



## Path analysis and expected response in indirect selection for grain yield in soybean

Ivana Marino Bárbaro<sup>1\*</sup>, Maria Aparecida Pessoa da Cruz Centurion<sup>2</sup>, Antonio Orlando Di Mauro<sup>2</sup>, Sandra Helena Unêda-Trevisoli<sup>3</sup>, Nair Helena Castro Arriel<sup>4</sup>, and Marcelo Marchi Costa<sup>2</sup>

Received 16 September 2005

Accepted 23 May 2006

**ABSTRACT** – *This study aimed to determine the best auxiliary trait for indirect selection of soybean grain yield, through path analysis and in avoidance of the adverse effects of multicollinearity and expected response. Seventy-nine F<sub>5</sub> soybean genotypes from the cross FT-Cometa x Bossier were used. The populations were distributed on the field as the families inserted with replicated controls. Primary and secondary traits of grain yield were evaluated in four phenotypically superior plants per family. The traits number of pods, height and number of nodes were considered as the most important, showing the best combination of direct effect and genotypic correlation. The number of pods achieved the highest expected gain through the estimation method based on the selection differential. On the other hand, plant height, by the method based on selection intensity, was not a good indicator of the most productive plants.*

**Key words:** agronomic traits, genotypic correlation, path analysis, collinearity, indirect selection.

### INTRODUCTION

In an improvement program it is essential to know the magnitude of association between traits, so breeders can understand how the selection of one trait can cause alterations in others (Johnson et al. 1955, Vencovsky and Barriga 1992). It is also possible to achieve an indirect improvement of a trait with complex inheritance and low heritability through the selection of another trait, with a more simple inheritance or higher heritability. Still, the quantification and interpretation of the magnitude of a correlation can lead to mistakes in the selection strategy, since a high correlation between

two traits can be result of the effect of a third trait or of a group of traits on them (Cruz and Regazzi 1994). To deepen the understanding on the reasons of the association between traits, Wright (1921) proposed a method called path analysis that partitions the estimated correlations in direct and indirect effects of traits on a basic variable. This method has recently been studied in some crops by Santos et al. (1995), Carvalho et al. (1999), Reis et al. (2001), Kurek et al. (2001), Carvalho et al. (2002) and others.

The comprehension of the system cause-effect between the variables lies precisely in the success of

<sup>1</sup> APTA, Polo Regional Alto Mogiana, Avenida Rui Barbosa, s/n, C.P. 35, 14.770-000, Colina, SP, Brasil. \*Email: imarino@aptaregional.sp.org.br

<sup>2</sup> Departamento de Produção Vegetal, Faculdade de Ciências Agrárias e Veterinárias (UNESP), Via de Acesso Prof. Paulo Donato Castellane, s/n, 14.884-900, Jaboticabal, SP, Brasil

<sup>3</sup> APTA, Polo Regional Centro Leste, Av. Bandeirantes, 2419, 14.030-670, Ribeirão Preto, SP, Brasil

<sup>4</sup> Embrapa Algodão, Rua Oswaldo Cruz, 1143, Centenário, C.P. 174, 58.107-720, Campina Grande, PB, Brasil

the path analysis (Schuster 1996). The partitioning of the correlations depends on a set of traits under study that is normally determined based on previous knowledge of the researcher, in view of their importance and possible inter-relations expressed in path diagrams (Cruz and Regazzi 1994). The estimation of the path coefficients can be adversely affected by the effects of multicollinearity between the traits, which appear when the random observations of the explanatory variables or linear combinations are correlated (Ferrari 1989). The problems caused by multicollinearity are not simply due to the presence, but rather to the degree in which it is manifested. When multicollinearity occurs in moderate to severe levels in a set of explanatory variables, it becomes very difficult to evaluate its influence on the response in the main variable, whereas the ignorance of its effects can bring forth undesirable results (Cruz and Carneiro 2003). In the presence of multicollinearity, the variances associated to the estimators of the path coefficients can therefore attain very high values, making the estimates little reliable (Carvalho 1995, Cruz and Carneiro 2003). Besides, the parameter estimates can assume values without any coherence with the biological phenomenon under study (Cruz and Carneiro 2003). Principally, in cases involving a large number of variables or when there is no previous knowledge on the association between them, results may be misvalued, owing to the multicollinearity leading to not very pertinent conclusions (Cruz and Carneiro 2003). The problem can be solved by diagnosing the multicollinearity of data before realizing the procedure. The criterion that is efficient is the one that, apart from informing on the existence of multicollinearity, shows the degree of severity and identifies the variables involved in the problem (Montgomery and Peck 1981). In order to lessen the adverse effects of multicollinearity, one can identify the variables that are causing the problems and eliminate them, to carry out the analysis with a smaller group. There is also the possibility of using an alternative methodology recommended by Carvalho (1995), denominated ridge regression analysis, which is used when the exclusion of variables is not in the breeder's interest.

