets, CV and s^2_ε were only available for 156/168 and 163/568 of the trials, respectively

Dataset	Trials (total)	Years (total)	Locations (total)	Genotypes (total)	Reps	Grain yield (t/ha)	CV (%)	s^2_ε $(t/ha)^2$
QDPI	144 (168)	1974/5–1990/1 (17)	1, 6.8, 14 (18)	4, 33.1, 71 (222)	3, 3.6, 4	0.49, 4.27, 9.79	4.5, 12.4, 49.3	0.005, 0.229, 0.797
Pacific Seeds	209 (568)	1983/4–1994/5 (12)	8, 17.4, 22 (28)	4, 8.9, 25 (91)	1, 1.8, 4	0.29, 3.59, 10.12	3, 11.4, 23.1	0.005, 0.247, 2.411
NSW	38 (39)	1983/4–1991/2 (9)	3, 4.2, 7 (14)	6, 19.1, 32 (20)	3, 3, 3	1.79, 4.34, 7.60	7.4, 14.8, 35.8	0.072 0.400 1.616

Table 2. Analyses of deviance for fitted log-linear model of error variance in three sorghum variety yield datasets

All sources of variation are significant at $P = 0.01$

Dataset	Source	d.f.	Deviance	Residual d.f.	Residual deviance
QDPI				155	3228.4
	Log (trial mean)	1	1253.4	154	1974.9
	Location	17	270.9	137	1704.0
	Year	16	345.7	121	1358.3
Pacific				162	1211.2
	Log (trial mean)	1	551.1	161	660.1
	Location	18	147.5	143	512.6
NSW				38	452.9
	Log (trial mean)	1	44.1	37	408.8
	Location	13	156.7	24	252.1
	Year	8	154.6	16	97.5

Taken across seasons and datasets, the importance of L_c is clear (Fig. 6). For each unit of increase in σ^2_{gl}, >70% is that due to an increase in L_c, i.e. environments become less and less correlated. The V effect increases at less than half of the rate observed for L_c.

Discussion

Inter-annual variation in σ^2_g and σ^2_{gl} is a pervasive characteristic of dryland hybrid trials of sorghum in north-eastern Australia. Over time, it becomes apparent that most of the G×E interaction is associated with σ^2_{gly}, but within an active selection program, plant breeders must deal with G×L interaction effects as they arise, usually over a fixed number of seasons. Although mean yields may differ 2- or 3-fold across seasons (Fig. 1), there was greater variation in the estimated variance components: σ^2_ε (Fig. 2), σ^2_g (Fig. 3*a*), and σ^2_{gl} (Fig. 3*b*). Fitted trial error variances ranged almost 500-fold across trials and up to 10-fold among locations within seasons (Fig. 2). The average trial error variance was greater in the NSW dataset than the others that were conducted predominantly in Queensland. Sheppard *et al*. (1996) found similar differences between datasets collected by the 2 States for wheat. Although mean yields were similar across the sorghum datasets, there had been fewer low-yielding trials in NSW. As there was a decrease in trial error variance with mean trial yield (Fig. 2*a*), a larger sample of low-yielding trials from NSW may have decreased the estimate of average trial error variance, but perhaps not down to the level observed in the other datasets.

In each dataset, estimates for both σ^2_g and σ^2_{gl} varied greatly from season to season. Significant correlation existed between the estimates of σ^2_g from the QDPI and Pacific datasets, which together had many locations in common. Components derived from the smaller testing region of the NSW dataset were not correlated with either of the other 2 datasets, even though the Pacific trials covered both States. The ratio of σ^2_g to $(\sigma^2_g + \sigma^2_{gl})$ was almost always <0.5, and averaged about 0.22 across all years and datasets (Fig. 3*c*). This contributed to low repeatabilities (usually <0.6) of a 'standard' MET of 10 locations with 3 replicates and indicated that the G×L interaction is complex. Assuming that the locations chosen each year are random, the large seasonal variations in the ratio of σ^2_g to $(\sigma^2_g + \sigma^2_{gl})$ imply that the sample of the target population of environments differs year to year.

