Efficiency of Single Seed Descent and Early Selection in the Breeding of Self-fertilizing Crops

Efficiëntie van Single Seed Descent en vroege selectie in de veredeling van zelfbevruchtende gewassen

Promotor:	dr. R.F. Hoekstra hoogleraar in de populatie- en kwantitatieve genetica
Co-promotor:	dr. ir. P. Stam universitair hoofddocent in de erfelijkheidsleer

A.J. van Oeveren

Efficiency of Single Seed Descent and Early Selection in the Breeding of Self-fertilizing Crops

Proefschrift

ter verkrijging van de graad van doctor
in de landbouw- en milieuwetenschappen
op gezag van de rector magnificus,
dr. H.C. van der Plas,
in het openbaar te verdedigen
op dinsdag 5 januari 1993
des namiddags te vier uur in de Aula
van de Landbouwuniversiteit te Wageningen.

ABSTRACT

Two selection procedures are examined and compared in the breeding for quantitative traits in self-fertilizing crops. They represent two more or less extreme breeding schemes: a) Early Selection (ES), with early generation cross selection, followed by line selection. The cross selection is based on F_3 estimates of the relevant genetic parameters predicting the distribution of F_∞ inbred lines; b) Single Seed Descent (SSD), where a quick advancement towards the F_5 is combined with line selection only in the F_6. Both field trials and computer simulation studies show that the early cross selection is not an efficient way of breeding. Cross prediction will often be erroneous due to severe bias on estimates of the genetic parameters. This bias is caused by non-additive genetic effects, environmental errors and, especially, intergenotypic competition. The procedure *of single seed descent can produce superior inbred lines in a more consistent, cheaper* and faster way. It appears though that, with maximum input, ES may produce better lines than SSD. Which procedure is more preferable therefore depends on the effort the plant breeder is willing to spend on a relatively small genetic gain.

CIP-gegevens Koninklijke Bibliotheek, Den Haag

Oeveren, A.J. van

Efficiency of single seed descent and early selection
in the breeding of self-fertilizing crops /
A.J. van Oeveren. - [S.l. : s.n.]
Proefschrift Wageningen. - Met lit. opg. - Met
samenvatting in het Engels.
ISBN 90-5485-044-2
Trefw.: plantenveredeling

Omslagontwerp: Ernst van Cleef

STELLINGEN

I. Plantenveredeling in de praktijk is in hoge mate een kansspel.

II. Het verdient aanbeveling om het promotiereglement van de Landbouwuniversiteit aan te passen aan de landelijke wetgeving voor wat betreft de verplichting tot stellingen.

III. De enige echte zin van het leven is niet 'aardbeientaart' doch "Wat is de zin van het leven?".

Loesje

IV. Discriminatie hangt nauw samen met hokjesgeest.

V. Het zou de levenshouding van de westerse consument zeer ten goede komen indien men minstens een week doorbracht in een door hongersnood of ander rampspoed getroffen gebied.

VI. Gezien de hogere levensverwachting van vrouwen dient de pensionering bij mannen op lagere leeftijd te beginnen.

VII. Bij de opmaak van de meeste proefschriften zou de aandacht vooral uit moeten gaan naar de rugzijde van de omslag aangezien dat deel het meest bekeken wordt.

Stellingen behorend bij het proefschrift "Efficiency of single seed descent and early selection in the breeding of self-fertilizing crops" door A.J. van Oeveren, te verdedigen op 5 januari 1993 te Wageningen.

VOORWOORD

Dit proefschrift doet verslag van mijn werkzaamheden als Assistent In Opleiding (AIO), zoals tegenwoordig een promotie-onderzoeker wordt genoemd. Het onderzoek was een samenwerkingsproject tussen de vakgroep Erfelijkheidsleer van de Landbouwuniversiteit (LU) en het Centrum voor Plantenveredelings- en Reproductieonderzoek van de Dienst Landbouwkundig Onderzoek (CPRO-DLO). In de tijd dat mijn AIO-schap begon (juni 1988), heette laatstgenoemde instituut nog de Stichting Voor Plantenveredeling (SVP). In het kader van reorganisatie en bezuiniging bij DLO vond eerst een fusie met het Instituut voor de Veredeling van Tuinbouwgewassen (IVT) en het Instituut voor Toepassing van Atoomenergie in de Landbouw (ITAL) plaats tot het CPO, waarna al snel een fusie met het Centrum voor Reproductieonderzoek en Zaadtechnologie (CRZ) volgde, uitmondend in het huidige CPRO.

Deze samenwerking hield in dat ik twee werkplekken had: één op de Wageningse berg en één ten noorden van Wageningen (de Haaff). Dit had zo zijn voordelen; altijd wel een collega om mee te kletsen en de mogelijkheid om van beide plekken de leuke graantjes mee te pikken. Natuurlijk waren er ook nadelen. Met dubbele werkbesprekingen en colloquia was een week al snel gevuld. Bovendien waren twee bureaus en twee (al dan niet completely) personal computers er de reden van dat de benodigde documenten of computerbestanden net op de andere locatie lagen. Dit betekende steeds een zware tas en de laatste files op flop mee naar huis.

Ook had ik wat meer collega's dan gebruikelijk, die gezamenlijk voor een uitstekende werksfeer zorgden. Met name mijn kamergenoten wil ik hiervoor bedanken. Op de vakgroep: Johan van Ooijen, die mij ook de eerste weg in dit onderzoek gewezen heeft, en later Maarten Nauta en Chris Maliepaard. Op het CPRO: Henk Holthof, die mij bovendien met filosofische E-mail communicatie bestookte. Piet Stam wil ik bedanken voor de tijd die hij wist vrij te maken voor de begeleiding van mijn onderzoek, ondanks zijn drukke werkzaamheden als afdelingshoofd bij het CPRO. Verder ben ik Rolf Hoekstra erkentelijk voor zijn bereidheid en inzet om op te treden als promotor.

Het onderzoek heb ik grotendeels alleen uitgevoerd, maar voor de veldproeven was toch de hulp van een groot aantal mensen onmisbaar. Ten eerste wil ik

Henk Oosting bedanken voor zijn enthousiaste en deskundige hulp. Zonder hem zou de (onderbemande) proefvelddienst op de Haaff het waarschijnlijk niet redden. Verder dank ik Marijke Broos-Reitsma, Peter van der Putten, Geert Scholten, Jan Menting, Mart-Jan Prins, Jos van Schaik, Henk Vleeming en de 'Broekema-jongens' Jan Witteveen, Gerrit Bunk en Jan Timmers voor al het, al dan niet vuile werk dat ze voor me opgeknapt hebben.

De kweekbedrijven van der Have en Zelder en met name de respectievelijke graanveredelaars Roothaan en Groenewegen ben ik zeer erkentelijk voor hun welwillendheid om gegevens ter beschikking te stellen voor een economische analyse.

Tot slot wil ik het Landbouw Export Bureau (LEB-fonds) bedanken voor de financiële bijdrage in de totstandkoming van dit proefschrift.

JAN

CONTENTS

1. INTRODUCTION

The research described in this thesis concerns some of the problems and possibilities of selection in the breeding of self-pollinating crops. The classical way of breeding is to develop new varieties by creating genetic variation and subsequently extracting from this new gene pool well performing genotypes by means of selection. In case of self-pollinating crops this genetic variation is created by crossing two or more homozygous parent varieties each exhibiting one or more good properties. The first generation hybrids (F_1) are selfed to produce segregating populations. Homozygous genotypes can be obtained after several generations of self-fertilization and these are the new potential varieties. The final goal in breeding is to obtain a genotype which will have all desirable properties from its parents accumulated and hopefully will perform better than all existing varieties. On the way to this goal, choices have to be made concerning the size of the successive generations and, accordingly, the selection intensity. Populations have to be large enough to avoid serious genetic sampling errors and small enough to enable a practical and economical programme.

Another problem is the time of selection; is it effective to select in an early generation, where plants are still heterozygous and, consequently, offspring lines are heterogeneous, or will it be more favourable to postpone selection to a late generation where plants are largely homozygous and very homogeneous within lines? Provided a correct identification of superior genotypes, the first option has the advantage that the number of genotypes to be propagated is reduced. The advantage of the second option is that, especially considering quantitative traits and line-based selection, line estimates in the late generations are much more accurate than in the early stage and selected lines are more stable over generations.

The present study deals with the problems of optimising the selection procedure. Two more or less extreme breeding procedures are examined and compared, as examples of the above options:

a) Early Selection (ES), with cross selection in the F_3, followed by line selection, and

b) Single Seed Descent (SSD), where a quick advancement towards the F_5 is combined with line selection only in the F_6.

The comparison between the two procedures comprises the difference between early and late generation selection as well as the effectiveness of selection among crosses. Only selection on a single, quantitative trait is considered. In the field trials this trait is substituted by yield as the most prominent example.

Early Selection

Some authors considered selection for yield in early generations a useful tool in plant breeding (Weber, 1984), or even obligatory for maximizing selection results (Townley-Smith et al., 1973; Sneep, 1977). A complete population, including all possible genotypes, can be grown in the early stage and thus the risk of losing desirable genes is limited. However, it is evident that there are several disadvantages to early selection; much time and labour has to be spent on the early generation trials and yield estimates may be biased by the effects of limited plot size and competition (Sneep, 1981). Further theoretical considerations led other authors to conclude that postponing selection for quantitative traits to later generations can be advantageous (Snape & Riggs, 1975; Jinks & Pooni, 1981a). Some field trials concerning this matter reported of a more or less successful prediction of inbred line performance from early yield trials in wheat (DePauw & Shebeski, 1973; Cregan & Busch, 1977; Seitzer & Evans, 1978). However, other trials showed no or non-significant correlation between F_3-yields on the one hand and F_5/F_6-yields on the other hand (Briggs & Shebeski, 1971; Whan et al., 1981).

In the perspective of a successful early yield testing, methods have been developed to enable a selection between crosses. This early cross prediction method is based on quantitative genetic theory (e.g. Mather & Jinks, 1971). According to this theory the genetical distribution of inbred lines (F_∞) derived from a certain cross can be predicted from the early generations. Several methods have been designed to obtain estimates, required for such predictions: the North Carolina Experiment III (Comstock & Robinson, 1952), the triple test cross (Kearsey & Jinks, 1968), a method using basic generations F_1, F_2, B_1 and B_2 (Jinks & Perkins, 1970) and one using only F_3 data (Jinks & Pooni, 1980). For practical purposes the latter is largely preferred to the others because it does not include large numbers of test crosses and the corresponding time and labour. This method uses estimates of the mean (m) and additive genetic variance (D) of a cross to predict the probability of obtaining superior inbred lines in the F_∞ progeny. Based on this probability, the most promising crosses are identified and propagated. Field trials on this method showed that it performs fairly well for highly heritable traits in tobacco (Jinks & Pooni, 1980), in barley (Caligari et al., 1985; Thomas et al., 1986) and wheat (Snape & Parker, 1986), whereas predictions concerning yield were less accurate. Van Ooijen (1989a; 1989b) tested the method for yield with mixtures of spring wheat. This author concluded that estimates of the genetic parameters are often severely biased leading to erroneous cross prediction.

The main sources of bias reflect differences between the F_3 generation and the predicted F_∞. Firstly, there is bias due to the simplifying assumptions of the genetical model. Parameters predicted from the F_3 and estimated from the F_∞ differ by non-additive genetic effects, which are neglected. They consist of dominance (h) and epistasis (i). Secondly, there are errors due to differences between the environments in which the two generations are grown: small plots with mixtures of genotypes in the F_3 and large plots with genotypes in a pure stand in the F_∞. These differences in growing conditions comprise residual plot errors and effects of intergenotypic competition.

Using the terminology of Mather & Jinks (1971), the F_3 cross mean, with expectation $m + \frac{1}{4}h$, is taken as prediction of F_∞ mean (m). Accordingly, the genetic variance of the F_∞ (D + I) is predicted by twice the F_3 between line variance: $2 \cdot \sigma^2_{BL} = D + \frac{1}{8}H + \frac{1}{2}I$ (any higher order of epistasis is neglected). Thus the prediction of the mean is overestimated by $\frac{1}{4}h$, whereas the genetic variance is overestimated by $\frac{1}{8}H$ and underestimated by $\frac{1}{2}I$. Therefore, the level of possible bias by dominance and epistasis depends on the relative size of both effects. Van Ooijen (1989a;1989b) concluded that dominance and epistasis were of only minor influence, whereas the major part of the bias was due to intergenotypic competition.

This competition can cause yield in a mixed stand of many different genotypes, like in the F_3 generation, to differ substantially from yield of the same genotype in a pure stand (Spitters, 1979). Especially in case of small grain crops like wheat, where the size of F_3 plots is limited by the amount of F_2 seed, this bias can be large. In order to minimize the effects of between-plot competition it is best to grow the single plant progeny on a 3-row plot of ±1.5 m^2 in two replicates (Kramer et al., 1982).

Single Seed Descent

SSD is a promising method for obtaining homozygous inbred lines in a fast way

(Goulden, 1939; Brim, 1966). As its name indicates, it is a method where each generation is derived from the former by taking from each parent plant only one seed. Plants are grown in the greenhouse and special growing conditions can force early flowering. For small grain crops like wheat, three or even four successive generations can be grown in one year. SSD can be practised from, for instance, the F_2 to the F_5 generation and a first selection can be performed in the F_6, consisting of single plant progenies. Advantages are the quick generation advancement and the accuracy of line estimates.

Field trials have been done on comparing SSD with conventional pedigree breeding concerning yield in several autogamous crops. Comparing unselected SSD-F_6 lines with a F_6 derived from pedigree selection, authors report of equal performances of both procedures and similar frequencies of desirable genotypes in barley (Park et al., 1976; Rossnagel et al., 1987). In other research, where SSD until the F_6 was followed by one or two generations of selection, again no consistent differences in performance were found between F_8 lines from both procedures in soybeans (Boerma & Cooper, 1975), and cowpea (Obisesan, 1992). As the above reports all involve various forms of pedigree breeding, concerning selection intensities and population sizes, it leaves insufficient ground for general conclusions. Computer simulation studies (Casali & Tigchelaar, 1975) have shown that SSD is more favourable at low heritability, while pedigree breeding performs better at higher heritabilities.

Some alternatives exist for a fast generation advancement towards homozygosity. Two of these are the Random Bulk (RB) and the Doubled Haploids technique (DH). Random Bulk consists of generation advancement by harvesting all plants in bulk and planting a random sample of the seeds in the next generation. This may also be done in the greenhouse and thus allows several generations in one year. The difference with SSD is that natural selection does play an important role in RB. When competitive ability is not or negatively correlated with monoculture yield, this may lead to the selection of undesirable genotypes. Field trials have been performed concerning a comparison between SSD and Random Bulk in soybeans (Empig & Fehr, 1971), lentils (Haddad & Muehlbauer, 1981) and wheat (Tee & Qualset, 1975). All authors conclude that SSD is better for maintaining genetic variance than RB. However, differences were small and SSD requires more effort.

With doubled haploids a fully homozygous situation can be achieved in one step. In theory this is a very efficient way to create inbred lines; however, recombination will be limited to a single (F_1-derived DH) or two meioses (F_2-derived DH). Thus, when linkage is present and genes are in repulsion phase, genetic variance will be small compared to the genetic variance between SSD-derived lines (Snape, 1976; Jinks & Pooni, 1981b; Snape & Simpson, 1984). This is confirmed by computer simulation results (Riggs & Snape, 1977). Yonezawa et al. (1987) concluded from their simulation study that DH can only compete with conventional methods when the production of doubled haploids is relatively easy. Furthermore, these authors concluded that not too many loci (<10) should be involved and these should not be linked too tightly. Some field trials report of lower generation means for DH than for SSD for a number of traits, including yield, in tobacco (Schnell et al., 1980) and barley (Rossnagel et al, 1987). Genetic variation between (unselected) lines from both procedures was equivalent in both trials. The difference in means may have been partly due to sub-optimal techniques for the production of doubled haploids. Other reports on barley breeding show no consistent differences between the two procedures (Park et al., 1976; Choo et al., 1982; Jinks et al., 1985; Powell et al, 1986; 1992; Caligari et al., 1987). Linkage may cause differences in performance between the two procedures, although Powell & Caligari (1986) concluded from a field trial comparing SSD and DH that it is of no importance. In order to achieve a higher fraction of recombinants, it is useful to derive doubled haploids from the F_2 instead of the F_1 (Jinks & Pooni, 1981b; Yonezawa et al., 1987). In that perspective both methods do not differ much in the genetical composition of their offspring. Which method is preferred depends on an economic comparison.

Aim and outline of this study

In this study a comparison is made between two more or less extremes of all possible selection procedures: a) early cross selection (ES) versus b) single seed descent (SSD). More precisely, the following schemes are considered, which are schematically visualized in Fig. 1.

a) ES is practised in the F_3 of the breeding programme of an autogamous crop, where the quantitative trait is determined on F_2-derived F_3 lines. Based on the probability of obtaining superior inbred lines, only the most promising crosses are selected. Line selection can be performed within the selected F_3 crosses and the best lines are increased to larger F_4 lines which give a more sound yield estimate. The best lines are selected and from each line a few plants are randomly chosen and propagated to single plant progeny F_5 plots.

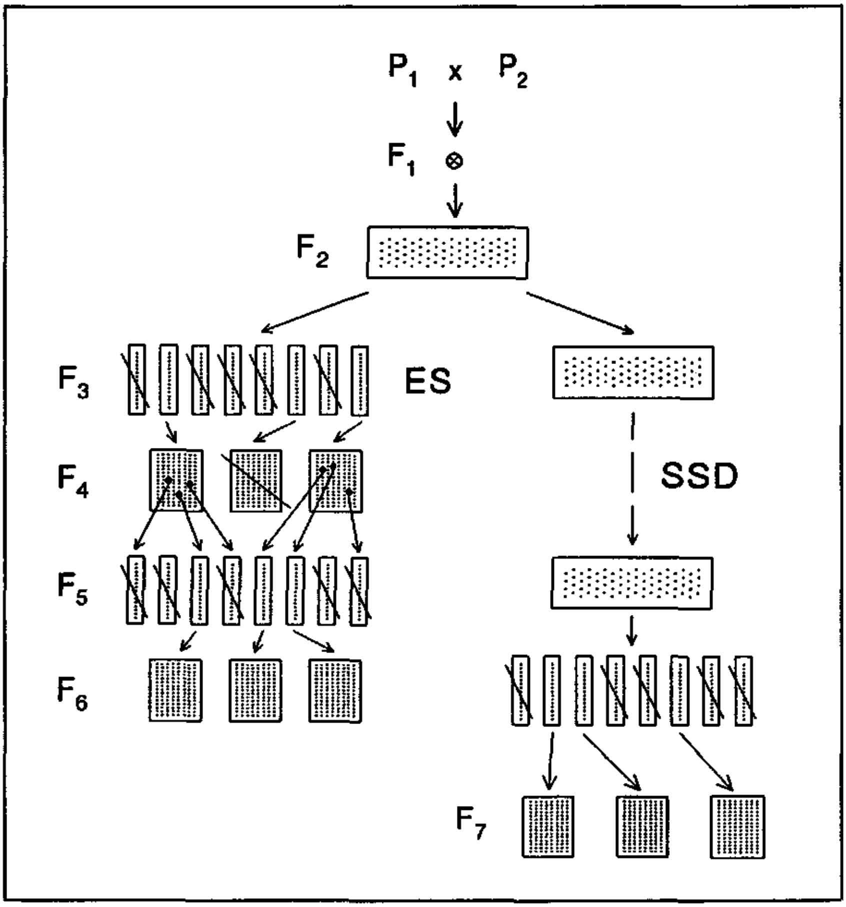

Fig. 1. Scheme of the two breeding procedures in case of one cross. Early Selection (left) and Single Seed Descent (right). When several crosses are involved, a selection between crosses is made in the F_3 with ES, whereas with SSD lines from all crosses are pooled.

These again offer a rough yield estimate and leave opportunity for line selection. Again selected lines are increased to the F_6 which consist of fields of the same size as the F_4. Thus F_6 lines are derived from a single F_4 plant.

b) SSD is practised on the progeny of the same crosses, starting with the F_2. From the SSD-F_3 an F_5 generation is derived by two more successive rounds of SSD. This F_5 is space planted and increased to single plant progeny F_6 lines. Yield estimates are obtained from the F_6 lines and only then selection is made between pooled lines of all crosses. The best lines are propagated to large F_7

lines, on the same scale as the ES-F_6.

A final evaluation is made comparing the SSD-F_7 with the ES-F_6. Although they differ by one generation it is the most logical comparison, considering the time schedule. Because the three SSD generations can all be grown in one year, the SSD-F_7 may even be one year ahead of the ES-F_6. Both generations consist of single plant progenies of two generations earlier and their level of heterozygosity will be very much alike.

Early cross selection has not yet been compared to SSD in any previous study. The efficiency of early line selection (ELS) on yield in the F_3 (no selection between crosses) in comparison with SSD has been investigated. Boerma & Cooper (1975) found no significant differences in selection results between the two procedures and pedigree breeding in soybeans. Because SSD is quicker and cheaper they concluded it is more favourable. Knott & Kumar (1975) found that ELS led to a higher mean than SSD in wheat. The 20% best lines from the SSD procedure, however, were better than those from the ELS. These authors also concluded that SSD can be an efficient way of breeding. In the present study ELS is also compared to ES to investigate the effectiveness of cross selection (chapter 6).

In chapters 2 to 4 ES and SSD are compared by means of field trials with mixtures of spring wheat varieties. These mixtures are used to mimic segregating populations. Growing the constituent varieties in monoculture enables the assessment of the genetic parameters in the F_∞. The latter are also predicted from the ES-F_3 and the SSD-F_6 mixtures. In this way the accuracy of the ES- and SSD-based predictions can be examined and sources of possible bias can be identified. This is an extension of the research by Van Ooijen (1989a;1989b), who merely compared F_3 predictions with the F_∞ parameters. Because of the use of variety mixtures the genetical constitution of the trait was limited to two loci (chapter 2) and three loci (chapter 4). Chapter 3 reports on the evaluative generations of the first trial from chapter 2.

Though the results of the experiments with simulated breeding generations are indicative as to the performance of the alternative selection methods, they do not allow firm conclusions that can be generalized. The limited set of crosses being mimicked and the uncontrollable factors such as heritability and genotype-environment interaction will prohibit a clear insight into the causes of different performance of the selection methods. Therefore, the same procedures were also investigated by means of computer simulations. With simulation the large number of replicate experiments that can be run provides a more secure basis for statements about the average performance of both procedures. There are, of course, some disadvantages to a computer simulation study: simplifying assumptions must be made. On the other hand, it allows variation of parameter values, such that the relative significance of the input variables on the selection results can be studied. This is described in chapter 5. An extended model is discussed in chapter 6 and an economically based comparison between the two procedures is considered. Accordingly, indications for an optimum breeding scheme are given.

References

Boerma H.R. & R.L. Cooper, 1975. Comparison of three selection procedures for yield in soybeans. Crop Science 15: 225-229.

Briggs K.G. & L.H. Shebeski, 1971. Early generation selection for yield and breadmaking quality of hard red spring wheat (*Triticum aestivum* L. em Thell.). Euphytica 20: 453-463.

Brim, C.A., 1966. A modified pedigree method of selection in soybeans. Crop Science 6: 220.

Caligari P.D.S., W. Powell & J.L. Jinks, 1985. The use of doubled haploids in barley breeding 2. An assessment of univariate cross prediction methods. Heredity 54: 353-358.

Caligari P.D.S., W. Powell & J.L. Jinks, 1987. A comparison of inbred lines derived by doubled haploidy and single seed descent in barley (*Hordeum vulgare*). Ann. appl. Biol. 111: 667-675.

Casali V.W.D. & E.C. Tigchelaar, 1975. Computer simulation studies comparing pedigree, bulk and single seed descent selection in self pollinated populations. J. Amer. soc. hort. sci. 100: 364-367.

Choo T.M., E. Reinbergs & S.J. Park, 1982. Comparison of frequency distributions of doubled haploid and single seed descent lines in barley. Theor. Appl. Genet. 61: 215-218.

Comstock, R.E. & H.F. Robinson, 1952. Estimation of average dominance of genes. In: Gowen, J.W. (Ed.), Heterosis. Iowa State College Press, Ames. pp. 494-516.

Cregan, P.B. & R.H. Busch, 1977. Early generation bulk hybrid yield testing of adapted hard red spring wheat crosses. Crop Science 17: 887-891.

DePauw, R.M. & L.H. Shebeski, 1973. An evaluation of early generation yield testing procedure in *Triticum aestivum*. Can. J. Plant Sci. 53: 465-470.

Empig, L.T. & W.R. Fehr, 1971. Evaluation of methods for generation advance in bulk hybrid soybean populations. Crop Science 11: 51-54.

Goulden, C.H., 1939. Problems in plant selection. Proceedings of 7th international genetical congress. Edinburgh. pp. 132-133.

Haddad, N.I. & F.J. Muehlbauer, 1981. Comparison of random bulk population and single-seed-descent methods for lentil breeding. Euphytica 30: 643-651.

Jinks, J.L., M.K.U. Chowdhury & H.S. Pooni, 1985. Comparison of the inbred lines derived from a hybrid tobacco (burley x flue cured) by dihaploidy and single seed descent. Heredity 55: 127-133.

Jinks, J.L. & J.M. Perkins, 1970. A general method for the detection of additive, dominance and epistatic components of variation. III. F_2 and backcross populations. Heredity 25: 419-429.

Jinks, J.L. & H.S. Pooni, 1980. Comparing predictions of mean performance and environmental sensitivity of recombinant inbred lines based upon F_3 and triple test cross families. Heredity 45: 305-312.

Jinks, J.L. & H.S. Pooni, 1981a. Comparative results of selection in the early and late stages of an inbreeding programme. Heredity 46: 1-7.

Jinks, J.L. & H.S. Pooni, 1981b. Properties of pure-breeding lines produced by dihaploidy, single seed descent and pedigree breeding. Heredity 46: 391-395.

Kearsey, M.J. & J.L. Jinks, 1968. A general method of detecting additive, dominance and epistatic variation for metrical traits. I. Theory. Heredity 23: 403-409.

Knott, D.R. & J. Kumar, 1975. Comparison of early selection yield testing and a single seed descent procedure in wheat breeding. Crop Science 15: 296-299.

Kramer, Th., J.W. van Ooijen & C.J.T. Spitters, 1982. Selection for yield in small plots of spring wheat. Euphytica 31: 549-564.

Mather, K. & J.L. Jinks, 1971. Biometrical genetics. Chapman and Hall, London, 382 pp.

Obisesan, I.O., 1992. Evaluation of pedigree and single seed descent selection methods for cultivar development in cowpea (*Vigna unguiculata* L. Walp). Plant Breeding 108: 162-168.

Ooijen, J.W. van, 1989a. The predictive value of quantitative genetic parameters in autogamous crops: Bias caused by intergenotypic competition. 1. $F\infty$-mean.Euphytica 44: 81-94.

Ooijen, J.W. van, 1989b. The predictive value of quantitative genetic parameters in autogamous crops: Bias caused by intergenotypic competition. 2. $F\infty$-variance. Euphytica 44: 95-108.

Park, S.J., E.J. Walsh, E. Reinbergs, L.S.P Song & K.J. Kasha, 1976. Field performance of doubled haploid barley lines in comparison with lines developed by the pedigree and single seed descent methods. Can. J. Plant Sci. 56: 467-474.

Powell, W. & P.D.S. Caligari, 1986. Investigations into the linkage of genes controlling individual quantitative characters in barley. Can. J. Genet. Cytol. 28: 63-68.

Powell, W., P.D.S. Caligari & W.T.B. Thomas, 1986. Comparison of spring barley lines produced by single seed descent, pedigree inbreeding and doubled haploidy. Plant Breeding 97: 138-146.

Powell, W., W.T.B. Thomas & D.M. Thompson, 1992. The agronomic performance of anther culture derived plants of barley produced via pollen embryogenesis. Ann. appl. Biol. 120: 137-150.

Riggs, T.J. & J.W. Snape, 1977. Effects of linkage and interaction in a comparison of theoretical populations derived by haploid and single seed descent methods. Theor. Appl. Genet. 49: 111-115.

Rossnagel, B.G., M.A. Sariah & K.N. Kao, 1987. Field evaluation of anther derived breeding materials compared to materials developed by the pedigree and single seed descent methods in barley (*Hordeum vulgare* L.). Barley Genetics V; 997-1004.

Schnell, R.J. II, E.A. Wernsman & L.G. Burk, 1980. Efficiency of single-seed-descent vs. anther-derived dihaploid breeding methods in tobacco. Crop Science 20; 619-622.

Seitzer, J.F. & L.F. Evans, 1978. Yield gains in wheat by the pedigree method of selection and two early yield tests. Plant breeding 80: 1-10.

Snape, J.W., 1976. A theoretical comparison of diploidised haploid and single seed descent populations. Heredity 36: 275-277.

Snape, J.W. & B. Parker, 1986. Cross prediction in wheat using F_3 data. In: Biometrics in plant breeding, Proceedings of the 6th meeting of the Eucarpia section biometrics in plant breeding, University of Birmingham, U.K., pp. 359-369.

Snape, J.W. & T.J. Riggs, 1975. Genetical consequences of single seed descent in the breeding of self-pollinating crops. Heredity 2: 211-219.

Snape, J.W. & E. Simpson, 1984. Early generation selection and rapid generation advancement methods in autogamous crops. In: Lange, W., A.C. Zeven & N.G. Hogenboom (Eds.), Efficiency in plant breeding: Proceedings 10th Congress of Eucarpia. Pudoc, Wageningen, pp. 82-86.

Sneep J., 1977. Selection for yield in early generations of self-fertilizing crops. Euphytica 26: 27-30.

Sneep, J., 1981. Some results of selection in early generations of small grains. Quantitative genetics and breeding methods. Gallais, A. (Ed.), Quantitative genetics and breeding methods: Proceedings 4th meeting of Eucarpia, section biometrics in plant breeding. Lusignan. pp. 85-90.

Spitters, C.J.T., 1979. Competition and its consequences for selection in barley breeding. Pudoc, Wageningen, 268 pp.

Thomas, W.T.B., C.R. Tapsell & W. Powell, 1986. The application of cross prediction in spring barley breeding. In: Biometrics in plant breeding, Proceedings of the 6th meeting of the Eucarpia section biometrics in plant breeding, University of Birmingham, U.K., pp. 80-89.

Tee, T.S. & C.O. Qualset, 1975. Bulk populations in wheat breeding: comparison of single seed descent and random bulk methods. Euphytica 24: 393-405.

Townley-Smith, T.F., E. Hurd & D.S. McBean, 1973. Techniques of selection for yield in wheat. In: Sears, E.R. & L.M.S. Sears (Eds.), Proceedings 4th International Wheat Genetics Symposium. University of Missouri, pp. 605-609.

Weber, W.E., 1984. Selection in early generations. In: Lange, W., A.C. Zeven & N.G. Hogenboom (Eds.), Efficiency in plant breeding: Proceedings 10th Congress of Eucarpia. Pudoc, Wageningen, pp. 72-81.

