Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Precision gain and sample size: use of precision statistics for soybean early-cycle traits

Rafael Rodrigues de Souza

rafael.r.de.s@gmail.com

Federal University of Santa Maria (UFSM)

Karina Chertok Bittencourt

Federal University of Santa Maria (UFSM)

Anderson Carvalho Mello

Federal University of Pelotas (UFPel)

Rafael Barbieri Bellé

Federal University of Santa Maria (UFSM)

Marcos Toebe

Federal University of Santa Maria (UFSM)

Research Article

Keywords: coefficient of variation, experimental planning, Glycine max, least significant difference

Posted Date: November 11th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11662473/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Precision gain and sample size: use of precision statistics for soybean early-cycle traits

Rafael Rodrigues de Souza[a,*], Karina Chertok Bittencourt[b], Anderson de Carvalho Mello[c], Rafael Barbieri Bellé[b], Marcos Toebe[b]

[a] Department of Plant Science, Federal University of Santa Maria (UFSM), Santa Maria, RS, Brazil. E-mail: rafael.r.de.s@gmail.com

[b] Department of Agronomic and Environmental Sciences, Federal University of Santa Maria (UFSM), Frederico Westphalen, RS, Brazil. E-mail: m.toebe@gmail.com, kari.cher.bitt@gmail.com, rafael-barbieri1@hotmail.com

[c] Federal University of Pelotas (UFPel), Pelotas, RS, Brazil. E-mail: andersondecarvalhomello@gmail.com

* Corresponding author at: Avenida Roraima, nº 1000, Bairro Camobi, CEP 97105-900, Santa Maria, RS, Brazil. +55 (55) 99620-2991. E-mail address: rafael.r.de.s@gmail.com.

ABSTRACT

Empirical decisions on the number of samples to be collected per experimental unit may cause experiments to obtain misleading results. Once little attention has been given to soybean early-cycle traits, the aim of this study was to find the optimal sample size to estimate the overall experimental mean, coefficient of variation, and least significant difference for soybean early-cycle traits; and to assess precision gain as a function of sample size. Two experiments were performed in Erval Seco, Rio Grande do Sul, Brazil, with 30 soybean cultivars in a randomized block design with three repetitions. Thirty-nine days after emergence, 20 plants per experimental unit were measured, totaling 3,600 sampled plants. Then, sampling scenarios of 1, 2, ..., 100

plants were considered and the statistics overall experimental mean, coefficient of experimental variation, and least significant difference at 5% error probability were estimated in each scenario. In general, more precise estimates were made for the three precision statistics as a higher number of plants per experimental unit was sampled. However, precision gain tends to stabilize at ~70%. Therefore, a sample size of 17 plants per experimental unit was sufficient to estimate the experimental precision statistics reliably for soybean early-cycle traits.
KEYWORDS: coefficient of variation, experimental planning, *Glycine max*, least significant difference.

1. INTRODUCTION

Soybean crop [*Glycine max* (L.) Merr.], one of the most relevant commodities worldwide, has been the target of several kinds of research (Zhang et al. 2019; Souza et al. 2022; Zhang et al. 2022), from ecophysiological studies that focus on providing technical recommendations (Oikawa and Ainsworth 2016; Zanon et al. 2016) to the finding of superior genotypes through plant breeding programs (Li et al. 2020; Mwiinga et al. 2020; Souza et al. 2021). However, as pointed out by Souza et al. (2022), sampling has not been a major research topic, neither for soybean nor other crops, even though it represents one of the main and most important stages of experiments (Siegel 2012; Anderson et al. 2017). The lack of recommendations on the number of observations to be sampled may lead to empirical decision-making (Piñera-Chaves et al. 2020), which can later result in biased inferences (Maxwell et al. 2008; Anderson et al. 2017; Alvez-Silva et al. 2018).

For experimental research to be valid, it must be conducted at known and high precision levels (Cargnelutti Filho et al. 2009a; Górdon-Mendoza and Camargo-Buitrago 2015). In this sense, the sample size used for measuring traits in soybean and other crops needs to be adequate for the correct estimation of means, variability, and complementary statistics (Souza et al. 2022). Confalonieri et al. (2009) state that sampling insufficiency may contribute to the non-rejection of the initially-established basic hypothesis because of a high associated error, once it increases the rate of type II error (Maxwell et al. 2008; Siegel 2012; Anderson et al. 2017). On the other hand, the sampling of an excessive number of plants may harm the conduction and evaluation of experiments with a high number of treatments, not only because these require larger experimental areas but also due to the need for more manpower, time, and financial resources (Toebe et al. 2015; Bittencourt et al. 2022), making sample size recommendations crucial to maintain result reliability.

