Field evaluation of type 2 modified augmented designs for non-replicated yield trials in the early stages of a wheat breeding program

C. H. A. SNIJDERS

Abstract[2]

At the early stages (F_7, F_8) of a wheat breeding program, a large number of test lines need to be evaluated for grain yield, but with limited seed supply per line. These lines are tested in non-replicated multi-location experiments. In this study the use of a modified augmented design for rectangular plots (type 2) was evaluated for grain yield in four multi-location experiments with each the same set of wheat lines, in total 10 experiments. Soil heterogeneity was investigated by ANOVA and existed for all experiments except one. Yield adjustment was based on either row-column effects (Method 1) or covariance (Method 3). Among 10 experiments, seven showed a good agreement between adjustment method as indicated by ANOVA and RE of adjustment. In all trials important changes in the ranking of the test lines did not occur. The MAD-2 provides a convenient and flexible means of measuring environmental heterogeneity, allowing yield adjustment and assessment of test lines for selection purposes. In case the RE does not support the choice for adjustment method based on ANOVA, further research of the experimental design, results and heterogeneity pattern is required.

Introduction

The goal of a wheat breeding program is to develop superior genotypes as a result of many years of selection. In *Figure 1* the general selection scheme for winter wheat as used at Zelder is illustrated. Early generation selection is based on visual observations of yield components (ear size, number of ears, number of spikelets per ear), disease resistance, tillering potential, lodging resistance and seed quality. For the three regions in Europe that is selected for, i.e. Germany/Netherlands/Denmark, France and UK, respectively, every year in total about 850 crosses are made. Between and within cross progeny selection for these regions is carried out in nurseries at Zelder in the Netherlands, at Compièg-

[2] **Abbreviations:** SSD, single seed descent; MAD, modified augmented design; ANOVA, analysis of variance; CV%, coefficient of variation; RE, relative efficiency

