Theor Appl Genet (2014) 127:1949–1962
DOI 10.1007/s00122-014-2351-6

ORIGINAL PAPER

Exploitation of yield stability in barley

**Jonathan Mühleisen · Hans-Peter Piepho ·
Hans Peter Maurer · Yusheng Zhao ·
Jochen Christoph Reif**

Received: 4 February 2014 / Accepted: 30 June 2014 / Published online: 24 July 2014

Abstract

Key message **Analyses of registration trials of winter barley suggested that yield and yield stability can be enhanced by developing hybrid instead of line varieties.**

Abstract Yield stability is central to cope with the expected increased frequency of extreme weather conditions. The objectives of our study were to (1) examine the dimensioning of field trials needed to precisely portray yield stability of individual winter barley (*Hordeum vulgare* L.) genotypes, (2) compare grain yield performance and yield stability of two-rowed lines with those of six-rowed lines and hybrids, and (3) investigate the association of various agronomic traits with yield stability. Static and dynamic yield stability as well as grain yield performance was determined in five series of 3-year registration trials of winter barley in Germany. Each series included 4 or 5 six-rowed hybrids, 40–46 six-rowed inbred lines, as well as 42–49 two-rowed inbred lines. The genotypes were evaluated in 10–45 environments, i.e. year-by-location combinations. We found that precise assessment of yield stability of individual genotypes requires phenotyping in at least 40 test environments. Therefore, selection for yield stability is not usually feasible since the required number of test environments exceeds the common capacity of barley breeding programs. Also, indirect improvement of yield stability by means of agronomic traits seemed not possible since there was no constant association of any agronomic trait with yield stability. We found that compared with line varieties, hybrids showed on average higher grain yield performance combined with high dynamic yield stability. In conclusion, breeding hybrid instead of line varieties may be a promising way to develop high yielding and yield stable varieties.

Introduction

Yield stability measures how reliable a genotype performs across varying growing conditions in target regions mostly defined as different locations and years. Yield of a genotype in a specific environment is an integrated result of many physiological and biochemical processes which take place in the crop plant (Ceccarelli et al. 1991). The growing conditions of the genotype as well as the genotype's genetic constitution influence the individual growth processes. Yield stability is therefore a highly complex product of the genotypes' reaction to changing environmental factors such as temperature, water and nutrient supply, radiation, and disease pressure.

Yield stability is commonly assessed by statistical approaches describing the highly complex reaction of genotypes to varying growing conditions with a single parameter (Piepho 1992). This single parameter is of use

---

Communicated by Emilio A. Carbonell.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-014-2351-6) contains supplementary material, which is available to authorized users.

J. Mühleisen · H. P. Maurer
State Plant Breeding Institute, University of Hohenheim,
70593 Stuttgart, Germany
e-mail: Jonathan.Muehleisen@uni-hohenheim.de

H.-P. Piepho
Bioinformatics Unit, Institute of Crop Science,
University of Hohenheim, 70593 Stuttgart, Germany

Y. Zhao · J. C. Reif (✉)
Leibniz Institute of Plant Genetics and Crop Plant Research
(IPK), Corrensstraße 3, 06466 Gatersleben, Germany
e-mail: reif@ipk-gatersleben.de

for selection and variety description. Several parameters have been suggested to measure yield stability, which can be assigned either to the static or to the dynamic concept of stability (Becker and Leon 1988). Genotypes with a constant yield performance across environments are stable according to the static concept. The main disadvantage of the static concept is that genotypes that honor favorable growing conditions in terms of increasing yield are not stable according to the static concept when the environments comprise varying fertility levels. Therefore, static stability is expected to be associated with low yield (Lin et al. 1986; Becker and Leon 1988). For genotypes that are stable according to the dynamic concept, changes in yield performance must correspond to the changes of the average performance of all genotypes across environments. The underlying assumption in the dynamic concept is that stable genotypes respond to changing environmental conditions, and that the response is to some extent predictable.

Dynamic stability can be estimated by Shukla's stability variance (Shukla 1972), which is the genotype-specific genotype-by-environment interaction variance. Static stability can be estimated by regressing individual genotype yields on the environmental yield levels (Finlay and Wilkinson 1963). A low value of the regression coefficient indicates that a genotype responds little to changing environmental conditions and is less sensitive to changing growing conditions and therefore stable according to the static concept. When, in addition to the regression coefficient, the variance of deviations from the regression is considered (Eberhart and Russell 1966); dynamic yield stability can be measured in a way that is conceptually very similar to Shukla's stability variance.

Estimates of heritability are important to judge the quality of the measurements of the trait under consideration. In one frequently applied method, repeatability of yield stability measures is determined by dividing the total number of environments into two subsets, calculating yield stability in both subsets, and estimating the rank correlation coefficient between yield stability values in both subsets. This rank correlation coefficient can be regarded as a measure closely related to, but not equal to the square root of heritability (for details see "Materials and methods", "Results" and "Appendix"). Subsets can be created grouping the environments by growing seasons, yield levels or based on random sampling (Eagles and Frey 1977). Applying the subdivision approach revealed that the heritability of static and dynamic stability parameters is substantially lower than that of grain yield (Pham and Kang 1988; Leon and Becker 1988; Jalaluddin and Harrison 1993; Sneller et al. 1997; Kumar et al. 1998; Robert 2002). As a consequence of the low heritability, yield stability is only rarely estimated in plant breeding trials and is no criterion for the official variety testing trials in Germany. Especially in the official variety testing, however, phenotypic data of a large number of environments are available, which may be suitable to estimate yield stability with sufficient precision. Coupled with the increasing demand to breed varieties with high yield stability coping with high seasonally averaged temperatures, changing patterns of rainfall and increasing incidence of extreme weather, the question arises whether yield stability can be included as relevant trait for official variety testing. Moreover, for studies investigating the genetic basis of yield stability, dimensioning of phenotyping is an important issue, but the current estimates of the number of environments needed to portray yield stability properly are ranging from 10 (Becker 1987) to 200 (Piepho 1998).

Our study was based on five series of 3-year registration trials of winter barley conducted in Germany. The objectives were to (1) examine the dimensioning of field trials needed to precisely assess yield stability of individual genotypes, (2) compare grain yield performance and yield stability of two-rowed lines with those of six-rowed lines and hybrids, and (3) investigate the association of various agronomic traits with yield stability.

Materials and methods

Field experiments and plant materials

The data comprised five series of winter barley registration trials conducted in Germany (Table 1). The series included three genotypic groups: six-rowed hybrids, six-rowed inbred lines, and two-rowed inbred lines. The Federal Plant Variety Office (Bundessortenamt, Hannover) tested winter barley variety candidates and check varieties in multi-location field trials in three consecutive years. After each year some candidates were rejected and check varieties could change. The number of environments was not constant across series since a varying number of trials were discarded by the Federal Plant Variety Office, e.g., due to frost damage or high field heterogeneity.

