Euphytica (2012) 186:489–500
DOI 10.1007/s10681-012-0646-8

Effect of region on the uncertainty in crop variety trial programs with a reduced number of trials

Johannes Forkman · Saeid Amiri · Dietrich von Rosen

Received: 5 October 2010 / Accepted: 16 February 2012 / Published online: 26 February 2012
© Springer Science+Business Media B.V. 2012

Abstract Results from crop variety trials may vary between geographical regions because of differences in climate and soil types. Results are usually presented at regional level. To evaluate the importance of the regions used in the Swedish variety trial programs, we examined which regions produced similar levels of yield and similar ratios in yield between cultivars; the amount by which variance could be reduced by division into regions or clusters of regions; and the amount of trials per region and year, replicates per trial, and trials per year required in order to fulfill specifications on the precision of results. Yield data from spring barley and winter wheat trials performed during 1997–2006 were studied using cluster analysis and variance component estimation. The objectives were (1) to discuss the effects of regions on precision when the number of trials has decreased; (2) to demonstrate the method; and (3) to report the results obtained. In spring barley, clusters of regions produced different levels of yield, but similar yield ratios between cultivars. In winter wheat, clusters of regions giving different yield ratios were identified. When the option of a single analysis was compared with that of region-wise analysis, the reduction in variance with the former, due to the larger number of trials, outweighed the reduction in variance with the latter due to decreased random interaction between trials and cultivars.

Keywords Cluster analysis · Crop variety evaluation program · Multi-environment trial · Replication · Sample size calculation · Soil-climate area · Variety trial

S. Amiri contributed to this article when employed at the Swedish University of Agricultural Sciences, Department of Energy and Technology.

J. Forkman (✉)
Department of Crop Production Ecology, Swedish University of Agricultural Sciences, P.O. Box 7043, 750 07 Uppsala, Sweden
e-mail: johannes.forkman@slu.se

S. Amiri
Department of Epidemiology and Public Health, University of Miami, Miami, FL 33136, USA

D. von Rosen
Department of Energy and Technology, Swedish University of Agricultural Sciences, P.O. Box 7032, 750 07 Uppsala, Sweden

Introduction

As is well known, results from crop variety trials may vary between geographical regions. Different cultivars may prefer different soil types and climates. For this reason, results of crop variety evaluation programs are often presented by region. Through the use of regions, local recommendations can be given. This paper discusses another aspect of the use of regions, one that is often overlooked: the increase in uncertainty in

comparisons of cultivars when the total number of trials is split into several regions.

In southern and central Sweden seven agricultural soil-climate regions are used, denoted A–G (Fig. 1). During recent decades, the number of variety trials in Sweden has gradually been decreased for financial reasons. Nowadays, for some crops, only a few trials are performed per year in each region (Table 1). Consequently, the precision in the results reported at regional level is often low, and the estimates can be misleading. This problem could possibly be solved by merging some of the present regions into larger regions. This would result in a larger number of trials per region, without needing to increase the total number of trials. However, it is necessary to know which regions produce similar results and could be merged. It is also essential to study the effects on variation of using larger regions instead of smaller, because genotype-by-environment interactions may increase within the pooled regions. Finally, sample size calculations should be carried out in order to determine the best options regarding numbers of regions and trials per region.

The seven regions A–G originate from a division of Sweden into 'natural' agricultural zones (Höijer 1921). This division considered differences in soil types, topography and climate. The zones were formed within the borders of the administrative provinces. Although the regions have been revised many times over the years (Larsson 2006), to our knowledge they have never been evaluated statistically. In view of the diminishing number of trials performed yearly, the appropriateness of the regions, and their value, has to be questioned.

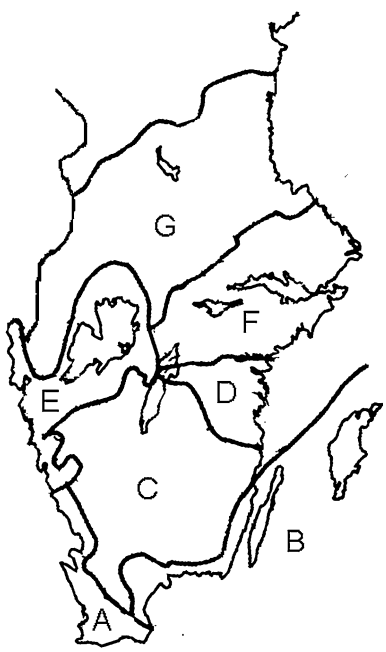

Fig. 1 Swedish agricultural regions

In brief, this study investigated the following questions:

1. Which of the seven regions A–G produce (1) similar levels of yield and (2) similar ratios in yield between cultivars?
2. How much is the standard deviation reduced by a division into regions or clusters of regions?
3. How many trials are needed per region and year, how many replicates per trial, and how many trials per year, in order to fulfill particular specifications on precision in the results?

The study retrospectively examined official trials in spring barley (*Hordeum vulgare* L.) and winter wheat (*Triticum aestivum* L.) performed in Sweden during the 10-year period 1997–2006.

The regions were evaluated through cluster analysis (e.g., Gordon 1999) and sample size calculations. Population models were used; that is the trials were regarded as samples from populations defined by the regions. The trials (i.e., the samples) provide information about the target region (i.e., the population). The sample size was considered in the evaluation because of its importance for inference about the target region.