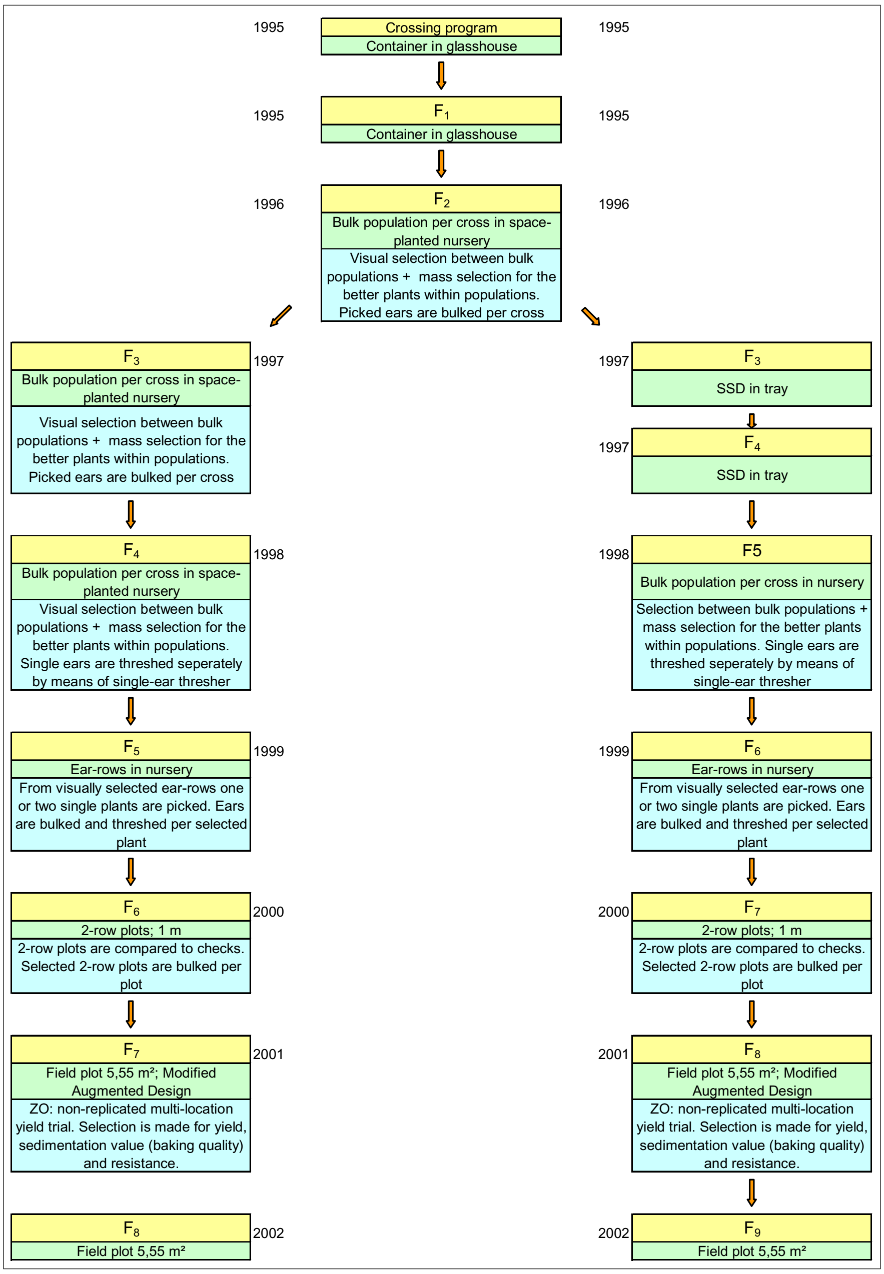

Figure 1: **Selection scheme for winter wheat as carried out at Zelder B.V.**

Autor: Dr. Charles H. A. SNIJDERS, Zelder B.V. Plant Breeders and Seedsmen, P.O. Box 26, NL-6590 AA GENNEP, charles.snijders@chello.nl

ne in France and at Banbury in the UK. In general this results in 7000-9000 lines in the 2-row plot stage (F_6 - F_7) at location Zelder, 3000 lines in Compiègne and 1200 lines in Banbury. From these lines 3,5% - 5% are continued and tested in a grain yield trial. Depending on whether the Single Seed Descent procedure is used or not, the lines are in that first yield trial at either the F_7 or F_8-stage, six years from the time the cross was made. The seed for this first yield trial is derived from a 2-row plot of 1 m long, i.e. about 20 plants. The maximum amount of seed available is 800 g. This amount of seed suffices for four field plots of about 5,5 x 1,5 m² net.

At this early stage of a wheat breeding program we have a large number of lines but limited seed supply per line. Therefore we have to make a choice between either a replicated trial at one location or a non-replicated trial at multi-locations. As adaptation to a wide range of environments is important in a cereal breeding programme, at Zelder lines in this stage have always been tested in a multi-location non-replicated trial. To adjust for environmental heterogeneity, in the past the set of new test lines was divided over many trials. Each non-replicated trial existed of 20 plots consisting of 16 test lines and four check varieties. Sister lines were kept together in a trial to make selection within one cross progeny easier. Relative grain yields and other selection criteria from four locations (Zelder treated and untreated, Dronten treated and untreated) were compared per trial. Although the same check varieties were found in every trial of 20 lines, results from trial to trial were not taken into account.

To test a large number of early lines in one non-replicated trial, Lin and Poushinsky (LIN and POUSHINSKY 1985) proposed a modified augmented design for rectangular plots (type 2). The design is structured as a split-plot with whole plots arranged in rows and columns. Rectangular subplots are arranged in parallel to keep distances between a central control plot and its surrounding test plots as uniform as possible, so that the homogeneity of a whole plot can be preserved. The shape of a whole plot should be relatively square. The control plots with a check variety are used to adjust for soil heterogeneity in the test plots and to obtain some measure of experimental error for line comparisons. A systematic placement of control plots ensures maximum efficiency in adjustment for soil heterogeneity. For estimating the subplot error, a number of whole plots is randomly chosen and two check varieties are assigned to randomly selected subplots in each of the selected whole plots: the control subplots. The subplot controls are used to measure the efficiency of adjustment. Test lines are randomised over the test plots. The percentage of plants required for control lines is relatively low and allows flexibility in methods of adjustment. Three methods using control plots to adjust for soil variation in a modified augmented design type 2 were studied by LIN and POUSHINSKY (1985): using the design structure (row and column correction factors) which is best when soil variation is relatively uniform in one or two directions (Method 1); adjustment using the control plot as fertility index (Method 2), and regression analysis (covariance adjustment) when the variation is multi-directional (Method 3). Method 2 was inferior because of loss of efficiency (LIN and POUSHINSKY 1985; LIN and VOLDENG 1989).