Since fungicides and growth regulators mask disease susceptibility and other traits like lodging, genotypes were tested applying a reduced crop protection scenario, besides standard crop protection. Therefore, the experimental design was a split plot design with level of crop protection as main-plot factor and genotypes as sub-plot factor. The main-plot factor crop protection had two intensity levels. Intensity level one entailed reduced crop protection, where no fungicides were applied and growth regulators only in exceptional cases at up to 50 % of the local custom application rate. In intensity level two, fungicides and growth regulators were applied according to locally adapted best

Table 1 Number of six-rowed hybrids, six- and two-rowed inbred lines of winter barley in five series of 3-year registration trials in Germany

Series	Complete set[a]			Balanced set			Total number of environments
	Six-rowed hybrids	Six-rowed lines	Two-rowed lines	Six-rowed hybrids	Six-rowed lines	Two-rowed lines	
2006–2008	4	43	48	0	13	14	44
2007–2009	4	46	42	1	7	11	43
2008–2010	5	40	49	0	10	12	45
2009–2011	4	40	46	1	7	8	42
2010–2012	5	45	44	3	10	13	39

The complete set comprised all genotypes included in the registration trials and the balanced set is a subset thereof and comprised those genotypes, which were tested in all environments

[a] In some federal states of Germany, resident agricultural departments (Länderdienststellen) included varieties of their interest in the registration trials. Such varieties were tested in only up to four environments and were therefore not part of this study. All other genotypes were tested in at least ten environments

practices. For grain yield, the performance of the genotypes under intensive crop protection was more relevant since farmers usually apply intensive crop protection. Therefore, we used only the yield data of the intensive crop protection for our study. Each trial comprised two replicates and 16–95 genotypes (Table 1). In exceptional cases three replicates were used or additional varieties were included due to the interest of the resident agricultural departments (Länderdienststellen). However, these varieties were not considered in the present study. The Federal Plant Variety Office provided for each location the genotype means and plot data for intensity level two, where fungicides and growth regulators were applied according to locally adapted best practices. Grain yield was adjusted to a moisture content of 14 % and converted to Mg ha^{-1}.

In addition, the Federal Plant Variety Office provided annual mean values for several agronomic traits. The traits were date of heading and date of maturity, which were converted to days after beginning of the years, visual scores from 1 to 9 for low to high susceptibility for lodging at harvest, stem breaking, ear breaking, powdery mildew caused by *Blumeria graminis*, net blotch caused by *Pyrenophora teres*, scald caused by *Rhynchosporium secalis*, leaf spot caused by *Ramularia collo-cygni*, physiological leaf spots and leaf rust caused by *Puccinia hordei*, plant height in cm, visual scores from 1 for no deficiencies after winter to 9 for all plants killed, and plant density in ears m^{-2}. For deficiencies after winter and plant density, genotype mean values were based on both intensity levels, for the other traits only on intensity level one. Use of values from trials without intensive crop protection for description of resistance to diseases and straw characteristics was reasonable, since differences between genotypes were more pronounced in trials without application of fungicides and growth regulators. We have to point out that deficiencies after winter were not necessarily caused by frost damage. This scoring comprised all visually detectable deficiencies observed in a plot after winter, independently of the reason.

Statistical analysis

All series were analyzed separately. We assumed that years as well as locations are interchangeable and that each year-by-location combination (environment) represents a random sample of the growing conditions for winter barley in Germany. The locations were not grouped into regions. We investigated the relationship between environments via cluster analysis but could not find distinct groups of environments. We expect that this is due to the relatively homogenous climatic conditions in Germany. Furthermore, division of Germany into regions would lead to a low number of environments per regions in our dataset and therefore to poor regional yield stability estimates. To provide an overview of the different sources of variation we fitted the following model:

$$y_{ijk} = \mu + g_i + u_j + (gu)_{ij} + r_{jk} + e_{ijk},$$

where y_{ijk} is the plot grain yield of the ith genotype of the kth replicate in the jth environment, μ is the intercept, g_i is the main effect of the ith genotype, u_j is the main effect of the jth environment, $(gu)_{ij}$ is the genotype-by-environment interaction effect of the ith genotype with the jth environment, r_{jk} is the effect of the kth replicate in the jth environment, and e_{ijk} is the residual of y_{ijk}. μ is assumed to be fixed, whereas all other effects are assumed to be random effects with independent normal distribution, zero mean and variances σ_g^2, σ_u^2, σ_{gu}^2, σ_r^2, and σ_ε^2, respectively. Note that it may be more realistic to allow for heterogeneity in the block and error variances between environments. But here our main purpose is only to obtain a general overview of the importance of different model terms; for this purpose, we consider the simplifying and more convenient assumption of homogeneous variance as sufficient.

Estimation of yield stability parameters

Estimation of average grain yield and yield stability parameters was done applying the linear mixed models for mean data suggested by Piepho (1999). Only the environmental variance model was not estimated using a mixed model with unstructured variance–covariance matrix, since for this purpose the number of environments has to be larger than the number of genotypes, which was not the case for the data at hand. In the models described by Piepho (1999), stability parameters were estimated for individual genotypes. Additionally, we estimated differences in yield stability between genotypic groups, i.e. differences between six-rowed hybrids, six-rowed inbred lines, and two-rowed inbred lines following Mühleisen et al. (2014). All models were fitted with restricted maximum likelihood (REML) using the MIXED procedure of the SAS System (SAS Institute Inc. 2011). REML is applicable for unbalanced data, i.e., when not each genotype was present in each environment. But it has to be assumed that the data meet the missing-at-random assumption, which in turn requires that all data used for selection decisions must be included in the dataset (Piepho and Möhring 2006). Therefore, all genotypes included in a series (complete set; Table 1) were analyzed together, regardless whether the individual genotypes were tested in one or more years. Mean grain yield (YIELD) of individual genotypes across environments was estimated with the following model (two-way ANOVA model in Piepho 1999):

$$\bar{y}_{ij} = \mu + g_i + v_j + f_{ij}$$

where $\bar{y}_{ij}$ is the mean grain yield of the ith genotype in the jth environment, μ is the intercept, g_i is the main effect of the ith genotype, v_j is the main effect of the jth environment confounded with the replicate effects, f_{ij} is the genotype-by-environment interaction effect of the ith genotype with the jth environment confounded with the residual effects. Effects μ and g_i are assumed to be fixed, whereas v_j and f_{ij} are random effects with independent normal distribution, zero mean and variances $\sigma_v^2 = \sigma_u^2 + \sigma_r^2/K$, and $\sigma_f^2 = \sigma_{gu}^2 + \sigma_\varepsilon^2/K$, respectively, where σ_u^2 is the environment variance, σ_r^2 the replicate variance, σ_{gu}^2 the genotype-by-environment interaction variance, σ_ε^2 the residual variance, and K the number of replicates per environment. Best linear unbiased estimators (BLUEs) of genotype means were calculated. For genotype group-specific estimation of grain yield, we modified the model to:

$$\bar{y}_{qij} = \mu + t_q + g_{qi} + v_j + (vd)_{qj} + f_{qij}$$

where t_q is the main effect of the qth genotypic group, g_{qi} is the main effect of the ith genotype within the qth genotypic group, and $(vd)_{qj}$ is the interaction effect of the qth genotypic group with the jth environment. Effects t_q are assumed to be fixed, whereas g_{qi} and $(vd)_{qj}$ are random effects with independent normal distribution, zero mean and variances σ_g^2, and σ_d^2, respectively, where σ_g^2 is the genotypic variance, and σ_{vd}^2 the group-by-environment interaction variance. Significant differences between genotypic group means were tested with a t test using the PDIFF option in the LSMEANS statement of the MIXED procedure of the SAS system (SAS Institute Inc. 2011). Adjustment of p values for multiple testing was done according to Kramer (1956) as cited in SAS Institute Inc. (2011).

Shukla's stability variance (SVAR) of individual genotypes was estimated in the following way: the genotype-by-environment interaction variance σ_{gu}^2 was assumed to be specific for each genotype ($\sigma_{gu(i)}^2$). The genotype-by-environment interaction variance of genotype i is the stability variance of genotype i. A smaller stability variance indicates a higher stability according to the dynamic concept. Due to the use of mean data, we could only estimate $\sigma_{f(i)}^2$ and termed this variance stability variance (SVAR). For group-specific estimates of stability variance, $\sigma_{f(i)}^2$ was replaced by $\sigma_{f(q)}^2$, which denotes the stability variance of the genotypic group q, i.e., the variance of the residual effect f_{qij} for group q.

