

# Number of trials to estimate the condition number in rye traits<sup>1</sup>

Ismael Mario Márcio Neu<sup>2\*</sup>, Alberto Cargnelutti Filho<sup>3</sup>, Daniela Lixinski Silveira<sup>2</sup>, Rafael Vieira Pezzini<sup>4</sup>,  
Cirineu Tolfo Bandeira<sup>5</sup>

**ABSTRACT** - Multicollinearity must be diagnosed in multivariate analyses. Among the indicators, the condition number can be used to quantify the degree of multicollinearity. Hence, this study sought to determine the number of measurements (trials) necessary to estimate the number of condition in linear correlation matrices between rye traits. Five uniformity trials were carried out with 'BRS Progresso' rye, and eight morphological traits and eight productive traits were evaluated, forming two groups. In each group of traits, six cases (combinations of traits) were planned and the multicollinearity diagnosis was performed. Repeatability analyses were performed using the following methods: analysis of variance, principal component analysis, and structural analysis, and the number of measurements (trials) was determined for different levels of precision. A higher condition number of repeatability coefficients was obtained by the principal component methods (based on correlation and variance and covariance matrices) and structural analysis based on the variance and covariance matrix. A greater number of measurements (trials) is necessary to estimate the number of conditions in productive traits compared to morphological ones. One trial is enough to efficiently estimate the condition number with a minimum accuracy of 80% in morphological and productive traits of rye, whereas at least three trials are required for 95% accuracy.

**Key words:** *Secale cereale* L. Repeatability analysis. Multicollinearity. Experimental planning.

---

DOI: 10.5935/1806-6690.20230046

Editor-in-Chief: Prof. Salvador Barros Torres - sbtorres@ufersa.edu.br

\*Author for correspondence

Received for publication 01/08/2022; approved on 25/01/2023

<sup>1</sup>Paper extracted from the thesis of the first author, presented to the graduate program in Agronomy of the Universidade Federal de Santa Maria (UFSM), Santa Maria-RS, Brasil. Study financed by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)

<sup>2</sup>Graduate Program in Agronomy, Universidade Federal de Santa Maria (UFSM), Santa Maria-RS, Brasil, ismaelmmneu@hotmail.com (ORCID ID 0000-0002-9186-2532); danilisl@gmail.com (ORCID ID 0000-0003-0993-0100)

<sup>3</sup>Department of Plant Science, Center for Rural Sciences, Universidade Federal de Santa Maria (UFSM), Santa Maria-RS, Brasil, alberto.cargnelutti.filho@gmail.com (ORCID ID 0000-0002-8608-9960)

<sup>4</sup>Agronomist and Master in Agronomy - Plant Production from the Universidade Federal de Santa Maria, Santa Maria-RS, Brasil, acting as Technical Sales Assistant, rypezzini@hotmail.com (ORCID ID 0000-0003-4134-2499)

<sup>5</sup>Agronomist from the Universidade Federal do Pampa, Master in Agronomy - Plant Production from the Universidade Federal de Santa Maria, Santa Maria-RS, Brasil, working as an Agronomist at Cotripal Agropecuária Cooperativa, cirineutolfobandeira@gmail.com (ORCID ID 0000-0001-5019-5959)

## INTRODUCTION

Multivariate analysis techniques allow researchers to better understand the phenomena of multiple measures of studied individuals. More reliable parameter estimates are obtained when assumptions are met, and in multivariate analyses, multicollinearity must be investigated (HAIR *et al.*, 2009). It can be understood as the linear relationship between traits, and when present at high levels, performance and parameter prediction decrease in most linear methods due to information sharing between the characteristics (DORMANN *et al.*, 2013). In most multivariate techniques, multicollinearity increases the variance of the estimated parameters (FIGUEIREDO FILHO *et al.*, 2011), resulting in parameter estimates of low reliability (HAIR *et al.*, 2009), overestimated statistics and excessive false positives (GOODHUE; LEWIS; THOMPSON, 2017), or even an inadequate interpretation of the results (ALVES; CARGNELUTTI FILHO; BURIN, 2017; DORMANN *et al.*, 2013; TOEBE; CARGNELUTTI FILHO, 2013).

The repeatability coefficient (*rc*) has been used to verify the correlation between measurements in the same individual and determine the number of measurements (ABEYWARDENA, 1972; CRUZ; REGAZZI; CARNEIRO, 2012). The repeatability analysis and determining the number of measurements have been performed in crops such as Tanzania grass (*Panicum maximum* Jacq.) (CARGNELUTTI FILHO *et al.*, 2004; FERNANDES *et al.*, 2017), elephant grass (*Pennisetum* spp.) (CAVALCANTE *et al.*, 2012; SOUZA *et al.*, 2017), wheat (*Triticum aestivum*) (PAGLIOSA *et al.*, 2014), palisade grass (*Urochloa brizantha*) (TORRES *et al.*, 2015), cabbage (*Brassica oleracea* var. *acephala*) (AZEVEDO *et al.*, 2016), strawberry (*Fragaria x ananassa*) (DIEL *et al.*, 2020), and soybean (*Glycine max*) (DUARTE; FERREIRA; SILVA, 2022). Given the above, this study sought to determine the number of measurements (trials) required to estimate the condition number in linear correlation matrices between rye traits.