Modified augmented designs with Method 1 and Method 3 adjustment were tested and recommended for breeding programmes in potato (SCHAALJE, LYNCH, and KOZUB 1987), soybean (LIN and VOLDENG 1989) and 2-row barley (MAY and KOZUB 1995; MAY, KOZUB, and SCHAALJE 1989).

The objectives of this study were to evaluate the use of the MAD-2 design for non-replicated yield trials in wheat. The interpretation of the statistical procedures, and the effect of adjustments of grain yield are discussed.

Materials and Methods

Field experiments

The efficiency of the MAD-2 was studied intensively for four groups of experiments, i.e. 4 multi-location experiments with each the same set of wheat test lines, in total 10 experiments. Trials were carried out in 2002 within the collaborative wheat breeding program of Zelder B.V. Plant Breeders and Seedsmen and Landbouwbureau Wiersum B.V. The germplasm used was developed from crosses made within the collaborative wheat program and following the scheme in *Figure 1*. The lines were either in F_7 or F_8 stage, depending on the use of SSD or not. In *Table 1* the experimental set-up for the trials is given. Type 2 MAD (LIN, POUSHINSKY, and VOLDENG 1985) was used for all experiments. The number of rows and columns differed for each experiment, but the number of subplots within a whole plot was always nine. Subplots were end-trimmed and differed in length depending on the location, but were always 1,5 m wide. Rows were separated by paths of alternating 1 m and 2,5 m width. Three check varieties were used: check A as whole plot control in all whole plots situated in the center subplot; checks B and C as subplot controls assigned randomly to a certain number of whole plots defined by the row x column design and number of test lines. Check varieties are shown in *Table 1*. The checks do not have to be necessarily the same for a multi-location trial. More important is to choose checks that are reliable at a given location. In the treated trials the test lines were randomised. In the untreated trials lines were not randomised to ease the visual selection between lines within a cross progeny. In all treated trials herbicides, fungicides and aphid-control were applied. In all untreated trials herbicides and aphid-control were applied. Except for the spring wheat trial SW02-300, all trials were treated with growth regulator (CCC).

Statistical analyses

Trials were designed and analysed with Agrobase Generation II (©Agronomix Software Inc., Winnipeg, Canada). Two adjustment methods for soil variation were evaluated: row-column (Method 1) and regression (Method 3) (LIN and POUSHINSKY 1985).

Soil heterogeneity was investigated by the analysis of variance (ANOVA) of the grain yield data. Row, column and row x column interaction (whole plot error) effects were calculated from the whole plot controls. Subplot error was calculated from the subplot controls. The row and column effects can be tested against the interaction, and the interaction can be tested against the subplot error. The number of degrees of freedom of the sub-

Table 1: **Experimental set-up of 10 MAD-2 wheat experiments in 2002. The checks in control plot and control subplots are given**

Multi-location experiment	Crop	Test location	# Test lines	Row x Column	# whole plots with subplots†	% test plots	Control plot	Control subplot 1	Control subplot 2
WW02-331	winter wheat	Zelder treated	360	10 x 5	20	80%	Bristol	Napier	Koch
		Zelder untreated	360	10 x 5	20	80%	Bristol	Napier	Koch
		Dronten treated	360	7 x 7	16	82%	Bristol	Napier	Vivant
WW03-301	winter wheat	Zelder treated	344	10 x 5	28	76%	Bristol	Napier	Koch
		Dronten treated	344	7 x 7	24	76%	Bristol	Napier	Vivant
		Dronten untreated	344	7 x 7	24	76%	Bristol	Napier	Vivant
WW02-351	winter wheat	Compiègne treated	270	8 x 5	25	75%	Hamac	Isengrain	Charger
		Compiègne untreated	270	8 x 5	25	75%	Hamac	Isengrain	Charger
SW02-300	spring wheat	Zelder spring sown treated	318	10 x 5	41	71%	Tybalt	Pasteur	Lavett
		Zelder autumn sown treated	318	9 x 5	21	79%	Drifter	Pasteur	Tybalt

