Theor Appl Genet (1988) 76:352–360



# Factor and principal component analyses as alternatives to index selection

**E. B. Godshalk * and D. H. Timothy**

Department of Crop Science, North Carolina State University, Raleigh, NC 27695-7620, USA

Received November 23, 1987; Accepted April 11, 1988
Communicated by A. R. Hallauer

**Summary.** Selections from factor and principal component analyses were compared with those from the Smith-Hazel index when selecting for several switchgrass (*Panicum virgatum* L.) traits. The objective of this study was to examine several alternatives to index selection. Such procedures would potentially eliminate problems of selection associated with Smith-Hazel indices, including errors in genetic parameter estimates and difficulty in assigning relative economic weights to traits. Selection was performed on 1,280 plants that were evaluated over 2 years at 1 location, in a randomized complete block design with 4 replicates. The plants were evaluated for forage yield and several forage quality traits. The comparisons of index selection with principal factor analysis, maximum-likelihood factor analysis and principal component analysis were made for three sets of traits (five traits per set) to estimate repeatability for the comparisons. Multivariate analyses were performed on both simple and genotypic correlation matrices. Comparisons were made by computing Spearman's rank correlations between selection index plant scores and scores computed from multivariate analysis and by determining the number of plants selected in common for the selection methods. Among the three multivariate analysis methods evaluated in this study, principal component analysis had the highest correlation with index selection. The high correlation for principal component analysis of simple correlation matrices indicates the potential for using this statistical method for selection purposes. This would permit the breeder to reduce field costs (e.g., time, labor, equipment) required to obtain the genetic parameter estimates necessary to construct selection indices.

* To whom all correspondence should be addressed; current address; Dekalb-Pfizer Genetics, 3100 Sycamore Road, Dekalb, IL 60115, USA

**Key words:** Index selection – Factor analysis – Principal component analysis

## Introduction

Index selection (Smith 1936; Hazel 1943) has been used effectively in selecting for improved genetic worth in animal and plant populations. This selection method consists of maximizing the correlation between the index and the aggregate genotype, and then selecting individuals according to their index scores. Applications of index selection to agronomic crop breeding have been reviewed by Baker (1986). Lin (1978) and Baker (1986) have described several alternatives to the Smith-Hazel index.

There are several limitations of index selection that reduce its effectiveness. Changing genetic parameters over generations, as a result of selection, may require re-estimation to obtain estimates that are relevant to the population being evaluated (Lin 1978). Accuracy of index selection is also dependent on the magnitudes of errors associated with genetic and phenotypic parameter estimates (Brim et al. 1959). Effects of parameter estimate errors on accuracy of index selection are dependent on the number of traits selected, the relative economic weights of those traits, the magnitude of parameter estimates and selection intensity (Lin 1978). Accuracy of index selection in *Tribolium castaneum* was more strongly affected by overestimation than by underestimation of narrow-sense heritability (Lin et al. 1979). To eliminate dependence on parameter estimates, Brim et al. (1959), Williams (1962) and Harris (1964) suggested using relative economic weights instead of index coefficients that are calculated in the

Smith-Hazel index. This base index, as referred to by Williams (1962), was shown by Elgin et al. (1970), Eagles and Frey (1974), and Campo and Rodriguez (1985) to be similar in efficiency to the Smith-Hazel index and generally easier to use than the Smith-Hazel index.

Another area of difficulty presented by the Smith-Hazel index is the assignment of relative economic weights to the traits comprising the index. Relative economic weights of a set of traits are not always obvious to the breeder. Andrus and McGilliard (1975) proposed multiple regression analysis to derive economic weights, where estimates of profit are regressed on phenotypic values. Young (1961) showed that the superiority of index selection to tandem selection and independent culling is greatest when the traits have equal relative importance. He computed relative importance as a product of the economic weight, narrow-sense heritability and phenotypic standard deviation of a given trait. Given estimates of narrow-sense heritability and phenotypic standard deviation, economic weights can be computed so that all traits have equivalent relative importance values. Another potential method of obtaining relative economic weights is from the index in retrospect, as described by Dickerson et al. (1954) and Allaire and Henderson (1966). Van Vleck (1974) presented formulae to compute the aggregate genotype and relative economic weights in retrospect.

Given these limitations of index selection, several alternatives to this selection method have been discussed by Baker (1986). Canonical discriminant analysis was used by Riggs (1973) for selection among barley (*Hordeum vulgare* L.) lines. Riggs (1973) briefly described the matrix manipulation required to perform canonical analysis.