## MATERIALS AND METHODS

Five uniformity trials (without applying treatments) were conducted with rye (*Secale cereale* L.); the cultivar utilized was 'BRS Progresso', which is intended for grain production (EMBRAPA, 2013). The experimental area belongs to the Department of Plant Science of the Federal University of Santa Maria (29°42'S, 53°49'W; 95 m altitude). The region climate is classified as humid subtropical Cfa with hot summers and no defined dry season, according to Köppen (ALVARES *et al.*, 2013). The soil

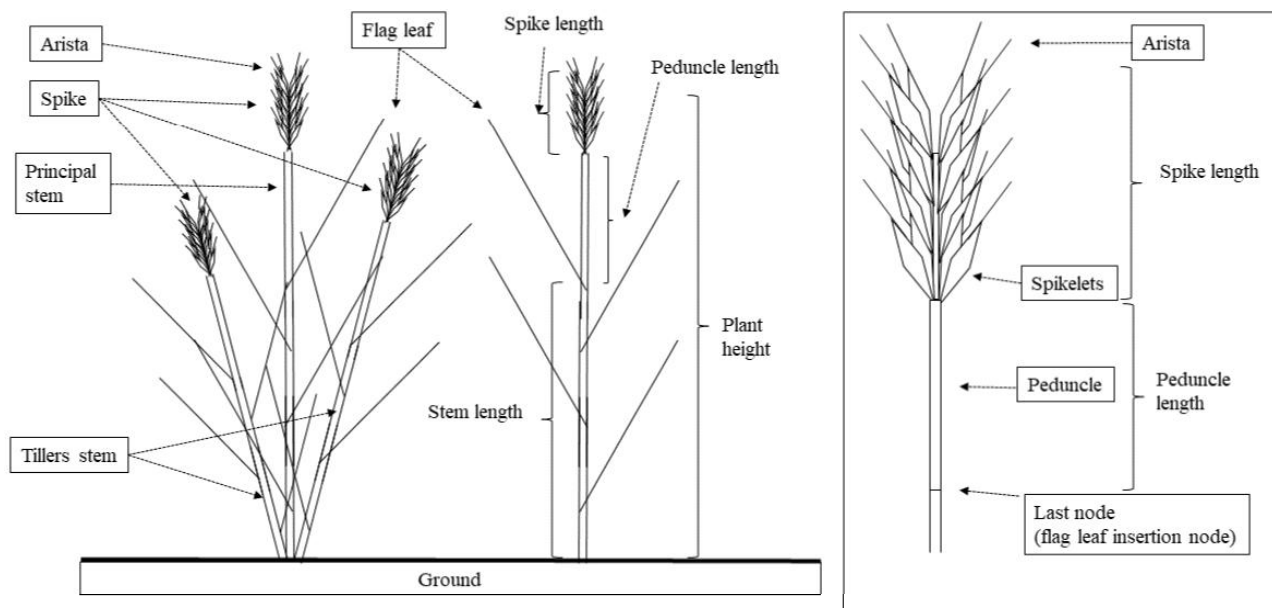
of the region is classified as Typical Dystrophic Bruno-gray Argisol (Argissolo Bruno-Acinzentado distrófico típico) (SANTOS *et al.*, 2018).

Conventional soil preparation was performed throughout the experimental area by harrowing. Soil fertility was corrected by applying 500 kg ha<sup>-1</sup> of fertilizer with the 5-20-20 formulation (NPK), corresponding to 25 kg ha<sup>-1</sup> of N, 100 kg ha<sup>-1</sup> of P<sub>2</sub>O<sub>5</sub>, and 100 kg ha<sup>-1</sup> of K<sub>2</sub>O.

Sowings in trials T1, T2, T3, T4, and T5 were performed on 03/05/2016, 25/05/2016, 07/06/2016, 22/06/2016, and 04/07/2016, respectively. Seeding was performed by broadcast seeding in a 320 m<sup>2</sup> area (20 × 16 m) in the first sowing season (T1). In contrast, in the other seasons, it was performed in a 375 m<sup>2</sup> area (25 × 15 m) at a density of 455 seeds m<sup>-2</sup>. Cover fertilization was performed when the plants were between the stages with three (V3) and four (V4) developed leaves using 25 kg ha<sup>-1</sup> of N. The other cultural treatments and management recommendations were carried out as needed for rye (BAIER, 1994).

Then, 100 plants were randomly collected and evaluated in each uniformity trial (sowing season), except in T4 (fourth season), in which 90 plants were evaluated, totaling 490 plants. The evaluations were performed on the stems of each plant collected (the primary stem and secondary stem or tiller), obtaining values for eight morphological and eight productive traits. In total, 1,136 stalks were evaluated (i.e., 193, 370, 242, 169, and 162 in T1, T2, T3, T4, and T5, respectively). The values were obtained by counting the number of nodes, spikelets, and spike grains<sup>-1</sup>; measuring the length of the stem, stalk, and spike (cm); and, weighing the fresh and dry mass of the stem, stalk, spike (grain and straw mass), and grain (g).

The following morphological traits were evaluated in each plant: 1) plant height (cm) obtained by the mean distance between the base of the plant to the last spikelet of all the stalks of the plant; 2) stem length (cm) obtained by the mean distance between the base of the plant until the flag leaf node of all the stalks of the plant; 3) peduncle length (cm) obtained by the mean distance between the flag leaf node and the spike insertion in the peduncle of all the stalks of the plant; 4) fresh mass of the aerial part (g) obtained by the mean mass of the aerial part of all the stalks of the plant; 5) total fresh mass of the aerial part (g) obtained by the sum of the mass of the aerial part of all the stalks of the plant; 6) the ratio between the mean of the fresh masses of stalk + leaves + peduncle on the total fresh mass of the aerial part; 7) the number of stalks obtained by the sum of the main stem + the number of tillers; and 8) number of nodes per stem obtained by dividing the number of nodes of the plant by the number of stalks.