The Finlay–Wilkinson as well as the Eberhart–Russell model can be written as factor analytic model in the following way:

$$\bar{y}_{ij} = \mu + g_i + \lambda_i w_j + f_{ij}$$

where λ_i is the sensitivity (SEN-FW for the Finlay–Wilkinson model; SEN-ER for the Eberhart–Russell model) of the ith genotype to a latent environmental variable w_j. The variance of w_j is σ_w^2. The term $\lambda_i w_j$ is over-parameterized and therefore we set $\sigma_w^2 = 1$ following Piepho (1999). The interpretation of sensitivity is straightforward. A genotype with a small sensitivity is stable according to the static concept. The effect f_{ij} denotes a random deviation and is confounded with the error variance. The variance of f_{ij} is $\sigma_f^2 = \sigma_d^2 + \sigma_\varepsilon^2$ for the Finlay–Wilkinson model and $\sigma_{f(i)}^2 = \sigma_{d(i)}^2 + \sigma_\varepsilon^2$ for the Eberhart–Russell model, where σ_d^2 is the deviation variance and σ_ε^2 the residual variance. In the Eberhart–Russell model the deviation variance is assumed to be specific for each genotype ($\sigma_{d(i)}^2$) and can be used as dynamic stability measure in analogy to the stability variance. Since the analysis was based on mean data, we could only estimate $\sigma_{f(i)}^2$ and termed this variance deviation variance (DVAR). For estimation of group-specific sensitivity and deviation variance, λ_i was replaced by λ_q and $\sigma_{f(i)}^2$ by $\sigma_{f(q)}^2$. Significant differences between group-specific stability parameters were tested with a Wald test using the COVTEST statement of the GLIMMIX procedure of the SAS system (SAS Institute Inc. 2011), where under the null hypothesis a common value for the stability parameters of the two genotypic groups under comparison was assumed, but in the alternative hypothesis a separate one.

For some genotypes, mostly for genotypes tested in a low number of environments, estimates of SVAR and DVAR had very large values, which were outside the plotting area of Figs. 3 and 4. Therefore, grain yield, stability parameters and description of these genotypes were given in the Supplementary Table S1.

Heritability of yield stability

One part of the present study was devoted to the heritability of grain yield and yield stability parameters and the association between magnitude of heritability and the number of test environments. For this purpose, we considered the correlation coefficient between genotypic values estimated based on a reference set of environments and genotypic values estimated based on the corresponding test set of different environments. We explained in the "Appendix" how this correlation was used to obtain an approximate estimate of heritability. Spearman's rank correlation was used instead of Pearson's correlation to make results more robust to outliers. Each pair of reference and test set was determined in the following way: environments were split in two halves: one half was defined as the reference set, and a sample of three or more environments (up to all the environments in that second half) from the other half was used as test set. If the number of environments was uneven, we used the larger half as the reference set. For each series, 1,000 runs of sampling were performed for each number of test environments. Analysis of the stability models described by Piepho (1999) is computationally demanding. To simplify the analysis, we used from each series only those genotypes which were tested in all environments (balanced set, Table 1).

The average Spearman rank correlation was calculated for each number of test environments. If the average Spearman rank correlation was negative it was set to zero. The average Spearman rank correlation is not an estimate of the square root of heritability, since the estimated values of grain yield and stability parameters in the reference set were not the true values but only estimates. To adjust for that underestimation, the correlations were divided by the square root of the correlations estimated with the maximum number of test environments. This measure was termed rank-correlation-heritability. For a detailed explanation of the adjustment see the Appendix. In addition, the regression-heritability was defined as the regression coefficient of the linear regression of the estimates obtained from the analysis of the test set on the estimates from the reference set. The derivation is also shown in the "Appendix". If the average regression coefficient was negative, it was set to zero. For grain yield we estimated heritability for each test set in addition by a traditional approach following Piepho and Möhring (2007):

$$\text{broad-sense heritability } (h^2) = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\overline{vd}}{2}}, \quad (1)$$

where σ_g^2 is the genotypic variance and $\overline{vd}$ the mean variance of a difference of two adjusted genotype means (BLUEs). Estimates of heritability based on the resampling approach were only available for up to 22 test environments. However, we would like to know how large the heritability is with more than 40 environments. Therefore, we developed a formula describing the relationship between heritability and number of test environments. Ignoring design effects, heritability (h^2) can be calculated with the following equation:

$$h^2 = \frac{\sigma_g^2}{\sigma_g^2 + \frac{\sigma_f^2}{n}} = \frac{1}{1 + \frac{\sigma_f^2}{n \times \sigma_g^2}}, \quad (2)$$

where σ_g^2 is the genotypic variance, σ_f^2 the genotype-by-environment interaction variance confounded with the error variance, and n the number of test environments. Assuming that $\frac{\sigma_f^2}{\sigma_g^2}$ is a constant ratio for a given population of environments and genotypes leads to the equation:

$$h^2 = \frac{1}{1 + \frac{c}{n}}. \quad (3)$$

For a given value of c, heritability can be estimated for any number of test environments. We assumed that c is specific for each series. To obtain estimates for c, a non-linear regression with formula (3) based on the estimated values of heritability was performed for each series for the parameter grain yield. For the stability parameters, n was defined as number of test environments minus two. The intuition behind this heuristic definition is that the minimum number of environments needed to estimate the stability variance with the MINQUE method is three (Shukla 1972). Also, we found empirically that this definition led partially to a better fit of the non-linear regression (3) than if we define n as number of test environments. Subsequently heritability was predicted based on the estimated values of c for up to 50 test environments.

Relationship between grain yield and yield stability

Spearman's rank correlation coefficient was calculated between grain yield and dynamic and static yield stability parameters for the five series. In addition, the association between grain yield and stability parameters was visually assessed by scatter plots.

Table 2 Estimates of variance components (σ_g^2 genotypic, σ_u^2 environment, σ_{gu}^2 genotype-by-environment interaction, σ_r^2 replicate, σ_ε^2 error) for the complete set (*c*) and the corresponding balanced set (*b*) of five series of multi-environment registration trials of winter barley evaluated for grain yield (Mg ha^{-1}) in Germany

Series	2006–2008		2007–2009		2008–2010		2009–2011		2010–2012	
Set	*c*	*b*	*c*	*b*	*c*	*b*	*c*	*b*	*c*	*b*
σ_g^2	0.125	0.100	0.125	0.090	0.116	0.071	0.102	0.086	0.113	0.060
σ_u^2	2.637	2.739	2.250	2.284	2.100	2.150	2.056	2.102	1.291	1.335
σ_{gu}^2	0.162	0.150	0.229	0.193	0.155	0.182	0.185	0.201	0.191	0.202
σ_r^2	0.052	0.061	0.036	0.038	0.046	0.051	0.053	0.059	0.031	0.034
σ_ε^2	0.153	0.163	0.200	0.160	0.191	0.176	0.197	0.185	0.188	0.191

Association of agronomic traits with yield stability

Each agronomic trait (heading, maturity, lodging, stem breaking, ear breaking, powdery mildew, net blotch, rhynchosporium, ramularia, physiological leaf spots, leaf rust, deficiencies after winter, plant height, and plant density) in each series was analyzed with the following model:

$$\bar{y}_{ij} = \mu + g_i + a_j + \varepsilon_{ij}$$

where $\bar{y}_{ij}$ is the mean of the agronomic trait of the ith genotype in the jth year, μ is the intercept, g_i is the effect of the i th genotype, a_j is the effect of the jth year, and ε_{ij} is the error associated with $\bar{y}_{ij}$. All effects except ε_{ij} were assumed to be fixed, and ε_{ij} was assumed to be random and independently and identically distributed with zero mean. BLUEs of genotypes were extracted and standardized in the following way:

$$z_i = (x_i - \bar{x})/s_i$$

where z_i is the standardized value, x_i is the BLUE of the ith genotype, $\bar{x}$ the mean of the BLUEs of all genotypes, and s_i the corresponding standard deviation.

For each series and the response variables YIELD, SVAR, DVAR, SEN-FW, and SEN-ER, a multiple linear regression model was fitted, where standardized values of all 14 agronomic traits were used as explanatory variables. For model selection we used the stepwise method. For each explanatory variable, a partial F statistic was calculated, which measures the association of the variable with the response and only variables with a p value less than or equal to 0.05 entered the model. When a new variable entered the model, p values of all variables, which had already been entered into the model, were calculated again and only those with p values still less than or equal to 0.05 remained in the model. The regression analysis was performed using the REG procedure of the SAS system (SAS Institute Inc. 2011). Intercept, regression coefficients and adjusted R^2 were reported.