Considering several traits simultaneously requires that the dimensionality of the data set be reduced. Factor analysis and principal component analysis are designed to do this. Factor analysis has been applied successfully to the study of relationships among an array of agronomic traits (Murty and Arunachalan 1967; Hsu and Walton 1971; Walton 1971; Walton 1972; Denis and Adams 1978; Bramel et al. 1984). Unlike principal component analysis, which explains the total variation in the original measures, factor analysis is used to account only for the covariation among variables (Dunteman 1984). Principal component analysis does not have the underlying model that is characteristic of factor analysis. Chatfield and Collins (1980) have discussed several problems inherent in principal component analysis and factor analysis, including the difficulty in interpreting results of analysis, and suggested that factor analysis be avoided in most practical situations. Rummel (1970), on the other hand, applauded factor analysis for its utility in understanding the patterned variation in a set of variables. He labelled factor analysis as "a scientific tool par excellence".

Despite the controversial nature of factor analysis and, to a lesser extent, principal component analysis, the purpose of this study was to evaluate the use of selection based on these multivariate analysis methods. Selection from factor or principal component analysis of the simple correlation matrix would eliminate the need to obtain reliable estimates of the genetic parameters that are required for most methods of index selection. The multivariate analysis selection methods were evaluated by comparison with selection on the Smith-Hazel index.

## Materials and methods

The Cycle 0 population comprised a selected group of 161 switchgrass (*Panicum virgatum* L.) plants of the lowland form representing 11 different germplasm sources. A total of 31 plants were selected among this group on the basis of the arithmetic product of forage yield and in vitro dry matter disappearance (IVDMD). These plants, in addition to two other plants that were selected to maintain their germplasm, were allowed to cross-pollinate at random. Their 660 half-sib progenies (Cycle 1) were evaluated for forage yield, IVDMD and nitrogen (N) concentration (Talbert et al. 1983) and 33 plants were selected for combinations of these traits. Selections were made using the three different indices described by Godshalk et al. (1988). One index produced one 16-clone synthetic, the second index produced one 16-clone synthetic and a 4-clone synthetic, and the third index produced 16-, 8-, and 4-clone synthetics, resulting in a total of 6 synthetics. A given plant (among the 33 plants selected) was frequently present in more than 1 synthetic. All 6 synthetics were planted in crossing blocks and open-pollinated in isolation of each another and other switchgrass. Open-pollinated progeny (Cycle 2) of these 33 plants were evaluated in the current study as half-sib families. As a result of varying representation (among the 6 synthetics) of the 33 plants selected from Cycle 1, family sizes of the progeny ranged from 20–100 members per half-sib family.

A total of 1,280 Cycle 2 plants, representing the 33 half-sib families, were evaluated in 1985 and 1986 at the Central Crops Research Station in Clayton/NC, in a randomized complete block design with 4 replicates. Each row contained seven plants, with data collection restricted to the middle five plants of a given row. Rows and plants within rows were spaced 1.07 m apart. At harvest, approximately five culms per plant were sampled to a 5 cm stubble and immediately immersed in liquid nitrogen to preserve the plant material. Samples were kept frozen (−10 °C), freeze-dried, ground in a Wiley mill to pass through a 1 mm sieve, and stored in the freezer until they were analyzed. The remainder of the plant was subsequently harvested and its weight recorded as forage yield. Yield was recorded on a dry weight basis by saving one plant per row after recording its green weight, drying it in a forced air drier for approximately 72 h, and reweighing the plant. Yields of the remaining four plants were then adjusted on the basis of percent dry matter of the plant representing that particular row. First harvests were taken on 11 June 1985 and 12 June 1986, while regrowth was harvested on July 23, 1985 and August 5, 1986. Plants were rated for maturity at harvest on a scale of 1 to 5, with 1 indicating no heads emerged and 5 corresponding to

100% emergence. All plants were in the boot or early emergence stages of maturity when initial growth and regrowth were harvested.

Near infrared reflectance spectroscopy (McClure and Hamid 1980) was used to measure quality-related traits on the forage samples. Equations presented by deRuiter (1984) were used to predict IVDMD, acid detergent fiber (ADF), neutral detergent insoluble residue monomers [arabinose (ARA), galactose (GAL), glucose (GLU) and xylose (XYL)], total hexose monomers (HEX), total monomers (SUG), monomeric ratios of XYL:GLU and XYL:GAL, pectin-free cellulose (PFC), and trifluoroacetic acid-hydrolyzed monomers (TFAHYD). The ARA:GAL ratio was computed from predictions of the respective monomers, and hemicellulose (HEMI) concentration was calculated as the difference between neutral detergent fiber and ADF.

Analyses of variance and covariance were performed according to a statistical model with families, replicates and years as random effects. Yield error variances were heterogeneous for the 2 years, according to Hartley's maximum F ratio test (Neter and Wasserman 1974), and were stabilized by natural logarithmic transformation. Expected mean squares for the analysis of variance were presented by Godshalk et al. (1986). Genotypic variance and covariance components were estimated by computing appropriate linear functions of the mean squares and mean cross products, respectively. The genotypic correlation coefficients were computed using procedures described by Mode and Robinson (1959). A genotypic correlation was considered to be significant if the absolute value of the coefficient exceeded three times its standard error (Fisher 1936; Hill and Leath 1975). Simple correlation coefficients were computed on an individual plant basis according to Steel and Torrie (1980).