Sampling the TPE

The TPE encompasses the range of growing conditions that can occur across all locations and years. To select efficiently for broad adaptation, a plant breeder aims to adequately sample this population with as few locations and years as possible. If the trials across all datasets are taken to represent the TPE, then it is evident that the individual years of testing differ greatly as adequate samples of the environments. However, in the combined analysis over years, much of the genotype by environment interaction was partitioned

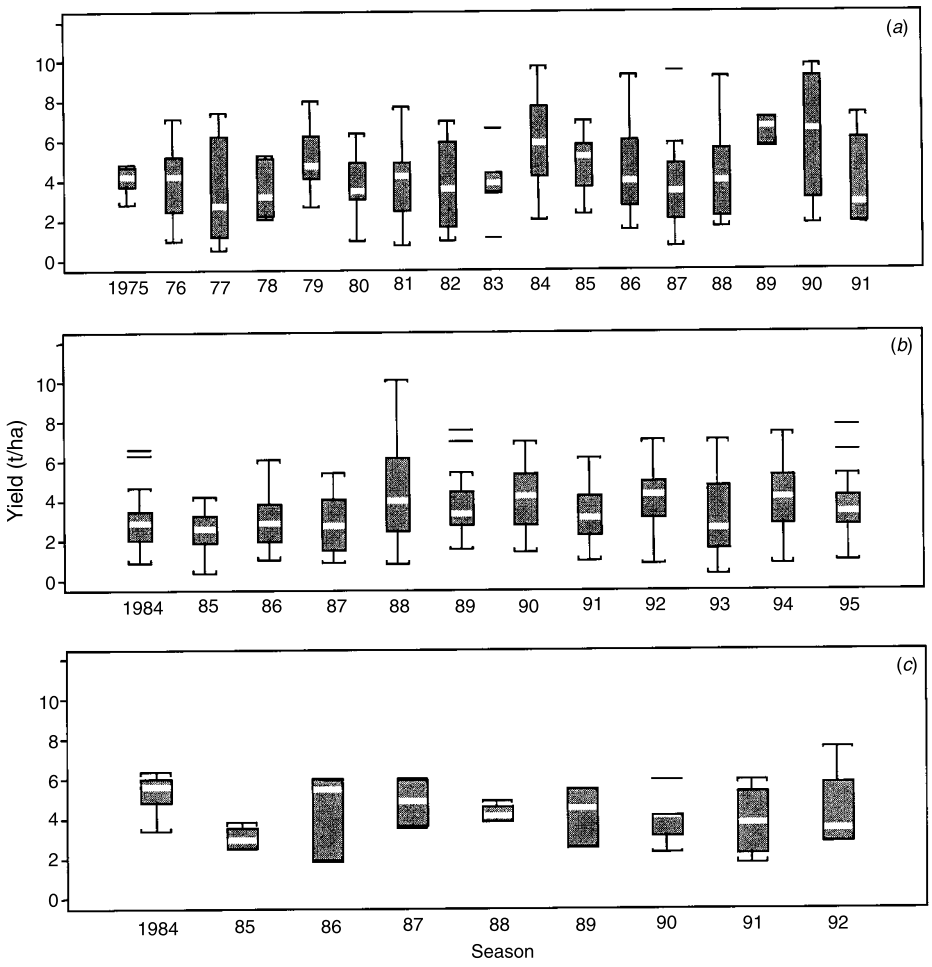

Fig. 1. Box plot of mean trial yields (t/ha) for each year of (*a*) QDPI, (*b*) Pacific Seeds, and (*c*) NSW Agriculture sorghum hybrid trial datasets. The centre line of the box plot is the median, whereas the lower and upper ends of the box are the first (1Q = 25%) and third (3Q = 75%) quartiles, respectively. The lower and upper ends of the vertical line are the minimum and maximum values, respectively, and not including outliers (dashes) that are >1.5 quartile ranges (3Q–1Q) below 1Q or above 3Q.

into the 3-way component (G×L×Y). This further suggests that the choice of locations or years to randomly sample such a large TPE is of little consequence. Although the 2-way interactions were significant in the QDPI dataset, their small size would lead them to both be considered as relatively 'unrepeatable' in the terminology of Allard and Bradshaw (1964). Despite these problems, if the trials could be classified as being members of different 'environment types' (regardless of location and year), then this might capture repeatability that exists in the interactions of genotypes with both locations and years. This concept is explored by Chapman *et al*. (2000*b*).

Ideally, given sufficient operating resources and the expectation of a normal distribution of environments from the TPE, a breeding program would be advised to test in a single year over a large number of geographically divergent locations. This would maximise throughput of germplasm for the program, but only if one could assume that, over all the locations, the testing season represented a balanced sample of the entire TPE. Unfortunately, in the variable climate of north-eastern Australia, this assumption would often be wrong as indicated by seasonal differences in estimates of σ^2_g and σ^2_{gl} (Fig. 3).

