Institute for Crop Production and Grassland Research
Department of Bioinformatics
Prof. Dr. H.-P. Piepho
University of Hohenheim

Biometrical approaches for analysing gene bank evaluation data on barley (*Hordeum* spec.)

Dissertation
submitted in fulfilment of requirements for the degree "Doktor der Agrawissenschaften"
(Dr.sc.agr. in Agricultural Sciences)

to the Faculty of Agricultural Sciences

presented by
Karin Hartung
born in Langen

Stuttgart Hohenheim, 2006

Tables of Contents

1 Abbreviations

AD	augmented design
ANOVA	analysis of variance
a.v.d.	average variance of a difference
BLUE	Best linear unbiased estimation
BLUP	best linear unbiased prediction
FE	folded exponential transformation
IPK	Institute of Plant Genetics and Crop Plant Research, Gatersleben
LSD	least significant differences
ML	maximum likelihood
P1	percentage rating scale using 1%-steps
P5	percentage rating scale using 5%-steps
R9	ordinal rating scale
PGR	plant genetic resources
QTL	quantitative trait loci
RE	relative efficiency
REML	restricted maximum likelihood
S1	scales based on a descriptive characterization of the trait only
S2	scales based on a underlying percentage or metric scale
S3	scales that are direct percentages themselves

2 General Introduction

One of the largest collections of plant seeds in the world – held at the N. I. Vavilov Institute of Plant Industry (VIR) in St. Petersburg – was created by Nikolai Ivanovich Vavilov (Николай Иванович Вавилов, Nov. 25, 1887 until Jan. 26, 1943), who was a prominent Russian botanist and geneticist and is regarded as the originator of gene banks (Anonymous A, 2006). In the wake of Soviet collecting missions several collectors from different countries appeared including Jack Hawkes, later one of the founders of the worldwide movement to conserve Plants Genetic Resources (PGR). In the 1970s small national gene banks were established around the world (Guarino et al., 1995, p. 1-11). And in 1998 over 6 million accessions were being conserved in more than 1300 gene banks (Koo et al., 2005).

2.1 Gene banks

The size and "organisation" of gene banks today is very diverse. There are huge gene banks like PGRC (Canada), NSGC (USA) or ICARDA (Syria) and small ones which conserve only some local plant species. The Food and Agriculture Organization of the United Nations (FAO) and the World Information and Early Warning System on Plant Genetic Resources (WIEWS) lists about 1,460 gene banks worldwide, including 465 in Europe, 468 in the Americas, and 298 in Asia (Hawtin and Cherfas, 2003). Financial conditions, numbers of employees and equipment are highly variable. Gene banks are financed mostly by governments and there are only few possibilities to raise money from other sources like research funds. Thus, the problem for many gene banks is that they run on small budgets, unsure whether the funding will continue, hoping that no additional costs arise, e.g. from machine damage or accidents (Hawtin and Cherfas, 2003). Even in the developed countries some gene banks do not have the capacity to conduct field trials, so they cooperate with breeders and farmers and leave the cultivation strategy to these partners. Nevertheless, evaluation and characterisation is often done by gene bank staff. In the extreme case the task of a gene banks is just the long-term cold storage of seeds, as is the case on the Norwegian island of Svalbard (Anonymous B, 2006).

The main task of a gene bank is to maintain accessions of crop species to preserve the existing agrobiodiversity for research and breeding. Therefore the aims are conservation of accessions, i.e. maintenance of germinability of seeds, and prevention of gene drift in the collection during seed multiplication (Ortiz, 2002; Anonymous C, 2006; Anonymous D, 2006). Through time germination capacity of seeds decreases, so sowings for reproduction are necessary. Up to the 1980s

it was necessary for cereals to multiply seeds every two to five years, but it is now common to store e.g. barley cooled down to temperatures of -15°C for over 15 years with unchanged fertility (Börner et al., 2000). Today the accessions that need seed reproduction are grown in unreplicated field trials with only few or no checks (standards). And even if checks are used, accessions and checks are normally cultivated without experimental field designs. While the focus is on reproduction, diverse characteristics of the accessions are assessed in these trials. Data of morphological traits are collected such as grain colour, thousand seed weight, plant height, and maturity date. Also, sometimes ordinal evaluation data are available like degree of lodging, resistance to pests and diseases. It is usually impossible to grow all accessions stored in one gene bank together in one year in a homogenous environment. For example the gene bank at the Institute of Plant Genetics and Crop Plant Research, Gatersleben (IPK) has an inventory of 20,000 different barley accessions (private communication, Knüpffer, 2006) and only around 500 plots per year to regenerate them. Overall the IPK stores 147,500 accessions from more than 2,700 plant species and 773 genera. Therefore it is one of the most comprehensive collections in the world and provides a major contribution towards preventing extinction (gene erosion) of both cultivated plants and their related wild species (Anonymous C, 2006).

2.2 Preservation of barley (*Hordeum* spec.)

Barley is the second largest crop represented in gene banks comprising 8% of world's accessions after wheat (13%) (FAO, 1996). Seed storage is relatively easy. Seeds sealed hermetically with a moisture content of 3.1% showed a germination of 90% after 110 years of storage at ambient temperatures (Steiner and Ruckenbauer, 1995). Even if held under open conditions in a temperate condition, seeds maintained germinability above 50% for over 7.2 years (Priestley, 1986). Under cool-storage (-20 to -15°C and 3% to 7% moisture) as recommended for long-term storage by FAO/IPGRI (1994) barley is expected to retain germinability for over 100 years. Barley regeneration is relatively easy for cultivated forms. Pollen contamination is usually very low since it is a self-pollinated crop (Hammer, 1975). Wild species show more problems regarding regeneration (Hintum and Menting, 2003).

The field design for regeneration of barley is very diverse for different gene banks ranging from single rows with lengths of 0.8 to 3 m to plots of a size of around 1.5 m^2 (built of 3 to 4 rows), while rows or plots are separated either by space or by another cereal, leading to a chessboard-like design (c.f. Paper 1, Figure 1). The number of barley accessions cultivated every year depends on the size of the gene bank, availability of equipment and the number of barley

accessions stored. A trial size of several hundred barley accessions seems to be common. In general, when cultivating accessions for rejuvenation, the accessions are regenerated without following an experimental field design. Only in rare cases, i.e. if there is a specific research question, field designs are used. A few gene banks cultivate checks in regularly spaced intervals every year, a larger number of gene banks has at least some replicated checks or accessions, e.g. on border plots (personal communications from several gene banks, 2003).

2.3 Objectives of gene banks

The intention of gene banks, like the IPK, is to improve management of their collections by investigating spatio-temporal patterns of genetic diversity, to analyse the population structures (Anonymous C, 2006), and to contribute to breeding and research programs by providing information about phenotypic traits, thus facilitating an informed choice among the available accessions. To reach the latter objective it is necessary to present the data in such a way that external users can easily find the desired information. This includes ensuring the greatest possible availability of data and information concerning PGR's (Ortiz, 2002), as for example in the European Barley Database at the IPK (Anonymous E, 2006). Another aim is to combine data over years and/or sites to obtain more reliable information. Standardised procedures for obtaining characterisation and evaluation data of accessions have already been recommended, but are not yet binding (IPGRI, 1994; Bundessortenamt, 2000). All these aims should be realisable without any or with only minor changes to the current system.

Furthermore there are different research activities at gene banks. For example at the IPK this includes the optimisation of in vitro and cryo-conservation, the use of DNA fingerprinting technology to monitor the genetic integrity of samples, and the analysis of population structures (Anonymous C, 2006). Identifying unknown duplicated accessions within a collection and between gene banks is important to avoid a waste of resources (Ortiz, 2002). Developing a core collection[1] improves the management and utilisation of a germplasm collection (Knüpffer and Hintum, 2003). Today gene banks benefit from new information technology and powerful computers, resulting in the opportunity to offer specific accessions with information on the relevant characteristics to research geneticists or applied plant breeders (Ortiz, 2002).

1 A core collection is a subset of a large germplasm collection, containing chosen accessions that capture most of the genetic variability in the entire collection.

2.4 Requirements to improve accuracy of information from field reproduction

In order to obtain valid data for a single trait of an accession the trait data assessed in field trials need to comply with several requirements:

(1) A sound and analysable experimental field design is required, comprising repeated entries for at least a certain number of entries. The experimental field design can either follow approaches where every entry has at least two replicates (e.g. incomplete blocks), or only a certain number of checks is repeated (e.g. augmented designs). The replication is necessary to obtain valid estimates of experimental error.

(2) The single trait data that are to be analysed need to be assessed as precisely as possible, preferably on a metric or percentage scale.

(3) If data are to be analysed over years or locations or both it has to be ensured that the data are connected (Searle, 1987, p139), i.e. some entries and/or checks need to be replicated across the trials that are to be analysed jointly.

(4) The data obtained then need to be analysed by a sound model that fits the chosen approach. These analyses can follow randomisation-based models or geostatistical models.

2.5 Problems with statistical analyses arising from field data generation as currently practised by gene banks

Up to now some gene banks spend a few plots to grow check varieties, but they normally do not use any of the standard experimental field designs (personal communication from different gene banks, 2003). With the large number of accessions that need to be grown each year, the most common design in agricultural trials, the complete block design, where standards and cultivars are fully replicated in each complete block, is not feasible (Federer and Raghavarao, 1975). Other designs such as augmented designs need fewer plots and therefore are one option to tackle the problem (Peterson, 1994; May et al., 1989). Another option is to find suitable designs using geostatistical (i.e. spatial) methods (Eccleston, 1998; Watson, 2000; Stroup, 2002). The former option has the advantage that less strong assumptions are needed for analysis than for spatial methods (Schabenberger and Gotway, 2005). But with large block sizes there is often heterogeneity within a block. This heterogeneity is due to competition between entries, heterogeneity of soil, crop diseases and insect dispersion as well as other influences. Thus, the latter option, the use of spatial methods, is more flexible and might handle the problem of complex field heterogeneity more effectively if a good design is found (Schabenberger and Gotway, 2005). In comparison to the unreplicated trials currently used by most gene banks, both sorts of design require additional space and costs associated with check plots.

Field designs and spatial models not only allow to properly analyse accessions of one year but also allow to analyse multi-year data sets if connecting checks or entries are used. Additionally this offers the possibility of combined analysis of different gene bank data provided that the data sets are connected, i.e. similar accessions and/or checks are cultivated. However, since in practice every gene bank cultivates its own checks and accessions in a certain year, it is not guaranteed that trials are connected, so an evaluation of accessions over different environments is usually difficult with data sets currently available.

Another problem – which always arises when assessing characteristics in evaluation trials – is the scale which should be used for measurement. The chosen scale should be appropriate regarding the question under research and the method to be used for analysis of a trial. Both from a statistical point of view regarding analysis and from a gene bank point of view regarding the amount of work, least problematic are traits that are already assessed on a metrical scale. Major difficulties – like unknown or changing thresholds, transformation problems, uncertainty towards statistical evaluation method – arise when data are assessed on an ordinal rating scale, which is less informative than data from a metric scale. In gene banks the majority of traits are visually assessed on ordinal rating scales during reproduction. Within this thesis ordinal rating scales will be subdivided into three groups:

(S1) scales based on a descriptive characterization of the trait only (very high, high, medium, …),

(S2) scales based on a underlying percentage or metric scale, and

(S3) scales that are direct percentages themselves.

Scales based on (S1) and (S2) range for example from 1 to 9. Scales based on (S3) always range from 0 to 100. If a descriptive ordinal rating scale (S1) is used to asses a certain trait, methods for ordinal data are preferable for analysis, such as rank-based methods (Brunner and Langer, 1999) or methods on generalised linear models (Agresti, 1984). If a trait is assessed on an underlying percentage scale (S2) or even better directly on a percentage scale (S3), analysis of variance can be used, even though percentages do not strictly meet the usual assumptions of homogeneity of variance (heteroscedasticity), normality (normal distribution of data), and linearity/ additivity (Thöni, 1985; Schumacher and Thöni, 1990). A further common option, if there is no value of zero or one hundred, is the logit-transformation which could provide data that can be analysed with ordinary statistical methods. The usual way to analyse percentages is to use generalised linear models (McCullagh and Nelder, 1989). With ordinal rating scales that are based on

an underlying percentage scale (S2), specific problems may occur. Thresholds for these ordinal rating scales are not always accurately defined and may change over time. The underlying percentage scale may have clearly defined class thresholds, but the true class means on that underlying scale are usually unknown. For example, let the thresholds be 10 and 20 then the arithmetical mean of 10 and 20 is 15, but the true mean of the class could either be 12 or 18. Furthermore the transformation of ordinal ratings back to percentages or absolute values is always difficult. If ordinal ratings are directly assessed as percentages (S3), the larger number of values with percentages than with ordinal ratings (e.g. one hundred versus nine) is expected to result in more accurate assessments.

Another problem is that ordinal rating scales (S1 and S2) used at a gene bank may change over years. This complicates summary of data per accession for one characteristic (trait) over years. The same problem arises if data are to be combined from several gene banks where different scales are used. For metric data (yield, thousand kernel weight, etc.) there are no such problems. The standard approach for such data is to use an appropriate linear model for the series of trials and to estimate least squares means per accession (Piepho, 2003a). Finally, an important consideration is the required computational capacity, which rises not only with complexity of analysis, but also with the size and quality of the database.

2.6 Topics covered by this thesis

The aims of this thesis were to find a way to efficiently analyse the current barley data (*Hordeum* spec.) of the gene bank at the IPK, as well as to investigate how and where optimisation in gene bank evaluation is possible.

The objective of Paper 1 was to estimate the genetic value of accessions in a single year as accurately as possible. Since field replicates and standards are rare in gene bank data, a geostatistical approach was used, where the empirical variogram was fitted by a non-linear regression using a visually chosen covariance model, followed by a mixed model analysis to estimate genetic effects by best linear unbiased prediction (BLUP).

With long-term data, ordinal rating scales often change. This complicates the joint analysis of data from several years. Thus, in order to allow for a joint analysis of data from different rating scales, a threshold model was developed with a common latent scale for the different rating systems (c.f. Paper 2).

Disease incidence and severity is assessed either on an ordinal rating scale or on a percentage rating scale. Paper 3 compares three different rating scales regarding accuracy, precision, and time needed for scoring. The paper is based on an experiment in which persons with different rating experience were asked to rate virtual leaves on three different scales.

Spatial designs (Kempton and Gleeson, 1997) and augmented designs (AD) are two design options for gene bank evaluation trials with unreplicated accessions. ADs seem particularly worthwhile due to the relatively straightforward analysis. Paper 4 provides a summary of important developments in design and analysis of ADs and sets these into context with gene bank needs.

Paper 5 uses spatial covariance models in order to find an optimal field layout for a simple kind of AD. Particular emphasis is given to the influence of the number of checks per block.

2.7 Data used in this thesis

2.7.1 Phenotypic data

For this thesis the IPK kindly provided all barley data that were electronically documented until the year 2002. These data covered the years 1946 to 1991 and 1993 to 2002 and comprised 13,396 different accessions. With every accession a large number of different traits (characteristics) was available, of which the following traits were considered: growth habit, lodging, plant density before and after winter, plant length, waxy cover of culm and leaves, brown rust, mildew.

For each trait, approximately 10,000 different accessions were available. Most accessions were repeated 1-10 times and the highest number of replicates for a single accession was 84.

2.7.2 A rating experiment

A survey was conducted to assess the quality of different rating scales. For this investigation rating values as well as the time needed for rating were obtained on three different rating scales.

2.7.3 Survey data

An e-mail questionnaire was sent to several gene banks that regenerate barley. The aim was to assess information on experimental designs used and evaluation of data, in particular statistical procedures used for analysis. The gene banks were:

Austria	Österreichische Genbanken für Kulturpflanzen
Croatia	Croatian Plant Genebank - (HBBG)
Canada	Plant Gene Resources of Canada (PGRC)/ Les Ressources Phytogénétiques du Canada (RPC)
Cyprus	National Genebank, Agricultural Research Institute, Ministry of Agriculture, Natural Resources and Environment
France	BRG - Collections de Ressources Génétiques Végétales
Hungary	Génbank Tápiószele
Ireland	Irish Genetic Resources Conservation Trust (IGRCT)
Israel	Israel Gene Bank
Japan	NIAS Genebank
Netherlands	Centrum voor Genetische Bronnen, Netherlands (CGN)
Portugal	Banco Portugues de Germoplasma Vegetal (BPGV)
Romania	Suceava Genebank
Russia	N. I. Vavilov Institute of Plant Industry (VIR)
Sweden	Nordiska Genbanken (NGB)
Switzerland	Station Federale de Recherches en Production Vegetale de Changins

3 Publications

3.1 Paper 1 (Abstract only): Analysis of genebank evaluation data by using geostatistical methods

Hartung K, Piepho H-P, Knüpffer H (2006)
Genetic Resources and Crop Evolution 53:737-751

Abstract

At gene banks several characteristics of accessions are assessed regularly in field trials. But unlike with conventional field trials, replicates and standards are rare. To cope with this problem, geostatistical approaches can be used for analysis. In this study eight characteristics, mostly ratings, of spring and winter barley are analysed. Ratings, with quasi-metric scales, were transformed using the folded exponential transformation. Two methods were compared to estimate the genetic component. Method 1: the variogram is fitted using non-linear regression. Then the selected spatial correlation is embedded into the mixed model analysis to estimate the genetic effect via Best Linear Unbiased Prediction (BLUP). Method 2: All data are re-estimated by kriging to correct for spatial correlation. These data then are analysed using mixed model analysis.

We suggest to use Method 1 (even though occasionally convergence problems occur) to obtain genetic effects via BLUP. The method has the following steps:

- fit the short range of the empirical variogram,
- visually choose the suitable covariance model,
- implement the covariance model and the initial values from non-linear regression fit with the mixed model, fixing the spatial parts at their starting values.

To obtain more valid genetic values, we recommend that, wherever possible, rating scales are replaced by metric scales or percentage scales without categories and to use more standards.

3.2 Paper 2 (Abstract only): A threshold model for multi-year genebank data based on different rating scales

Hartung K, Piepho H-P (2005)
Crop Science 45:1045-1051

Abstract

Plant characteristics are often assessed on an ordinal rating scale. For example at gene banks accessions are surveyed routinely every year using such scales. A problem arising from long-term data is that scales may change over time. In this paper we present a method for joint analysis of data from different rating scales, assuming a threshold model with a common latent scale for the different rating systems. Mean scores on any of the rating scales can be derived. This is illustrated with a long-term series of evaluation data on barley.

3.3 Paper 3 (Abstract only): Are ordinal rating scales better than percent ratings? - A statistical and "psychological" view

Hartung K, Piepho H-P (2007)
Euphytica, 155:15-26

Abstract

Characteristics, like disease incidence or severity, are often assessed on an ordinal rating scale (e.g. scores from 1 to 9) or a percentage rating scale. In this paper three different scales [1%-steps and 5%-steps and an 9-point rating scale] are compared regarding accuracy, precision and time needed for scoring. Persons with different rating experience rated pictograms of diseased cereal leaves using the three rating scales. The pictograms simulated cereal leaves diseased with mildew and were generated following a right skewed beta-distribution. The thresholds of the 9-point rating scale followed a logarithmic pattern with respect to the underlying percentage scale. The transformed value of the estimated disease severity and the transformed time needed per leaf estimate were documented and evaluated using mixed models. In general both percent ratings performed better than the ordinal rating scale. The 9-point rating scale performed better for time needed with untrained raters. With trained raters the 5%-steps performed best. The raters, especially the untrained, mostly preferred the 9-point rating scale. Nevertheless, the results suggest that P5 can be recommended in terms of accuracy.

3.4 Paper 4:
Development in augmented designs and their potential for gene banks – a review

Hartung K, Piepho H-P (2006)

Abstract

Gene banks collect morphological and evaluation data during cultivating accessions for reproduction. Since in general accessions are cultivated without replication or experimental designs and only some gene banks use checks, it is nearly impossible to analyse data even from the same environment and year. So it cannot be decided to what degree the observed differences among entries are based upon genetical or environmental causes. As an approach to test many unreplicated entries Federer (1956) introduced the augmented design (AD), by which unreplicated entries can be corrected for environmental heterogeneity by replicated checks. In this article a short introduction to ADs is given, some of the important developments in design and analysis of ADs are summarised, and analysis over time and between different environments is considered, which is possible if trials are connected via identical checks or accessions. ADs are sets into context with gene bank needs, i.e., the reproduction of thousands of seed-samples. In relation to the gain of information, the additional effort for gene banks to implement ADs is small. Therefore we recommend ADs for gene banks, and expect that in the near future ADs and associated analysis software will be adapted to the specific needs of gene banks.

Introduction

The aim of gene banks is to preserve biodiversity. Hence, they store large numbers of accessions over long periods. Through time germination capacity of seeds decreases so that from a certain point on it is necessary to cultivate the material to receive enough fertile seeds for further storage. So the main focus in gene banks is on reproduction of seeds, i.e. the conservation of accessions. Some additional data of morphological traits are collected such as thousand seed weight, plant height and maturity date. Also, ordinal evaluation data are sometimes available, such as degree of lodging, resistance to pests and diseases. Up till the eighties (of the twentieth century), with barley and other cereals it was necessary to cultivate seeds every two to five years. From then on it was possible to cool-store seeds. Due to progress in research and technology it is now common to store barley at temperatures of -15 °C (5 °F) for up to over 15 years with unchanged fertility (Börner et al., 2000). This increases the absolute number of storable accessions. Hence, it is now impossible to cultivate all accessions together in one year in a homogeneous environment. As there are only limited capacities in manpower, field area, and money, and as it is only necessary to cultivate parts of the collection, only a small selection of accessions is cultivated every year without replication. Even though some gene banks use checks, these checks are cultivated without experimental designs. Therefore it is nearly impossible to analyse accessions

cultivated even in the same environment and year based on a sound model. An assessment of environmental heterogeneity is not possible, thus it can not be decided to what degree the observed differences are based upon genetical or environmental causes.

Federer (1956) introduced the augmented design (AD) for testing many unreplicated test lines in early breeding stages – a task that is similar to that faced by gene banks. In relation to the gain of information, the additional effort for gene banks to implement ADs is small. That this additional effort is necessary and desirable, emerges from articles like that of Haussmann et al. (2004), where it is stated that „The use of PGR [plant genetic resources] in crop improvement could be facilitated by systematic evaluation and documentation of the acquired data." The evaluation of gene bank data in their current form is only possible with difficulties and high computational effort (Hartung and Piepho, 2005, 2006). Through the inclusion of checks and proper experimental design, ADs offer an economical and easy-to-implement option for sound testing of many entries, which are denoted as accessions in gene bank context. Especially in times of decreasing financial resources, it is worthwhile to gain the largest benefit from evaluation data routinely collected by gene banks.