**Figure 1** - Representation of a rye (*Secale cereale* L.) plant and details of the evaluated parts

The following productive traits were evaluated in each plant: 1) spike length (cm) obtained by the mean length of the spikes on the plant; 2) grain mass (g) obtained by summing the grain mass of all spikes on the plant; 3) 100-grain mass (g); 4) the number of grains obtained by summing the number of grains on all spikes on the plant; 5) the number of spike grains<sup>-1</sup> obtained by dividing the number of grains on the plant by the number of spikes on the plant; 6) the number of spikelets obtained by summing the number of spikelets in all spikes on the plant; 7) number of spike spikelets<sup>-1</sup> obtained by dividing the number of spikelets on the plant by the number of spikes on the plant; and 8) the ratio of the mass of grains per stem to the total fresh mass of the aerial part.

Six cases were planned for each trait group (morphological and productive) and were formed by combinations of eight traits ( $p = 8$  traits) taken by  $p_i$  in  $p_i$  ( $C_{(p,p_i)}$  with  $i = 2, 3, 4, 5, 6,$  and  $7$  traits). That is, in each trait group, in the first case identified as case 2, 28 combinations of eight traits were taken as two by two ( $C_{(p,p_i)} = C_{(8,2)} = 28$  combinations). In the following cases, by adding one trait, combinations with three ( $C_{(8,3)}$ ), four ( $C_{(8,4)}$ ), and so forth were obtained until the last case with seven combined traits ( $C_{(8,7)}$ ). Therefore, 28, 56, 70, 56, 28, and 8 combinations were obtained for the cases containing 2, 3, 4, 5, 6, and 7 traits, respectively. A total of 492 combinations were obtained, with 246 combinations belonging to the cases for the morphological trait group and another 246 combinations belonging to the cases for the productive trait group.

Next, the condition number ( $CN$ ) estimates were obtained for each combination within each case, trait group, and trial. The  $CN$  was obtained by the ratio of the largest ( $\lambda_{\max}$ ) and lowest eigenvalue ( $\lambda_{\min}$ ) of Pearson's linear correlation matrix between the traits (CRUZ; REGAZZI; CARNEIRO, 2012; GUJARATI; PORTER, 2011). As a rule of thumb, the  $CN$  indicator divides multicollinearity into classes: weak ( $CN \leq 100$ ); moderate to strong ( $100 < CN \leq 1,000$ ); and severe ( $CN > 1,000$ ) (MONTGOMERY *et al.*, 2012).

Repeatability analysis for  $CN$  was performed in each case (combined traits) and trait group (morphological and productive), totaling 12 repeatability analyses (six cases  $\times$  two trait groups). The different combinations of traits within each case and trait group were considered to be the observed "subjects" and the trials (sowing seasons) the "repeated measures".

Considering the example of estimating the repeatability coefficient ( $rc$ ) of  $CN$  in case 2 and the morphological trait group, the 140  $CN$  estimates (28 combinations of eight traits taken two by two  $\times$  five trials) were considered, forming a matrix of 28 rows (combinations) and 5 columns (trials). The same number of estimates was obtained for the productive trait group because it also presents eight traits. Therefore, for each character group, 140, 280, 350, 280, 140, and 40  $CN$  estimates were obtained for the cases with 2, 3, 4, 5, 6, and 7 combined traits, respectively.

For each of the cases with 2, 3, 4, 5, 6, and 7 combined traits and trait groups (morphological and productive), the  $rc$  and coefficient of determination

( $R^2$ ) were estimated by analysis of variance (ANOVA), principal components based on the correlation matrix (PCR), principal components based on variance and covariance matrix (PCS), structural analysis based on the theoretical eigenvalue of the correlation matrix (SAR), and structural analysis determined based on the theoretical eigenvalue of the variance and covariance matrix (SAS) (CRUZ; REGAZZI; CARNEIRO, 2012).

In the ANOVA method, the model was considered:

$$CN_{ij} = m + C_i + E_j + \varepsilon_{ij} \quad (1)$$

Where:  $CN_{ij}$  is the estimate of the condition number referring to the  $i$ -th combination and the  $j$ -th trial;  $m$  is the general mean;  $C_i$  is the effect of the  $i$ -th combination under the influence of the trial;  $E_j$  is the effect of the trial at the  $j$ -th measurement; and  $\varepsilon_{ij}$  is the experimental error established by the effects of the  $j$ -th trial in the  $i$ -th combination.

The mean  $rc$  and  $R^2$  of the cases were compared between trait groups (morphological and productive) within each method by the Student's  $t$ -test for independent samples at a 5% significance level.

For each case, method, and trait group, the number of measurements or trials ( $\eta_m$ ) to estimate the condition number with different determination coefficients ( $R^2 = 0.80, 0.85, 0.90, 0.95, \text{ and } 0.99$ ) were determined using the equation below (CRUZ; REGAZZI; CARNEIRO, 2012):

$$\eta_m = \frac{R^2(1-rc)}{(1-R^2)rc} \quad (2)$$

where:  $\eta_m$  is the number of measurements (trials),  $rc$  is the repeatability coefficient, and  $R^2$  is the coefficient of determination ( $R^2 = 0.80, 0.85, 0.90, 0.95, \text{ and } 0.99$ ). The means of  $\eta_m$  of the cases were compared between the trait groups (morphological and productive) within each method and  $R^2$  by the Student's  $t$ -test for independent samples at a 5% significance level. The analyses were performed using Microsoft Excel® and R software (R CORE TEAM, 2021).