Due to limited resources within a single season and the risk of failure (e.g. a broad-scale drought), breeding programs typically employ several years of testing. Assuming 2 years with 3 replicates per trial and the variance components given in Table 3, at least 5 (Pacific) to 7 (QDPI data) locations would be required to obtain a repeatability of discrimination among hybrids of >0.5. Ten (Pacific) to 33 (QDPI) locations would be needed to surpass a repeatability of 0.7. Note, though, that each 'location' in the Pacific dataset had an average of 3 farm trials, so that the G×L component includes the effects of both 'genotype × location' and 'geno-

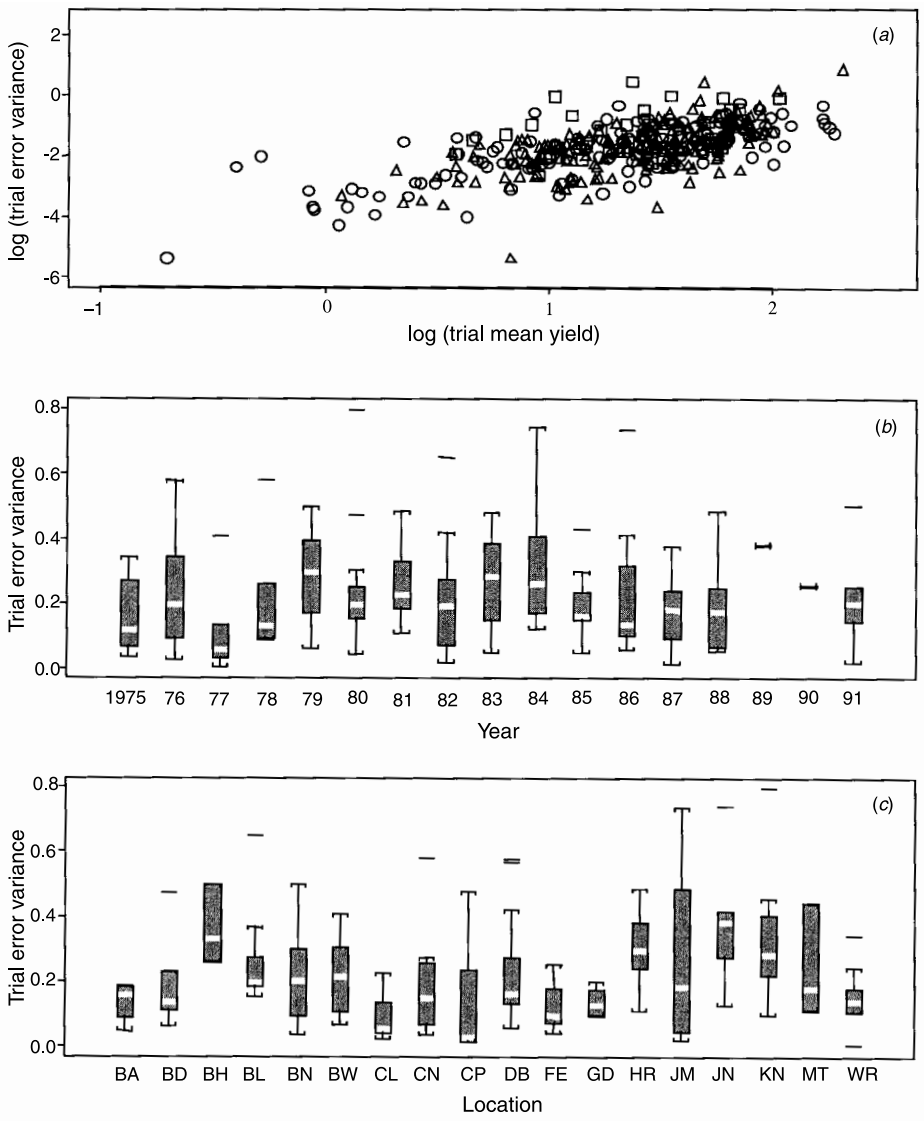

Fig. 2. (*a*) Trial error variance (s_ε^2) versus trial mean yield (log scaled) for 3 sorghum variety trial datasets (circles, QDPI; triangles, Pacific Seeds; squares, NSW Agriculture), and box plots (see Fig. 1 for description) of s_ε^2 for (*b*) years and (*c*) locations in the QDPI dataset.

type × farm within location'. Hence, in practice, more than 30 trials (over the 10 locations) would still be required to obtain a repeatability of 0.7 in the Pacific dataset, i.e. effectively both breeding programs would need about 30 trials.

There is a further qualifier to conclusions about sampling based on the analysis over locations and years. Fig. 3*d* shows that the repeatability would be particularly dependent upon which seasons were encountered during testing. For example, if the 2 seasons of testing were 1983 and 1984, repeatabilities of >0.8 would have been obtained in each year with 10 locations and 3 replicates. If the years of testing were 1981 and 1982, the repeatability of the results from season to season would have been poor. A valid question would be: are differences in repeatability due to differences in weather from season to season or due to changes in the set of locations sampled? Unfortunately, with these datasets we cannot easily determine whether either set of years is 'more representative' of the TPE, given the lack of balance in the locations sampled each year.