The objective of the present study was to evaluate the direct and indirect effects of agronomic traits on soybean grain yield by path analysis, with a view to help breeders in the selection process of plants by the

prediction of the genetic gain and through the indication of the best correlated trait for use.

## MATERIAL AND METHODS

The experiment was conducted in the growing season 2003/04 on an experimental area of the Departamento de Produção Vegetal of the FCAV-UNESP, campus of Jaboticabal/SP. Data obtained from a soybean population ( $F_5$  generation), derived from the cross FT-Cometa x Bossier were analyzed, conducted by the modified pedigree method and consisting of 79 families.

The population on the field was distributed in families alternated with controls; each plot consisted of one five meter long row which represented one family (progeny of a plant selected in the previous generation). The rows were spaced 0.5 m and the mean density was 15 to 20 plants per meter. Two standard cultivars (COODETEC- 205 and BRS/MG 68 (Vencedora)) were used, which were alternated at every 10 experimental rows with the above-mentioned experimental plot structure.

At maturation, four phenotypically superior plants were picked from each family for an evaluation of the following agronomic traits considered primary for yield (GY): number of pods per plant (NP), number of seeds per pod (NS) and weight of one seed in g/seed (W1S) and as secondary traits: number of days to flowering (NDF), number of days to maturation (NDM), plant height at maturation (PHM), insertion height of the first pod (IHP) and number of nodes (NN).

The estimates of genotypic correlations between families were obtained as suggested by Vencovsky and Barriga (1992) apart from the partitioning of these correlations in direct and indirect effects of the traits of agronomic importance for soybean (independent variables of the regression model) on the grain yield (basic or dependent variable) using the path analysis developed by Wright (1921).

The mathematical yield model is given by:  $PG = NS \times NV \times P1S$ , that is, it is multiplicative. With the logarithmization this model becomes additive, since one of the premises of the path analysis is the additivity (Santos et al. 1995). To perform path analysis and obtain the direct and indirect effects, the matrix  $X'X$  must be well-established. In the presence of moderate to severe multicollinearity between the variables, the least squares estimator, obtained from  $X'X = \beta = X'Y$ , can be associated

to a very high variance, whose estimates are little reliable. Since it would not be of interest to discard the variables that are causing the problem, an alternative methodology proposed by Carvalho (1995) and Carvalho and Cruz (1996) was used, to minimize this adverse effect through the introduction of constant K in the diagonal of matrix  $X'X$ , similarly to the ridge regression analysis proposed by Hoerl and Kennard (1970a) in slight modification of the system of normal equations. The path coefficients were obtained this way by means of the equation  $(X'X + KIP)\theta = X'Y$ , where  $X'X$  is the correlation matrix between the independent variables of the regression model; K is a small quantity added to the elements of the diagonal of matrix  $X'X$ ;  $I_p$  is the matrix identity;  $\theta$  is the vector of the estimators of the path coefficients and  $X'Y$  is the correlation matrix between the dependent variable with each independent variable of the regression model.

The value obtained for constant K was determined in this trial by the graphical examination of the ridge trace (Hoerl and Kennard 1970b), which consists of a graphic where the Y axis is represented by the  $\beta_j$  ( $j = 1.2...p$ ) values, obtained based on a set of K values represented in the X axis. The ridge trace is found by plotting the path coefficients in function of K values in the interval  $0 < k < 1$ . In these conditions, the K value was chosen for the genotypic effects that presented multicollinearity, which was the lowest possible value, for which the majority of the path coefficients, associated to several traits, would be stabilized.

Two criteria proposed by Montgomery and Peck (1981) were used for a diagnosis of the degree of multicollinearity of the matrix  $X'X$ , which are based on an examination of the determinant of the correlation matrix and the analysis of the eigenvalues and eigenvectors through an evaluation of the number of condition (NC), which is the ratio of the highest and the lowest eigenvalue of these matrixes.