Mass selection was performed among the 1,280 Cycle 2 plants on the basis of a Smith-Hazel index, principal factor analysis (Harman 1976), maximum-likelihood factor analysis (Lawley and Maxwell 1971) and principal component analysis (Hotelling 1933). In index selection, the vector of index coefficients (b) was computed as $b = V_P^{-1} V_G\, a$, where $V_P^{-1}$ is the inverse of the phenotypic variance-covariance matrix, $V_G$ is the genotypic variance-covariance matrix and a is a vector of relative economic weights. Phenotypic variances and covariances for mass selection were computed according to Hallauer and Miranda (1981). Genotypic variances and covariances on an individual plant basis were computed by multiplying the among half-sib family variances and covariances by four, since among half-sib family variance represents one-quarter of the additive genetic variance. All traits were assigned equal economic weights of 100, assuming that the traits under selection were of equivalent exonomic value. Index coefficients for each trait were multiplied by their respective trait values and the resulting arithmetic products were then summed over traits to give an index score for each plant. The index scores were ranked in descending order and the plants corresponding to the 16 largest scores were selected.

Plants selected from principal component, principal factor and maximum-likelihood factor analyses were compared with selections made from Smith-Hazel indices. Principal component and factor analyses were performed using PROC FACTOR in SAS Institute (1985). Analyses were performed with and without varimax rotation (Kaiser 1958). In this type of orthogonal rotation, the loadings that were relatively small prior to rotation are reduced in magnitude as a result of rotation and larger loadings are forced closer to a value of one. In each multivariate analysis procedure, scoring coefficients were computed for each trait, with a vector of scoring coefficients corresponding to each factor in the specific model. The scoring coefficients were also added across the individual factor vectors, resulting in a vector of scoring coefficient sums. Orthogonality of factors permitted the scoring coefficients to be summed over factors. The scoring coefficients were used in the same manner as the index coefficients derived from selection indices. Total score values were computed for each plant.

Multivariate analyses were performed on genotypic and simple correlation matrices. Genotypic correlations among traits were analyzed to determine the effect of removing nongenetic variances and covariances from the correlation coefficients on the multivariate analysis techniques used for selection. Simple correlations among traits were analyzed to investigate the potential for selecting on the basis of the raw data and avoiding the costs involved in obtaining genetic parameter estimates. Multivariate analysis of correlation coefficients avoided scaling effects caused by differences in the units of measurements among the traits.

Simple and genotypic correlation matrices for three independent sets of traits were analyzed to test the repeatability of the comparison of index selection with multivariate analysis. Trait set A consisted of initial growth yield (trait #1) and regrowth XYL:GLU (trait #2), GAL (trait #3), GLU (trait #4) and ARA:GAL (trait #5). Trait set B included regrowth IVDMD (trait #1), PFC (trait #2), HEMI (trait #3), HEX (trait #4) and SUG (trait #5). Regrowth ARA (trait #1), XYL:GAL (trait #2), TFAHYD (trait #3), initial growth maturity rating (trait #4) and regrowth ADF (trait #5) constituted trait set C. There was no a priori biological basis for assembling the trait sets that were evaulated. The trait sets were intended to represent statistical relationships among traits that are typically encountered by plant breeders.

Effectiveness of selection by factor analysis and principal component analysis was estimated in terms of similarity to index selections, assuming that the "correct" set of plants was chosen by index selection. Godshalk et al. (1988) have already shown the effectiveness of index selection when attempting to develop switchgrass populations with desired characteristics. The demonstrated utility of index selection, as well as the relatively low standard errors of variance (Godshalk 1987) and covariance component estimates obtained in this study, indicates that index selection is a reasonably reliable control in evaluating the multivariate analysis methods of selection.

Spearman's rank correlation coefficients were computed between the index scores and the multivariate analysis scores to determine similarity of plant rankings over the methods of selection. The index selection and multivariate analysis methods were also compared for the number of plants, from a total of 16 plants, that were selected in common.

## Results and discussion

Correlation coefficients varied considerably in their signs and magnitudes for the three sets of traits (Table 1). Narrow-sense heritability estimates of the traits were generally low (Table 2), although 10 of the 15 traits had heritability estimates that were at least twice their standard errors. The variability of correlations and relatively low heritability estimates of these traits were intended to simulate situations typically encountered by plant breeders.