## RESULTS AND DISCUSSION

The number of combined traits provided different estimates for the condition number ( $CN$ ), with the highest values in cases with the highest number of combined traits (Table 1). In combinations with two traits (case 2), weak multicollinearity was observed, with means of 2.61 ( $1.0 \leq CN \leq 16.6$ ) and 5.85 ( $1.0 \leq CN \leq 78.3$ ) in the morphological and productive trait groups, respectively. In case 7, nevertheless, the  $CN$  means were roughly 163 and 80 times higher than the means in case 2, with values of 426.78 ( $75.6 \leq CN \leq 1,067.5$ ) and 468.25 ( $144.7 \leq CN \leq 1,170.8$ ), respectively. The

percentage of combinations with  $CN \leq 100$  decreased toward the cases with a higher number of combined traits. An extreme case was observed in case 7 (seven traits combined) and the productive trait group, with no combination with  $CN \leq 100$ .

Regardless of the trait group, we observed that, on average, the  $CN$  increased as more traits were used in the multicollinearity diagnosis. Additionally, the amplitudes were larger in the cases with more combined traits. This may be related to using all possible combinations in each case. The greater the number of traits present in the group under analysis, the greater the chance of strongly related traits. In an extreme case, high  $CN$  ( $CN \geq 144.7$ ) was verified in all combinations of productive traits in case 7. Multicollinearity can be understood as the linear relationship between traits and information sharing (DORMANN *et al.*, 2013). The high magnitude of the correlation between traits can be used to indicate the presence of high multicollinearity levels, and researchers must pay more attention to when correlations are above  $|r| \geq 0.7$  (DORMANN *et al.*, 2013).

Estimates of  $CN > 100$  are classified as moderate to strong or severe multicollinearity (MONTGOMERY *et al.*, 2012). In such cases, trait elimination has been used as a measure to reduce the degree of multicollinearity, and this practice has been reported elsewhere with maize (*Zea mays*) (ALVES *et al.*, 2016a, 2016b; OLIVOTO *et al.*, 2017; TOEBE *et al.*, 2017; TOEBE; CARGNELUTTI FILHO, 2013), showy crotalaria (*Crotalaria spectabilis*) (TOEBE *et al.*, 2017), and black oats (*Avena strigosa* S.) (MEIRA *et al.*, 2019).

Based on five trials, repeatability coefficients ( $rc$ ) equal to or greater than 0.707 were observed for  $CN$  estimation, with a minimum accuracy of 92.4% in predicting its real value ( $R^2 = 0.924$ ), regardless of the number of combined traits (cases), trait group, and method of repeatability (Table 2). A high value of  $R^2$  indicates that the mathematical model used to determine repeatability was efficient (CAVALCANTE *et al.*, 2012).

Regardless of the method and trait group, the mean  $R^2$  of the cases ranged between 0.959 and 0.997. These high  $R^2$  values indicate that all methods accurately estimated  $rc$ . Differences between the mean  $R^2$  of the cases among the trait groups were verified by the Student's  $t$ -test for independent samples at a 5% significance level when the repeatability analysis was performed using PCR, PCS, and SAR. Hence, these methods gave us higher accuracy to estimate  $CN$  repeatability observed in morphological traits compared to productive ones. No studies of repeatability analysis have been found for the rye crop nor for estimating  $CN$ . Similar high values and magnitudes of  $R^2$  for the same trait were observed

**Table 1** - Minimum, mean, median, maximum, and range (maximum-minimum) of the condition number and percentage of combinations with weak multicollinearity (PCWM) in combinations of morphological and productive traits (cases) in five trials of ‘BRS Progresso’ rye (*Secale cereale* L.)

Case <sup>1</sup>	NCT <sup>2</sup>	No. of Trials	No. of Comb. <sup>3</sup>	Minimum	Mean	Median	Maximum	Range	PCWM (%) <sup>4</sup>
----- Morphological traits -----									
2	28	5	140	1.0	2.61	1.45	16.6	15.6	100.00
3	56	5	280	1.1	11.07	3.20	377.1	376.0	98.21
4	70	5	350	1.5	37.30	6.43	853.0	851.6	92.00
5	56	5	280	2.7	100.20	22.52	965.2	962.5	76.79
6	28	5	140	3.6	223.81	94.05	1043.7	1,040.1	50.71
7	8	5	40	75.6	426.78	352.37	1067.5	991.9	20.00
----- Productive traits -----									
2	28	5	140	1.0	5.85	2.25	78.3	77.3	100.00
3	56	5	280	1.5	18.95	6.61	170.3	168.9	97.14
4	70	5	350	2.6	50.57	27.67	519.5	516.9	85.43
5	56	5	280	7.1	120.00	69.54	723.0	716.0	49.29
6	28	5	140	23.6	251.10	198.46	1008.5	984.9	24.29
7	8	5	40	144.7	468.25	384.32	1170.8	1,026.2	0.00

<sup>1</sup>Case: number of traits combined. <sup>2</sup>NCT: number of combinations in each case and trial. <sup>3</sup>No. of Comb. = number of combinations in each case × number of trials. <sup>4</sup>Percentage of combinations with weak multicollinearity (PCWM):  $CN \leq 100$  (MONTGOMERY *et al.*, 2012)

**Table 2** - Coefficients of repeatability (*rc*) and determination (*R*<sup>2</sup>) for the condition number (*CN*) obtained in different cases, trait groups (morphological [MORP] and productive [PROD]), and by different methods based on five trials of ‘BRS Progresso’ rye (*Secale cereale* L.)