As the determinant approaches zero, multicollinearity becomes more intense and its effects more harmful, although this criterion does not identify the traits that cause multicollinearity. However, the detection becomes possible in an analysis of the elements of the eigenvectors together with the eigenvalues, as cited by Belsley et al. (1980). When  $NC \leq 100$ , multicollinearity is considered weak and does not pose a serious problem in the analysis.

When  $100 < NC < 1000$ , multicollinearity is considered moderate to strong, so that  $NC \leq 1000$  represents an indicator of severe multicollinearity.

The determination coefficient for the analysis of the traits on the grain yield (variable basic) is given by  $R^2 = p_1 r_{1y} + p_2 r_{2y} + \dots + p_n r_{ny}$ , where  $r_{iy}$  = correlation between the main variable (y) and the  $i^{th}$  explanatory variable.;  $p_i$  = measure of the direct effect of variable i on the main variable and  $p_j r_{ij}$  = measure of the indirect effect of variable i through variable j, on the main variable. In turn, the residual effect is expressed by  $p_E = \sqrt{1 - r}$ .

The following criteria were adopted to compute gain prediction: 50% selection among families; selection for optimized values of the traits GY, NS, NP, WIS, PHM, IHP and NN; selection towards lower values for NDF and NDM.

Two methodologies of expected gain estimation were used:

*Based on the selection intensity* by the formula: , where:  $GS_i = pk\sigma_{gi}h_i$  p: parental control (assumed to be equal to 1); k: selection intensity or differential of selection in units of the phenotypic standard deviation;  $\sigma_{gi}$ : additive genetic standard deviation of trait i;  $h_i$ : square root of heritability at the level of family means, for trait i.

The indirect gain in trait j through selection for trait i is given by:  $GS_{j(i)} = pkh_i r_{gi} \sigma_{gi}$ , where:  $r_{gi}$ : genetic correlation between the traits i and j;

*Based on the differential of selection* by the formula: ,  $GS_i = (X_{Si} - X_{oi})h_i^2 = DS_i h_i^2$  where:  $X_{Si}$ : mean of the plants selected for trait i;  $X_{oi}$ : original mean of the population;  $DS_i$ : differential of selection realized in the population;  $h_i^2$ : heritability at the level of family means for trait i.

Indirect gain in trait j through selection for trait i is given by:  $GS_{j(i)} = DS_{j(i)} h_j^2$ , where:  $DS_{j(i)}$ : differential of indirect selection obtained in function of the mean of the trait of the plants whose superiority was evidenced based on the other trait, which is the target of direct selection.

All statistical analyses of the present study were performed using the software Genes (Cruz 2001).

## RESULTS AND DISCUSSION

The only directly measurable correlation between traits in a trial is the phenotypic correlation, for genetic and environmental reasons. However, only the genetic correlation involves an association of inheritable nature,

which can consequently be used in the orientation of improvement programs. In this study only genotypic correlation estimates were considered, to make it easier to take decisions regarding the efficacy of indirect selection.

The estimates of the direct and indirect effects of the explanatory variables considered primary (NS, NP and W1S) on the basic variable (GY) are presented in Table 1.

The determination coefficient  $R^2$  was 0.992 and the residual effect 0.009. The adopted explanatory model was therefore satisfactory to explain the relation cause-effect in the primary variables. The traits NS, NP and W1S were slightly multicollinear (determinant of the matrix  $X'X = 0.329$  and  $NC = 5.45525, 0.54229$  and  $1.02801$ , respectively).

The most favorable situation for improvement was verified for the primary trait NP, which presented the highest and positive value in the genotypic correlation (0.921) and in the direct effect (0.586), compared with the traits NS and W1S, in agreement with results obtained by Santos et al. (1995) and Reis et al. (2001).

**Table 1.** Estimates of the direct and indirect effects of the primary traits (NS, NP and W1S) on grain yield, by means of the ridge path analysis in a  $F_5$  soybean population

Traits	Effect
Direct effect of NS on GY	0.267
Indirect effect via NP	-0.077
Indirect effect via W1S	-0.039
Total	<b>0.150</b>
Direct effect of NP on GY	0.586
Indirect effect via NS	-0.035
Indirect effect via W1S	0.369
Total	<b>0.921</b>
Direct effect of W1S on GY	0.452
Indirect effect via NS	-0.023
Indirect effect via NP	0.478
Total	<b>0.908</b>
Determination coefficient ( $R^2$ )	<b>0.992</b>
Residual effect ( $p_E$ )	<b>0.009</b>
Determinant of the matrix	<b>0.329</b>
Collinearity of the matrix	<b>Weak</b>
K	<b>0</b>

NS = number of seeds per pod; NP = number of pods per plant; W1S = weight of one seed and GY = grain yield.