Principal component analysis of the genotypic and simple correlation matrices accounted for is the variance

of the variables (Table 3). However, factor analysis, by definition, only accounted for the variance of the common factors (Harman 1976). Common factors represent unobservable variables that explain variation of at least two observable variables. Variances unique to individual variables are those not explained by the factor analysis model. The proportions of variance accounted for by principal factor and maximum-likelihood factor analysis models were, as a result, less than those corresponding to principal component analysis. The number of factors included in each factor analysis model was specified to impose the maximum number of factors possible for a particular correlation matrix. A unique solution was not possible if the number of factors exceeded this maximum value, as determined by the correlation matrix rank. Prior communality estimation methods (SAS Institute 1985) were chosen to build a model accounting for a maximum proportion of the total variance.

For all three sets of traits, at least one of the Spearman's rank correlations for principal component analysis was larger than the correlations of the two methods of factor analysis (Table 4). In trait set A, maximum-likelihood factor analysis was less than satisfactory in simulating results of index selection. The largest rank correlation between maximum-likelihood factor analysis of the genotypic correlation matrix and index selection ($r=0.35$*) resulted from varimax rotation. The one-factor model in maximum-likelihood analysis of trait set B had high rank correlations with index selection scores ($r=0.80$** for analysis of the simple correlation matrix and $r=0.84$** for analysis of the genotypic correlation matrix). Principal factor analysis rank correlations for trait set B were similar to those of maximum-likelihood factor analysis and slightly lower than the principal component analysis rank correlations. In trait set C, the rank correlation coefficients were similar for the two factor analysis methods.

When ranking the plant scores from highest to lowest and selecting the top 16, none of the plants selected from maximum-likelihood analysis of trait set A were selected as a result of index selection (Table 5). This was also the case for principal factor analysis of the simple correlation matrix. However, for trait set A, as many as nine plants were selected in common when using principal component analysis of the simple correlation matrix and index selection. Rank correlations in Table 4 were indicative of the number of plants selected in common over the methods compared.

Principal component analysis of the simple and genotypic correlation matrices resulted in rank correlation coefficients that were similar in sign and magnitude for the two types of correlation matrices (Table 4). The correlation matrices were also similar in terms of the number of plants selected in common with index se-

**Table 1.** Genotypic (*above diagonal*) and simple (*below diagonal*) correlations for three different sets of switchgrass traits

| Trait number | Trait number | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| Trait set A | | | | | |
| 1 | – | 0.51 | 0.77[a] | −0.23 | −0.78[a] |
| 2 | 0.18[c] | – | 0.39 | −0.81[a] | −0.42 |
| 3 | 0.05 | −0.29[c] | – | −0.34 | −0.99[a] |
| 4 | −0.06[b] | −0.67[c] | 0.28[c] | – | 0.30 |
| 5 | 0.03 | 0.39[c] | −0.87[c] | −0.29[c] | – |
| Trait set B | | | | | |
| 1 | – | −0.73[a] | −0.09 | 0.78[a] | −0.54 |
| 2 | −0.59[c] | – | −0.31 | −0.51 | 0.27 |
| 3 | −0.28[c] | −0.03 | – | −0.35 | 0.46 |
| 4 | 0.33[c] | −0.30[c] | 0.02 | – | −0.28 |
| 5 | −0.56[c] | 0.47[c] | 0.57[c] | 0.05 | – |
| Trait set C | | | | | |
| 1 | – | −0.77[a] | 0.48 | −0.60 | 0.08 |
| 2 | −0.39[c] | – | −0.11 | 0.42 | −0.09 |
| 3 | 0.64[c] | −0.03 | – | −0.45 | 0.69[a] |
| 4 | 0.07[b] | 0.14[c] | 0.12[c] | – | −0.07 |
| 5 | 0.11 | 0.34[c] | 0.33[c] | 0.17[c] | – |

[a] Absolute value of genotypic correlation exceeds three times its standard error

[b, c] Significant at the 0.05 and 0.01 probability levels, respectively

**Table 2.** Narrow-sense heritability estimates and their standard errors, on an individual plant basis for several switchgrass traits

| Trait set | Narrow-sense heritability<br>Trait number | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| A | 0.31±0.14 | 0.58±0.18 | 0.24±0.09 | 0.28±0.14 | 0.22±0.09 |
| B | 0.48±0.18 | 0.21±0.10 | 0.16±0.09 | 0.44±0.16 | 0.21±0.09 |
| C | 0.07±0.09 | 0.15±0.09 | 0.21±0.11 | 0.19±0.14 | 0.46±0.17 |