Improving selection for broad adaptation in variable environments given limited resources

The large component of G×L×Y would seem to indicate that, in choosing hybrids, both farmers and breeders have to 'take conditions as they come'. Narrowly adapted hybrids would only be useful where there is an ability to predict the upcoming season, perhaps assisted by knowledge of soil conditions at planting. There already are some relationships between seasonal weather predictors (the El Niño Southern Oscillation Index) and season type that may

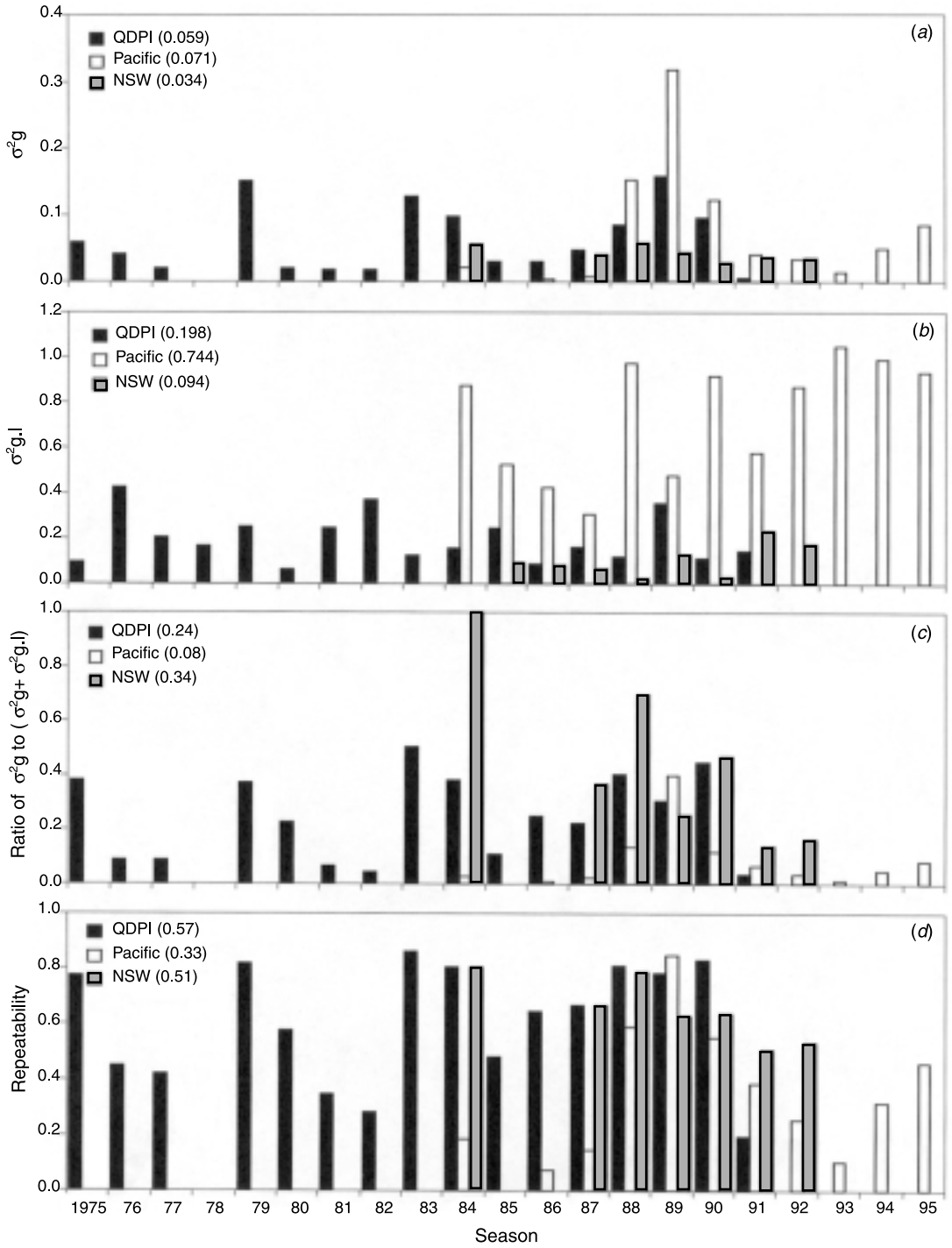

Fig. 3. For each year of sorghum variety trials conducted by QDPI, Pacific Seeds, or NSW Agriculture, REML estimates of (*a*) genotypic variance components (σ^2_g); (*b*) genotype by location variance components (σ^2_{gl}); (*c*) the ratio of σ^2_g to (σ^2_g + σ^2_{gl}); and *(d)* respeatability (h^2), as estimated using the given variance components and fitted error variances for a MET of 10 locations with 3 replicates per location.