† number of randomly chosen whole plots in which subplots are allocated

plot error is calculated as *(3-1)m* where *m* is the number of whole plots in which subplot controls are allocated (LIN and VOLDENG 1989). If row, column or row x column interaction effects are not significant, the soil variation can be considered homogeneous and no adjustment is needed. If there is soil heterogeneity in one or two directions, but the interaction is not significant, an additive model can describe the soil variation and Method 1 can be used for adjustment. If the row x column interaction is significantly greater than the subplot error, then the soil variation is described by a non-additive model and Method 3 is used for adjustment (LIN and POUSHINSKY 1985).

Adjustment of test lines is on a whole plot basis. When $Y_{ij(k)}$ is the observed value of the *k*th test line in the whole plot of the *i*th row (*i*=1,...,*r*) and the *j*th column (*j*=1,...,*c*) and $X_{ij(A)}$ is the observed value of the control plot (cultivar A) in the *ij*th whole plot, for Method 1 the adjusted value is

$$Y'_{ij(k)} = Y_{ij(k)} - R_i - C_j$$

where

$$R_i = \sum_{j=1}^{c} X_{ij(A)} / c - \overline{X}_A$$

and

$$C_j = \sum_{j=1}^{r} X_{ij(A)} / r - \overline{X}_A$$

For Method 3 the adjusted value is

$$Y'_{ij(k)} = Y_{ij(k)} - b(X_{ij(A)} - \overline{X}_A)$$

where *b* is the regression coefficient of the mean of two control subplots on the corresponding control plot (LIN and POUSHINSKY 1985).

The effectiveness of the design expressed as relative efficiency (RE) is calculated from the ratios of pooled within-line variances from unadjusted control subplot data to those from adjusted data

$$RE = \frac{MS_{unadjusted\_control\_subplot\_data}}{MS_{adjusted\_control\_subplot\_data}} x100\%$$

Changes in ranking due to adjustment are expressed in the rank correlation between unadjusted and adjusted yields within a trial.

Contour plots were produced using the DCONTOUR directive in Genstat 5 Release 4.1.

Results

In *Table 2* the control plots average (Check A), CV% and ANOVA for grain yield are given for the 10 experiments. In general grain yield was high and CV%'s low. Soil heterogeneity existed for all experiments except for experiment WW02-351 at Compiègne untreated, and yield adjustments are necessary. In *Table 3* the relative efficiencies of the adjustments methods are given together with the rank correlation between adjusted and unadjusted yields. In *Table 4* it is shown which of the two adjusted methods should be chosen based on ANOVA and based on RE, and which method was actually used to adjust the yield data.

Trial WW02-331 at location Zelder treated has the highest CV%. The ANOVA mean squares show that row, column and row x column effect are significant. This indicates that there was two-directional

Table 2: **Average grain yield (kg/ha) and CV% for the control variety, and ANOVA mean squares for grain yields based on control and sub-control plots incorporated in the 10 MAD-2 wheat experiments.**

Multi-location experiment	Crop	Test location	Control plot mean (kg/ha)	CV%	Row MS	Row df	Column MS	Column df	row x column MS	row x column df	Subplot error MS	Subplot error df
WW02-331	winter wheat	Zelder treated	9.991	4,6	489.190*	9	565.191*	4	213.535*	36	106.131	40
		Zelder untreated	9.463	2,9	194.333*	9	532.731*	4	74.874	36	54.331	40
		Dronten treated	10.981	2,5	152.581†	6	136.980	6	77.075	36	74.752	32
WW03-301	winter wheat	Zelder treated	9.465	3,9	638.200*	9	107.636	4	132.711	36	124.358	56
		Dronten treated	10.979	1,9	334.332*	6	107.449*	6	44.715	36	86.020	48
WW02-351	winter wheat	Compiègne treated	10.210	3,0	676.366*	7	1.445.801*	4	94.257	28	62.919	50
		Compiègne untreated	8.861	4,0	140.847	7	11.873	4	123.752	28	72.600	50
SW02-300	spring wheat	Zelder treated	8.776	2,6	519.344*	9	81.222	4	51.230	36	46.386	82
		Zelder autumn sown treated	9.618	3,8	77.875	9	224.758	4	134.318*	36	40.420	42