Whan, B.R, A.J. Rathjen & R. Knight, 1981. The relation between wheat lines derived from the F_2, F_3, F_4 and F_5 generations for grain yield and harvest index. Euphytica 30: 419-430.

Yonezawa, K., T. Nomura & Y. Sasaki, 1987. Conditions favouring doubled haploid breeding over conventional breeding of self-fertilizing crops. Euphytica 36: 441-453

2. A COMPARISON BETWEEN SINGLE SEED DESCENT AND EARLY CROSS SELECTION IN WHEAT BREEDING

This chapter is published in Euhpytica 58: 275-287

Summary

Two selection procedures in wheat breeding were compared on the basis of their ability to supply high yielding inbred lines. The first procedure consists of an early selection between crosses in the F_3 generation, based on predictions of the cross mean and the between-line variance. In the second procedure selection is postponed until the F_6, which is derived by single seed descent. The two procedures are evaluated in a two year test, using pseudo-lines of spring wheat. These pseudo-lines consist of mixtures of varieties and enable an estimation of the exact genetic parameters. In this way the accuracy of the predictions can be examined.

In case of early selection, it appears that the predictions of the cross mean and especially the between-line variance are very inaccurate. This is caused by the effects of plot size, intergenotypic competition and, to a lesser extent, dominance and/or epistasis. It results in an erroneous ranking of the crosses and the discarding of the potentially best cross. The SSD-F_6 line estimates are much more accurate and thus the better lines are indeed selected. A first comparison between the two selection procedures therefore indicates a preference to the SSD method.

Introduction

Various methods are used to obtain high performing inbred lines in the breeding of self-pollinating crops. Selection can be applied in different stages of the breeding process and with varying intensities. In order to get a better view on the efficiency of different selection procedures two contrasting selection methods are examined.

The first is a method, based on a quantitative genetic theory, which has been developed to predict the genetic potential of a certain cross in an early breeding generation (Mather & Jinks, 1971). With this technique, which uses estimates of the mean (m) and additive genetic variance (D) of a cross, it should be possible to identify the most promising crosses in an early generation. Only those crosses are retained and in the subsequent generations line selection can be performed. This procedure will be referred to as 'Early Selection' (ES).

From experimental studies on 'early selection' (Jinks & Pooni, 1980; Van Ooijen, 1989a, 1989b) it was found that estimating the additive variance (D) by using only F_3 lines is by far the most practical application of this procedure. Estimates based on F_3 lines however may be biased by the effects of intergenotypic competition. Especially in small grain crops this bias can be large, because plots are small and the level of heterogeneity within plots is high. Van Ooijen (1989a, 1989b) concluded from his studies with mixtures of spring wheat varieties that the estimates of the genetic parameters for yield are indeed severely biased, leading to unreliable cross predictions. This confirmed earlier research by Spitters (1979) from experiments with binary mixtures of barley.

A completely different method is to avoid selection in an early stage and wait until a high degree of both homogeneity and homozygosity has been reached. At this stage, selection between lines can be made, without regard to the pedigree. An example of this is the method of 'Single Seed Descent' (SSD). One great advantage of this technique is the opportunity to grow three successive generations in one year (spring cereals).

The SSD-method has proven to be a fast breeding procedure, but tests on the advantages compared to early generation selection methods are quite inconsistent. Knott & Kumar (1975) found in their field experiment with spring wheat that an early yield test procedure (EYT) produced lines with a significantly higher average yield than did SSD. But the yield level of the 20% best lines did not differ significantly for EYT and SSD. Computer simulation studies showed that, especially at low heritability, SSD performs just as well as pedigree selection (Casali & Tigchelaar, 1975).

The goal of the present study was to evaluate the differences in selection response between these two 'extreme' breeding approaches. On the one hand it makes a comparison between early and late selection and on the other hand the effectiveness of a selection among crosses is investigated. Yield in spring wheat is used as the character of interest.

Materials and methods

A field trial was performed, using the so called 'pseudo-lines' technique, instead of real crosses (Van Ooijen, 1989a;1989b). This technique provides a way to create any breeding generation that we like. Besides this, it provides the unique opportunity of an exact estimation of the genetic parameters. It is assumed that yield, as an example of a quantitative trait, segregates in a number of crosses at two independent loci A/a and B/b. For each cross this will give nine different genotypes, each of which is represented by a specific variety. So nine varieties are used to compose all lines in one cross. Each cross is represented by a different set of varieties and in this way different levels of additive, dominance and epistatic effects are created. These genetic effects are related over generations by changes in the variety frequencies. Segregating lines are represented by mixtures of the appropriate varieties.

Table 1. Varieties and pure breeding lines used for the composition of the pseudo-lines, numbered according to their expected yield capacity (1 =high, 23 =low) and their mean yields in both the pseudo-line experiment (1.35 m^2 plots) and the monoculture experiment (10.5 m^2 plots) in kg ·ha^{-1}.

	pseudo		mono	
variety	yield	rank	yield	rank
1. Stratos	2775	10	2687	15
2. G 74010	2721	11	2806	13
3. vdH 1166	2267	20	2668	18
4. vdH 1132	2217	21	2385	22
5. ZESC 1963	3175	3	3241	3
6. TK 2832 3	2976	6	2892	11
7. vdH 3132	2775	12	2683	16
8. G 8005	2336	17	2715	14
9. Minaret	2523	14	2860	12
10. Heros	2443	18	2608	20
11. Darima	2704	8	3045	8
12. TK 2832 2	2285	2	2661	19
13. Ralle	2037	23	2303	23
14. Sunnan	2921	4	3008	9
15. Bastion	2613	16	2527	21
16. Spartacus	2712	7	3215	4
17. Adonis	2696	13	3159	6
18. Melchior	2815	9	3109	7
19. Axona	2953	5	3457	1
20. Kokart	2484	19	2680	17
21. Sicco	3244	1	3429	2
22. Wembley	3160	2	3189	5
23. Prinqual	2571	15	2967	10
mean :	2703		2882	
s.e. :	153		82	

In addition to these pseudo-lines, all varieties used in the simulated crosses are grown in pure stand in large yield trials. These monoculture yields are considered as the true genotypic values of components of the mimicked crosses. They enable a comparison between the estimates of m and D from the pseudo-crosses and the corresponding 'true' values obtained from the monoculture yields. Thus this method provides a way to single out bias due to growing in small plots in a mixed stand from other potential sources of error, such as dominance. For a more detailed description of the pseudo-lines technique: see Van Ooijen (1989a and 1989b).

Both an F_3 and an F_6 generation were created by means of pseudo-lines. Fifteen crosses were mimicked, assuming different sets of two homozygous parents. Hypothetically both the F_3 and the F_6 of each cross are derived from the same diheterozygous F_1. The F_3 consisted of F_2-derived lines. In this F_3 early selection was applied. The F_6 lines were supposed to be derived from F_5 plants, which originated from the F_2 by means of single seed descent. Because both F_3 and F_6 consisted of single plant derived lines, field size was limited to 3-row plots of ± 1.4 m^2 in two replicates. Such plots are assumed to be large enough to give a rough estimate of the yielding potential of that particular F_3 or F_6 line (Kramer et al., 1982).

Twentythree varieties and pure breeding lines of spring wheat (*Triticum aestivum* L.) were used to compose the crosses (Table 1). Each cross was represented by a different combination of varieties as given in Table 2. The field trial was performed in 1989 at one location: Prof. Broekemahoeve, CPRO-DLO, Lelystad.

The desirable number of lines in both generations was approximately 70. To exclude sampling errors from causing an additional bias to the estimates of the genetic parameters, the distribution of lines within crosses was kept as close as possible to the theoretical Mendelian segregation ratios. This resulted in 70 F_3 lines

Table 2. Composition of the pseudo crosses. The numbers of the varieties correspond to the numbering in Table 1.

	Genotype								
Cross	g_1 AABB	g_2 AaBB	g_3 aaBB	g_4 AABb	g_5 AaBb	g_6 aaBb	g_7 AAbb	g_8 Aabb	g_9 aabb
A	1	3	9	4	8	18	10	17	22
B	2	6	14	7	11	20	15	21	23
C	1	5	10	4	7	16	9	17	21
D	3	1	9	2	5	14	10	15	18
E	2	4	8	5	6	9	7	10	13
F	11	9	17	10	12	19	16	20	22
G	8	11	15	12	13	21	14	20	23
H	13	14	17	15	16	19	18	21	20
I	14	19	8	3	16	17	13	21	2
J	23	7	12	15	9	22	11	1	6
K	5	4	20	18	10	6	14	9	13
L	18	20	1	11	17	2	7	22	15
M	16	5	12	21	23	10	4	8	19
N	12	13	6	5	21	14	15	9	2
O	3	10	23	1	22	16	20	19	8

and 72 F_6 lines for each cross. Assuming a two locus model, the theoretical distribution of the F_3 lines is: (1:2:1:2:4:2:1:2:1)/16. This is closely fitted by (4:9:4:9:18:9:4:9:4)/70. For the 72 F_6 lines the theoretical distribution is (225:30:225:30:4:30:225:30:225)/1024 and the closest fitting distribution is (16:2:16:2:0:2:16:2:16)/72. The segregating ratios within lines all follow exactly the expected Mendelian frequencies.

The lines of one cross were grouped in a superplot. Each superplot was grown in two replicates. The F_3 and F_6 populations were grown in the same trial, so as to keep the growing conditions as similar as possible. As an extra check, four superplots were added in both replications, which contained all varieties in three neighboring plots. This was done in order to obtain an estimate of the variety yield on small plots, without the effect of competition, by using only the data of the central plots. So each replicate contained 15 F_3 populations, 15 F_6 populations and 4 variety-blocks, 34 superplots in all. Crosses were completely randomized among superplots, as well as lines were randomized within superplots. All plots were sown as 3.0 m long fields, with no spacing at the front and the rear. They consisted of three rows, 17.5 cm apart and the between-plot distance was 40 cm. This resulted in an effective plot width of 75 cm. Sowing density was 200 seeds m^{-2}. Shortly before harvest the front and rear 60 cm of each plot were mowed. This resulted in an average net plot length of 1.8 m, but varied significantly per plot. The exact length of all plots was measured.

The monoculture experiment consisted of eight replicates of all 23 varieties in a completely randomized block design. The size of these plots was 8.5 m long and 1.5 m wide and they were trimmed back shortly before harvest to 7.0 m long. They consisted of 10 rows, 12.5 cm apart and 37.5 cm between neighbouring plots. Sowing density was again 200 seeds m^{-2}.

Each field was simultaneously mowed and threshed with a combine harvester. After drying the total grain weight of each plot was measured and corrected for plot length. Samples were taken to determine the moisture content and all weights were converted to 0% moisture yields in kg per ha.

Analysis of the data

Pseudo-line experiment

The following statistical model can be applied to the pseudo-line trial:

$$\underline{y}_{ijk} = m + a_i + \underline{r}_j + \underline{s}_{ij} + g_{k(i)} + \underline{e}_{ijk},$$

with $\underline{y}_{ijk}$ = realized yield of line k of cross i in replicate j

m = over-all mean

a_i = fixed effect of cross i $\quad i_{F3}=1,2,..,15 \quad i_{F6}=16,17,..,30 \quad \Sigma_i a_i = 0$

$\underline{r}_j$ = effect of replicate j $\quad j=1,2 \quad \underline{r}_j \simeq N(0,\sigma_r^2)$

$\underline{s}_{ij}$ = effect of superplot ij $\quad \underline{s}_{ij} \simeq N(0,\sigma_s^2)$

$g_{k(i)}$ = fixed effect of line k of cross i $\quad k_{F3}=1,2,..,70 \quad k_{F6}=1,2,..,72 \quad \Sigma_k g_{k(i)} = 0$

$\underline{e}_{ijk}$ = residual effect of plot ijk $\quad \underline{e}_{ijk} \simeq N(0,\sigma_e^2)$

(random effects are underlined; all effects are assumed mutually stochastic independent)

F_3 and F_6 crosses were placed in the same model to keep the effects of residuals, superplots and replicates similar.

The F_3 and F_6 means, as predictions of the cross mean in the final breeding stage (F_∞), were calculated as common means over all 140 resp. 144 plots of each cross. The error variances of these means are:

$$\hat{\text{var}}\ \bar{y}_{i..F3} = \frac{1}{2}s_r^2 + \frac{1}{2}s_s^2 + \frac{1}{140}s_e^2$$

$$\hat{\text{var}}\ \bar{y}_{i..F6} = \frac{1}{2}s_r^2 + \frac{1}{2}s_s^2 + \frac{1}{144}s_e^2$$

The between-line variance is calculated for each cross separately by:

$$\sigma^2_{BLi} = \frac{MSBL_i - MSE}{r} \cdot \frac{l-1}{l}$$

When absence of epistasis and linkage is assumed, the between-line variance, which will be indicated as V_1, can be defined as (Mather & Jinks, 1971):

$$V_{1F3i} = \frac{1}{2}D + \frac{1}{16}H \quad \text{and}$$

$$V_{1F6i} = \frac{15}{16}D + \frac{15}{1024}H\ ,$$

where D is the additive and H the dominance component of the genetic variance. Van Ooijen (1989c) found that D can be most accurately estimated by ignoring the dominance components. In this way the best estimators for D are:

$$D_{F3psi} \approx 2V_{1F3i} \quad \text{and}$$

$$D_{F6psi} \approx \frac{16}{15}V_{1F6i}\ .$$

Assuming a normal distribution of the random effects, the mean squares are χ^2 distributed and can be estimated according to Van Ooijen (1989b).

Monoculture experiment

A straightforward two-way classification model is applicable to the monoculture trial. The variety effect is assumed to be fixed, whereas both block and residual effects are random. The variance components of blocks (σ_b^2) and residuals (σ_e^2) are estimated from the ANOVA (Table 4). All effects are assumed mutually stochastic independent. The mean for each variety is calculated as the common mean over all 8 blocks. The error variance of these means is estimated by:

$$\hat{\mathrm{var}}\ \bar{y}_{i.} = \frac{1}{8}s_b^2 + \frac{1}{8}s_e^2 = s_v^2$$

The same model applies to the variety plots in the pseudo-line trial.

The average for each cross was calculated directly from the frequencies of the different varieties (= genotypes; the indices of the genotypes refer to the ordering in Table 2) used in that specific cross. For each F_3 population this amounts to:

$$m_{F3mo} = \frac{1}{70}\{(4\cdot 1 + 9\cdot\frac{1}{2} + 18\cdot\frac{1}{16})\cdot\sum_{i=1,3,7,9} g_i + (9\cdot\frac{1}{2} + 18\cdot\frac{1}{8})\cdot\sum_{i=2,4,6,8} g_i + 18\cdot\frac{1}{4}g_5\} =$$

$$\frac{1}{560}\{77\cdot\sum_{i=1,3,7,9} g_i + 54\cdot\sum_{i=2,4,6,8} g_i + 36\cdot g_5\}$$, with error variance:

$$\hat{\mathrm{var}}\ m_{F3mo} = \frac{1}{560^2}\{77^2\cdot 4 + 54^2\cdot 4 + 36^2\}\cdot s_v^2 \approx 0.117\cdot s_v^2$$

For each F_6 population this amounts in a similar way to:

$$m_{F6mo} = \frac{1}{72}\{17\cdot\sum_{i=1,3,7,9} g_i + \sum_{i=2,4,6,8} g_i\}$$, with error variance:

$$\hat{\mathrm{var}}\ m_{F6mo} = \frac{1}{72^2}\{17^2\cdot 4 + 4\}\cdot s_v^2 \approx 0.224\cdot s_v^2$$

The mean of the F_∞-generation, in which the mere four homozygote genotypes are left over, is estimated by:

$$m_{F\infty mo} = \frac{1}{4}\cdot\sum_{i=1,3,7,9} g_i$$, with error variance:

$$\hat{\mathrm{var}}\ m_{F\infty mo} = \frac{1}{4}s_v^2 \ .$$

The between-line variances within each cross, as a measure for the additive variance component, can also be directly estimated. For the F_∞ this works out as follows:

$$\sigma_{BLF\infty}^2 = \frac{1}{4}\cdot\sum_{i=1,3,7,9} g_i^2 - (\frac{1}{4}\cdot\sum_{i=1,3,7,9} g_i)^2 = \frac{3}{16}\cdot\sum_{i=1,3,7,9} g_i^2 - \frac{1}{8}(g_1g_3 + g_1g_7 +$$

$$g_1g_9 + g_3g_7 + g_3g_9 + g_7g_9)$$

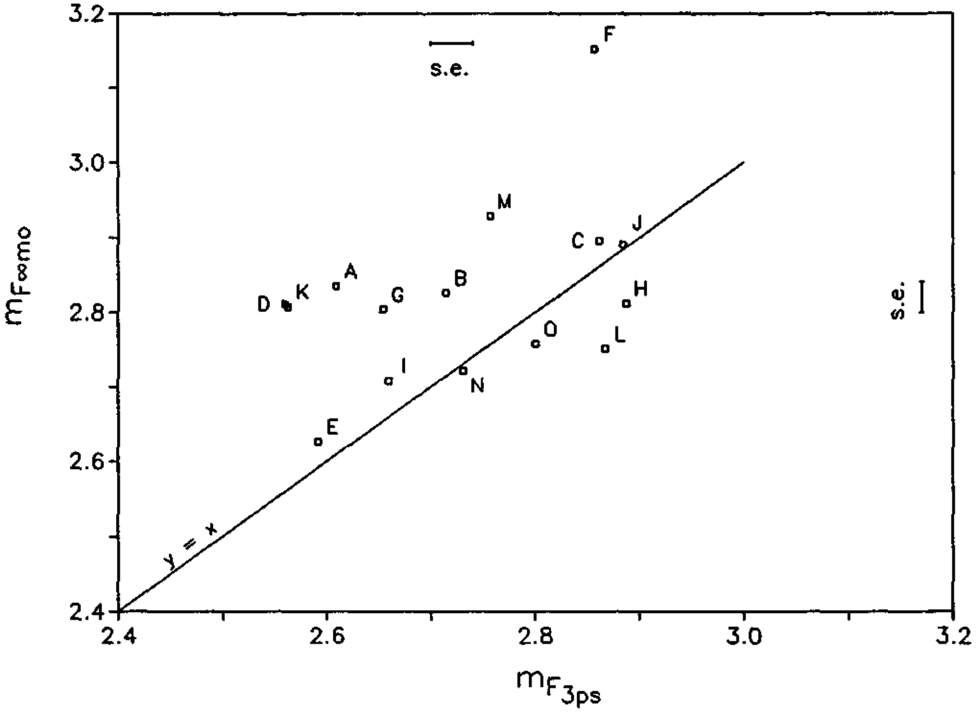

Fig. 1. Scatter diagram of the prediction of the F_∞-mean from the monoculture data against the estimate of the F_3 mean from the pseudo-line experiment for the crosses A to O. Yields in 10^3 kg·ha^{-1}. The standard errors of both parameters are given in the error bars on the side.

For the F_3 and F_6 generations this will lead in a similar way to :

$$\begin{aligned}\sigma^2_{BLF3} = \frac{1}{560^2}\{&17346\cdot\sum_{i=1,3,7,9} g_i^2 + 8424\cdot\sum_{i=2,4,6,8} g_i^2 + 3744\cdot g_5^2 - 6188\cdot(g_1g_3 + \\ &g_1g_7 + g_3g_9 + g_7g_9) - 11228\cdot(g_1g_9 + g_3g_7) - 3312\cdot(g_2g_4 + g_2g_6 + g_2g_8 + g_4g_6 + \\ &g_4g_8 + g_6g_8) + 3024\cdot(g_1g_2 + g_1g_4 + g_2g_3 + g_3g_6 + g_4g_7 + g_6g_9 + g_7g_8 + g_8g_9) - \\ &7056\cdot(g_1g_6 + g_1g_8 + g_2g_7 + g_2g_9 + g_3g_4 + g_3g_8 + g_4g_9 + g_6g_7) - \\ &3024\cdot\sum_{i=1,3,7,9} g_ig_5 + 1152\cdot\sum_{i=2,4,6,8} g_ig_5\}\end{aligned}$$ and

$$\begin{aligned}\sigma^2_{BLF6} = \frac{1}{72^2}\{&881\cdot\sum_{i=1,3,7,9} g_i^2 + 35\cdot\sum_{i=2,4,6,8} g_i^2 - 560\cdot(g_1g_3 + g_1g_7 + g_3g_9 + \\ &g_7g_9) - 578\cdot(g_1g_9 + g_3g_7) - 2\cdot(g_2g_4 + g_2g_6 + g_2g_8 + g_4g_6 + g_4g_8 + g_6g_8) + \\ &2\cdot(g_1g_2 + g_1g_4 + g_2g_3 + g_3g_6 + g_4g_7 + g_6g_9 + g_7g_8 + g_8g_9) - 34\cdot(g_1g_6 + g_1g_8 + \\ &g_2g_7 + g_2g_9 + g_3g_4 + g_3g_8 + g_4g_9 + g_6g_7)\}.\end{aligned}$$

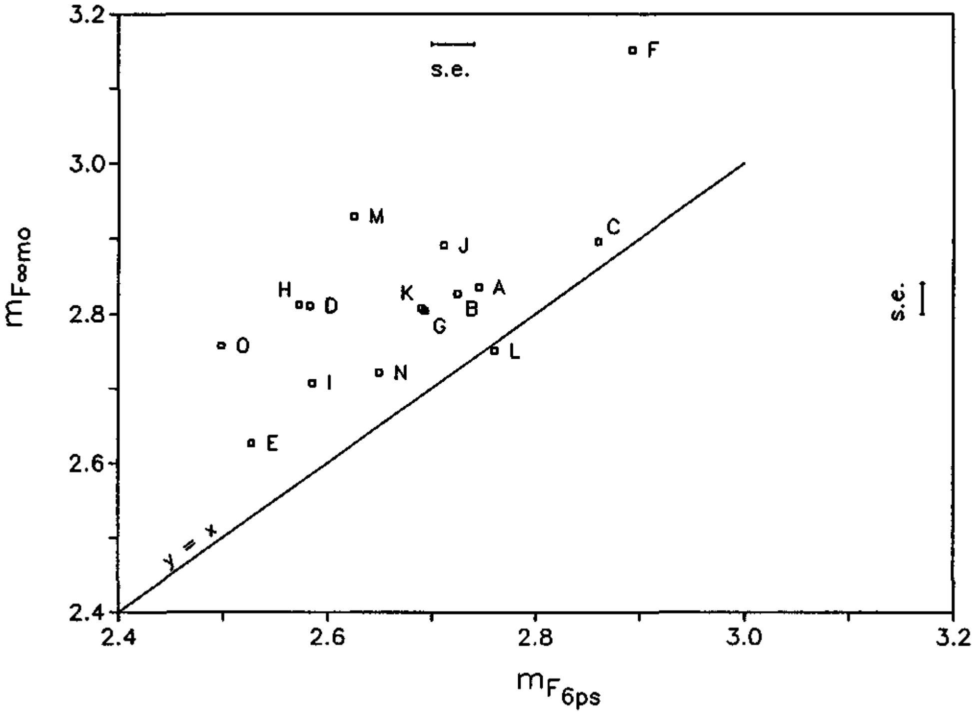

Fig. 2. Scatter diagram of the prediction of the F_∞-mean from the monoculture data against the estimate of the F_6 mean from the pseudo-line experiment for the crosses A to O. Yields in 10^3 kg·ha^{-1}. The standard errors of both parameters are given in the error bars on the side.

By using values from the monoculture experiment, the yield of genotype g_i is estimated as $v_i + \underline{e}$. The expectation of $\underline{e}$ equals zero and its variance is s_v^2. This has no effect on the given means but has a definite effect of overestimation on the between-line variance. The true estimators for the between-line variances are therefore given by:

$$V_{1F3} = \sigma^2_{BLF3} - \frac{13353}{39200}\cdot s_v^2 = \frac{1}{2}\cdot D_{F3mo} \ ,$$

$$V_{1F6} = \sigma^2_{BLF6} - \frac{229}{324}\cdot s_v^2 = \frac{15}{16}\cdot D_{F6mo} \quad \text{and}$$

$$V_{F\infty} = \sigma^2_{BLF\infty} - \frac{3}{4}\cdot s_v^2 = D_{F\infty mo} \ .$$

A comparison was made between the parameters directly estimated from the pseudo-line experiment (m_{F3ps}, m_{F6ps}, D_{F3ps} and D_{F6ps}) on the one hand and the parameters predicted from the monoculture data (m_{F3mo}, m_{F6mo}, $m_{F\infty mo}$, D_{F3mo},

D_{F6mo} and $D_{F\infty mo}$) on the other hand. From these means and the corresponding standard deviations, the probability of a superior inbred line in each cross was calculated. It was assumed that the distribution of the inbred lines in the F_∞ is normal with mean m and variance D. Superior lines are considered to be those lines with an average yield at least equal to a certain threshold value T, for which the highest yielding parent was taken.

$$P_i = P(m_i + \sqrt{D_i} \cdot \chi > T), \qquad \text{with } \chi \simeq N(0,1).$$

The probabilities and the resulting ranking of the crosses for F_{3ps} and F_{6ps} are compared to the $F_{\infty mo}$-data. Based on the P_{F3ps} the best three crosses were selected.

Furthermore an evaluation is made between the two selection procedures. From the F_6 population a selection of the best lines is made, taking no regard of their pedigree. The selected populations will be evaluated in later generations, using larger yield trials. From the results of 1989 an early comparison can be made between both selection procedures.

Results

The analysis of variance for the pseudo-line test and the monoculture experiment are presented in Tables 3 and 4. Table 1 shows the average variety yields calculated both from the monoculture trial and from the variety plots in the pseudo-line experiment. The correlation between the large and small plots is quite high (0.81), but obviously plot size has a definite effect. Remarkable is the higher level of the monoculture yields, even though sowing density was equal in both trials.

Table 3. Analysis of variance of the pseudoline experiment.

Source of variation	d.f.	Mean Square	F prob	Expected Mean Square
Replicates	1	$0.141 \cdot 10^6$ (MSR)		$\sigma_e^2 + l \cdot \sigma_s^2 + l \cdot c \cdot \sigma_r^2$
Crosses	29	$2.046 \cdot 10^6$ (MSBF)	<0.001	$\sigma_e^2 + l \cdot \sigma_s^2 + \frac{l \cdot r}{c-1} \cdot \sum_{i=1}^{c} a_i^2$
Superplots (Res 1)	29	$0.341 \cdot 10^6$ (MSS)	<0.005	$\sigma_e^2 + l \cdot \sigma_s^2$
Lines within crosses	2100	$0.261 \cdot 10^6$ ($MSBL_i$)	<0.001	$\sigma_e^2 + \frac{r}{l-1} \cdot \sum_{k=1}^{l} g_{ki}^2$
Plots (Res 2)	2100	$181.1 \cdot 10^3$ (MSE)		σ_e^2

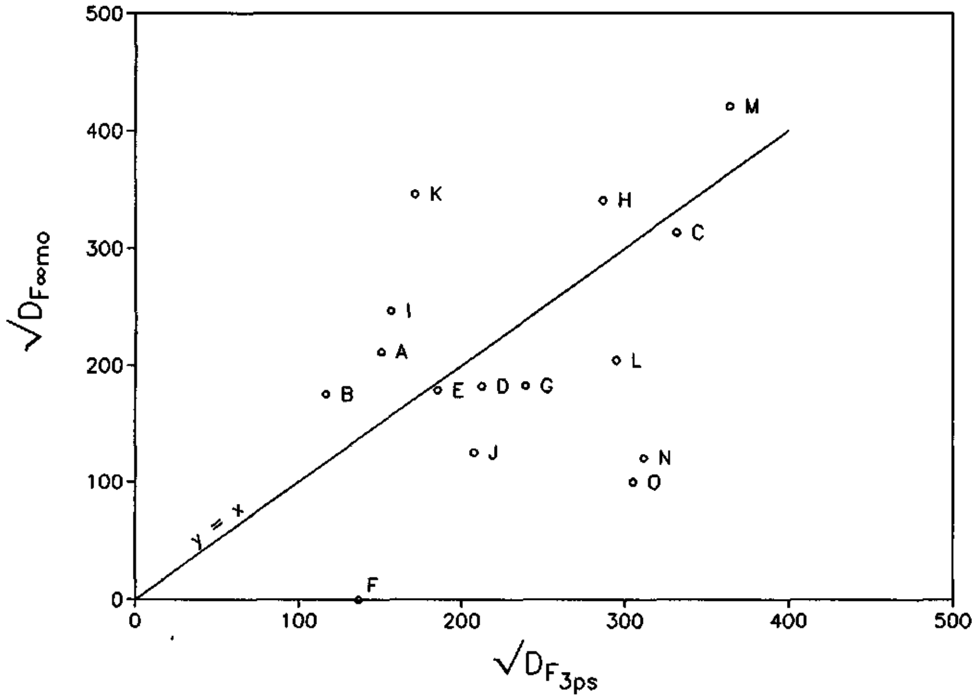

Fig. 3. Scatter diagram of the prediction of the genotypic standard deviation in the F_∞ from the monoculture data against the estimates of the F_3 standard deviation in the pseudo-line experiment for the crosses A to O. (kg·ha^{-1}).

This effect is probably due to the better conditions during germination in the monoculture experiment.

Correlation coefficients between the F_3 and F_6 generation means for each cross, both estimated and predicted values, as well as the predicted F_∞-mean are given in Table 5.The most important comparisons are those between the estimated means from the pseudo-line experiment (m_{F3ps} and m_{F6ps}) on the one hand and the predicted value for the F_∞ ($m_{F\infty mo}$) on the other. They are plotted in Figs. 1 and 2. M_{F3ps} and $m_{F\infty mo}$ show a very poor correlation; a somewhat better fit is found between m_{F6ps} and $m_{F\infty mo}$. This is not surprising because both effects of dominance and

Table 4. Analysis of variance of the monoculture experiment.