Similarities between regions are often investigated through singular value decomposition and biplot techniques (e.g., Rodriguez et al. 2008). Yan and Tinker (2006) reviewed biplot analysis for multi-environmental variety trials. Yan and Holland (2009) contributed an adjusted biplot for evaluation of environments. Although biplot analysis is extensively used, it has also been criticized: Yang et al. (2009) questioned the adequacy of the analysis and whether the biplot can be more than a simple descriptive technique.

Biplot analysis requires a complete genotype-environment table. Cultivars that have not been trialed in all regions cannot be included in a biplot analysis. This may not be a problem in a planned series of variety trials for the investigation of differences

Table 1 Number of variety trials performed in the Swedish regions A–G in 1970, 1980, 1990, 2001 and 2009

Year	Winter wheat	Spring barley	Spring wheat	Oats	Peas	Winter rape
1970	70	154	44	87	15	38
1980	96	204	36	121	18	48
1990	55	130	25	94	49	28
2001	51	63	20	39	18	23
2009	37	39	14	22	9	18

between environments, but in a retrospective study such as the present it is a major obstacle. The data sets of the study were highly unbalanced with regard to regions, years, and cultivars. This is often the case in variety testing, because every year new cultivars enter the market and are included in the trials, and older, less well performing cultivars are phased out. In addition, some regions are considered more important for spring barley or winter wheat cultivation than others, and most trials are carried out in those regions. Using biplot analysis, 93% of the spring barley observations and 83% of the winter wheat observations would have been discarded because of missing values. For this reason, we opted for direct cluster analyses on yield and log ratios in yield. This method allowed all available data to be taken into account. Illustrating differences between regions does not suffice. The potential number of trials must be taken into account. Variance components estimates were used to solve for numbers of years, trials, and replicates. While this is not a new concept, it is, however, not widely practiced.

Roßberg et al. (2007) and Graf et al. (2007) used cluster analysis for the definition of soil-climate areas in Germany. Municipalities with similar soils, temperatures, and precipitation were clustered into regions that were assumed to be homogeneous with regard to crop production. These regions comprised a basis for defining agricultural zones for a regionalized field trial system including variety trials. Different zones can be expected to produce different levels of yield, but similarities in yield ratios remain to be investigated.

The Euclidean distance in the cultivar space can be used as a measure of the discrepancy between two environments (e.g., Thomason and Phillips 2006). Cluster analyses have been applied by Collaku et al. (2002) for grouping of locations for winter wheat trials, and by Gutiérrez et al. (2009) for clustering of breeding programs in barley. Hageman et al. (2012) showed how to use two-mode clustering of data arranged in genotype–environment matrices. Estimates of variance components are central for the design of series of variety trials. Cullis et al. (2000) and Laidig et al. (2008) reported estimated genotypic, environmental, and genotype-by-environment interaction variability for common crops in Australia and Germany, respectively.

Swallow and Wehner (1989) and Brennan et al. (1998) considered optimal designs for series of breeding trials, taking financial costs into account. Cullis et al. (2000) discussed the accuracy of alternative schemes, with different numbers of replicates, locations, and years. These articles note the importance of number of years. Conaghan et al. (2008) estimated the genotype-by-environment interaction variability in perennial ryegrass sward plots and calculated optimal allocation of replicates, locations, and years based on the least significant difference. Atlin et al. (2000) discussed effects of regions on precision from a breeder perspective and used variance component estimates for comparison of heritability in undivided target regions with heritability in subregions.

Materials and methods

Data sets

All trials were carried out during the 10-year period 1997–2006. The spring barley data set comprised 7,582 yields from 480 trials, with a total of 250 cultivars, while the winter wheat data set comprised 7,041 yields from 402 trials, with a total of 218 cultivars. Maximal doses of fungicides were used in the trials, primarily as protection against powdery mildew. The trials were designed as complete or incomplete randomized block experiments, usually

with two replicates. The data sets consisted of means estimated in each single trial. The test locations varied between years. Many of the cultivars included in the data sets have now disappeared from the market, but they still convey information about regions and sources of variation between and within regions (variance components).

Cluster analysis

Swedish regions A–G were clustered according to similarities in yield and in log ratios in yield, respectively. Log ratios were analyzed instead of ratios, because the size of a log ratio is not dependent on which cultivar is in the numerator and which is in the denominator. As discussed by Cole (2003), the log scale is the natural scale on which to express percentage differences. Regions that are similar with regard to log ratios are also similar with regard to ratios.

In the cluster analyses based on yield, each combination of year and cultivar formed a variable. Thus, the yield of a specific cultivar observed in 1 year was considered as one variable, and the yield of the same cultivar observed in another year was considered as another variable. The spring barley data set included 615 combinations of years and cultivars. The cluster analysis of spring barley was consequently performed on a data set of seven objects (A–G) and $N = 615$ variables. For each combination of object and variable, the average yield over locations was calculated (when observations were not missing) and used in the cluster analysis. The winter wheat data set included $N = 516$ combinations of years and cultivars, and accordingly the cluster analysis of winter wheat was made on a data set of seven objects (A–G) and 516 variables. For the cluster analyses based on log ratio, in each trial all pairwise differences in yield between cultivars were calculated. In a trial including v cultivars, there are $v(v - 1)/2$ pairwise differences in yield. The objects were classified according to similarity in pairwise differences in log yield, as calculated by year. The pairwise differences were averaged over locations using arithmetic means. In the spring barley data set there were $N = 8{,}684$ combinations of years and pairs of cultivars, and in the winter wheat data set there were $N = 8{,}386$ combinations. The distances between the regions were accordingly measured in 8,684-dimensional and 8,386-dimensional spaces, respectively.