* Significant at the 0,05 probability level
† Significant at the 0,10 probability level

Table 3: **Relative efficiency (%) and rank correlation between adjusted and unadjusted yields measured by subplot controls for Methods 1 and 3 for the 10 MAD-2 wheat experiments**

Multi-location experiment	Crop	Test location	Relative efficiency (%) Method 1	Relative efficiency (%) Method 3	Rank correlation Method 1	Rank correlation Method 3
WW02-331	winter wheat	Zelder treated	160	160	0,91	0,94
		Zelder untreated	96	130	0,94	0,98
		Dronten treated	167	149	0,96	0,97
WW03-301	winter wheat	Zelder treated	178	218	0,92	0,85
		Dronten treated	106	108	0,95	0,99
		Dronten untreated	130	134	0,94	0,98
WW02-351	winter wheat	Compiègne treated	161	265	0,77	0,88
		Compiègne untreated	68	112	0,99	1,00
SW02-300	spring wheat	Zelder treated	242	211	0,94	0,94
		Zelder autumn sown treated	110	148	0,95	0,94
mean			142	163		

soil variation and that the effects were non-additive. Method 3 based on regression analysis should therefore be the best method of adjustment. The RE does confirm this conclusion. The adjustment does not have much effect on the rankings of these lines. In the untreated yield trial WW02-331 at location Zelder with the same set of test lines, the checks Bristol and Koch did not suffer from any diseases. Check variety Napier was infected by *Septoria tritici*. There were highly significant row and column effects, but no significant interaction between row and column. ANOVA means squares suggest to use Method 1. RE showed no effect of adjustment by Method 1, however a RE of 130% for adjustment Method 3. The same set of winter wheat lines tested at location Dronten treated shows a lower value for CV%. Both row-effect and column effects are considerable but not significant at a probability of 5%.

ANOVA of experiment WW03-301 at location Zelder treated showed that there was an important row-effect, indicating to use adjustment method 1. The RE however was highest for method 3. For the same set of wheat lines at location Dronten treated the CV% is very low.

Although ANOVA shows significant row and column effects, the RE shows that adjustments are not necessary. The results and conclusion for WW03-301 at location Dronten untreated is similar.

Trial WW02-351 at location Compiègne treated showed strong row effects and even stronger column effects. The bi-directional effect is clearly illustrated in *Figure 2*. RE however suggests to use adjustment method 3. The ANOVA results on yield data of the same set of test lines at location Compiègne untreated did not show any significant effects implying no need for adjustment. RE of both adjustment methods were poor and rank correlation between adjusted and unadjusted yield data were about 1. Therefore no adjustment of grain yield data was applied.

Table 4: **Adjustment methods for grain yield as indicated from control and subplot control plots data by ANOVA and RE, respectively, of 10 MAD-2 wheat experiments**

Multi-location experiment	Crop	Test location	Best method ANOVA	Best method RE%	Method used
WW02-331	winter wheat	Zelder treated	3	3	3
		Zelder untreated	1	3	3
		Dronten treated	1	1	1
WW03-301	winter wheat	Zelder treated	1	3	3
		Dronten treated	1	1 or 3	1
		Dronten untreated	1	1 or 3	1
WW02-351	winter wheat	Compiègne treated	1	3	3
		Compiègne untreated	-	-	-
SW02-300	spring wheat	Zelder spring sown treated	1	1	1
		Zelder autumn sown treated	3	3	3

For the spring-sown spring wheat trial SW02-300 at location Zelder treated there was a strong row effect. ANOVA suggested that adjustment method 1 would be best which the RE confirmed. The ANOVA on the yield data of same set of spring wheat lines sown in autumn did not show any significant row- or column-effects. The row x column interaction however was highly significant.