Source of variation	d.f.	MS	F prob	EMS
Blocks	7	$4.68 \cdot 10^5$	«0.001	$\sigma_e^2 + 23 \cdot \sigma_b^2$
Varieties	22	$7.90 \cdot 10^5$	«0.001	$\sigma_e^2 + {}^8/_{22} \cdot \Sigma v_i^2$
Res. error	154	34586		σ_e^2

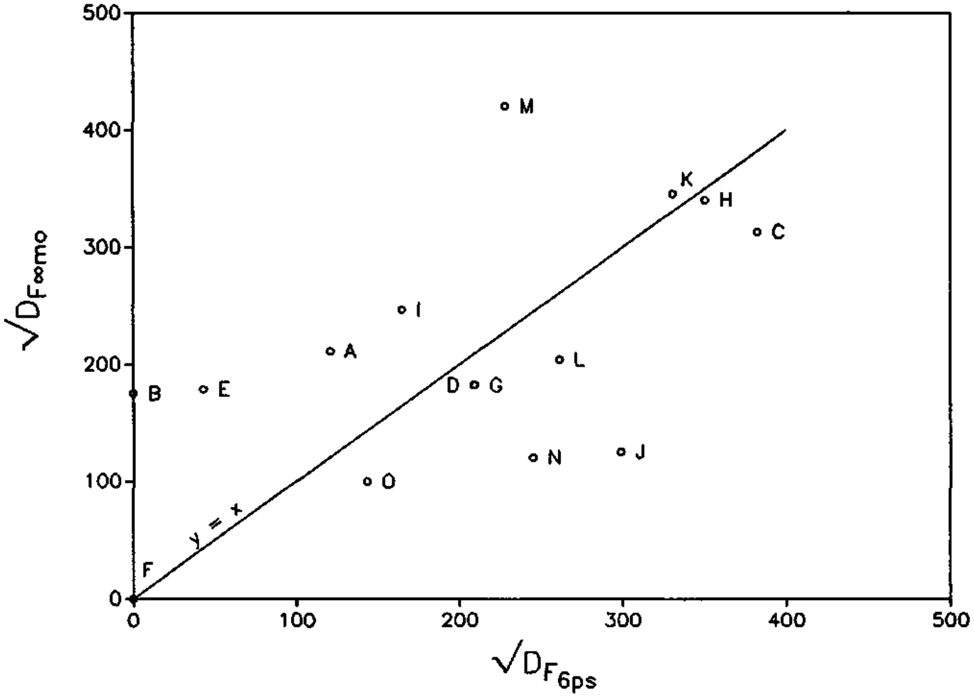

Fig. 4. Scatter diagram of the prediction of the genotypic standard deviation in the F_∞ from the monoculture data against the estimates of the F_6 standard deviation in the pseudo-line experiment for the crosses A to O. (kg·ha^{-1}).

within-plot competition, which are completely absent in the F_∞, are much smaller in the F_6 than in the F_3. Still, the effect of competition between F_6 plots can be substantial, because the between-line variance in the F_6 is relatively high. When we consider the correlation between the respective predicted values from the monoculture test, $m_{F\infty mo}$ and m_{F6mo} show a very good fit, but a much smaller correlation is found between the former and m_{F3mo}. This indicates a definite effect of dominance and/or epistasis. Van Ooijen (1989a) did not detect a great dissimilarity between $m_{F\infty mo}$ and m_{F3mo}. Obviously the importance of dominance depends on the composition of the simulated crosses.

The same trend, albeit to a lesser extent, is found from the relation

Table 5. Coefficients of correlation between estimated and predicted means.

	m_{F6ps}	m_{F3mo}	m_{F6mo}	$m_{F\infty mo}$
m_{F3ps}	0.35	0.62	-	0.43
m_{F6ps}		-	0.68	0.69
m_{F3mo}			0.77	0.73
m_{F6mo}				1.00

between the additive variance components. The correlation coefficients between the square root of D for all five situations are presented in Table 6. It is obvious that the connection between estimated values from the pseudo-crosses and the predictions from the monoculture trial is even worse than that for the means. Especially $\sqrt{D_{F3ps}}$ and $\sqrt{D_{F3mo}}$ show hardly any relationship at all. The higher correlation between $\sqrt{D_{F3ps}}$ and $\sqrt{D_{F\infty mo}}$ probably is a mere coincidence. Apparently the effects of competition and dominance and/or epistasis have a larger impact on the variance than they do on the mean. In Figs. 3 and 4 $\sqrt{D_{F\infty mo}}$ is plotted against $\sqrt{D_{F3ps}}$ and $\sqrt{D_{F6ps}}$ respectively.

Table 6. Correlation coefficients between estimates and predictions of the genotypic standard deviation.

	$\sqrt{D_{F6ps}}$	$\sqrt{D_{F3mo}}$	$\sqrt{D_{F6mo}}$	$\sqrt{D_{F\infty mo}}$
$\sqrt{D_{F3ps}}$	0.61	0.27	-	0.38
$\sqrt{D_{F6ps}}$		-	0.55	0.53
$\sqrt{D_{F3mo}}$			0.82	0.76
$\sqrt{D_{F6mo}}$				0.99

For the three situations F_{3ps}, F_{6ps} and $F_{\infty mo}$ the probabilities of superior inbred lines and their ranking are given in Table 7. Only one of the three best crosses, indicated by the $P_{F\infty mo}$, is actually identified as being most promising according to the estimated data from the pseudo-F_3. The best three crosses of F_{6ps} and those of $F_{\infty mo}$ have two in common. The potentially best cross (M) is not selected on grounds of the F_3 data, nor would it have been selected if cross selection was

Table 7. X-values, calculated as $(T-m)/\sqrt{D}$ for each cross, the corresponding probability of finding superior inbred lines in the F_∞ and their ranks.

	F_{3ps}			F_{6ps}			$F_{\infty mo}$		
cross	x	$P(\chi \geq x)$	rank	x	$P(\chi \geq x)$	rank	x	$P(\chi \geq x)$	rank
A	4.25	<.0001	14	4.16	<.0001	11	2.95	0.0016	5
B	4.58	<.0001	15	∞	0	14.5	3.62	0.0002	10
C	1.17	0.1210	1	1.02	0.1539	1	1.79	0.0367	2
D	3.25	0.0006	10	3.19	0.0007	9	3.57	0.0002	8
E	3.55	0.0002	11	16.79	<.0001	13	4.68	<.0001	12
F	2.86	0.0021	9	∞	0	14.5	∞	0	15
G	2.49	0.0064	8	2.65	0.0040	7	3.61	0.0002	9
H	1.27	0.1020	2	1.93	0.0268	5	1.89	0.0294	4
I	3.76	0.0001	12	4.02	<.0001	10	3.05	0.0011	6
J	1.76	0.0392	7	1.80	0.0359	3	4.61	<.0001	11
K	4.01	<.0001	13	1.69	0.0455	2	1.88	0.0301	3
L	1.29	0.0985	3	1.87	0.0307	4	3.47	0.0003	7
M	1.35	0.0885	4	2.76	0.0029	8	1.25	0.1056	1
N	1.66	0.0485	6	2.44	0.0073	6	6.23	<.0001	13
O	1.47	0.0708	5	5.24	<.0001	12	7.22	<.0001	14

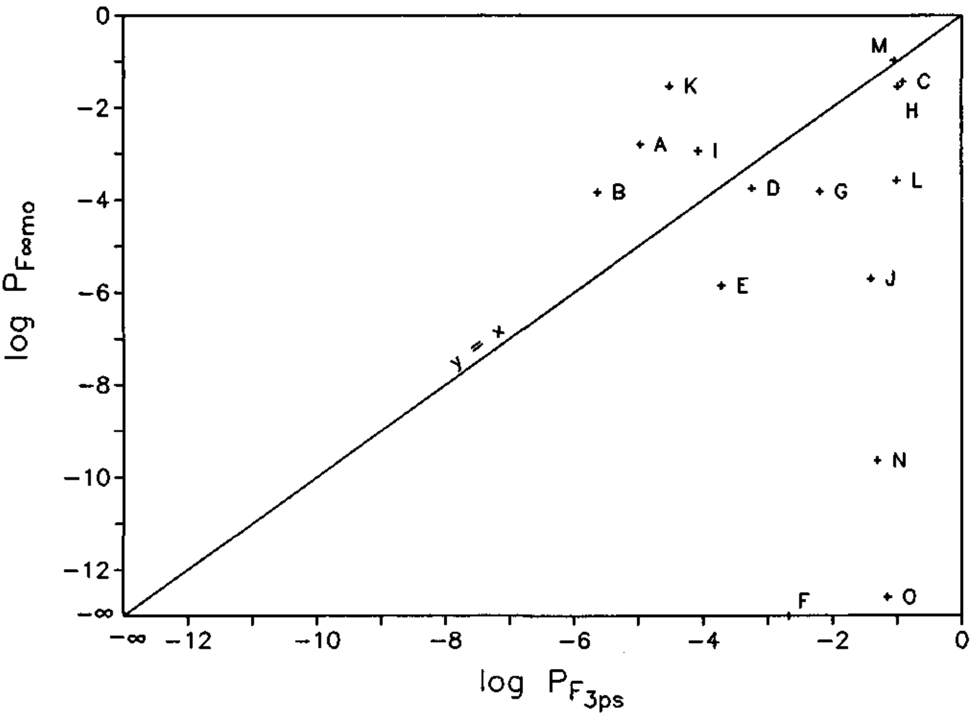

Fig. 5. Scatter diagram of the logarithm of the probability $P_{F\infty mo}$ against the logarithm of P_{F3ps} for the crosses A to O.

practised in the SSD-F_6.

The correlation between the estimated P_{F3ps} and the predicted $P_{F\infty mo}$ appears to be very low (Table 8) and not significant ($\alpha = 0.10$). The relationship between P_{F6ps} and $P_{F\infty mo}$ shows a better fit but is still very poor. This is not surprising, considering the low correlation between the parameters from which it is derived. For both cases this is illustrated in Figs. 5 and 6. Evidently, cross prediction in the F_3 can lead to very erroneous results.

Furthermore, a comparison between the two selection procedures is made. The best F_6 SSD-lines are compared to the best lines of the three F_3 crosses, selected according to the P_{F3ps}. The average

Table 8. Spearman rank correlation coefficients between estimates and predictions of the probability of finding superior inbred lines.

	P_{F6ps}	P_{F3mo}	P_{F6mo}	$P_{F\infty mo}$
P_{F3ps}	0.54	0.30	-	0.18
P_{F6ps}		-	0.44	0.53
P_{F3mo}			0.62	0.54
P_{F6mo}				0.97

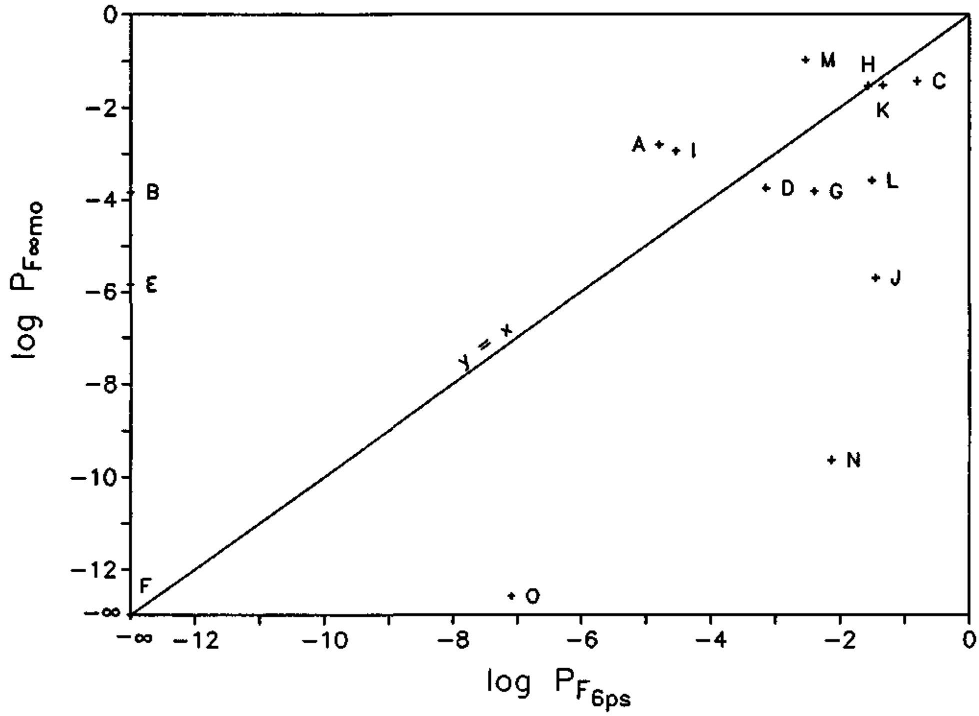

Fig. 6. Scatter diagram of the logarithm of the probability $P_{F\infty mo}$ against the logarithm of P_{F6ps} for the crosses A to O.

of the 96 best SSD-lines equals 3455 kg·ha^{-1}. The best 32 lines in each of the three selected crosses (also 96 lines in all) yield an average of 3184 kg·ha^{-1}. Taking account of the overall means for both generations (2733 kg·ha^{-1} for F_{3ps} and 2675 kg·ha^{-1} for F_{6ps}), this amounts to a direct selection differential of 16.5% for ES and 29.2% for the SSD method.

Discussion and conclusions

The use of simulated crosses may rise some objections. First there is the question whether mixtures of pure inbred lines are representative for real breeding populations consisting of unselected heterozygous genotypes. In view of the aim of the pseudo-line approach, i.e. to identify the effects of F_3 (and F_6) lines being grown in small plots in a mixed stand, this question is only relevant if the reaction of real heterozygotes to being grown in mixture differs from the reaction of pure lines. As far as we know there is neither experimental evidence nor a theoretical basis to support the idea of such a differential reaction pattern.

Secondly, the underlying two locus model for a quantitative trait is an oversimplification. It should be realized however that its first aim is not to perfectly mimic a polygenic character. Together with the pseudo-line approach this model primarily serves as an instrument to identify specific sources of error in cross prediction. As such it proves to be a useful tool. Remains the question whether a simulated cross based on a two locus system will generate a genetic diversity similar to real crosses. This possible discrepancy between simulated and real crosses was checked by Van Ooijen (1989a,1989b), who found that the level and range of parameter estimates obtained from pseudo-crosses were very similar to those obtained from real crosses grown in the same trial field. This not only shows that a supposed two locus model is, for the present purpose, representative for real crosses, but also justifies the use of (selected) pure lines.

As was found by Van Ooijen (1989a; 1989b), the predictions and F_3 estimates show a poor correlation. This can be caused by three sources of bias:

1. plot size (environmental error),
2. intergenotypic competition and
3. non-additive genetic interaction (dominance and epistasis).

The effect of the first is shown by the correlation between F_3/F_6 and monoculture plots: $r=0.81$. The effect of dominance and epistasis is indicated by the relation between F_{3mo} and $F_{\infty mo}$: $r \approx 0.75$ for both m and $\sqrt{D}$. The relation between F_{3ps} and F_{3mo} includes both effects of differences in plot size and intergenotypic competition. If their correlation is regarded as the product of both separate effects, the mere effect of competition would lead to a correlation of 0.77 for m and 0.33 for $\sqrt{D}$. We conclude that competition causes a large portion of bias especially on D.

Under these circumstances the differences in growing conditions between the selection and the goal environment (intergenotypic competition and, to a lesser extent, plot size) as well as the level of dominance lead to an erroneous cross prediction. We have no reason to believe that these findings do not apply to a situation where a real polygenic model is involved.

The F_3/F_6 based comparison shows a preference to the SSD procedure, especially considering the amount of time and labour which is spent in the early selection method. However, this is not a equitable comparison, because the F_3 is still very heterogenic and will be segregating in subsequent generations. During these generations both plant and line selection can be exercised. In the F_6 generation a large field trial will be used to evaluate the better lines together with the

F_7 lines derived from the SSD-procedure. Only then a final comparison between the two selection procedures can be made.

Acknowledgements

I would like to thank Dr P. Stam and Dr J.W. van Ooijen for their advise during my research and their useful comments on the manuscript.

References

Casali, V.W.D. & E.C. Tigchelaar, 1975. Computer simulation studies comparing pedigree, bulk and single seed descent selection in self pollinated populations. J. Am. Soc. Hort. Sci. 100: 364-367.

Jinks, J.L. & H.S. Pooni, 1980. Comparing predictions of mean performance and environmental sensitivity of recombinant inbred lines based upon F_3 and triple test cross families. Heredity 45: 305-312.

Knott, D.R. & J. Kumar, 1975. Comparison of early generation yield testing and a single seed descent procedure in wheat breeding. Crop science 15: 295-299.

Kramer, Th., J.W. van Ooijen & C.J.T. Spitters, 1982. Selection for yield in small plots of spring wheat. Euphytica 31: 549-564.

Mather, K. & J.L. Jinks, 1971. Biometrical genetics. Chapman and Hall, London, 382 pp.

Ooijen, J.W. van, 1989a. The predictive value of quantitative genetic parameters in autogamous crops: bias caused by intergenotypic competition. 1. F_∞-mean. Euphytica 44: 81-94.

Ooijen, J.W. van, 1989b. The predictive value of quantitative genetic parameters in autogamous crops: bias caused by intergenotypic competition. 2. F_∞-variance. Euphytica 44: 95-108.

Ooijen, J.W. van, 1989c. Estimation of additive genotypic variance with the F_3 of autogamous crops. Heredity 63: 73-81.

Spitters, C.J.T., 1979. Competition and its consequences for selection in barley breeding. Pudoc, Wageningen, 268 pp.

3. A COMPARISON BETWEEN SINGLE SEED DESCENT AND EARLY CROSS SELECTION IN WHEAT BREEDING.

II. THE EVALUATIVE GENERATIONS

This chapter is in press for Euphytica

Summary

Two selection procedures in wheat breeding were compared on the basis of their ability to supply high yielding inbred lines. The first procedure consists of an early selection between crosses in the F_3 generation, based on cross performance. In the second procedure selection is postponed until the F_6, which is derived by single seed descent. The two procedures are evaluated in a two year test, using pseudo-lines of spring wheat. These pseudo-lines consist of mixtures of varieties and enable an estimation of the true genetic parameters. In this way the accuracy of the predictions can be examined.

The first year of this comparison comprised the actual selection process and in the second year an F_6 and F_7 generation were mimicked, based on the selected lines of the F_3 and F_6 respectively, and evaluated in a large yield test. This paper gives the final results after the second year. From the first year it was found that the F_3 cross predictions were very inaccurate, whereas selection in the SSD-F_6 appeared to perform well. This resulted in a higher yielding set of lines evolving from the SSD-F_6. Despite a pronounced genotype-year interaction this difference in yield level was also found between the F_7 and the F_6 lines, derived from the selected crosses in the F_3. We conclude that the early selection procedure is not advantageous compared to the fast SSD procedure.

Introduction

A previous paper (Van Oeveren, 1992) evaluated the first year of a trial, which was carried out to compare two different selection procedures in the breeding of small autogamous cereals. Grain yield in spring wheat was used as the quantitative character of interest.

The first procedure is referred to as the 'Early Selection' procedure (ES) (Jinks

& Pooni, 1980). In the F_3 generation a prediction is made, based on the estimation of the cross mean and the between-line variance. Based on the predicted probabilities of superior inbred lines in the F_∞, the most promising crosses are selected and propagated. In the subsequent generations pedigree selection can be applied.

A completely different method is to avoid selection in an early stage and wait until a high degree of homozygosity, and thus homogeneity within lines, has been reached. At this stage, selection between lines can be made, without regard to the pedigree. An example of this is the method of 'Single Seed Descent' (SSD). One great advantage of this technique is the possibility to grow three successive generations in one year (spring cereals).

By using so called pseudo-lines instead of real crosses, we could calculate predictions of the genetic parameters, in absence of intergenotypic competition. In accordance to earlier research (Van Ooijen, 1989a; 1989b) it was found that the cross predictions were very inaccurate. They were severely biased by environmental variation, intergenotypic competition and dominance, which led to an erroneous cross prediction (Van Oeveren, 1992).

In order to make a fair comparison between the two selection procedures, the selected crosses and lines were advanced to a generation with a high degree of homozygosity. From both procedures an equal number of selected lines was yield tested on a large scale. The comparison of the best lines from both procedures gives an indication of the best breeding strategy.

Materials and methods

Pseudo-lines were used to simulate 15 segregating crosses in both 1989 and 1990. This technique provides a way to create any desired breeding generation. Besides, it offers an opportunity for an estimation of the true genetic parameters. Basically this technique is as follows. It is assumed that yield, as an example of a quantitative trait, segregates in each cross at two unlinked loci A/a and B/b. For each cross this results in nine distinct genotypes, each of which is represented by a specific variety. So nine varieties are used to compose all lines in one cross. Segregating lines consist of mixtures of these varieties. Twenty-three varieties and pure breeding lines of spring wheat (*Triticum aestivum* L.) were used (Table 1). Each cross was represented by a different combination of varieties. The composition of the crosses is described in Table 2. Further details and rationale behind this technique can be found in Van Ooijen (1989a) and Van Oeveren

Table 1. Varieties and pure breeding lines used for the composition of the pseudo-lines, numbered according to their expected yield capacity (1 = high, 23 = low), and their mean yields from the monoculture experiments of both years 1989 and 1990 (10.5 m^2 plots) and the F_6/F_7-trial (6.0 m^2 plots) in kg·ha^{-1} and their ranks.

		MONO '89		MONO '90		F_6/F_7'90	
	Variety	yield	rank	yield	rank	yield	rank
1	Stratos	2687	15	5506	7	5152	10
2	G 74010	2806	13	5304	11	5254	8
3	vdH 1166-76	2668	18	5068	16		
4	vdH 1132	2385	22	5048	18		
5	ZESC 1963-6	3241	3	6002	1	5914	1
6	TK 2832 3,10	2892	11	5843	4		
7	vdH 3132	2683	16	4797	21	4655	17
8	G 8005	2715	14	5172	14	4905	13
9	Minaret	2860	12	4869	20	4906	12
10	Heros	2608	20	5059	17	4949	11
11	Darima	3045	8	5605	6	5580	3
12	TK 2832 2,11	2661	19	5695	5		
13	Ralle	2303	23	4592	22	4638	18
14	Sunnan	3008	9	5163	15	5291	7
15	Bastion	2527	21	5347	10	5518	5
16	Spartacus	3215	4	5476	8	5505	6
17	Adonis	3158	6	5283	13	4700	15
18	Melchior	3109	7	5960	2	5713	2
19	Axona	3456	1	5898	3		
20	Kokart	2680	17	5389	9	5564	4
21	Sicco	3429	2	5291	12	5229	9
22	Wembley	3189	5	4916	19	4854	14
23	Prinqual	2967	10	4484	23	4673	16
	mean:	2882		5294		5167	

(1992).

In 1989 both an F_3 and an F_6 generation was grown from each cross. The F_3 lines were simulated as F_2-derived lines and the F_6 as F_5-derived SSD lines. All lines were yield tested on equally sized plots of 1.4 m^2, in two replications. Early selection was applied to the F_3, selecting the three best crosses (C,H and L). Among the SSD-F_6 lines normal line selection was applied (Van Oeveren, 1992).

An F_6 generation was simulated as being derived from the ES-F_3. It was also composed of pseudo-lines and thus the F_4 and F_5 generation could be skipped; the F_6 was created as follows. As the character of interest is assumed to be controlled by only two loci, the maximum number of distinct genotypes appearing is nine. A common way to apply pedigree selection in the intermediate generations is to augment the selected F_3 lines to larger F_4 lines. Selection between and within

Table 2. Composition of the pseudo crosses. The numbers of the varieties correspond to the numbering in Table 1.

	genotype								
cross	AABB	AaBB	aaBB	AABb	AaBb	aaBb	AAbb	Aabb	aabb
A	1	3	9	4	8	18	10	17	22
B	2	6	14	7	11	20	15	21	23
C	1	5	10	4	7	16	9	17	21
D	3	1	9	2	5	14	10	15	18
E	2	4	8	5	6	9	7	10	13
F	11	9	17	10	12	19	16	20	22
G	8	11	15	12	13	21	14	20	23
H	13	14	17	15	16	19	18	21	20
I	14	19	8	3	16	17	13	21	2
J	23	7	12	15	9	22	11	1	6
K	5	4	20	18	10	6	14	9	13
L	18	20	1	11	17	2	7	22	15
M	16	5	12	21	23	10	4	8	19
N	12	13	6	5	21	14	15	9	2
O	3	10	23	1	22	16	20	19	8

these F_4 lines can then be applied and via an intermediate small F_5 line an F_6 line can be derived from every selected F_4 plant. Considering a situation where no selection is applied, the probability of the double heterozygous genotype in the F_4 equals ${}^{4}/_{256}$. For each single heterozygote the probability is ${}^{14}/_{256}$ and for each fully homozygote it equals ${}^{49}/_{256}$ (see Fig. 1a). The most favourable situation in the F_6 is one where none of the genetic variance is lost. This is closely mimicked when ample F_6 lines are distributed according to their expected frequencies in case of no selection. As the chance of retrieving the double heterozygote is very small and the genotype is of no direct interest for selection, it was left out. When 32 lines of each selected cross are grown, the expected distribution will be closely fitted by 4 x 6 lines, derived from fully homozygous F_4 plants, and 4 x 2 lines, derived from single heterozygous F_4 plants (Fig. 1b). The latter are expected to show a (3:2:3)/8 segregation in genotypes and hence were composed in accordance with these ratios. So although normally line selection is expected in the F_3, F_4 and F_5 of the ES procedure, it was not performed in this trial.

From the 96 phenotypic best SSD-F_6 lines of the 1989 trial an F_7 generation, on the same scale as the ES-F_6, was simulated using pseudo-lines. The F_7 consisted of 84 fully homogeneous lines and 12 segregating lines, which were all composed as F_5-derived lines. The segregating lines were therefore also composed according to a (3:2:3)/8 ratio.

From both procedures the 96 lines were yield tested in 1990 at one location: Prof. Broekemahoeve, Lelystad (CPRO-DLO). This was done on 6 m^2 plots in four replicates. An $\alpha(0,1)$-design was used to minimize the effects of field heterogeneity (Patterson et al., 1978). Each replicate consisted of 12 blocks of size 8. All plots were sown as 5.5 m long and 1.5 m broad fields, without spacing on the front and the rear. They consisted of 10 rows with a row distance of 12.8 cm and a between-plot distance of 34.8 cm. Sowing density was 300 seeds m^{-2}. Shortly before harvest the front and rear 75 cm of each plot were mowed. The data were analysed according to their balanced incomplete block design using the REML procedure in GENSTAT 5, release 2.1.

	AA	Aa	aa			
BB	49	14	49	112		14
Bb	14	4	14	32	=	4
bb	49	14	49	112		14
				256		32

a

	AA	Aa	aa	
BB	6	2	6	14
Bb	2	0	2	4
bb	6	2	6	14
				32

b

Fig. 1a,b. Distribution of F_4 genotypes. *a.* Expected distribution. *b.* Assumed distribution, represented in 32 pseudo-F_6 lines.

An additional trial was laid out to obtain accurate estimates of the variety yields in monoculture and this will be referred to as the 'monoculture experiment'. Eight replicates of all 23 varieties were grown in a completely randomized block design. The plots were 8.5 m long and 1.5 m wide and they were trimmed back shortly before harvest to 7.0 m long. They consisted of 10 rows with a between-row distance of 12.8 cm and a between-plot distance of 34.8 cm. Sowing density was again 300 seeds m^{-2}. Each field was mowed and threshed with a combine harvester. After drying, the total grain yield of each plot was measured. Samples were taken to determine the dry matter content and all data were converted to 0% moisture yields in $kg{\cdot}ha^{-1}$. A straightforward two-way classification model was used for the statistical analysis.

With the aid of the monoculture data the true potential yield of all completely

Table 3. Analysis of variance of yield of the F_6/F_7-trial ($kg^2 \cdot ha^{-2}$).

		F_6		F_7	
Source of Variation	d.f.	MS	F prob	MS	F prob
Replicates	3	$6.66 \cdot 10^5$	<0.001	$19.48 \cdot 10^5$	<0.001
Blocks	44	$1.57 \cdot 10^5$	0.002	$1.56 \cdot 10^5$	0.001
Lines	95	$4.51 \cdot 10^5$	<0.001	$4.54 \cdot 10^5$	<0.001
Residual	241	86135		81396	

homogeneous lines, which are represented by just one variety, can be measured. In this way selection of the best lines can be checked on its accuracy. A final comparison between both selection procedures is made, based on the genotypic values of the 10 best lines in each procedure.

Results

The analysis of variance for balanced incomplete block designs for the F_6 and F_7 is presented in Table 3. There is a highly significant effect of lines for both cases. The adjusted line means are calculated from this analysis.

Out of the 192 F_6/F_7 lines, 156 are fully homogeneous and thus consist of only one variety. In this way yields can be estimated from the F_6/F_7 trial for a subset of the varieties (i.e. the selected ones). An exact estimate for each variety yield is obtained from the monoculture experiment. Values from both experiments, together with the monoculture yields of the previous year (1989), are given in Table 1 and plotted in Fig. 2. There appears to be a very high variety-year interaction. The correlation coefficient between the variety yields of both years is low: 0.44. The variety ranking, calculated from the F_6/F_7 trial is very similar to the one from the monoculture experiment. The correlation coefficient is high: 0.88. Obviously the size of the F_6/F_7 trial is adequate for the estimation of the line means.

The 156 selected, completely homogeneous lines from both selection procedures can be portrayed in a graph as the varieties they consist of. The ratio of the number of selected, homogeneous lines of a certain variety to the number of lines of that variety if selection would not have been applied is called the selected fraction. In case of SSD it is calculated as: the number of selected SSD lines consisting of the specific variety divided by the total number of SSD-F_6 lines of that variety. For instance, in the SSD-F_6 48 lines were present which consisted of

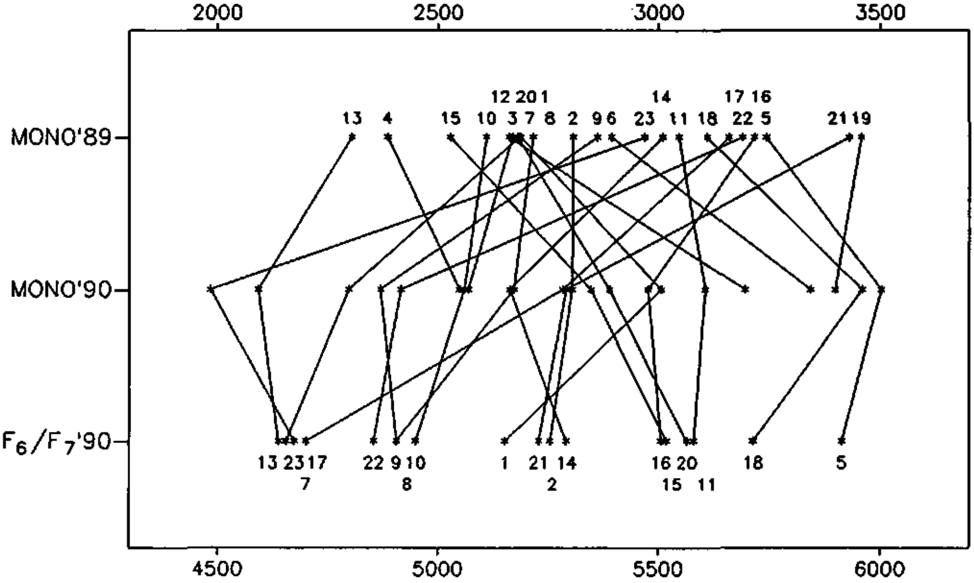

Fig. 2. Variety yields estimated from three experiments: the monoculture experiment in 1989, the monoculture experiment in 1990 and the F_6/F_7-trial. The varieties are indicated according to the numbering in Table 1. **N.B.**: The values for the MONO'89 experiment should be read from the upper scale, whereas the other two are connected to the lower scale (both in $kg \cdot ha^{-1}$).

variety 9 alone. Five of these were selected, so the selected fraction equals $^5/_{48}$. In case of ES it is calculated as: the number of ES-F_6 lines (only three crosses selected) consisting of that variety divided by the total number of lines of that variety which would been present if all 15 crosses had been grown and had been represented by the same number of lines. For instance, three of the 15 crosses would produce homogeneous lines of variety 9. One of these crosses is selected, so the selected fraction equals $^1/_3$. It should be noted that in this way the overall selection intensity is only 10% for SSD, whereas it is 20% for ES.