Case <sup>1</sup>	ANOVA <sup>2</sup>		PCR		PCS		SAR		SAS	
	MORP	PROD	MORP	PROD	MORP	PROD	MORP	PROD	MORP	PROD
<i>rc</i> - repeatability coefficient										
2	0.923	0.817	0.948	0.885	0.954	0.872	0.948	0.882	0.923	0.817
3	0.859	0.849	0.983	0.879	0.991	0.873	0.983	0.877	0.859	0.849
4	0.838	0.885	0.979	0.917	0.988	0.914	0.979	0.916	0.838	0.885
5	0.825	0.880	0.980	0.933	0.989	0.932	0.980	0.933	0.825	0.880
6	0.812	0.831	0.983	0.937	0.992	0.942	0.983	0.937	0.812	0.831
7	0.792	0.707	0.988	0.961	0.994	0.969	0.988	0.961	0.792	0.707
Mean	0.842 a <sup>3</sup>	0.828 a	0.977 a	0.918 b	0.985 a	0.917 b	0.977 a	0.918 b	0.842 a	0.828 a
<i>R</i> <sup>2</sup> - coefficient of determination										
2	0.984	0.957	0.989	0.975	0.991	0.972	0.989	0.974	0.984	0.957
3	0.968	0.966	0.997	0.973	0.998	0.972	0.997	0.973	0.968	0.966
4	0.963	0.975	0.996	0.982	0.998	0.981	0.996	0.982	0.963	0.975
5	0.959	0.973	0.996	0.986	0.998	0.986	0.996	0.986	0.959	0.973
6	0.956	0.961	0.997	0.987	0.998	0.988	0.997	0.987	0.956	0.961
7	0.950	0.924	0.998	0.992	0.999	0.994	0.998	0.992	0.950	0.924
Mean	0.963 a	0.959 a	0.996 a	0.982 b	0.997 a	0.982 b	0.996 a	0.982 b	0.963 a	0.959 a

<sup>1</sup>Case: number of combined traits. <sup>2</sup>Repeatability analysis methods: ANOVA: analysis of variance; PCR: principal components based on the correlation matrix; PCS: principal components based on variance and the covariance matrix; SAR: structural analysis based on the correlation matrix; SAS: structural analysis based on variance and the covariance matrix. <sup>3</sup>*rc* and *R*<sup>2</sup> means with distinct letters in the row (comparison between groups of morphological and productive traits within each method) differ from each other at a 5% significance by the Student’s t-test for independent samples with 10 degrees of freedom (exception for *R*<sup>2</sup> by the PCS method with 6.21 degrees of freedom)

among the methods used in repeatability analysis. As an example, similar magnitudes of  $R^2$  values were observed in the repeatability analysis of plant height in Tanzania grass ( $0.919 \leq R^2 \leq 0.970$ ) (CARGNELUTTI FILHO *et al.*, 2004), cabbage ( $0.994 \leq R^2 \leq 0.998$ ) (AZEVEDO *et al.*, 2016), dry mass of the aerial part in elephant grass ( $0.96 \leq R^2 \leq 0.97$ ) (CAVALCANTE *et al.*, 2012) and  $0.734 \leq R^2 \leq 0.798$  (SOUZA *et al.*, 2017), and soybean ( $0.942 \leq R^2 \leq 0.944$ ) (DUARTE; FERREIRA; SILVA, 2022).

When analyzing the  $rc$  estimates between cases in each trait group and method, no pattern was observed with the decreasing or increasing number of combined traits. However, lower  $rc$  magnitudes were only observed when estimated by ANOVA and SAS and in case 7. Similar to what was observed with the mean  $R^2$ , higher  $rc$  means were observed in the group of morphological traits compared to the group of productive traits according to the Student's t-test for independent samples at a 5% significance level when  $rc$  was obtained by PCR, PCS, and SAR.

In summary, the highest repeatability coefficients of condition number were estimated by principal component methods (based on the correlation and variance and covariance matrices – PCR and PCS, respectively) and structural analysis based on the variance and covariance matrix (SAS). The average estimates of the repeatability coefficient do not differ between the groups of morphological and productive characters by analysis of variance (ANOVA) and structural analysis based on the matrix of variances and covariances (SAS), but with a higher mean for the group of morphological traits when the coefficient is estimated by the principal components based on the correlation matrix (PCR), principal components based on variance and covariance matrix (PCS), and structural analysis based on the theoretical eigenvalue of the correlation matrix (SAR).

The present study obtained the highest repeatability values for  $CN$  estimation using principal component methods (PCR and PCS) and structural analysis based on the correlation matrix (SAR). These methods seem suitable for  $rc$  estimation for  $CN$  in rye traits. High repeatability estimates indicate that with a relatively small number of measurements, it is possible to estimate the true value of a given trait (CARGNELUTTI FILHO *et al.*, 2004). This is because the higher the  $rc$  estimate, the greater the predictability that values very close to the estimates of previous events will occur in subsequent measurements (CRUZ; REGAZZI; CARNEIRO, 2012).

Lower  $rc$  estimates by ANOVA were also observed in studies with agronomic traits of soybean (DUARTE; FERREIRA; SILVA, 2022; MATSUO *et al.*, 2012), elephant grass (CAVALCANTE *et al.*, 2012; SOUZA *et al.*, 2017), palisade grass (TORRES *et al.*, 2015), cabbage (AZEVEDO *et al.*, 2016), and strawberry (DIEL *et al.*, 2020). By using the principal component methods (PCR and PCS), the highest magnitudes of  $rc$  were verified in studies with agronomic characteristics in Tanzania grass (CARGNELUTTI FILHO *et al.*, 2004), wheat (PAGLIOSA *et al.*, 2014), palisade grass (TORRES *et al.*, 2015), and strawberries (DIEL *et al.*, 2020).