The distance between two clusters was measured as the Euclidean distance between their centroids or means, that is by using the centroid method (e.g., Everitt et al. 2011). The distances were adjusted for missing values. Let y_{ij} denote the observed average yield or average log ratio for variable i in region j. Let N denote the number of variables and let $M(p, q)$ denote the total number of observed pairs (y_{ip}, y_{iq}) of average yields or average log ratios from regions p and q. Because of missing values, $M(p, q) < N$. The adjusted Euclidean distance $d(p, q)$ between regions p and q is defined as

$$d(p,q) = \sqrt{\frac{N}{M(p,q)} \sum_{i=1}^{M(p,q)} (y_{ip} - y_{iq})^2}.$$

The clustering process, as commonly applied, is described by a dendrogram (Everitt et al. 2011). The standardized distance between two clusters of regions is given on the y-axis. This is the adjusted Euclidean distance between the centers of the clusters divided by the square root of the average of all pairwise squared adjusted Euclidean distances between regions. In the Euclidean space, the distance between two objects, say A and B, can be larger than the distance between the centroid of those objects (i.e., the centroid of Cluster AB) and a third object or centroid of objects, C, although the distance between A and B is smaller than the distances between A and C and between B and C. In this case, A and B are merged at a larger distance than the distance at which Cluster AB is merged with C. This results in crossed lines in the dendrogram. For example, in Fig. 4, the (standardized Euclidean) distance between Regions A and B is approx. 0.81. This distance is larger than the distance (approx. 0.80) between the centroids of Clusters AB and DF. The distances between A and DF and between B and DF are larger than the distance between A and B. Thus, Cluster AB is established, but not Clusters ADF or BDF, and Cluster AB is merged with Cluster DF at a smaller distance than Region A is merged with Region B.

The calculations were performed using the DISTANCE, CLUSTER and TREE procedures in SAS 9.1.3 (SAS Institute 2004).

Sample size calculation

The yields from a series of n trials including v cultivars, performed during the same year, can be

described by a statistical model with two factors: trial and cultivar. In the following, the yields are assumed to be normally distributed and the significance level is 5%. The difference between two cultivars can be compared with the least significant difference (LSD), denoted L. Let s^2 be the error variance, and t the 97.5th percentile of a t distribution with $(n-1)(v-1)$ degrees of freedom. Then LSD is

$$L = t\sqrt{\frac{2s^2}{n}}. \tag{1}$$

When the number of degrees of freedom is large $t \approx 2$, and the number of trials can be written as a function of LSD:

$$n \approx \frac{8s^2}{L^2} \tag{2}$$

Moreover, when the expected difference between two specific cultivars is Δ, then, due to the symmetric distribution of the yields, the probability of a significant result is 50% when LSD is Δ. In other words, the power of the test is only 50%. As a rule of thumb, in order to obtain a higher power, we may require the standard error in the difference between two cultivars to be smaller than $\Delta/3$, where Δ is the true difference between the cultivars (Mead 1988, p. 126). Through this rule,

$$n = \frac{18s^2}{\Delta^2} \tag{3}$$

trials are needed. This rule of thumb gives a power of approx. 80%. Equations 2 and 3 imply $\Delta = 1.5\ L$.

In a series of trials performed over several years, the differences between the cultivars usually vary between years. This can be modeled as a random interaction between cultivars and years. The statistical model then includes three factors: cultivar, year, and trial, and also the cultivar-by-year interaction. If the series includes v cultivars investigated during a years with n trials per year, the variance for the cultivar-by-year interaction is s_A^2, and the residual variance is s^2, then the LSD is

$$L = t\sqrt{2\left(\frac{s_A^2}{a} + \frac{s^2}{a \cdot n}\right)}, \tag{4}$$

where t is now the 97.5th percentile of a t distribution with $(a-1)(v-1)$ degrees of freedom.

Each trial includes several replicates. Let r denote the number of replicates (in a complete block design r is the number of blocks), and let s_E^2 denote the residual error variance. Statistical analyses of series of variety trials are often performed on the averages estimated in the single trials. These averages have variance s_E/r^2. As a result,

$$s^2 = s_B^2 + \frac{s_E^2}{r}. \tag{5}$$

In Eq. 5, s_B^2 is the random interaction between cultivars and trials within years. Together, Eq. 4 and 5 give the following equation for the LSD as a function of the number of years (a), trials per year (n), and replicates per trial (r):

$$L = t\sqrt{2\left(\frac{s_A^2}{a} + \frac{s_B^2}{a \cdot n} + \frac{s_E^2}{a \cdot n \cdot r}\right)}, \tag{6}$$

where t is the 97.5th percentile of a t distribution with $(a-1)\ (v-1)$ degrees of freedom. Using Eq. 6, we generalize to other locations and years than those observed. When all trials are carried out in the same year, Eqs. 1 and 5 give:

$$L = t\sqrt{2\left(\frac{s_B^2}{n} + \frac{s_E^2}{n \cdot r}\right)}, \tag{7}$$

where t is the 97.5th percentile of a t distribution with $(n-1)(v-1)$ degrees of freedom. Using Eq. 7, we generalize to other locations than those observed, but not to other years.