The selected fraction of each variety is plotted against the variety yield in Figs. 3a and 3b for the '89 and the '90 yields respectively. Fig. 3a shows that the SSD-F_7 lines cover indeed a great deal of the potentially best lines. Apart from variety number 19, which is emerging in only one cross, the best varieties are selected most frequently. The 3 selected crosses from the ES procedure though consist of high and low yielding varieties. Due to the large genotype-year interaction this picture becomes completely different when the variety yields of 1990 are used (Fig. 3b). The selected lines in both procedures appear a random sample of the

Table 4. The best 10 lines from both selection procedures with their phenotypic yield, line number, variety they are composed of and the 2-year mean monoculture yield of that variety as an indication for the genotypic value (kg·ha^{-1}).

	ES-F_6				SSD-F_7			
	yield	line	var	genotype	yield	line	var	genotype
1	5928	90	18	4534	6066	1	18	4534
2	5859	9	18	4534	6057	40	5	4622
3	5737	54	18/20*	4303	5995	29	5	4622
4	5736	26	20	4034	5985	19	18	4534
5	5712	3	18	4534	5975	53	5	4622
6	5703	19	18	4534	5956	59	18	4534
7	5701	49	18	4534	5863	23	5	4622
8	5674	4	18	4534	5856	63	18	4534
9	5670	20	20	4034	5801	25	5	4622
10	5640	93	15	3937	5797	87	11	4325
mean	5736			4351	5935			4557
mean of all 96 lines:			5130				5348	

* segregating line, consisting of $^3/_8$ of variety 18, $^3/_8$ of variety 20 and $^2/_8$ of variety 21.

range of varieties. Neither of the two selection methods appears better than the other. The best variety though (number 5), occurs only in the SSD-F_7.

The best 10 lines from both procedures, together with their 'pedigree' are given in Table 4. At first sight the best SSD lines seem definitely better than the best ES lines. But because these phenotypic values are influenced by environmental effects, it is best to look at their 'genotypic' values. These genotypic values can be directly estimated from the monoculture yields of the varieties they are composed of. The mean variety yield of both years was taken to exclude (a part of) the genotype-year interaction. It appears that the 10 best SSD lines consist of better varieties than the ES lines. Accordingly the mean genotypic value of the 10 best SSD lines is much higher than that of ES.

Conclusions and discussion

The use of pseudo-lines puts some restrictions on the genetic models that can be simulated and possibly on the conclusion when comparing the outcome of ES and SSD. Because a simple two-locus model was used as the genetic basis for a quantitative trait, only a very small number (4) of different, fully homozygous lines can be formed in the late breeding generations. Therefore, the chance of a certain

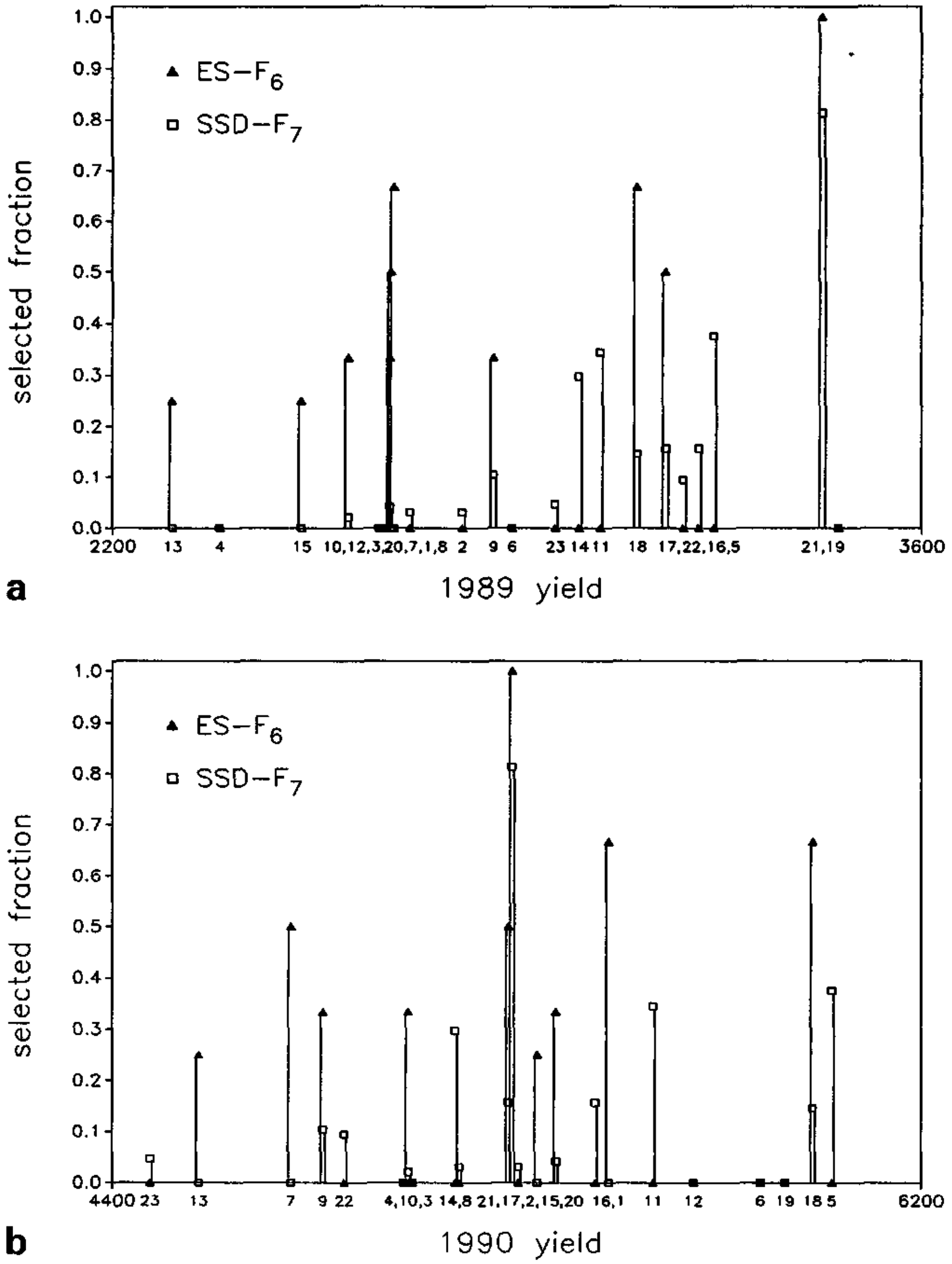

Fig. 3a,b. The selected fraction of completely homogeneous lines, expressed as the varieties they are composed of, in both the ES-F_6 and the SSD-F_7. The varieties are plotted according to their monoculture yield in 1989 (*a*) and 1990 (*b*) and indicated with their corresponding number from Table 1.

genotype being lost is accordingly very low and the potentially best genotype of a selected cross will always be retrieved (see the above results). This holds for both ES and SSD.

For a given cross, the probability of a specific, fully homozygous genotype in the F_5 equals $\frac{225}{1024}$ ($\approx$ 22%). When 10 segregating loci are involved, however, more than one thousand different homozygous genotypes can be formed and the probability of a specific one will be very small (0.05%). It can be questioned if, in that case, SSD will still perform best. Under polygenic circumstances there will probably be a positive effect of line selection (in addition to selection between crosses) in the ES-F_3, F_4 and F_5, whereas this effect is fully absent in the present situation. This could lead to a higher probability of well performing lines in the ES-F_5, compared to the balanced distribution of lines, which was used in this trial. Thus ES could have an advantage over SSD.

Two factors oppose on this matter: in the early generations, plots are small and the level of heterogeneity within plots is high. This can lead to serious bias on monoculture yield estimates, caused by intergenotypic competition, dominance effects and a relatively large portion of environmental error (Van Oeveren, 1992; Van Ooijen, 1989b). Secondly, genotype-year interaction can completely annul the results of selection or even lead to a negative selection result. This is clearly shown by Figs. 3a and 3b. It should be stated however that the 1989 trial had an extremely rough growing season, reflected in a low yield level (Table 1). This may have caused the genotype-year interaction to be large. Possibly an alternative trial would show a higher correlation between variety yields of two succeeding years and thus a higher selection response.

The above results indicate that under these circumstances early selection does not pay off. These findings are in accordance with another field trial concerning a comparison between SSD and early yield testing (no cross selection) in wheat (Knott & Kumar, 1975). It is possible that in other cases early selection and subsequent line selection will lead to better results. It is likely though that comparable selection results will be obtained from the speedy and accurate SSD method.

Acknowledgements

I would like to thank ir Henk Holthof for his help on the α-design. Dr Piet Stam is thanked for his advice and comments on the manuscript.

References

Jinks, J.L. & H.S. Pooni, 1980. Comparing predictions of mean performance and environmental sensitivity of recombinant inbred lines based upon F_3 and triple test cross families. Heredity 45: 305-312.

Knott, D.R. & J. Kumar, 1975. Comparison of early generation yield testing and a single seed descent procedure in wheat breeding. Crop science 15: 295-299.

Oeveren, A.J. van, 1992. A comparison between single seed descent and early cross selection in wheat breeding. Euphytica 58: 275-287.

Ooijen, J.W. van, 1989a. The predictive value of quantitative genetic parameters in autogamous crops: bias caused by intergenotypic competition. 1. F_∞-mean. Euphytica 44: 81-94.

Ooijen, J.W. van, 1989b. The predictive value of quantitative genetic parameters in autogamous crops: bias caused by intergenotypic competition. 2. F_∞-variance. Euphytica 44: 95-108.

Patterson, H.D., E.R. Williams & E.A. Hunter, 1978. Block design for variety trials. J. Agric. Sci. 90: 395-400.

4. A COMPARISON BETWEEN SINGLE SEED DESCENT AND EARLY CROSS SELECTION IN WHEAT BREEDING.

III. A 3-LOCUS MODEL

Summary

Two selection procedures in wheat breeding were compared on the basis of their ability to supply high yielding inbred lines. The first procedure consists of an early selection between crosses in the F_3 generation, based on predictions of the cross mean and the between line variance. In the second procedure selection is postponed until the F_6, which is derived by single seed descent. The two procedures were evaluated in a two year test, using pseudo-lines of spring wheat. These pseudo-lines consist of mixtures of varieties, mimicking segregating generations, and enable estimation of the true genetic parameters. In this way the accuracy of the predictions can be examined.

In case of early selection, it appears that the predictions of the cross mean and especially the between line variance are inaccurate. This is caused mainly by intergenotypic competition but also by the effects of plot size and dominance and/or epistasis. Bias due to line sampling is of no importance. Contrary to the results of a former field test, the ranking of the crosses according to their F_3 predictions is almost similar to their real potential. The SSD-F_6 line estimates are even more accurate and thus the better lines are indeed selected. When the selected lines are evaluated in the second year a large genotype-year interaction completely annuls the effects of selection. Nevertheless, the SSD lines yield higher than the ES lines and we conclude that the effort spent on early cross prediction does not pay off.

Introduction

Various methods are used to obtain high performing inbred lines in the breeding of self-pollinating crops. Selection can be applied in different stages of the breeding process and with varying intensities. In order to get a better view on the efficiency of different selection procedures two contrasting selection methods are examined.

The first is a method, based on a quantitative genetic theory, which has been developed to predict the genetic potential of a certain cross in an early breeding generation (Mather & Jinks, 1971). With this technique, which uses estimates of the mean (m) and additive genetic variance (D) of a cross, it should be possible to identify the most promising crosses in an early generation. Only those crosses are retained and in the subsequent generations line selection can be performed. This procedure will be referred to as 'Early Selection' (ES).

From experimental studies on Early Selection (Jinks & Pooni, 1980; Van Ooijen, 1989a, 1989b) it was found that estimating the additive variance (D) by using only F_3 lines is by far the most practical application of this procedure. Estimates based on F_3 lines however may be biased by the effects of intergenotypic competition. Especially in small grain crops this bias can be large, because plots are small and the level of heterogeneity within plots is high. Van Ooijen (1989a, 1989b) concluded from his studies with mixtures of spring wheat varieties that the estimates of the genetic parameters for yield are indeed severely biased, leading to unreliable cross predictions. This confirmed earlier research by Spitters (1979) from experiments with binary mixtures of barley.

A completely different method is to avoid selection in an early stage and wait until a high degree of both homogeneity and homozygosity has been reached. At this stage, selection between lines can be performed, without regard to the pedigree. An example of this is the method of 'Single Seed Descent' (SSD). One great advantage of this technique is the opportunity to grow three successive generations in one year (spring cereals).

The SSD-method has proven to be a fast breeding procedure, but tests on the advantages compared to early generation selection methods are quite inconsistent. Knott & Kumar (1975) found in their field experiment with spring wheat that an early yield test procedure (EYT) produced lines with a significantly higher average yield than did SSD. But the yield level of the 20% best lines did not differ significantly for EYT and SSD. Computer simulation studies showed that, especially at low heritability, SSD performs just as well as pedigree selection (Casali & Tigchelaar, 1975).

The goal of the present study was to evaluate the differences in selection response between these two 'extreme' breeding approaches. On the one hand it makes a comparison between early and late selection and on the other hand the effectiveness of a selection among crosses is investigated. Yield in spring wheat is used as the character of interest.

Materials and methods

A field trial was performed, using the so called 'pseudo-lines' technique, instead of real crosses (Van Ooijen, 1989a;1989b). This technique provides a way to create any breeding generation that we like. Besides this, it provides the unique opportunity of an exact estimation of the genetic parameters. It is assumed that yield, as an example of a quantitative trait, segregates in a number of crosses at two or three independent loci. For the two loci crosses this will give nine different genotypes and the three loci crosses comprise 27 genotypes, each of which is represented by a specific variety. Distinct crosses are composed by different sets of varieties; in this way different levels of additive, dominance and epistatic effects are created. Segregating lines in successive generations are represented by appropriate mixtures of the composing varieties.

In addition to these pseudo-lines, all varieties used in the simulated crosses are grown in pure stand in large yield trials. These monoculture yields are considered as the true genotypic values of components of the mimicked crosses. They enable a comparison between the estimates of m and D from the pseudo-crosses and the corresponding 'true' values obtained from the monoculture yields. Thus this method provides a way to identify the several sources of bias. For a more detailed description of the pseudo-lines technique: see Van Ooijen (1989a and 1989b).

Thirty-five varieties and pure breeding lines of spring wheat (*Triticum aestivum* L.) were used to compose fifteen crosses (Table 1). Each cross was represented by a different combination of varieties as given in Table 2; five crosses were composed according to the three locus model and ten according to two loci. The field trial was performed in 1990 and 1991 at one location: Prof. Broekemahoeve, CPRO-DLO, Lelystad.

First year

Both an F_3 and an F_6 generation were created by means of pseudo-lines. All fifteen crosses were mimicked, assuming different sets of two homozygous parents. Hypothetically both the F_3 and the F_6 of each cross are derived from the same heterozygous F_1. The F_3 consisted of F_2-derived lines. In this F_3 early selection was applied. The F_6 lines were supposed to be derived from F_5 plants, which originated from the F_2 by means of single seed descent. Because both F_3 and F_6 consisted of single plant derived lines, field size was limited to 3-row plots of ± 1.5 m^2 in two replicates. Such plots are assumed to be large enough to give a rough estimate of the yielding potential of that particular F_3 or F_6 line (Kramer et

Table 1. Varieties and pure breeding lines used for the composition of the pseudo-lines and their mean yields in the monoculture experiment of both years 1990 and 1991 (10.5 m^2 plots) and estimated both from the F_3/F_6 experiment of 1990 (1.5 m^2 plots) and the F_6/F_7 experiment of 1991 (6 m^2 plots) in $kg{\cdot}ha^{-1}$.

		mono'90		F_3/F_6'90		mono'91		F_6/F_7'91	
	variety	yield	rank	yield	rank	yield	rank	yield	rank
1	Stratos	5506	9	4201	23	6044	7	6012	5
2	G 74010	5304	18	4381	18	5895	10	5834	7
3	vdH 1166-76	5068	25	4228	22	5829	11	5837	6
4	vdH 1132	5048	27	3970	30	6064	6		
5	ZESC 1963-6	6002	2	4980	6	6015	9	6060	2
6	TK 2832 3,10	5843	5	5092	2	5568	29	5611	15
7	vdH 3132	4797	32	4184	24	5630	25		
8	G 8005	5172	23	3945	31	5561	30		
9	Minaret	4869	31	4235	21	5597	27	5381	20
10	Heros	5059	26	4060	27	6115	4		
11	Darima	5605	8	4760	11	6105	5	6051	3
12	TK 2832 2,11	5695	6	5034	4	5732	19	5768	10
13	Ralle	4592	34	3906	32	5802	12		
14	Sunnan	5163	24	4528	15	5724	20		
15	Bastion	5347	16	4785	10	5591	28	5433	19
16	Spartacus	5476	10	4707	13	6158	2	6017	4
17	Adonis	5283	20	3864	33	5788	14		
18	Melchior	5960	3	4989	5	5460	33	5610	16
19	Axona	5898	4	5040	3	6134	3	6115	1
20	Kokart	5389	14	4858	8	5713	21	5793	9
21	Sicco	5291	19	4747	12	5557	31	5661	13
22	Wembley	4916	30	4096	26	5681	22	5481	18
23	Prinqual	4484	35	4045	28	4785	35	4824	21
24	JVO 1063	5374	15	4698	14	5543	32	5813	8
25	JVO 1064	5673	7	4483	17	6043	8		
26	JVO 1065	4945	29	3830	34	6322	1		
27	JVO 1070	4748	33	4292	20	5389	34		
28	JVO 2293	5034	28	3984	29	5757	18		
29	JVO 2296	5440	13	4500	16	5769	16		
30	JVO 2302	5469	11	4959	7	5782	15	5624	14
31	JVO 2311	5282	21	4125	25	5757	17		
32	JVO 3531	6153	1	5359	1	5636	24	5494	17
33	JVO 3537	5305	17	4316	19	5801	13		
34	JVO 3539	5183	22	3725	35	5607	26	5730	11
35	JVO 3549	5444	12	4811	9	5650	23	5673	12
	mean:	5309		4449		5760		5706	

al., 1982).

The number of lines in both generations was 42. All lines were sampled from their expected distributions. Assuming a two locus model, the theoretical distribution of the F_3 lines is (1:2:1:2:4:2:1:2:1)/16, for a three locus model it is (1:2:1:

Table 2. Composition of the pseudo crosses. The numbers of the varieties correspond to the numbering in Table 1.

cross	tri-heterozygote g_{14}	di-heterozygote g_5 g_{11} g_{13} g_{15} g_{17} g_{23}	mono-heterozygote g_2 g_4 g_6 g_8 g_{10} g_{12} g_{16} g_{18} g_{20} g_{22} g_{24} g_{26}	homozygote g_1 g_3 g_7 g_9 g_{19} g_{21} g_{25} g_{27}
A	7	21 19 32	6 2 12 27 5 14	1 11 16 9
		8 29 13	4 22 17 10 15 25	30 20 35 23
B	14	5 30 1	21 34 28 33 16 17	19 6 32 4
		20 7 23	24 26 3 35 13 10	2 18 31 22
C	8	22 35 14	30 7 4 32 3 10	2 16 5 1
		12 13 15	18 27 21 31 26 17	34 11 19 6
D	17	16 10 27	22 28 18 13 26 31	33 23 20 15
		3 34 11	4 14 35 7 12 29	9 8 24 19
E	23	30 18 31	19 15 22 10 14 4	33 32 24 17
		5 8 20	7 28 13 21 9 6	29 27 2 12
F		2	34 6 1 16	19 21 30 11
G		30	11 21 19 1	34 32 5 2
H		12	14 27 18 3	31 26 7 10
I		17	31 4 26 14	3 12 27 18
J		15	25 23 35 29	33 8 13 28
K		20	8 33 9 24	25 15 35 23
L		16	22 23 15 1	14 13 4 29
M		4	11 24 14 5	7 22 28 10
N		16	23 18 27 10	26 17 14 24
O		5	6 28 20 21	35 32 22 3

2:4:2:1:2:1:2:4:2:4:8:4:2:4:2:1:2:1:2:4:2:1:2:1)/64. For the F_6 lines the theoretical distributions are (225:30:225:30:4:30:225:30:225)/1024 and (3375:450:3375:450:60:450:3375:450:3375:450:60:450:60:8:60:450:60:450:3375:450:3375:450:60:450:3375:450:3375)/32768. The segregating ratios within lines all follow exactly the expected Mendelian frequencies.

The lines of one cross were grouped in a superplot. Each superplot was grown in two replicates. The F_3 and F_6 populations were grown in the same trial, so as to keep the growing conditions as similar as possible. Each replicate contained 15 F_3 populations and 15 F_6 populations, 30 superplots in all. Crosses were completely randomized among superplots, as well as lines were randomized within superplots. All plots were sown as 3.5 m long fields, with no spacing at the front and the rear. They consisted of three rows, 17.5 cm apart and the between plot distance was 40 cm. This resulted in an effective plot width of 75 cm. Sowing density was 230 seeds m^{-2}. Shortly before harvest the front and rear 75 cm of each plot were mowed.

The monoculture experiment consisted of eight replicates of all 35 varieties (plus one dummy variety to complete the blocks) in a completely randomized block design. The size of these plots was 8.5 m long and 1.5 m wide and they were trimmed back shortly before harvest to 7.0 m long. They consisted of 10 rows, 12.5 cm apart and 37.5 cm between neighbouring plots. Sowing density was 300 seeds m^{-2}.

Each field was simultaneously mowed and threshed with a combine harvester. After drying the total grain weight of each plot was measured and corrected for plot length. Samples were taken to determine the moisture content and all weights were converted to 0% moisture yields in kg per ha.

Analysis of the pseudo-line experiment

The following statistical model can be applied to the F_3/F_6 trial:

$$\underline{y}_{ijk} = m + a_i + \underline{r}_j + \underline{s}_{ij} + \underline{g}_{k(i)} + \underline{e}_{ijk},$$

with $\underline{y}_{ijk}$ = realized yield of line k of cross i in replicate j

m = over-all mean

a_i = fixed effect of cross i $\quad i_{F3} = 1,2,..,15 \quad i_{F6} = 16,17,..,30 \quad \Sigma_i a_i = 0$

$\underline{r}_j$ = effect of replicate j $\quad j = 1,2 \quad \underline{r}_j \approx N(0,\sigma_r^2)$

$\underline{s}_{ij}$ = effect of superplot ij $\quad \underline{s}_{ij} \approx N(0,\sigma_s^2)$

$\underline{g}_{k(i)}$ = effect of line k of cross i $\quad k = 1,2,..,42 \quad \underline{g}_{k(i)} \approx N(0,\sigma_{BLi}^2)$

$\underline{e}_{ijk}$ = residual effect of plot ijk $\quad \underline{e}_{ijk} \approx N(0,\sigma_e^2)$

(random effects are underlined; all effects are assumed mutually stochastic independent)

F_3 and F_6 crosses were placed in the same model to keep the effects of residuals, superplots and replicates similar.

The F_3 and F_6 means, as predictions of the cross mean in the final breeding stage (F_∞), were calculated as common means over all 84 plots of each cross. The between line variance is calculated for each cross separately by:

$$\sigma_{BLi}^2 = \frac{MSBL_i - MSE}{r}$$

When absence of epistasis and linkage is assumed, the between line variance, which will be indicated as V_1, can be defined as (Mather & Jinks, 1971):

$$V_{1F3i} = \frac{1}{2}D + \frac{1}{16}H \text{ and}$$

$$V_{1F6i} = \frac{15}{16}D + \frac{15}{1024}H ,$$

where D is the additive and H the dominance component of the genetic variance. Van Ooijen (1989c) found that D can be most accurately estimated by ignoring the dominance components. In this way the best estimators for D are:

$$D_{F3psi} \approx 2V_{1F3i} \text{ and}$$

$$D_{F6psi} \approx \frac{16}{15}V_{1F6i} .$$

Assuming a normal distribution of the random effects, the mean squares are χ^2 distributed and the error variances of the above estimators can be estimated according to Van Ooijen (1989b).

Analysis of the monoculture experiment

A straightforward two-way classification model is applicable to the monoculture trial. The variety effect is assumed to be fixed, whereas both block and residual effects are random. The variance components of blocks (σ_b^2) and residuals (σ_e^2) are estimated from the ANOVA. All effects are assumed mutually stochastic independent. The mean for each variety is calculated as the common mean over all 8 blocks. The error variance of these means is estimated by :

$$\hat{\text{var}}\ \bar{y}_{i.} = \frac{1}{8}s_b^2 + \frac{1}{8}s_e^2 = s_v^2$$

The same model applies to the variety plots in the pseudo-line trial.

The average for each cross was calculated directly from the frequencies of the different varieties (= genotypes; for the indices of the genotypes see Table 2) used in that specific cross. For each F_3 and F_6 population this amounts to:

$$m_{mo} = \sum_{i=1}^{n} f_i \cdot g_i , \quad \text{where } f_i = \frac{1}{42} \cdot \sum_{l=1}^{42} f_{il} \quad \text{and}$$

g_i is the value of genotype i appearing in that cross,

f_{il} and f_i are the frequencies of that genotype within each line l and within the cross resp.

$n = 3^{(\# \text{ loci})}$.

The mean of the F_∞ generation, in which the mere four homozygote genotypes are left over, reduces to:

$$\text{2 loci: } m_{F\infty mo} = ¼ \cdot (g_1 + g_3 + g_7 + g_9)$$

3 loci: $m_{F\infty mo} = ⅛\cdot(g_1 + g_3 + g_7 + g_9 + g_{19} + g_{21} + g_{25} + g_{27})$.

The between line variances within each cross, as a measure for the additive variance component, can then be estimated as (see van Ooijen, 1989b):

$$\sigma^2_{BL} = \frac{1}{42}\cdot\sum_{l=1}^{42}\{\sum_{i=1}^{n} f_{il}\cdot g_i\}^2 - m^2_{mo}$$

By using values from the monoculture experiment, the yield of genotype g_i is estimated as $v_i + \underline{e}$. The expectation of $\underline{e}$ equals zero and its variance is s_v^2. This has no effect on the given means but has a definite effect of overestimation on the between line variance. The true estimators for the between line variances are therefore given by:

$$V_{1F_G} = \sigma^2_{BLF} - \{\frac{1}{42}\cdot\sum_{i=1}^{n}\sum_{l=1}^{42} f_{il}^2 - \sum_{i=1}^{n} f_i^2\}\cdot s_v^2 \text{ , which will equal } \{1 - \frac{1}{2^{G-2}}\}\cdot D_{F_{Gmo}} \text{ ,}$$

where G is the generation (3,6 or ∞).

Additionally the genetic parameters were calculated from the monoculture data, given the expected line distributions in both the F_3 and F_6. These enable an estimation of the bias caused by sampling errors. A comparison was now made between the parameters directly estimated from the pseudo-line experiment (m_{F3ps}, m_{F6ps}, D_{F3ps} and D_{F6ps}) on the one hand and the parameters predicted from the monoculture data (m_{F3mo}, m_{F3exp}, m_{F6mo}, m_{F6exp}, $m_{F\infty mo}$, D_{F3mo}, D_{F3exp}, D_{F6mo}, D_{F6exp} and $D_{F\infty mo}$) on the other hand. From all these means and the corresponding standard deviations, the probability of finding a superior inbred line in each cross was calculated. It was assumed that the distribution of the inbred lines in the F_∞ is normal with mean m and variance D. Superior lines are considered to be those lines with an average yield at least equal to a certain threshold value T, for which the highest yielding parent was taken.

$$P_i = P(m_i + \sqrt{D_i}\cdot\chi > T),$$
$$\text{with } \chi \simeq N(0,1).$$

The probabilities and the resulting ranking of the crosses for F_{3ps} and F_{6ps} are compared to the $F_{\infty mo}$-data. Based on the P_{F3ps} the best three crosses were selected.

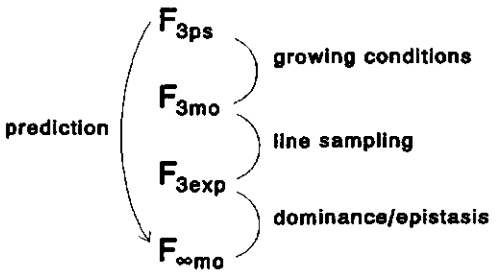

Fig. 1. Relations between the four estimated and predicted situations and the corresponding sources of bias.

The F_3 prediction of the F_∞ situation will be biased by three main sources of error: differences in growing conditions (including plot size and competition), line sampling effects and non-additive gene effects (dominance and epistasis). Correlations between appropriate pairs of the above parameters will express the different sources of bias, as is depicted in Fig. 1. An equal pathway is valid for the F_6. In this way the level of bias of the respective error sources can be analysed.

Second year

From the selected crosses in the ES-F_3 an F_6 generation was created. Because of the use of pseudo-lines, growing the F_4 and F_5 generation could be skipped; the F_6 was created as follows. As the character of interest is assumed to be controlled by only two or three loci, the maximum number of distinct genotypes appearing is 9 or 27. A common way to apply pedigree selection in the intermediate generations is to augment the selected F_3 lines to larger F_4 lines. Selection between and within these F_4 lines can then be applied and via an intermediate small F_5 line an F_6 line can be derived from every selected F_4 plant. Considering a situation where no selection is applied, the probabilities of finding the fully homozygous genotype in the F_4 equals $^{49}/_{256}$ (2 loci) or $^{343}/_{4096}$ (3 loci). For each single heterozygote the probability is $^{14}/_{256}$ and $^{98}/_{4096}$ and for the double heterozygote it equals $^{4}/_{256}$ and $^{28}/_{4096}$ for 2 and 3 loci respectively. The probability of finding the triple heterozygous genotype in case of three segregating loci is $^{8}/_{4096}$. 24 lines of each selected cross were sampled from these distributions. This resulted in an F_6 consisting of 52 fully homozygous lines and 20 heterogeneous lines, derived from single heterozygous F_4 plants. The latter will be expected to show a (3:2:3)/8 segregation in genotypes and hence were composed in accordance with these ratios.

The 72 phenotypic best SSD-F_6 lines from the 1990 trial were propagated to SSD-F_7 lines on the same scale as the ES-F_6. They consisted of 67 fully homogeneous lines and 5 segregating lines. These segregating lines were supposed to be derived from single heterozygous F_5 plants and therefore also were composed according to a (3:2:3)/8 ratio.