**Table 3.** Proportion of variance explained by each factor in the model, when selecting for three different sets of traits using mass selection

| Correlation matrix | Factors[a] in model | Proportion of variance Set of traits | | |
|---|---|---|---|---|
| | | A | B | C |
| | | Principal component | | |
| Genotypic | Total | 1.000 | 1.000 | 1.000 |
| | First | 0.652 | 0.531 | 0.514 |
| | Second | 0.251 | 0.286 | 0.283 |
| | Third | 0.075 | 0.133 | 0.130 |
| | Fourth | 0.021 | 0.034 | 0.063 |
| | Fifth | 0.001 | 0.014 | 0.010 |
| Simple | Total | 1.000 | 1.000 | 1.000 |
| | First | 0.481 | 0.467 | 0.365 |
| | Second | 0.245 | 0.263 | 0.301 |
| | Third | 0.185 | 0.158 | 1.178 |
| | Fourth | 0.065 | 0.070 | 0.103 |
| | Fifth | 0.023 | 0.042 | 0.053 |
| | | Principal factor | | |
| Genotypic | Total | 0.978 | 0.950 | 0.982 |
| | First | 0.719 | 0.647 | 0.651 |
| | Second | 0.259 | 0.303 | 0.331 |
| Simple | Total | 0.972 | 1.033 | 0.917 |
| | First | 0.691 | 0.731 | 0.525 |
| | Second | 0.281 | 0.303 | 0.392 |
| | | Maximum-likelihood | | |
| Genotypic | Total | 0.990 | – | 0.798 |
| | First | 0.925 | 0.531 | 0.515 |
| | Second | 0.066 | – | 0.283 |
| Simple | Total | – | – | 0.666 |
| | First | 0.890 | 0.468 | 0.365 |
| | Second | – | – | 0.301 |

[a] 'Total' is a sum over the factors included in the model

lection (Table 5). This indicates the potential for avoiding estimation of genetic parameters, as required to construct Smith-Hazel indices, and the associated errors caused by these estimates (Lin 1978). It may be possible to select on the basis of simple correlation coefficients in conjunction with principal component analysis. When selecting, the breeder could use the set of scoring coefficients that correspond to a particular trait or combination of traits. This method of selection resembles the base index, in which index coefficients are assigned in accordance with relative importance of the traits. Assigning these coefficients to forage crop traits is often very difficult, since the relative importance of the traits determining yield and quality is not readily apparent. Principal component analysis may serve to alleviate this problem, by allowing the breeder to examine the scoring coefficient vectors and select the vector that contains values consistent with the goals of the breeding program.

The effect of varimax rotation on correlations between index selection and selection based on multivariate analysis was also observed (Table 4). Varimax rotation of factors derived from principal component analysis had the effect of redistributing rank correlations among the factors. The rank correlation sign of the first principal component analysis factor was only changed, as a result of rotation, for analysis of the simple correlation matrix of trait set C. Rotation of factors resulting from maximum-likelihood factor analysis had little effect on rank correlations. However, varimax rotation caused substantial redistribution of the rank correlations for principal factor analysis of genotypic correlations of trait set A and principal factor analysis of simple correlations of trait set B.

It is interesting to note that there was no apparent relationship between rank correlations (Table 4), or the number of plants selected in common (Table 5), and the proportion of variance explained by a given factor (Table 3). For example, maximum-likelihood factor analysis of the simple correlation matrix for trait set B accounted for only 46.8% of the total variance. However, the rank correlation between factor analysis scores and index scores was 0.80**, and six plants were selected in common over the two selection methods. This supports the use of multivariate analysis as a selection method, with the objective being to extract as many factors as possible and use the factor with scoring coefficients that emphasize the particular traits of interest. Principal component analysis generally provides the largest number of factors, since it accounts for 100% of the variance, and may be most appropriate in this regard.

Correlations between multivariate analysis selection and index selection were sensitive to conflicting simple and genotypic correlation coefficient signs. There were large positive and negative correlation coefficients for trait set A, whereas in the other two sets of traits the simple and genotypic coefficients tended to be lower in magnitude (Table 1). This resulted in Spearman's rank correlations for trait sets B and C that were somewhat larger than the rank correlations of trait set A. This pattern existed for the three methods of multivariate analysis. According to these results, selection based on multivariate analysis is most appropriate when there are few negative correlations or when the mixture of positive and negative correlations does not contain correlations of extremely large magnitude (i.e., greater than 0.80).