Even though we advocate AD for use by gene banks in the present paper, there are alternatives, e.g. the use of geostatistical methods such as next neighbouring. Cullis et al. (1989) adapted a geostatistical procedure for a field design used in Australian plant breeding. Isaaks and Srivastava (1989) give a good introduction in geostatistics and Schabenberger and Pierce (2002) offer detailed information for spatial methods and their computational implementation in SAS. Edmondson (2005) provides a review on design and analysis of crop experiments, which also covers spatial methods.

In the following, we will first give a short introduction to ADs, and then summarise some of the important developments in the design and analysis of ADs. Finally, we discuss ADs in the context of gene banks.

What is an Augmented design (AD)?

An AD is based on any standard design and its randomisation principles, e.g. complete or incomplete block design, or lattice square for the checks. The checks are randomised according to the chosen design and therefore are replicated. These designs then are augmented by entries (treatments), which are in general not replicated. Due to the presence of checks, the entries (= acces-

sions) can be adjusted for environmental heterogeneity. Kempton and Gleeson (1997) supply a good introduction to AD, as does Petersen (1994), who shows for one example how to analyse augmented block designs by ANOVA.

With every AD different minimum constraints are required to ensure that the experiment can be analysed. For example, with an augmented randomised complete block design two blocks and two different checks per block are the minimum constrains. Following Federer (1961), an augmented block design for two blocks (I and II), three checks (checks A to C) and 12 accessions (entries 1 to 12) could be constructed as shown in Figure 1. ADs combine the necessity to test many unreplicated entries with a well-known statistically analysable design that includes checks connecting the blocks (or row and columns).

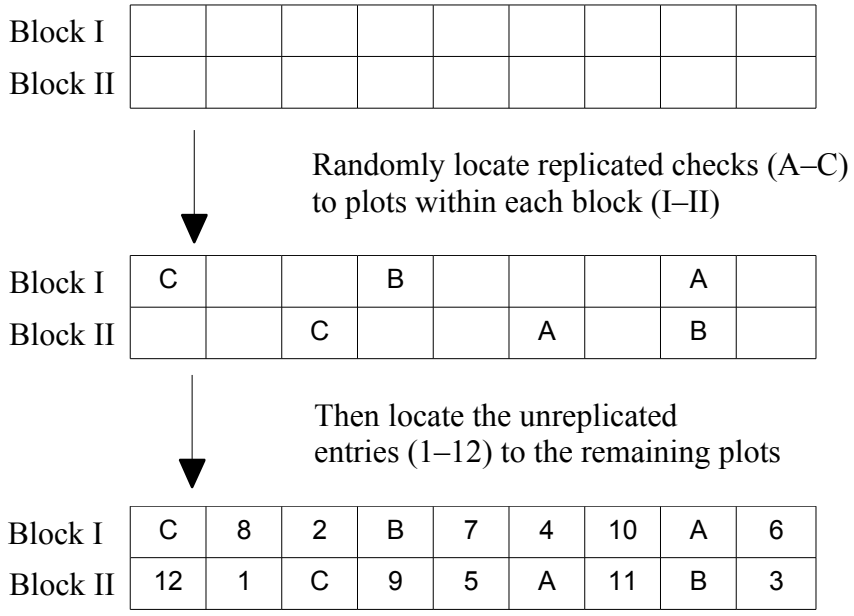

Figure 1: Construction of an augmented randomised complete block design with 2 blocks (I and II) with 3 checks (A to C) and 12 entries (1 to 12).

Articles published on AD

Possible designs for AD

Since Federer published his Article "Augmented (or Hoonuiaku) designs" in 1956, a lot of additional research results have been published. Federer (1961) extensively illustrated arithmetically and algebraically an augmented randomized complete block design and an augmented balanced lattice design. For both designs he considered analyses with and without recovery of inter-block information and provided some discussion on unequally sized incomplete blocks. He pointed out that sufficient replications of checks need to be included to have sufficient degrees of freedom

for estimating experimental error variance and effects of blocking used to control field heterogeneity (Federer and Raghavarao, 1975). He also gave a precise introduction to some augmented row-column designs (Federer et al., 1975; see Figure 2) and to the construction and analysis of augmented lattice square designs (Federer, 2002).

Lin et al. (1983) published a simulation study of three adjustment methods for AD based on a lattice square design and compared their AD with a balanced lattice square design. They named their design "modified AD" (MAD (Type1)), (see Figure 2b and 3a). This design, as well as MAD (Type2) (see Figure 3b), is broadly used in the literature (Auclair et al., 2004; Snijders, 2003; Casler et al., 2002; Calhoun, 1997).

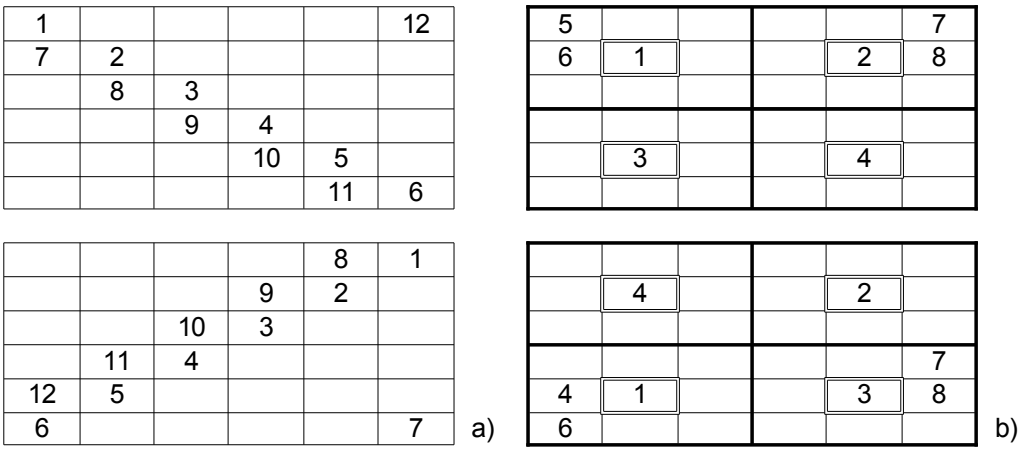

Figure 2: Two designs of ADs
a) AD following Federer et al. (1975) for row-column designs with 2 checks per row and column; labels 1-12 are checks b) modified AD (Type 1) following Lin and Poushinsky (1983); labels 1-4 are checks in the control plot, labels 5-8 are checks in control subplots, where 2 control subplots each are allocated randomly within a sufficient number of whole plots (delineated by solid line); subplots within whole plots are delineated by twin- lines.

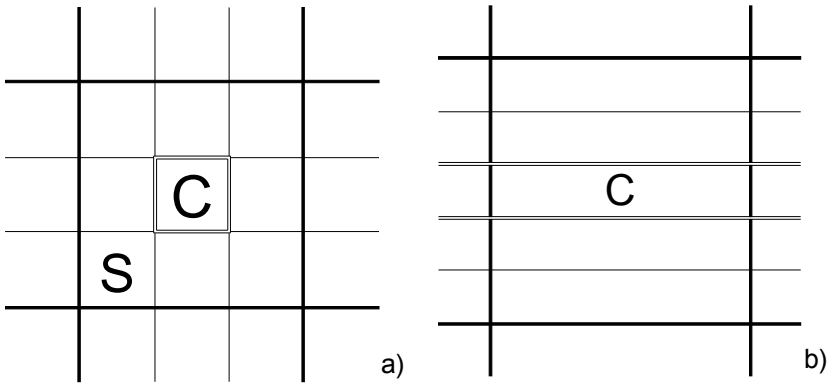

Figure 3: One whole plot of a modified augmented design (MAD) as recommended by Lin and Poushinsky (1983, 1985); square whole plot (bold line), with subplots (twin- and fine lines) either be rectangular (Type 1; a) or elongated (Type 2; b). Centre plot (twin-lines) for check is marked with a C and is surrounded by subplots. Here with the MAD (Type 1) (Figure 3a) one randomly chosen subplot per whole plot is used as control subplot (S).

In contrast to the designs mentioned by Federer, in the design by Lin et al. (1983) the checks are placed systematically on the field. The configuration of their MAD is based on square or near-square whole plots, which are subdivided in 3x3 square or near-square subplots. The centre sub-plot within each whole plot contains the check and is surrounded by eight subplots, which have approximately equal distance from the centre plot. So homogeneity of within-whole-plot correlations among the non-central sub-plots can be maintained. This design applies where square plots are used, but crops like cereals are mostly planted in long rows. To cope with this situation, Lin and Poushinsky (1985) developed the MAD (Type2) (Figure 3b). Within each square whole plot, they proposed five or any other odd number of rectangular subplots arranged in parallel rows within a whole plot. Again the check is placed in the centre subplot (control plot) and the entries are randomly allocated to the remaining plots. With both MADs (Type 1 and 2), Lin et al. (1983) proposed to additionally randomise control subplots for estimating subplot error. The control plot (check) always remains the centre subplot. Additionally, an arbitrary number of whole plots are chosen and supplementary check varieties are assigned randomly to at least one of the subplots. It is not important whether the different checks are cultivated in the control plot (Lin et al., 1983) or the control subplots (Lin and Poushinsky, 1985). It is also possible to replicate entries. Data from these entries can be used to assess the relative efficiency of adjustment (Lin et al., 1983). Lin et al. (1983) proposed three methods of adjustment for environmental heterogeneity, using the design structure (Method 1), the fertility index of the control plots (Method 2), and regression analysis (Method 3). Their results indicated that Method 1 and Method 3 are both effective depending on the environmental heterogeneity. Method 1 is preferable if variation is unidirectional, otherwise Method 3 is recommended. Method 2 is usually inferior to the other two methods. To decide which method is to be preferred, Lin and Poushinsky (1983) propose ANOVA to test for environmental heterogeneity and to choose the method on that basis as well as based on the relative efficiency (RE) of the adjustment method versus no adjustment (Lin et al., 1983). May et al. (1989) apply MAD (Type 2) to screen barley lines. In the discussion they stress that decisions made based on ANOVA are not always consistent with RE. But rank correlations between yields adjusted for Method 1 and yields adjusted for Method 3 were fairly high, so that in both cases the same entries would have been selected. In conclusion their data indicate that adjustment for environmental heterogeneity is required for entries and can have a considerable influence on the selection. If conclusions drawn from ANOVA and RE are not consistent, which can be due to the small number of control plots and control subplots, Lin and Voldeng (1989) suggest to investigate whether the disagreement arises from uneven distribution of control subplots or on outliers. If no apparent reason can be found, they recommend ANOVA to decide whether or not to adjust

the data and then to look at RE of subplot controls to choose the method. In borderline cases they propose Method 3 as it seldom over-adjusts. Schaalje et al. (1987) successfully used the MAD (Type1) in potato breeding, pointing out the ease of handling and ability to investigate environmental variation. As others (Kempton and Gleeson, 1997) also do, they caution that if checks do not respond in a similar fashion as entries, the adjustment might lead to misleading estimates. May and Kozub (1995) arrive at a similar conclusions regarding the usefulness of MAD as Lin and Voldeng (1989) and May et al. (1989). In their discussion they point out that replication of "MAD in more than one environment would increase the likelihood of best lines being advanced to subsequent trials." But they also point out that "selecting [...] in a specific environment does not imply that they [entries] will yield well in other environments."

The ADs as proposed by Federer and those proposed by Lin et al. have advantages and disadvantages. For the ADs proposed by Federer, the randomisation theory is valid and the analysis is based on a sound linear model. This has the advantage of an unbiased error estimate. But the random location of checks is also a problem. If checks are randomly located, e.g., to plots of a block, their distribution pattern can become very patchy over the experimental site (see e.g. Figure 2a). So assessment of soil heterogeneity can become unreliable for single randomisations. That is why Lin and Poushinsky (1983) hold that "the primary object […] is to obtain gross estimates of genotypic values […], rather than to test […] differences critically, effective adjustment is more important than unbiased error estimation." As second problem they point out that the "effectiveness of adjustment for environmental heterogeneity depends on the basic assumption of homogeneity of the plots within a block." This may not hold with standard randomisations, if few checks are used to adjust many entries and/or if the distances between checks and entries show pronounced heterogeneity. So their MAD, where checks are always placed in the centre plot of every whole plot, prevents the problem with the experimental design. The uncertainty with their ADs lies with the analysis, for there is no obvious best model to estimate the genetic effect of an entry. None of the three proposed methods of analysis is based on an explicit statistical model, and therefore there is no well-founded decision criterion to select between them.

A recent development in ADs are so-called alpha-alpha-designs (Williams and John, 2003), which are advanced compared to augmented lattice square designs in that in every (long) row or column the number of different checks is equal. Thus, if a whole (long) row or column of a trial is lost, more than one check variety remains and therefore their information can be used.

Computational analysis of ADs

In the nineties, when personal computers and statistical software became readily available, data analysis of AD became straightforward with linear models as a basis. Scott and Milliken (1993) published a SAS program for analysing augmented randomized complete block designs, which conducts ANOVA using a generalised linear model (GLM) procedure, and Restricted Maximum Likelihood (REML) is used to combine intra- and inter-block information via mixed models. Wolfinger et al. (1997) published an article with a similar focus. They offer SAS codes for ADs based on randomized complete and incomplete block designs and other designs such as row-column designs. The codes to obtain only intra-block and intra-variety (= intra-entry) analyses as well as codes for recovering inter-block information are given. With row-column designs they propose orthogonal polynomial regressions (Arfken, 1985) to model the environmental heterogeneity.

Santos et al. (2002) conducted a simulation study on the efficiency of four types of mixed model for the augmented block design including effects for entries and blocks as well as a random error. The models differed in terms of the nature of the effects of blocks and entries, being either fixed (F) or random (R): FF, FR, RF, RR. In the simulation the number of entries (50, 100, 200), the heritability h^2 (0.2, 0.5, 0.8), the soil heterogeneity b (0.1, 0.5, 0.9) as suggested by Bearzoti (1997), and the number of blocks (5% or 20% of the number of entries) were varied. With the variance components there were two assumptions: In one investigation they were considered to be known and in the other they were assumed to be unknown. Using the mean squared error (MSE) as decision criterion, for known as well as unknown variance components and high residual variation (low h^2 and high b), RR is to be recommended. RR is also preferable with known variance components if Spearman correlation between predictions and actual genetic effect of entries is the criterion. With unknown variance components and Spearman correlation as decision criterion, RF was always superior. The RR approach presented the smallest elite bias in 85.2% of all cases. The elite bias is defined as the bias in the estimation of the percentage of treatments superior to the best check. These results suggest that if selection is non-truncated, the recovery of inter-block and inter-variety information is suitable.

Series of expriments – Combining AD results

Federer et al. (2002) considered the combination of results from ADs over sites. The authors investigated a case with three different sites. At sites 1 and 2 the design had 15 rows by 12 columns with two checks, each of which was replicated 30 times, and 120 unreplicated entries.

At the third site the design was laid out as an incomplete block design with v = 120 treatments (the entries from site 1 and 2), incomplete block size k = 8, and 15 incomplete blocks within each of r = 2 replicates. No checks were included. The SAS codes for analysis are given as well as two methods to combine results over sites. The authors point out that "these methods do not require the same experimental design at each site, do not require the same check treatments at each site, do not require the same response model at each site, and do not require homogeneity of error variances from site to site. […] Apart from the saving of land and management costs, reducing the number of repetitions at a site and increasing the number of sites, improves the efficiency with which certain traits are evaluated by reducing the time to characterize a family of lines."

Generally, the combination of results from different sites is straightforward, when powerful mixed model procedures are used (Smith et al., 2005), also when designs differ among sites. Historically, a two-stage method has been used in most instances, in which adjusted means are computed per trial in the first stage, which are combined using a suitable mixed model (Cochran and Cox, 1957). With powerful software, a fully efficient single-stage analysis is now feasible with minimal computational effort (Smith et al., 2005).

ADs in context with gene bank needs

In the last years ADs have been enhanced, and due to high performance PCs and software programs, complex analysis procedures have become available. So ADs represent a well manageable tool to separate genetic and environmental effects. Nowadays ADs are used commonly in different areas of research, also when gene bank material is investigated (Padilla et al., 2005; Châtel et al. 2004; Sharma et al., 2003; Anstaltsleitung der BAZ, 2004). But so far ADs have not found their way into standard cultivation of accessions in gene banks. As a minimum requirement, the cultivation of checks with a sufficient number of replications is desirable (Hartung and Piepho, 2006). The associated design can and must be adapted to the examined crop and the conditions of every single gene bank. An experimenter, who wants to implement AD, first has to decide, which design fits his needs. In most cases the main environmental factors influencing plant growth – like slope, shadow, or main direction of infestation – are known. Therefore the experimenter decides on the experimental design of a gene bank trial on the same basis as he would with any other experiment. With unidirectional variation in field he might choose designs like a complete or incomplete block design. With field heterogeneity in more than on direction row-column designs are advisable. One should try to use square whole plots, as recommended by Lin and Poushinsky (1985) while subplots can either be rectangular (Type 1) or elongated (Type 2)

(Figure 3). If that is not possible, all subplots should equally be affected by the stronger environmental gradient, i.e. laid out orthogonal to this factor. The minimum number of checks and replications depends on the chosen design, with the entries of the basic design corresponding to checks in the AD. For example if checks are laid out in a randomised complete block design, there must be at least two blocks and two checks per block. Regarding the number of replicates for checks, Yates (1932) proposed $n^{½}$ where n is the number of entries, while Kempton and Gleeson (1997) hold that a "higher frequency of checks (say grater than one in five) is unwise unless the spatial heterogeneity is large and very local so that correlation among plot yields fall off rapidly with separation". Assuming that 400 accessions are to be tested, then the number of plots for entries (= accessions) and checks following Kempton and Gleeson is 480 plots, following Lin it is at least 450 plots and following Yates it is 421 plots. So for 400 entries a design with 21 to 80 check plots should be chosen. The number of different check varieties has to fit the demands of the design. In gene bank context it also might be interesting to replace some check varieties by accessions that are in higher demand due to a larger frequency of inquiries to the gene bank. These accessions could be replicated in the same way as checks.

The possibility to link data of diverse environments, i.e. between different gene banks, even though different checks and accessions were cultivated, might be of special interest in gene bank context. In order to come up with a meaningful analysis, however, it is reasonable to have as many connecting checks and accessions in all environments as possible. The collected data can than be set into a larger context and offer a better basis for decision as to which accession is of interest for a certain purpose. It is well known that testing an entry in one more environment is more efficient than one more replication in the same environment (Talbot, 1997), for it reduces not only the impact of the error variance but also that of the variance of interactions involving environmental effects. The same is truth for replication in time.

A general requirement for a joint analysis is that all data are connected (Searle, 1987). This is important, no matter whether a single experiment or a series of experiments over time or different environments is analysed, because experiments are only comparable if at least one accession (= entry) or check is tested in two years or two sites to be combined. The more similar the conditions under which checks and/or accessions are cultivated, the more meaningful and efficient the analysis will be.

Conclusion

ADs are very flexible and have been applied in different areas, mostly in field experiments, but also in other fields such as poultry research (Boyle and Montgomery, 1996). Powerful computers and software necessary for analysis are fully developed and available. It is likely that over time ADs and associated analysis software will be adapted to the gene banks needs. The programs for gene banks should include methods to select additional accessions that insure the connection of experiments over time in one place and within on year over sites, i.e. coordination between gene banks, if desired. It also might become possible to determine the optimal number of different checks and replications, based on experiences of previous years.

This article arose from a research project focusing on the statistical analysis of gene bank data. In this context, we felt it desirable to call attention to the need of replication (checks) and good experimental design. Both requirements are satisfied if ADs are used. It is not only desirable but also economically reasonable to invest the additional costs for the more efficient design. The information content and the explanatory power of the data increases by use of good experimental design.

Literature

Anstaltsleitung der BAZ (Hrsg.). 2004. Jahresbericht 2003 der Bundesanstalt für Züchtungsforschung an Kulturpflanzen (BAZ), Koch-Druck, Halberstadt. 99-109 http://www.bafz.de/JB2003/pgr.pdf

Arfken G. 1985. Orthogonal Polynomials. Mathematical Methods for Physicists, 3rd ed. Orlando, FL: Academic Press, p. 520-521.

Auclair J, Boland GJ, Cober E, Grael GL, Steadman JR, Zilka J, Rajcan I. 2004. Development of a new field inoculation technique to assess partial resistance in soybean to Sclerotinia sclerotiorum. Canadian Journal of Plant Science 84:57-64.

Börner A, Freytag U, Specht CE. 2000. Ergebnisse zur Langzeitlagerung von Samen verschiedener Kulturpflanzen in der Genbank Gatersleben. S. 29-32. Bericht über die 51. Arbeitstagung der Vereinigung Österreichischer Pflanzenzüchter BAL Gumpenstein.

Boyle CR, Montgomery RD. 1996. An application of the augmented randomized complete block design to poultry research. Poultry Science 75:601-607.

Calhoun DS. 1997. An evaluation of modified augmented designs for single-replication yield testing in cotton. p.460-464. In: Dugger P, Richter D (ed.) Proceedings Beltwide Cotton Conferences, New Orleans, LA, USA.

Casler MD, Fales SL, McElroy AR, Hall MH, Hoffman LD, Undersander DJ, Leath KT. 2002. Half-sib family selection for forage yield in orchardgrass. Plant Breeding 121:43-48.

Châtel M, Lorieux M, Martinez C, Borrero J, de Carvalho Neves P, Peixoto de Morais O. 2004.

Assemble and document information on selected entries of CIAT, CIRAD, IRD and EMBRAPA rice working collections. In: Improved Rice Germplasm for Latin America and the Caribbean Project, 2004 Annual Report. CIAT. p. 13-14 http://www.ciat.cgiar.org/riceweb/pdfs/rice_2004_output_1.pdf

Cochran WG, Cox DR. 1957 Experimental designs. 2nd edition. Wiley, New York.

Cullis BR, Lill WJ, Fischer JA, Read BJ. 1989. A new procedure for the analysis of early generation variety trials. Applied Statistics 38:361-75

Edmondson RN. 2005. Centenary Review: Past developments and future opportunities in the design and analysis of crop experiments. Journal of Agricultural Science 143: 27-33.

Federer WT, Raghavarao D. 1975. On augmented designs. Biometrics 31:29–35.

Federer WT. 1961. Augmented designs with one-way elimination of heterogeneity. Biometrics 17:447-473.

Federer WT. 2002. Construction and analysis of an Augmented lattice square designs. Biometrical Journal 44:251-257.

Federer, WT, Nair RC, Raghavarao D. 1975. Some augmented row-column designs. Biometrics 31:361–373.

Federer WT, Reynolds M, Crossa J. 2001. Combining results from augmented designs over sites. Agronomy Journal 93:389-395.

Hartung K, Piepho HP. 2005. A threshold model for multi-year genebank data based on different rating scales. Crop Science 45:1045-1051.