To assess how transferable the estimated intercept and regression coefficients were across series, estimated values of intercept and each regression coefficient were averaged across four series. The regression model formed in this way was used to predict the remaining series. The R^2 for this prediction, henceforth denoted as transferable R^2, was calculated by squaring the Pearson's correlation coefficient between observed and predicted values. This was done for each series and each response parameter.

Results

In all five series the contribution of the genotype-by-environmental interaction variance was substantial (Table 2). In the complete sets, the genotype-by-environment interaction variance on average was 1.6 times larger than that of the genotypic variance and in the balanced sets even 2.4 times larger. Therefore, utilization of this source of variation in breeding by selecting genotypes with high yield stability may help to increase barley production.

Heritability of grain yield and yield stability

Spearman's rank correlation between yield stability parameters estimated in the test and in the reference set was mostly positive (Fig. 1). With increasing number of test environments, the correlation increased for all stability parameters in all series except for SVAR in the series 2008–2010. The magnitude of rank correlations and the improvement with increasing number of test environments were substantially lower and less consistent across series for yield stability compared with grain yield. The median of the correlations of grain yield between test and reference sets was larger than 0.50 in all series already with three test environments. In contrast, the medians of the observed correlations of yield stability between test and reference sets were not larger than 0.50 except for SEN-FW and SEN-ER in the series 2006–2008 and for SVAR and DVAR in the series 2007–2009. For Shukla's stability variance model, convergence criteria were frequently not met, when a small number (<10) of test environments was used for estimation.

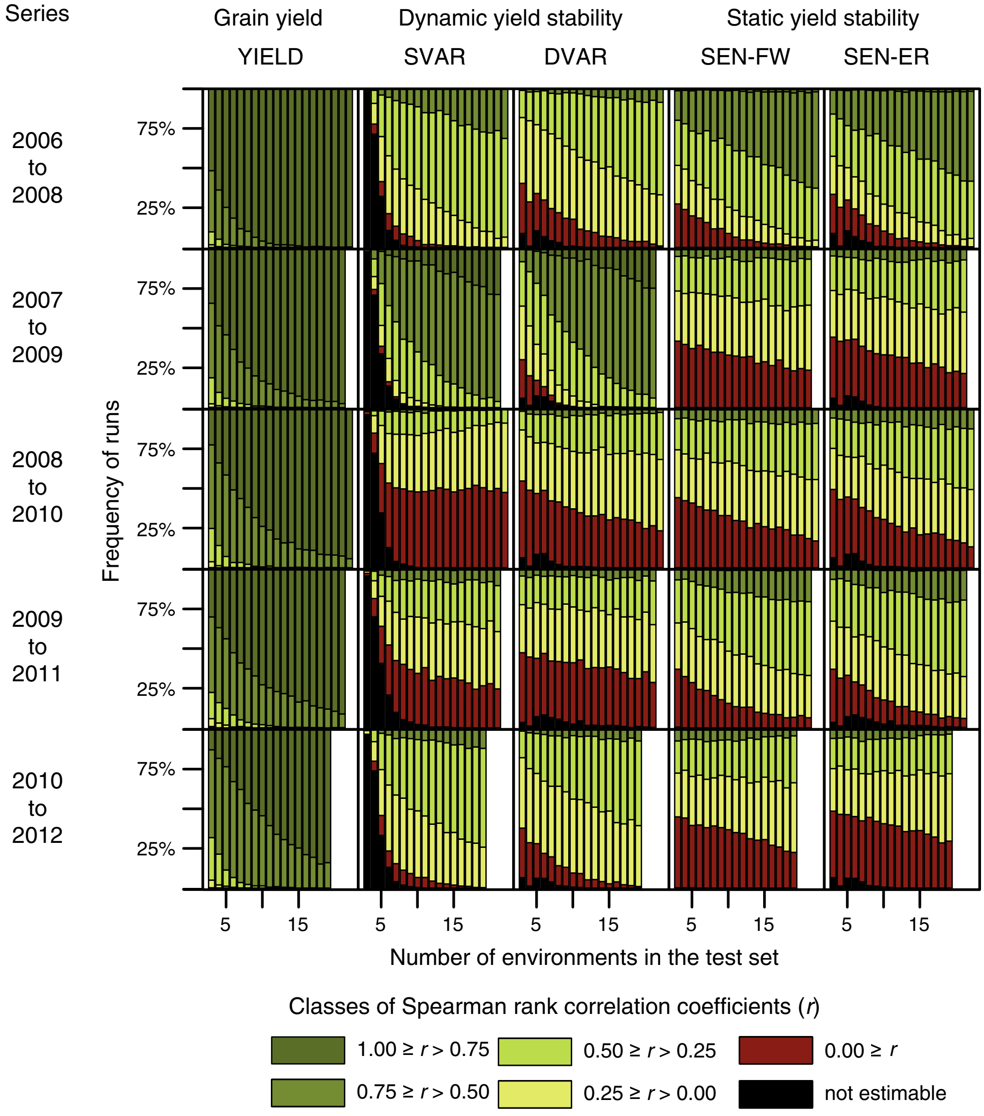

Fig. 1 Spearman rank correlation coefficients between grain yield and yield stability (*SVAR* stability variance, *DVAR* deviation variance, *SEN-FW* sensitivity estimated with the Finlay–Wilkinson model, *SEN-ER* sensitivity estimated with the Eberhart–Russell model) estimated in varying reference and test sets of environments

In the other stability models, convergence problems appeared less frequently.

For grain yield, the rank-correlation-heritability and the regression-heritability approximated broad-sense heritability well (Fig. 2). The squared rank correlation was constantly lower than the three heritabilities, indicating that the squared rank correlation cannot be directly used as a precise measure of heritability. Across series the magnitude as well as the increase of heritability with rising number of test environments was similar.

Rank-correlation-heritability and the regression-heritability were generally lower for stability parameters compared with grain yield (Fig. 2). There were only marginal differences between both heritabilities, except for SVAR and DVAR in the series 2010–2012, where the regression-heritability was substantially higher. We found that the high regression-heritability was caused by one extremely unstable genotype. Hence, this definition of heritability is strongly influenced by individual genotypes with extreme values. Rank-correlation-heritability is a preferable measure here because of its robustness against outliers. There was an increase in magnitude of the heritabilities for all stability parameters with increasing number of test environments in all series except for SVAR in the series 2008–2010. With a small number of test environments, the increase in heritabilities was lower for stability parameters than for grain yield. But while increase in heritabilities for grain yield became marginal after ten test environments, the heritabilities of stability parameters increased further.