The very low correlation between NS and GY on the other hand is provoked by the indirect effects, mainly of trait NP.

In general, the information on the effects of the secondary variables on the basic variable confirmed the information obtained in the path analyses of the secondary component effects on the primary and of these on the basic variable (Santos et al. 1995); the estimates of the direct and indirect effects of the secondary on the primary variables are presented in Table 2.

The inclusion of trait NN in the genotypic correlation matrix (NP x NN, NS x NN and W1S x NN) was the main reason for the appearance of severe multicollinearity (determinant of the matrix  $X'X = -0.001$  and  $NC = 255227652226.258$ ). To attenuate the adverse multicollinearity effects, the alternative methodology recommended by Carvalho (1995) was used; the K value used in the present study was 0.3415408 for the estimation of the path coefficients. Due to presence of severe multicollinearity, these became more stable with the cited K value. Our results were concordant with those obtained by Santos et al. (1995). Therefore, the exclusion of variables from the regression model would not necessarily be a viable alternative to minimize the adverse multicollinearity effects in the present study.

The best combinations for the traits NS and W1S, i.e., a direct positive effect associated with a high and positive correlation, apart from of the indirect positive effect with the secondary trait NDM, was observed with trait PHM. Thus, the increases in the number of seeds per pod and weight of one seed can be achieved through selection correlated with this trait. Trait NP however presented PHM and NN as the most important variables, partly in line with Santos et al. (1995), who verified the most important traits in the process of indirect selection for a higher soybean grain yield through secondary components, through the values of the direct effects of NN and NDF on NS, of PHM on W1S and of the number of nodes at flowering (NNF) and NDM on NP. Likewise, they partly agreed with the results obtained by Carvalho et al. (2002), who recognized the traits NDM and plant height at flowering (APF), individually, as most important or the combination NDM, APF and NN, depending on the time of planting.

The study of the influences of secondary components on yield components can have great practical importance, since these traits are generally of

**Table 2.** Estimates of the direct and indirect effects of the secondary on the primary traits, by means of the ridge path analysis<sup>1</sup> in a F<sub>5</sub> soybean population