allow this to become a reality in the future, e.g. Hammer *et al*. (1996). The obvious trait to be targeted for this approach would be maturity type. Removal of this effect in the variance analysis greatly increased the estimates of the σ^2_g compared with the interaction components (Table 3), although the $\sigma^2_{g\ ly}$ was still a little more than twice σ^2_g. However, until seasonal predictors are in frequent use and are reliable enough to modify choice of hybrids, broad

Table 3. Variance components with standard error in parentheses derived from the genotype (g), locations (l), and years (y) model applied to sorghum hybrid yield trials conducted by QDPI, Pacific Seeds, or NSW Agriculture

Variance components have been determined after modelling the trial error variance (see Table 2 and text). Also shown for each dataset is the ratio of genotypic variance to all sources of genotype variation (genotypic + interactions)

Dataset	σ^2_g	σ^2_{gl}	σ^2_{gy}	σ^2_{gly}	σ^2_ε	Ratio: $\sigma^2_g/(\sigma^2_g+\sigma^2_{gl}+\sigma^2_{gy}+\sigma^2_{gly})$
QDPI	0.033 (0.006)	0.020 (0.004)	0.020 (0.004)	0.180 (0.006)	0.229	0.149
QDPI[A]	0.047 (0.007)	0.015 (0.003)	0.018 (0.003)	0.139 (0.005)	0.229	0.214
Pacific	0.100 (0.024)	—[B]	—	0.782 (0.019)	0.247	0.128
NSW	0.030 (0.013)	—	0.008 (0.014)	0.097 (0.019)	0.400	0.283

[A] Analysis of the QDPI data using anthesis date (residual cf. each trial mean) as a co-variate in main and interaction effects for yield.

[B] Estimates were small, and effectively equal to 0.0.