The minimum number of measurements or trials ( $\eta_m$ ) for  $CN$  estimation varied according to the method, the case (number of traits combined), the level of precision ( $R^2$  - coefficient of determination), and the trait group (morphological and productive) (Table 3). By comparing the  $\eta_m$  mean of cases between morphological and productive traits in each method and  $R^2$ , we observed the significant superiority of  $\eta_m$  in the group of productive traits using the PCR, PCS, and SAR methods in all  $R^2$  ( $R^2 = 0.80, 0.85, 0.90, 0.95, \text{ and } 0.99$ ) according to the Student's t-test for independent samples at a 5% significance level.

**Table 3** - Number of trials (measurements) associated with different coefficients of determination ( $R^2 = 0.80, 0.85, 0.90, 0.95, \text{ and } 0.99$ ) for estimating condition number ( $CN$ ) on combinations of traits (morphological and productive) in rye (*Secale cereale* L.)

Case <sup>1</sup>	Morphological traits					Productive traits				
	Coefficient of determination ( $R^2$ )									
	0.80	0.85	0.90	0.95	0.99	0.80	0.85	0.90	0.95	0.99
	ANOVA - Analysis of variance									
2	0.3	0.5	0.8	1.6	8.3	0.9	1.3	2.0	4.3	22.2
3	0.7	0.9	1.5	3.1	16.3	0.7	1.0	1.6	3.4	17.6
4	0.8	1.1	1.7	3.7	19.1	0.5	0.7	1.2	2.5	12.9
5	0.8	1.2	1.9	4.0	21.0	0.5	0.8	1.2	2.6	13.5
6	0.9	1.3	2.1	4.4	22.9	0.8	1.1	1.8	3.9	20.1
7	1.1	1.5	2.4	5.0	26.0	1.7	2.3	3.7	7.9	40.9
Mean	0.76 a <sup>2</sup>	1.08 a	1.72 a	3.63 a	18.92 a	0.86 a	1.21 a	1.93 a	4.07 a	21.21 a

Continuation Table 3

----- PCR - Principal components based on the correlation matrix -----										
2	0.2	0.3	0.5	1.0	5.4	0.5	0.7	1.2	2.5	12.9
3	0.1	0.1	0.2	0.3	1.7	0.6	0.8	1.2	2.6	13.7
4	0.1	0.1	0.2	0.4	2.1	0.4	0.5	0.8	1.7	9.0
5	0.1	0.1	0.2	0.4	2.0	0.3	0.4	0.6	1.4	7.1
6	0.1	0.1	0.2	0.3	1.7	0.3	0.4	0.6	1.3	6.7
7	0.0	0.1	0.1	0.2	1.2	0.2	0.2	0.4	0.8	4.0
Mean	0.10 b	0.13 b	0.21 b	0.45 b	2.35 b	0.36 a	0.51 a	0.81 a	1.71 a	8.90 a
----- PCS – Principal components based on the covariance matrix -----										
2	0.2	0.3	0.4	0.9	4.7	0.6	0.8	1.3	2.8	14.5
3	0.0	0.1	0.1	0.2	0.9	0.6	0.8	1.3	2.8	14.4
4	0.0	0.1	0.1	0.2	1.2	0.4	0.5	0.8	1.8	9.3
5	0.0	0.1	0.1	0.2	1.1	0.3	0.4	0.7	1.4	7.2
6	0.0	0.0	0.1	0.2	0.8	0.2	0.3	0.6	1.2	6.1
7	0.0	0.0	0.1	0.1	0.6	0.1	0.2	0.3	0.6	3.1
Mean	0.07 b	0.09 b	0.14 b	0.30 b	1.56 b	0.37 a	0.52 a	0.83 a	1.75 a	9.10 a
----- SAR – Structural analysis based on the correlation matrix -----										
2	0.2	0.3	0.5	1.0	5.5	0.5	0.8	1.2	2.5	13.2
3	0.1	0.1	0.2	0.3	1.7	0.6	0.8	1.3	2.7	13.9
4	0.1	0.1	0.2	0.4	2.1	0.4	0.5	0.8	1.7	9.1
5	0.1	0.1	0.2	0.4	2.0	0.3	0.4	0.6	1.4	7.1
6	0.1	0.1	0.2	0.3	1.7	0.3	0.4	0.6	1.3	6.7
7	0.0	0.1	0.1	0.2	1.2	0.2	0.2	0.4	0.8	4.0
Mean	0.10 b	0.13 b	0.21 b	0.45 b	2.36 b	0.36 a	0.51 a	0.82 a	1.73 a	9.00 a
----- SAS - Structural analysis based on the covariance matrix -----										
2	0.3	0.5	0.8	1.6	8.3	0.9	1.3	2.0	4.3	22.2
3	0.7	0.9	1.5	3.1	16.3	0.7	1.0	1.6	3.4	17.6
4	0.8	1.1	1.7	3.7	19.1	0.5	0.7	1.2	2.5	12.9
5	0.8	1.2	1.9	4.0	21.0	0.5	0.8	1.2	2.6	13.5
6	0.9	1.3	2.1	4.4	22.9	0.8	1.1	1.8	3.9	20.1
7	1.1	1.5	2.4	5.0	26.0	1.7	2.3	3.7	7.9	40.9
Mean	0.76 a	1.08 a	1.72 a	3.63 a	18.92 a	0.86 a	1.21 a	1.93 a	4.07 a	21.21 a