Description of the effects	Primary components					
	With collinearity			Ridge path analysis		
	NS	W1S	NP	NS	W1S	NP
<b>Direct effect of NDF-</b>	1.260	-0.922	0.786	0.269	0.093	0.620
Indirect effect via NDM	2.268	1.531	0.515	-0.803	-0.472	-0.610
via PHM	0.279	0.205	2.551	-0.594	-0.527	-0.606
via IHP	-2.451	-1.358	0.677	0.293	0.239	0.435
via NN	0.630	-0.070	-4.593	0.208	0.020	-0.116
<b>Total</b>	<b>-0.534</b>	<b>-0.614</b>	<b>-0.064</b>	<b>-0.534</b>	<b>-0.614</b>	<b>-0.064</b>
<b>Direct effect of NDM</b>	2.798	1.888	0.635	-0.990	-0.583	-0.752
Indirect effect via NDF	-1.026	-0.747	0.637	0.218	0.075	0.503
via PHM	0.145	0.106	1.326	-0.309	-0.274	-0.315
via IHP	-3.294	-1.830	0.911	0.395	0.322	0.586
via NN	0.527	-0.058	-3.841	0.174	0.017	-0.097
<b>Total</b>	<b>-0.850</b>	<b>-0.641</b>	<b>-0.332</b>	<b>-0.850</b>	<b>-0.641</b>	<b>-0.332</b>
<b>Direct effect of PHM</b>	-0.434	-0.319	-3.966	0.924	0.820	0.942
Indirect effect via NDF	0.814	0.593	-0.505	-0.173	-0.060	-0.399
via NDM	-0.936	-0.632	-0.211	0.331	0.195	0.251
via IHP	2.251	1.251	-0.622	-0.270	-0.220	-0.401
via NN	-0.847	0.094	6.177	-0.279	-0.027	0.157
<b>Total</b>	<b>0.848</b>	<b>0.987</b>	<b>0.873</b>	<b>0.848</b>	<b>0.987</b>	<b>0.873</b>
<b>Direct effect of IHP</b>	-3.389	-1.882	0.937	0.406	0.332	0.603
Indirect effect via NDF	-0.912	-0.667	0.567	0.194	0.067	0.448
via NDM	2.719	1.835	0.617	-0.962	-0.566	-0.731
via PHM	0.288	0.212	2.632	-0.614	-0.545	-0.626
via NN	0.683	-0.075	-4.979	0.225	0.022	-0.126
<b>Total</b>	<b>-0.611</b>	<b>-0.577</b>	<b>-0.226</b>	<b>-0.611</b>	<b>-0.577</b>	<b>-0.226</b>
<b>Direct effect of NN</b>	-0.861	0.095	6.282	-0.284	-0.027	0.159
Indirect effect via NDF	0.925	0.674	-0.575	-0.197	-0.068	-0.454
via NDM	-1.712	-1.155	-0.388	0.606	0.356	0.460
via PHM	-0.427	-0.314	-3.902	0.910	0.807	0.927
via IHP	2.689	1.494	-0.748	-0.322	-0.263	-0.479
<b>Total</b>	<b>0.614</b>	<b>0.794</b>	<b>0.669</b>	<b>0.614</b>	<b>0.794</b>	<b>0.669</b>
<b>R<sup>2</sup></b>	-0.531	0.202	0.267	1.059	0.912	1.003
<b>P<sub>E</sub></b>	1.238	0.893	0.856	-	0.295	-
<i>Collinearity of the matrix</i>	<i>severe</i>	<i>severe</i>	<i>severe</i>	<i>weak</i>	<i>weak</i>	<i>weak</i>
<b>Determinant of the matrix</b>	-0.001	-0.001	-0.001	0.288	0.288	0.288
<b>K</b>	0	0	0	0.3415	0.3415	0.3415

<sup>1</sup> alternative methodology recommended by Carvalho (1995), to be used when the exclusion of variables is not in the breeder's interest; NDF = number of days to flowering; NDM = number of days to maturation; PHM = plant height at maturity; IHP = insertion height of the first pod; NN = number of nodes; NS = number of seeds per pod; W1S = weight of one seed and NP = number of pods per plant.

less complex heritability, of high heritability and therefore easier to be selected (Cruz and Regazzi 1994).

The values of the direct effects of NDF and IHP on the traits NS, W1S and NP presented the opposite

situation to that obtained by the coefficient of genotypic correlation. This can be an indication of inadequacy of the correlation in such situations. The results show the value of the path analysis in presenting



the true genotypic relations of cause-effect between the evaluated traits.

One explanation for the negative coefficients among NS x NDF, W1S x NDF, NP x NDF, NDM x NS, NDM x W1S and NDM x NP was the high incidence of Asian or Australasian soybean rust (*Phakopsora pachyrhizi* Sidow) in the later genotypes, in the growing season 2003/04. Santos et al. (1995) found a similar situation in the values of the direct effects of the traits NNF and PHM on NS, NNF on W1S and NDF on NP. This fact evidences the influence of other variables on the relationship under study and the inadequacy of the correlation in these situations. Carvalho et al. (2002) observed positive correlations and direct negative or null effects of NDF on GY in soybean crosses.

The direct and indirect effects of the secondary traits on the variable grain yield, by the primary components, are presented in Table 3. In the total of the effects, variables PHM and NN presented the best combination of direct effect and correlation; the situation of these variables is best when the path of NP is considered.

Table 4 presents the estimates of the heritabilities at the level of family means; original means of the grain yield and families selected with a view to grain yield via auxiliary trait; differential of selection; gains with direct and indirect selection; percentage gains; and efficacy of indirect selection through the methodology of estimation based on the selection differential.

The results obtained in Table 4 indicate more promising possibilities of gain for grain yield when the families are selected for a higher NP, due to its importance as auxiliary trait. This fact can be observed in the methodology of expected gain estimation based on the selection differential. With this selection a gain can be obtained of up to 84.7% of the one expected when GY is directly selected. The presence of the coefficient of high genotypic correlation (0.921) was also verified and heritability at the level of family means of over 50%. Thereafter, for the same methodology, expected gains of 72.2% and 72.0% were obtained, respectively, for the traits NN and W1S. Reis et al. (2001) also obtained evidences of a greater efficacy of the primary component NP as indicator of more productive plants.