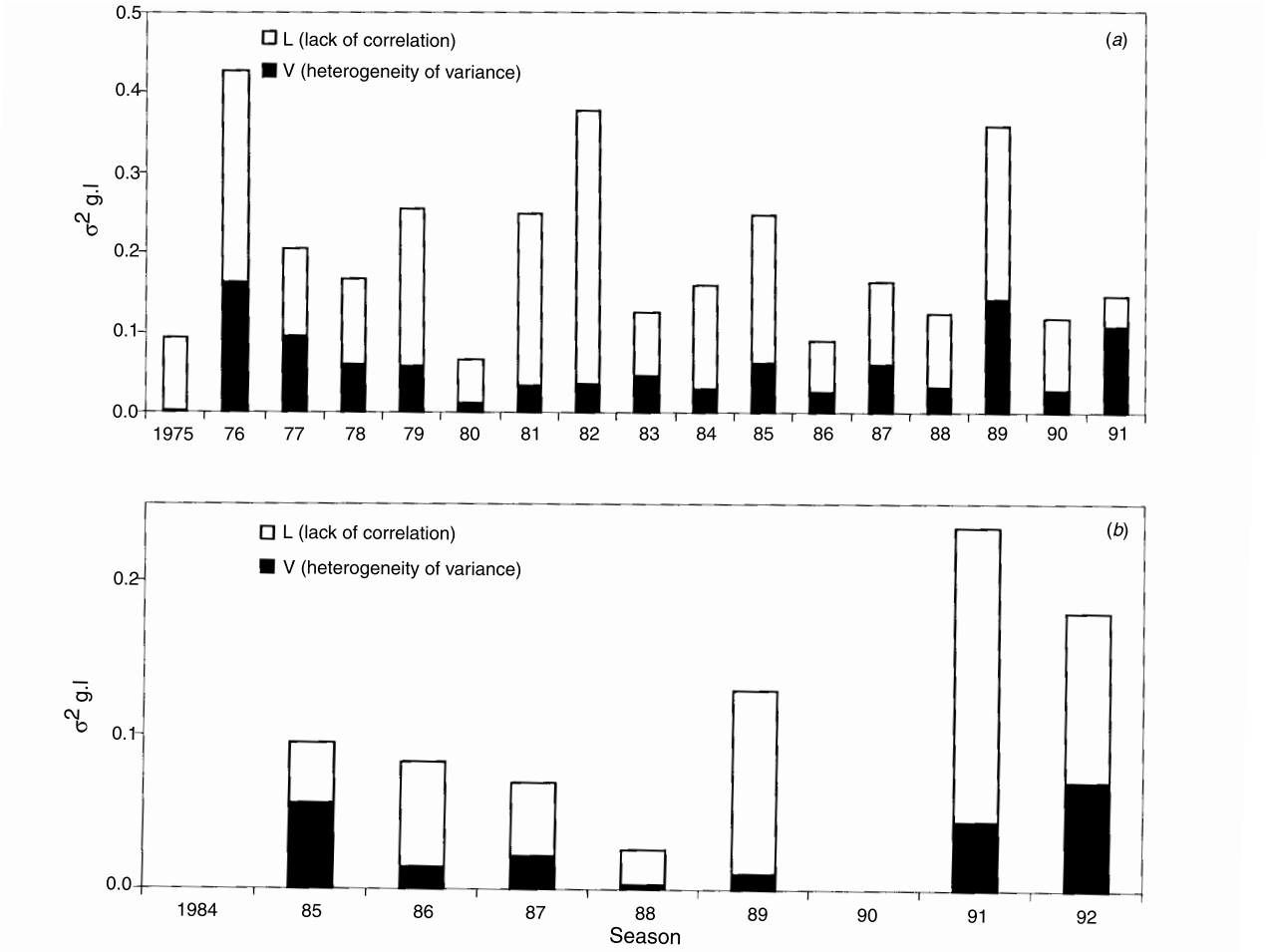

Fig 4. Genotype by location variance components (σ^2_{gl}) for each year divided into that part due to heterogeneity of variance and that due to lack of correlation among environments for (*a*) QDPI and (*b*) NSW Agriculture sorghum variety trials.

adaptation to drought conditions using any given maturity type will remain important.

Although selection for broad adaptation is difficult, there are opportunities to be more pro-active in the use of hybrid testing results. Cooper *et al*. (1996) have demonstrated theoretically how mismatches between the frequencies of environment types sampled in the MET and their true frequencies in the TPE can reduce genetic improvement for the TPE. The lack of genetic correlation among locations that exists from season to season could generate a 'yo-yo' effect (Rathjen 1994) whereby adaptation might advance and even recede in successive seasons, depending on how representative the

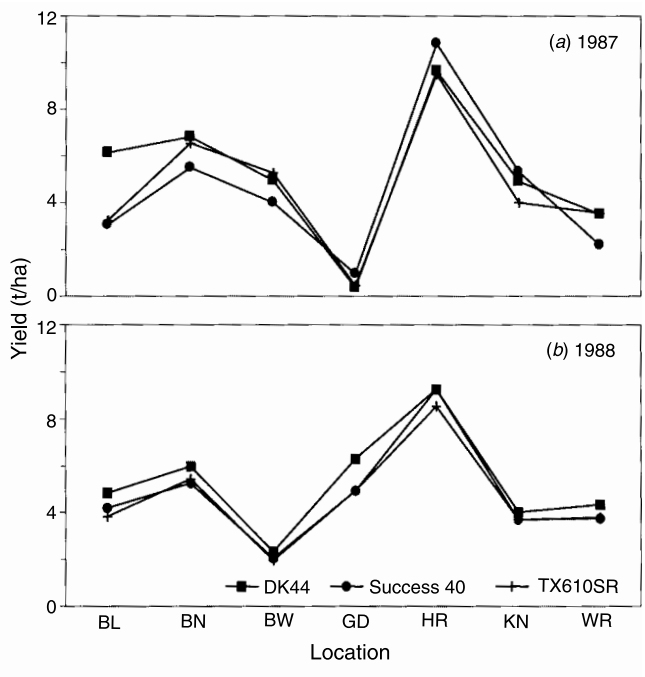

Fig 5. Yields (t/ha) of 3 genotypes grown in multi-environment trials in 7 locations in (*a*) 1987 and (*b*) 1988.

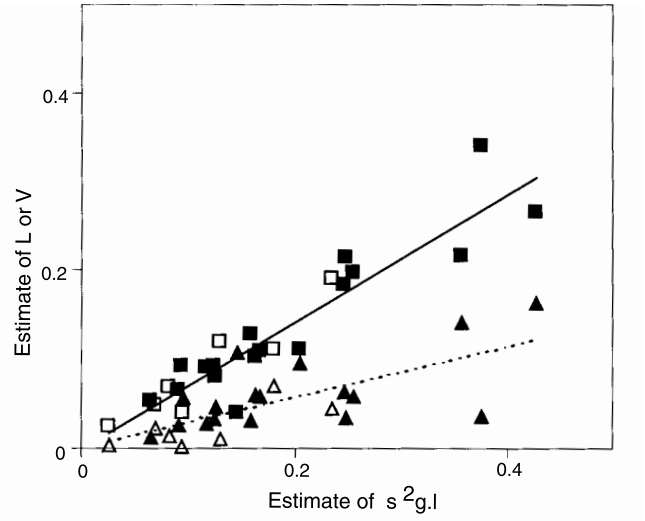

Fig. 6. Regressions of estimates of the variance components due to heterogeneity of variance (V; triangles) or lack of correlation among environments (L; squares) against their sum (the G×L interaction component) for each season of both the QDPI (closed symbols) and NSW datasets (open symbols). The regressions are $y = 0.28x$ ($r^2 = 0.50$) for V and $y = 0.72x$ ($r^2 = 0.87$) for L.

season is of the TPE. To control these effects, Podlich and Cooper (1998), using a simulation model of gene action, showed that selection gain over several generations of testing can be increased by weighting the results from the MET to account for the mismatch with the TPE.