Hartung K, Piepho HP, Knüpffer H. 2006. Analysis of gene bank evaluation data using geostatistical methods. Genetic Resources and Crop Evolution 53:737-751.

Hartung K, Piepho HP. 2006 Optimizing an augmented design using geostatistical methods. submitted to Canadian Journal of Plant Science.

Haussmann BIG, Parzies HK, Presterl T, Susic Z, Miedaner T. 2004. Plant genetic resources in crop improvement (Review). Plant Genetic Resources:characterization and utilization, 2:3-21.

Isaaks E. and Srivastava M. 1989. Applied geostatistics. Oxford University Press, Oxford.

Kempton RA, Gleeson AC. 1997. Unreplicated trials. p.86-100. In Kempton RA, Fox PN (ed.). Statistical methods for plant variety evaluation. Chapman & Hall, London.

Lin CS, Poushinsky G, Jui PY. 1983. Simulation study of three adjustment methods for augmented design and comparison with the balanced lattice square design. Journal of Agricultural Science 100:527-534.

Lin CS, and Voldeng HD. 1989. Efficiency of type 2 modified augmented design in soybean variety trials. Agronomy Journal 81:512–517.

Lin CS, Poushinsky G. 1983. A modified augmented design for an early stage of plant selection involving a large number of test lines without replication. Biometrics 39:553-561.

Lin CS, Poushinsky G. 1985. A modified augmented design (Type 2) for rectangular plots. Canadian Journal of Plant Science 65:743-749.

May KW, Kozub GC, Schaalje GB. 1989. Field evaluation of a modified. augmented design (Type 2) for screening. barley lines. Canadian Journal of Plant Science 69:9-15.

May KW, Kozub GC. 1995. Success of a selection program for increasing grain yield of two-row barley lines and evaluation of the modified augmented design (Type 2). Canadian Journal of Plant Science 75:795-799.

Padilla G, Cartea ME, Rodriguez VM, Ord A. 2005. Genetic diversity in a germplasm collection of Brassica rapa subsp. rapa L. from northwestern Spain. Euphytica 145: 171-180.

Peterson RG. 1994. Agricultural field experiments: design and analysis. New York, USA. Marcel Dekker Inc.

Santos AH, Bearzoti E, Ferreira FD. 2002. Simulation of mixed models in augmented design. Scientia Agricola 59:483-489.

Schaalje GB, Lynch DR, Kozub GC. 1987. Field evaluation of a modified augmented design for early stage selection involving large number of test lines without replications. Potato research: Zeitschr. d. Europaeischen Gesellschaft fuer Kartoffelforschung 30:35-45.

Schabenberger O, Pierce FJ. 2002. Contemporary statistical models for the plant and soil sciences. CRC Press, Boca Raton.

Scott RA, Milliken GA. 1993. A SAS program for analysing augmented randomized complete block design. Crop Science 33:865-867.

Searle SR. 1987. Linear models for unbalanced data. Wiley, New York.

Sharma KC, Sharma RK, Singhania DL, Singh D. 2003. Variation and character association in fodder yield and related traits in pearl millet (Pennisetum glaucum (L.) R. Br.). Indian Journal of Genetics & Plant Breeding 63:115-118.

Smith AB, Cullis BR, Thompson R. 2005. The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. Journal of Agricultural Science 143:449-462.

Snijders CHA. 2003. Field evaluation of type 2 modified augmented designs for non-replicated yield trials in the early stages of a wheat breeding program. Bericht ueber die Arbeitstagung 2002 der Vereinigung der Pflanzenzuchter und Saatgutkaufleute Osterreichs gehalten in Gumpenstein. p. 59-64.

Talbot M. 1997. Resource allocation for selection systems. p. 162-174. In Kempton RA, Fox PN (ed.). Statistical methods for plant variety evaluation. Chapman & Hall, London.

Williams ER, John JA. 2003. A note on the design of unreplicated trials. Biometrical Journal 45:751-757.

Wolfinger RD, Federer DW, Cordero-Brana O. 1997. Recovering information in augmented designs, using SAS Proc GLM and Proc MIXED. Agronomy Journal 89:856-59.

Yates F. 1936. A new method of arranging variety trials involving a large number of varieties. Journal of Agricultural Science 26:424-455.

3.5 Paper 5:
Optimizing an augmented design using geostatistical methods

Hartung K, Piepho H-P (2006)

Abstract

Spatial analysis can increase the efficiency e.g. compared to a standard complete block analysis, when within-block homogeneity is low, as can be the case in gene banks and early breeding stages, when block sizes are large. In this paper spatial analysis is used to find an optimal field layout for an augmented design, i.e. a layout that yields small least significant differences (LSD). The average variance of a difference (a.v.d.) and the average squared LSD are used to compare competing designs, using a theoretical approach based on variations of two anisotropic models (nearly isotropic, strongly anisotropic) and different rotations of anisotropy axes towards field reference axes. The a.v.d. is influenced by the number of blocks, plots and checks, with checks as dominating factor. Based on our calculations, up to five checks per block are recommended. The nearly isotropic combinations lead to designs with large quadratic blocks. The strongly anisotropic combinations display effects resulting from the combination of degree of anisotropy and rotation of anisotropy axes. Without rotation small elongated blocks are preferred. The closer the rotation is to 45° the more square blocks and the more checks are approprate. The squared LSD is a meaningful optimization criterion, which can and must be set into context with the practitioners needs.

Introduction

In field experiments conditions in plots closer together are likely to be more similar than in plots further apart. Trials in gene banks often include only few or no check varieties and large numbers of genotypes. Block sizes, if blocks are used at all, tend to be large and so within block homogeneity often is low. This is due to competition between entries, heterogeneity of soil, crop diseases and insect dispersion as well as other influences. The same problem occurs in early breeding stages. Augmented designs have been proposed by Federer (1956) as one way to address these problems. To construct an augmented design, checks are randomized according to a common replicated design, e.g. a block design, which is then modified by augmenting blocks with unreplicated treatments (entries). Numerous investigations have been conducted for this type of design (Wolfinger et al., 1997; Scott and Milliken, 1993; Schaalje et al., 1987), and Kempton and Gleeson (1997) and Peterson (1994) give a general introduction. If mixed models are used to analyse an augmented design, treatments and/or blocks can analysed as fixed or random factors (Santos et al., 2002) and analysis over sites is possible (Federer et al., 2001). Besides augmented designs, spatial methods, which analyse data with respect to spatial covariance, have been available since Papadakis (1937). Up to now numerous spatial methods have been

developed (Edmondson, 2005). In this investigation we used geostatistical methods to optimize block size and shape with a classical linear model analysis in mind. The purpose of this article is to help the researcher to find the best field layout for an augmented design with regard to:

- Number of entries to be tested (a), number of checks (s) to be used per block, ratio between numbers of entries and checks;
- plot size and shape as well as their position on the experimental site
- available number of plots (n_s), number of plots necessary (n_f) for entries plus checks, number of plots per block (k), number of blocks (b) and block shape;
- spatial information about the experimental site including degree of anisotropy, type of spatial model (e.g. Gaussian or exponential), spatial parameter values (sill, nugget, range), angle between major principal axis of anisotropy and reference axis of experimental site.

These aspects interact with on another and therefore need to be considered simultaneously.

Material and Methods

Gene banks and breeders are interested to detect differences between genotypes. A good design, i.e. one leading to a small (average) least significant difference (LSD), depends on a number of aspects, e.g. block shape and composition, number of checks, number of unreplicated entries, and spatial field variation.

Experimental site

The experimental site is assumed to be a symmetric rectangular grid with n_s plots laid-out in r rows and c columns. Every plot is assumed to be of the same shape and size. The grid is defined via x- and y-coordinates with the x-axis as main reference axis. Every plot is referenced by the coordinates of its midpoint. If only a certain area of the experimental site is used, this will be called an experimental field with n_f plots, where $n_f < n_s$. Blocks laid out on the experimental site are assumed to be rectangular and to have the same number of plots (k) (see Figure 1).

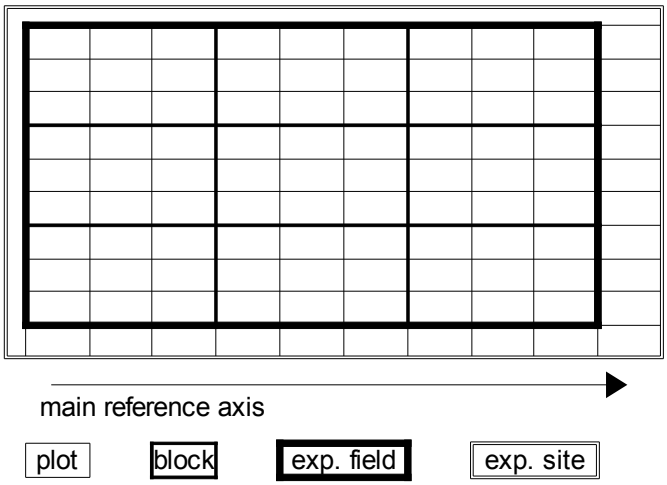

Figure 1: Experimental site with n_s= 100 plots laid out with 9 blocks of 3x3 plots leading to a experimental field of n_f = 89.

Geostatistical models

Geostatistical models assume that the covariance of values measured at two points becomes larger with decreasing distance, since the environmental influence becomes more similar. The spatial relationship may be represented by the semivariogram (for details see Schabenberger and Pierce, 2002; Hartung and Piepho, 2006). Semivariogram and covariance structure are two different representations of the spatial correlation structure, which can be transferred one into another (Table 1; Schabenberger and Pierce, 2002). In our investigation the geostatistical model includes variogram model, isotropy or geometric anisotropy, and rotation (only with geometric anisotropy). We consider the linear, spherical, exponential and Gaussian variogram model (Table 1), and assume second order stationarity for modelling the covariance structure.

Table 1: Isotropic variogram models and covariance structures used in this paper.

	Variogram model	Covariance structures
Spherical	$\sigma_N^2+\sigma^2\left\{1.5\left(\frac{h}{R}\right)-0.5\left(\frac{h}{R}\right)^3\right\}$ *for* $h\le R$	$\sigma^2\left[1-\left\{1.5\left(\frac{h}{R}\right)-0.5\left(\frac{h}{R}\right)^3\right\}\right]$ *for* $h\le R$
	σ_N^2 *for* $h\le R$	σ_N^2 *for* $h\le R$
Exponential	$\sigma_N^2+\sigma^2\left\{1-\exp\left(\frac{-h}{R}\right)\right\}$	$\sigma^2\left\{\exp\left(\frac{-h}{R}\right)\right\}$
Gaussian	$\sigma_N^2+\sigma^2\left\{1-\exp\left[-\left(\frac{h}{R}\right)^2\right]\right\}$	$\sigma_N^2+\sigma^2\left\{\exp\left[-\left(\frac{h}{R}\right)^2\right]\right\}$

σ_N^2 : nugget variance; σ^2: variance of partial sill (sill minus nugget); h: distance; R: range

When the semivariogram does not vary with direction, i.e. the variance between two points only depends on their distance and therefore the ratio of ranges is 1, it is said to be isotropic and isolines for correlation in the *x*-*y*-plane are perfect circles. If the variance is influenced not only by the distance but also by the direction, the semivariogram is said to be anisotropic. A special case of anisotropy occurs when the values for sill and nugget are identical in each direction but the range changes with direction, which is called geometric anisotropy. This leads to an elliptical contour of isolines of correlation. The ellipse can be described by using the direction of the largest range as major principal axis and their orthogonal axis (see Figure 2). The more elongated the ellipses, i.e. the more different the ranges of the two principal axes, the stronger is the degree of anisotropy. By linear transformation of axes, geometric anisotropy can be transformed into isotropy.

If main reference axis of experimental site and major principal axis of anisotropy are rotated against each other, the coordinates of the main reference axis need to be transformed so that the orientation is the same for anisotropy axes and coordinate axes of plots. The new coordinates here denoted *x*' and *y*', then represent the plot coordinates with respect to the new axes. With *x* and *y* being the coordinates of plot midpoints relative to the field main reference axes, we use the following transformation:

$$x' = x^* \cos \alpha + y^* \sin \alpha$$
$$y' = x^* (-\sin \alpha) + y^* \cos \alpha \ ,$$

where α is the counter-clockwise angle between *x*-axis of field and major principal anisotropy axis (Figure 2; Schabenberger and Gotway, 2000).

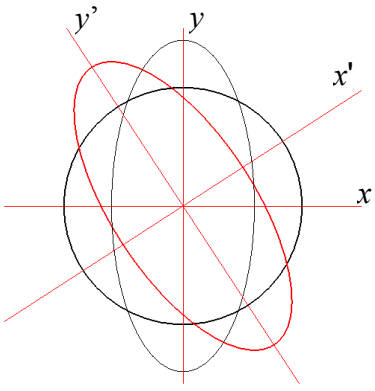

Figure 2: Isotropy (black circle), geometric anisotropy (black ellipse) and rotated anisotropy [red ellipse; major principal axis (*x*') is rotated by 15° regarding the main reference axis (*x*)]; circle and ellipses are isolines with same semivariance value.

Deriving the error and block variance

Assume a simple linear model written as

$$y_{ij} = \mu + \tau_{ij} + d_{ij},$$

where μ is a general mean, τ_{ij} is the effect of the treatment on the ij-th plot (in i-th row and j-th column of field layout) and d_{ij} is the ij-th plot effect, whose distribution has zero mean and a spatial variance-covariance structure Σ, and the spatial model for d_{ij} is stationary with constant marginal variance σ^2. If the experimental field is divided into blocks, plot effects will be partitioned into block and residual error effects. It is convenient to consider a null model without treatment effects:

$$y_{ij} = \mu + d_{ij}.$$

Assuming a block design, the deviation d_{ij} can be partitioned into block and plot effects, and the null model can be expressed as

$$y_{ij} = \mu + \beta_{h(ij)} + e_{ij},$$

where $\beta_{h(ij)}$ is the h-th block effect and it is assumed that the ij-th plot has been assigned to the h-th block. Let X_β be the design matrix for blocks and y the observation vector, assumed to be sorted by blocks. The error sum of squares (SS) of a standard block ANOVA for the null model is computed as

$$SS_E = \boldsymbol{y}'\left(\boldsymbol{I}_{n_f} - \boldsymbol{X}_\beta(\boldsymbol{X}'_\beta\boldsymbol{X}_\beta)^- \boldsymbol{X}'_\beta \right)\boldsymbol{y}$$

on (n_f - b) d.f., where I_{n_f} is an $n_f \times n_f$ identity matrix, b is the number of blocks, n_f the number of observed plots ($n_f = b*k$, with k the block size). Under a spatial model with variance-covariance structure Σ, the error mean square (MS_E)

$$\text{MSE} = \text{SSE}/(n_f - b)$$

has expected value

$$\text{E}(MS_E) = (n_f - b)^{-1} \text{ trace } \{\text{A}\Sigma\},$$

where $\boldsymbol{A} = \boldsymbol{I}_{n_f} - \boldsymbol{X}_\beta(\boldsymbol{X}'_\beta\boldsymbol{X}_\beta)^- \boldsymbol{X}'_\beta$ Searle et al., 1992, Appendix S) .

We set the error variance of the model $y_{ij} = \mu + \beta_{h(ij)} + e_{ij}$ equal to $\sigma_e^2 = \text{E}(MS_E)$.

In order to simplify the expression for σ_e^2, note that the design matrix can be written as $X_\beta = I_b \otimes 1_k$, where I_b is an $b \times b$ identity matrix, 1_k is an column vector of k ones, and ⊗ denotes the Kronecker product. Also it can be that

$$\text{A} = I_b \otimes R_k,$$

where $R_k = (I_k - (\frac{1}{k})J_k)$, and J_k is $k \times k$ matrix of ones everywhere.

With the spatial variance-covariance matrix Σ partitioned as

$$\Sigma = \begin{pmatrix} \Sigma_0 & \Sigma_{12} & . & . & \Sigma_{1r} \\ \Sigma_{21} & \Sigma_0 & & & \Sigma_{2r} \\ . & & \Sigma_0 & & . \\ . & & & \Sigma_0 & . \\ \Sigma_{1r} & \Sigma_{2r} & . & . & \Sigma_0 \end{pmatrix},$$

where Σ_0 is the spatial variance-covariance matrix within a block, it can be shown that

$$\text{trace}(A\Sigma) = b\ \text{trace}(R_k\Sigma_0)\ ,$$

whence

$$\sigma_e^2 = (k\text{-}1)^{-1}\ \text{tracε}\ (P_\kappa\Sigma_0)\ .$$

Computing the average variance of a difference (a.v.d.) and the squared least significant difference (squared LSD) for an augmented design

Assume an augmented design where *s* checks are laid out in complete blocks and these blocks are augmented by $a=b(k\text{-}s)$ entries. The variance of a difference of entries tested in the same block is σ_e^2 . With *s* checks and block size *k*, there are *k–s* entries per block and thus (*k–s*)(*k–s*–1)/2 direct comparisons among entries per block or *b*(*k–s*)(*k–s*–1)/2 direct comparisons overall. For an indirect comparison, the variance is $2\sigma_e^2(s+1)/s$. For each entry in a particular block, there are (*b*–1)(*k–s*) indirect comparisons. Thus, the total number of unique indirect comparisons is *b*(*k–s*)(*k–s*–1)2/2. The variance of a difference depends on the compared entries. An average variance of a difference (a.v.d.) can be calculated as follows:

$$\text{a.v.d} = \sigma_e^2 \frac{2*b(k-s)(k-s-1)/2 + 2(s+1)/s*b(b-1)(k-s)^2/2}{b(k-s)(k-s-1)/2 + b(b-1)(k-s)^2/2}$$

$$= 2\sigma_e^2 \frac{(k-s-1) + (s+1)(b-1)(k-s)/s}{b(k-s)-1}$$

We also need to consider the fact that the d.f. for estimating error will be limited. Thus, it is useful to compute the average squared LSD as

$$\text{LSD}^2 = \text{a.v.d.} \times t^2\ ,$$

where t is the appropriate critical value of a t-distribution with (*b*–1)(*s*–1) d.f.
For the following it is helpful to write a.v.d. as

$$a.v.d. = 2\sigma_e^2 * f(b,k,s)$$

with

$$f(b,k,s) = \frac{(k-s-1)+(s+1)(b-1)(k-s)/s}{b(k-s)-1}$$

Theoretical view on the influence of *b*, *k*, and *s* on squared LSD and *f(b,k,s)*

The squared LSD is a function of spatial variance [$\sigma_e^2(b, \Sigma_0)$] , the influence of the block effect [*f*(*b,k,s*)], and the critical t-value [t(*k,s*)]. These factors are influenced by *b*, *k* or *s*. Each of these three variables influences two factors. In order to achieve small squared LSD, all factors should be small:

- To minimize spatial variance, small block sizes are preferable. The optimal shape of a block, assuming square plots, depends on the degree of anisotropy (isotropy: ratio of ranges equals 1) and if anisotropy (ratio ≠ 1) is present, it depends on the angle between major principal anisotropy axis (larger range) and reference axis of experimental site.
  With isotropy square blocks are always optimal. With anisotropy the optimal shape depends on the angle. If the angle is 0° + *m** 90°, where *m* is an integer, the shapes elongate the more the ratio of both ranges differs from 1. This effect decreases the closer the angle is to 45° + *m** 90°, in which case again square blocks are optimal.
- To minimize *f*(*b,k,s*) few large blocks and many checks are best.
- To minimize the critical t-value, the degree of freedom (*b*-1)(*s*-1), i.e. the number of blocks and checks, should be high.

As can be seen from above considerations, the spatial variance and the critical t-value on the one side and the function *f*(*b,k,s*) lead to contradicting demands regarding the optimal number of blocks (*b*). In addition, the number of checks per block and number of blocks need to fit the experimental site, i.e. the available number of plots (n_s). Likewise the number of entries that need to be tested might be fixed.

The factor *f*(*b,k,s*) was examined to get more detailed information about the influence of the number of checks as well as the number of entries per block and the number of blocks. To asses the influence of every variable in (*b,k,s*), certain combinations of (*b,k,s*), where always two variables were kept constant, were examined (see Table 3). For example, when *b* was assessed, *k* and *s* were fixed at a constant level *k*=6 and *s*=2, and *b* was replaced by *a*=*b*/(*k*-*s*) in *f*(*b,k,s*) and then plotted versus *a*. To describe the magnitude of the influence of a variable in (*b,k,s*), the derivative of *f*(*b,k,s*) with respect to the variable divided by *f*(*b,k,s*), was plotted for every variable (f_b'/f_b; f_k'/f_k; f_s'/f_s).

The number of entries (*a*) was chosen as basis for evaluating alternative designs due to the fact that the number of entries to be tested might be the most relevant reference. Additionally the number of entries is a reference that can be used for all three variables of interest by converting number of checks (*s*), block size (*k*) and number of blocks (*b*) into that unit. For this purpose we

used the relation $a=b*(k-s)$ to express the factors determining squared LSD as a function of the number of entries (a). This leads to a certain way of interpreting the figures: When reading a figure where the number of entries is the unit on the abscissa, the total number of plots depends on the number of checks and block size chosen. For a given number of entries or value of squared LSD one finds the corresponding value of b, k, or s by solving $a=b*(k-s)$ for the variable in question and inserting the given values for the remaining variables. If for example the value of b corresponding to the function $f_b(b,6,2)$ from Figure 3a for 8 entries is of interest, the relation is $8 = b*(6-2)$ therefore $b = 8/4$ and so $b = 2$.

Testing the influence of some spatial combinations on squared LSD

To get insights regarding the spatial effects, some of the results of Pringle (2002) were used. He investigated variograms from published and unpublished reports. If variogram models were not spherical, an equivalent spherical model was estimated and used as approximation (McBratney and Pringle, 1999). Results from these isotropic spherical variograms were combined. To simulate anisotropy, nugget or partial sill were made to agree by averaging the values if they differed. Additionally, the values of partial sill, which equals the difference of sill minus nugget, and nugget were varied using the highest rounded-off values from Pringle (2002) (see Table 2).

Table 2: Parameter values of major and minor range, partial sill and nugget tested with the nearly isotropic and the strongly anisotropic spatial combinations taken from Pringle (2002).

Spatial combinations	Parameter	Values			
nearly isotropic	major range	57	57	57	57
	minor range	51	51	51	51
	partial sill	0.09	0.09	0.5 [a]	0.5 [a]
	nugget	0	0.2 [a]	0	0.2 [a]
strongly anisotropic	major range	500	500	500	500
	minor range	27	27	27	27
	partial sill	0.065	0.065	0.5 [a]	0.5 [a]
	nugget	0.02	0.2 [a]	0.02	0.2 [a]

[a] Rounded-off highest value reported in Pringle (2002)

An experimental field of size 480 x 480 m^2 with plots of size 20x20 m^2 was assumed, following the conditions of experimental sites described in Pringle (2002). Spatial models were parameterized with ranges, partial sill and nugget as shown in Table 2. Further restrictions were: the number of checks was between two and eight; the ratio between numbers of entries and checks was 2:1; due to the requirement of a ratio of 2 entries to 1 check and a minimum of 2 checks the block size was set to be between 6 and 24; all plots had to be used, so there had to be 384 plots for entries and 192 plots for checks. All combinations were tested for angles of rotation of 0°, 15° and 45°.