It has been shown that potential exists for using principal component analysis as a tool for selection and as a replacement for index selection. Most of the among

**Table 4.** Spearman's rank correlation between plants selected by index selection and plants selected from principal component analysis and two methods of factor analysis

| Varimax rotation | Factors[a] in model | Correlation coefficient[b] | | | | | |
|---|---|---|---|---|---|---|---|
| | | Principal component | | Principal factor | | Maximum-likelihood | |
| | | Index vs Factor-S | Index vs Factor-G | Index vs Factor-S | Index vs Factor-G | Index vs Factor-S | Index vs Factor-G |
| | | | | Trait set A | | | |
| Yes | Total | 0.73[d] | 0.66[d] | 0.002 | 0.57[d] | – | −0.41[c] |
| | First | −0.14 | 0.25 | 0.024 | −0.46[d] | – | 0.35[c] |
| | Second | 0.67[d] | 0.54[d] | 0.0004 | 0.56[d] | – | −0.42[c] |
| | Third | 0.47[d] | 0.11 | – | – | – | – |
| | Fourth | 0.52[d] | 0.75[d] | – | – | – | – |
| | Fifth | −0.33 | 0.56[d] | – | – | – | – |
| No | Total | 0.59[d] | 0.68[d] | −0.001 | −0.52[d] | – | −0.41[c] |
| | First | −0.42[c] | −0.35[c] | −0.012 | 0.58[d] | −0.18 | 0.16 |
| | Second | −0.38[c] | 0.48[d] | 0.002 | −0.55[d] | – | −0.42[c] |
| | Third | 0.61[d] | 0.79[d] | – | – | – | – |
| | Fourth | 0.76[d] | 0.73[d] | – | – | – | – |
| | Fifth | −0.54[d] | 0.57[d] | – | – | – | – |
| | | | | Trait set B | | | |
| Yes | Total | 0.94[d] | 0.92[d] | −0.68[d] | 0.85[d] | – | – |
| | First | −0.28 | 0.88[d] | 0.17 | 0.83[d] | – | – |
| | Second | 0.61[d] | 0.35[c] | −0.82[d] | −0.67[d] | – | – |
| | Third | −0.54[d] | 0.26 | – | – | – | – |
| | Fourth | 0.86[d] | 0.03 | – | – | – | – |
| | Fifth | 0.74[d] | 0.83[d] | – | – | – | – |
| No | Total | 0.89[d] | 0.88[d] | 0.28 | 0.87[d] | – | – |
| | First | −0.73[d] | 0.78[d] | −0.66[d] | 0.81[d] | 0.80[d] | 0.84[d] |
| | Second | 0.66[d] | 0.46[d] | 0.85[d] | −0.46[d] | – | – |
| | Third | −0.18 | −0.06 | – | – | – | – |
| | Fourth | 0.88[d] | 0.90[d] | – | – | – | – |
| | Fifth | −0.49[d] | 0.80[d] | – | – | – | – |
| | | | | Trait set C | | | |
| Yes | Total | 0.69[d] | 0.62[d] | 0.77[d] | 0.07 | 0.78[d] | 0.42[c] |
| | First | 0.22 | 0.17 | 0.47[d] | 0.42[c] | 0.40[c] | −0.50[d] |
| | Second | −0.34[c] | 0.82[d] | 0.75[d] | −0.19 | 0.78[d] | 0.47[d] |
| | Third | 0.87[d] | −0.25[d] | – | – | – | – |
| | Fourth | −0.71[d] | −0.37[c] | – | – | – | – |
| | Fifth | −0.57[d] | 0.42[c] | – | – | – | – |
| No | Total | −0.67[d] | 0.69[d] | 0.77[d] | −0.16 | 0.78[d] | 0.44[c] |
| | First | 0.76[d] | 0.75[d] | 0.60[d] | 0.42[d] | 0.22 | 0.07 |
| | Second | 0.72[d] | 0.84[d] | 0.74[d] | −0.25 | 0.78[d] | 0.47[d] |
| | Third | −0.86[d] | 0.72[d] | – | – | – | – |
| | Fourth | −0.78[d] | −0.62[d] | – | – | – | – |
| | Fifth | −0.40[c] | 0.46[d] | – | – | – | – |

[a] 'Total' refers to selection by summing over scores of factors included in the model

[b] Index = index selection, Factor-S = factor analysis on simple correlation matrix, and Factor-G = factor analysis on genotypic correlation matrix

[c,d] Significant at the 0.05 and 0.01 probability levels, respectively

half-sib family variance component estimates were considered to be significant or even highly significant (Godshalk 1987), suggesting that index selection errors resulting from errors in parameter estimates may have been minimal. Assuming that the indices constructed in this study were valid standards for comparison, the high correlations of principal component analysis selection with index selection indicate that selection may be accomplished merely on the basis of principal component analysis of simple correlation coefficients. This would