Some breeding programs (e.g. Fischer *et al*. 1989) generate artificial frequencies of well-watered and droughted environments using management, but this requires locations where drought is assured and irrigation is available. The approach being suggested here is one of taking the data from variety trials and utilising the information more efficiently. 'Low stress' seasons do still occur at a reasonable frequency in this region (Chapman *et al*. 2000*b*) and are often the years in which farmers make the greatest profit. Therefore, to sample these seasons, breeders should still include at least one partially irrigated environment in any set of variety trials as a minimum. This is often done, but usually to determine potential yield or ensure against disaster, rather than to adequately sample the TPE. For the same reason, low-yielding sites due to poor seasonal conditions should not be discarded.

The successful selection of superior hybrids in this complex TPE is dependent not only on choosing 'representative locations' (Chapman *et al*. 2000*a*), but on recognising 'representative seasons' (or combinations thereof) and weighting information in ways which reflect the TPE, i.e. data should be considered more reliable in seasons when the locations generate differences that are representative of the longer term effects. In a pro-active method of testing, one might utilise data immediately if a set of locations in a season is suitably representative; otherwise, extra seasons of testing would be undertaken or environment weightings applied. The pro-active process of extra seasons of testing is only currently used when trials are wiped out by catastrophe, but should equally be applied in years when the data are largely unrepresentative of the TPE. The appropriateness and degree of weightings applied are being further investigated for sorghum and other crops in continuing research. Some seasons may be reasonable random samples of the TPE, but we cannot be sure of this until environments (locations and years together) can be placed precisely within the TPE. The objective of on-going work in companion papers using simulation models of crop growth is to further define the sorghum TPE, and to explain the longer term differences observed among locations (Chapman *et al*. 2000*a*).

Acknowledgments

Thanks are due to Pacific Seeds, Toowoomba (Neil Muller, Brian Hare), and NSW Agriculture, Tamworth (Tony Dale, Bob Murchison), for the supply of variety trial data. We also acknowledge the assistance of Arthur Gilmour in using ASREML to do the variance analyses. We also thank the reviewers for their valuable comments. This research was conducted with funding support from the Grains Research Development Corporation (GRDC) for project DAQ230.

References

Allard RW, Bradshaw AD (1964) Implications of genotype–environmental interactions in plant breeding. *Crop Science* **4**, 503–508.

Bolaños J, Edmeades GO (1993*a*) Eight cycles of selection for drought tolerance in lowland tropical maize. I. Responses in grain yield, biomass, and radiation utilization. *Field Crops Research* **31**, 233–252.

Bolaños J, Edmeades GO (1993*b*) Eight cycles of selection for drought tolerance in lowland tropical maize. II. Responses in reproductive behaviour. *Field Crops Research* **31**, 253–268.

Ceccarelli S, Grando S (1996) Drought as a challenge for the plant breeder. In 'First INTERDROUGHT International Conference, Montpellier, France, 1995. Drought tolerance in higher plants: genetical, physiological and molecular biological analysis'. *Plant Growth Regulation* **20**, 149–155.

Chapman SC, Cooper M, Hammer GL, Butler DG (2000*a*) Genotype by environment interactions affecting grain sorghum. II. Frequencies of different seasonal patterns of drought stress are related to location effects on hybrid yields. *Australian Journal of Agricultural Research* **51**, 209–222.

Chapman SC, Edmeades GO, Crossa J (1996) Pattern analysis of gains from selection for drought tolerance in tropical maize populations. In 'Plant adaptation and crop improvement'. (Eds M Cooper, GL Hammer) pp. 513–527. (CAB International: Wallingford, UK)

Chapman SC, Hammer GL, Butler DG, Cooper M (2000*b*) Genotype by environment interactions affecting grain sorghum. III. Temporal sequences and spatial patterns in the target population of environments. *Australian Journal of Agricultural Research* **51**, 223–233.

Cockerham CC (1963) Estimation of genetic variances. In 'Statistical genetics and plant breeding'. (Eds WD Hanson, HF Robinson) pp. 53–94. NAS-NRC Publication 982. (NAS-NRC: Washington DC)

Comstock RE (1977) Quantitative genetics and the design of breeding programs. In 'Proceedings of the International Conference on Quantitative Genetics'. 16–21 August 1976. pp. 705–718. (Iowa State University Press: Ames, IA)

Cooper M, DeLacy IH (1994) 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.