Estimation of variance components

Several different groupings of geographical regions were formed, based on the results of the cluster analyses. Tables 2 and 3 present the alternatives formed for spring barley and winter wheat, respectively. For each alternative, year, and group of regions, the mean squared error was estimated in a two-way analysis of variance of log yield (g m^{-2}), with cultivar and trial as explanatory factors. The mean squared errors were averaged over years and groups of regions, thus producing one average per alternative. The square roots of these averages were calculated and reported in Tables 2 and 3 as estimated standard deviations, s. The calculations were made in SAS 9.2 using the GLM procedure (SAS Institute 2004).

The variance estimates in Eq. 4 were calculated, on the logarithmic scale, through fitting a linear model with

Table 2 Spring barley: alternative clusters of regions: 0, I,…, VI

N	0	I	II	III	IV	V	VI
121	A	A	A	A	AB		
84	B	B	B	B		ABDF	
51	D	D	D	DF	DF		
121	F	F	F				ABCDEFG
51	E	EG					
26	C	C	CEG	CEG	CEG	CEG	
26	G	EG					
s	0.063	0.062	0.060	0.060	0.060	0.063	0.062

Number of trials (*N*) per region. Estimated error standard deviation (*s*) in log yield (g m^{-2}), per cluster, calculated as described in "Estimation of variance components" section

Table 3 Winter wheat: alternative clusters of regions: 0, I,…, VII

N	0	I	II	III	IV	V	VI	VII
112	A	A	AB	AB	AB	AB		
56	B	B						
44	D	D	D	D			ABCDE	
10	C	CE	C	CE	CDEF	CDE		ABCDEFG
75	E		E					
99	F	F	F	F		FG	FG	
6	G	G	G	G	G			
s	0.098	0.098	0.104	0.105	0.108	0.108	0.115	0.117

Number of trials (*N*) per region. Estimated error standard deviation (*s*) in log yield (g m^{-2}), per cluster, calculated as described in "Estimation of variance components" section

three factors: cultivar, year, and trial. Cultivar was modeled as a fixed factor. Year, cultivar-by-year interaction, and trial were modeled as random independent normally distributed factors with expected value 0. On the logarithmic scale, standard deviations are approximately equal to coefficients of variation on the original scale. For an introduction to linear models with random effects, see Galwey (2006). The model was fitted in SAS 9.2, using the REML method (see e.g., Robinson 1987) and the MIXED procedure (SAS Institute 2004).

In the data sets there was no information about the variance s_E^2 within the trials, because the data sets only included estimated averages. An additional extraction from the same data base, comprising recorded coefficients of variations, was made from trials performed during the years studied. The square root of the mean squared coefficients of variation was calculated, as suggested by Forkman (2009), for spring barley and winter wheat, respectively. These means were used as estimates of the residual standard deviations, s_E, on the logarithmic scale.

Results

Similarities and dissimilarities of the regions are investigated by "Cluster analyses" section. The regions that produce similar yield ratios are identified. Regions producing similar yield ratios could possibly be merged into larger clusters of regions. The effects of divisions into regions are presented in "Variance reduction" section. Different groupings of regions, proposed in "Cluster analyses" section are compared with regard to reduction in error variance. A successful division into regions reduces the error variance. In statistical analyses performed per region, the average variance is expected to be smaller than the variance in a single statistical analysis performed on the complete data set of trials from all regions. In "Sample size calculation" section, the calculations are performed.

Cluster analyses

For spring barley, the differences between the regions were similar with regard to log ratios (Fig. 2). In other

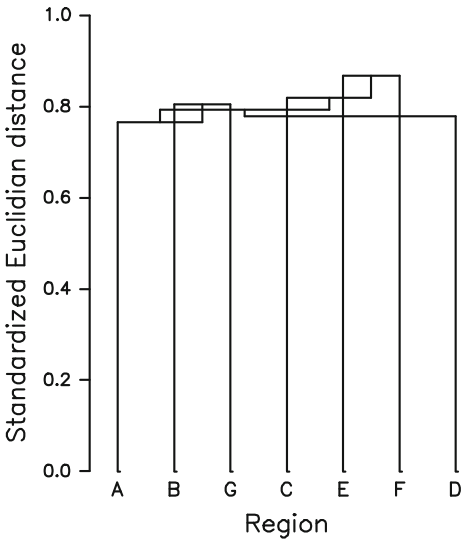

Fig. 2 Spring barley: clustering of Swedish agricultural regions based on similarities in yield log ratios. *Horizontal lines* indicate distances between regions or clusters of regions (distances within regions are not illustrated)