As a second example all restrictions were the same as before except for the ratio between numbers of entries and checks. The ratio was changed to be at least 2 to 1, hence designs with at least 384 entries are comprised.

The ranks of competing designs according to the squared LSD values were determined for exactly 384 entries (ratio exact 2:1) and additionally for 384 and more entries (ratio ≥ 2:1 $\rightarrow a \geq 384$) (see Table 4 and 7)

Computational Implementation

To find a good design a macro was written, which is available from the authors. Geostatistical models cover isotropy and geometric anisotropy and the linear, spherical, exponential and Gaussian model as well as an option for rotation of the major principal axis of anisotropy. The experimental site is defined by plot midpoints. The semivariogram parameters to be entered include nugget, sill and range for major and minor principal axes. In case of isotropy, the range is the same for both axes and the angle of rotation is zero. With anisotropy the angle between major principal axis of anisotropy and reference axis of field has to be given counter clockwise (see Figure 2). The range, the *x*- and *y*-coordinates have to be specified in the same unit.

Results

Consideration of *f*(*b*,*k*,*s*)

The function $f(b,k,s)$ as theoretically examined for the influence of b, k, or s is denoted as $f_b(b,k,s)$, $f_k(b,k,s)$, and $f_s(b,k,s)$, respectively. Table 3 gives for each selected function of $f_b(b,k,s)$, $f_k(b,k,s)$, $f_s(b,k,s)$ the limes, the smallest possible value regarding the constraints of the corresponding variable b, k, or s and the minimum and maximum number of entries for b, k, and s. This is because the number of s is restricted by k so the maximum number of entries is of interest and the outcome of this is a right hand side asymptote. The graphs of $f(b,k,s)$ are shown in Figure 3 a-c. The ratios of f'/f, where f' is the derivate of f, are shown in Figure 4 a-c.

As can be seen when combining information of Figures 3 a-c and 4 a-c as well as Table 3 there are no major changes in $f(b,k,s)$ within the constraints. It emerges that the influence of checks outweighs that of block size and number of blocks. These results are obvious when looking at the limes in Table 3 and in Figure 3 a-c and 4 a-c. The asymptote of every function with regard to one variable differs depending on the fixed values of the other two variables, e.g. in Figure 3c it is 60 with $f_s(6,10,s)$ and 360 with $f_s(18,20,s)$.

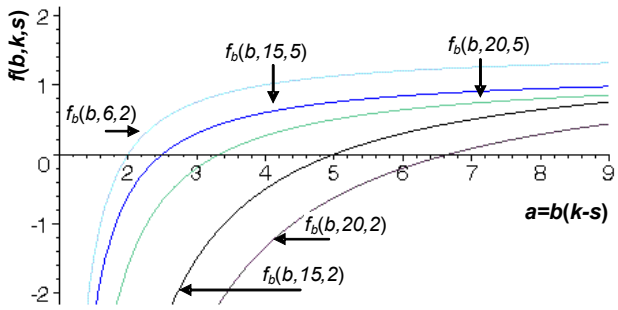

Figure 3 a: Five examples for the theoretical investigation of $f_b(b,k,s)$. In brackets the fixed values for k and s are given. The non-asymptotic area is shown.

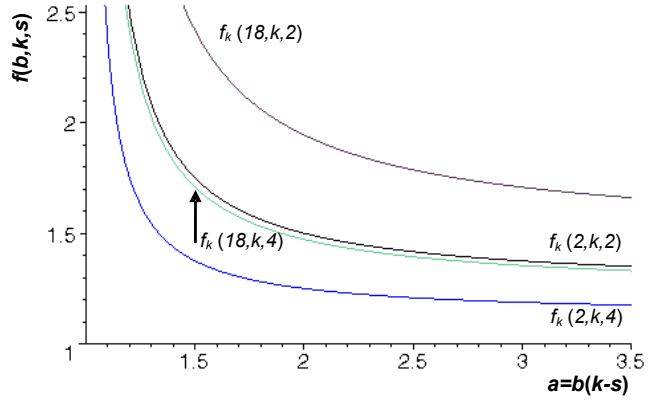

Figure 3 b: Four examples for the theoretical investigation of $f_k(b,k,s)$. In brackets the fixed values for b and s are given. The non-asymptotic area is shown.

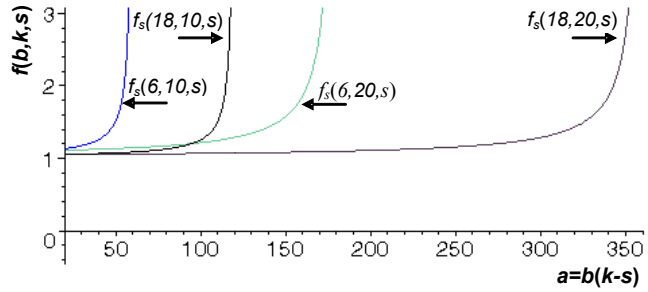

Figure 3 c: Four examples for the theoretical investigation of $f_s(b,k,s)$. In brackets the fixed values for b and k are given.

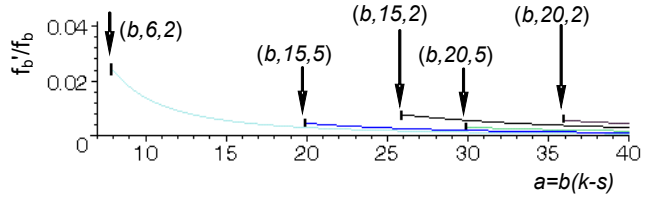

Figure 4a: Defined domain up to 40 entries of five examples for the theoretical investigation of f_b'/f_b. Fixed values of f_b'/f_b are given in brackets.

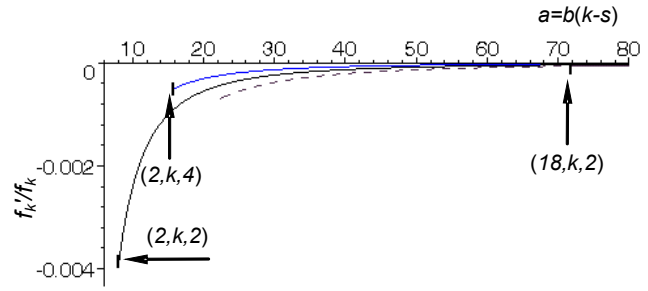

Figure 4b: Defined domain up to 80 entries of three examples for the theoretical investigation of f_k'/f_k. Not defined area of small values shown for (18,k,2) as broken line to visualise the curves progression. (18,k,4) not shown for defined domain is from 144 up and therefore in the asymptotic area. Fixed values of f_k'/f_k are given in brackets.

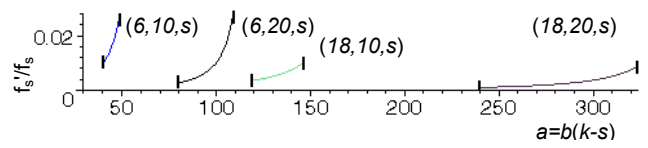

Figure 4c: Defined domain up to 340 entries of four examples for the theoretical investigation of f_s'/f_s. Fixed values of f_s'/f_s are given in brackets.

To give an impression of the complexity of $f(b,k,s)$, Figure 5 shows all values for entries plotted against $f(b,k,s)$ for all combinations of s between 2 and 7, k from 2 to 15 and b between 2 and 21. Constraints for Figure 5 were at least two checks per block, the number of plots per block was twice the number of checks and there were at least two blocks. For illustration the three functions f_b, f_k, and f_s are displayed each with one example.

To add more checks has a larger impact than to reduce the number of blocks or to enlarge the block size. The effect of using 4 instead of 2 checks, which also can be seen in Fig 4 a and b, is more pronounced than doubling from 4 to 8 checks, which has not nearly the same effect (see Fig 5).

Table 3: Examined functions (f_b, f_k, f_s), variable of interest, investigated combinations of other variables and the associated values of allowed minimum of the observed value of the variable (b,k,s) and minimum and maximum value of entries (a) are shown as well as the limes.

Function	var	Relation of interesting variable and no. of entries (a)	Number of blocks b	plots k	checks s	allowed minimum var	minimum a	maximum a	Limes[II] $f(b.k.s)$
$f_b(b,s,k)$	b	$b = a/(k-s)$		6	2	2	8	∞	1.5
				15	2	2	26	∞	1.5
				20	2	2	36	∞	1.5
				15	5	2	20	∞	1.2
				20	5	2	30	∞	1.2
$f_k(b,s,k)$	k	$k = s+(a/b)$	2		2	6	2	∞	1.25
			18		2	6	18	∞	1.47
			2		4	12	2	∞	1.13
			18		4	12	18	∞	1.24
$f_s(b,s,k)$	s	$s = k-(a/b)$	6	10		2	40	48	1
			18	10		2	120	144	1
			6	20		2	80	108	1
			18	20		2	240	324	1

I : for s the minimum number of entries is defined as $a = b * k$
II : limes $\rightarrow \infty$ for b, limes $\rightarrow \infty$ for k, limes $\rightarrow 0$ for s
var : related variable, i.e. b for f_b, k for f_k, s for f_s
a : entries

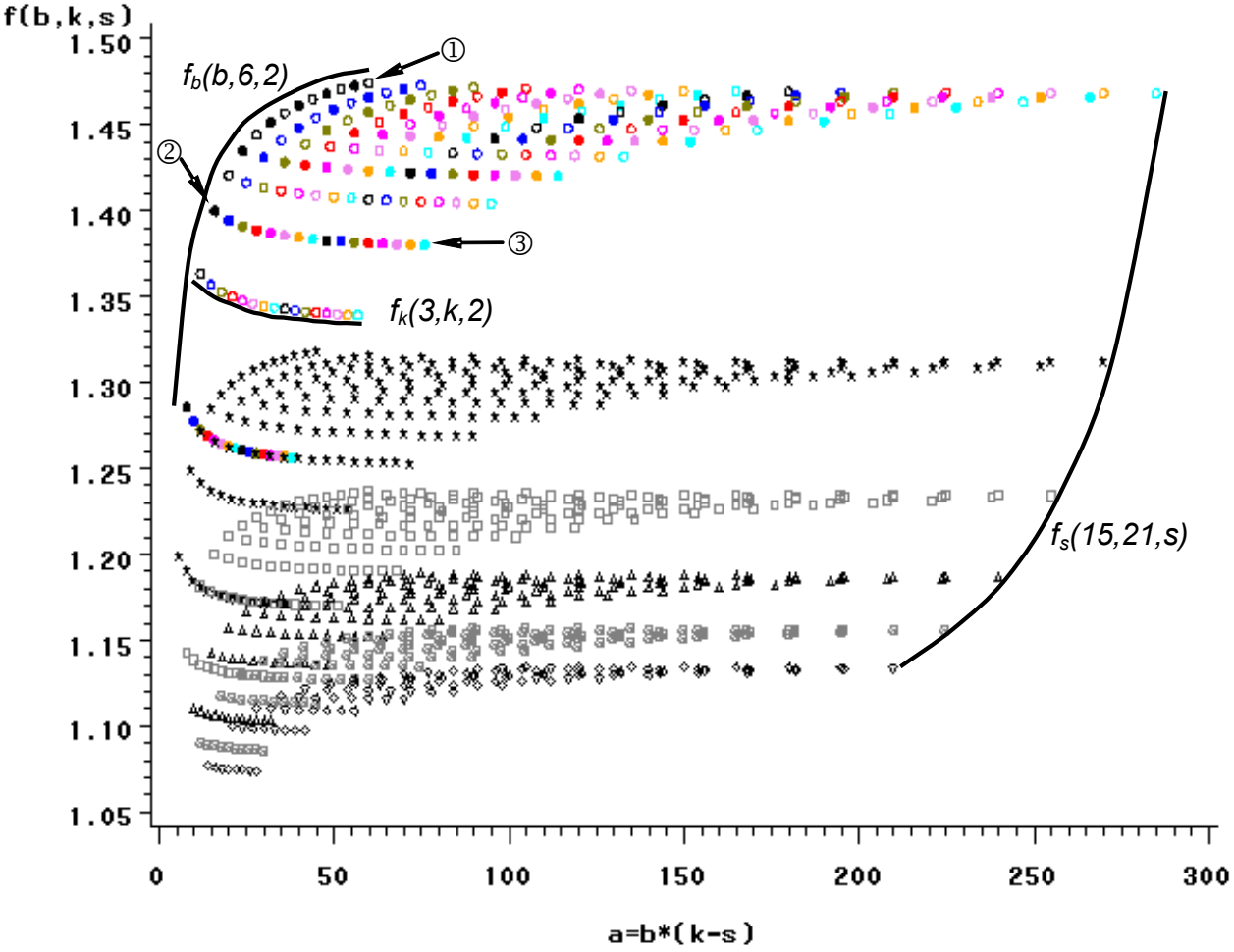

Figure 5: scatter plot of entries plotted against *f(b,k,s)* for all combination of *b* (2 to 15), *k* (6 to 21) and *s* (2 to 7) within the permissible range. Different symbols show different numbers of checks. For a given number of checks every point represents a combination of *b* and *k*.
For example with 2 checks (circle) for the same colour the alternating symbol (filled and empty circle) represents changing number of blocks and the colour represents changing number of plots per block within a symbol. One example for every type of function f_b, f_k, and f_s is given as solid line. ①–③ are three examples for values of points: ①(15,6,2), ②(4,6,2), ③(4,21,2)

Results for tested spatial combinations

Number of possible designs

The number of possible designs changes depending on the aspects mentioned in the introduction (e.g. ratio, checks) and the level of change. For example if only the ratio of entries and checks is allowed to be two or above (entries ≥ 384), while all other constraints are kept constant, the number of possible designs increases from 22 for exactly 384 entries to 107 for entries ≥ 384, while the smallest squared LSD value is the same in both cases. Keeping all constraints constant, except the size of experimental field (n_f), the number of possible designs increases from 22 to 51 and from 107 to 196 if size of experimental field changes from 480 to 400 plots. Reducing the

maximum number of checks used per block from 8 to 5, the number of possible designs decreases to 11 (exact 384 entries) and 80 (entries ≥ 384).

'Nearly isotropic' spatial combinations (major range 57/ minor range 51)

The nearly isotropic (57/51) spatial combinations all lead to designs with large blocks and shapes as quadratic as possible. The row-column-combination (RCC), defined as the number of rows (*r*) and columns (*c*) of a block, for at least the first six ranks are identical for exactly 384 entries and entries ≥ 384 within 0°, 15° and 45° rotation (not shown).

When looking at the database (not shown) of Table 4 the spatial combination with nugget of 0.0 and partial sill of 0.09 as well as 0.5 both behave quite similar. Compared to the other combinations, with these combinations a quadratic shape is more important than the number of checks. Therefore, small square blocks appear within the ten best RCC for exactly 384 entries and there is only a slight difference between the ranks of RCC for exactly 384 entries and entries ≥ 384. As shown in Table 4 the optimal RCC with 0° and 15° rotation has 8 checks and 6 rows by 4 columns and if rotation is 45° it has 4 rows by 6 columns. Hence, this combination is different from the other combinations and the number of checks is the most important criterion for a good design. This emerges from the fact that up to rank 8 all designs use 8 checks. With partial sill 0.5 and nugget 0.2 the ranks for designs are comparable to those for a partial sill 0.09 and a nugget 0.2, but shape is more important, hence certain designs with more square-like shapes but fewer checks get better ranks. With both combinations that have a nugget of 0.2, there are several further designs for entries ≥ 384 in between designs with rank 1 and 10 for exact 384 entries.

Table 4: Squared LSD and ranks for a selection of row-column-combinations (RCC) for near-isotropy for exactly 384 (=384) and at least 384 (≥ 384) entries for rotation of 0°, 15° and 45°. The selection is from the 10 best RCC according to the squares LSD among designs with exactly 384 entries. Of these RCC, only the best RCC for every number of checks are shown and values for rank 10. Additionally for every RCC, the rank for at least 384 entries is given and RCC with numbers of checks different to exactly 384 but within rank 10 are presented.

			0° Rotation						15° Rotation						45° Rotation					
Range major/minor	Partial sill	Nugget	RCC r	RCC c	LSD²	Number checks	rank with entries 384	rank with entries ≥ 384	RCC r	RCC c	LSD²	Number checks	rank with entries 384	rank with entries ≥ 384	RCC r	RCC c	LSD²	Number checks	rank with entries 384	rank with entries ≥ 384
57/51																				
	0.09	0.0	6	4	0.698	8	1	1	6	4	0.698	8	1	1	4	6	0.699	8	1	1
			6	3	0.702	6	4	4	6	3	0.702	6	4	4	3	6	0.704	6	5	5
			4	3	0.704	4	5	5	4	3	0.704	4	5	5	3	4	0.705	4	7	7
			3	3	0.708	3	9	9	3	3	0.709	3	9	9	3	3	0.709	3	9	9
			4	4	0.709	5	-	10	4	4	0.709	5	-	10	4	4	0.709	5	-	10
			12	2	0.712	8	10	13	12	2	0.713	8	10	13	2	12	0.716	8	10	15
	0.09	0.2	6	4	2.445	8	1	1	6	4	2.446	8	1	1	4	6	2.446	8	1	1
			6	4	2.489	7	-	7	6	4	2.489	7	-	7	4	6	2.490	7	-	7
			6	3	2.516	6	9	15	6	3	2.516	6	9	15	3	6	2.517	6	9	15
			3	6	2.519	6	10	16	3	6	2.519	6	10	16	6	3	2.517	6	10	16
	0.5	0.0	6	4	3.879	8	1	1	6	4	3.879	8	1	1	4	6	3.883	8	1	1
			6	3	3.900	6	4	4	6	3	3.902	6	4	4	3	6	3.910	6	5	5
			4	3	3.910	4	5	5	4	3	3.912	4	5	5	3	4	3.919	4	7	7
			3	3	3.936	3	9	9	3	3	3.936	3	9	9	3	3	3.937	3	9	9
			4	4	3.938	5	-	10	4	4	3.938	5	-	10	4	4	3.939	5	-	10
			12	2	3.958	8	10	13	12	2	3.960	8	10	13	2	12	3.976	8	10	15
	0.5	0.2	6	4	5.626	8	1	1	6	4	5.627	8	1	1	4	6	5.630	8	1	1
			6	3	5.714	6	6	6	6	3	5.715	6	6	6	3	6	5.724	6	7	7
			6	4	5.727	7	-	7	6	4	5.727	7	-	7	4	6	5.730	7	-	9
			4	3	5.857	4	9	16	4	3	5.859	4	9	16	3	4	5.866	4	9	16
			3	4	5.873	4	10	19	3	4	5.872	4	10	19	4	3	5.866	4	10	17

As can be seen in Table 5 the squared LSD values differ with every combination but the smallest squares LSD value is the same irrespective of the number of entries, i.e. the best design is the same (see Table 4). In principle the squared LSD value rises with rising partial sill and nugget. Within one spatial combination the rotation has little influence on the smallest squared LSD value, while with the maximum squared LSD value the value reduces with rotation.

Table 5: squared LSD values for nearly isotropic spatial combinations. The smallest squared LSD value (min), which is the same for exact 384 entries and for ≥ 384 entries, as well as the highest value (max) for exact 384 entries and for ≥ 384 entries are shown.

				Squared LSD		
Range	Parameter		Rot.	min	max	
major/minor	p. sill	nug.			*a*=384	*a*≥ 384
57/51	0.09	0	0	0.698	0.887	1.092
	0.09	0	15	0.698	0.885	1.092
	0.09	0	45	0.699	0.870	1.087
	0.09	0.2	0	2.445	3.245	3.626
	0.09	0.2	15	2.446	3.243	3.625
	0.09	0.2	45	2.446	3.229	3.621
	0.50	0	0	3.879	4.926	6.068
	0.50	0	15	3.879	4.915	6.064
	0.50	0	45	3.883	4.836	6.042
	0.50	0.2	0	5.626	7.285	8.601
	0.50	0.2	15	5.627	7.273	8.598
	0.50	0.2	45	5.630	7.195	8.575

p.sill: partial sill; nug: nugget; rot: rotation

The 'strongly anisotropic' spatial combinations (major range 500/ minor range 27)

The strongly anisotropic (500/27) spatial combinations displays the effects resulting from combining anisotropy and rotation of axis. In general without rotation small elongated blocks are preferred, the block width being one plot. The closer the rotation is to 45° the more square the optimal blocks are and the more checks are optimal. More checks are also adequate with rising nugget. Overall when the number of entries is ≥ 384, there are many more designs up to the design with rank 10 with exactly 384 entries and it is often only the first or the first two designs that are identical (not shown).

In Table 6 the influence of checks and the RCC becomes obvious. For example with 0° rotation, partial sill of 0.065 and nugget of 0.02 the best combinations have elongated blocks and few checks. With 45° they are squared and have many checks. When the nugget is changed to 0.2 (twice the partial sill) and the rotation is 0°, the best block design is 24 rows by 1 column and 8 checks. With 45° rotation, a design with 4 rows by 6 columns with 8 checks is best. On rank nine the first design with 6 checks (6 rows by 3 columns) occurs and no design with 4 checks is

within the first 10 ranks. The combinations perform very differently when the angle of rotation differs from zero. The behaviour depends on the relation of partial sill and nugget. As with the nearly isotropic combinations, the combination where the nugget outweighs the partial sill differs from the other combinations. This can be seen with zero rotation, where, even though the shape is elongated, the block size is big due to the need of many checks.

When examining the squared LSD values (see Table 7) with the strongly anisotropic combinations, basically the same behaviour is found as with the nearly isotropic combinations: maximum values decrease towards 45° rotation, values increase with increasing partial sill and nugget, but more rapidly. The behaviour differs when looking at minimum squared LSD and small nugget. With 0° rotation the squared LSD is much smaller than with 15° or 45° rotation. The values for 0° rotation, partial sill of 0.065 and nugget of 0.02 are between 0.343 and 0.979 compared to partial sill of 0.5 and nugget of 0.2 were they are between 1.057 and 5.951.

Table 7: Squared LSD values for strongly anisotropic spatial combinations. The smallest squared LSD (min), which is the same for exact 384 entries and for ≥ 384 entries, as well as the highest value (max) for exactly 384 entries and for ≥ 384 entries are shown.