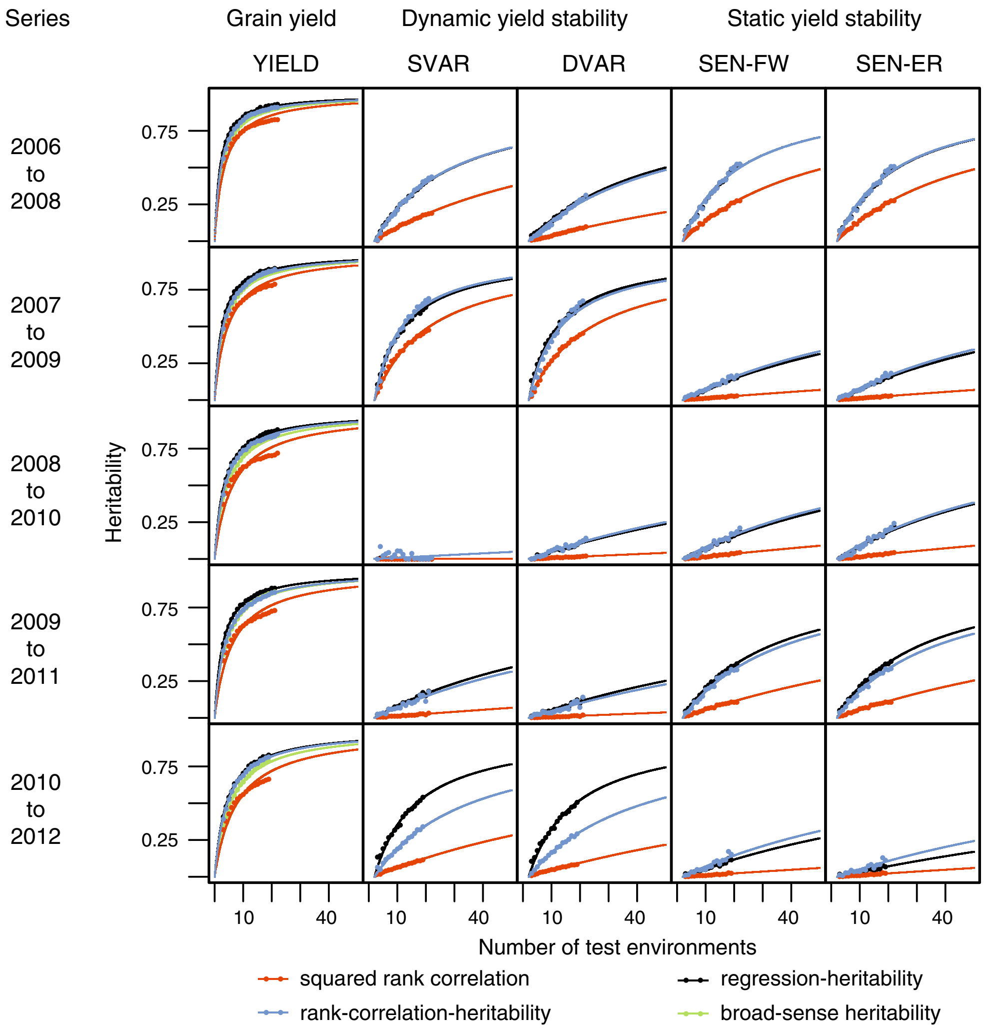

Fig. 2 Observed (*dots*) and predicted (*lines*) values of the squared rank correlation and up to three types of heritability of *YIELD* grain yield, *SVAR* stability variance, *DVAR* deviation variance, *SEN-FW* sensitivity estimated with the Finlay–Wilkinson model, *SEN-ER* sensitivity estimated with the Eberhart–Russell model for an increasing number of test environments in five series of multi-environment barley trials

Grain yield and yield stability of six-rowed hybrids, six- and two-rowed inbred lines

Grain yield and yield stability parameters of individual genotypes revealed large overlaps among six-rowed hybrids, six-rowed lines and two-rowed lines (Fig. 3). For grain yield there were consistent tendencies in all five series, suggesting that yield performance of six-rowed hybrids was highest, followed by six-rowed lines and two-rowed lines. In contrast to grain yield, tendencies between genotypic groups were less visible for stability parameters (Fig. 3). However, estimation of group-specific stability parameters revealed that there were significant differences in yield stability and between the genotypic groups (Table 3). In addition, yield stability of genotypic groups could be determined more precisely than that of individual genotypes, because standard errors of stability estimates were smaller for genotypic groups than for individual genotypes (Supplementary Table S2). Six-rowed hybrids showed high dynamic yield stability, expressed in low values for SVAR and DVAR. Neither six-rowed lines nor two-rowed lines had significantly smaller values in any series. In terms of static yield stability (SEN-FW and SEN-ER), two-rowed lines were most stable. Six-rowed lines were superior neither in static nor in dynamic yield stability.

Relationship between grain yield and yield stability

Grain yield and SVAR were significantly correlated ($r = 0.28$; $p < 0.05$) only in the series 2006–2008. In the other series grain yield and SVAR as well as grain yield and DVAR were not significantly correlated. Between

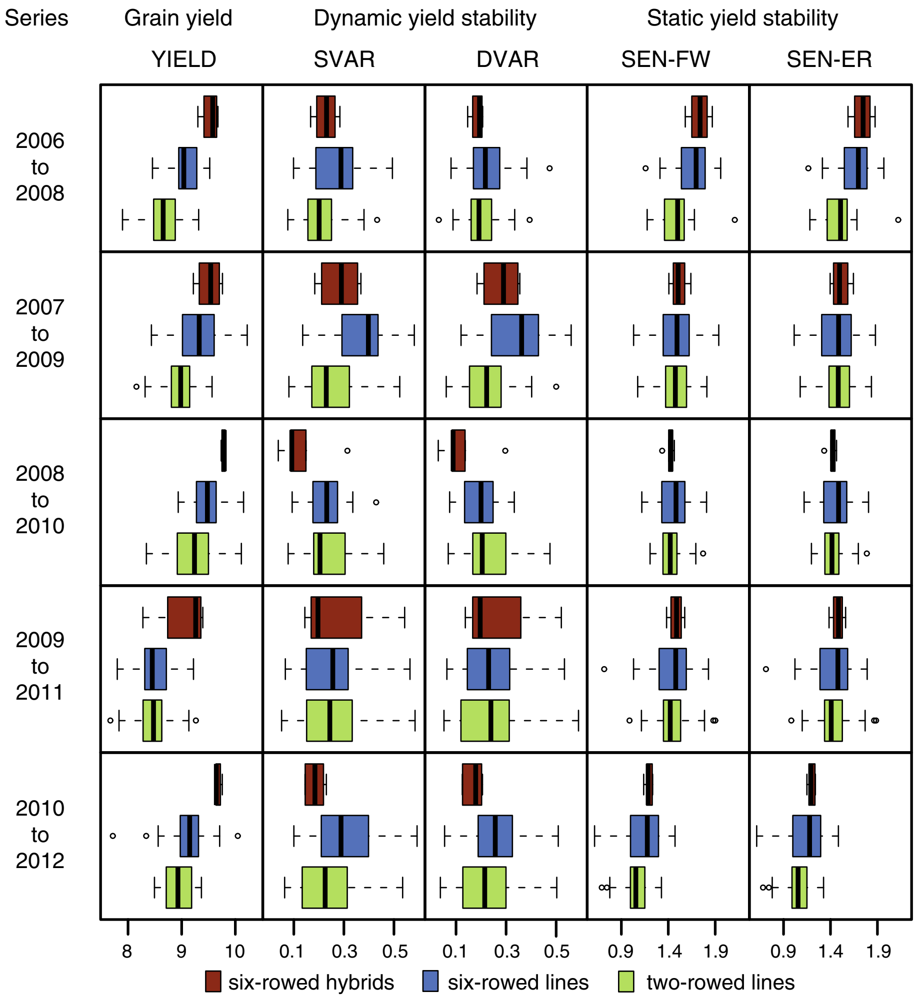

Fig. 3 Distribution of estimated values for *YIELD* grain yield, *SVAR* stability variance, *DVAR* deviation variance, *SEN-FW* sensitivity estimated with the Finlay–Wilkinson model, *SEN-ER* sensitivity estimated with the Eberhart–Russell model of six-rowed hybrids, six-rowed inbred lines, and two-rowed inbred lines of winter barley evaluated for grain yield (Mg ha^{-1}) in five series of multi-environment trials

Table 3 Estimates for *YIELD* grain yield across environments, *SVAR* stability variance, *DVAR* deviation variance, *SEN-FW* sensitivity estimated with the Finlay–Wilkinson model, *SEN-ER* sensitivity estimated with the Eberhart–Russell model of the genotypic groups of winter barley evaluated for grain yield (Mg ha^{-1}) in five series of multi-environment trials