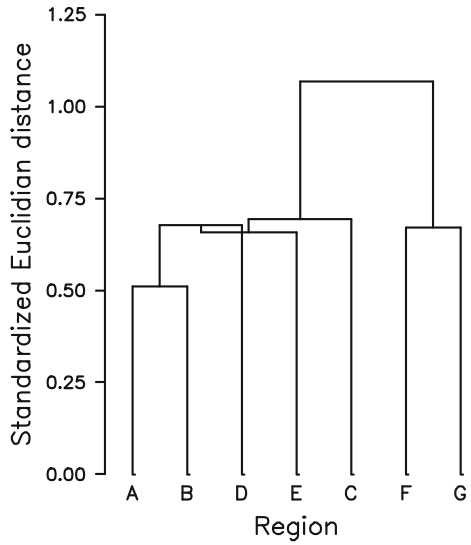

Fig. 3 Winter wheat: clustering of Swedish agricultural regions based on similarities in yield log ratios. *Horizontal lines* indicate distances between regions or clusters of regions (distances within regions are not illustrated)

words, there was no apparent grouping of regions into clusters of regions producing similar ratios between spring barley cultivars. Regions A and B produced similar ratios between winter wheat cultivars (Fig. 3). In addition, the cluster analyses on log ratios suggest clustering of Regions F and G (Fig. 3).

Cluster analyses on yields were made for comparison with the cluster analyses performed on log ratios. The cluster analyses of spring barley indicated similar levels of yield in Regions C, E, and G (Fig. 4). Two groups of regions can possibly be formed: one with Regions {A, B, D, F} and the other with Regions {C, E, G}. The yield should be homogeneous within these two groups of regions. This is not surprising, because Regions A, B, D and F are lowlands producing high yields, whereas Regions C and G are highlands. Region E, west of Lake Vättern, is known to give lower yields than Region D, east of Lake Vättern. Regions C and E give the most similar levels of winter wheat yield, although Regions A and B also produce similar levels (Fig. 5).

Variance reduction

Based on the results of the cluster analyses performed on the spring barley data set, seven different groupings of the regions were further investigated. These are presented in Table 2. Alternative 0 represents the original regions. The average variance in statistical analyses performed per region (Alternative 0) was expected to be smaller than that in an analysis made on the complete data set without division into regions (Alternative VI). Yet the division into regions resulted in approximately the same (the estimate was slightly larger) standard deviation: 0.063 compared with 0.062. Nothing was gained in precision by making one statistical analysis per region compared with making one single statistical analysis. The differences between the standard deviations of the various alternatives are small. This result confirms that the use of regions or clusters of regions, based on similarities in levels of yield, may not reduce the variance in the estimates of the differences between the cultivars.

For winter wheat, as defined in Table 3, eight different groupings were investigated. Alternative 0 is the same for winter wheat as for spring barley (the present regions), and Alternative VII for winter wheat is the same as Alternative VI for spring barley (no division into regions). Table 3 includes the estimated standard deviations for the different groupings. Without any division, i.e., if one single analysis is carried out including all trials from all regions, the estimated standard deviation is 0.117. If instead one analysis is performed per region, the estimated standard deviation is 0.098. Table 3 indicates that Region C can be

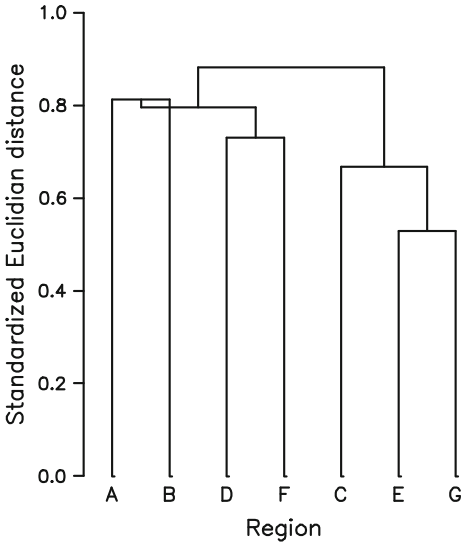

Fig. 4 Spring barley: clustering of Swedish agricultural regions based on similarities in yield levels. *Horizontal lines* indicate distances between regions or clusters of regions (distances within regions are not illustrated)

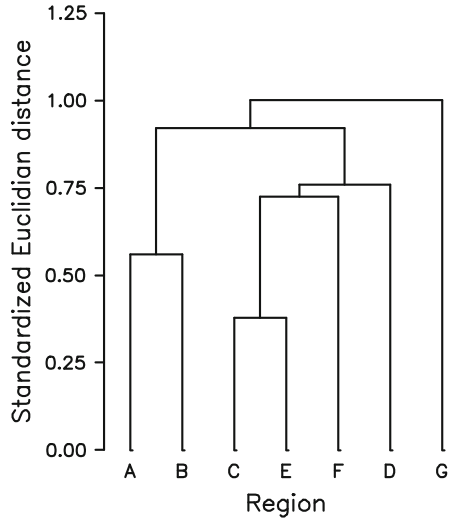

Fig. 5 Winter wheat: clustering of Swedish agricultural regions based on similarities in yield levels. *Horizontal lines* indicate distances between regions or clusters of regions (distances within regions are not illustrated)

merged with Region E (Alternative I) without increasing the variance in the results. A merge of Regions A and B (Alternatives II and III) increases the variance relative to Alternatives 0 and I. The alternatives IV and V, which comprise three groups, and Alternative VI, which comprises only two groups, increase the variance further, although they result in smaller variance than Alternative VII (no division into groups).