Table 5 shows the heritability estimates at the level of family means; genetic correlations between grain yield and the other traits; genotypic variances; standardized selection differential; yield gains due to direct and indirect selection through the methodology of

estimation based on selection intensity; gains with selection in percent; and efficacy of indirect selection in a F<sub>5</sub> soybean population.

The performance of efficacy of the traits NP, NN and W1S was not in the same order in the methodology based on selection intensity for estimation of the expected gain. The gains of PHM, NN, NP and W1S were 218.0, 87.4, 18.1 and 6.0%, respectively (Table 5). Reis et al. (2001) suggested that the use of this methodology may be less effective when the selected units do not reflect the correlation observed in the population under selection; this can lead to over or underestimated gains with selection. Besides, primary yield traits must contribute more than secondary traits, even if they are controlled by several genes, more influenced by the environment, generally with lower heritabilities and more complicated for measurements.

The responses correlated in view of the secondary traits IHP for the two methodologies of gain prediction (based on the differential of selection and selection intensity), and NDF and NDM only in the estimation methodology based on selection intensity, were negative, since later plants with higher insertion heights did not coincide with the more productive. It was verified that the grain yield mean, when selecting for the auxiliary trait IHP, was lower than the original population mean and consequently entailed a negative selection differential. Besides, the coefficients of negative genotypic correlations obtained in the three traits resulted in negative responses in the methodology of gain prediction based on selection intensity. These observations coincide with the results of Reis et al. (2001), who observed that the plants selected indirectly for the traits NDF and NDM were not representative of the base population, making the population parameters used to calculate the expected response in such trees inconsistent. In the case of this study, possible causes would be the rigorous selection for earliness and greater insertion heights of the first pod of the genotypes in the previous and present generations, apart from the above-mentioned occurrence of Asian soybean rust, as responsible agent for reducing the maturation cycle of the evaluated progenies.

#### ACKNOWLEDGEMENTS

The Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) supported this study with funds and the technician Geraldo Mangela of Assis helped us with the realization of the experiment.

**Table 3.** Estimates of the direct and indirect effects of secondary traits on grain yield. by means of the ridge path analysis in a F<sub>5</sub> soybean population.

Description of the effects	Primary components			Total
	NS	W1S	NP	
<b>Direct secondary effect of NDF</b>	0.072	0.042	0.363	0.477
Indirect effect via NDM	-0.214	-0.214	-0.357	-0.785
via PHM	-0.159	-0.238	-0.355	0.752
via IHP	0.078	0.108	0.255	0.441
via NN	0.055	0.009	-0.068	-0.004
<b>Total</b>	<b>-0.168</b>	<b>-0.293</b>	<b>-0.162</b>	<b>-0.623</b>
<b>Direct secondary effect of NDM</b>	-0.264	-0.264	-0.441	-0.969
Indirect effect via NDF	0.058	0.034	0.295	0.387
via PHM	-0.083	-0.124	-0.185	-0.392
via IHP	0.111	0.153	0.360	0.624
via NN	0.046	0.007	-0.057	-0.004
<b>Total</b>	<b>-0.132</b>	<b>-0.194</b>	<b>-0.028</b>	<b>-0.354</b>
<b>Direct secondary effect of PHM</b>	0.247	0.371	0.552	1.170
Indirect effect via NDF	-0.046	-0.027	-0.234	-0.307
via NDM	0.088	0.088	0.147	0.323
via IHP	-0.072	-0.100	-0.235	-0.407
via NN	-0.079	-0.013	0.097	0.005
<b>Total</b>	<b>0.138</b>	<b>0.319</b>	<b>0.327</b>	<b>0.784</b>
<b>Direct secondary effect of IHP</b>	0.108	0.150	0.353	0.611
Indirect effect via NDF	0.052	0.030	0.262	0.344
via NDM	-0.270	-0.269	-0.449	-0.988
via PHM	-0.164	-0.246	-0.367	-0.777
via NN	0.060	0.010	-0.074	-0.004
<b>Total</b>	<b>-0.214</b>	<b>-0.325</b>	<b>-0.275</b>	<b>-0.814</b>
<b>Direct secondary effect of NN</b>	-0.076	-0.012	0.093	0.005
Indirect effect via NDF	-0.053	-0.031	-0.266	-0.350
via NDM	0.162	0.161	0.270	0.593
via PHM	0.257	0.386	0.575	1.218
via IHP	-0.086	-0.119	-0.280	0.485
<b>Total</b>	<b>0.204</b>	<b>0.385</b>	<b>0.392</b>	<b>0.981</b>