Series	Group	YIELD	SVAR	DVAR	SEN-FW	SEN-ER
2006–2008	Six-rowed hybrids	9.53^{A}	0.22^{AB}	0.19^{AB}	1.79^{B}	1.80^{B}
	Six-rowed lines	9.07^{B}	0.28^{B}	0.26^{B}	1.73^{B}	1.73^{B}
	Two-rowed lines	8.70^{C}	0.20^{A}	0.21^{A}	1.54^{A}	1.54^{A}
2007–2009	Six-rowed hybrids	9.61^{A}	0.32^{AB}	0.31^{AB}	1.55^{A}	1.56^{A}
	Six-rowed lines	9.19^{AB}	0.42^{B}	0.42^{B}	1.52^{A}	1.52^{A}
	Two-rowed lines	9.05^{B}	0.25^{A}	0.25^{A}	1.48^{A}	1.49^{A}
2008–2010	Six-rowed hybrids	9.82^{A}	0.13^{A}	0.13^{A}	1.45^{AB}	1.45^{AB}
	Six-rowed lines	9.51^{A}	0.23^{B}	0.23^{B}	1.49^{B}	1.50^{B}
	Two-rowed lines	9.26^{B}	0.27^{C}	0.28^{C}	1.44^{A}	1.44^{A}
2009–2011	Six-rowed hybrids	9.10^{A}	0.22^{A}	0.22^{A}	1.44^{A}	1.44^{A}
	Six-rowed lines	8.61^{B}	0.32^{B}	0.32^{B}	1.47^{A}	1.47^{A}
	Two-rowed lines	8.39^{B}	0.27^{A}	0.27^{AB}	1.45^{A}	1.45^{A}
2010–2012	Six-rowed hybrids	9.68^{A}	0.19^{A}	0.18^{A}	1.19^{B}	1.20^{B}
	Six-rowed lines	9.16^{B}	0.34^{B}	0.33^{C}	1.19^{B}	1.18^{B}
	Two-rowed lines	8.95^{B}	0.25^{A}	0.25^{B}	1.08^{A}	1.08^{A}

Values within one column and series without common letter were significant different at $p < 0.05$

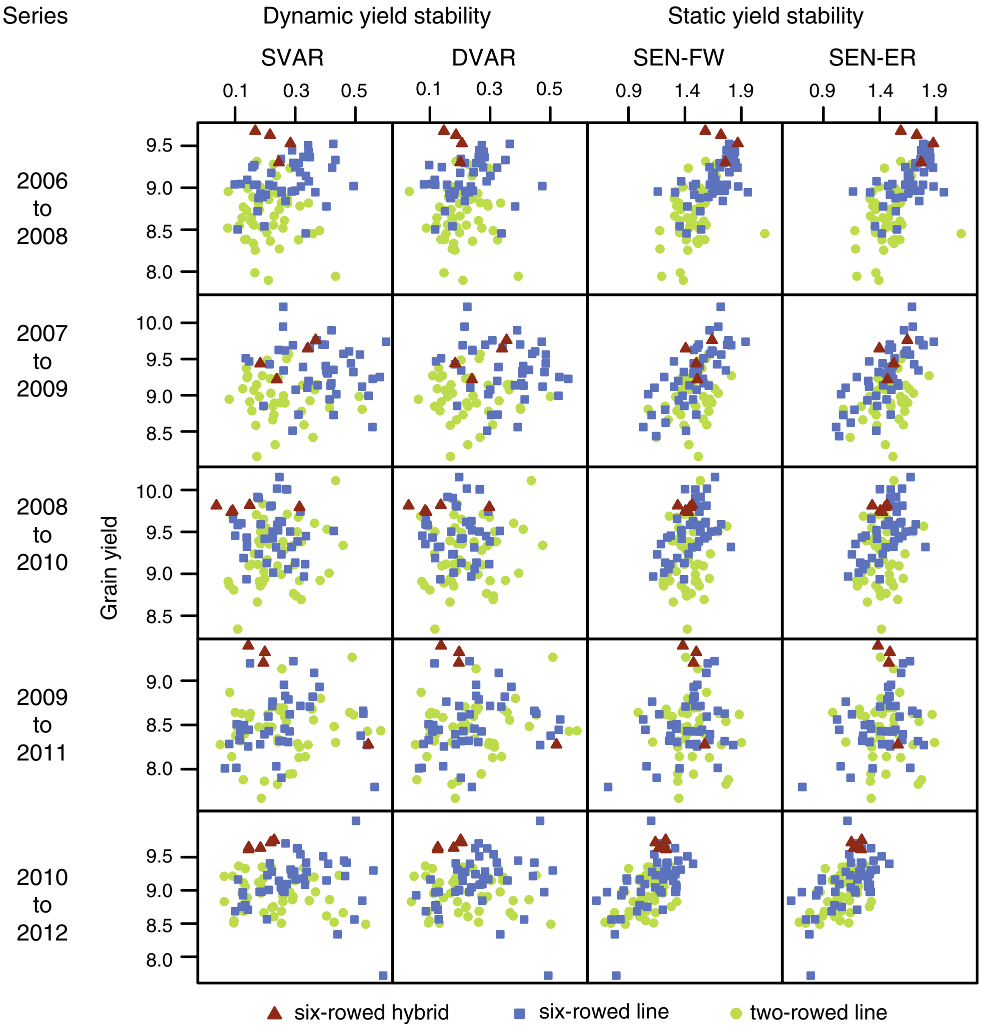

Fig. 4 *Scatter plots* between grain yield across environments and yield stability parameters (*SVAR* stability variance, *DVAR* deviation variance, *SEN-FW* sensitivity estimated with the Finlay–Wilkinson model, *SEN-ER* sensitivity estimated with the Eberhart–Russell model) of six-rowed hybrids, six-rowed inbred lines, and two-rowed inbred lines of winter barley evaluated for grain yield (Mg ha^{-1}) in five series of multi-environment trials

grain yield and the static stability parameters, SEN-FW and SEN-ER, there was a significant ($p < 0.05$) correlation above 0.3 in all series, except in the series 2009–2011 where the correlation was below 0.2 and not significant. Scatter plots showed that hybrids combined high grain yield with high dynamic yield stability in all series (Fig. 4).

Association of agronomic traits on yield stability

Standardized regression coefficients of individual series revealed that an association of agronomic traits with grain yield and yield stability was present (Table 4). However, the regression coefficients were frequently not constant across series. The adjusted R^2 ranged between 0.22 and 0.46 for grain yield, between 0.04 and 0.26 for dynamic stability and between 0.05 and 0.50 for static stability. The transferable R^2 was similar to the adjusted R^2 only for grain yield (0.18–0.33). For static and dynamic stability the transferable R^2 was mostly close to zero ($\leq$0.02) and never larger than 0.07.

Discussion

We used static and dynamic yield stability measures in our study. In the static concept, genotypes have high yield stability if they perform equally across environments. Our and also previous findings (Pham and Kang 1988; Duarte and Zimmermann 1995; Sneller et al. 1997; Mekbib 2003), however, clearly underlined that high static yield stability is frequently associated with low grain yield. This result is not surprising, because genotypes should be able to exploit

Table 4 Estimated intercept, standardized regression coefficients, adjusted and transferable R^2 of the multiple regression models regressing different response variables (*YIELD* grain yield, *SVAR* stability variance, *DVAR* deviation variance, *SEN-FW* sensitivity estimated with the Finlay–Wilkinson model, *SEN-ER* sensitivity estimated with the Eberhart–Russell model) on standardized values of agronomic traits in five series of 3-year registration trials of winter barley in Germany