				Squared LSD		
Range	Parameter		Rot.	min	max	
major/minor	p.sill	nug.			*a*=384	*a*≥ 384
500/27	0.065	0.02	0	0.343	0.979	1.07
	0.065	0.02	15	0.619	0.974	1.069
	0.065	0.02	45	0.655	0.929	1.057
	0.065	0.20	0	2.005	3.102	3.351
	0.065	0.20	15	2.192	3.097	3.349
	0.065	0.20	45	2.227	3.052	3.337
	0.50	0.02	0	1.057	5.951	6.538
	0.50	0.02	15	3.544	5.916	6.529
	0.50	0.02	45	3.866	5.571	6.436
	0.50	0.20	0	5.626	7.285	8.601
	0.50	0.20	15	5.627	7.273	8.598
	0.50	0.20	45	5.63	7.195	8.575

p.sill: partial sill; nug: nugget; rot: rotation

Table 6: Squared LSD and ranks for a selection of row-column-combination (RCC) for strongly anisotropy for exactly 384 (=384) and at least 384 (≥ 384) entries for rotation of 0°. 15° and 45°. The selection is from the 10 best RCC according to the squares LSD among designs with exactly 384 entries. Of these RCC. only the best RCC for every number of checks are shown and values for rank 10. Additionally for every RCC the rank for at least 384 entries is given and RCC with numbers of checks different to exactly 384 but within rank 10 are presented.

			0° Rotation						15° Rotation						45° Rotation					
Range major/minor	Partial sill	Nugget	RCC r	RCC c	LSD²	Number checks	rank with entries 384	rank with entries ≥ 384	RCC r	RCC c	LSD²	Number checks	rank with entries 384	rank with entries ≥ 384	RCC r	RCC c	LSD²	Number checks	rank with entries 384	rank with entries ≥ 384
500/ 27																				
	0.065	0.02	6	1	0.343	2	1	1	8	3	0.619	8	1	1	4	6	0.655	8	1	1
			12	1	0.355	4	2	2	8	2	0.627	5	-	3	4	6	0.666	7	-	3
			8	1	0.375	2	-	3	6	2	0.630	4	3	4	3	6	0.676	6	5	8
			24	1	0.433	8	3	5	6	3	0.633	6	4	6	3	4	0.691	4	7	14
			6	3	0.597	6	7	23	6	1	0.690	2	10	27	12	2	0.696	8	10	17
			4	6	0.661	8	10	36												
	0.065	0.2	24	1	2.005	8	1	1	8	3	2.192	8	1	1	4	6	2.227	8	1	1
			24	1	2.041	7	-	2	8	3	2.231	7	-	4	4	6	2.267	7	-	5
			12	1	2.108	4	3	5	6	3	2.265	6	7	10	3	6	2.308	6	9	13
			6	3	2.229	6	6	14	3	6	2.332	6	10	21	6	3	2.308	6	10	14
			6	2	2.302	4	10	26												
	0.5	0.02	6	1	1.057	2	1	1	6	2	3.544	4	1	1	4	6	3.866	8	1	1
			8	1	1.296	2	-	2	8	2	3.575	5	-	2	3	6	3.986	6	3	6
			12	1	1.431	4	2	3	8	3	3.595	8	2	3	3	4	4.013	4	7	10
			24	1	2.158	8	3	6	6	3	3.658	6	4	5	4	6	4.029	6	-	12
			6	3	3.380	6	7	25	6	1	3.726	2	5	8	3	3	4.107	3	9	16
			4	3	3.654	4	10	36	24	1	4.036	8	10	26	2	12	4.186	8	10	24
	0,5	0,2	6	1	3.179	2	1	1	8	3	5.168	8	1	1	4	6	5.439	8	1	1
			12	1	3.184	4	2	2	8	2	5.257	5	-	3	4	6	5.536	7	-	3
			24	1	3.730	8	3	5	6	3	5.290	6	3	5	3	6	5.618	6	5	7
			12	1	3.422	3	-	3	6	2	5.296	4	4	7	3	4	5.766	4	9	16
			6	3	5.012	6	7	22	12	1	5.725	4	10	27	4	3	5.766	4	10	17
			4	6	5.488	8	10	36												

Discussion

In simple screening experiments as performed e.g. at gene banks for preserving germination capacity, it is advisable to carefully choose a suitable design such as an augmented design. Using an augmented design with at least two checks per block allows for statistical analysis of collected data, including adjustments for local trend, and therefore the collected data become more informative for researchers and breeders.

Our study involved scrutiny of the properties of the function *f(b,k,s)* as *b*, *k* or *s* changes. One possibility to obtain information about a function is to inspect the area under the curve. To compute an area, limits of integration must be defined. For this purpose one might consider coordinates axes, limes and asymptotes. In this paper the investigated functions converged rarely against a limes, so that for each function an arbitrary point would have had to be set, for which the limes is considered to be reached. The asymptotes and axes for each function are located elsewhere. Additionally they are located beyond the range for which the function is admissible given its constraints. With the type of function investigated when looking at *k* or *s* it is also impossible to identify the value on the abscissa for which the function reaches half of its maximum. Hence for each function individual limits would have had to be defined. These problems led to the choice of *f'/f* for describing the function. From Figure 4 a-c it becomes evident that the functions are defined only in ranges with small slopes. The function *f(b,k,s)* has steep slope only for values of *b*, *k* or *s* outside the permissible range. Hence, no major changes occur within the definition range of the function when changing value of *b*, *k* or *s*, and it is difficult to predict their influence.

In addition to the basic advice to use small blocks with high spatial variability, more elongated blocks with anisotropy, more rectangular blocks with rotated anisotropy axes towards reference axes of field or isotropy, a more precise advice can be given regarding the checks. It is obvious that the influence of the checks is stronger than that of block size or number of blocks, so that it seems advisable to use many checks, although the efficiency decreases with each additional check. In context with this investigation it is recommended to use up to five checks per block.

Depending on the spatial variability, the differences in squared LSD can be marginal. The more designs need to be tested, the more computation time is needed. Hence the smaller the ranges of values for every aspect (e.g. entries to be tested), the fewer designs need to be tested, and the less choices among designs are necessary afterwards. If a design is needed when there are many un-

certainties regarding aspects such as block size, shape, number of blocks and plots, other methods may be used for investigation like tabu or simulated annealing (Eccleston and Chan, 1998) or linear programming (e.g. Bertsimas and Tsitsiklis, 1997; Cook et al., 1997).

Apart from the LSD and costs there seems to be no other meaningful criterion in context with this investigation. It is difficult to generally weight the influencing factors or specify the costs, because in individual cases it can, e.g., be more important that the number of plots on the experimental field (n_f) equals the number of plots on the experimental site (n_s) or that the number of blocks or the number of plots per block are more important than the number of entries due to costs or requirements of field equipment. So it appears unpractical to specify an objective criterion other than the squared LSD. Nevertheless, the squared LSD can and must be set into context with practitioners needs and costs. This is mostly in accordance with Paterson and Hunter (1983) who in their investigation pointed out that "the choice of block size still remains a matter of judgement". The method investigated in this paper can help in fine-tuning an augmented design. The squared LSD can be used as a decision criterion if a selection has to be made between different designs, which in terms of costs and complexity are one a par with another.

Acknowledgements

Thanks are due to the German Research Foundation (DFG) for supporting this research, grant number PI 377/5.

References

Bertsimas, D. and Tsitsiklis, J. 1997 Introduction to Linear Optimization. Athena Scientific, Belmont, Mass.

Cook, W.J., Cunningham, W.H., Pulleyblank, W.R. and Schrijver, A. 1997. Combinatorial Optimization, Wiley Interscience, New York.

Eccleston, J. and Chan, B. 1998. Design algorithms for correlated data. Pages 41-52. In: COMPSTAT 1998, Proceedings in Computational Statistics.

Edmondson, R.N. 2005. Centenary Review: Past developments and future opportunities in the design and analysis of crop experiments. J. Agric. Sci. (Camb.) 143: 27-33.

Federer, W.T., Reynolds, M. and Crossa, J. 2001. Combining results from augmented designs over sites. Agron. J. 93: 389-395.

Federer, W.T. 1956. Augmented (or hoonuiaku) designs. Hawaii. Plant Rec. 2: 191–208.

Gilmour, A.R., Cullis, B.R. and Verbyla, A.P. 1997. Accounting for natural and extraneous variation in the analysis of field experiments. J. Agric. Biol. Envir. S. 2: 269-293.

Grondona, M.O. and Cressie, N. 1991. Using spatial considerations in the analysis of. Experiments. Technometrics 33: 381-392.

Hartung, K., Piepho, H.-P. and Knüpffer, H. 2006. Analysis of genebank evaluation data by

using geostatistical methods. Genec. Resour. Crop. Ev. 53: 737-751.

McBratney, A.B. and Pringle, M.J. 1999. Estimating Average and Proportional Variograms of Soil Properties and Their Potential Use in Precision Agriculture. Precision Agriculture 1 : 219-236.

Papadakis, J.S. 1937. Méthode statistique pour les expériences en champ. Bull. Scientifique de l'Institut d'Amélioration des Plantes de Thessalonique. 23.

Patterson, H.D. and Hunter, E.A. 1983. The efficiency of incomplete block designs in National list and recommended list cereal variety trials. J. Agric. Sci. 101: 427-433.

Payne, R. and Green, P. (Eds.) 1998. COMPSTAT 1998 - Proceedings in Computational Statistics. 13th Symposium held in Bristol, Great Britain.

Peterson, R.G. 1994. Agricultural field experiments – design and analysis. Marc Dekker Inc. New York.

Pringle, M.J. 2002. Field-Scale Experiments and Site-Specific Crop Management. PhD Thesis. The University of Sydney.

Santos, A.H., Bearzoti, E., Ferreira, D.F. and da Silva Filho, J.L. 2002. Simulation of mixed models in augmented block design. Sci. agric. (Piracicaba, Braz.). 59: 483-489. [http://www.scielo.br/pdf/sa/v59n3/10580.pdf]

Schaalje, G.B., Lynch, D.R. and Kozub, G.C. 1987. Field evaluation of a modified augmented design for early stage selection involving large number of test lines without replications. Potato res. 30: 35-45.

Schabenberger, O. and Pierce, F.J. 2002. Contemporary statistical models for the plant and soil sciences. CRC Press LLC, Boca Raton, FL.

Scott, R.A. and Milliken, G.A. 1993. A SAS program for analyzing augmented randomized complete-block designs. Crop Sci. 33 : 865-867.

Solorzano, V., Gilmour, S.G., Phelps, K. and Kennedy, R. 1997. Assessment of suitable designs for field experiments involving airborne diseases. J. Agric. Sci. 129 : 249-256.

Watson, S. 2000. Spatial dependence and block designs in spaced plant herbage trials. J. Agric. Sci. 134 : 245-258.

Wolfinger, R.D., Federer, W.T. and Cordero-Brana, O. 1997. Recovering information in augmented designs, using SAS Proc GLM and Proc MIXED. Agron. J. 89: 856-859.

4 General Discussion

Initially gene banks were installed primarily to preserve biodiversity, so accessions were described mainly to control gene drift. For this purpose many phenotypic traits were used and still are today. Complex analyses, as possible with current powerful computers and software, e.g. assessing information for accessions over different years and sites, is a recent development, as is the use of molecular methods. With the rising possibilities, the demands by potential users of accessions, e.g. breeders, increased. But, unfortunately, the budgets of gene banks decrease, even in developed countries, as do the governmental budgets. The tight financial resources of gene banks lead to reduced capacity, so that more and more activities are reduced to the absolutely necessary: preserving germinability and preventing gene drift. Therefore the usefulness of gene bank databases for potential users deminishes. The fundamental potential of accessions regarding breeding still remains. Evaluation on a genetical level is cost-intensive. So a good statistical analysis of phenotypic data is advisable in order to make the best possible use of both phenotypic and genetic information. But the phenotypic data, as available today, are often not fully satisfactory for proper statistical evaluation: standards and experimental design are missing; randomness of yearly chosen accession is assumed, but may sometimes be insufficient, e.g. after collection missions; the scales used to asses ratings are mostly ordinal. The objective of the investigations of this thesis therefore was to develop and assess methods to better exploit the phenotypic information of accessions.

The discussion starts with a view on fixed and random effects in mixed models and then turns to geostatistical methods used for analysing gene bank data. As next topic the possibility to use augmented designs is considered, followed by a comparison of both approaches and a glance on the use of geostatistical methods for obtaining optimal designs. Statistical problems with ratings are the next topic, followed by a discussion of the issue of connecting gene bank data over environments. A brief look at multivariate methods and quantitative traits are the last topic before the conclusion.

4.1 Accessions and blocks as fixed or random effect in the mixed model

A common method is to analyse experimental data by mixed model analysis, where the response is modeled by fixed and random explanatory factors plus a residual error that represents all the variability of the response not accounted for by the explanatory variables. With each mixed

model that is developed to analyse a data set, the question arises whether an effect is fixed or random. Conventionally, an effect is called fixed if the levels in the study represent all possible levels of the factor, or at least all levels about which inference is to be made. This can include regression models where the observed values of the explanatory variable cover the entire region of interest. Effects are random if they are used in the study to represent only a sample (ideally a random sample) of a larger set of potential levels with a probability distribution (Littell et al., 1996; Piepho et al., 2003). It is well recognised, however, that the assumption of random effects may be advantageous also in settings were interest is only in the observed levels of a factor but the number of levels is large (Piepho et al., 2006). One should think carefully whether to consider a certain effect as random, if it is represented by only few levels, say less than 10, e.g. if only two types of fertiliser are tested. In such cases, variance components can be only be estimated poorly, resulting in inefficient effect estimates. On the other hand if there are very many levels, the large number of fixed parameters might lead to estimation problems, e.g. when a threshold model is used (c.f. Paper 2).

Within this thesis accessions were taken to be random in Paper 1 for it is assumed that accessions grown within one year are a random sample from the accessions stored at the gene bank. This assumption does not strictly hold in years were, for example after a collecting mission, the collected accessions are grown for multiplication. With Paper 5 accessions were taken as fixed, mainly because this facilitated derivation of good experimental designs, as theory is much better developed in a fixed effects setting. Piepho and Williams (2006) show that designs optimal for a fixed effects analysis may also be advantageous when analysis is done by BLUP (see below). There are some suggestions for optimal design when treatment effects are random (Bueno and Gilmour, 2003; Cullis et al., 2006), but this is a recent development. A practical difficulty with these approaches is that strong assumptions need to be made regarding the genetic variance structure. With the analysis of the barley collection over several years by a threshold model in Paper 2, it was not feasible to take accessions as fixed, due to the large number of parameters to be estimated and the resulting breakdown of asymptotic theory (McCulloch and Searle, 2001). Thus, accessions had to be taken as random. This discussion shows that the decision as to whether an effect is fixed or random often needs to be made based on a number of different, possibly conflicting considerations. Hence, in practice the final decision will often need to be a pragmatic one, while adhering to very strict rules is not generally feasible. Nevertheless, for gene banks, where the number of accessions is very large, it will almost invariably be advantageous to take accessions as random when it comes to analysis.

Best linear unbiased estimation (BLUE) and best linear unbiased prediction (BLUP) are standard methods to estimate fixed and random effects, respectively. Both BLUP and BLUE assume known variance components (Searle et al., 1992). In practice variance components are replaced by their estimates, obtained preferably by Restricted Maximum Likelihood (REML), resulting in empirical BLUP and empirical BLUE. Besides the decision criteria mentioned above, BLUE is preferable if unbiased estimates of differences between specific pairs of the explanatory variable (accessions, rater, etc.) are of interest, since BLUP of a specific difference is biased. By contrast, if the aim is to rank the estimated effects, BLUP is advisable (Searle et al., 1992; Smith et al., 2005) since it maximises the probability of a correct ranking (Searle et al., 1992), and also has the property of shrinkage towards the mean and thus involves bias.

Within this discussion the block effect takes up an exceptional position. In field experiments with complete blocks there are usually only few blocks (<10), so it is not advisable to take block effects as random. By contrast, when incomplete blocks are used, the number of blocks is typically so large that it is preferable to take blocks as random. An exception are augmented designs, where the number of incomplete blocks may be rather limited, unless the number of accessions is very large compared to block sizes. If the block effect is taken as fixed, only the intra-block-information can be used. If the block effect is assumed to be random, one can use both the intra- and the inter-block-information (Paterson and Thompson, 1971) thus increasing efficiency. The question of whether to take blocks as a fixed or a random effect is discussed widely in the literature (Samuels et al., 1991). As the main task of blocking is the reduction of the unexplained (random) error, so that the differences of treatments are estimable with minimal error, it is advisable for gene banks to use blocks. And it is reasonable to use fixed block effects if the experimental field design has only few blocks (less than 10) and interest lies on the genetic value of individual accessions.

4.2 Geostatistical methods for optimising usage of gene bank data

If interest lies on the assessment of the genetic value of an accession, it is necessary to correct the value observed in the field for environmental effects. This can be achieved by using a proper experimental design. Since the barley data of the gene bank at the IPK had no experimental design and only few standards, a geostatistical approach was used to correct the obtained values for field heterogeneity by estimating the environmental effects from the neighbouring accessions (e.g. Isaaks and Srivastava, 1989). The complete model was formulated including information from standards and accessions, partitioning variance into a genetic part, an environmental part,

and a random error, each of the three assumed to be independent of one another. The environmental values were assumed to be spatially correlated. The selected model for spatial correlation was embedded into a mixed model, which allowed an estimation of the genetic effect by BLUP (Stroup and Mulitze, 1991). Usually the choice of a covariance model was not critical, as the considered models differed only slightly in most cases. The spherical and the exponential model were preferable in most cases. This is in contrast to the findings of Pilarczyk and Tomaszyk (2006), who found the linear model to be the most suitable model. This might be due to the shape of the plots in their trials and the fact that all plots were laid out in long single rows. The plots were elongated and the direction of spatial analysis was unidirectional and perpendicular to the direction of cultivation. With the barley data of the IPK the spatial analysis was isotropic and the plots much smaller and of square shape. For variogram estimation and adjustment, the first points of the empirical variogram should be explained well by the fitted theoretical variogram, as it is normally impossible to properly explain the entire variogram. Reducing the maximal lag distance improves the fit for short distances when using non-linear regression. Since spatial mixed models use all of the original data, excluding large distances is not possible. Therefore the spatial component obtained by non-linear regression was considered as fixed within the mixed model analysis, thus leading to better adjustment of the lower distances of the theoretical variogram and to better convergence behaviour. A nugget effect (residual error) was needed for nearly all traits. For the method to work well with unreplicated accessions, a good estimate of the two variance components associated with the nugget is essential (i.e., genetic and non-genetic), and this requires replications of at least one check. The small number of check plots within the available data made it difficult to accurately dissect genetic and non-genetic components in the nugget effect. Field designs for spatial analysis that allow for checks, as used with Australian breeding programs (Cullis et al., 1989; Eccleston, 1998), improve the analysis. But only a few gene banks cultivate checks in regularly spaced intervals every year. A larger number of gene banks have at least some checks grown without an underlying design or grow some replicated accessions, e.g. on border plots, that could be used as checks (personal communications from several gene banks, 2003) to better estimate the unreplicated accessions. If checks are not available or sparsely replicated, one may still use geostatistical methods introduced in Paper 1 to remove spatial effects. This leads to accession estimates that include both genetic and non-genetic effects. While this will be less informative than BLUP of genetic effects, it may still be more efficient than using uncorrected raw data. If the use of a mixed model is abandoned, however, the possibility of separating the genetic effect from the residual error is lost.

An alternative approach, not investigated in this thesis, is the analysis of first differences as recommended by, e.g. Besag and Kempton (1986), Williams (1986), and recently Wu and Dutilleul (1999). The ANOFT-Program (Schwarzbach, 1984), which is based on second differences, is used by some German plant breeders. Wu and Dutilleul (1999) point out that a model with nugget is advisable, which is in agreement with the analysis in Paper 1.

It should be stressed, however, that a geostatistical analysis can be of some help when the chosen design has some deficiencies, but it cannot be expected to salvage an entirely inappropriate or missing design.

4.3 Augmented designs for optimising gene bank data

Another option to conduct experiments with unreplicated accessions are augmented designs. Federer (1956) introduced augmented design for the similar problem of testing many unreplicated test lines in early breeding stages. Augmented designs combine the option of testing many unreplicated entries with a well-known statistically analysable design that includes checks connecting the blocks (or rows and columns). Thus it can be decided to which degree the observed differences are based upon genetical or environmental causes. An augmented design is based on any of the common replicated designs and its randomisation principles. The checks are randomised according to the chosen design. This design is then augmented by entries (i.e. accessions), which are in general not replicated. Kempton and Gleeson (1997) and Petersen (1994) supply a good introduction to augmented design. The chosen design must be adapted to the examined crop and the conditions of the gene bank. Regarding the number of replicates for checks, Yates (1932) proposed the square root of n, where n is the number of entries, while Kempton and Gleeson (1997) recommend a frequency of one check in five or more accessions. In gene bank context, it might also be of interest to replace some check varieties by accessions that are in higher demand due to a larger frequency of inquiries to the gene bank. Cullis et al. (2006) recently proposed the p-rep design that allocates a certain percentage of each block for repeated test lines and therefore does not need special check varieties. But checks may be required to ensure connection of data across trials. In the recent years augmented designs have been enhanced, e.g. by mixed model analysis (Federer et al., 1998), and due to high performance computers and software programs, complex analysis procedures have become available. So augmented designs represent a well manageable tool to separate genetic and environmental effects. It is likely that over time augmented designs and associated analysis software will be adapted to gene banks needs. Design tools for gene banks should include methods for selection of additional accessions

that ensure the connection of experiments over time in one place and within one year over sites, i.e. coordination between gene banks, if desired. It also might become possible to determine the optimal number of different checks and replications, based on experiences obtained in previous years. In relation to the information gained, the additional effort for gene banks to implement augmented designs is small. That this additional effort is necessary and desirable emerges from articles like that of Haussmann et al. (2004). Especially in times of decreasing financial resources it is worthwhile to obtain the largest benefit possible from evaluation data routinely collected by gene banks. Through the inclusion of checks and proper experimental design, augmented designs offer an economical and easy-to-implement option for sound testing of many accessions.

4.4 Similarities and differences between design and analysis of geostatistical methods and augmented designs

If a chosen field design is based on randomisation theory, e.g. an augmented design, the model for statistical analysis derives from randomisation. With augmented designs, as with other field experimental designs, the large scale trend is handled through blocking and the small scale trend at the scale of the experimental unit within blocks is coped with through randomisation. If, regarding the spatial field trend, an inadequate design is chosen, the error variance increases and therefore it becomes difficult to detect treatment differences (Schabenberger and Gotway, 2005). In other words: randomization ensures that unaccounted effects, such as spatial trends, are balanced out. Another way to account for spatial trends follows the assumption that two experimental units closer together are more similar and that similarity decreases with increasing distance. Thus the assumption of independence of two experimental units as arises in randomisation-based models is abandoned. The researcher can model the fixed effects by adding or deleting terms by means of which the large scale trend is described and through modelling the covariance structure of the random error term all small scale trends can be handled. This includes the necessity to develop a suitable model and select among several options. But whether the chosen model is “the optimal choice” can never be stated without doubt. The gain in efficiency when using geostatistics instead of classical field experimental designs depends on how well the large and the small scale trends can be assessed via modelling (Schabenberger and Pierce, 2002).