NDF = number of days to flowering; NDM = number of days to maturation; PHM = plant height at maturity; IHP = insertion height of the first pod; NN = number of nodes; NS = number of seeds per pod; W1S = weight of one seed and NP = number of pods per plant

**Table 4.** Heritability estimates at the level of family means ( $h_i^2$ ), original means of the grain yield ( $X_{oi}$ ) and families selected for grain yield via auxiliary trait ( $X_{si}$ ), direct ( $DS_i$ ) and indirect ( $DS_{j(i)}$ ) selection differential. Gains with direct ( $GS_i$ ) and indirect ( $GS_{j(i)}$ ) selection through the estimation methodology based on the selection differential, gains in percentage ( $GS\%$ ) and efficacy of indirect selection (EIS) in an  $F_5$  soybean population

Auxiliary trait	Main trait (GY)						
	$h_i^2$ or $h_j^2$	$X_{oi}$	$X_{si}$	$DS_i$ or $DS_{j(i)}$	$GS_i$ or $GS_{j(i)}$	$GS\%$	EIS
GY	62.921	13.486	17.763	4.277	2.691	19.954	-
NDF	78.674	13.486	16.181	2.695	1.696	12.576	0.630
NDM	89.872	13.486	15.584	2.098	1.320	9.788	0.491
PHM	53.448	13.486	15.334	1.848	1.163	8.624	0.432
IHP	55.793	13.486	10.199	-3.287	-2.068	-15.334	-0.768
NN	83.290	13.486	16.572	3.086	1.942	14.400	0.722
NP	55.689	13.486	17.110	3.624	2.280	16.906	0.847
NS	52.933	13.486	15.777	2.291	1.4421	0.693	0.536
W1S	73.3281	3.4861	6.564	3.078	1.9371	4.361	0.720

GY = grain yield; NDF = number of days to flowering; NDM = number of days to maturation; PHM = plant height at maturity; IHP = insertion height of the first pod; NN = number of nodes; NP = number of pods per plant; NS = number of seeds per pod and W1S = weight of one seed.

**Table 5.** Estimates of heritability at the level of family means ( $h_i^2$ ); genetic correlations ( $r_g$ ) between grain yield and other traits; genotypic variances ( $\sigma_G^2$ ); standardized selection differential (K); yield gains due to direct and indirect selection through the methodology of estimation based on selection intensity (GS); gains in percentage per selection ( $GS\%$ ) and efficacy of indirect selection (EIS), in an  $F_5$  soybean population

	Target trait								
	GY	NDF	NDM	PHM	IHP	NN	NS	NP	W1S
$h_i^2$ (%)	62.921	78.674	89.872	53.448	55.793	83.290	52.933	55.689	73.328
$h_i$	0.793	0.887	0.948	0.731	0.747	0.913	0.728	0.746	0.856
$r_g$	-	-0.623	-0.354	0.784	-0.814	0.981	0.150	0.921	0.908
$\sigma_G^2$	18.069	25.209	35.496	164.499	12.102	10.820	0.015	0.785	0.068
$\sigma_G$	4.251	5.021	5.958	12.826	3.479	3.289	0.122	0.886	0.261
K	0.800	0.800	0.800	0.800	0.800	0.800	0.800	0.800	0.800
GS	2.697	-2.220	1.600	5.880	-1.692	2.357	0.011	0.487	0.162
$GS\%$	19.999	-16.462	-11.864	43.600	-12.546	17.477	0.082	3.611	1.201
EIS	-	-0.823	-0.593	2.180	-0.627	0.874	0.004	0.181	0.060

GY = grain yield; NDF = number of days to flowering; NDM = number of days to maturation; PHM = plant height at maturity; IHP = insertion height of the first pod; NN = number of nodes; NS = number of seeds per pod; NP = number of pods per plant and W1S = weight of one seed.