Sample size calculations

Sample size calculations were carried out based on the results of "Variance reduction" section. The number of trials necessary is dependent on the required precision. As an example, it was assumed that LSD should be less than 0.040 on the logarithmic scale, given the estimated variance components. This corresponds to approximately 4% difference on the original scale (Cole 2003). We also assumed that the number of cultivars, v, is large, so that we could use $t = 2$ in Eqs. 4, 6 and 7.

The multiplicity due to regions

In spring barley, without any division into regions the standard deviation was estimated to be $s = 0.062$ (Cluster VI in Table 2). Using the present division into seven regions (Alternative 0 in Table 2), the standard deviation is approximately the same: $s = 0.063$. According to Eq. 2, 20 trials are required to obtain LSD < 0.040. Then, according to Eq. 3, there is a fair chance of a significant result when the true difference is $1.5 \times 0.040 = 0.060$. If seven regions are used, altogether $7 \times 20 = 140$ trials are required.

For winter wheat, the standard deviation was estimated to be $s = 0.117$ without division into regions, and $s = 0.098$ with the present division into seven regions (Table 3). According to Eq. 2, 69 trials are needed when $s = 0.117$, and 48 when $s = 0.098$. Thus, without division into regions (i.e., when $s = 0.117$), 69 trials are required, but with a division into seven regions (i.e., when $s = 0.098$), a total of $7 \times 48 = 336$ trials is required to meet the specification in all regions. The decrease in variance due to the division into regions does not outweigh the introduced multiplicity (i.e., the factor 7). Alternative V, with three clusters: AB, CDE, and FG, produced a comparatively low standard deviation: 0.108 (Table 3). With this alternative, 59 trials are required per cluster, i.e., 177 trials in total.

Series of experiments

We now assume that there is no division into regions, and we calculate the required sample sizes for

Table 4 Estimated standard deviations s_A, s_B and s_E, for the cultivar-by-year interaction, the cultivar-by-trial interaction (within years) and the residual error (within trials), respectively, on the logarithmic scale (log g m^{-2})

	Spring barley	Winter wheat
s_A	0.029	0.037
s_B	0.050	0.110
s_E	0.056	0.062

Table 5 Number of years (a), number of trials per year (n) and number of replicates per trial (r) required to fulfill the specification LSD < 0.040 on the logarithmic scale, in multiple-year series ($a > 1$) and single-year series ($a = 1$)

	Spring barley				Winter wheat	
Multiple-year series						
a	5	5	6	6	8	8
n	26	21	12	10	61	57
r	2	4	2	4	2	4
Single-year series						
n	21		17		71	66
r	2		4		2	4

multiple-year series and single-year series. Table 5 presents the results of these calculations. In the mixed model analysis of spring barley, the standard deviation for the cultivar-by-year interaction was estimated to be $s_A = 0.029$ (Table 4), and the standard deviation for the residual error $s = 0.064$, which corresponds well to the estimate $s = 0.062$ for Cluster VI in Table 2. The CV within spring barley trials was estimated to be 5.6%, which corresponds to $s_E \approx 0.056$ on the logarithmic scale (Table 4) In Swedish variety trials, the number of blocks, r, is usually 2. Equation 5 gives $s_B^2 = 0.064^2 - 0.056^2/2 = 0.050^2$ (Table 4). According to Eq. 6, in a 5-year series, approximately 26 trials are needed yearly to obtain LSD = 0.040, provided each trial includes two replicates (Table 5). If, instead, each trial comprises four replicates, 21 trials per year should suffice to give LSD < 0.040. Similarly, in a 6-year series, 12 trials with two replicates, or 10 trials with four replicates, are needed yearly. In spring barley series with trials from a single year, according to Eq. 7, a total of 21 trials are needed to obtain LSD < 0.040, given two replicates per trial (Table 5). Note that in this situation the conclusions are limited to the specific year. If the trials included four replicates, then 17 trials would suffice.

The standard deviation for the interaction between years and cultivars was estimated to be $s_A = 0.037$ (Table 4), and the residual standard deviation $s = 0.118$. In this analysis, regions were not used, and the estimate $s = 0.118$ was consequently close to the estimate $s = 0.117$ for Alternative VII in Table 3. The standard deviation within winter wheat trials was estimated, through observed coefficients of variations, to be $s_E = 0.062$ (Table 4). According to Eq. 5, $s_B^2 = 0.118^2 - 0.062^2/2 = 0.110^2$, with $r = 2$ (Table 4). In an 8-year series, 61 trials with two replicates per trial or 57 trials with four replicates per trial are needed yearly, based on Eq. 6 (Table 5). When all trials in the series are performed in the same year and the statistical inference aims at that particular year, according to Eq. 7, 71 trials with two replicates are required (Table 5). This result agrees approximately with the previous conclusion using Eq. 2: 69 trials, but it is slightly larger because of the higher estimate of the residual standard deviation. With four replicates, 66 trials are required.

Discussion

For financial reasons, nowadays fewer variety trials are performed than before. As a consequence, we cannot estimate differences between varieties in subdivided target regions without increased random errors in the results. Fortunately, as this study shows, the differences in yield ratios between regions may not be large. As is well known, yield varies between regions. This study confirmed the existence of persistent similarities and differences in yield between regions. However, variety trials do not aim at estimating absolute levels of yield. Rather, the objective is determination of differences or ratios in yield between cultivars.