Discussion

Among 10 experiments, seven showed good agreement between the two assessment methods (ANOVA and RE). For two experiments however, ANOVA indicated Method 1 as best adjustment method while RE indicated to use Method 3. As ANOVA of this type of design is not intended as a rigid test of significance, but to provide an overall view of the variation (LIN and VOLDENG 1989), for these experiments the RE was taken as decisive. Method 3 seldom overadjusts (LIN and VOLDENG 1989). Our experiments showed that the superiority of a method depends on variability patterns. For 50% of the experiments Method 1 was the best adjustment, for the other 50% Method 3. On average, the gain in RE was 42% for Method 1 and 63% for Method 3. This also indicates that no method is superior over the other.

In all our trials, important changes in the ranking of the test lines for yield did not occur. The RE and rank correlations were for some trials that high that although ANOVA indicated so adjustment was in fact not necessary. CV%'s were also relatively low. This indicates that the environmental heterogeneity for these trials was low. MAD studies in potato (SCHAALJE, LYNCH, and KOZUB 1987) and barley (MAY, KOZUB, and SCHAALJE 1989) reported high changes in ranking due to adjustment, indicating that adjustment could lead to different selections being made in the breeding program. In our study the lowest rank correlations were found for experiment WW02-351 at Compiègne treated and WW02-301 at Zelder treated.

In our selection we do not use rank, but select those lines that have a relative yield compared to the check varieties larger than 100%. If no adjustment had made for WW02-351 at Compiègne treated, we

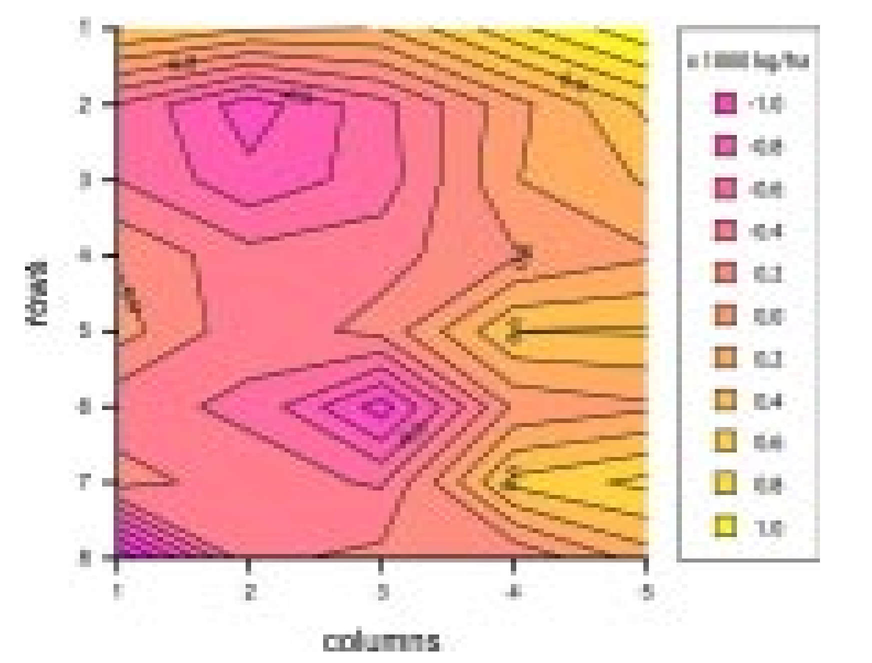

Figure 2: **Contour plot illustrating soil heterogeneity in the treated winter wheat experiment WW02-351 for location Compiègne treated based on unadjusted grain yield data of the control plots with check Hamac. Contour lines connect points with the same deviation from the mean control plot yield in 1000 kg/ha. Intervals are 0.2.**

would have missed 15 lines out of 70 selected based on adjusted yields. If no adjustment had made for WW02-301 at Zelder treated, we would have missed 17 lines out of 50 selected based on adjusted yields.