**Table 5.** Number of plants selected in common, out of a total of 16, by index selection, principal component analysis and two methods of factor analysis

| Varimax rotation | Factors[a] in model | Number of plants[b] | | | | | |
|---|---|---|---|---|---|---|---|
| | | Principal component | | Principal factor | | Maximum-likelihood | |
| | | Index vs Factor-S | Index vs Factor-G | Index vs Factor-S | Index vs Factor-G | Index vs Factor-S | Index vs Factor-G |
| | | | | Trait set A | | | |
| Yes | Total | 7 | 8 | 0 | 7 | – | 0 |
| | First | 0 | 0 | 0 | 0 | – | 0 |
| | Second | 6 | 4 | 0 | 7 | – | 0 |
| | Third | 1 | 0 | – | – | – | – |
| | Fourth | 4 | 8 | – | – | – | – |
| | Fifth | 0 | 6 | – | – | – | – |
| No | Total | 6 | 7 | 0 | 0 | – | 0 |
| | First | 0 | 0 | 0 | 7 | 0 | 0 |
| | Second | 0 | 3 | 0 | 0 | – | 0 |
| | Third | 5 | 7 | – | – | – | – |
| | Fourth | 9 | 9 | – | – | – | – |
| | Fifth | 0 | 6 | – | – | – | – |
| | | | | Trait set B | | | |
| Yes | Total | 8 | 8 | 0 | 6 | – | – |
| | First | 1 | 5 | 0 | 6 | – | – |
| | Second | 2 | 1 | 0 | 0 | – | – |
| | Third | 0 | 1 | – | – | – | – |
| | Fourth | 6 | 1 | – | – | – | – |
| | Fifth | 4 | 6 | – | – | – | – |
| No | Total | 7 | 6 | 0 | 8 | – | – |
| | First | 0 | 6 | 0 | 6 | 6 | 6 |
| | Second | 2 | 1 | 7 | 0 | – | – |
| | Third | 0 | 0 | – | – | – | – |
| | Fourth | 5 | 8 | – | – | – | – |
| | Fifth | 0 | 6 | – | – | – | – |
| | | | | Trait set C | | | |
| Yes | Total | 7 | 6 | 8 | 0 | 9 | 2 |
| | First | 0 | 1 | 1 | 3 | 1 | 0 |
| | Second | 0 | 7 | 9 | 0 | 9 | 2 |
| | Third | 7 | 0 | – | – | – | – |
| | Fourth | 0 | 0 | – | – | – | – |
| | Fifth | 0 | 3 | – | – | – | – |
| No | Total | 0 | 6 | 9 | 0 | 8 | 2 |
| | First | 7 | 7 | 3 | 2 | 0 | 0 |
| | Second | 7 | 9 | 9 | 0 | 9 | 2 |
| | Third | 0 | 4 | – | – | – | – |
| | Fourth | 0 | 0 | – | – | – | – |
| | Fifth | 0 | 3 | – | – | – | – |

[a] 'Total' refers to selection by summing the scores of factors included in the model

[b] Index = index selection, Factor-S = factor analysis on simple correlation matrix, and Factor-G = Factor analysis on genotypic correlation matrix

eliminate the need to allocate field resources (e.g., time, labor, equipment) required to obtain reliable genetic parameter estimates. Principal component analysis of simple correlations may allow the breeder more flexibility in determining the number of families and family sizes to be evaluated. In some cases, substantially more genetic gain may be expected from selection within families than from among family selection. However, large family sizes and a small number of families would likely result in genetic parameter estimates with large

standard errors. The breeder could use the multivariate selection methods by first determining the combination of traits that constitute an "ideal" plant. By plotting the principal components or factors that are considered to be important, plants close to the ideal plant on the two-dimensional plots would be chosen. Principal components or factors may be deemed important if their associated scoring coefficients are of relative magnitude or sign consistent with breeding objectives.

Given this apparent potential for using principal component analysis, further work is needed to compare multivariate analysis methods with index selection in terms of actual gains achieved from selection. A number of methods of factor analysis and (orthogonal or oblique) rotation, in addition to those included in this study, should be carefully considered when designing the comparison. Variations of factor analysis and factor rotation are described thoroughly by Harman (1976). In making this comparison, consideration should be given to costs involved in estimating genetic parameters that are required to construct the Smith-Hazel index.

## References


Allaire FR, Henderson CR (1966) Selection practiced among dairy cows. II. Total production over a sequence of lactations. J Dairy Sci 49:1435–1440

Andrus DF, McGilliard LD (1975) Selection of dairy cattle for overall excellence. J Dairy Sci 58:1876–1879

Baker RJ (1986) Selection indices in plant breeding. CRC Press, Boca Raton/FL

Bramel PJ, Hinz PN, Green DE, Shibles RM (1984) Use of principal factor analysis in the study of three stem termination types of soybean. Euphytica 33:398–400

Brim CA, Johnson HW, Cockerham CC (1959) Multiple selection criteria in soybeans. Agron J 51:42–46

Campo JL, Rodriguez MC (1985) Experimental comparison of methods for simultaneous selection of two correlated traits in *Tribolium*. Theor Appl Genet 71:93–100

Chatfield C, Collins AJ (1980) Introduction to multivariate analysis. Chapman and Hall, London New York