Regions differ less in ratios than in absolute values, especially in spring barley. For example, in spring barley, there are differences in yield between the clusters {C, E, G} and {A, B, D, F} (Fig. 4). Consequently, Regions C, E, and G produce different levels of yield than Regions A, B, D, and F. The cluster analyses do not reveal which cluster produces more and which produces less. The analyses only provide the information that Regions C, E, and G usually

produce similar yields, and that Regions A, B, D, and F usually produce similar yields. In some years, Regions C, E, and G may give lower yields than Regions A, B, D, and F, but in other years they may give higher yields. Interestingly, the two clusters {C, E, G} and {A, B, D, F} are not distinguishable in log ratio (Fig. 2). There may be differences in log ratios between the regions, but the log ratios between the cultivar yields obtained in Regions C, E, and G do not consistently differ from the log ratios obtained in Regions A, B, D, and F. This conclusion agrees with the results reported by Navabi et al. (2006), who studied grouping of testing sites for spring wheat. They concluded that 'yearly site groupings did not generally follow a repeatable pattern over years', and 'genotype × environment patterns' were 'inconsistent over the years, mainly because of complex, highly variable, and unpredictable year × location effects.'

In winter wheat, Regions A and B give similar levels of yield (Fig. 5) and similar log ratios (Fig. 3). Regions F and G also produce similar ratios between the cultivars (Fig. 3). It is not surprising that differences in ratios between the regions are revealed in winter wheat, which is sown in the autumn, but not in spring barley, which is sown in the spring. There are regional differences in the winter climate, and some cultivars tolerate hard weather conditions better than others. Yan and Hunt (2001) found that cold winter and hot summer temperatures were the major environmental causes of genotype-by-environment interactions in winter wheat.

The random variations were expected to be smaller within regions than in the whole country. This was the case for winter wheat (compare the standard deviation with Alternative 0 with that of Alternative VII in Table 3), but not for spring barley (Table 2). In practice, division into small regions has the serious effect that the required number of trials is multiplied. As noted by Smith et al. (2001) 'If varietal information is to be reported at a regional level the variance component for variety-by-region interaction should be large', which 'is rarely the case'. Michel et al. (2007) argued: 'When estimating yield for a particular region, inclusion of data from other regions, while reducing variance, is a potential source of bias, because rankings in different regions need not be identical. If the bias incurred by neighbouring regions is larger than the reduction in variance, using the neighbouring data would not be beneficial. If bias is more than offset by the reduction in variance, however, using the additional data is worthwhile.' In this study, we did not see any large persistent variety-by-region interactions, especially not for spring barley. In our opinion, when comparing the option of making a single analysis with the option of making region-wise analyses, the reduction in variance due to the larger number of trials in the former outweighs the reduction in variance due to decreased random interaction between trials and cultivars in the latter. Atlin et al. (2000) concluded similarly: 'selection based on means resulting from testing over a wide geographical area will often produce a greater response within local subdivisions of that area than will selection based on within-subregion means only'.

If Sweden is to be divided into a few larger regions for the analysis and presentation of winter wheat trials, we recommend that Regions A and B; Regions C, D, and E; and Regions F and G be merged into three clusters, although this alternative does not have the smallest standard error (Table 3). The present trial resources can scarcely be divided across more than two or three regions without a large reduction in precision per region. We cannot see any need for regions in the analyses and presentations of Swedish spring barley trials.

Piepho and Möhring (2005) discussed breeding for local vs. broad adaptation. When the objective is broad adaptation, the regions are ignored and all data are used, but when the objective is local adaptation, only the data from the targeted region is used. As they remarked, this view is problematic, because 'selecting or recommending for local adaptation requires almost the same amount of resources within each subregion as would be needed for assessing broad adaptation with the same accuracy.' Piepho and Möhring (2005) dealt with this problem by proposing a method for weighted inclusion of results from adjacent regions. Their analysis is careful and theoretically sound. However, their method could be difficult to popularize, because the users of the results may find it difficult to accept the outcomes for a specific region being influenced by trials performed outside that region. Even so, the method proposed by Piepho and Möhring (2005) has been included in the Hohenheim-Gülzow method for analysis of series of variety trials (Michel et al. 2007). We prefer enlarging the target population and letting the trials performed within the large population constitute the complete sample.

The high frequency of missing values may have influenced the conclusions of this study. The cultivars were not missing randomly. If all cultivars had been included in all regions, larger differences between the regions would presumably have been detected. The cultivars were trialed in regions where they were expected to perform well. Nevertheless, a division into regions implies a division of testing resources, which increases the random errors in the results.

Large interactions between cultivars and years were observed in this study. In other words, differences between cultivars changed randomly between years. However, there may be long-term systematic differences between the cultivars. Because of these interactions, many years are needed in order to distinguish the systematic differences. Series of Swedish variety trials often run for 5 years, but according to the results presented here, more than 5 years may be required. Unfortunately this theoretical conclusion is impractical. After more than 5 years, there are hopefully new and better performing cultivars available, at which point there is little interest in the older cultivars. Even so, estimation of long-term differences in a short time is impossible.

Acknowledgments This research was supported by the Swedish Farmers' Foundation for Agricultural Research (SLF). We thank the anonymous reviewers for suggestions that improved the manuscript.