From both procedures the 72 lines were yield tested on 6 m^2 plots in four replicates. An $\alpha(0,1)$-design was used to minimize the effects of field heterogeneity (Patterson et al., 1978). Each replicate consisted of 9 blocks of size 8. All plots were sown as 5.5 m long and 1.5 m broad fields, without spacing on the front and the rear. They consisted of 10 rows with a row distance of 12.8 cm and a between plot distance of 34.8 cm. Sowing density was 250 seeds m^{-2}. Shortly

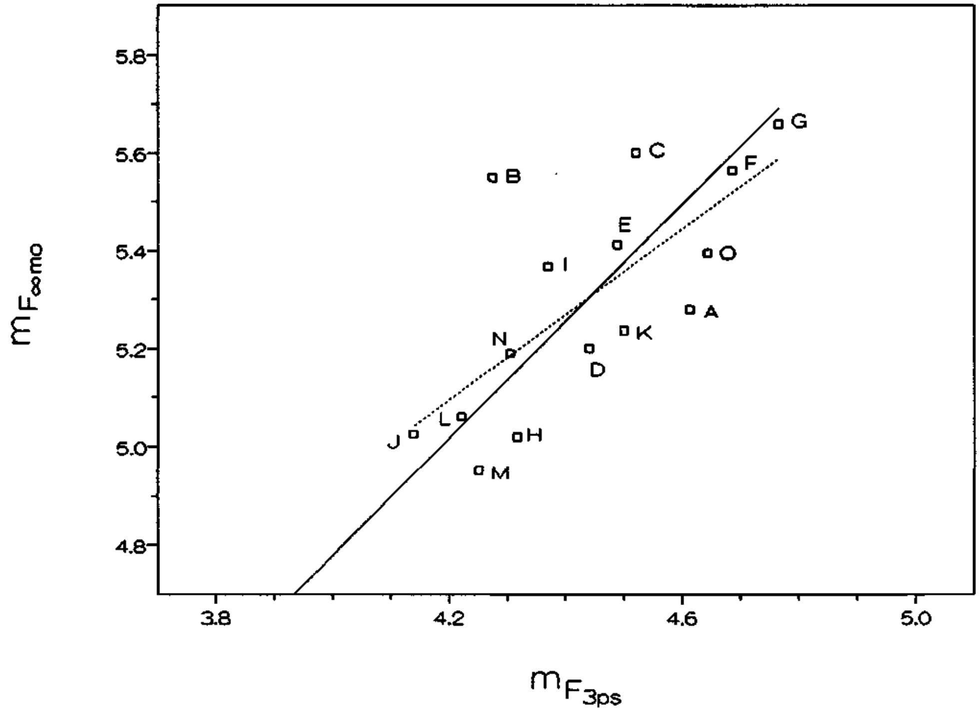

Fig. 2. Scatter diagram of the estimated F_{∞}-mean from the monoculture data against the predicted mean from the F_3 pseudo-lines for the crosses A to O. Yields in 10^3 kg·ha^{-1}. Given is the regression line (dotted) and the regression line with intercept zero (straight line).

before harvest the front and rear 75 cm of each plot were mowed. The data were analysed according to their balanced incomplete block design using the REML procedure in GENSTAT (Payne et al., 1990).

A monoculture trial was laid out, identical to the trial in 1990. With the aid of the monoculture data the true potential yield of all completely homogeneous lines, which are represented by just one variety, can be measured. In this way selection of the best lines can be checked on its accuracy. A final comparison between both selection procedures is made, based on the genotypic values of the 10 best lines in each procedure.

Results

The analysis of variance of the F_3/F_6 experiment in 1990 is presented in Table 3. There appears to be a highly significant effect between crosses and between lines

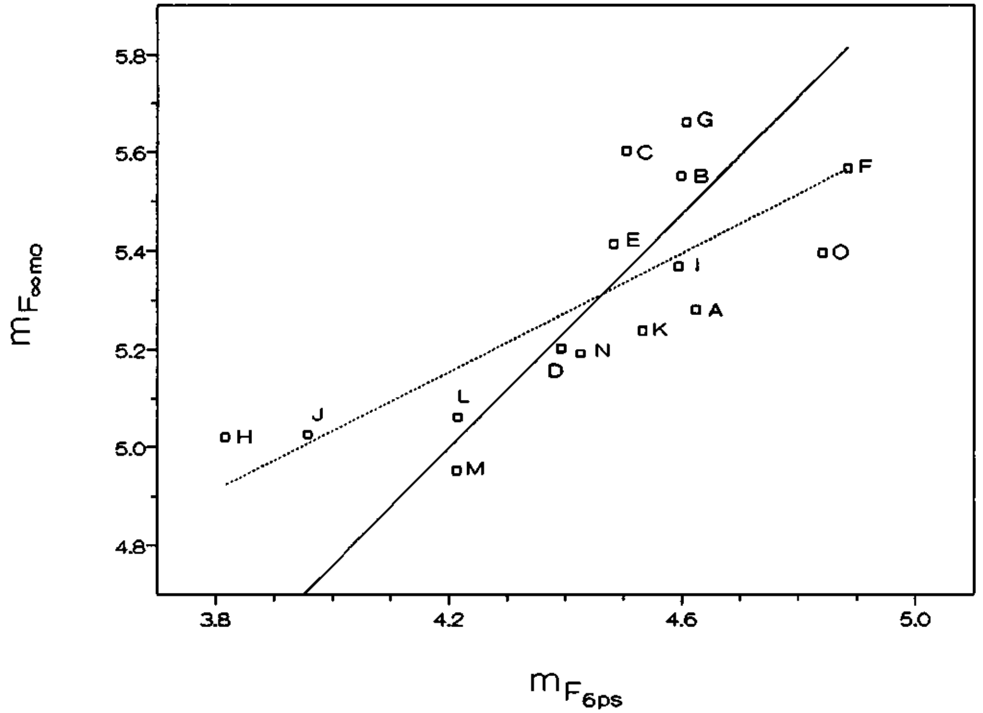

Fig. 3. Scatter diagram of the estimated F_∞-mean from the monoculture data against the predicted mean from the F_6 pseudo-lines for the crosses A to O. Yields in 10^3 kg·ha^{-1}. Given is the regression line (dotted) and the regression line with intercept zero (straight line).

within crosses. Variety yields are obtained from the monoculture experiment and these values are used to calculate the expected cross means and variation. From

Table 3. Analysis of variance of the pseudoline experiment.

Source of Variation	d.f.	Mean Square	F prob	Expected Mean Square
Replicates	1	$2.67\cdot 10^6$ (MSR)	≈ 0.10	$\sigma_e^2 + 42\cdot\sigma_s^2 + 42\cdot 30\cdot\sigma_r^2$
Crosses	29	$4.93\cdot 10^6$ (MSBF)	<0.001	$\sigma_e^2 + 42\cdot\sigma_s^2 + \frac{42\cdot 2}{30-1}\cdot\sum_{i=1}^{30} a_i^2$
Superplots (Res 1)	29	$0.88\cdot 10^6$ (MSS)	<0.001	$\sigma_e^2 + 42\cdot\sigma_s^2$
Lines within crosses	1230	$0.31\cdot 10^6$ ($MSBL_i$)	<0.001	$\sigma_e^2 + \frac{2}{42-1}\cdot\sum_{k=1}^{42} g_{ki}^2$
Plots (Res 2)	1230	$0.15\cdot 10^6$ (MSE)		σ_e^2

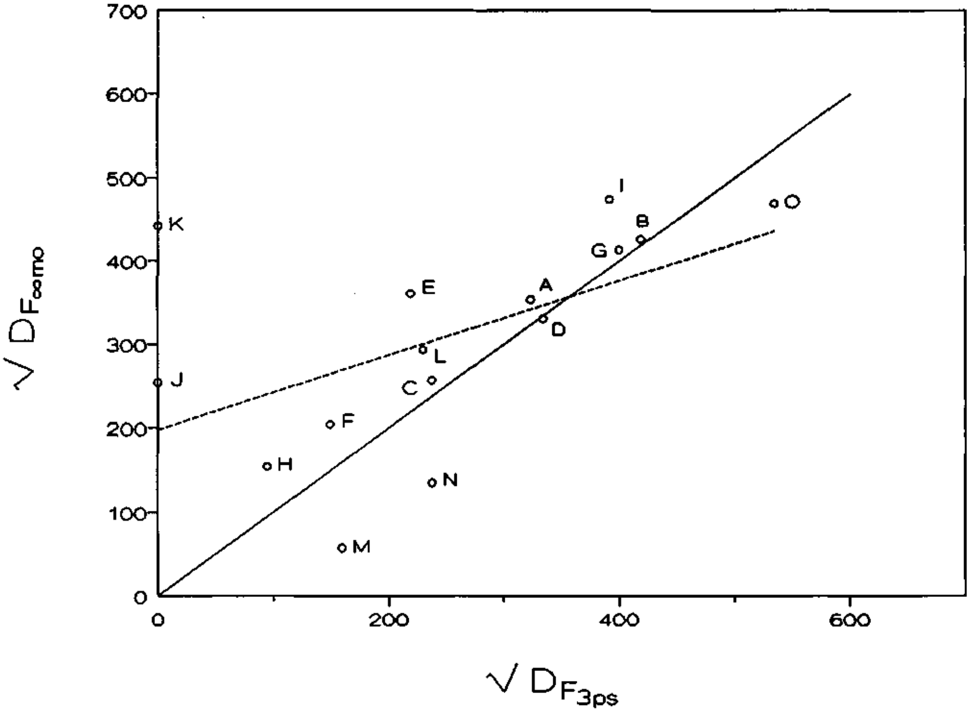

Fig. 4. Scatter diagram of the estimate of the genotypic standard deviation in the F_∞ from the monoculture data against the predicted standard deviation from the F_3 pseudo-line experiment for the crosses A to O. (kg·ha^{-1}). Given is the regression line (dotted) and the line y = x.

the fully homogeneous lines in the pseudo-line trial, which are represented by a single variety, variety yields can also be estimated. Both yields are presented in Table 1. Correlation between both values is quite high (Table 4). But there is an obvious discrepancy which will be caused by environmental error and intergenotypic competition among small plots.

Cross means are calculated for all seven situations: the normal means from the pseudo-line trial (m_{F3ps} and m_{F6ps}), the expected means, estimated from the monoculture data, both for the expected line distribution (m_{F3exp} and m_{F6exp}) and the realized distribution (m_{F3mo} and m_{F6mo}) and the expected line distribution in the

Table 4. Coefficients of correlation between the variety monoculture yields of years 1990 and 1991, the estimated variety yields from the F_3/F_6-trial (all based on 35 varieties) and the F_6/F_7-trial (based on 20 varieties).

	F_3/F_6'90	F_6/F_7'91	mono'91
mono'90	0.79	0.56	0.29
mono'91	-	0.91	

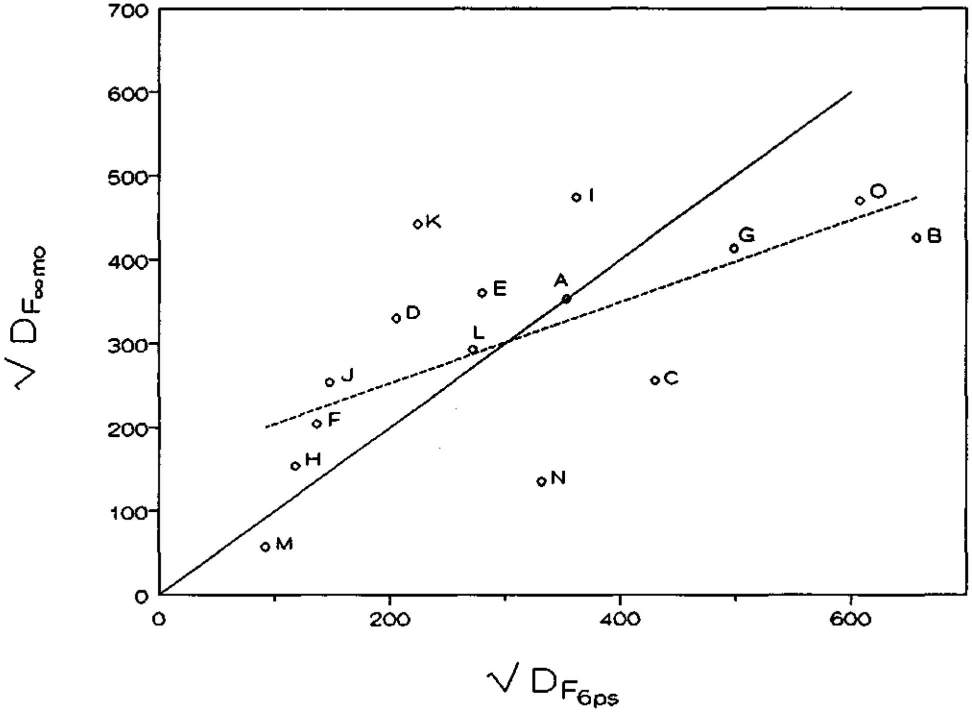

Fig. 5. Scatter diagram of the estimate of the genotypic standard deviation in the F_∞ from the monoculture data against the predicted standard deviation from the F_6 pseudo-line experiment for the crosses A to O. (kg·ha^{-1}). Given is the regression line (dotted) and the line y = x.

F_∞ ($m_{F\infty mo}$). There mutual correlation coefficients are given in Table 5. The most important comparisons are those between the estimated means from the pseudo-line experiment on the one hand and the predicted value for the F_∞ on the other. They are plotted in Figs. 2 and 3.

Table 5. Coefficients of correlation between estimated and predicted means.

	m_{F3mo}	m_{F6mo}	$m_{F\infty mo}$
m_{F3ps}	0.83	-	0.71
m_{F6ps}	-	0.75	0.77
m_{F3exp}	0.96	-	0.89
m_{F6exp}	-	0.97	1.0
$m_{F\infty mo}$	0.87	0.96	

Table 6. Coefficients of correlation between estimates and predictions of the genotypic standard deviation.

	$\sqrt{D_{F3mo}}$	$\sqrt{D_{F6mo}}$	$\sqrt{D_{F\infty mo}}$
$\sqrt{D_{F3ps}}$	0.42	-	0.53
$\sqrt{D_{F6ps}}$	-	0.68	0.65
$\sqrt{D_{F3exp}}$	0.88	-	0.77
$\sqrt{D_{F6exp}}$	-	0.98	0.99
$\sqrt{D_{F\infty mo}}$	0.72	0.98	

M_{F3ps} and $m_{F\infty mo}$ show a fairly high correlation, whereas m_{F6ps} and $m_{F\infty mo}$ are even closer correlated. This is not surprising because both effects of dominance and within plot competition, which are completely absent in the F_{∞}, are much smaller in the F_6 than in the F_3. Correlation between m_{F3exp} and $m_{F\infty mo}$ shows that there is an effect of dominance and/or epistasis in the F_3 ($r=0.89$), whereas it is fully absent in the F_6 ($r=1.0$). The effect of line sampling within each cross appears to be of only minor influence as the correlation between m_{exp} and m_{mo} show (0.96 for the F_3 and 0.97 for the F_6).

Correlations between the additive genetic standard deviances are much lower but the trend is the same (Table 6). Bias caused by all factors is obviously much larger on the variance than on the mean. $\sqrt{D_{F6ps}}$ is higher correlated with $\sqrt{D_{F\infty mo}}$ ($r=0.65$) than $\sqrt{D_{F3ps}}$ is ($r=0.53$). This is depicted in Figs. 4 and 5.

For the three situations F_{3ps}, F_{6ps} and $F_{\infty mo}$ the probabilities of finding superior inbred lines and their ranking are given in Table 7. The ranking of the crosses in the three situations does not differ much. Correlations are given in Table 8. Two of the three potentially best crosses (according to $P_{F\infty mo}$: G, B and O) are also identified as being most promising according to the P_{F3ps}. In the F_{6ps}, all three crosses are identified. $P_{F\infty mo}$ is plotted against P_{F3ps} and P_{F6ps} in Figs. 6 and 7 respectively.

From each of the three selected crosses (A, G and O) 24 F_6 lines were derived. The 72 highest yielding lines were selected in the SSD-F_6 and were propagated to F_7 lines. Both these SSD-F_7 and ES-F_6 lines were yield tested in the second year.

Out of the 144 F_6/F_7 lines, 119 are fully homogeneous and thus consist of only one variety. In this way yields can be estimated from the F_6/F_7 trial for a subset of the varieties (i.e. the selected ones). An exact estimate for each

Table 7. X-values from which the probabilities of superior inbred lines for each cross are derived and their ranks.

	F_{3ps}		F_{6ps}		$F_{\infty mo}$	
cross	x	rank	x	rank	x	rank
A	1.67	3	1.50	4	2.47	8
B	2.10	5	0.84	2	1.41	2
C	2.67	7	1.51	5	2.15	7
D	2.14	6	3.70	12	2.88	9
E	3.04	8	2.40	9	2.05	5
F	3.14	9	1.97	7	2.90	10
G	0.97	2	1.09	3	1.19	1
H	8.82	13	1.35	15	7.44	14
I	2.00	4	1.55	6	1.66	4
J	∞	14.5	8.10	13	4.46	12
K	∞	14.5	2.77	10	2.07	6
L	4.06	11	3.46	11	3.74	11
M	5.67	12	10.21	14	23.25	15
N	3.57	10	2.19	8	7.24	13
O	0.95	1	0.51	1	1.61	3

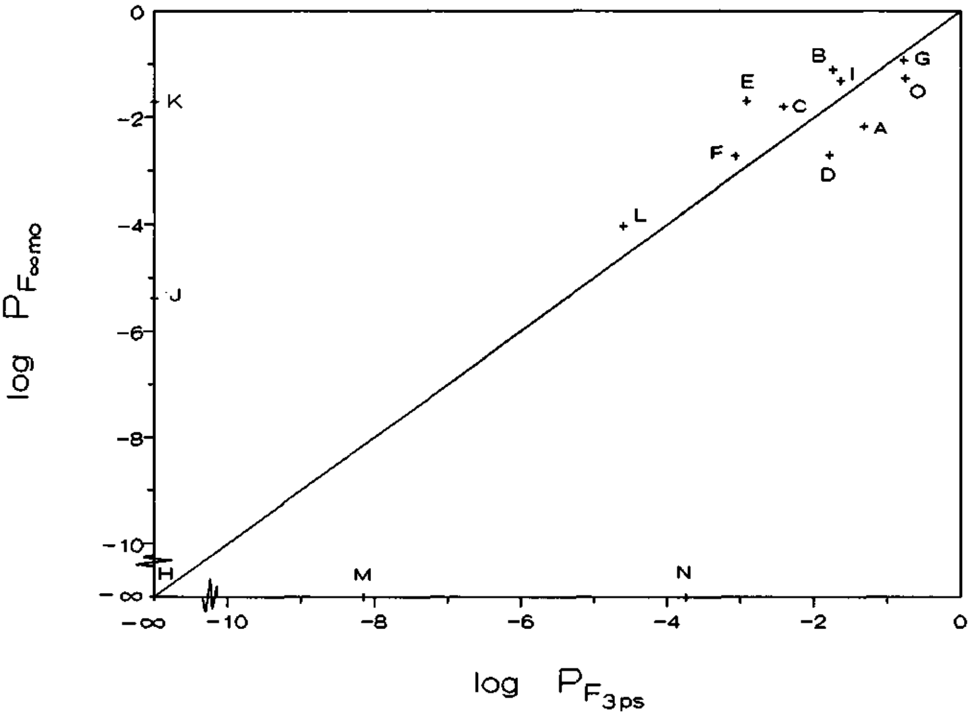

Fig. 6. Scatter diagram of the logarithm of the probability $P_{F\infty mo}$ against the logarithm of P_{F3ps} for the crosses A to O.

variety yield is obtained from the monoculture experiment. Values from both experiments are given in Table 1. There appears to be a very high variety-year interaction. The correlation coefficient between the variety yields of both years is very low: 0.29 (Table 4). The variety ranking, calculated from the F_6/F_7 trial is very similar to the one from the monoculture experiment. The correlation coefficient is high: 0.91. Obviously the size of the F_6/F_7 trial is adequate for the estimation of the line means.

The best 10 lines from both procedures, together with their 'pedigree' are given in Table 9. At first sight the best SSD lines seem better than the best ES lines. But because

Table 8. Spearman rank correlation coefficients between estimates and predictions of the probability of finding superior inbred lines.

	P_{F3mo}	P_{F6mo}	$P_{F\infty mo}$
P_{F3ps}	0.59	-	0.72
P_{F6ps}	-	0.84	0.83
P_{F3exp}	0.85	-	0.93
P_{F6exp}	-	0.98	0.99
$P_{F\infty mo}$	0.72	0.97	

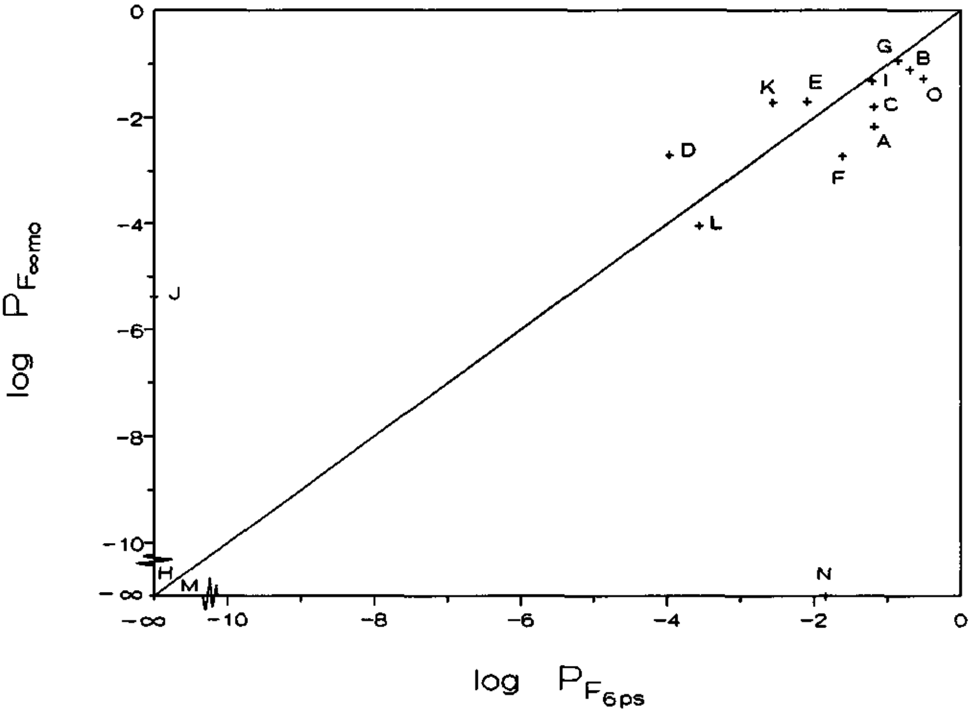

Fig. 7. Scatter diagram of the logarithm of the probability $P_{F\infty mo}$ against the logarithm of P_{F6ps} for the crosses A to O.

these phenotypic values will also be influenced by environmental effects, it is best to look at their 'genotypic' values. These genotypic values can be directly estimated from the monoculture yields of the varieties they are composed of. The mean variety yield of both years 1990 and 1991 was taken to exclude (a part of) the genotype-year interaction. It appears that the 10 best SSD lines consist of higher yielding varieties than the ES lines. Accordingly the mean genotypic value of the 10 best SSD lines is much higher than that of ES.

Conclusions and discussion

The use of pseudo-lines puts some restrictions on the genetic models that can be simulated and possibly on the conclusion when comparing the outcome of ES and SSD. This is discussed in Van Oeveren (1992) and the previous chapter. The 3-loci crosses were added to create a more truthful situation: a larger number of different genotypes could be present in the F_6/F_7 generation, which could possibly cause an

Table 9. The best 10 lines from both selection procedures with their phenotypic yield, line number, variety they are composed of and the 2-year mean monoculture yield of that variety as an estimate for the genotypic value ($kg \cdot ha^{-1}$).

	ES-F_6				SSD-F_7			
	yield	line	var	genotype	yield	line	var	genotype
1	6373	36	5	6008	6338	47	19	6016
2	6163	21	1/5/30*	5777	6245	56	19	6016
3	6123	22	16/4/35*	5651	6227	62	19	6016
4	6097	47	5/1/2*	5797	6208	48	11	5855
5	6087	64	3	5449	6195	49	5	6008
6	6072	7	11	5855	6180	40	16	5817
7	6025	42	2	5600	6157	59	5	6008
8	6023	35	5	6008	6150	25	19	6016
9	6012	2	1	5775	6140	18	19	6016
10	6012	6	1	5775	6120	61	5	6008
mean	6099			5770	6196			5978
mean of all 72 lines :			5713				5763	

* segregating lines, consisting of three different varieties in a 3:2:3 ratio.

advantage to the ES procedure compared to the purely 2-locus model.

Still the overall conclusions are the same as those of the previous field trial. Because the predictions of the several parameters in the F_{3ps} are much more accurate than those of the former year, the cross selection was much better. However, this did not result in a higher average of the selected lines in the ES-F_6 compared to the SSD-F_7. Due to a severe genotype-year interaction the response to selection was very small. Identical to the 2-loci experiment, SSD produced higher yielding lines. So we conclude that the extra effort in ES does not pay off. The SSD procedure will produce lines at least as good with less effort in quicker way.

When we have a closer look at the relations between the genetic parameters from the 1990-trial, some more information about the different sources of bias influencing them, can be obtained. The difference between the parameter estimates from the F_3/F_6 experiment and the $F_{\infty mo}$ data is caused by the following sources of bias:

1. line sampling,
2. dominance and epistasis and
3. differences in growing conditions: - plot size (environmental error) and
   - intergenotypic competition.

For each set of parameters (m, $\sqrt{D}$ and P, for both the F_3 and the F_6) correlations are tabulated according to the relevant sources of bias (Table 10). Corresponding

Table 10. Coefficients of correlation as measures of bias in the pseudo-line experiment (F_3/F_6) between the parameters (m, $\sqrt{D}$ and P) of the relevant situations (see Fig. 1).

	m		$\sqrt{D}$		P	
Source of bias	F_3	F_6	F_3	F_6	F_3	F_6
line sampling	0.96	0.97	0.88	0.98	0.85	0.98
dominance/epistasis	0.89 (0.73)*	1.0 (1.0)	0.77 (0.76)	0.99 (0.99)	0.93 (0.54)	0.99 (0.97)
growing conditions (plot size & competition)	0.83 (0.62)	0.75 (0.68)	0.42 (0.27)	0.68 (0.55)	0.59 (0.30)	0.84 (0.44)
competition	(0.77)	(0.84)	(0.33)	(0.68)		

* values between brackets are from the 1989-trial.

values from the previous field trial (Van Oeveren, 1992) are also given. It shows that line sampling has only little effect on the F_3 data and almost none on the F_6. Non-additive genetic effects can be of pronounced influence, but only on the F_3. Difference in growing conditions appears to be the major source of bias, especially inflicting on the (F_3) between line variance. From the 1989 data it was also possible to estimate the mere effect of plot size ($r=0.81$). This enabled an estimation of the effect of intergenotypic competition (Van Oeveren, 1992). We conclude that an important part of the bias is caused by the relative large portion of environmental error effecting the small plots, but most likely the major part is caused by competition. These findings are in accordance with the results from Van Ooijen (1989b).

References

Casali, V.W.D. & E.C. Tigchelaar, 1975. Computer simulation studies comparing pedigree, bulk and single seed descent selection in self pollinated populations. J. Am. Soc. Hort. Sci. 100: 364-367.

Jinks, J.L. & H.S. Pooni, 1980. Comparing predictions of mean performance and environmental sensitivity of recombinant inbred lines based upon F_3 and triple test cross families. Heredity 45: 305-312.

Knott, D.R. & J. Kumar, 1975. Comparison of early generation yield testing and a single seed descent procedure in wheat breeding. Crop science 15: 295-299.

Kramer, Th., J.W. van Ooijen & C.J.T. Spitters, 1982. Selection for yield in small plots of spring wheat. Euphytica 31: 549-564.

Mather, K. & J.L. Jinks, 1971. Biometrical genetics. Chapman and Hall, London, 382 pp.

Oeveren, A.J. van, 1992. A comparison between single seed descent and early cross selection in wheat breeding. Euphytica 58: 275-287.

Ooijen, J.W. van, 1989a. The predictive value of quantitative genetic parameters in autogamous crops: bias caused by intergenotypic competition. 1. F_∞-mean. Euphytica 44: 81-94.

Ooijen, J.W. van, 1989b. The predictive value of quantitative genetic parameters in autogamous crops: bias caused by intergenotypic competition. 2. F_∞-variance. Euphytica 44: 95-108.

Ooijen, J.W. van, 1989c. Estimation of additive genotypic variance with the F_3 of autogamous crops. Heredity 63: 73-81.

Payne, R.W., Lane, P.W., Ainsley, A.E., Bicknell, K.E., Digby, P.G.N., Harding, S.A., Leech, P.K., Simpson, H.R., Todd, A.D., Welham, S.J., White, R.P., Paterson, L.J., Thompson, R., Tunnicliffe Wilson, G. & L.G. Underhill, (1990). Genstat 5 release 2, reference manual supplement, Oxford, NAG, pp. 86-100.

Spitters, C.J.T., 1979. Competition and its consequences for selection in barley breeding. Pudoc, Wageningen, 268 pp.

5. COMPARATIVE SIMULATION STUDIES ON THE EFFECTS OF SELECTION FOR QUANTITATIVE TRAITS IN AUTOGAMOUS CROPS; EARLY SELECTION VERSUS SINGLE SEED DESCENT

Co-author: P. Stam. This chapter is published in Heredity 69: 342-351

Summary

A comparison is made between two contrasting breeding procedures for self-pollinating crops by means of computer simulation studies. The first is an early cross selection method, based on cross prediction by F_3 line estimates of the cross mean and between line variance. Subsequently, line selection is performed. In the second procedure selection is postponed to a more homozygous, F_6 generation, which is obtained by single seed descent. Only then is line selection performed, regardless of the pedigree.

The two procedures are compared for the cases of one and five crosses under selection. If only one cross is concerned, the early selection method reduces to early line selection and it was found to supply better inbred lines than SSD. But when more crosses are involved, the SSD procedure performed just as well as the early selection method and, at low heritability, even better because the early cross prediction was often poor.

Dominance appears to be of very little influence on either selection procedure.

Introduction

Various methods are used to obtain high performing inbred lines in the breeding of self-pollinating crops. Selection can be applied in different stages of the breeding process and with varying intensities. In order to get a better view of the efficiency of different selection procedures two contrasting selection methods are examined.

The first is a method, based on quantitative genetic theory, which has been developed to predict the genetic potential of a certain cross in an early breeding generation (Mather & Jinks, 1971). This technique, which makes use of estimates of the mean (m) and additive genetic variance (A) of a cross, should, in theory, allow the selection of those crosses most likely to produce superior inbred lines. Then, only the most promising crosses would be retained in the breeding programme. After this stage there is an opportunity to perform pedigree selection in the subsequent generations. This is referred to as 'early selection' (ES). In this form ES comprises both between- and within-cross selection.

A second, completely different method avoids selection in the early generations and waits until a high degree of both homogeneity and homozygosity has been reached. Genotypic differences between lines are then more pronounced and stable. At this stage selection between lines can be performed, without regard to the pedigree. An example of this is the method of single seed descent (SSD), and for crops which can be forced to a short generation cycle (e.g. spring cereals), SSD results in a quick advancement towards homozygosity.

It was found from experimental studies on ES (Jinks & Pooni, 1980; Van Ooijen, 1989a,b) that using F_3 lines to estimate the additive variance (A) is by far the most practical method for self-pollinating species. Estimates based on F_3 lines, however, may be biased by the effects of intergenotypic competition and in small grain crops this bias can be particularly large because plot size is limited by the amount of seed. Van Ooijen (1989a,b) concluded from his studies with mixtures of spring wheat varieties that the estimates of the genetic parameters for yield are in fact severely biased, leading to unreliable cross predictions. In addition, early selection is very demanding in terms of labour and the trial field area.