Response	Series	Intercept	Standardized regression coefficients														R^2 adj.	R^2 trans.
			Heading	Maturity	Lodging	Stem breaking	Ear breaking	Powdery mildew	Net blotch	Rhynchosporium	Ramularia	Physiological leaf spots	Leaf rust	Deficiencies after winter	Plant height	Plant density		
YIELD	06/08	8.89				0.13			−0.12					−0.06		−0.21	0.46	0.35
YIELD	07/09	9.15	−0.16	0.11		0.16				−0.10		−0.09				−0.13	0.43	0.38
YIELD	08/10	9.36	−0.19	0.14						−0.11					0.17		0.33	0.22
YIELD	09/11	8.48		0.10		0.16			−0.08					−0.09			0.22	0.13
YIELD	10/12	9.06				0.14		−0.09				−0.07					0.28	0.22
SVAR	06/08	0.24							−0.03			0.03				−0.04	0.19	0.07
SVAR	07/09	0.37						0.10				0.09			0.09		0.26	0.02
SVAR	08/10	0.23					−0.02										0.04	0.00
SVAR	09/11	0.27											−0.03				0.04	0.02
SVAR	10/12	0.28				0.04											0.06	0.00
DVAR	06/08	0.21								0.02							0.06	0.03
DVAR	07/09	0.34						0.09				0.08			0.09		0.26	0.00
DVAR	08/10	0.22		0.03										−0.03			0.13	0.00
DVAR	09/11	0.25		0.04													0.08	0.01
DVAR	10/12	0.25				0.03											0.06	0.02
SEN-FW	06/08	1.57			0.09			−0.04			0.03					−0.08	0.50	0.00
SEN-FW	07/09	1.48	−0.06							−0.04							0.15	0.06
SEN-FW	08/10	1.43								−0.04		−0.03					0.11	0.01
SEN-FW	09/11	1.44	−0.13										−0.05				0.35	0.00
SEN-FW	10/12	1.09						−0.04									0.05	0.02
SEN-ER	06/08	1.57			0.09			−0.04			0.03					−0.08	0.50	0.00
SEN-ER	07/09	1.47	−0.06							−0.04							0.15	0.07
SEN-ER	08/10	1.43	−0.03							−0.04		−0.03					0.14	0.02
SEN-ER	09/11	1.43	−0.12										−0.05				0.35	0.02
SEN-ER	10/12	1.09						−0.04									0.05	0.02

yield benefits offered by better conditions. The static yield stability concept is thus only of minor relevance for the rather favorable Central European environments.

The dynamic concept of yield stability aims for genotypes that reliably exploit favorable growing conditions. By this definition an intrinsic conflict in combining high yield stability with high grain yield is not given in contrast to the static yield stability. This is also reflected by the frequent absence of a significant correlation between grain yield and dynamic yield stability in our study. Therefore, we will focus in the discussion on dynamic yield stability, if not stated otherwise.

Precise determination of yield stability requires intensive phenotyping

Becker (1987) estimated the heritability of the regression slope and of the deviation variance based on their variance components, which he estimated using registration trials of wheat and barley. He suggested that precise determination of the deviation variance requires phenotyping in at least 10–15 test locations. In addition, his results showed that two and three years evaluation of genotypes improved the heritability considerably compared to evaluation in a single year. Lin and Binns (1991) investigated if there was an additive genetic contribution to various stability parameters based on a diallel evaluated in 12 environments. For the environmental variance, a static stability parameter, they found a genetic influence. For stability and deviation variance, they reported that there was no additive genetic contribution and concluded that these measures were nongenetic. We would rather conclude from this finding that the respective stability estimates based on 12 environments were not precise enough to detect the genotypic influence. Piepho (1998) considered the expected values of the coefficient of variation (CV) of the estimated environmental variance, and concluded that 50 test environments are needed to obtain a CV of the environmental variance of 20 %.

We applied subsampling based on experimental data sets and developed an approach to estimate the expected heritability for an arbitrary number of environments. Our results suggested that at least 40 environments are required to reach a heritability of 0.5 for the different yield stability parameters (Fig. 2). It is important to point out that we observed a high heterogeneity in heritability estimates of yield stability parameters across the different series, which was not the case for grain yield. Consequently, for official variety testing of barley, use of yield stability as a selection criterion cannot be recommended despite the large phenotyping intensity of the final set of genotypes at more than 40 environments.

The final set of genotypes in registration trials is strongly pre-selected across years not only during the official variety testing but also before that in the course of multi-stage selection. In addition, the number of genotypes within each series tested across 3 years is limited with a maximum of 27 genotypes (Table 1). Both led potentially to a narrowing of the genetic variation of yield stability for the population under consideration (compare Becker 1987). Consequently, for investigating the genetic basis of yield stability in a larger and unselected panel of genotypes, a lower number of environments are likely required to reach a heritability of 0.5.

On average, barley hybrids outyielded inbred lines and displayed higher yield stability

Despite substantial overlap of grain yield performance at the individual level (Fig. 3), we observed on average that six-rowed hybrids significantly ($p < 0.05$) outyielded six-rowed lines in three series and two-rowed lines in all five series (Table 3). The observed differences between the average performances of the three groups coincided well with previous studies investigating grain yield of hybrids versus lines (Mühleisen et al. 2013) as well as of two- and six-rowed lines in Western and Central Europe (Aufhammer and Kübler 1987; Le Gouis 1992, 1999; Maidl et al. 1996).

Six-rowed lines showed lower dynamic yield stability in comparison with two-rowed lines (Table 3) on average in almost all series, despite substantial overlap at the individual level (Fig. 3). Interestingly, in all series six-rowed hybrids were in the top class with highest dynamic yield stability. The higher yield stability of hybrids compared with lines was in accordance with previous findings for six-rowed genotypes (Mühleisen et al. 2013, 2014).

Precise phenotyping of yield stability requires testing at a large number of environments as shown above. The required test resources are more than those commonly available in barley breeding programs. Therefore, use of yield stability as selection criterion is not recommended. Consequently, choice of the most yield stable type of variety (hybrids) represents a more promising strategy to breed strongly desired high yielding and yield stable barley genotypes.

Yield stability cannot be improved by indirect agronomic traits

We observed a significant association of multiple agronomic traits with grain yield (Table 4), despite intensive treatment with fungicides and growth regulators. High grain yield tended to be associated with early heading, late maturity, high susceptibility to stem breaking, low susceptibility to leaf diseases, few deficiencies after winter, high plant height, and low plant density. The association of agronomic traits with grain yield depended on the series but R^2 values transferable across series still amounted to 0.26 on average (Table 4). Therefore, agronomic traits such as leaf disease resistances are of use as indirect trait for grain yield in early stages of selection.

Yield stability was not consistently associated with any agronomic trait (Table 4). The association of agronomic traits with yield stability depended heavily on the series with R^2 values transferable across series amounting to 0.02 on average (Table 4). Yield stability might be caused by a combination of agronomic traits (Ceccarelli et al. 1991). In this case, the impact of individual agronomic traits can be masked by the higher interaction contributions of the individual components. Consequently, no indirect trait could be suggested to efficiently select for high yield stability.

Prospects of breeding for high yield and enhanced yield stability

Selection for improved dynamic yield stability is hardly possible due to the large phenotyping requirements with a minimum of 40 test environments. Also the indirect improvement of yield stability by modifying individual traits seemed not feasible, since the association of agronomic traits with yield stability was not transferable across series. Based on the presented results, a promising way breeding high yielding and yield stable genotypes is the development of hybrid instead of line varieties.

Acknowledgments We gratefully acknowledge the permission by the Federal Plant Variety Office (Bundessortenamt, Hannover) to use the barley data for this study. We thank Yong Jiang for developing the heritability formula used for extrapolation. J. Mühleisen was supported by BMELV/BLE within the „Züchtung von Triticalesorten für extreme Umwelten – eine Frage des Sortentyps?" project (Grant ID: 2814502410) and by BMBF within the HYWHEAT project (Grant ID: FKZ0315945D).

Conflict of interest The authors declare that they have no conflict of interest.

Appendix

For a given parameter (mean yield or yield stability) in a given series, the estimated values from one test set may be denoted as S_{test}, the estimated values from the corresponding reference set as S_{ref}, and the true values as S_{true}.