References

Atlin GN, Baker RJ, McRae KB, Lu X (2000) Selection response in subdivided target regions. Crop Sci 40:7–13

Brennan JP, Kingwell RS, Thomson FM, Cullis BR (1998) Optimal strategies for regional cultivar testing. Aust J Agr Resour Econ 42:209–226

Cole TJ (2003) Sympercents: symmetric percentage differences on the 100 $\log_e$ scale simplify the presentation of log transformed data. Stat Med 19:3109–3125

Collaku A, Harrison SA, Finney PL, Van Sanford DA (2002) Clustering of environments of southern soft red winter wheat region for milling and baking quality attributes. Crop Sci 42:58–63

Conaghan P, Casler MD, McGilloway DA, Kiely PO, Dowley LJ (2008) Genotype × environment interactions for herbage yield of perennial ryegrass sward plots in Ireland. Grass Forage Sci 63:107–120

Cullis BR, Smith A, Hunt C, Gilmour A (2000) An examination of the efficiency of Australian crop variety evaluation programmes. J Agric Sci 135:213–222

Everitt BS, Landau S, Leese M, Stahl D (2011) Cluster analysis, 5th edn. Wiley, Chichester

Forkman J (2009) Estimator and tests for common coefficients of variation in normal distributions. Commun Stat Theory Methods 38:233–251

Galwey NW (2006) Introduction to mixed modeling: beyond regression and analysis of variance. Wiley, Chichester

Gordon AD (1999) Classification, 2nd edn. Chapman & Hall/CRC, Boca Raton

Graf R, Michel V, Roßberg D, Neukampf R (2007) Regionalized field trial system by defining crop-specific cultivation areas. In: Piepho H-P, Bleiholder H (eds) Agricultural field trials—today and tomorrow. Proceedings of international symposium, Stuttgart-Hohenheim, Germany, 8–10 Oct. 2007. Verlag Grauer, Stuttgart, pp 44–47

Gutiérrez L, Nason JD, Jannink J-L (2009) Diversity and mega-targets of selection from the characterization of a barley collection. Crop Sci 49:483–497

Hageman JA, Malosetti M, van Eeuwijk FA (2012) Two-mode clustering of genotype by trait and genotype by environment data. Euphytica 183:349–359

Höijer E (1921) Sveriges uppdelning på naturliga jordbruksområden. Nord. bokh., Stockholm (in Swedish)

Laidig F, Drobek T, Meyer U (2008) Genotypic and environmental variability of yield for cultivars from 30 crops in German official variety trials. Plant Breed 127:541–547

Larsson S (2006) Sveriges jordbruksområden. En redovisning av jordbruksområden och växtzoner i svenskt jord- och trädgårdsbruk. Aktuellt från VPE 1, SLU, Uppsala (in Swedish)

Mead R (1988) The design of experiments. Statistical principles of practical application. Cambridge University Press, Cambridge

Michel V, Zenk A, Graf R, Möhring J, Büchse A, Piepho H-P (2007) The "Hohenheim-Gülzow-method" for analysis of series of trials as basic procedure for PIAFStat and SAS in a regionalized field trial system. In: Piepho H-P, Bleiholder H (eds) Agricultural field trials—today and tomorrow. Proceedings of the international symposium, Stuttgart-Hohenheim, Germany. 8–10 Oct. 2007. Verlag Grauer, Stuttgart, pp 136–141

Navabi A, Yang R-C, Helm J, Spaner DM (2006) Can spring wheat-growing megaenvironments in the northern great plains be dissected for representative locations or niche-adapted genotypes? Crop Sci 46:1107–1116

Piepho H-P, Möhring J (2005) Best linear unbiased prediction of cultivar effects for subdivided target regions. Crop Sci 45:1151–1159

Robinson DL (1987) Estimation and use of variance components. The Statistician 36:3–14

Rodriguez M, Rau D, Papa R, Attene G (2008) Genotype by environment interactions in barley (*Hordeum vulgare* L.): different responses of landraces, recombinant inbred lines and varieties to Mediterranean environment. Euphytica 163:231–247

Roßberg D, Graf R, Michel V, Neukampf R (2007) Definition of soil-climate-areas for Germany. Nachrichtenbl Deut Pflanzenschutzd 59:155–161 (In German, with English abstract.)

SAS Institute (2004) SAS/STAT user's guide. SAS Institute Inc., Cary

Smith A, Cullis B, Gilmour A (2001) The analysis of crop variety evaluation data in Australia. Aust N Z J Stat 43:129–145

Swallow WH, Wehner TC (1989) Optimum allocation of plots to years, seasons, and replications, and its application to once-over-harvest cucumber trials. Euphytica 43:59–68

Thomason WE, Phillips SB (2006) Methods to evaluate wheat cultivar testing environments and improve cultivar selection protocols. Field Crop Res 99:87–95

Yan W, Holland JB (2009) A heritability-adjusted GGE biplot for test environment evaluation. Euphytica 171:355–369

Yan W, Hunt LA (2001) Interpretation of genotype × environment interaction for winter wheat yield in Ontario. Crop Sci 41:19–25

Yan W, Tinker NA (2006) Biplot analysis of multi-environmental trial data: principles and applications. Can J Plant Sci 86:623–645

Yang R-C, Crossa J, Cornelius PL, Burgueño J (2009) Biplot analysis of genotype × environment interaction: proceed with caution. Crop Sci 49:1564–1576