The SSD method has proved to be a fast breeding procedure but tests on the advantages compared to other methods are quite inconsistent. Knott & Kumar (1975) found in their field experiment with spring wheat that an early yield test procedure (EYT) produced lines with a significantly higher average than did SSD. The yield level of the 20% best lines did not, however, differ significantly for EYT and SSD. Boerma & Cooper (1975) also compared EYT with SSD and Pedigree Selection (PS) within crosses of soybean. They found no consistent differences in selection results between the three procedures and therefore regarded the rapid SSD method as most efficient. Computer simulation studies showed that, especially with low heritability, SSD performs just as well as pedigree selection (Casali & Tigchelaar, 1975); however, they considered only one population on a strictly additive model and selection was based on individual plant performance.

A field experiment was carried out to compare both selection procedures, using pseudo-lines of spring wheat (Van Oeveren, 1992). It was concluded that ES can easily lead to erroneous cross selection and that SSD is to be preferred for this particular set of crosses.

It is risky to depend solely on field trials when evaluating the effectiveness of the two different breeding strategies. They will only show a sample of what could happen. In this way, occasional random effects can misrepresent the average actual situation. It can therefore be very useful to examine the very same methods by means of computer simulations: the large number of replicate experiments that can be run provides a more secure base for statements about the average performance of both procedures. There are, of course, some disadvantages to this approach; simplifying assumptions must be made. On the other hand, simulation models allow variation of parameter values, such that a realistic range of input variables can easily be tested; and more importantly, the relative significance of the input variables on the selection results can be studied.

Materials and methods

A computer program was written in PASCAL which could simulate a breeding programme from the F_1 to the F_7.

A relatively simple genetic model was assumed; a varying number of loci (up to 100) was supposed to determine a true quantitative trait. The F_1 could be segregating for 30 of these loci, at most. All loci were assumed to have equal effect and to show no interaction (i.e. no epistasis) and no linkage. The two different breeding strategies are schematically visualized in Fig. 1. Starting with a certain F_1, which is heterozygous for a number of loci, an F_2 of size NF_2 is created by selfing. Two sets of progeny are created from this F_2 . All F_2 plants are advanced to F_3 lines for the early selection procedure and for the SSD procedure just a single F_3 plant is derived from each F_2 plant. Every plant is supposed to give an equal number of progeny.

The F_2 plant progenies will be large enough to allow an early yield testing. Each line is grown in a 3-row plot in two replications. Based on the F_3 cross mean and the additive variance, which is estimated as twice the between-line variance, the probability for each cross of retaining superior inbred lines in the F_∞ is calculated (see Van Ooijen, 1989b). Only the most promising crosses are propagated and from these a mild selection (50%) is made among lines. These best lines are again increased to larger F_4 lines which give a more sound yield estimate. Each line is grown in a larger plot in four replicates. The 20 best lines are selected. From each line five plants are randomly chosen and they are each propagated to two 3-row F_5 plots. These again offer a rough yield estimate and leave opportunity

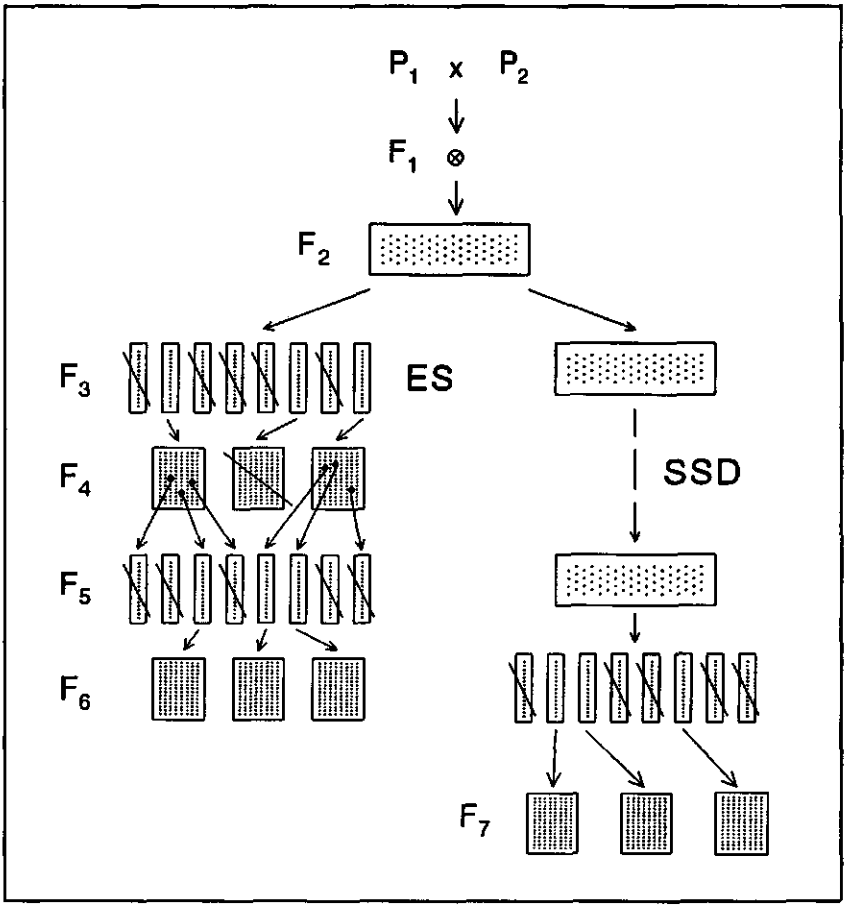

Fig. 1. Scheme of the two breeding procedures in case of one cross. Early Selection (left) and Single Seed Descent (right).

for line selection. Again the best 50% are increased to the F_6 which consist of fields of the same size as the F_4. The final evaluation takes place in this generation. The genotypic mean, maximum and standard deviation of the 10 phenotypic best lines are recorded.

An F_5 generation is derived

Table 1. Numbers of lines in each generation for both selection procedures as a function of the number of F_2 plants. Except for the ES generations F_4 to F_6 these are numbers of lines per cross.

	ES				SSD	
NF_2	F_3	F_4	F_5	F_6	F_3 .. F_6	F_7
25	25	20	100	50	25	20
50	50	25	100	50	50	25
100	100	50	100	50	100	50
200	200	100	100	50	200	100
400	400	200	100	50	400	100

from the SSD-F_3 by two more successive rounds of SSD. This F_5 is space planted and increased to NF_2 F_6 lines. Yield estimates are obtained from the F_6 lines and the best 50% are again propagated to large F_7 lines in four replicates. The 10 phenotypic best lines are finally evaluated and compared to those of the ES-F_6. The number of lines in each generation, relative to the number of F_2 plants are given in Table 1.

Phenotypic values are simulated for the two types of plots: the 3-row plots (ES-F_3, ES-F_5 and SSD-F_6), which consist of 100 plants and the large plots (ES-F_4, ES-F_6 and SSD-F_7), which consist of 600 plants. The model for the phenotype of plant l of line k in replicate j is as follows:

$y_{jkl} = \mu + r_j + E_{jk} + ic_{jk} + g_{jkl} + e_{jkl}$, where

μ = overall mean,

r_j = effect of replicate j $\quad r_j \simeq N(0,\sigma_r^2)$,

E_{jk} = between-plot environmental error of plot jk $\quad E_{jk} \simeq N(0,\sigma_{eb}^2)$,

ic_{jk} = between-plot competition effect of plot jk $\quad ic_{jk} \simeq N(0,\sigma_{ic}^2)$,

g_{jkl} = genotype of plant jkl,

e_{jkl} = within-plot environmental error of plant jkl $\quad e_{jkl} \simeq N(0,\sigma_{ew}^2)$.

The parameter μ is entered as an input variable, as well as the additive (a) and dominance (d) effects, which are simply summed over all loci to give the genotypic value g.

The heritability on a per plant basis can be derived from the input variables H^2 (F_3 between line heritability), a, d and n (number of plants per F_3 plot):

$$h^2 = \frac{H^2 \cdot n\sqrt{n} - (n-1) \cdot (\sqrt{n}-1)}{n\sqrt{n} - H^2 \cdot (n-1) \cdot (\sqrt{n}-1)} \; .$$

As the genetic variance between F_3 lines equals: $V_{1F3} = \frac{1}{2}A + \frac{1}{16}D$ (i) (in the absence of epistasis; A and D are the sum of the quadratic effects of additivity and dominance respectively over all segregating loci (Mather & Jinks, 1971)) and (ii) $h^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2}$, then the environmental error variance can be calculated as:

$$\sigma_e^2 = \frac{1 - h^2}{h^2} \cdot (\frac{1}{2}A + \frac{1}{16}D) .$$

This variance can be divided into a between- and a within-plot component. Assuming a soil heterogeneity index of 0.5 (Fairfield Smith, 1938) this amounts to:

$$\sigma_{eb}^2 = \frac{\sigma_e^2}{\sqrt{n}} \text{ (between plots) and}$$

$$\sigma_{ew}^2 = \sigma_e^2 - \sigma_{eb}^2 \text{ (within plots).}$$

The replicate effect is assessed as:

$$\sigma_r^2 = \sigma_{eb}^2 \cdot n \cdot m \,,$$

where m stands for the number of plots in one replicate.

Finally, an extra random effect was added to the plotsums to cover the effects of intergenotypic competition. From earlier research (A.J. van Oeveren, unpublished), simulating both yield and competition, it was found that competition effects within plots were negligible compared to both the between-plot competition effects and the environmental error. A normally distributed error term could well mimic the between-plot competition and the size of the effect was found to be related to both the number of plants per plot and the between-line variance, which varies with the level of heterogeneity of the generation. The following formula was found to perform well:

$$\sigma_{ic}^2 = 2.25 \cdot n \cdot G_F \qquad \text{, where}$$

$$G_F = \begin{cases} 1 \text{ in ES-F}_3 \,, \\ 1.7 \text{ in ES-F}_5 \text{ and} \\ 2 \text{ in SSD-F}_6 \,. \end{cases}$$

Results

One Cross

At first, simulations were performed to compare ES with SSD when only one cross was considered. In fact this results in early line selection instead of cross selection. As the possible negative effects of selecting the wrong crosses are completely absent, it is expected that ES will perform better than SSD, especially at high heritability. The quantitative trait of interest is assumed to be determined by 30 loci and the F_1 is heterozygous for 10 of them in the range 16-25. This indicates that 15 loci are positive homozygous and 5 are negative homozygous. Heritability and number of F_2 plants (NF_2) appear to be the two main sources of influence and results have been obtained in relation to these two variables. The first is varied

Table 2. Genotypic mean, standard error of that mean and maximum of the 10 phenotypic best lines in both the ES-F_6 and the SSD-F_7. Mean and maximum expressed as a percentage of the difference between the theoretical maximum and minimum genotype. Selection in only one cross with 10 segregating loci. The number of runs varies from 150 to 300.

		ES			SSD		
H^2	NF_2	mean	s.e.	max	mean	s.e.	max
0.02	25	64.0	4.9	81.9	55.7	4.8	78.1
	50	65.2	4.7	82.3	60.8	4.5	80.2
	100	67.4	4.2	84.7	64.0	4.2	83.5
	200	69.0	4.3	86.3	66.6	4.0	85.4
	400	71.2	4.5	87.9	67.8	4.1	86.5
0.05	25	68.5	4.7	83.9	58.0	4.5	78.7
	50	70.4	4.6	86.0	64.6	4.0	82.4
	100	73.3	3.6	88.0	68.9	3.8	84.8
	200	75.0	3.8	89.5	71.9	3.5	87.7
	400	78.0	4.0	90.7	73.7	3.7	90.0
0.1	25	71.3	4.6	84.4	59.5	4.3	78.7
	50	73.7	4.5	87.1	66.7	3.9	82.8
	100	77.7	3.5	90.0	71.7	3.4	86.4
	200	79.6	3.8	91.2	75.1	3.5	88.7
	400	81.4	3.5	92.8	78.0	3.1	91.0
0.25	25	74.2	4.0	85.5	62.2	4.0	79.7
	50	77.7	3.7	88.2	69.2	3.5	82.9
	100	80.8	3.5	90.7	74.3	2.9	86.9
	200	83.3	2.6	93.0	78.1	2.5	89.1
	400	85.7	2.8	94.8	81.8	2.6	92.3
0.5	25	75.7	3.9	85.6	62.8	3.8	79.5
	50	79.4	3.2	88.5	70.3	3.4	83.0
	100	82.6	3.3	91.0	76.0	3.4	87.5
	200	86.0	2.5	94.2	80.3	2.4	89.7
	400	88.1	2.5	95.7	83.4	2.4	91.8
1	25	76.4	3.9	85.3	64.0	4.0	79.5
	50	80.1	3.2	88.7	70.7	3.1	82.9
	100	83.9	2.7	91.3	76.2	2.9	86.9
	200	86.5	2.4	93.8	80.6	2.6	90.2
	400	89.0	2.2	95.3	84.1	2.1	92.5

from 0.02 up to 1.0 and NF_2 is varied from 25 to 400 plants. The results are presented in Table 2. The genotypic plot totals are given, expressed as a percentage of their maximum possible genotypic value, compared to the minimum value. Thus when all 10 segregating loci are positive homozygous the genotype

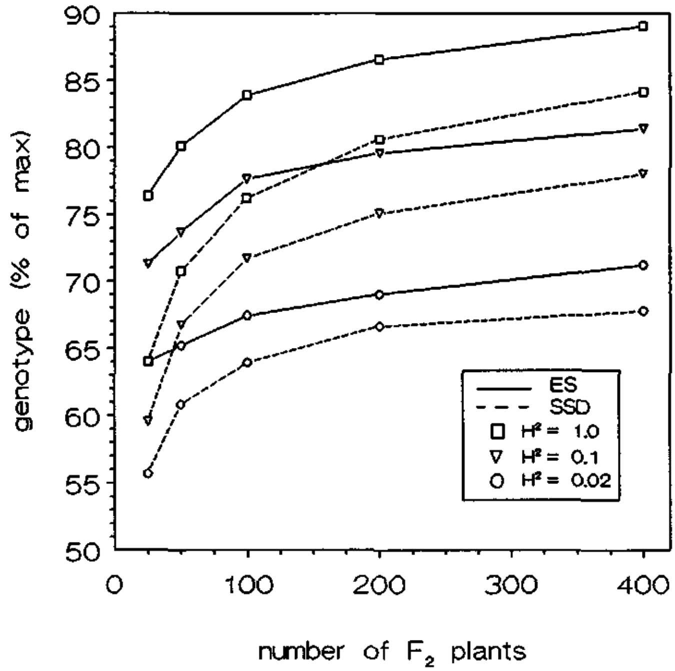

Fig. 2. The effects of the number of F_2 plants (= number of ES-F_3 and SSD-F_6 lines) on the genotypic mean of the 10 best lines resulting from both selection procedures at three different levels of heritability.

equals 100% and with all segregating loci negative homozygous it equals 0%. The mean of the phenotypic ten best lines is also shown graphically in Fig. 2.

The influence of the number of segregating loci is also tested. Two other crosses are examined, which are identical to the one described above except for the number of heterozygous loci in the F_1. Cross two segregates for loci 8 to 25 (18 loci) and cross three for range 1 to 25 (25 loci).

Table 3. Genotypic mean, standard error of that mean and maximum of the 10 phenotypic best lines in both the ES-F_6 and the SSD-F_7 at varying numbers of segregating loci. H^2 equals 1. 100 runs.

Segre-gating loci	NF_2	ES			SSD		
		mean	s.e.	max	mean	s.e.	max
25	25	65.5	2.7	71.4	57.9	2.5	67.4
	50	68.3	2.3	73.8	62.1	2.2	70.6
	100	70.5	1.6	76.0	65.6	1.8	72.9
	200	72.5	2.0	77.3	68.5	1.4	75.4
	400	74.1	1.5	79.1	70.9	1.5	76.9
18	25	69.9	2.6	76.7	60.8	2.5	72.5
	50	72.9	2.6	79.0	65.6	2.6	75.1
	100	75.8	2.3	82.0	69.7	1.9	78.2
	200	78.1	2.0	83.7	72.8	2.0	80.1
	400	80.1	1.9	85.4	76.2	1.7	82.6
10	25	76.4	3.9	85.3	64.0	4.0	79.5
	50	80.1	3.2	88.7	70.7	3.1	82.9
	100	83.9	2.7	91.3	76.2	2.9	86.9
	200	86.5	2.4	93.8	80.6	2.6	90.2
	400	89.0	2.2	95.3	84.1	2.1	92.5

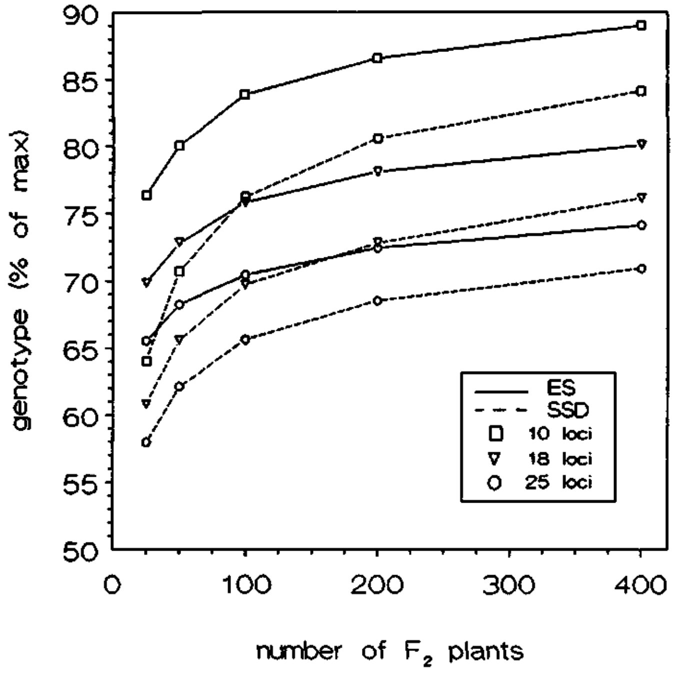

Fig. 3. The effects of the number of F_2 plants and the number of segregating loci on the genotypic mean of the 10 best lines resulting from both selection procedures. Heritability equals 1.

The results for all three crosses at a heritability of 1.0 (which is the most discriminating) and for a varying number of NF_2 are presented in Table 3; the genotypic means are plotted in Fig. 3.

Five crosses

Since ES comprises both selection between and within crosses, it is not fair to compare ES and SSD by only one cross; therefore, a breeding programme initiated with five crosses is considered. More crosses could be considered (and will be in an actual breeding programme), but our investigations were restricted to a subset of the potentially best crosses, which do not differ widely in population mean and variance. These five crosses have varying numbers and ranges of segregating loci. The characteristics of the crosses are given in Table 4.

Table 4. Genetic construction of the five crosses and the expected probability of retrieving well performing inbred lines in the F_∞ with a given threshold value (see text).

	Segregating loci	range	P
Cross A	4	11-14	0.008
Cross B	6	9-14	0.003
Cross C	6	10-15	0.025
Cross D	8	8-15	0.008
Cross E	10	7-16	0.016

The threshold value beyond which a

Table 5a. Genotypic mean and maximum of the 10 phenotypic best lines of both selection methods. Two situations: 5 crosses and 1 cross; for the first the frequency of selection of each cross is also given. Mean and maximum expressed as a percentage of the difference between the theoretical maximum and minimum genotype. 50 F_2 plants. Number of runs = 300.

	H^2	1	0.5	0.25	0.1	0.05	0.02
5 crosses :							
mean	ES	82.3	79.9	76.8	73.2	69.6	65.4
	SSD	80.9	80.0	77.8	74.5	71.2	66.5
	cross A	79.4	77.8	75.4	71.8	69.6	66.6
	cross B	75.6	72.8	71.7	67.7	63.8	60.6
	cross C	83.4	82.7	79.5	76.9	73.5	70.1
	cross D	79.7	77.6	74.4	71.3	66.8	63.0
	cross E	81.8	80.5	78.0	74.7	70.5	65.0
max	ES	88.1	86.1	84.5	83.5	81.3	79.5
	SSD	88.0	88.0	88.0	86.3	85.1	82.5
	cross A	80.0	80.0	80.5	79.6	79.1	77.7
	cross B	80.0	79.1	79.6	77.9	75.5	74.1
	cross C	88.7	88.7	87.5	86.9	84.9	84.5
	cross D	85.3	85.6	82.9	82.7	80.4	79.1
	cross E	90.0	90.1	88.4	88.1	84.9	83.1
freq	cross A	0.07	0.22	0.25	0.25	0.30	0.31
	cross B	0.01	0.07	0.10	0.11	0.14	0.17
	cross C	0.54	0.37	0.37	0.28	0.26	0.21
	cross D	0.10	0.12	0.14	0.17	0.13	0.19
	cross E	0.28	0.23	0.15	0.19	0.17	0.13
1 cross :							
mean	ES	80.1	79.4	77.7	73.7	70.4	65.2
	SSD	70.7	70.3	69.2	66.7	64.6	60.8
max	ES	88.7	88.5	88.2	87.1	86.0	82.3
	SSD	82.9	83.0	82.9	82.8	82.4	80.2

recombinant inbred line is considered to perform well was set at 80% of the best conceivable genotype (with all positive alleles accumulated). The expected probabilities of each cross exceeding this threshold value are also given in Table 4. It can be seen that cross C has the highest probability with cross E second, although the latter can in fact deliver the highest yielding inbred line. Only the one best cross is selected and propagated with continuing line selection. The results of both breeding methods are given in Tables 5a and 5b, for various heritabilities and two different numbers of F_2 plants: 50 and 100. Means of the 10 best lines for both selection methods in case of 50 F_2 plants and for both one and five crosses are plotted in Fig. 4.

Table 5b. Genotypic mean and maximum of the 10 phenotypic best lines of both selection methods. Two situations: 5 crosses and 1 cross; for the first the frequency of selection of each cross is also given. Mean and maximum expressed as a percentage of the difference between the theoretical maximum and minimum genotype. 100 F_2 plants. Number of runs = 200.

	H^2	1	0.5	0.25	0.1	0.05	0.02
5 crosses :							
mean	ES	86.8	83.6	79.9	76.1	72.6	68.2
	SSD	82.8	82.2	79.9	75.6	72.2	67.3
	cross A	80.0	79.2	77.0	74.4	72.6	68.7
	cross B	78.7	77.0	73.7	69.5	67.3	63.5
	cross C	87.9	85.9	83.6	80.0	77.2	72.8
	cross D	83.1	80.5	79.0	75.0	70.9	65.1
	cross E	87.8	86.5	82.6	79.4	74.7	70.3
max	ES	89.7	88.1	86.0	84.8	82.6	81.3
	SSD	89.6	90.4	89.0	86.8	85.9	82.7
	cross A	80.0	80.0	80.0	79.8	79.6	78.2
	cross B	80.0	80.0	79.6	77.6	77.4	76.3
	cross C	89.9	89.7	89.3	88.4	87.3	85.2
	cross D	88.6	87.0	86.7	85.1	82.6	80.4
	cross E	93.9	94.4	91.3	91.3	88.4	87.8
freq	cross A	0.11	0.20	0.20	0.25	0.27	0.27
	cross B	0.01	0.07	0.13	0.14	0.22	0.16
	cross C	0.62	0.37	0.29	0.23	0.22	0.21
	cross D	0.04	0.10	0.19	0.17	0.13	0.22
	cross E	0.23	0.27	0.18	0.21	0.17	0.15
1 cross :							
mean	ES	83.9	82.6	80.8	77.7	73.3	67.4
	SSD	76.2	76.0	74.3	71.7	68.9	64.0
max	ES	91.3	91.0	90.7	90.0	88.0	84.7
	SSD	86.9	87.5	86.9	86.4	84.8	83.5

Dominance

In Addition, the effect of dominance was tested on the selection results, in both cases of one and five crosses. Only unidirectional dominance was considered as it might produce the largest possible effect. It was applied on an intermediate ($d = ½a$) and a complete dominance level ($d = a$) and results were compared with the situation where dominance was absent (Table 5a). Results are given in Tables 6a and 6b.

Conclusions and discussion

It appears both from Table 2 and Fig. 2 that better lines are retained when herit-

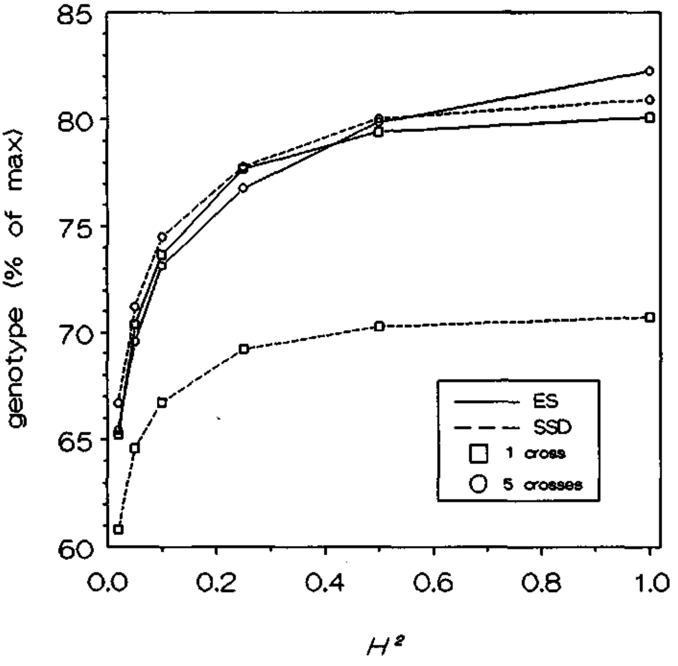

Fig. 4. The difference in selection results between ES and SSD in case of selection within one cross and between and within five crosses, in dependence of the heritability. 50 F_2 plants.

ability is high, as was expected. Likewise, results are higher when more F_2 plants (and therefore F_3 and F_6 lines) are taken. The advantage of ES over SSD is largest when heritability is high and the number of F_2 plants is low. Obviously the ES procedure is relatively inefficient when the F_3 line estimates are influenced by a large environmental error. On the other hand, the SSD programme leads to poor results when less than 100 F_2 plants are taken.

The number of segregating loci that characterize the trait appears to have quite a large impact on the selection response. A much better line can be retained if only 10 loci are involved (90% of the maximum versus 70% with 25 loci). This is not surprising because the chance of retrieving a genotype with all positive alleles accumulated will be much smaller when 25 instead of 10 loci are segregating. Accordingly, the absolute differences between ES and SSD are slightly smaller when more loci are involved. The general trend is the same, however: if only one cross is considered, ES performs better than SSD, at least, with equal numbers of F_3/F_6 lines.

These findings do agree with those of Casali & Tigchelaar (1975) concerning different procedures of plant selection in one population. They also concluded that an (early) pedigree selection was more efficient than SSD at heritabilities varying

Table 6a. Genotypic mean and maximum of the 10 phenotypic best lines in both the ES-F_6 and the SSD-F_7 and the selection frequencies of each cross. Mean and maximum expressed as a percentage of the difference between the theoretical maximum and minimum genotype. Dominance $d = ½a$, NF_2 = 50, 200 runs.

		H^2					
		1	0.5	0.25	0.1	0.05	0.02
mean	ES	83.2	80.1	77.6	73.7	70.1	65.4
	SSD	81.0	80.1	78.4	74.6	71.6	67.1
	cross A	79.3	77.7	75.5	72.8	70.3	66.8
	cross B	-	72.7	70.9	66.3	65.3	59.3
	cross C	84.1	82.8	80.3	76.8	73.5	70.1
	cross D	78.8	76.8	75.3	71.4	67.6	63.2
	cross E	82.5	80.3	79.4	75.5	71.0	66.1
max	ES	89.3	86.7	85.4	84.2	82.2	79.7
	SSD	87.8	87.7	88.1	87.0	84.9	82.8
	cross A	80.0	80.0	79.9	79.7	79.6	78.0
	cross B	-	80.0	78.2	76.8	75.7	73.2
	cross C	89.1	89.0	88.2	86.7	85.4	83.3
	cross D	85.8	84.3	83.6	83.1	80.9	79.6
	cross E	90.6	89.3	89.7	89.2	87.2	85.6
freq	cross A	0.02	0.17	0.19	0.20	0.24	0.25
	cross B	0.00	0.04	0.06	0.11	0.15	0.21
	cross C	0.59	0.37	0.32	0.30	0.26	0.23
	cross D	0.05	0.11	0.22	0.18	0.16	0.15
	cross E	0.34	0.32	0.22	0.23	0.20	0.15

from 0.1 to 1.0.

It is evident from Tables 5a and 5b that the advantage of ES over SSD in the case of one cross is completely absent in the more realistic situation of several crosses under selection. In the case of the one best line, the SSD method is even superior to ES at low and moderate heritability. As can be seen from the frequencies with which the different crosses are selected, the cross selection is close to random when heritability is low. When heritability increases the better crosses (C and E) are selected more frequently and the worst crosses (B and D) less frequently. It can be concluded that at low heritabilities the advantage of ES is completely lost due to erroneous cross selection. This was also found from field trials concerning ES (Van Ooijen, 1989b; Van Oeveren, 1992).

The effect of dominance is small. At moderate and high heritabilities ($H^2 > 0.1$) genotypic F_6 and F_7 values are slightly higher compared to the corresponding situation where dominance is absent. This effect increases with the number of

Table 6b. Genotypic mean and maximum of the 10 phenotypic best lines in both the ES-F_6 and the SSD-F_7 and the selection frequencies of each cross. Mean and maximum expressed as a percentage of the difference between the theoretical maximum and minimum genotype. Dominance $d=a$, NF_2 = 50, 200 runs.

		H^2					
		1	0.5	0.25	0.1	0.05	0.02
mean	ES	83.3	80.9	77.3	73.6	70.4	66.1
	SSD	81.4	80.1	78.2	74.5	71.1	66.5
	cross A	80.0	77.9	75.7	72.3	70.0	67.8
	cross B	-	72.6	71.7	66.4	65.1	60.4
	cross C	84.5	82.7	80.2	77.3	74.1	69.6
	cross D	79.4	78.4	76.2	71.4	67.6	64.4
	cross E	82.7	81.5	78.8	74.8	72.5	66.2
max	ES	89.6	88.0	85.4	84.6	82.6	79.6
	SSD	88.4	88.3	88.2	87.0	85.1	82.1
	cross A	80.0	79.9	79.8	79.6	79.6	78.0
	cross B	-	78.8	79.1	77.5	76.6	74.1
	cross C	89.4	88.9	88.1	87.4	85.7	83.1
	cross D	85.4	87.1	85.8	84.5	81.1	78.8
	cross E	90.6	90.5	88.9	88.7	87.5	84.6
freq	cross A	0.02	0.11	0.18	0.21	0.22	0.27
	cross B	0.00	0.03	0.12	0.09	0.16	0.18
	cross C	0.44	0.39	0.29	0.29	0.26	0.23
	cross D	0.05	0.13	0.18	0.20	0.16	0.18
	cross E	0.50	0.36	0.24	0.22	0.20	0.15

segregating loci. Because the level of heterozygosity in the F_6/F_7 is not very high, the genotypic line values do not differ much from those in the completely additive situation. The effect of dominance on the SSD procedure is therefore small. It could have an effect on the ES programme if the cross prediction changes. As can be seen from the frequencies of selected crosses, there is a tendency towards a higher selection frequency of the crosses with the largest variances. Because they are, in this case, also the most promising crosses, the selection results could be higher than results in the completely additive situation. The ES-F_7 line genotypes are indeed slightly higher at high and moderate heritabilities but it is not clear to what extent this is due to the selection of a better cross or to the fact that dominance increases the value of heterozygotes. If the crosses with the largest variance are not the most promising ones, dominance could have a negative effect on the final selection result. In any case, the overall effect is small.