## Análise de trilha e resposta esperada na seleção indireta para produtividade de grãos em soja

**RESUMO** - Este trabalho objetiva determinar através da análise de trilha, contornando os efeitos adversos da multicolinearidade e resposta esperada, o melhor caráter auxiliar a ser utilizado na seleção indireta para produtividade de grãos. Utilizaram-se 79 genótipos de soja, de geração  $F_5$ , do cruzamento FT-Cometa x Bossier, sendo que a distribuição da



população no campo foi a de famílias intercaladas com testemunhas. Avaliaram-se caracteres primários e secundários em quatro plantas fenotipicamente superiores por família e considerando os efeitos diretos sobre a produtividade de grãos, os caracteres número de vagens, altura e número de nós figuram como os mais importantes, apresentando a melhor combinação de efeito direto e correlação genotípica. O número de vagens proporcionou maior ganho esperado quando utilizou-se a metodologia de estimação baseada no diferencial de seleção. Por sua vez, a altura de plantas foi ineficiente como indicadora de planta mais produtiva, pela metodologia baseada na intensidade de seleção.

**Palavras-chave:** caracteres agronômicos, correlação genotípica, análise de trilha, colinearidade, seleção indireta.

## REFERENCES

- Belsley DA, Kuh E and Welch RE (1980) **Regression diagnostics: identifying data and sources of collinearity**. John Wiley & Sons, New York, 292p.
- Carvalho SP (1995) **Métodos alternativos de estimação dos coeficientes de trilha e índices de seleção sob multicolinearidade**. Editora UFV, Viçosa, 163p.
- Carvalho SP and Cruz CD (1996) Diagnosis of multicollinearity assessment of the condition of correlation matrices used in genetic studies. **Revista Brasileira de Genética** **19**: 479-484.
- Carvalho CGP, Oliveira VR, Cruz CD and Casali VWD (1999) Análise de trilha sob multicolinearidade em pimentão. **Pesquisa Agropecuária Brasileira** **34**: 603-613.
- Carvalho CGP, Arias CAA, Toledo JFF, Oliveira MF and Vello NA (2002) Correlações e análise de trilha em linhagens de soja semeadas em diferentes épocas. **Pesquisa Agropecuária Brasileira** **37**: 311-320.
- Cruz CD and Regazzi AJ (1994) **Modelos biométricos aplicados ao melhoramento genético**. Editora UFV, Viçosa, 390p.
- Cruz CD (2001) Programa Genes – Versão Windows – Aplicativo computacional em genética e estatística. **Editora UFV, Viçosa, 648p**.
- Cruz CD and Carneiro PCS (2003) **Modelos biométricos aplicados ao melhoramento genético**. Editora UFV, Viçosa, 585p.
- Ferrari F (1989) **Estimadores viesados para modelos de regressão em presença de multicolinearidade**. Editora USP-ESALQ, Piracicaba, 127p.
- Hoerl AE and Kennard RW (1970a) Ridge regression: applications to monorthogonal problems. **Technometrics** **12**: 69-82.
- Hoerl AE and Kennard RW (1970b) Ridge regression: biased estimation for monorthogonal problems. **Technometrics** **12**: 55-68.
- Johnson HW, Robinson HF and Comstock RE (1955) Genotypic and phenotypic correlations in soybeans and their implications in selection. **Agronomy Journal** **47**: 477-483.
- Kurek AJ, Carvalho FIF, Assmann IC, Marchioro VS and Cruz PJ (2001) Análise de trilha como critério de seleção indireta para rendimento de grãos em feijão. **Revista Brasileira de Agrociência** **7**: 29-32.
- Montgomery DC and Peck EA (1981) **Introduction to linear regression analysis**. John Wiley & Sons, New York, 504p.
- Reis EF, Reis MS, Cruz CD, Sediya T and Sediya CS (2001) Resposta esperada à seleção correlacionada em uma população  $F_6$  de soja. **Revista Ceres** **48**: 169-179.
- Santos CAF, Reis MS, Cruz CD, Sediya CS and Sediya T (1995) Adequação de modelos no estudo do coeficiente de trilha nos componentes primários e secundários de progênies  $F_6$  de soja (*Glycine max* (L.) Merrill). **Revista Ceres** **42**: 111-121.
- Schuster I (1996) **Correlações coeficientes de trilha, composição de gluteninas e qualidade do trigo para panificação**. Editora UFV, Viçosa, 98p.
- Vencovsky R and Barriga P (1992) **Genética biométrica no fitomelhoramento**. Revista Sociedade Brasileira de Genética, Ribeirão Preto, 486p.
- Wright S (1921) Correlation and causation. **Journal of Agricultural Research** **20**: 557-585