Cooper M, DeLacy IH, Basford K (1996) Relationships among analytical methods used to analyse genotypic adaptation in multi-environment trials. In 'Plant adaptation and crop improvement'. (Eds M Cooper, GL Hammer) pp. 193–224. (CAB International: Wallingford, UK)

Cooper M, Woodruff DR, Eisemann RL, Brennan PS, DeLacy IH (1995) A selection strategy to accommodate genotype-by-environment interaction for grain yield of wheat: managed-environments for selection among genotypes. *Theoretical and Applied Genetics* **90**, 492–502.

Cullis BR, Thomson FM, Fisher JA, Gilmour AR, Thompson R (1996) The analysis of the NSW wheat variety database. I. Modelling trial error variance. *Theoretical and Applied Genetics* **92**, 21–27.

DeLacy IH, Cooper M, Basford KE (1996) Relationships among analytical methods used to study genotype-by-environment interactions and evaluation of their impact on response to selection. In 'Genotype-by-environment interaction: new perspectives'. (Eds MS Kang, HG Gauch) pp. 51–84. (CRC Press: Boca Raton, FL)

Dickerson GE (1962) Implications of genetic-environmental interaction in animal breeding. *Animal Production* **4**, 47–63.

FAO (1997) 'FAO production yearbook.' (Food and Agriculture Organisation of the United Nations: Rome)

Fehr WR (1987) 'Principles of cultivar development.' Vol. 1. Theory and technique. (MacMillan: New York) 536 pp.

Fischer KS, Edmeades GO, Johnson EC (1989) Selection for the improvement of maize yield under moisture-deficits. *Field Crops Research* **22**, 227–243.

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

Gilmour AR, Cullis BR, Welham SJ, Thompson R (1998) ASREML program user manual, NSW Agriculture Biometrics Bulletin.

Haldane JBS (1947) The interaction of nature and nuture. *Annals of Eugenics* **13**, 197–205.

Hammer GL, Holzworth DP, Stone RC (1996) The value of skill in seasonal climate forecasting to wheat crop management in a region with high climatic variability. *Australian Journal of Agricultural Research* **47**, 717–737.

Henzell RG (1992) Grain sorghum breeding in Australia: current status and future prospects. In 'Proceedings of the Second Australian Sorghum Conference'. Gatton, 4–6 February 1992. Vol. 1. (Eds MA Foale, RG Henzell, PN Vance) pp.70–82. (Australian Institute of Agricultural Science: Melbourne)

Henzell RG, Hare BW (1996) Sorghum breeding in Australia—public and private endeavours. In 'Proceedings of the Third Australian Sorghum Conference'. Tamworth, 20–22 February 1996. (Eds MA Foale, RG Henzell, JF Kneipp) pp. 159–171. (Australian Institute of Agricultural Science: Melbourne)

Mirzawan PDN, Cooper M, Hogarth DM (1993) The impact of genotype multiply environment interactions for sugar yield on the use of indirect selection in southern Queensland. *Australian Journal of Experimental Agriculture* **33**, 629–638.

Patterson HD, Thompson R (1975) Maximum likelihood estimation of components of variance. In 'Proceedings of the 8th International Biometrics Conference'. (Ed. LCA Corsten) pp. 197–207. 1974, Constanta, Rumania. (Editura Academiei Republicii Socialiste: Bucharest, Romania)

Podlich D, Cooper M (1998) QU-GENE: a simulation platform for quantitative analysis of genetic models. *Bioinformatics* **14**, 632–653.

Rathjen AJ (1994) The biological basis of genotype by environment interaction—its definition and management. In 'Proceedings of the Seventh Assembly of the Wheat Breeding Society of Australia'. (Eds J Paull, IS Dundas, KJ Shepherd, GJ Hollamby) pp. 13–17. (Wheat Breeding Society of Australia, The Australian National University: Canberra)

Sheppard JA, Ratnasiri WGA, DeLacy IH, Butler DG, Cooper M, Brennan PS (1996) Classification of Queensland test sites into zones for yield evaluation using multi-environment trials from 1972–1994. In 'Proceedings of the Eighth Wheat Breeding Assembly'. (Eds RA Richards, CW Wrigley, HM Rawson, GJ Rebetzke, JL Davidson, RIS Brettell) pp. 213–217. (Wheat Breeding Society of Australia, The Australian National University: Canberra)

Smith AB, Cullis BR, Gilmour AR, Thompson R (1998) Multiplicative models for genotype by environment in spatial mixed model analyses of multi-environment trial data. In 'Proceedings of XIXth International Biometric Conference'. pp. 91–101. IBC98, Invited Papers, Cape Town.

Manuscript received 4 February 1999, accepted 8 October 1999