<sup>1</sup>Case: number of combined traits. <sup>2</sup>Means of the number of trials (measurements) with different letters in the row (comparison between groups of morphological and productive traits within each method and level of accuracy -  $R^2$ ) differ at a 5% significance level according to the Student's t-test for independent samples with 10 degrees of freedom (except in the PCS method with 6.15 degrees of freedom)

Regardless of the method, case, and character group, one trial ( $\eta_m = 1$  trial) is enough to estimate the  $CN$  with at least 80% accuracy except in combinations with seven traits (case 7) when the  $rc$  was estimated by the ANOVA or SAS methods. The lowest  $\eta_m$  values were observed for the cases, methods, and trait groups with the highest  $rc$  because the coefficient is used to determine  $\eta_m$  (Equation 2) and, therefore, the lowest  $\eta_m$  were estimated by the PCR, PCS, and SAR methods. The principal component methods

(PCR and PCS) have been used in repeatability analysis because, in most cases, higher  $rc$  and high accuracy are observed. In these methods, the cyclic behavior of the trait is considered, containing the eigenvector elements of the same sign and similar magnitudes, expressing the tendency of the genotypes (in this study, the combinations) to maintain their positions in successive measurements, thereby being recommended in repeatability estimation because of the higher accuracy (ABEYWARDENA, 1972).

Given the recommendation to use principal component methods and the high  $rc$  and  $R^2$  values obtained in this study, the principal component methods were used in inferences for the number of trials ( $\eta_m$ ) to estimate  $CN$ . It should be emphasized that, in this study, the estimates of  $\eta_m$  were similar to each other by the PCR, PCS, and SAR methods (Table 3).

Using the PCR, PCS, and SAR implies that a single assay is enough to estimate  $CN$  in morphological traits with at least 95% accuracy. In productive traits and the same level of accuracy (minimum  $R^2$  of 0.95), up to three assays are needed depending on the number of traits. The number of trials obtained in this study is lower than those reported in other crops as necessary to evaluate agronomic traits. Three to 12 evaluation cycles were found to be necessary (CAVALCANTE *et al.*, 2012) and 11 to 49 measurements in elephant grass genotypes (SOUZA *et al.*, 2017), two evaluations in wheat (PAGLIOSA *et al.*, 2014), and two to 18 measurements in soybean (MATSUO *et al.*, 2012), with 95% accuracy.

Fewer trials can be used, although the researcher must give up accuracy. Therefore, a single trial to estimate  $CN$  with 80% accuracy can be used in almost all cases and trait groups. The exception occurs in cases with seven morphological and productive traits and when the  $rc$  of the  $CN$  is determined by ANOVA and SAS methods.

It is up to the researcher to choose the adequate number of measurements (trials), considering the availability of material, manpower, and the desired precision. When defining the number of trials, the results of previous experiments and studies of sample size, plot size, relationships between traits, multicollinearity diagnosis, and other information about the crop must be considered.

Using a greater number of traits may result in greater predictability in  $CN$  estimation. On the other hand, it may result in lower precision in the estimates of the parameters of multivariate analysis because the researcher must be aware that the use of a higher number of traits will also lead to a higher degree of multicollinearity, requiring some procedure to reduce  $CN$  to values below 100. In most multivariate techniques, parameter estimates become unreliable in the presence of multicollinearity (HAIR *et al.*, 2009) or there is a misinterpretation of the results (ALVES; CARGNELUTTI FILHO; BURIN, 2017; DORMANN *et al.*, 2013; TOEBE; CARGNELUTTI FILHO, 2013); this is because the information is shared among the traits, consequently increasing the variance of the estimated parameters (FIGUEIREDO FILHO *et al.*, 2011).

A larger number of trials is needed to estimate  $CN$  in productive traits compared to the group of morphological traits, although using different numbers of

trials is not practical. Thus, in conducting experiments in rye or any other crop, a single number of trials facilitates the planning and experimental conduct. Adopting the highest  $\eta_m$  value enables minimum precision to be obtained, regardless of the trait group.

For the morphological and productive traits of 'BRS Progresso' rye, a single trial is enough to estimate the  $CN$  with 80% accuracy except for the case with seven combined traits and when repeatability analysis is performed using ANOVA and SAS methods. When one seeks to obtain higher accuracy values, at least three trials are required to estimate  $CN$  with 95% accuracy, regardless of the number of traits and trait group.

## CONCLUSIONS

1. Fewer trials are needed for the cases with a higher number of combined traits. However, the larger the number of traits, the larger the condition number estimate will also be;
2. One trial is enough to estimate the condition number with at least 80% accuracy in morphological and productive traits of rye. At least three trials are necessary for 95% accuracy.

## ACKNOWLEDGMENTS

To the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) - Processos 304652/2017-2 and 146258/2019-3, for granting scholarships to the authors. We would also like to thank Atlas Assessoria Linguística for language editing.

## REFERENCES

- ABEYWARDENA, V. An application of principal component analysis in genetics. **Journal of Genetics**, v. 61, n. 1, p. 27-51, 1972.
- ALVARES, C. A. et al. Köppen's climate classification map for Brazil. **Meteorologische Zeitschrift**, v. 22, n. 6, p. 711-728, 2013.
- ALVES, B. M. *et al.* Correlações canônicas entre caracteres agrônomicos e nutricionais proteicos e energéticos em genótipos de milho. **Revista Brasileira de Milho e Sorgo**, v. 15, n. 2, p. 171-185, 2016b.
- ALVES, B. M. *et al.* Linear relations among phenological, morphological, productive and protein-nutritional traits in early maturing and super-early maturing maize genotypes. **Journal of Cereal Science**, v. 70, p. 229-239, 2016a.