It is interesting to examine the total trial field area needed for both procedures.

Table 7. Total number of F_3-type (m_1) and F_6-type (m_2) lines for both selection procedures, in relation to the number of F_2 plants (NF_2), in the case of 1 and 5 crosses. Calculated from these are the total field trial area (t.a.) occupied by them, expressed as the number of F_3-type plots, and the according ratio.

			NF_2				
			25	50	100	200	400
1 cross	ES	m_1	125	150	200	300	500
		m_2	70	75	100	150	250
		t.a.	1370	1500	2000	3000	5000
	SSD	m_1	25	50	100	200	400
		m_2	20	25	50	100	100
		t.a.	370	500	1000	2000	2400
	ES/SSD		3.7	3	2	1.5	2.1
5 crosses	ES	m_1	225	350	600	1100	2100
		m_2	70	75	100	150	250
		t.a.	1570	1900	2800	4600	8200
	SSD	m_1	125	250	500	1000	2000
		m_2	62	100	100	100	100
		t.a.	1242	2100	2600	3600	5600
	ES/SSD		1.3	0.9	1.1	1.3	1.5

The size of a large yield plot (F_7-type) will be about four times the size of a small 3-row plot (F_3-type). Additionally, the F_7-type lines are grown in four replicates instead of two, so they will occupy eight times as much space as the F_3-type lines. The total number of lines used for the F_3 to the F_6/F_7 generation is given in Table 7, together with the total area needed, expressed as the number of F_3-type plots. It can be seen that, with a single cross, ES occupies much more space than SSD. When equal trial sizes are engaged for both procedures, the advantage of ES over SSD will be less obvious and maybe even absent. In the case of five crosses, the ES trial is larger than the SSD trial except for $NF_2 = 50$. In most cases this implies an even larger advantage of SSD over ES when equal trial sizes are used. Of course greenhouse area is also necessary for the early SSD generations. This is partly compensated in the above comparison by the fact that the last SSD generation is one ahead of the last ES generation. When a true economically based comparison is made, other factors have to be regarded, such as the amount of labour and materials.

Some other aspects have not been considered. For instance, a genotype-environment (-year) interaction can have a large potential bias on selection results. It would probably most influence the ES procedure, because the chance of discarding potentially good crosses or lines due to a bad performance in an occasional selection environment would be high. This would increase the advantage of SSD. Another aspect is the number of selected crosses. The results of ES are likely to improve when a second or third cross is kept.

References

Boerma, H.R. & R.L. Cooper, 1975. Comparison of three selection procedures for yield in soybeans. Crop Science 15: 225-229.

Casali, V.W.D. & E.C. Tigchelaar, 1975. Computer simulation studies comparing pedigree, bulk and single seed descent selection in self pollinated populations. J. Am. Soc. Hort. Sci. 100: 364-367.

Fairfield Smith, H., 1938. An empirical law describing heterogeneity in the yields of agricultural crops. J. Agr. Sci. 28: 1-23.

Jinks, J.L. & H.S. Pooni, 1980. Comparing predictions of mean performance and environmental sensitivity of recombinant inbred lines based upon F_3 and triple test cross families. Heredity 45: 305-312.

Knott, D.R. & J. Kumar, 1975. Comparison of early generation yield testing and a single seed descent procedure in wheat breeding. Crop science 15: 295-299.

Mather, K. & J.L. Jinks, 1971. Biometrical genetics. Chapman and Hall, London, 382 pp.

Oeveren, A.J. van, 1992. A comparison between single seed descent and early cross selection in wheat breeding. Euphytica 58: 275-287.

Ooijen, J.W. van, 1989a. The predictive value of quantitative genetic parameters in autogamous crops: bias caused by intergenotypic competition. 1. F_∞-mean. Euphytica 44: 81-94.

Ooijen, J.W. van, 1989b. The predictive value of quantitative genetic parameters in autogamous crops: bias caused by intergenotypic competition. 2. F_∞-variance. Euphytica 44: 95-108.

6. OPTIMIZING SINGLE-TRAIT SELECTION IN THE BREEDING OF AUTOGAMOUS CROPS; A SIMULATION STUDY

Summary

Two contrasting breeding methods are evaluated on their ability to supply well performing inbred lines by means of computer simulation studies. The simulation model is restricted to selection on a single, quantitative trait in self-pollinating small grain crops. The first procedure is an early cross selection method, based on cross performance. Subsequently, line selection is performed. In the second procedure selection is postponed to a more homozygous, F_6 generation, which is obtained by single seed descent. Only then line selection is performed, regardless of the pedigree. An economical comparison and optimization for the two procedures is made.

In the early selection procedure more than one cross needs to be selected to achieve the optimum selection result. Even at the optimum, results are no better than a pure early line selection method, where no cross selection is made. However, the results of the early selection procedure do outyield the SSD method in most cases. When costs for both procedures are taken into account it appears that SSD is much more efficient, leading to better selection results with equal costs.

Genotype-year interaction leads to a considerable reduction in selection efficiency; the reduction is of the same order of magnitude for the two procedures.

Introduction

Various methods are being used to obtain high performing inbred lines in the breeding of self-pollinating crops. Selection can be applied in different stages of the breeding process, and with varying intensities. In order to obtain a better view of the efficiency of different selection procedures two contrasting selection methods have been examined and their optimum breeding schemes and accompanying costs are discussed.

The first is a method, based on quantitative genetic theory, which has been developed to predict the genetic potential of a certain cross in an early breeding generation (e.g. Mather & Jinks, 1971). This technique, which makes use of estimates of the mean (m) and the additive genetic variance (A) of a cross, should, in theory, allow the selection of those crosses most likely to produce superior inbred lines. Then only the most promising crosses would be retained in the breeding programme. After this stage there is an opportunity to perform pedigree selection in the subsequent generations. This is referred to as 'early selection' (ES). In this form ES comprises both between and within cross selection.

A second, completely different method avoids selection in the early generations and waits until a high degree of both homozygosity and within-line homogeneity has been reached. Genotypic differences between lines are then more pronounced and stable. At this stage selection between lines can be performed, without regard to the pedigree. An example of this is the method of single seed descent (SSD); for crops which can be forced to a short generation cycle (e.g. spring cereals), SSD results in a quick advancement towards homozygosity.

Experimental studies on ES, with selection based on F_3 line estimates, have been performed with mixtures of spring wheat (Van Ooijen, 1989a; 1989b; Van Oeveren, 1992). It was concluded from these studies that the estimates will be severely biased by the effects of intergenotypic competition and relatively large environmental errors. This can lead to unreliable cross predictions. In addition, early selection is very demanding in terms of labour and trial field area.

The SSD method has proved to be a fast breeding procedure, but tests on the advantages compared to other methods are quite inconsistent. Knott & Kumar (1975) found in their field experiment with spring wheat that an early yield test procedure (EYT) produced lines with a significantly higher average than did SSD. But the yield level of the 20% best lines did not differ significantly for EYT and SSD. Boerma & Cooper (1975) also compared EYT with SSD and Pedigree Selection (PS) within crosses of soybean. They found no consistent differences in selection results between the three procedures and therefore regarded the rapid SSD method as most efficient. Computer simulation studies showed that, especially with low heritability, SSD performs just as well as pedigree selection (Casali & Tigchelaar, 1975). These authors, however, considered only a single cross, with strictly additive genetic effects and selection was based on individual plant performance.

A previous simulation study, considering a comparison between the two selection procedures (Van Oeveren & Stam, 1992), indicated a preference to the early selection procedure when only one population was considered. However, when one out of more crosses was to be selected the SSD method was more advantageous. It further showed that dominance was of no significant influence.

Materials and methods

A computer program was written in PASCAL which could simulate a breeding programme from the F_1 to the F_7.

A relatively simple genetic model was assumed; a varying number of loci (up to 100) was supposed to determine a single quantitative trait. The F_1 could be segregating for 30 of these loci at most. All loci were assumed to have equal effect and to show no interaction (i.e. no epistasis) and no linkage. Starting with a certain F_1, which is heterozygous for a number of loci, an F_2 of size NF_2 is created by selfing. Two sets of progeny are created from this F_2. All F_2 plants are advanced to F_3 lines for the Early Selection procedure and for the SSD procedure just a single F_3 plant is derived from each F_2 plant. Every plant is supposed to give an equal number of progeny.

The F_2 plant progenies will be large enough to allow an early yield testing. Each line is grown in a 3-row plot in two replications. Based on the F_3 cross mean and the additive variance, which is estimated as twice the between-line variance, the probability for each cross of retaining superior inbred lines in the F_∞ is calculated (see Van Ooijen, 1989b). Only the most promising crosses are propagated and from these a mild selection ($SelF_3$) is made among lines. These best lines are again increased to larger F_4 lines which give a more sound yield estimate. Each line is grown in a larger plot in four replicates. The $SelF_4$ best lines are selected. From each line $Seln$ plants are randomly chosen and they are each propagated to two 3-row F_5 plots. These again offer a rough yield estimate and allow the opportunity for line selection. Again the best $SelF_5$ lines are increased to the F_6 which consist of fields of the same size as the F_4. The final evaluation takes place in this generation. The genotypic mean, maximum and standard deviation of the 10 phenotypically best lines are recorded.

An F_5 generation is derived from the SSD-F_3 by two more successive rounds of SSD. This F_5 is space planted and increased to NF_2 F_6 lines. Yield estimates are obtained from the F_6 lines and the best $SelF_6$ are again propagated to large F_7 lines in four replicates, equal to the ES-F_6. The 10 phenotypically best lines are finally evaluated and compared to those of the ES-F_6. A diagrammatical representation of the two procedures is given by Van Oeveren & Stam (1992).

Phenotypic values are simulated for the two types of plots: the 3-row plots

(ES-F_3, ES-F_5 and SSD-F_6), which consist of n plants and the large plots (ES-F_4, ES-F_6 and SSD-F_7), which consist of $6n$ plants. The model for the phenotype of plant m of line k in replicate j in year l is as follows:

$y_{jklm} = \mu + r_{jl} + ic_{jkl} + g_{jkm} + GY_{klm} + e_{jklm}$, where

μ = overall mean,
r_{jl} = effect of replicate j in year l $r_{jl} \simeq N(0,\sigma_r^2)$,
E_{jkl} = between-plot environmental error of plot jkl $E_{jkl} \simeq N(0,\sigma_{eb}^2)$,
ic_{jkl} = between-plot competition effect of plot jkl $ic_{jkl} \simeq N(0,\sigma_{ic}^2)$,
g_{jkm} = genotype of plant jkm,
GY_{jklm} = interaction between genotype jkm and year l $GY_{jklm} \simeq N(0,\sigma_{gy}^2)$,
e_{jklm} = within-plot environmental error of plant $jklm$ $e_{jklm} \simeq N(0,\sigma_{ew}^2)$.

Because no comparison between years was made, the year effect was left out.

The parameter μ is entered as an input variable, as well as the additive (a) and dominance (d) effects, which are simply summed over all loci to give the genotypic value g.

The heritability on a per plant basis can be derived from the input variables H^2 (F_3 between-line heritability), a, d and n (number of plants per F_3 plot) :

$$h^2 = \frac{H^2 \cdot n\sqrt{n} - (n-1)\cdot(\sqrt{n}-1)}{n\sqrt{n} - H^2\cdot(n-1)\cdot(\sqrt{n}-1)} \ .$$

As the genetic variance between F_3 lines equals: $V_{1F3} = \frac{1}{2}A + \frac{1}{16}D$ (i) [in the absence of epistasis; A and D are the sum of the quadratic effects of additivity and dominance respectively over all segregating loci (Mather & Jinks, 1971)] and (ii) $h^2 = \dfrac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2}$, then the environmental error variance can be calculated as:

$$\sigma_e^2 = \frac{1 - h^2}{h^2} \cdot \left(\frac{1}{2}A + \frac{1}{16}D\right) \ .$$

This variance can be divided into a between- and a within-plot component. Assuming a soil heterogeneity index of 0.5 (Fairfield Smith, 1938) this amounts to:

$\sigma_{eb}^2 = \dfrac{\sigma_e^2}{\sqrt{n}}$ (between plots) and

$\sigma^2_{ew} = \sigma^2_e - \sigma^2_{eb}$ (within plots).

The replicate effect is assessed as:

$\sigma^2_r = \sigma^2_{eb} \cdot n \cdot k$,

where k stands for the number of plots in one replicate.

To cover the effects of intergenotypic competition between F_3 type plots an extra random effect was added to the plot sums. From earlier research (A.J. van Oeveren, unpublished), simulating both yield and competition, it was found that competition effects within plots were negligible compared to both the between-plot competition effects and the environmental error. A normally distributed error term could well mimic the between-plot competition and the size of the effect was found to be related to both the number of plants per plot and the between-line variance, which varies with the level of heterogeneity of the generation. The following formula was found to perform satisfactorily:

$\sigma^2_{ic} = 2.25 \cdot n \cdot G_F$, where

$$G_F = \begin{cases} 1 \text{ in ES-}F_3 , \\ 1.7 \text{ in ES-}F_5 \text{ and} \\ 2 \text{ in SSD-}F_6 . \end{cases}$$

The genotype-year interaction was accounted for by adding a random term to each genotype, drawn from a normal distribution (with variance σ^2_{gy}). Its magnitude could be varied and was indicated as a fraction of the environmental variance (σ^2_e).

Costs

Data were obtained from two main cereal breeding companies in the Netherlands, concerning the costs of comparable breeding schemes as described above. These companies were Zelder and VanderHave. The data covered all costs from preparation and analysis of all trials, costs of machinery to trial field area and labour. The latter was classified into field worker, assistant and plant breeder hours. From Zelder information was also available on the comparable costs of an SSD generation. Costs were classified as either fixed or variable to enable a cost calculation per line or plant. This resulted in the final costs per generation as given in Table 1.

As the fixed costs for the single-plant progeny generations differed significantly for the two companies, the mean was calculated and would serve as an estimate

Table 1. Fixed and variable costs, for three different stages in the breeding process (in dutch guilders). Obtained from two sources: Zelder and VanderHave and their corresponding mean.

Generation	Source	Fixed		Variable
SSD	Zelder	ƒ8,000	+	ƒ0.46
single plant progeny (F_3 type)	Zelder	ƒ24,000	+	ƒ2.60
	VanderHave	ƒ9,600	+	ƒ3.60
	mean	ƒ16,800	+	ƒ3.10
line progeny (F_7 type)	Zelder	ƒ31,000	+	ƒ52.-
	VanderHave	ƒ4,500	+	ƒ38.-
	mean	ƒ17,750	+	ƒ45.-

for the true costs of the F_3 type fields. The same was done for the line progenies, as basis for the F_7 type fields. With these three formulas an indication of the total costs of the two breeding schemes could be given and accordingly an economically comparison and optimization could be made.

Results

Optimum number of crosses and lines

At first, ES is compared with SSD when a variable number of crosses is selected in the F_3. Ten different crosses are simulated, with varying numbers and ranges of segregating loci as specified in Table 2. Of each cross 100 lines (NF_2) are grown in both the ES-F_3 and the SSD-F_6. From the first a total of 75 lines is selected, originating from a varying number of crosses, with equal numbers of lines per cross. $SelF_4$ equals 15 and $Seln$ equals 10. This results in 150 F_5 lines. The 75 best are selected ($SelF_5$). From the 10 × 100 SSD-F_6 lines the phenotypic best 150 are selected ($SelF_6$). So both procedures consist of an equal amount of F_7 type fields. Effects of intergenotypic competition were added but genotype-year interaction was not taken into account in these simulations. Results for heritabilities of 0.05 and 0.10 are given

Table 2. Genetic construction of the ten crosses and the expected probability P of retrieving well performing inbred lines in the F_∞.

	segregating loci	range	P
Cross 1	12	13-24	.000
Cross 2	12	15-26	.001
Cross 3	12	17-28	.032
Cross 4	10	14-23	.000
Cross 5	10	16-25	.001
Cross 6	10	17-26	.004
Cross 7	10	18-27	.021
Cross 8	8	15-22	.000
Cross 9	8	17-24	.000
Cross 10	8	19-26	.012

Table 3. Genotypic mean and maximum of the 10 phenotypic best lines in both the ES-F_6 and the SSD-F_7. Expressed as a percentage of the difference between the theoretical maximum and minimum genotype; their standard errors are given between brackets. Selection of a variable number of crosses. $NF_2=100$, $n=200$, $SelF_4=15\times10$, $SelF_5=75$, $SelF_6=150$. 200 runs.

		ES		SSD	
H^2	$SelF_3$	mean	max	mean	max
0.1	1×75	76.9 (7.0)	84.2 (7.8)	78.4 (2.0)	88.1 (3.5)
	2×37	79.8 (4.7)	87.7 (5.7)	78.6 (2.2)	87.9 (3.8)
	3×25	80.4 (4.0)	88.3 (4.8)	78.4 (2.2)	88.2 (4.2)
	5×15	81.4 (3.0)	89.4 (4.0)	78.4 (2.2)	88.0 (3.7)
	7×11	81.0 (3.2)	89.0 (4.3)	78.2 (2.1)	87.7 (3.8)
0.05	1×75	72.7 (8.0)	81.2 (8.5)	75.3 (2.4)	86.6 (4.2)
	2×37	75.7 (5.6)	84.7 (6.5)	75.2 (2.6)	86.3 (4.1)
	3×25	76.5 (4.4)	85.9 (5.6)	75.3 (2.7)	86.6 (4.2)
	5×15	77.7 (3.9)	87.1 (4.7)	75.3 (2.6)	86.4 (4.5)
	7×11	77.4 (3.9)	86.8 (4.8)	75.3 (2.5)	86.6 (4.3)

in Table 3.

Based on the optimum number of 5 selected crosses, additional simulations were done to investigate the influence of the number of F_2 plants (and accordingly ES-F_3 and SSD-F_6 lines) on selection. Results for $H^2=0.10$ are given in Table 4.

Additionally, cross prediction is compared to a situation where merely line selection is performed in the F_3 (comparable to selection in the SSD-F_6). This is done for various combinations of different heritabilities and selection intensities. Results of both ES and early line selection (ELS), together with corresponding results from the SSD procedure are given in Table 5. The results for $H^2=0.05$, to-

Table 4. Relative genotypic mean and maximum of the 10 phenotypic best lines in both the ES-F_6 and the SSD-F_7 and the accompanying costs. Selection of a variable number of F_2 plants with 5 crosses selected in the ES-F_3. $n=100$, $SelF_4=15\times10$, $SelF_5=25$, $SelF_6=SelF_3$, $H^2=0.1$, 200 runs.

		ES			SSD		
NF_2	$SelF_3$	mean	max	costs	mean	max	costs
10	5×10	74.2	83.3	73250	70.6	81.5	61248
25	5×20	77.6	86.3	75965	74.6	84.6	64170
50	5×30	79.3	87.5	78990	76.7	86.8	67540
100	5×40	80.8	89.0	82790	78.4	87.9	72030
200	5×40	81.5	89.6	85890	79.3	88.5	76510

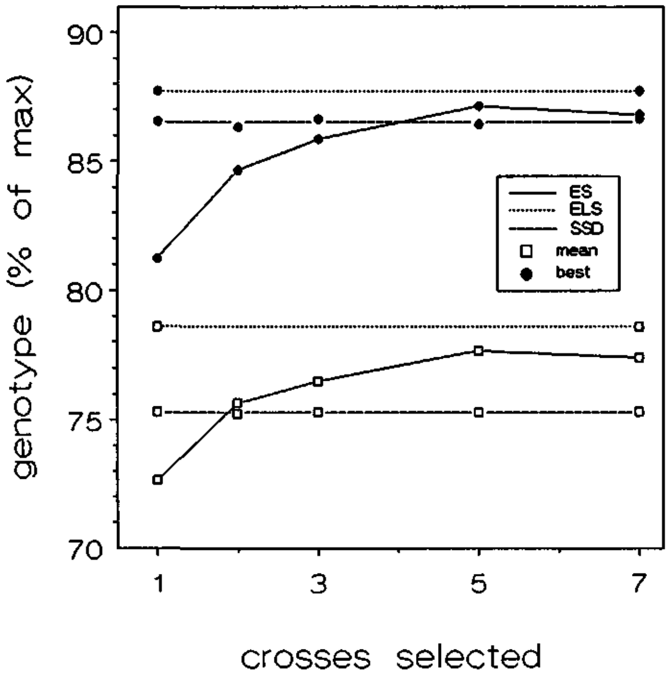

Fig. 1. Relative selection result in the F_6 and F_7 of resp. the ES and SSD procedure as a function of the number of crosses selected (out of the 10). The selection result is indicated by both the genotypic mean and maximum of the phenotypic 10 best lines, expressed as a percentage of the difference between the theoretical maximum and minimum. $SelF_3 = 75$. $H^2 = 0.05$, $NF_2 = 100$, $SelF_4 = 15 \times 10$, $SelF_5 = 75$, $SelF_6 = 150$.

gether with the corresponding data from Table 3 are depicted in Fig. 1.

Based on the same model, the optimum number of ES-$SelF_5$ and SSD-$SelF_6$ was determined. The number of selected crosses was three so $SelF_3 = 3 \times 25$. $SelF_4$ was kept constant to 15×10. $SelF_5$ and $SelF_6$ were varied from 10 to 75 and 20 to 150 respectively. Results are given in Fig. 2.

Table 5. Genotypic mean and maximum of the 10 phenotypic best lines from the ES procedure, an early line selection procedure (ELS) and SSD at various situations. Expressed as a percentage of the difference between the theoretical maximum and minimum genotype. $NF_2 = 100$, $n = 200$, $SelF_4 = 15 \times 10$, 200 runs.

		ES		ELS		SSD	
H^2	$SelF_3/F_5/F_6$	mean	max	mean	max	mean	max
0.05	75/75/150	see Table 3		78.6	87.7	75.3	86.5
0.10	75/75/150	see Table 3		81.7	89.3	78.4	88.0
	3x25/25/100	79.8	88.0	80.4	88.9	78.1	87.7
0.25	3x25/25/100	83.6	90.3	84.4	91.2	81.1	88.8
0.5	2x37/76/150	87.1	92.2	87.8	92.9	83.0	89.4

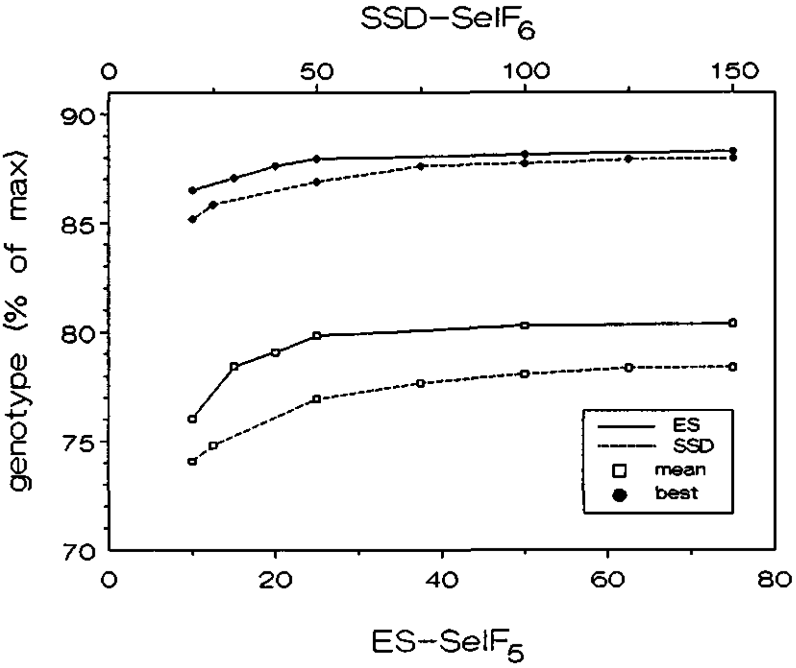

Fig. 2. Relative selection result of both selection procedures as a function of the number of lines selected in the ES-F_5 and the SSD-F_6. $H^2=0.10$, $NF_2=100$, $SelF_3=3\times 25$, $SelF_4=15\times 10$.

Genotype-year interaction

The effect of genotype-year interaction was determined by changing the level of interaction. For both a heritability of 0.10 and 0.25, with three crosses selected in the ES-F_3, results are given in Table 6. The degree of interaction was varied for both heritabilities to the same extent, which effected in a interaction by environmental error variance ratio of 0 to 0.46 for $H^2=0.10$ and 1.39 for $H^2=0.25$. The results from both selection procedures, as a function of this ratio for $H^2=0.25$, are depicted in Fig. 3.

Economical analysis

The three formulas obtained from Table 1 enable a calculation of selection result per unit of cost. This ratio will be referred to as 'relative gain'. This relative gain was simply derived as the genotypic line mean (in percentage) divided by the total costs. The costs for the ES and the SSD procedure in dependence of NF_2 is given in Table 4. The relative gain is plotted against NF_2 in Fig. 4. The same is done for

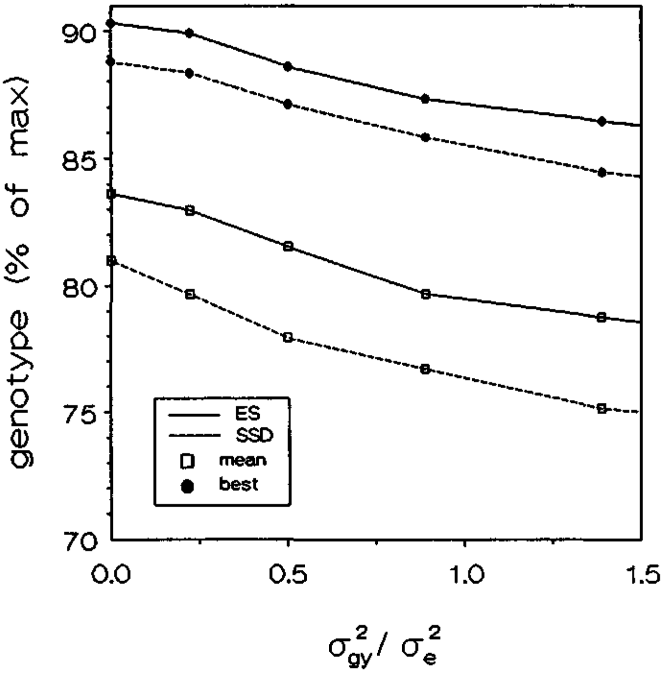

Fig. 3. Relative selection result of both selection procedures in dependence of a genotype-year interaction effect. The level of interaction is given as the interaction by environmental error variance ratio. $H^2 = 0.25$, $NF_2 = 100$, $SelF_3 = 3 \times 25$, $SelF_4 = 15 \times 10$, $SelF_5 = 25$, $SelF_6 = 100$.

a situation with varying heritability. This is depicted in Fig 5.

Another way to compare both selection methods is to plot the selection result against the costs. This is done in Fig. 6 for a varying NF_2 with 5 crosses selected.

Table 6. Relative genotypic mean and maximum of the 10 phenotypic best lines in both the ES-F_6 and the SSD-F_7. Influence of different levels of genotype-year interaction at two heritabilities. The standard errors of both the mean and maximum are given between brackets. $NF_2 = 100$, $SelF_3 = 3 \times 25$, $SelF_4 = 15 \times 10$, $SelF_5 = 25$, SSD-$SelF_6 = 100$. 400 runs.

	σ^2_{gy}/σ^2_e	ES		SSD	
		mean	max	mean	max
$H^2 = 0.1$	0	79.8 (4.1)	88.0 (5.1)	78.1 (2.3)	87.7 (3.9)
	0.07	78.9 (4.3)	87.4 (5.1)	76.9 (3.0)	86.9 (4.6)
	0.17	78.0 (4.8)	86.7 (5.3)	75.7 (3.9)	86.0 (5.0)
	0.30	77.0 (5.3)	85.8 (5.9)	74.2 (4.9)	84.9 (6.1)
	0.46	76.4 (5.6)	85.1 (6.8)	73.4 (5.3)	84.3 (6.7)
$H^2 = 0.25$	0	83.6 (3.0)	90.3 (4.1)	81.0 (1.9)	88.8 (3.6)
	0.22	83.0 (3.7)	89.9 (4.2)	79.7 (2.8)	88.3 (4.0)
	0.50	81.5 (4.5)	88.6 (4.9)	78.0 (3.9)	87.1 (5.0)
	0.89	79.8 (5.5)	87.3 (6.4)	76.7 (5.0)	85.8 (6.2)
	1.39	78.8 (5.8)	86.5 (6.3)	75.2 (5.8)	84.5 (7.3)

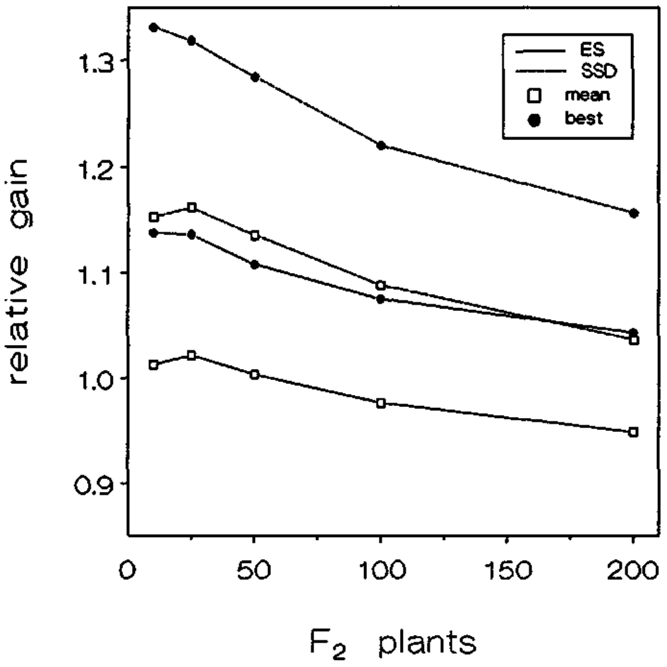

Fig. 4. The selection results of the two procedures per unit of cost (=relative gain) as a function of the number of F_2 plants per cross, with 5 crosses selected. $H^2 = 0.10$.

Conclusions and discussion

It appears from Table 3 and Fig. 1 that an optimum exists for the number of selected crosses in the ES procedure. As was indicated by the previous study (Van Oeveren & Stam, 1992), it is inefficient to select only one cross, due to unreliable cross prediction. It is better to propagate more crosses; with an optimum of five, when ten crosses are involved. This optimum will depend on the total number of crosses, the number of lines per cross (NF_2) and the heritability. At the optimum, ES results in higher yielding lines than SSD. With less than four crosses selected SSD produces the best maximum genotype. The early line selection method outyields both the other procedures in all given situations, as can also be concluded from Table 5. Ergo, cross prediction is of little value, if any.