We know that REML is asymptotically unbiased (it is a consistent estimator). Therefore, we can use the approximate model

$$\begin{cases} S_{\text{test}} = S_{\text{true}} + e_1 \\ S_{\text{ref}} = S_{\text{true}} + e_2 \end{cases} \tag{4}$$

where e_1 and e_2 are independent errors associated with S_{test} and S_{ref}, respectively. For a given set of environments, the errors for different genotypes are not independent, nor are they homoscedastic, when the stability variance is considered. Working with the exact joint distribution of the stability variance estimates is intractable, however, so to obtain an approximate result, we make the simplifying assumption that errors are identically and independently distributed with zero mean.

Under these simplifying assumptions, we can define heritability for a given test set as $h^2 = \text{var}(S_{\text{true}})/\text{var}(S_{\text{test}})$, where var() denotes the variance. We can exploit the fact that under the model (4), we have

$$\text{cov}(S_{\text{test}}, S_{\text{true}}) = \text{cov}(S_{\text{ref}}, S_{\text{true}}) = \text{cov}(S_{\text{test}}, S_{\text{ref}}) = \text{var}(S_{\text{true}}), \tag{5}$$

where cov() denotes the covariance, from which we find that

$$h^2 = \text{cov}(S_{\text{test}}, S_{\text{ref}})/\text{var}(S_{\text{test}}), \tag{6}$$

suggesting that h^2 can be estimated by a regression of S_{ref} on S_{test}.

We now derive a second estimator of h^2 that is a function of correlations only. Heritability may be defined as

$$h^2 = [\text{cor}(S_{\text{test}}, S_{\text{true}})]^2 = \text{var}(S_{\text{true}})/\text{var}(S_{\text{test}}), \tag{7}$$

where cor() denotes Pearson's correlation coefficient, which under our assumed model is equivalent to (6). Now consider the squared correlation between S_{ref} and S_{test}. Using (5), we readily find that

$$[\text{cor}(S_{\text{test}}, S_{\text{ref}})]^2 = [\text{var}(S_{\text{true}})]^2/[\text{var}(S_{\text{test}}) \cdot \text{var}(S_{\text{ref}})]. \tag{8}$$

We will now seek a correction that converts (8) into h^2. To this end, we consider the case when S_{test} is based on the maximum number of test environments, i.e. the same number of test environments as in the reference set (or one less). In this case, the estimators in the test set will be denoted as $S_{\text{test*}}$. Because of the equality of sample sizes between reference set and maximum test set, we have $\text{var}(S_{\text{test*}}) = \text{var}(S_{\text{ref}})$, from which we find that

$$cor(S_{\text{test*}}, S_{\text{ref}}) = \text{var}(S_{\text{true}})/\text{var}(S_{\text{ref}}). \tag{9}$$

Now dividing the left-hand and right-hand sides of Eq. (8) by those of Eq. (9), respectively, we obtain

$$[\text{cor}(S_{\text{test}}, S_{\text{ref}})]^2/\text{cor}(S_{\text{test*}}, S_{\text{ref}}) = \text{var}(S_{\text{true}})/\text{var}(S_{\text{test}}) = h^2. \tag{10}$$

This shows that under model (4), h^2 can be defined as a function of the two Pearson correlations cor(S_{test}, S_{ref}) and cor($S_{\text{test*}}$, S_{ref}), both of which are estimable from the data. To robustify the estimate, we may replace the Pearson correlation with Spearman's rank correlation.

References

Aufhammer W, Kübler E (1987) Zur Leistungsfähigkeit von Gerste in Abhängigkeit von Form und Sorte sowie von Standort und

Produktionstechnik. I. Die Ertragspotenz verschiedener Gerstenformen und deren Nutzung unter differenzierten Aufwuchsbedingungen. Bodenkultur 38:119–133 (in German)

Becker HC (1987) Zur Heritabilität statistischer Maßzahlen für die Ertragssicherheit. Vortr Pflanzenzüchtg 12:134–144 (in German)

Becker HC, Leon J (1988) Stability analysis in plant breeding. Plant Breed 101:1–23

Ceccarelli S, Acevedo E, Grando S (1991) Breeding for yield stability in unpredictable environments: single traits, interaction between traits, and architecture of genotypes. Euphytica 56:169–185

Duarte JB, Zimmermann MJDO (1995) Correlation among yield stability parameters in common bean. Crop Sci 35:905–912

Eagles HA, Frey KJ (1977) Repeatability of the stability-variance parameter in oats. Crop Sci 17:253–256

Eberhart ST, Russell WA (1966) Stability parameters for comparing varieties. Crop Sci 6:36–40

Finlay KW, Wilkinson GN (1963) The analysis of adaptation in a plant-breeding programme. Crop Pasture Sci 14:742–754

Jalaluddin MD, Harrison SA (1993) Repeatability of stability estimators for grain yield in wheat. Crop Sci 33:720–725

Kramer CY (1956) Extension of multiple range tests to group means with unequal numbers of replications. Biometrics 12:307–310

Kumar S, Singh O, Rheenen HV, Rao KVS (1998) Repeatability of different stability parameters for grain yield in chickpea. Plant Breed 117:143–146

Le Gouis J (1992) A comparison between two- and six-row winter barley genotypes for above-ground dry matter production and distribution. Agron Sustain Dev 12:163–171

Le Gouis J, Delebarre O, Beghin D, Heumez E, Pluchard P (1999) Nitrogen uptake and utilisation efficiency of two-row and six-row winter barley cultivars grown at two *N* levels. Eur J Agron 10:73–79

Leon J, Becker HC (1988) Repeatability of some statistical measures of phenotypic stability—correlations between single year results and multi years results. Plant Breed 100:137–142

Lin CS, Binns MR (1991) Genetic properties of four types of stability parameter. Theor Appl Genet 82:505–509

Lin CS, Binns MR, Lefkovitch LP (1986) Stability analysis: where do we stand? Crop Sci 26:894–900

Maidl FX, Panse A, Dennert J, Ruser R, Fischbeck G (1996) Effect of varied *N* rates and *N* timings on yield, *N* uptake and fertilizer *N* use efficiency of a six-row and a two-row winter barley. Eur J Agron 5:247–257

Mekbib F (2003) Yield stability in common bean (*Phaseolus vulgaris* L.) genotypes. Euphytica 130:147–153

Mühleisen J, Maurer HP, Stiewe G, Bury P, Reif JC (2013) Hybrid breeding in barley. Crop Sci 53:819–824

Mühleisen J, Piepho HP, Maurer HP, Longin CFH, Reif JC (2014) Yield stability of hybrids versus lines in wheat, barley and triticale. Theor Appl Genet 127:309–316

Pham HN, Kang MS (1988) Interrelationships among and repeatability of several stability statistics estimated from international maize trials. Crop Sci 28:925–928

Piepho HP (1992) Vergleichende Untersuchungen der statistischen Eigenschaften verschiedener Stabilitätsmaße mit Anwendungen auf Hafer, Winterraps, Ackerbohnen sowie Futter- und Zuckerrüben. Doctoral thesis (unpublished), Kiel

Piepho HP (1998) Methods for comparing the yield stability of cropping systems. J Agron Crop Sci 180:193–213

Piepho HP (1999) Stability analysis using the SAS system. Agron J 91:154–160

Piepho HP, Möhring J (2006) Selection in cultivar trials—is it ignorable? Crop Sci 46:192–201

Piepho HP, Möhring J (2007) Computing heritability and selection response from unbalanced plant breeding trials. Genetics 177:1881–1888

Robert N (2002) Comparison of stability statistics for yield and quality traits in bread wheat. Euphytica 128:333–341

SAS Institute Inc. (2011) SAS/STAT® 9.3 user's guide. SAS Institute Inc., Cary

Shukla GK (1972) Some statistical aspects of partitioning genotype–environmental components of variability. Heredity 29:237–245

Sneller CH, Kilgore-Norquest L, Dombek D (1997) Repeatability of yield stability statistics in soybean. Crop Sci 37:383–390