Denis JC, Adams MW (1978) A factor analysis of plant variables related to yield in dry beans. I. Morphological traits. Crop Sci 18:74–78

deRuiter JM (1984) Characterization and prediction of cell wall carbohydrate composition of the subtropical forage grasses: *Pennisetum flaccidum, Panicum amarum, Panicum amarulum* and *Panicum virgatum*. PhD thesis, North Carolina State University, Raleigh/NC

Dickerson GE, Blunn CT, Chapman AG, Kottman RM, Krider JL, Warwick EJ, Whatley JA Jr, Baker ML, Lush JL, Winters LM (1954) Evaluation of selection in developing inbred lines of swine. Missouri Res Bull 551, University of Missouri, Columbia, MO

Dunteman GH (1984) Introduction to multivariate analysis. Sage Publications, Beverly Hills, Calif

Eagles HA, Frey KJ (1974) Expected and actual gains in economic value of oat lines from five selection methods. Crop Sci 14:861–864

Elgin JH Jr, Hill RR Jr, Zeiders KE (1970) Comparison of four methods of multiple trait selection for five traits in alfalfa. Crop Sci 10:190–193

Fisher RA (1936) Statistical methods for research workers. Oliver and Boyd, London

Godshalk EB (1987) Evaluation of heritability of cell wall monosaccharides, index selection and multivariate analysis for switchgrass improvement. PhD thesis, North Carolina State University, Raleigh/NC

Godshalk EB, Burns JC, Timothy DH (1986) Selection for in vitro dry matter disappearance in switchgrass regrowth. Crop Sci 26:943–947

Godshalk EB, Timothy DH, Burns JC (1988) Effectiveness of index selection for switchgrass forage yield and quality. Crop Sci (In Press)

Hallauer AR, Miranda JB Fo (1981) Quantitative genetics in maize breeding. Iowa State University Press, Ames/IA

Harman HH (1976) Modern factor analysis. University of Chicago Press, Chicago, Ill

Harris DL (1964) Expected and predicted progress from index selection involving estimates of population parameters. Biometrics 20:46–72

Hazel LN (1943) The genetic basis for constructing selection indices. Genetics 476–490

Hill RR Jr, Leath KT (1975) Genotypic and phenotypic correlations for reaction to five foliar pathogens in alfalfa. Theor Appl Genet 45:254–258

Hotelling H (1933) Analysis of a complex of statistical variables into principal components. J Ed Psych 24:417–441m 498–520

Hsu P, Walton PD (1971) Relationship between yield, its components and structures above the flag leaf node in spring wheat. Crop Sci 11:190–193

Kaiser HF (1958) The varimax criterion for analytic rotation in factor analysis. Psych 23:187–200

Lawley DN, Maxwell AE (1971) Factor analysis as a statistical method. McMillan, New York

Lin CY (1978) Index selection for genetic improvement of quantitative characters. Theor Appl Genet 52:49–56

Lin CY, Pevzner I, Friars GW (1979) Experimental investigation of errors of heritability estimates in index selection. Can J Genet Cytol 21:303–308

McClure WF, Hamid A (1980) Rapid NIR measurement of the chemical composition of foods and food products. Part one: Hardware. Am Lab, September, pp 57–69

Mode CJ, Robinson HF (1959) Pleiotropism and the genetic variance and covariance. Biometrics 15:518–537

Murty BR, Arunachalam V (1967) Factor analysis of diversity in the genus *Sorghum*. Indian J Genet Plant Breed 27:123–135

Neter J, Wasserman W (1974) Applied linear statistical models. RD Irwin, Homewood/IL

Riggs TJ (1973) The use of canonical analysis for selection within a population of spring barley. Ann Appl Biol 74:249–258

Rummel RJ (1970) Applied factor analysis. Northwestern University Press, Evanston/IL, pp 16–17

SAS Institute (1985) SAS User's Guide: Statistics, version 5 edn. SAS Institute, Cary/NC

Smith HF (1936) A discriminant function for plant selection. Ann Eugen, London 7:240–250

Steel RGD, Torrie JH (1980) Principles and procedures of statistics, 2nd edn. McGraw-Hill, New York

Talbert LE, Timothy DH, Burns JC, Rawlings JO, Moll RH (1983) Estimates of genetic parameters in switchgrass. Crop Sci 23:725–728

Vleck LD van (1974) Notes on the theory and application of selection principles for the genetic improvement of animals. Department of Animal Science, Cornell University

Walton PD (1971) The use of factor analysis in determining characters for yield selection in wheat. Euphytica 20:416–421

Walton PD (1972) Factor analysis of yield in spring wheat (*Triticum aestivum* L.). Crop Sci 12:731–733

Williams JS (1962) The evaluation of a selection index. Biometrics 18:375–393

Young SSY (1961) A further examination of the relative efficiency of three methods selection for genetic gains under less restricted conditions. Genet Res 2:106–121