It can be seen from Table 3 that the standard errors of both the genotypic mean and maximum of the 10 best lines are high. Especially in case of early selection of just one cross. This is partly due to occasional erroneous cross selection. It is in accordance with the fact that field trials concerning comparisons between different selection methods lead to contrasting results.

In the previous simulation study, the effect of varying NF_2 on the selection result was already discussed (Van Oeveren & Stam, 1992). However, this concerned selection within a single cross. There seems to be no different reaction to

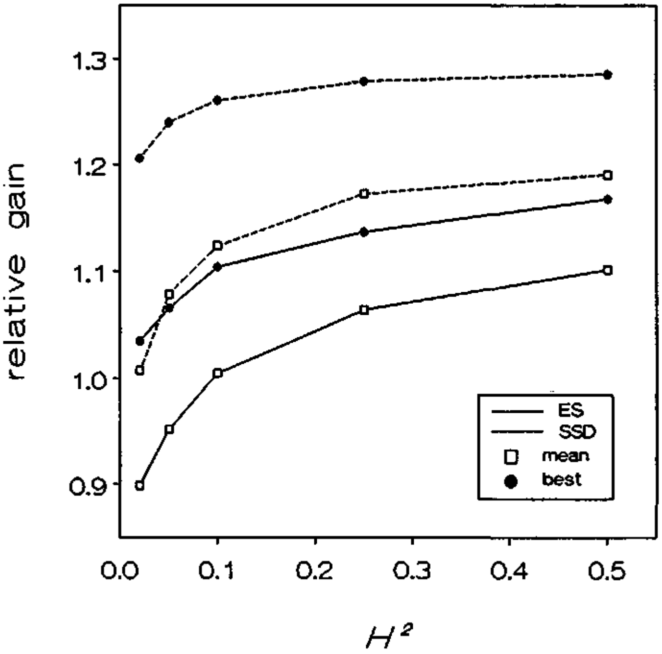

Fig. 5. The relative gain of both selection procedures in dependence of the heritability. NF_2 = 100, $SelF_3$ = 2×37, $SelF_4$ = 15×10, $SelF_5$ = 76, $SelF_6$= 150.

changes of NF_2 when selecting five out of ten crosses (Table 4).

It can be seen from Fig. 2 that selection results get slightly better when more lines are selected in the late generations. The incline is relatively large at the left side of both curves, but beyond ES-$SelF_5 = 25$ and SSD-$SelF_6 = 100$ there is almost no progress. Therefore, in additional simulations where NF_2 is taken 100, $SelF_5$ and $SelF_6$ were kept equal to 25 and 100 respectively.

From Table 6 and Fig. 3 it shows that introducing a genotype-year interaction does not cause a differential reaction pattern for the two selection procedures. Of course there is a decline in selection result when the level of interaction increases. Selection obviously becomes less efficient in both procedures in an equal way. This decline seems to be somewhat larger when heritability is higher (Table 6). At low heritabilities the interaction effect may be partly overshadowed by a large environmental error variance. Again it can be seen that standard errors are large. This effect increases when the interaction factor gets larger. Because ES and SSD do not react differentially to genotype-environment interaction, it was not included in additional simulations.

When the relative gain is considered as a function of NF_2 (Fig. 4), it appears that SSD is more efficient in creating well performing inbred lines than ES. A somewhat lower selection result is more than compensated by a much cheaper proce-

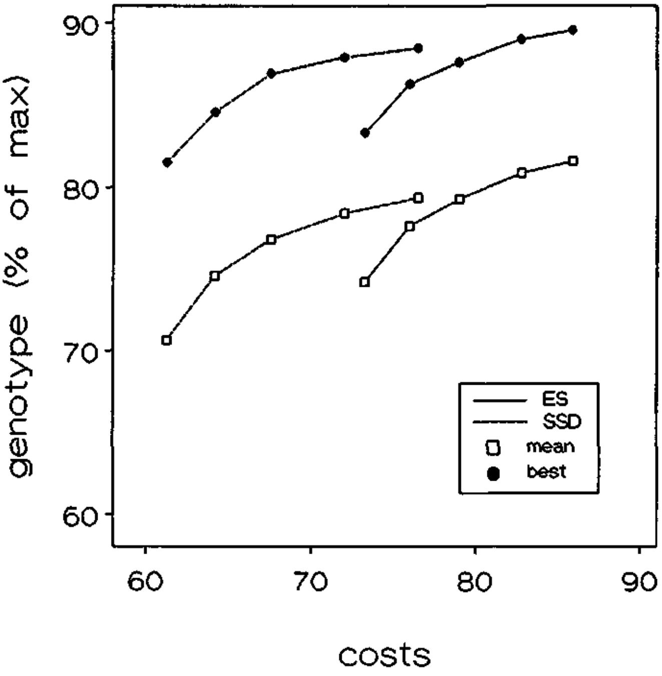

Fig. 6. Selection results (genotypic line mean) of both procedures as a function of the costs (in thousands of dutch guilders) for varying number of NF_2 (see Table 4). $H^2=0.10$, $SelF_4 = 15\times10$, $SelF_5 = 25$.

dure. Furthermore, the optimum gain seems to be reached when very few F_2 plants are taken. The absolute selection result in those cases though are relatively low and probably not sufficient for the breeder. The difference between ES and SSD gets smaller when NF_2 gets larger. This is caused by a relatively large increase in input for SSD when more F_2 plants are taken. The same can be concluded from Fig. 5. SSD appears to be more economical for all heritabilities, although ES becomes relatively more efficient when heritability increases.

Fig. 6 gives another view on the fact that SSD can produce high performing lines in a cheaper way than ES. It appears though that, with maximum input, ES will produce better lines than SSD. Which procedure is more preferable depends on the effort the plant breeder is willing to spend on a relatively small gain.

It is important though to acknowledge the fact that the given costs are only an indication for the respective procedures. It is impossible to obtain truly comparable figures because the costs will depend strongly on the methods used by the individual breeder. Therefore, costs will vary between different breeding programmes (see Table 1). One can consider a situation where the number of F_7 type fields for both procedures are kept equal to each other (as has been in most of the previous simulations) as well as NF_2. When the fixed costs for both field type generations fall below *f*10,000.-, while the other costs remain equal to those in Table 1, ES will cost less than SSD. Accordingly, when all remaining costs are equal to Table 1, ES will get cheaper when the variable costs of SSD will exceed 120% of the costs of each F_3 type field. Thus robustness of the economical comparison appears to be rather large.

A second remark concerns the limited number of crosses. In practice it is not unusual to have 100 or more crosses included in the breeding programme. A (correct) early cross selection of five out of 80 or more, would be cheaper (with the given costs) than an SSD procedure, propagating all 80 crosses to the F_6. An alternative could be to visually select crosses on their F_1 or F_2 appearances and discard the non-promising ones.

On the other hand, the length of the two breeding procedures was not taken into account. SSD can be at least one year quicker and, although it is hard to give it an exact figure, this year can be of inestimable value to the breeder.

Acknowledgements

I am very grateful to ir L.J.M. Groenewegen from Zelder and ir M.E. Roothaan from VanderHave for their willingness to supply the data for the economical analysis. Dr Piet Stam is thanked for his advice and the critical reading of the manuscript.

References

Boerma, H.R. & R.L. Cooper, 1975. Comparison of three selection procedures for yield in soybeans. Crop Science 15: 225-229.

Casali, V.W.D. & E.C. Tigchelaar, 1975. Computer simulation studies comparing pedigree, bulk and single seed descent selection in self pollinated populations. J. Am. Soc. Hort. Sci. 100: 364-367.

Fairfield Smith, H., 1938. An empirical law describing heterogeneity in the yields of agricultural crops. J. Agr. Sci. 28: 1-23.

Jinks, J.L. & H.S. Pooni, 1980. Comparing predictions of mean performance and environmental sensitivity of inbred lines based upon F_3 and triple test cross families. Heredity 45: 305-312.

Knott, D.R. & J. Kumar, 1975. Comparison of early generation yield testing and a single seed descent procedure in wheat breeding. Crop science 15: 295-299.

Mather, K. & J.L. Jinks 1971. Biometrical genetics. Chapman and Hall, London, 382 pp.

Oeveren, A.J. van, 1992. A comparison between single seed descent and early cross selection in wheat breeding. Euphytica 58: 275-287.

Oeveren, A.J. van & P. Stam, 1992. Comparative simulation studies on the effects of selection for quantitative traits in autogamous crops: early selection versus single seed descent. *Heredity.*

Ooijen, J.W. van, 1989a. The predictive value of quantitative genetic parameters in autogamous crops: bias caused by intergenotypic competition. 1. F_∞-mean. Euphytica 44: 81-94.

Ooijen, J.W. van, 1989b. The predictive value of quantitative genetic parameters in autogamous crops: bias caused by intergenotypic competition. 2. F_∞-variance. Euphytica 44: 95-108.

7. SUMMARY AND DISCUSSION

Two selection procedures are examined and compared in the breeding for quantitative traits in self-fertilizing crops. They represent two more or less extreme breeding schemes:

a) Early Selection (ES), with early generation cross selection, and

b) Single Seed Descent (SSD), where selection is postponed to a more homozygous generation.

In the early selection procedure, F_3 estimates of the cross mean and between line variance are used as predictions for the distribution of inbred lines to be derived from that cross in the F_∞. Based on the predicted probability of superior inbred lines, a selection between crosses is performed in the F_3. In subsequent generations line selection is performed. Provided a correct identification of the superior genotypes, this early selection should reduce the risk of losing desirable genes. A disadvantage is that much time and labour is required on the early generation trials.

In the single seed descent procedure a quick advancement towards the F_5 is combined with line selection only in the F_6, regardless of the pedigree. Advantages are the relative speed of the procedure (three generations in one year) and the accuracy of line estimates in the first selection cycle.

In this study the final evaluation is made between ES-F_6 lines, derived from single F_4 plants, and SSD-F_7, derived from single F_5 plants. Considering the time schedule, the size of the plots and their level of heterozygosity, those are the most logical generations to compare. The comparison between the two procedures comprises the difference between early and late generation selection as well as the effectiveness of selection among crosses. The study involves both field trials and computer simulation studies.

Bias on predictions of the genetic parameters

Two 2-year field trials were performed, using mixtures of spring wheat varieties to simulate segregating generations (chapters 2, 3 and 4). The constituent varieties were also grown in monoculture to enable the assessment of the genetic parameters in the F_∞. In the first year of both trials, an ES-F_3 and an

SSD-F_6 generation were grown to predict the same parameters. Results showed that the F_3 predictions of the cross mean and the additive genetic variance were severely biased.

When a comparison is made between the predictions of the genetic parameters from both the ES-F_3 and the SSD-F_6 mixtures and the F_∞ values the various sources of bias can be detected and analysed. The main sources of bias reflect differences between the F_3 generation and the predicted F_∞, where all possible fully homozygous genotypes are equally represented. They include:

- bias due to the simplifying assumptions of the genetical model. Parameters predicted from the F_3 and those estimated from the F_∞ differ by non-additive genetic effects, which are neglected. They consist of dominance (h/H) and epistasis (i/I),
- differences in growing conditions. The F_3 consists of mixtures of genotypes, grown on small plots, whereas the F_∞ consists of large plots of genotypes in a pure stand. Intergenotypic competition and residual plot errors contribute to these differences.

Other possible sources of error are:

- genetic sampling and
- genotype-environment interaction.

Dominance and epistasis

The F_∞ mean is overestimated by the F_3 prediction by $\frac{1}{4}$h and by the F_6 prediction by only $^1/_{32}$h. Thus, dominance will cause a larger bias on the F_3 predictions than on the F_6. This is confirmed by results from both field trials (see chapter 4, Table 10). In the first trial the influence of dominance on the total bias was clearly present. In the second trial its influence was much less obvious.

The genetic variance of the F_∞ is predicted by twice the F_3 between line variance and is overestimated by $\frac{1}{8}$H and underestimated by $\frac{1}{2}$I. Accordingly, the F_6 predictions overestimate by $^1/_{64}$H and underestimate by $^1/_{16}$I. Indeed there was substantial influence of the non-additive effects on the F_3 predictions and none whatsoever on the F_6 predictions. The relative roles of dominance and epistasis in this influence are difficult to establish.

Van Ooijen (1989a; 1989b) found much less influence of dominance and epistasis in comparable studies. Obviously the importance of these non-additive effects on the bias depends on the composition of the simulated crosses. Biasing effects on the mean are likely to be smaller when more loci are involved.

Unless dominance on all loci is unidirectional, the summed effect will be relatively small compared to the sum of all additive effects. However, the effect of both dominance and epistasis on the variance may be large when many loci are involved, because the signs of the individual effects are irrelevant. Indeed the correlations between m_{F3mo} and $m_{F\infty mo}$ from the 3-loci trial are higher than those from the 2-loci trial, whereas they are similar for D_{F3mo} and $D_{F\infty mo}$.

This single comparison is a somewhat narrow basis for general conclusions; however, comparable results have been obtained from the simulation studies. It appears from the results in chapter 5 that dominance is of little influence on the final selection results, comparing the 10 phenotypic best lines from both the ES-F_6 and the SSD-F_7. A tendency was found towards a higher selection frequency of crosses with a high level of between line variance when dominance was present. This is only favourable if the crosses with large variances are also the most promising crosses. Otherwise, the overestimation of D may lead to erroneous cross selection and thus to negative effects on the selection results. At low heritability this effect was negligible.

Growing conditions

The biasing effect of growing conditions is obviously of more importance. In case of small grain crops like wheat, the size of F_3 plots is limited by the amount of F_2 seed. The high level of heterogeneity among plants within plots and between plots, together with a small plot size, causes competition effects and environmental error to be large (see chapter 4, Table 10). Their joint influence is great, especially inflicting on the predictions of the genetic variance. Except for the mean in the second field trial (1990), bias on the F_3 predictions was in all cases larger than the bias on F_6 predictions. With plot sizes being the same for the F_3 and the F_6, this difference can only be due to different levels of intergenotypic competition. As the F_6 lines will be much more homogeneous, the within-plot competition will be much smaller than in the F_3. On the other hand, differences between F_6 lines will be more pronounced, thus between-plot competition is expected to be larger in the F_6.

As was derived from earlier studies (A.J. van Oeveren, unpublished) it is expected that, considering 3-row plots, within-plot competition is negligible compared to between-plot competition. Therefore, only competition between plots was effected in the simulation model. This resulted in higher effects of competition in the F_6 than in the F_3, because genetic differences between lines are more pronounced in the F_6. Additionally, simulation results display a much smaller influence of competition. Especially when heritability is low ($h^2<0.25$) it is largely overshadowed by the environmental effects between plots. On this point there is a discrepancy between results from the field trials and results from the simulation studies. This may have two causes. Firstly, the field trials may be obscured by occasional random effects. This could explain the fact that not all comparisons point in the same direction. Secondly, within-plot competition could be more important than was assumed in the simulation model.

From the first trial an estimate was obtained for the effect of mere plot size (chapter 2). Although plot size had a considerable effect on the bias, the effect of competition appeared to be even larger. These findings agree with results from Van Ooijen (1989a; 1989b), who concluded that intergenotypic competition is the main source of bias.

Genetic sampling

An additional error can be formed by genetic sampling when the number of F_2 plants (= F_3 lines) per cross is limited. Results from chapter 4 showed, in case of 42 lines per cross in the 3-loci model, that this is of only minor influence. Sampling errors are likely to become larger when more loci are involved. The probability of losing a specific genotype will then be higher.

Genotype-environment interaction

Another important form of bias is the effect of genotype-environment interaction, including both genotype-year and genotype-location interaction. A genotype selected as best performing in one year may well be ranked much lower in another year. This can lead to a strong reduction in selection result as is clearly shown by simulation results from chapter 6. One might expect that this will cause a disadvantage to the ES. A high level of interaction together with a relatively high selection intensity, could easily cause the discarding of the most promising crosses in the F_3. On the other hand ES offers the opportunity for selection in several successive generations. Thus the identification of genotypes performing well in various environments is enabled, provided selection intensity is kept relatively low. Perhaps there was a levelling out of both possible effects, because in the simulation results the reduction appeared to be of the same relative magnitude for both ES and SSD.

A high level of genotype-year interaction was found from the field trials.

Correlations between variety yields in successive years were low and in one case even non-significant. Because both the selection and the F_∞ environment were set in the same year and on the same location, interaction had no influence on the predictions of the genetic parameters.

Comparing ES with SSD

The two 2-year field trials showed better results from SSD compared to ES. The F_3 predictions of the cross mean and additive genetic variance were severely biased in both trials. In the first one this led to an erroneous cross prediction, resulting in the discarding of the potentially best cross (chapter 2). In the second trial (chapter 4) bias from growing conditions was less severe. Additional coincidence may have caused the good agreement between the ranking of the crosses from both the F_3 and F_∞ situations. Indeed the better crosses were retained in the ES procedure. However, postponed selection in the SSD procedure appeared to be even more accurate, and the 10 best SSD-F_7 lines showed a much higher genotypic mean than the 10 best ES-F_6 lines, although the maximum genotype derived from both procedures did not differ much (chapter 3 and 4).

The two field trials suggested a more positive view concerning SSD than the simulation studies (chapter 5 and 6). This will be partly caused by the fact that no line selection was performed in the ES-F_3 to the ES-F_5 of the field trials. Considering a situation with many loci determining the trait of selection, there will most likely be a positive effect of line selection in the intermediate generations of the ES procedure. This may lead to better selection results from ES, as is indicated by the simulation studies. The success of the ES procedure appears to depend heavily on the number of selected crosses in the F_3. Whereas three out of fifteen crosses were selected in both field trials, simulation showed that it is very inefficient to select only the one best cross out of five or ten. In that case SSD will lead to better results. However, when three or more crosses are selected, the selected lines of ES may outdo those of the SSD. Considering ten crosses under selection, there appeared to be an optimum at five selected crosses. Early cross prediction was also compared to an early line selection method (ELS) where no selection between complete crosses is made, but merely among all lines (like in the SSD-F_6). It appeared that ELS in all cases is more efficient than ES, although the difference is small when the optimum number of crosses is selected in ES. This is another indication that early cross selection is not efficient (chapter 6).

Additionally, results from both breeding procedures were considered in relation to the heritability, the number of F_2 plants (= ES-F_3 and SSD-F_6 lines) and the number of segregating loci, which seem to be the three main controlling factors. As expected, results for both selection procedures were higher when heritability was higher and more F_2 plants were taken. ES is relatively more efficient when h^2 is high and the number of F_2 plants is low. Conversely, SSD becomes more efficient when h^2 is low and many F_2 plants are taken. The number of segregating loci that characterize the trait appears to have a large impact on the selection response. A much better line can be retained if only 10 loci are involved (90% of the maximum versus 70% with 25 loci). This is not surprising, because the chance of retrieving a genotype with all positive alleles accumulated will be much smaller when 25 instead of 10 loci are segregating. Accordingly, the absolute differences between ES and SSD are slightly smaller when more loci are involved (chapter 5).

Simulation results also show that final selection results have large standard errors. When ES leads to better results in one run, SSD can easily turn out better in the next. Only numerous replications can give a clear picture of which method is actually best. This once again shows that differences are not large. It also explains the phenomenon of contrasting results from comparable field trials, as mentioned in the introduction of this thesis. Standard errors are especially large in ES, when few crosses are selected. Among SSD results they are relatively small and therefore SSD proves to be a more consistent procedure than ES.

Economic comparison

The previous comparisons all concerned breeding schemes with the total field areas occupied by both procedures being approximately equal. It is useful to include the accompanying costs in this comparison. This is done in chapter 6 and it shows that SSD is a much more economic procedure. The relative gain (selection result divided by the costs) is much higher for SSD than for ES. It appears though that, with maximum input, ES will produce better lines than SSD. Which procedure is more preferable therefore depends on the effort the plant breeder is willing to spend on a relatively small genetic gain.

Conclusions

When the results of the previous chapters are put together it becomes clear that early cross selection (ES) is not an efficient way of breeding. Cross prediction will often be erroneous due to severe bias on predictions of the genetic parameters. There may be a substantial amount of bias due to the neglect of the non-additive genetic effects, but the main source of error is the difference in growing conditions between the F_3 selection environment and the predicted F_∞ environment. The latter mainly consists of intergenotypic competition.

The procedure of single seed descent (SSD) can produce superior inbred lines in a more consistent, cheaper and faster way. Early line selection (ELS; without cross selection) is always better than ES. In some cases ELS can, when heritability is intermediate, give a breeder a somewhat higher chance than SSD on obtaining the best possible genotype. This will require additional costs.

References

Ooijen, J.W. van, 1989a. The predictive value of quantitative genetic parameters in autogamous crops: Bias caused by intergenotypic competition. 1. F_∞-mean. Euphytica 44: 81-94.

Ooijen, J.W. van, 1989b. The predictive value of quantitative genetic parameters in autogamous crops: Bias caused by intergenotypic competition. 2. F_∞-variance. Euphytica 44: 95-108.

SAMENVATTING

Dit proefschrift behandelt een onderzoek naar een deel van de mogelijkheden en moeilijkheden van selectie bij de veredeling van zelf-bevruchtende gewassen. Het doel van de plantenveredelaar is om nieuwe rassen te ontwikkelen door eerst genetische variatie te creëren en vervolgens uit die grote verzameling genotypen de beste te halen door middel van selectie. Bij zelfbevruchters wordt de genetische variatie gecreëerd door het onderling kruisen van twee of meer volledig homozygote ouderlijnen. Deze lijnen vertonen elk één of meer gunstige eigenschappen en in de nakomelingschap wordt gezocht naar het genotype waarin al deze eigenschappen verenigd zijn. Uit de kruising tussen de ouderlijnen ontstaat de F_1 hybride, die na zelfbevruchting een uitsplitsende F_2-generatie oplevert. Na verscheidene rondes van herhaalde zelfbevruchting ontstaan weer homozygote genotypen: de potentiële nieuwe rassen. Om uit die groep genotypen de allerbeste te kunnen halen dient er geselecteerd te worden. Het probleem is nu om in elke generatie genoeg planten aan te houden zodat de beste genotypen aanwezig zijn en ook herkend kunnen worden. Aan de andere kant zijn er economische factoren om de omvang van het veredelingsprogramma te beperken.

Een tweede probleem is het moment van selectie; is het effectief om in een vroege generatie (b.v. F_3) te selecteren, als de planten nog sterk heterozygoot zijn en hun nakomelingschappen daardoor heterogeen? Of is het beter om selectie uit te stellen tot een latere generatie met zeer sterk homozygote planten en homogene nakomelingen? De eerste optie heeft als voordeel dat geselecteerd wordt in een stadium met relatief geringe genetische variatie tussen lijnen. Zodoende is de kans om het superieure genotype te missen klein en is het nog niet nodig om een overdreven grote populatie aan te houden. Het voordeel van uitstel van selectie is dat, in het geval van kwantitatieve eigenschappen, de schattingen van plot-totalen veel nauwkeuriger zijn dan in een vroege generatie. Bovendien is het geselecteerde materiaal min of meer genetisch stabiel.

Twee veredelingsschema's zijn in dit onderzoek onderzocht en vergeleken, als voorbeeld van bovenstaande, contrasterende methoden:

a) Early Selection (ES: vroege selectie), en

b) Single Seed Descent (SSD), waarbij selectie uitgesteld wordt tot een latere generatie.

In de vroege selectie procedure worden in de F_3 schattingen gedaan van het kruisingsgemiddelde en de tussen-lijn variantie. Deze schattingen worden gebruikt voor een voorspelling van de verdeling van inteeltlijnen die uiteindelijk uit die kruising verkregen kunnen worden (in de F_∞). Er wordt daarna een selectie tussen kruisingen gemaakt, gebaseerd op de voorspelde kans op het vinden van superieure genotypen. In de navolgende generaties kan lijnselectie uitgeoefend worden. Bij SSD wordt uitgegaan van dezelfde F_2, waarna een snelle homozygotering tot de F_5 volgt. Lijnselectie vindt plaats in de F_6 ongeacht de afstamming. De twee veredelingsprogramma's staan schematisch afgebeeld in figuur 1 van hoofdstuk 1.

De vergelijking tussen deze twee uiteenlopende procedures heeft betrekking op enerzijds het verschil tussen *selectie in vroege en late generaties* en anderzijds de effectiviteit van kruisingsselectie. Deze studie behelst zowel veldproeven als computer-simulaties.

Twee 2-jarige veldproeven zijn uitgevoerd met rassenmengsels van zomertarwe welke uitsplitsende generaties nabootsen. De F_3-voorspellingen van het gemiddelde en de additieve genetische variantie van elke kruising blijken in beide veldproeven sterk vertroebeld. Deze vertroebeling wordt veroorzaakt door de volgende foutenbronnen:

- Non-additieve genetische effecten, die met de veronderstellingen van het genetische model verwaarloosd worden. Ze bestaan uit dominantie en epistasie.
- Verschillen in de teeltomstandigheden tussen het selectie-milieu (F_3) en het voorspelde milieu (F_∞). Deze teeltomstandigheden hebben invloed op de grootte van milieu- en concurrentie-effecten.
- Toevalseffecten veroorzaakt door het telen van slechts een steekproef van genotypen in elke generatie.
- Genotype-milieu interactie.

Bij het vergelijken van de voorspellingen met de echte waarden van de genetische parameters, blijkt dat er een duidelijk verstorend effect is van dominantie en waarschijnlijk ook van epistasie. Dit kan gedeeltelijk te wijten zijn aan het gebruikte model van slechts enkele loci die de eigenschap bepalen. Uit de computer-simulatie met veel loci blijkt nauwelijks invloed van dominantie.

Het effect van verschil in teeltomstandigheden op de voorspellingen is zeer sterk. Omdat het selectie-milieu bestaat uit kleine veldjes met een sterk heterogene samenstelling, is er een grote invloed van milieuvariatie en vooral concurrentie-effecten. Het F_∞-milieu daarentegen bestaat uit grote velden met planten in monocultuur, waardoor die effecten nauwelijks een rol spelen.

Toevalsvariatie in de verdeling van genotypen blijkt nauwelijks van invloed te zijn op de juistheid van de voorspellingen in het in deze studie gebruikte model.

Genotype-milieu interactie echter kan een zeer groot verstorend effect hebben, zoals blijkt uit de computer-simulaties. Een genotype dat het goed doet in het ene milieu kan gemakkelijk veel slechter scoren in een ander jaar en/of op een andere locatie. Op de voorspellingen in deze veldproeven was dit echter van geen invloed aangezien het selectie-milieu en het voorspelde milieu in hetzelfde jaar op dezelfde locatie werden gerealiseerd.

In de eerste veldproef leidden de bovenstaande verstorende invloeden op de voorspellingen van de genetische parameters tot een foutieve kruisingsvoorspelling. Dit resulteerde in het verloren gaan van de potentieel beste kruising. In de tweede proef was met name de verstorende invloed van concurrentie in de F_3 minder groot, resulterend in een gering verschil tussen de voorspelde rangorde van de kruisingen en hun F_∞ rangorde. Zodoende werden de beste kruisingen inderdaad geselecteerd. In beide gevallen bleek echter de selectie in de SSD-procedure veel nauwkeuriger: de 10 beste lijnen uit de SSD vertoonden een veel hoger genotypisch gemiddelde dan de 10 beste ES lijnen. Ook het allerbeste genotype was bij SSD beter dan bij ES, hoewel het verschil minder groot was.

Uit het simulatie-onderzoek kwam een iets minder positief oordeel over SSD naar voren. Dit komt gedeeltelijk doordat in de veldproef geen rekening is gehouden met een eventueel positief effect van lijnselectie in de ES-generaties. In het simulatie-onderzoek is dit wel gemodelleerd en blijken de resultaten van ES dichter bij SSD te liggen of deze zelfs te overtreffen.

Indien slechts één kruising uit 5 of 10 geselecteerd wordt in de ES-F_3, blijkt het selectieresultaat van SSD veel beter te zijn dan dat van ES, doordat vaak de verkeerde kruising geselecteerd wordt. Worden meerdere kruisingen geselecteerd, dan blijkt er een optimum (5 uit 10) te zijn en in dat geval is ES beter dan SSD. Wordt ES vergeleken met een methode waar wel vroege lijnselectie wordt toegepast, maar niet tussen complete kruisingen wordt geselecteerd, dan blijkt deze laatste methode het nog net iets beter te doen, zelfs bij het voor ES optimale aantal geselecteerde kruisingen. Vroege selectie tussen kruisingen lijkt daarom niet erg zinvol.

Voorts is een vergelijking gemaakt tussen ES en SSD in afhankelijkheid van heritability (erfelijkheidsgraad), aantal F_2-planten en aantal uitsplitsende loci. SSD bleek relatief efficiënter bij een lage heritability en veel F_2 planten, terwijl ES juist relatief efficiënter was bij een hoge heritability en weinig F_2-planten. Bij een toename van het aantal uitsplitsende loci was een afname van het uiteindelijke selectieresultaat te bemerken. Dit wordt veroorzaakt door het feit dat de kans op extreem gunstige recombinante genotypen bij veel loci veel kleiner is dan bij enkele loci. De verschillen tussen de beide selectieprocedures waren iets geringer bij veel loci.

Er blijkt bovendien dat er een grote variatie bestaat in de selectieresultaten: terwijl de ene keer ES beter is, kan evengoed de volgende keer, onder identieke omstandigheden, SSD beter zijn. Dit toont aan dat de onderlinge verschillen gering zijn. De resultaten bij SSD blijken meer constant te zijn.

Wanneer ook de kosten van de verschillende veredelingsmethoden meegenomen worden, blijkt dat SSD een veel goedkopere procedure is. Het behaalde selectieresultaat per eenheid kosten is bij SSD veel hoger dan bij ES. Echter, bij maximale input en het optimum aantal te selecteren kruisingen, zijn de kansen bij ES op het verkrijgen van een superieur genotype hoger dan bij SSD. Dit vereist dan wel de nodige extra kosten.

De belangrijkste conclusies zijn:

- Vroege kruisingsselectie (ES) in de F_3 is niet effectief. Door verschillende foutenbronnen zijn de voorspellingen van de kruisingen onbetrouwbaar. Dit wordt veroorzaakt door sterk verstorende invloeden van non-additieve genetische effecten en met name tussen-plant concurrentie.
- Single Seed Descent (SSD) is een veel goedkopere en snellere methode dan ES in het kweken van superieure genotypen. Echter, bij maximale input en de bijbehorende extra kosten, kan ES een beter selectieresultaat opleveren.

CURRICULUM VITAE

Jan van Oeveren jr. is geboren op 18 juni 1964 in het Zeeuwse Kortgene. Voor zijn middelbare-school opleiding mocht hij elke dag 30 km fietsen naar het Goese Lyceum. Ondanks zijn meer dan acht gemiddelde op zijn gymnasium diploma werd hij uitgeloot voor diergeneeskunde en belandde toen in 1982 in Wageningen. Daar startte hij de studie plantenveredeling aan de Landbouwuniversiteit, die hij eind juni 1988 succesvol af wist te ronden. Zijn afstudeervakken betroffen de plantenveredeling en de wiskundige statistiek. Een maand daarvoor was hij reeds te werk gesteld als Assistent in Opleiding aan de vakgroep Erfelijkheidsleer van de Landbouwuniversiteit. Na zijn promotie gaat Jan voor een half jaar op avontuur in Afrika.