Both methods, classical field experimental designs (e.g. augmented design) and geostatistical approaches, try to reduce random error to optimise the obtained information about the genetical effect of an accession. With both methods checks are necessary and the amount of check

varieties and replicates of every check variety is similar to obtain comparable results. With geostatistical methods it is indeed possible to omit checks, but this reduces the accuracy when accessions are unreplicated. Combining both methods seems to be a promising route in gene banks because accession material is diverse and many influences, e.g. pests, are not under the researcher's control.

4.5 Using geostatistical models for finding optimal designs

So far in this discussion, geostatistical methods were considered for analysis of evaluation trials. It is also possible to optimise an experimental design with a particular geostatistical analysis in mind. Finding good designs requires valid parameterization of the geostatistical model. It is not guaranteed that the selected design will still perform well, when the assumed geostatistical model does not hold. It is therefore worthwhile to follow an intermediate path by using geostatistical methods to optimise block size and shape with a classical linear model analysis in mind. This approach is similar in spirit to Williams et al. (2006), who optimise row-column designs assuming a linear spatial model, but allow for an efficient classical row-column analysis as a fall-back option, when the assumed spatial model does not hold for the data at hand after the trial. The purpose of Paper 5 was to help the researcher to find the best field layout for an augmented design, i.e. a layout that yields small least significant differences (LSD) with regard to different aspects of an assumed geostatistical model, e.g. degree of anisotropy and rotation of the major axis of anisotropy (c.f. Paper 5). As these aspects interact with one another, they need to be considered simultaneously. Based on the calculations performed, up to five checks per block are recommended. The nearly isotropic combinations lead to designs with large quadratic blocks. The strongly anisotropic combinations display effects resulting from the combination of the degree of anisotropy and rotation of anisotropy axes. Without rotation, small elongated blocks are preferred. The closer the rotation is to 45° the more square blocks and the more checks are appropriate. The squared LSD is a meaningful optimization criterion, which can and must be set into context with the practitioners needs. The method investigated in Paper 5 can help in fine-tuning an augmented design. The squared LSD can be used as a decision criterion if a selection has to be made between different designs, which in terms of costs and complexity are on par with another.

4.6 Ratings

Plant disease severity and other traits are often visually scored, using either a percentage scale or

an ordinal scale, and it is not always obvious which scale is preferable. Conceptual and computational difficulties are caused by ordinal ratings. Thresholds for these scales are rarely accurately defined but mostly descriptive and may change over time (more or fewer or even different thresholds). In the most favourable case they are based on an underlying genuinely metric scale with clearly defined class thresholds, but the true class means on that underlying scale are usually unknown. The compatibility of two scales may not always be given. Finally, most statistical methods as used for metric data are not strictly valid. Therefore ordinal data need to be analysed by less informative nonparametric methods that are available (Shah and Madden, 2004). With percentage ratings most problems associated with ordinal ratings do not occur, even though one needs to account for heteroscedasticity and non-normal distribution of data (Piepho, 1999; Shah and Madden, 2004). Furthermore, one uses a larger number of values with percentages than with ordinal ratings (e.g. one hundred versus nine) which was shown to result in more accurate disease assessment.

Ratings were considered in Paper 1 and 2: most ratings were assessed on nominal or ordinal scales; the scales were mainly descriptive and only some had an underlying percentage scale; some rating values were given as ranges, e.g. 2-3, so an *ad hoc* decision had to be made when converting these data to single scores, which always is problematic; scales required complex transformation and normalisation to fit standard statistical procedures. So the third paper directly addressed the problem of assessment of three rating methods (1%-steps, 5%- steps, 9-point rating) assessed by untrained persons (Group A) and persons experienced in rating (Group B). Every person had to rate several computer-generated pictograms of diseased grain leaves. The estimates of Group B were always closer to the real disease value than those of Group A. The highest accuracy was found with Group B using the 1%-scale and with Group A using the 5%-scale. This is comparable to the findings of Hau and Kranz (1996). And the results of Group A correspond to the results of Schumacher et al. (1995) and Hau et al. (1989). They found that with a percentage scale many raters tend to use values that are multiples of 5% or 10%. But raters will, at least to a certain degree, use the entire range of the 1%-rating scale. For the time needed per leaf assessment the trained group was fastest when using the 5% rating scale. One decides faster when there are fewer possibilities due to wider distances between class thresholds and/or the scale feels “familiar” (e.g. 5%-rating scale). Moreover, raters feel uncomfortable with uncommon thresholds as given with log-divided scales and tend to linearise scale intervals (Forbes and Korva, 1994). Most raters preferred 9-point rating, especially when untrained. Also the chance to give the right answer is felt to increase with decreasing number of possible answers

(here: 9 versus 100). But from a statistical point of view both percent ratings performed better than the ordinal rating scale. Generally the closer the scale of collected data is to a ratio scale with normal distribution, the more powerful methods are available for analysis. Therefore, percentages perform better than ordinal ratings and more ordered classes are better than fewer. It is also obvious that the scale used needs to fit the demands of the investigation (Campbell and Madden, 1990, p. 112). But one should keep in mind that if interpretation is to be possible even after years or over locations, percentages are unique and therefore more informative especially if statistical analysis is required, while the definition of ordinal rating scales might change over time and be forgotten later. The smaller the intervals of the used scale are, the better the statistical properties of the resulting data. One may have to overcome one's inhibitions to decide on a definite number, but statistically the possible error made by the rater is calculable and usually smaller than with rougher methods, which is a main result of Paper 3. So directly rating percentages whenever possible leads to smaller overall estimation errors, and with proper training, e.g. using an appropriate computer program, accuracy and precision (see below) can be further improved.

Once percentage estimates are available, one can always derive the associated ordered class for a rating scale, if needed. Conversely, transforming ordinal data back to a metric scale is strictly impossible and can be implemented only by making some *ad hoc* assumptions. An added advantage of using percentage estimates is that the comparison between data sets of different origin (by different experimenters) becomes easier and meaningful. Therefore it is strongly suggested that gene banks, as well as researchers in general, assess traits on a percentage scale wherever possible. The results of this thesis suggest that a 5%-rating scale can be recommended regarding rater's preference and in terms of accuracy.

4.6.1 Ratings in phytopathological context (accuracy and precision)

The influence of rating scale, rater, and disease on the obtained rating value is a main objective in phytopathology. Although it is generally well known by phytopathologists that direct percentages or even better metrical data should be generated, many phytopathological investigations are done using ordinal ratings with underlying percentage scales. In Paper 3, accuracy and precision as attributes of disease assessment where considered. Accuracy describes the closeness of a sample estimate (E) to the true value (T), whereas precision refers to the repeatability (Campbell and Madden, 1990, p.110). These terms are comparable to variance (precision) and bias (accuracy) in statistics. Together variance and bias determine the mean squared error (MSE) which in

statistics is frequently used to assess the performance of an estimator. The MSE is defined as

$$\text{MSE}(E) = \text{variance}(E) + (\text{bias}(E))^2$$ (Agresti, 2002, p.85).

The mixed model analysis utilised in Paper 3 can be looked upon as an extension of correlation or regression analysis which are often done on a single rater basis (Nutter et al., 1993; Newton and Hackett, 1994; Nita et al., 2003) in phytopathology. The mixed model analysis applied to the transformed data resulted in significant influences of leave, rater, and method. With Group B both percentage ratings led to smaller variance estimates. With Group A the method had no significant influence, but also the percent ratings had the smallest variance estimates. A further extension are the generalised linear mixed models that can be applied when transformation of data fails to fulfil the requirements for mixed model analysis (Madden, 2002). Aspects as shape, size, colour, and intensity of disease that are also influencing the rating value (Hau and Kranz, 1996) were not investigated.

4.7 Connection over years and locations

Joint analysis of trials over years and locations is widely used in plant breeding (Hill and Rosenberger, 1985; Yan and Rajcan, 2003). This type of analysis may be of special interest in a gene bank context, as it is possible even if different designs were used and different checks and accessions cultivated (Federer et al., 2001). And since more efficient utilisation of gene bank accessions in breeding programs is worthwhile (Duvick, 1984; Williams, 1991), joint analyses are desirable to obtain more reliable estimates for accessions, which therefore become more interesting to breeders. A general requirement for such a joint analysis is that all data are connected (Searle, 1987), no matter whether a single experiment or a series of experiments over time or different locations are analysed. It is desirable to have as many similar checks and connecting accessions as possible. Connected data offer a better basis for decision as to which accession is of interest for a certain purpose, even though for two locations the rank correlation between accessions may be very different for different traits (Annicchiarico et al., 2000). Different rating scales, e.g. due to changes over years, cause problems. This complicates the integration of multi-year data into a single score per accession. For metrical data an appropriate linear model can be fitted, for ordinal rating data a threshold model (McCullagh and Nelder, 1989; Piepho and Kalka, 2003) is applicable. In this thesis the standard threshold model (c.f. Paper 2) was extended to jointly analyse data from two different ordinal rating scales, and it was based on the assumption that the two rating scales are anchored in a common threshold. The threshold model may also be used when a metric scale, such as a percentage scale (Piepho, 2002), underlies the observed

ratings. In this case, the thresholds need not be estimated, but follow directly from the definition of the rating scale. A critical point then is the distribution assumption on the percentage scale. While for a latent scale, it can be assumed that there is a monotone transformation to normality, this assumption is not generally tenable, when the metric scale is not latent. If transformation of the percentage scale to normality can be found (Piepho, 2003b), the thresholds may be transformed accordingly. Adaptive quadrature (Pinheiro and Bates, 1995) was used to numerically integrate the likelihood over the random effects in the linear predictor. In the mildew example (c.f. Paper 2) using adaptive quadrature a single quadrature point was adequate, which is mainly due to the large number of accessions tested each year. In general it is recommendable to increase the number of quadrature points until the change in parameter estimates becomes negligible. Accession-by-year interaction was not separated from experimental error, as this would require replicate data for at least some accessions and/or a geostatistical model for within-trial variation (Hartung et al., 2006). Furthermore computing time would increase dramatically due to the need to integrate not only the genetic effect, but also the random accession-by-year interaction out of the likelihood. So if gene banks are interested in joined analysis, the data and thresholds of rating scales need to be connected. So in order to be able to use the threshold procedure, care should be exercised when redefining rating scales. It will generally be advantageous to have partial agreement between categories on all used scales. Further research on threshold models is necessary, if one is interested in analysing more complex data structures, e.g. due to incomplete blocking or specific pedigree structure (Piepho and Pillen, 2004).

4.8 Multivariate methods and mapping of quantitative traits

Multivariate methods are used widely, e.g. to analyse genomic data. In plant breeding, multivariate methods can be used to group genotypes according to genetic distance. The most prominent example is hybrid breeding, where multivariate methods can be used to form heterotic groups (Reif et al., 2005). With gene banks the main interest is in controlling and preventing gene drift and developing core collections (Knüpffer and Hintum, 2003, p. 260). Cluster analysis and principal component analysis, which are multivariate methods, are used to group accessions. Grouping can be done based on phenotypic data or genotypic data or a mixture of both. For example cluster analysis and principal component analysis are used to investigate qualitative traits in PGRs (Rojas et al., 2000; Ortiz et al., 2001; Knezović et al., 2005) but are much more common with genetical markers (Schena, 2003; Wang et al., 2006). Phenological and morphological traits, which are assessed on several metrical, ordinal and/or nominal scales, are the basis of

these investigations. To obtain an accurate grouping, a sound database is fundamental. It is advisable to have similar measurement levels for different characteristics to prevent bias and to ensure similar influence of all variables on the outcome of a multivariate analysis. So the data may need transformation (Digby and Kempton, 1987; Backhaus et al., 2000, p.345). If phenotypic trait data are the basis of such an analysis, their measurement level and variance can be very diverse. In addition phenotypic data are typically very noisy due to multiple sources of random variation, including genotype-by-environment interaction. Poor phenotypic data can lead to erroneous grouping of genotypes and thus to suboptimal decisions in the management of core collections, exacerbating the problem of genetic drift.

While metrical data are preferable to ordinal data, many traits are often assessed on an ordinal scale. In these circumstances, a threshold model is a promising option for extracting information. During the course of this thesis some barley traits were investigated by cluster and principal component analysis. Similarity of measurement levels was attained by restriction to ordinal traits, by use of the threshold model to obtain effect estimates on a latent scale (leading to similar scales for all traits) and by the standardisation to unit variance of effect estimates obtained from the threshold model. Since the results showed no clear grouping, they were not presented. It is expected, however, that analyses that use estimates of effects on the latent scale of the threshold model, are more powerful than direct analyses of ordinal data by multivariate methods.

Another important use of phenotypic information from gene banks is quantitative trait loci (QTL) mapping (Lynch and Walsh, 1998) and association mapping (Yu et al., 2006). While most QTL mapping procedures are based on classical linkage analysis for a segregating population derived from a single cross, association mapping procedures, as developed mainly in the field of human and animal genetics, exploit marker-QTL associations within a larger pedigree, thereby achieving more power compared to classical linkage mapping methods. To fully exploit the power of these methods, it is very worthwhile to study large populations. In this context, gene bank data are expected to gain increased importance in the future. The most critical issue, again, is reliability of the phenotypic data. Association mapping procedures search for associations among the genotype, as assessed by markers or more recently by gene expression profiles (Kendziorski and Wang, 2006; Kendziorski et al., 2006), and the phenotype. The more reliablly the phenotype can be measured, the more powerful any mapping approach will be. Conversely, when the phenotypic database is poor, any attempt to detect marker-trait associations is doomed to fail. So not only the precise interpretation of the analysis of the genome (microarrays, QTL-

mapping, etc.) is important, but first and foremost the precise generation and evaluation of phenotypic data is crucial.

4.9 Conclusion

To obtain data on phenotypic traits of accessions that are satisfactory for proper statistical evaluation checks, experimental design, and randomness of yearly chosen accession have to be warranted. Additionally, metrical and percentage scales should be used.

5%-rating scales are advisable in respect to rater's preferences and in terms of accuracy, but a wider range of scales should be investigated in further studies. The focus should be on scales composed of different threshold ranges for different infestation intensities, e.g. 1%-steps from 0% to 15% and 5%-steps up to 100% infestation. Aspects such as shape, size, color, and intensity of disease lesions that are also influencing the rating value (Hau and Kranz, 1996) could be included. The effect of computer-supported training of raters is worth more intensive study, as it is known to improve ratings. It is also advisable to give raters insights as to their effect on accuracy and precision of the rating particularly from a statistical point of view. This will help raters to understand and accept the necessity of small distances between thresholds.

Mixed model analysis offers precise assessments of genetic effects as well as of rater effects. This method is also recommendable if gene banks are interested in joint analysis of their phenotypic data. In combination with the thresholds model for categorical rating scales presented in Paper 2, the data become more reliable. In order to be able to use a joint analysis it is important to ensure connectedness of the design by ensuring a certain number of accession and checks replicated over years. It is also important to have an agreement of scales within a gene bank over years and between gene banks that want to cooperate. The squared LSD, as can be obtained by mixed model analysis, can be a help in selecting accessions interesting to breeders. In addition, as shown in Paper 5, this criterion can help to decide between different designs, which in terms of costs and complexity are on par with another.

The implementation of augmented designs in gene banks seems promising as their application is established and likewise is their analysis. Further research is necessary regarding the optimal number of different checks and replications per check, e.g. based on experience gained in previous years. When using geostatistical methods, there is no certainty whether the right model is chosen and quality depends on how well large and small scale trend are modelled. The advantage

of geostatistical methods is their ability to reflect and exploit the spatial variability. It is difficult so state which of the two approaches is preferable. Both are expected to result in considerable gains compared to the current practice of no or only rudimentary analyses of phenotypic data as practised by many gene banks. Combining both methods seems to be a promising route for gene banks, as accession material is diverse and many influences, e.g. pests, are not under the researcher's control. Regardless which method is chosen, geostatistical methods, classical field experimental designs or a combination of both, it is essential to adopt one of these options. Especially in times of decreasing financial resources it is worthwhile to invest money to gain the largest benefit possible from evaluation data routinely collected by gene banks.

The geostatisical approach of first differences (Besag and Kempton, 1986; Williams, 1986; Wu and Dutilleul, 1999), which was not investigated in this thesis, should be explored in context with gene bank data, since it is an alternative method to account for spatial variance. Its main advantage is the simplicity of the resulting mixed model, which often involves a linear variance structure that is straightforward to fit. Due to this advantage, it may be preferable over other geostatistical methods that require more choices and selection steps on the part of the user.

Research on suitability of multivariate methods like cluster analysis or principal component analysis for gene bank data should be continued. The use of the threshold model as well as percentage or metric scales is expected to improve the power of these methods, when applied to phenotypic gene bank data. Since the aim of many genome analyses is to predict phenotypic expression from genomic data, it is obvious that less precise phenotypic data lead to poor analysis, and therefore to rougher differentiation between genotypes. So precise assessment of field data is as essential as accurate molecular data. Good experimental design as well as use of checks and suitable scales are prerequisites to achieve this aim.

5 Complete reference list

Agresti A (1984) Analysis of ordinal categorical data. Wiley, New York.

Agresti A (2002) Categorical data analysis. Wiley, New York.

Annicchiarico P, Pecetti L, Boggini G, Doust MA (2000) Repeatability of large-scale germplasm evaluation results in durum wheat. Crop Science 40:1810-1814.

Anonymous A 2006. Biography of Nikolai I. Vavilov. http://www.vir.nw.ru/history/vavilov.htm .....Last visit: 04/10/2006

Anonymous B 2006. Press Release 19 June – Svalbard. http://www.nordgen.org/ngb/morenews.php?n_id=15&lng=en ..... Last visit: 04/10/2006

Anonymous C 2006. Department: Genebank – Activities. http://www.ipk-gatersleben.de/en/02/02/ ..... Last visit: 04/10/2006

Anonymous D 2006 Conservation by the numbers: Reducing genetic drift in crop gene bank collections. http://www.cimmyt.org/english/wps/news/2006/jan/conserv_byNumbrs.htm Last visit: 04/10/2006

Anonymous E 2006. The European barley database. http://barley.ipk-gatersleben.de/ebdb.php3 ..... Last visit: 04/10/2006

Anstaltsleitung der BAZ (Hrsg.) (2004) Jahresbericht 2003 der Bundesanstalt für Züchtungsforschung an Kulturpflanzen (BAZ), Koch-Druck, Halberstadt. 99-109. http://www.bafz.de/JB2003/pgr.pdf ..... Last visit: 04/10/2006

Arfken G (1985) Orthogonal polynomials. Mathematical methods for physicists, 3rd ed. Orlando, FL: Academic Press, p. 520-521.

Auclair J, Boland GJ, Cober E, Grael GL, Steadman JR, Zilka J, Rajcan I (2004) Development of a new field inoculation technique to assess partial resistance in soybean to Sclerotinia sclerotiorum. Canadian Journal of Plant Science 84:57-64.

Backhaus K, Erichson B, Plinke W, Weiber R (2000) Multivariate Analysemethoden. 9.Auflage, Springer, Berlin.

Bertsimas D, Tsitsiklis J (1997) Introduction to linear optimization. Athena Scientific, Belmont.

Besag J, Kempton R (1986) Statistical analysis of field experiments using neighbouring plots. Biometrics 42, 231-251.

Börner A, Freytag U, Specht CE (2000) Ergebnisse zur Langzeitlagerung von Samen verschiedener Kulturpflanzen in der Genbank Gatersleben. S. 29-32. Bericht über die 51. Arbeitstagung der Vereinigung Österreichischer Pflanzenzüchter BAL Gumpenstein.

Boyle CR, Montgomery RD (1996) An application of the augmented randomized complete block design to poultry research. Poultry Science 75:601-607.

Brunner E, Langer F (1999) Nichtparametrische Analyse longitudinaler Daten. R. Oldenburg Verlag, München.

Bundessortenamt (2000) Richtlinien für die Durchführung von landwirtschaftlichen Wertprüfungen und Sortenversuchen. Landbuch Verlag, Hannover, Deutschland.

Burnham K P, Anderson DR (1998) Model selection and inference: A practical information-theoretic approach. Springer-Verlag, New York, New York, USA.

Calhoun DS. 1997. An evaluation of modified augmented designs for single-replication yield testing in cotton. p.460-464. In: Dugger P, Richter D (ed.) Proceedings Beltwide Cotton Conferences, New Orleans, LA, USA.

Campbell CL, Madden LV (1990) Introduction to plant disease epidemiology. Wiley, New York.

Casler MD, Fales SL, McElroy AR, Hall MH, Hoffman LD, Undersander DJ, Leath KT (2002) Half-sib family selection for forage yield in orchardgrass. Plant Breeding 121:43-48.

Châtel M, Lorieux M, Martinez C, Borrero J, de Carvalho Neves P, Peixoto de Morais O (2004) Assemble and document information on selected entries of CIAT, CIRAD, IRD and EMBRAPA rice working collections. In: Improved rice germplasm for Latin America and the Caribbean Project, 2004 Annual Report. CIAT. p. 13-14. http://www.ciat.cgiar.org/riceweb/pdfs/rice_2004_output_1.pdf ... Last visit: 04/10/2006

Cochran WG, Cox DR (1957) Experimental designs. 2nd ed. Wiley, New York.

Consultative Group on International Agricultural Research (CGIAR) (1999) CGIAR statements on genetic resources, intellectual property rights and biotechnology.

Cook WJ, Cunningham WH, Pulleyblank WR, Schrijver A (1997) Combinatorial optimization. Wiley, New York.

Cullis BR, Lill WJ, Fischer JA, Read BJ (1989) A new procedure for the analysis of early generation variety trials. Applied Statistics 38:361-75.

Cullis BR, Smith AB, Coombes NE (2006) On the design of early generation variety trials with correlated data. Journal of Agricultural, Biological, and Environmental Statistics 11:1-13.

Digby PGN, Kempton RA (1987) Multivariate analysis of ecological communities. Chapman & Hall, London.

Duveiller E (1994) A pictorial series of disease assessment keys for bacterial leaf streak of cereals. Plant Disease 78:137-141.

Eccleston J, Chan B (1998) Design algorithms for correlated data. p. 41-52. In: Payne R, Green P (eds) (1998) COMPSTAT - Proceedings in Computational Statistics. 13th Symposium held in Bristol, Great Britain.

Edmondson RN (2005) Centenary Review: Past developments and future opportunities in the design and analysis of crop experiments. Journal of Agricultural Science 143:27-33.

Falconer DS, Mackay TFC (2001) Introduction to quantitative genetics. 4th ed. Pearson Education, Harlow.