The advantages of the MAD-2 are that a systematic placement of control plots with checks ensures maximum efficiency in adjustment for soil heterogeneity, and its flexibility to hold different number of test lines and whole plot arrangements in rows and columns. Also the percentage of plants required for control lines is relatively low. In our experiments percentage of test plots varied from 71% to 82%. This percentage depends on the number of sub-plots per whole plot, and the number of whole plots in which control subplots are allocated. A whole plot consisting of nine subplots with the center plot being the control plot is a good choice when testing wheat in plots of this size. Including the paths, the dimensions of a whole plot were in our case 13,5 m x 7,3 m, which is more rectangular than square, where square is in fact a prerequisite for this kind of design (LIN and POUSHINSKY 1985). However, in a practical breeding program the number of subplots per whole plot will always be a compromise between what is theoretically sound and acceptable for practice. Many years of practice and knowledge of the land used for trials has led to an average CV of 3,0% for untreated replicated trials (2000-2002, 31 experiments) and an average CV of 2,6% for treated replicated trials (2000-2001, 34 experiments; SNIJDERS, C.H.A., unpublished data). Therefore although the whole plots are rectangular in form, we may assume that the subplots within a whole plot are homogeneous with respect to soil variability and equal adjustment can be applied to all subplots within a whole plot.

The number of whole plots in which control subplots are allocated should be at least twice the number of columns. The more control subplots, the more accurate the RE can be estimated, but the lower the percentage of test lines in the field. A considerable number of checks is necessary at this early stage of testing lines. Not only for grain yield, but also other agronomic characters like plant length, earliness, lodging tolerance, diseases resistances have to be measured related to check varieties. Whatever design is used, in trials with test lines at this early stage one on four to five plots should always be a check variety.

A disadvantage of MAD-2 lays in the practical aspects: a MAD-2 design determines your complete field layout. As a MAD-2 consumes a large, rather square part of land, all other experiments have to be planned around it. As all whole plots are arranged in rows and columns, it is possible to harvest the trial in two steps. Any systematic error that would arise from this can be corrected for, as long as the borderline between the two harvest times lies between columns or rows. It is not advisable to harvest in two steps, but rain may sometimes force you to.

The MAD-2 test is used to select potentially high yielding lines for further evaluation in a replicated trial. The breeder is interested in the top lines, not in the line differences per se. The adjustment was only applied for grain yield. For all other observations on agronomic and resistance traits unadjusted data were used. Per set of test lines, selections were made based on the adjusted grain yield data and all other observed traits for all locations. Correlations between locations within a multi-location experiment are in general rather poor at this stage of trials. Locations are chosen because of their different differentiating potential. It is up to the breeder to select those lines that are good for all important characters. Average selection intensity over the four sets used in these experiments was 5%. Yield data at this stage in the selection process are regarded as a good indication. It is only in the year following the MAD-2 trial that in a replicated multi-location trial a reliable estimate of the grain yield potential of a lines is received.

The MAD-2 provides a convenient means of measuring environmental heterogeneity, allowing yield adjustment and assessment of test lines for selection purposes. In case the RE does not support the choice for adjustment method based on ANOVA, further research of the experimental design and results are required before a decision for adjustment method is made. A contour plot will certainly help to interpret the results.

Acknowledgements

The Cereal department and farm of Zelder B.V., and the Landbouwbureau Wiersum B.V. team are thanked for their support.

References

LIN, C. S. and G. POUSHINSKY, 1985: A modified augmented design (type 2) for rectangular plots. *Canadian Journal of Plant Science* 65:743-749.

LIN, C. S. and H. D. VOLDENG, 1989: Efficiency of type 2 modified augmented designs in soybean variety trials. *Agron J* 81:512-517.

LIN, C. S., G. POUSHINSKY and H. D. VOLDENG, 1985: Design and model for investigating competition effects from neighbouring test plots. *Canadian Journal of Plant Science* 65 (4):1073-1077.

MAY, K. W. and G. C. KOZUB, 1995: Success of a selection program for increasing grain yield of two-row barley lines and evaluation of the modified augmented design (type 2). *Canadian Journal of Plant Science* 75 (4):795-799.

MAY, K. W., G. C. KOZUB and G. B. SCHAALJE, 1989: Field evaluation of a modified augmented design (type 2) for screening barley lines. *Canadian Journal of Plant Science* 69 (1):9-15.

SCHAALJE, G. B., D.R. LYNCH and G. C. KOZUB, 1987: Field evaluation of a modified augmented design for early stage selection involving a large number of test lines without replication. *Potato Research* 30:35-45.