- ALVES, B. M.; CARGNELUTTI FILHO, A.; BURIN, C. Multicollinearity in canonical correlation analysis in maize. **Genetics and Molecular Research**, v. 16, n. 1, p. 1-14, 2017.
- AZEVEDO, A. M. *et al.* Estudo da repetibilidade genética em clones de couve. **Horticultura Brasileira**, v. 34, n. 1, p. 54-58, 2016.
- BAIER, A. C. **Centeio**. Passo Fundo: Embrapa Trigo, 1994. Disponível em: <http://ainfo.cnptia.embrapa.br/digital/bitstream/item/164511/1/FL-06193.pdf>. Acesso em: 20 set. 2019.
- CARGNELUTTI FILHO, A. *et al.* Análise de repetibilidade de caracteres forrageiros de genótipos de *Panicum maximum*, avaliados com e sem restrição solar. **Ciência Rural**, v. 34, n. 3, p. 723-729, 2004.
- CAVALCANTE, M. *et al.* Coeficiente de repetibilidade e parâmetros genéticos em capim-elefante. **Pesquisa Agropecuária Brasileira**, v. 47, n. 4, p. 569-575, 2012.
- CRUZ, C. D.; REGAZZI, A. J.; CARNEIRO, P. C. S. **Modelos biométricos aplicados ao melhoramento genético vegetal**. 4. ed. Viçosa, MG: Editora UFV, 2012. v. 1, 514 p.
- DIEL, M. I. *et al.* Repeatability coefficients and number of measurements for evaluating traits in strawberry. **Acta Scientiarum. Agronomy**, v. 42, n. e43357, p. 1-9, 2020.
- DORMANN, C. F. *et al.* Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. **Ecography**, v. 36, n. 1, p. 27-46, 2013.
- DUARTE, A. B.; FERREIRA, D. de O.; SILVA, F. L. da. Repeatability and the optimal number of measurements for screening of soybean cultivars under water deficit. **Revista Ciência Agronômica**, v. 53, n. 2, p. 1-13, 2022.
- EMBRAPA. **Centeio**: BRS Progresso. Passo Fundo, 2013. Disponível em: <https://www.embrapa.br/trigo/busca-de-solucoes-tecnologicas/-/produto-servico/1969/centeo---brs-progresso>.
- FERNANDES, F. D. *et al.* Repeatability, number of harvests, and phenotypic stability of dry matter yield and quality traits of *Panicum maximum* Jacq. **Acta Scientiarum. Animal Sciences**, v. 39, n. 2, p. 149-155, 2017.
- FIGUEIREDO FILHO, D. *et al.* O que fazer e o que não fazer com a regressão: pressupostos e aplicações do modelo linear de mínimos quadrados ordinários (MQO). **Revista Política Hoje**, v. 20, n. 1, p. 44-99, 2011.
- GOODHUE, D. L.; LEWIS, W.; THOMPSON, R. Multicollinearity and measurement error statistical blind spot: correcting for excessive false positives in regression and PLS. **MIS Quarterly**, v. 41, n. 3, p. 667-684, 2017.
- GUJARATI, D. N.; PORTER, D. C. **Econometria básica**. 5. ed. Porto Alegre: AMGH Editora, 2011. 920 p.
- HAIR, J. F. *et al.* **Análise multivariada de dados**. 6. ed. Porto Alegre: Bookman, 2009. 688 p.
- MATSUO, É. *et al.* Análise da repetibilidade em alguns descritores morfológicos para soja. **Ciência Rural**, v. 42, n. 2, p. 189-196, 2012.
- MEIRA, D. *et al.* Multivariate analysis revealed genetic divergence and promising traits for indirect selection in black oat. **Revista Brasileira de Ciências Agrárias**, v. 14, n. 4, p. 1-7, 2019.
- MONTGOMERY, D. C. *et al.* **Introduction to linear regression analysis**. 5. ed. New Jersey: John Wiley & Sons, 2012. 672 p.
- OLIVOTO, T. *et al.* Multicollinearity in path analysis: a simple method to reduce its effects. **Agronomy Journal**, v. 109, n. 1, p. 131-142, 2017.
- PAGLIOSA, E. S. *et al.* Repeatability of pre-harvest sprouting in wheat. **American Journal of Plant Sciences**, v. 5, n. 11, p. 1607-1613, 2014.
- R CORE TEAM. **R**: a language and environment for statistical computing. Vienna, Áustria: R Foundation for Statistical Computing, 2021. Disponível em: <https://www.r-project.org/>. Acesso em: 10 ago. 2021.
- SANTOS, H. G. dos *et al.* **Sistema brasileiro de classificação de solos**. 5. ed. Brasília, DF: Embrapa: Embrapa Solos, 2018. 590 p.
- SOUZA, Y. P. de *et al.* Repeatability and minimum number of evaluations for morpho-agronomic characters of elephant-grass for energy purposes. **Revista Brasileira de Ciências Agrárias**, v. 12, n. 3, p. 391-397, 2017.
- TOEBE, M. *et al.* Dimensionamento amostral e associação linear entre caracteres de *Crotalaria spectabilis*. **Bragantia**, v. 76, n. 1, p. 45-53, 2017.
- TOEBE, M.; CARGNELUTTI FILHO, A. Não normalidade multivariada e multicolinearidade na análise de trilha em milho. **Pesquisa Agropecuária Brasileira**, v. 48, n. 5, p. 466-477, 2013.
- TORRES, F. E. *et al.* Minimum number of measurements for accurate evaluation of qualitative traits in *Urochloa brizantha*. **Journal of Agronomy**, v. 14, n. 3, p. 180-184, 2015.



This is an open-access article distributed under the terms of the Creative Commons Attribution License