FAO/IPGRI (1994) Genebank standards. Food and agriculture organisation of the United Nation, Rome, International Plant Genetic Resources Institut, Rome 13 p.

Federer WT (1956) Augmented (or hoonuiaku) designs. Hawaii. Plant Research 2: 191–208.

Federer WT (1961) Augmented designs with one-way elimination of heterogeneity. Biometrics 17:447-473.

Federer WT (2002) Construction and analysis of an augmented lattice square designs. Biometrical Journal 44:251-257.

Federer WT, Nair RC, Raghavarao D (1975) Some augmented row-column designs. Biometrics 31:361–373.

Federer WT, Raghavarao D (1975) On augmented designs. Biometrics 31:29–35.

Federer WT, Reynolds M, Crossa J (2001) Combining results from augmented designs over

sites. Agronomy Journal 93:389-395.

Federer WT, Wolfinger RD (1998) SAS code for recovering intereffect information in experiments with incomplete block and lattice rectangle designs. Agronomy Journal 90:545-551

Food and Agriculture Organization of the United Nations (FAO) (1996) Report on the state of the world's plant genetic resources for food and agriculture. FAO, Rome, Italy.

Forbes GA, Jeger MJ (1987) Factors affecting the estimation of disease intensity in simulated plant structures. Journal of Plant Diseases and Protection 94:113-120.

Forbes GA, Korva JT (1994) The effect of using a Horsfall-Barratt scale on precision and accuracy of visual estimation of potato late blight severity in the field. Plant Pathology 43:675-682.

Franco J, Crossa J, Taba S, Shands H (2003) A multivariate method for classifying cultivars and studying group x environment x trait interaction. Crop Science 43:1249-1258.

Garett KA, Madden LV, Hughes G, Pfender WF (2002) New applications of statistical tools in Plant Pathology. Phytopathology 94:999-1003.

Gilmour AR, Cullis BR, Verbyla AP (1997) Accounting for natural and extraneous variation in the analysis of field experiments Journal of Agricultural, Biological, and Environmental Statistics 2:269-293.

Gringarten E, Deutsch CV (2001) Variogram interpretation and modelling. Mathematical Geology 33:507-534.

Grondona MO, Cressie N (1991) Using spatial considerations in the analysis of Experiments. Technometrics 33:381-392.

Guarino L, Ramanatha Rao V, Reid R (eds) (1995) Collecting plant genetic diversity. Technical Guidelines. CAB International on behalf of IPGRI in association with FAO/IUCN/UNEP, Wallingford, UK.

Hammer K (1975) Die Variabilität einiger Komponenten Allogamieneigung bei Kulturgerste (Hodeum vulgare L. s.l.). Kulturpflanze 23:167-180.

Hartung K, Piepho HP (2006) Optimizing an augmented design using geostatistical methods. submitted to Canadian Journal Plant Science.

Hartung K, Piepho HP (2005) A threshold model for multi-year genebank data based on different rating scales. Crop Science 45:1045-1051.

Hartung K, Piepho HP, Knüpffer H (2006) Analysis of gene bank evaluation data using geostatistical methods. Genetic Resources and Crop Evolution 53:737-751.

Hau B, Kranz J (1989) Fehler beim Schätzen von Befallsstärken bei Pflanzenkrankheiten. Journal of Plant Diseases and Protection 96:649-674.

Haussmann BIG, Parzies HK, Presterl T, Susic Z, Miedaner T (2004) Plant genetic resources in crop improvement (Review). Plant Genetic Resources: Characterization and utilization 2:3-21.

Hawtin CH, Cherfas J (2003) Plant genebanks: Food security. http://www.actionbioscience.org/biodiversity/hawtin_cherfas.html#Primer ..... Last visit: 04/10/2006

Hill RR Jr, Rosenberger JL (1985) Methods for combining data from germplasm evaluation trials. Crop Science 25:467-470.

Hock J, Kranz J, Renfro BL (1992) Test of standard diagrams for field use in assessing the tarspot disease complex of maize (Zea mays). Tropical Pest Management 38:314-318.

Horsfall JG, Barratt RW (1945). An improved grading system for measuring plant diseases. Phytopathology 35:655.

IPGRI (1994) Descriptors for barley (*Hordeum vulgare* L.). International Plant Genetic Resources Institute, Rome, Italy.

Isaaks E, Srivastava M (1989) Applied geostatistics. Oxford University Press, Oxford.

James WC (1974) Assessment of plant diseases and loss. Annual Review of Phytopathology 12:27-48.

Jenkins SF Jr, Wehner TC (1983) A system for the measurement of folia diseases of cucumber. Cucurbit Genetics Cooperative Report 6:10-12.

Kempton RA, Gleeson AC (1997) Unreplicated trials. p. 86-100. In: Kempton RA, Fox PN (eds) Statistical methods for plant variety evaluation. Chapman and Hall, London.

Kempton RA, Gleeson AC (1997) Unreplicated trials. p. 86-100. In: Kempton RA, Fox PN (eds) Statistical methods for plant variety evaluation. Chapman & Hall, London.

Kendziorski C, Wang P (2006) A Review of Statistical Methods for Expression Quantitative Trait Loci Mapping. Mammalian Genome 17:509-517

Knezović Z, Gunjača J, Šatović Z, Kolak I (2005) Comparison of different methods for classification of gene bank accessions. Agriculturae Conspectus Scientificus 70:87-91.

Knüpffer H (2001) Handling of characterization and evaluation data in crop databases. In: Maggioni L, Spellman O (eds), Report of a Network Coordinating Group on Cereals. Ad hoc meeting, 7-8 July 2000, Radzików, Poland. International Plant Genetic Resources Institute, Rome, Italy, p. 58-65

Knüpffer H, Terentyeva I, Hammer K, Kovaleva O, Sato K (2003) Ecogeographical diversity – a Vavilovian approach. p. 53-76. In: von Bothmer R, van Hintum Th, Knüpffer H, Sato K (eds). Diversity in barley (*Hordeum vulgare*). Elsevier Science, Amsterdam, The Netherlands.

Knüpffer H, van Hintum Th (2003) Summarised diversity – the Barley Core Collection. p.259-267. In: von Bothmer R, van Hintum Th, Knüpffer H, Sato K (eds). Diversity in barley (*Hordeum vulgare*). Elsevier Science, Amsterdam, The Netherlands.

Koo B, Pardey PG, Wright BD, with Bramel P, Debouck D, van Dusen ME, Jackson MT, Rao NK, Skovmand B, Taba S, Valkoun J(2004) Saving seeds: The economics of conserving genetic resources ex-situ in the Future Harvest Centres of the CGIAR. Wallingford, UK: CABI Publishing.

Lesser W (1998) Sustainable use of genetic resources under the convention of biological diversity - Exploring access and benefit sharing issues. CAB International, Wallingford, UK.

Lin CS, Poushinsky G (1983) A modified augmented design for an early stage of plant selection involving a large number of test lines without replication. Biometrics 39:553-561.

Lin CS, Poushinsky G (1985) A modified augmented design (Type 2) for rectangular plots. Canadian Journal of Plant Science 65:743-749.

Lin CS, Poushinsky G, Jui PY (1983) Simulation study of three adjustment methods for augmented design and comparison with the balanced lattice square design. Journal of

Agricultural Science, Camb. 100:527-534.

Lin CS, Voldeng HD (1989) Efficiency of type 2 modified augmented design in soybean variety trials. Agronomy Journal 81:512–517.

Littell RC, Milliken GA, Stroup WW, Wolfinger RD (1996) SAS system for mixed models. SAS Institute Inc. Cary.

Little RJA, Rubin DB (2002) Statistical analysis with missing data. Wiley, New York.

Lynch M, Walsh B (1998) Genetics and analysis of quantitative traits. Sinauer, Sunderland.

Madden LV (2002) Evaluation of generalized mixed models of analyzing disease incidence data obtained in designed experiments. Plant Disease 86:316-325.

May KW, Kozub GC (1995) Success of a selection program for increasing grain yield of two-row barley lines and evaluation of the modified augmented design (Type 2). Canadian Journal of Plant Science 75:795-799.

May KW, Kozub GC, Schaalje GB (1989) Field evaluation of a modified augmented design (Type 2) for screening barley lines. Canadian Journal of Plant Science 69:9-15.

McBratney AB, Pringle MJ (1999) Estimating average and proportional variograms of soil properties and their potential use in precision agriculture. Precision Agriculture 1:219-236.

McCullagh P, Nelder J (1989) Generalized linear models. 2nd ed. Chapman and Hall, London.

McCulloch CE, Searle SR (2001) Generalized, linear, and mixed models. Wiley, New York.

Mead R (1997) Design of plant breeding trials. p. 40-67. In: Kempton RA, Fox PN (eds). Statistical methods for plant variety evaluation. Chapman & Hall, London.

Newton AC, Hackett CA (1994) Subjective components of mildew assessment on spring barley. European Journal of Plant Pathology 94:395-412.

Nita M, Ellis MA, Madden LV (2003) Reliability and accuracy of visual estimation of phomopsis leaf blight of strawberry. Phytopathology 93:995-1005.

Nutter FW, Gleason ML, Jenco JH, Christians NC (1993) Assessing the accuracy, intra-rater repeatability, and inter-rater reliability of disease assessment systems. Phytopathology 83:806-812.

Nutter FW, Schultz PM (1995) Improving the accuracy and precision of disease assessments: selection of methods and use of computer-aided training programs. Canadian Journal of Plant Pathology 17:74-184.

O'Brien RD, van Bruggen AHC (1992) Accuracy, precision, and correlation to yield loss of disease severity Scales for corky root of lettuce. Phytopathology 82:91-96.

Ortiz R, Mohamed SF, Madsen S, Weibull J, Christiansen JL (2001) Assessment of Phenotypic Variation in Winter Barley in Scandinavia. Acta Agriculturæ Scandinavica (Section B Soil and Plant Sciences) 51:151-159

Ortiz R. 2002. No just seed repositories: a more pro-active role for gene banks. http://www.geneconserve.pro.br/artigo_6.htm .....Last visit: 04/10/2006

Padilla G, Cartea ME, Rodriguez VM, Ord A (2005) Genetic diversity in a germplasm collection of Brassica rapa subsp. rapa L. from northwestern Spain. Euphytica 145:171-180.

Papadakis JS (1937) Méthode statistique pour les expériences en champ. Bull. Scientifique de l'Institut d'Amélioration des Plantes de Thessalonique. 23.

Patterson HD (1997) Analysis of series of variety trails. p. 139-161. In: Kempton RA, Fox PN (eds) Statistical methods for plant variety evaluation. Chapman & Hall, London.

Patterson HD, Hunter EA (1983) The efficiency of incomplete block designs in National list and recommended list cereal variety trials. Journal of Agricultural Science 101:427-433.

Patterson HD, Thompson R (1971) Recovery of inter-block information when block sizes are unequal. Biometrika 58(3):545-554.

Peterson RG. (1985) Design and analysis of experiments. Marcel Dekker Inc., New York.

Peterson, R.G. (1994) Agricultural field experiments – design and analysis. Marcel Dekker Inc., New York.

Piepho HP (1994) Best linear unbiased prediction (BLUP) for regional yield trials: A comparison to additive main effects multiplicative interaction (AMMI) analysis. Theoretical and Applied Genetics 89:647-654.

Piepho HP (1998) Empirical best linear unbiased prediction in cultivar trials using factor analytic variance-covariance structures. Theoretical and Applied Genetics 97:195-201.

Piepho HP (1999) Analysing disease incidence data from designed experiments by generalized nonlinear mixed-effect models. Plant Pathology 48:668-674.

Piepho HP (2002) Auswertung von Bonituren mit der SAS Prozedur NLMIXED. Zeitschrift für Agrarinformatik 10:30-41.

Piepho HP (2003) Folded exponential transformation. The Statistician 52:575-589.

Piepho HP (2003a) Model-based mean adjustment in germplasm evaluation data. Genetic Resources and Crop Evolution 50:281-290.

Piepho HP (2003b) The folded exponential transformation for proportions. The Statistician 52:575-589.

Piepho HP, Büchse A, Emrich K (2003) A hitchhiker's guide to the mixed model analysis of randomized experiments. Journal of Agronomy and Crop Science 189:310-322.

Piepho HP, Kalka E (2003) Threshold models with fixed and random effects for ordered categorical data. Food Quality and Preference 14:343-357.

Piepho HP, Möhring J, Büchse A (2006) BLUP for phenotypic selection in plant breeding and variety testing. Submitted to Euphytica.

Piepho HP, Pillen K (2004) Mixed modelling for QTL x environment interaction analysis. Euphytica 137:147-153.

Pilarczyk W, Tomaszyk K (2006) Search for the best spatial model in variety testing trials on wheat. Colloquium Biometryczne, 36:41-50.

Pinheiro JC, Bates DM (1995) Approximations to the log-likelihood function in the nonlinear mixed effects model. Journal of Computational and Graphical Statistics 4:12-35.

Priestley DA (1986) Seed aging. Cornell University Press, Ithaca.

Pringle MJ (2002) Field-scale experiments and site-specific crop management. PhD Thesis. The University of Sydney.

Reif JC, Melchinger AE, Frisch M (2005) Genetical and mathematical properties of similarity and dissimilarity coefficients applied in plant breeding and seed bank management. Crop Science 45:1-7.

Rice JA (1995) Mathematical statistics and data analysis. Duxbury Press 2nd ed. Wadsworth

Publishing Co Inc., Duxbury Press, Belmond, California.

Robinson GK (1991) That BLUP is a good thing: The estimation of random effects. Statistical Science 6:15-51.

Rojas W, Barriga P, Figueroa H (2000) Multivariate analysis of the genetic diversity of Bolivian quinua germplasm. PGR-Newsletter 122:16-23.

Samuels ML, Casella G, McCabe GP (1991) Interpreting blocks and random factors (with discussion). Journal of the Statistical Association 86:798-821.

Santos AH, Bearzoti E, Ferreira DF, da Silva Filho JL (2002) Simulation of mixed models in augmented block design. Science agricola (Piracicaba, Braz.). 59:483-489. http://www.scielo.br/pdf/sa/v59n3/10580.pdf .....Last visit: 04/10/2006

Santos AH, Bearzoti E, Ferreira FD (2002) Simulation of mixed models in augmented design. Scientia Agricola 59:483-489.

SAS Institute Inc. (1997) SAS/STAT Software: Changes and enhancements through release 6.12. SAS Institute Inc., Cary.

SAS Institute Inc. (1999) SAS/STAT User's Guide, Version 8. SAS Institute, Cary.

Schaalje GB, Lynch DR, Kozub GC (1987) Field evaluation of a modified augmented design for early stage selection involving large number of test lines without replications. Potato Research (Zeitschrift d. Europaeischen Gesellschaft für Kartoffelforschung) 30:35-45.

Schabenberger O, Gotway CA (2005) Statistical methods for spatial data analysis. Chapman & Hall, Boca Raton.

Schena M (2003) Microarray Analysis. 1ed. Wiley, Hoboken.

Schabenberger O, Pierce FJ (2002) Contemporary statistical models for the plant and soil sciences. CRC Press, Boca Raton.

Schumacher E, Bleiholder H, Thöni H (1995) Methodische Untersuchungen zur biometrischen Analyse von Boniturwerten aus Freilandversuchen mit Herbiziden. 9th EWRS Symposium, Budapest 1995, Proceding Vol. 1:283-290.

Schumacher E, Thöni H (1990) Auswertung von Boniturwerten. Agrarinformatik 18:51-62.

Schwarzbach E (1984) A new approach in the evaluation of field trials: the determination of the most likely genetic ranking of varieties. Vorträge für Pflanzenzüchtung. 6:249-259

Scott RA, Milliken GA (1993) A SAS program for analysing augmented randomized complete block design. Crop Science 33:865-867

Searle SR (1987) Linear models for unbalanced data. Wiley, New York.

Searle SR, Casella G, McCulloch CE (1992) Variance components. Wiley, New York.

Shah DA, Madden LV (2004) Nonparametric analysis of ordinal data in designed factorial experiments. Phytopathology 94:33-43.

Sharma KC, Sharma RK, Singhania DL, Singh D (2003) Variation and character association in fodder yield and related traits in pearl millet (Pennisetum glaucum (L.) R. Br.). Indian Journal of Genetics and Plant Breeding 63:115-118.

Sherwood RT, Berg CC, Hoover MR, Zeiders KE (1983) Illusions in visual assessments of Stangonospora Leaf Spot of orchardgrass. Phytopathology 73:173-177.

Smith AB, Cullis BR, Thompson R (2005) The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. Crop Science 46:1722-1733.

Smith AB, Cullis BR, Thompson R (2005) The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. Journal of Agricultural Science 143:449-462.

Snijders CHA (2003) Field evaluation of type 2 modified augmented designs for non-replicated yield trials in the early stages of a wheat breeding program. Bericht über die Arbeitstagung 2002 der Vereinigung der Pflanzenzüchter und Saatgutkaufleute Osterreichs gehalten in Gumpenstein. p. 59-64.

Solorzano V, Gilmour SG, Phelps K, Kennedy R (1997) Assessment of suitable designs for field experiments involving airborne diseases. Journal of Agricultural Science 129:249-256.

Steiner AM, Ruckenbauer P (1995) Germination of 110-year-old cereal and weed seeds, the Viennan samples of 1877. Verification of effective ultra-dry storage at ambient temperature. Seed Science Research 5:195-199.

Stroup WW (2002) Power analysis based on spatial effects mixed models: a tool for comparing design and analysis strategies in the presence of spatial variability. Journal of Agricultural, Biological, and Environmental Statistics 7:491-511.

Stroup WW, Baenzinger PS, Mulitze DK (1994) Removing spatial variation from yield trials: A comparison of methods. Crop Science 86:62-66.

Stroup WW, Mulitze DK (1991) Nearest neighbour adjusted best linear unbiased prediction. The American Statistician 45:194-200.

Talbot M (1997) Resource allocation for selection systems. p.162-174. In: Kempton RA, Fox PN (eds). Statistical methods for plant variety evaluation. Chapman & Hall, London.

Thöni H (1985) Auswertung von Bonituren: Ein empirischer Methodenvergleich. EDV in Medizin und Biologie 16:108-114.

Tomerlin JR, Howell TA (1988) Distrain: A computer program for training people to estimate diseases severity on cereal leaves. Plant Disease 72:455-459.

van Hintum T, Menting F (2003) Diversity in ex situ genebank collections of barley. p. 247-257. In: von Bothmer R, Van Hintum T, Knüpffer H, Sato K (eds). Diversity in barley (*Hordeum vulgare*). Elsevier, Amsterdam, The Netherlands.

Vollmann J, Buerstmayer H, Ruckenbauer P (1996) Efficient control of spatial variation in yield trials using neighbour plot residuals. Experimental Agriculture 32:185-197.

Vonesh EF, Chinchilli VM (1997) Linear and nonlinear models for the analysis of repeated measurements. Marcel Dekker, New York.

Wang ML, Dean R, Erpelding JE, Pederson GA (2006) Molecular genetic evaluation of sorghum germplasm differing in response to fungal diseases: rust (puccinia purpurea) and anthracnose (collectotrichum graminicola). Euphytica 148:319-330

Watson S (2000) Spatial dependence and block designs in spaced plant herbage trials. Journal of Agricultural Science 134:245-258.

Williams ER (1986) A neighbour model for field experiments. Biometrika 73:279-287

Williams ER, John JA (2003) A note on the design of unreplicated trials. Biometrical Journal 45:751-757.

Winer BJ, Brown DR, Michels KM (1991) Statistical principles in experimental design. 3^{ed} ed. McGraw Hill, Boston.

Wolfinger RD, Federer DW, Cordero-Brana O (1997) Recovering information in augmented

designs, using SAS Proc GLM and Proc MIXED. Agronomy Journal 89:856-59.

Wu T, Dutilleul P (1999) Validity and efficiency of neighbour analyses in comparison with classical complete and incomplete block analyses of field experiments. Agronomy Journal. 91:721-731.

Wu T, Mather DE, Dutilleul P (1998) Application of geostatistical and neighbour analyses to data from plant breeding trials. Crop Science 38:1545-1553.

Yan WK, Rajcan I (2003) Prediction of cultivar performance based on single- versus multiple-year test in soybean. Crop Science 43:549-555.

Yates F (1936) A new method of arranging variety trials involving a large number of varieties. Journal of Agricultural Science 26:424-455.

Yu J, Pressoir G, Briggs W, Vroh B, Yamasaki M, Doebley JF, Mcmullen MD, Gaut BS, Holland JB, Kresovich S, Buckler ES (2006) A unified mixed-model method for association mapping accounting for multiple levels of relatedness. Nature Genetics 38:203-208.

6 Summary

A main interest in plant breeding today is to analyse the genome at the molekular level. Microarray and QTL analyses become more and more important. The aim of these genome analyses is to predict the phenotypic characteristics of a trait using genomic data. These analyses are based on phenotypic data, so therefore an imprecise assessment of the phenotypic data results in imprecise genomic analyses. Hence, to obtain proper genomic information, highly precise assessment of field data is as important as accurate work at the genetic level. The aim of the present thesis was to explore methods to statistically analyse phenotypic data of gene banks for cultivated plants. Selected traits of the barley data (*Hordeum* spec.) of the gene bank of the Leibniz-Institut für Pflanzengenetik und Kulturpflanzenforschung (IPK) were evaluated. The data of years 1948 to 1991 and 1993 to 2002 were available. Within this period the ordinal scale changed from a 0 to 5 scale between the years 1991 and 1993 to a 1 to 9 scale after 1993.

At most gene banks, evaluation during seed reproduction of accessions is currently done without any experimental design. With the data of a single year there are only few replications of a single check for winter and summer barley and only rarely do accessions have replications. The data of 2002 were analysed separately for winter and summer barley using geostatistical methods. For the traits analysed (plant density before winter, plant density after winter, waxy cover of culm and leaves, lodging, plant length, growth habit, mildew, brown rust) four different types of variogram model (linear, spherical, exponential and Gaussian) were fitted to the empirical variogram using non-linear regression. The spatial parameters obtained by non-linear regression for every variogram model then were implemented in a mixed model analysis and the four model fits compared using Akaike's Information Criterion (AIC). Best linear unbiased predictors (BLUP) of accessions were generated for the best model according to AIC. The approach to estimate the genetical parameter by Kriging can not be recommended, since this approach often estimated a genetic variance of zero. The BLUPs obtained by the first approach and the original data were investigated using Spearman’s rank correlation. The values were between 0.79 and 0.96. The heritability was between 0.008 and 0.95. The most common well-fitting geostatistical models were the spherical and the exponential model. Usually the choice of a covariance model was not critical, as the considered models differed only slightly in most cases. It is normally impossible to properly model the whole variogram. The first points of the empirical variogram should be explained well by the fitted theoretical variogram, as these represent most of the pairwise distances between plots and are most crucial for neighbour adjustments. A nugget effect (residual

error) was needed for nearly all traits. The small number of check plots for the available data made it difficult to accurately dissect the genetical effect from environmental effects.

A method to analyse multi-year data assessed on an ordinal scale is the threshold model investigated in this thesis. Here the complete data set from 1948 to 2002 was analysed. The threshold model allows for joint analysis of data from different rating scales, assuming a common latent scale for the different rating systems. With the barley data of the IPK empirical Bayes estimates based on the threshold model yielded a rank order of accessions which was quite similar to the ranking by BLUPs based on observed scores (r_s = 0.985), though there were a considerable number of rank changes. This suggests that a mixed model analysis which treats ordinal scores as metric data will yield meaningful results, but that the gain in efficiency is higher when using a threshold model. The threshold model may also be used when there is a metric scale, such as a percentage scale, underlying the observed ratings. The Laplace approximation (a single quadrature point) as a numerical method to integrate the log-likelihood for random effects worked well, but it is recommended to increase the number of quadrature points until the change in parameter estimates becomes negligible. For more complex data structures, e.g. due to incomplete blocking or specific pedigree structure, further research is necessary.

Since ratings are important in the gene bank context, their statistical performance was analysed. Three rating methods (1%-steps, 5%- steps, 9-point rating) were assessed by untrained persons (Group A) and persons experienced in rating (Group B). Every person had to rate several computer-generated pictograms of diseased grain leaves. The estimates of Group B were always closer to the real disease value than those of Group A. The highest accuracy was found with Group B using the 1%-scale and with Group A using the 5%-scale. With a percentage scale untrained raters tended to use values that are multiples of 5% or 10%. For the time needed per leaf assessment the trained group was fastest when using the 5% rating scale. The raters mostly preferred 9-point rating, especially when untrained. From a statistical point of view both percent ratings performed better than the ordinal rating scale. Generally, raters felt uncomfortable with uncommon thresholds for 9-point ratings as given with log-divided scales. The closer the scale of collected data is to a ratio scale with normal distribution, the more powerful methods are available for analysis. And if interpretation is to be possible even after years or over locations, percentages are unique and therefore more informative especially if statistical analysis is required, while the definition of ordinal rating scales might change over time and be forgotten later. For the rater it might be inconvenient to decide on a definite number, but statistically the

possible error made by the rater is calculable and usually smaller than with ratings by rougher methods. So directly rating percentages whenever possible leads to smaller overall estimation errors, and with proper training, e.g. using an appropriate computer program, accuracy and precision can be further improved.

One way to make gene bank data more reliable is to use a proper experimental field design. The augmented designs as proposed by Federer in 1956 offer themselves. Over the past 50 years a lot of additional research results have been published, e.g. on the construction and analysis of augmented lattice square designs or the so-called alpha-alpha-designs. Another widely used method is the modified augmented design, which was introduced by Lin and his co-workers. The augmented designs as proposed by Federer and those proposed by Lin have advantages and disadvantages. The augmented designs proposed by Federer have the advantage of an unbiased error estimate. But the random allocation of checks is a problem, as the distribution pattern can become very patchy over the experimental site and therefore the assessment of environmental estimation can become unreliable. The augmented design by Lin et al. always places checks in the centre plot of every, if possible square-shaped, whole plot. The uncertainty with their augmented designs lies with the analysis, for there is no obvious best model to estimate the genetic effect of an entry. So none of the methods of analysis proposed by Lin and co-workers is based on an explicit statistical model, and therefore there is no well-founded decision criterion to select between them. Computational analysis is now available for all designs as is the combination of results from augmented designs over sites.

Spatial analysis can increase the efficiency, e.g. compared to a standard complete block analysis, when within-block homogeneity is low, e.g. when block sizes are large. In this thesis spatial analysis is also used to find an optimal field layout for an augmented design, i.e. a layout that yields small least significant differences (LSD). The average variance of a difference (a.v.d.) and the average squared LSD are used to compare competing designs, using a theoretical approach based on variations of two anisotropic models (nearly isotropic, strongly anisotropic) and different rotations of anisotropy axes towards field reference axes. The a.v.d. is mainly influenced by the number of blocks, plots and checks. Based on theoretical calculations, up to five checks per block are recommended. The nearly isotropic combinations lead to designs with large quadratic blocks. The strongly anisotropic combinations display effects resulting from the combination of degree of anisotropy and rotation of anisotropy axes. Without rotation small elongated blocks are preferred. The closer the rotation is to 45° the more square blocks and the

more checks are appropriate.

Augmented designs represent a cost-effective, efficient and well-manageable tool to separate genetic and environmental effects. But so far they have not found its way into standard regeneration of accessions. In a gene bank context it might be interesting to replace some check varieties by accessions that are in higher demand. The possibility to link data of diverse environments might be of special interest. But a general requirement then is that all data need to be connected, so it is important to have connecting checks and accessions in all environments. Powerful computers and software necessary for analysis are fully developed and available and it is likely that over time augmented designs and associated analysis software will be adapted to the gene banks' needs. Computer programs for gene banks should include methods to select additional accessions that insure the connection of experiments over time in one place and within one year over sites, i.e. coordination between gene banks, if desired. Further research is necessary regarding the optimal number of different checks and replications per check.

Especially in times of decreasing financial resources it is worthwhile to invest money to gain the largest benefit possible from evaluation data routinely collected by gene banks. Combining geostatistical methods and classical field experimental designs, like augmented designs, seems to be a promising route in gene banks, for accession material is diverse and many influences, e.g. pests, are not under the researcher's control. Regardless which method is chosen, geostatistical methods, classical field experimental designs or a combination of both, it is essential to adopt one of these methods. The geostatisical approach of first differences (Besag and Kempton, 1986; Williams, 1986; Wu and Dutilleul, 1999), which was not investigated in this thesis, should be explored in context with gene bank data, since it is an alternative method to account for spatial variance.

The results presented in this thesis may be summarised as follows:
To achieve phenotypic trait data of accessions that are satisfying for statistical analysis, it is necessary that the cultivation for regeneration of accessions is based on a meaningful and statistically analysable experimental field design. The design needs to include checks and a random sample of accessions from the gene pool held at the gene bank. Furthermore it is advisable to utilise metric or percentage rating scales. It can be expected that using a threshold model with ordinal ratings as well as applying metric or percentage rating scales wherever possible increases the quality of multivariate analysis and association mapping studies based on phenotypic gene bank data.

7 Zusammenfassung

Großes Interesse im Bereich der Pflanzenzüchtung liegt heute auf der Genomanalyse. Microarry- und QTL-Analyse spielen dabei eine immer größere Rolle. Dabei ist das Ziel vieler Genomanalysen, die phänotypische Ausprägung eines Merkmale mittels genomischer Daten vorherzusagen. Diese Analysen basieren auf phänotypischen Daten, so dass eine ungenaue Erhebung der phänotypischen Daten zu ungenauen Ergebnissen bei der Genomanalyse führt. Daher ist für die Ermittlung genomischer Informationen eine so präzise wie mögliche Erhebung der Felddaten genauso wichtig, wie akkurate Laborarbeit. Die vorliegende Arbeit hatte zur Aufgabe, sich mit der statistischen Auswertbarkeit von phänotypischen Kulturpflanzen-Genbankdaten zu beschäftigen, mit dem Ziel, den genetischen Effekt möglichst genau zu schätzen. Exemplarisch wurden hierzu die Boniturdaten verschiedener Merkmale von Gerste (*Hordeum* spec.) der Genbank des Leibniz-Instituts für Pflanzengenetik und Kulturpflanzenforschung (IPK) ausgewertet. Zur Verfügung standen die Daten der Jahre 1948-1991 und 1993-2002. Innerhalb dieses Zeitraumes war die Skalierung der Ordinal-Skalen zwischen den Jahren 1991 und 1993 von 0-5 auf 1-9 Intervalle umgestellt worden.

Dem Erhaltungsanbau lag kein Versuchsdesign zu Grunde. Die Daten je eines Jahres hatten nur wenige Wiederholungen je eines einzigen Standards (eine Akzession) innerhalb der Winter- bzw. Sommergerste, von anderen Akzessionen gab es nur ganz vereinzelt Wiederholungen. Daher wurde der Datensatz des Jahres 2002 getrennt für Sommer- und Wintergersten mit geostatistischen Verfahren ausgewertet. An jedes Merkmal (Bereifung, Pflanzenlänge, Mehltau und Zwergrost (Krankheitsbonituren), Lagerneigung vor der Reife, Wuchsform sowie Bestand vor und nach Winter) wurden vier Variogramm-Modelle (linear, sphärisch, exponentiell und Gauß) mittels nichtlinearer Regression an das jeweilige empirische Variogramm angepasst. Die so gewonnenen geostatistischen Parameter wurden in ein Gemischtes Modell integriert und danach die 4 Modelle anhand des Akaikeschen Informationskriterium (AIC) verglichen. Für das nach AIC beste Modell wurden die Besten Linearen Unverzerrten Prädiktoren (BLUP) der Akzessionenseffekte geschätzt. Mit dem Ansatz, die genetischen Effekte mittels Kriging zu schätzen, konnte jedoch häufig kein genetischer Effekt gefunden werden. Daher kann er nicht empfohlen werden. Die mit dem ersten Ansatz erhaltenen BLUPs wurden mit den ursprünglichen Boniturdaten mittels Spearmanscher Rangkorrelation verglichen. Die Werte lagen bei mindesten 0,79. Die Heritabilität lag zwischen 0,008 und 0,95. Als optimale geostatistische Modelle erwiesen sich das Sphärische und das Exponentielle. Grundsätzlich war die Wahl des Kovarianzmodells

nicht kritisch, denn die Modelle unterschieden sich meist nur geringfügig. Ein Nugget-Effekt wurde häufig gebraucht. Da das gesamte Variogramm selten zufriedenstellend beschrieben werden kann, ist es sinnvoll, den vorderen Bereich des Variogramms, der von besonderem Interesse ist, gut anzupassen. Wegen der geringen Zahl an Standards und Wiederholungen war es insgesamt jedoch schwer, den Nugget und damit den genetischen Effekt gut zu schätzen.

Ein Möglichkeit, mehrjährige ordinale Daten auszuwerten, hier der gesamte Datensatz von 1948-2002, ist das in dieser Arbeit untersuchte Schwellenwertmodell. Es erlaubt die gemeinsame Auswertung von Daten, die mit zwei verschiedenen Bonitursklalen erhoben wurden. Die bei der Anwendung dieser Methode auf die Gerstendaten des IPK erhaltenen empirischen Bayes Schätzer wiesen eine hohe Spearmansche Rankorrelation mit den BLUPs auf, die direkt aus den beobachteten Werten gewonnen wurden (r_s=0,985). Es ergab sich jedoch eine Vielzahl von unterschiedlichen Rängen. Dies deutet darauf hin, dass eine Analyse mit einem gemischten Modell zwar sinnvolle Ergebnisse liefert, das Schwellenwertmodell jedoch noch zu besseren Ergebnissen führt. Das Schwellenwertmodell kann auch verwendet werden, wenn den vergebenen Boniturnoten metrische Daten zu Grunde liegen, z.B. Prozentwerte. Die verwendete Laplace-Approximation (ein einzelner Quadraturpunkt) zur numerischen Integration der log-Likelihood über die zufälligen Effekte erwies sich hier als geeignete Methode. Die Anzahl der Quadraturpunkte sollte jedoch so lange erhöht werden, bis die Änderung der Parameter vernachlässigbar ist. Für komplexere Datenstrukturen, z.B. aufgrund unvollständiger Blöcke oder zusätzlicher Verwandschaftsverhältnisse zwischen den Genotypen, sind weitere Untersuchungen nötig.

Da Bonituren im Genbank-Kontext eine wichtige Rolle spielen, wurden ihre statistischen Eigenschaften untersucht. Geübte und ungeübte Boniteure wandten drei unterschiedliche Bonitursklalen (1%-, 5%-Schritte, 9er Bonitur) auf computergenerierte Bilder von Getreideblätter mit Mehltaubefall an. Die genauesten Schätzungen gelangen den Geübten mit der 1% Skala und den Ungeübten mit der 5% Skala. Bei der 1% Skalierung neigten die Ungeübten dazu, Vielfache von 5 häufiger als andere Werte zu vergeben. Bezüglich der zum Bonitieren benötigten Zeit war die trainierte Gruppe eindeutig mit der 5% Bonitur am schnellsten. Die meisten Boniteure, besonders die Untrainierten, bevorzugten die 9er Bonitur, fanden die logarithmische Unterteilung jedoch unangenehm. Aus statistischer Sicht sind die beiden Prozentbonituren angemessener, da metrische Skalen den Vorteil haben, auch nach Jahren oder über verschiedene Versuchstandorte hinweg, eindeutig und identisch zu sein und mehr Information zu beinhalten. Daher sind sie, besonders wenn statistische Auswertung der Daten gewünscht ist, den Boniturnoten vorzuziehen,

zumal die Skalierung von Boniturnoten sich über die Jahre hinweg verändern oder verloren gehen kann. Dem Boniteur mag es unangenehm sein, sich auf eine bestimmte Prozentzahl festzulegen, von statistischer Seite ist es jedoch günstiger, da der Fehler des Boniteurs berechenbar und in der Regel kleiner ist, als der, der bei groberer Unterteilung einer Ordinalskala entsteht. Daher führt das Bonitieren mit Prozentskalen zu geringeren Schätzfehlern. Und Schätzübungen mittels geeigneter Computerprogramme erhöhen die Genauigkeit und Präzision.

Eine Möglichkeit, Genbankdaten aussagefähiger zu machen, ist, beim Erhaltungsanbau einem Versuchsdesign zu folgen. Dazu bieten sich Augmented Designs an, die 1956 von Federer vorgeschlagen und seit dem weiterentwickelt wurden, u.a. durch Methoden zur Konstruktion und Analyse auf der Basis eines Gitterquadrates und das so genannte Alpha-Alpha-Design. Eine ebenfalls weit verbreitete Methode ist das Modified Augmented Design, das von Lin und Kollegen vorgestellt und weiterentwickelt wurde. Die Augmented Designs, wie sie von Federer und Lin vorgeschlagen wurden, haben Vor- und Nachteile. Federers Designs schätzen den Fehler unverzerrt. Die zufällige Verteilung der Standards kann jedoch im Einzelfall sehr unregelmäßig und dadurch die Schätzung des Umwelteffekts unzuverlässig werden. Das Modified Augmented Design vermeidet dies durch Platzierung der Standards in die Mitte der möglichst quadratischen Großparzelle (whole plot). Da das Modified Augmented Designs nicht auf einer expliziten statistischen Methode beruht, gibt es jedoch kein offensichtlich bestes Modell zur Schätzung der Effekte. Computeranalyse von Augmented Designs ist möglich, ebenso wie die gemeinsame Auswertung verschiedener Umwelten.

Die Verwendung räumlicher Statistik kann, verglichen mit einer einfachen vollständigen Blockanlage, die Effizienz erhöhen, wenn die Homogenität innerhalb eines Blocks gering ist, z.B. bei großen Blöcken. In dieser Arbeit wurden geostatistische Methoden genutzt, um Augmented Designs zu optimieren, d.h. eine möglichst kleine Grenzdifferenz (LSD) zu erhalten. Die durchschnittliche Varianz einer Differenz (a.v.d.) und die durchschnittliche quadrierte LSD wurden zum Designvergleich genutzt. Hierzu wurde ein theoretischer Ansatz gewählt, der auf zwei anisotropen Modellen (fast isotrop, stark anisotrop) und verschiedenen Rotationen der Anisotropie-Achse zur Hauptachse des Feldversuchs beruhte. Die a.v.d. wird hauptsächlich von der Zahl der Blöcke, Parzellen und Standards beeinflusst. Bis zu fünf Standards je Block scheinen empfehlenswert. Liegt nahezu Isotropie vor, sind große quadratische Blöcke empfehlenswert. Bei Anisotropie ist die Blockform von der Intensität der Anisotropie und der Rotation der Achsen zueinander abhängig. Ohne Rotation sind schmale lange Blocks empfehlenswert. Je näher die

Rotation bei 45° liegt, um so quadratischer sollte der Block sein und umso mehr Standards sollten Verwendung finden.

Augmented Designs stellen für Genbanken vom Standpunkt der Kosten und des Arbeitsaufwands ein gut handhabbares Werkzeug dar, das seinen Weg jedoch noch nicht in den Vermehrungsanbau gefunden hat. Für Genbanken interessant ist, dass Standards auch durch Akzessionen ersetzt werden können, für die z.B. eine größere Nachfrage besteht. Auch die Möglichkeit, verschiedene Umwelten zusammen auswerten zu können, ist im Genbankkontext interessant. Hierzu müssen die Datensätze jedoch verbunden sein (connected), wobei es sinnvoll ist, so viele gleiche Standards und Akzessionen wie möglich in mehreren, der zu vergleichenden Umwelten, zu haben. Die für solche Auswertungen notwendige Soft- und Hardware steht zur Verfügung und es ist anzunehmen, dass in Zukunft spezielle Design- und Analyseprogramme für Genbanken entwickelt werden. Diese sollten die Akzessionen so auswählen können, dass die Datensätze über verschiedene Orte und Jahre verbunden sind, so dass eine gemeinsame Auswertung in Kooperation zwischen Genbanken möglich ist. Weitere Forschung hinsichtlich der Schätzung der optimalen Anzahl verschiedener Standards im Verhältnis zu deren Wiederholungen ist nötig.

Gerade in finanziell schwierigen Zeiten ist es sinnvoll, soviel Informationen wie möglich aus Routineerhebungen der Genbanken zu ziehen. Die Kombination geostatistischer Methoden und des klassischen Felddesigns, z.B. Augmented Design, scheint ein viel versprechender Weg für Genbanken, da das genetische Material divers und viele Einflüsse, wie z.B. Krankheiten, nicht vorhersagbar sind. Es ist weniger ausschlaggebend, ob geostatistische Methoden, klassisches Felddesign oder eine Kombination von beiden gewählt wird, aber es ist sinnvoll eine dieser Optionen zu wählen.

Der geostatistische Ansatz der Ersten Differenzen (first differences), der in dieser Arbeit nicht berücksichtigt wurde, sollte im Zusammenhang mit Genbankdaten untersucht werden, da er eine weitere Möglichkeit darstellt räumliche Varianz zu modellieren.

Zusammenfassend kann gesagt werden:
Um phänotypische Merkmalsdaten von Akzessionen zu erhalten, die für statistische Auswertung geeignet sind, ist es nötig, dass der Erhaltungsanbau auf einem sinnvollen und statistisch auswertbaren Versuchsdesign beruht, dass wiederholte Standards und dass eine zufällige Auswahl der angebauten Akzessionen aus der Gesamtheit garantiert ist. Des Weiteren ist es sinnvoll metrische oder Prozentbonitурskalen zu verwenden.

Es ist davon auszugehen, dass die Anwendung des Schwellenwertmodells bei Boniturnoten sowie die Verwendung von metrischen oder Prozentskalen die Qualität multivariater Auswertungen sowie Assoziationsstudien mit phänotypischen Genbankdaten verbessern.

8 Acknowledgements

First of all I would like to thank my academic supervisor Prof. Dr. Hans-Peter Piepho for his confidence and constant support. He offered me the opportunity to do a PhD on a topic that was developed in cooperation with Prof. Dr. Karl Hammer.

My thank also goes to Prof. Dr. Karl Hammer for his help and for arranging the contact with the IPK as well as his willingness to act as examiner 0n the examining board of this dissertation, as did Prof. Dr. Claus Zebitz.

Prof. Dr. Claus Zebitz also supported this thesis by offering the possibility to use his lectures to obtain rating data for the third paper.

I am very grateful to my colleagues at the Institute for Crop Production and Grassland Research, Section Bioinformatics, for their support, fruitful discussions and friendship, especially Dr. Katharina Emrich, Barbara Keller, Dr. Irina Kuzyakova and Jens Möhring.

Thanks are due to the staff at the IPK for making the barley data available. Special thanks go to Dr. Helmut Knüpffer and to Michael Grau for offering advice and knowledge to me whenever there where questions about barley or gene banks.

Furthermore I want to thank all persons from the gene banks world wide who replied to my e-mail questionnaire. Especially I like to thank Dr. Theo van Hintum at the Centre for Genetic Resources (CGN), The Netherlands, and his colleagues, Roel Hoekstra and Loek van Soest, for giving me insights into the CGN and for informative discussions.

I am indebted to the staff at the Ihinger Hof, who helped in conducting the uniformity trial, and to all persons that kindly took part in the rating experiment.

Also I like to thank Dr. Helge Kreutzmann for proofreading and all my friends for their encouragement and friendship.

I am deeply grateful for the encouragement and support of my family.

Sincere thanks!

Without the financial support of the German Research Foundation (DFG) this work never would have been accomplished.

KARIN HARTUNG

Persönliche Angaben

Geburtsdaten: 17.12.1970, Langen (Hessen)

Staatsangehörigkeit: deutsch

Schulischer Werdegang:

Grundschule: Friedens-Schule, Essen (NRW)	Aug 1977 – Juni 1981
Gymnasium: Maria-Wächtler-Schule, Essen (NRW)	Aug 1981 – Mai 1991

Studium

Biologie (Diplom)

Grundstudium: Ruhr-Universität, Bochum		WS 91/92 – WS 94/95
Hauptstudium: Ernst-Moritz-Arendt Universität, Greifswald		SoSe 95 – WS 98/99
Prüfungsfächer:	Ökologie,	
	Landschaftsökologie & Naturschutz,	
	Geologie (Schwerpunkt: Paläontologie)	
Diplomarbeitsthema:	Lumbriziden und Bodeneigenschaften des Demeterhofes Marienhöhe.	

Ökologische Landwirtschaft (Master of Science)

Qualifikationsstudium: Ruhr-Universität, Witzenhausen		SoSe 99 – WS 01/02
Prüfungsfächer:	Theorie nachhaltiger Pflanzennutzungssysteme	
	Theorie ländlicher Entwicklung	
	Soziokulturelle Dimension der Entwicklung	
	Tiere in landwirtschaftlichen Nutzungssystemen	
Diplomarbeitsthema:	„Was hält Lebensgemeinschaften zusammen?“	
	Konflikte und Lösungen in alternativen Lebensgemeinschaften.	

Erklärung

Stuttgart, November 2006

Hiermit erkläre ich eidesstattlich, die vorliegende Arbeit selbst verfasst zu haben und alle zur Hilfe genommenen Quellen angegeben zu haben. Wörtlich oder inhaltlich übernommene Stellen wurden als solche gekennzeichnet.

Die vorliegende Arbeit wurde in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde oder Institution vorgelegt.

Ich erkläre, weder früher noch gleichzeitig einen Antrag auf Eröffnung eines Promotionsverfahrens unter Vorlage der vorliegenden Dissertation gestellt zu haben.

Karin Hartung