Defining sample size requires the selection of criteria that are associated with experimental bias. In that regard, Cargnelutti Filho et al. (2009a) highlighted the overall experimental mean, coefficient of variation, and least significant difference as directly associated with the experimental error in experiments with soybean, which was later corroborated by Górdon-Mendoza and Camargo-Buitrago (2015) for maize (*Zea mays* L.). In addition, Souza et al. (2022) verified such precision statistics are influenced by the sample size used. Nevertheless, although studies on experimental planning regarding sample size have been performed for soybean (Cargnelutti Filho et al. 2009b), including the analysis of the response of precision statistics to the number of plants sampled per experimental unit (Souza et al. 2022), an adequate sample size for soybean early-cycle traits has not been reported in the literature, especially focusing on

experimental precision statistics. This approach becomes relevant once sample size can also variate depending on the phenological stage of the plant (Confalonieri et al. 2006).

Early-cycle traits are extremely important for evaluating the vigor and early performance of soybean cultivars (Zhang et al. 2019; Liu et al. 2022; Zhang et al. 2022), for which reliable results are necessary to avoid misleading agronomical recommendations (Confalonieri et al. 2009; Anderson et al. 2017). However, in the absence of a standardized reference number of plants per experimental unit for this type of experiment, studies have been carried out with different sample sizes (Oikawa and Ainsworth, 2016; Zhang et al. 2019; Liu et al. 2022; Zhang et al. 2022), unknowing how this affects the precision of their results. Therefore, the aim of this study was to find the optimal sample size to estimate the overall experimental mean, coefficient of variation, and least significant difference for soybean early-cycle traits; and to assess precision gain as a function of sample size.

2. MATERIALS AND METHODS

2.1. Experimental design and conduction

Two experiments with soybean were carried out in the municipality of Erval Seco, Rio Grande do Sul, Brazil, during the 2017/2018 growing season. According to the Köppen climatic classification, the local climate is of the Cfa type, humid subtropical with no dry season defined and warm summers (Wrege et al. 2012), and with a Red Latosol soil (Santos et al. 2013). The experiments were conducted in a randomized block design with three replications, of which the

first experiment was sown on 10/24/2017, and the second, on 11/15/2017, using 30 soybean cultivars.

Plots were composed of five three-meter rows, spaced 0.45 m apart. Sowing was performed manually, with posterior thinning of plants so that a final population of 300,000 plants ha^{-1} was achieved. Base fertilization was applied according to the recommendations for the crop, aiming at a yield of 5,400 kg ha^{-1}. All cultural practices were performed uniformly within the experimental area according to the technical recommendations for the crop.

2.2. Evaluations and bootstrap resampling per experimental unit

In each experiment, 39 days after emergence, plant height and the number of nodes in the main stem were measured in 20 plants of the central row of each plot, totaling 1,800 plants per experiment, and 3,600 plants in total. Next, data were tabulated and 100 sample sizes were simulated (n = 1, 2, 3, ..., 100 plants per experimental unit). For this, the experiments were previously stratified into experimental units for each planned sample size. Also, for each sample size, 10,000 resamplings were performed with repositioning. In each resampling, the overall experimental mean (m) was estimated and a one-way analysis of variance was performed according to the following mathematical model for randomized block designs:

$$Y_{ij} = m + G_i + \beta_j + \varepsilon_{ij}$$

where Y_{ij} is the mean observed value of the response variable in plot ij; G_i is the fixed effect of level i of the cultivar (i = 1, 2, ..., 30); β_j is the random effect of the block (j = 1, 2, and 3); and ε_{ij} is the fixed effect of the experimental error, supposedly normal and independently distributed, with a mean of zero and a common variance σ^2 (Storck et al., 2016). Furthermore, the coefficient

of experimental variation (CVe), and the least significant difference (LSD) between cultivars were calculated in each resampling through Tukey's test at a 5% error probability, according to the following equations:

$$CVe = 100\frac{\sqrt{EMS}}{m}$$

$$LSD = q_{\alpha(i;DF_E)}\sqrt{\frac{EMS}{j}}$$

where *EMS* is the error mean square; $q_{\alpha(i;DF_E)}$ is the critical value to use in Tukey's test; *i* is the number of cultivars (i = 30); j is the number of repetitions (j = 3); and DF_E is the number of degrees of freedom of the error [$DF_E = (i-1)(j-1) = 58$].

2.3. Descriptive statistics and sample size definition

Statistical analyses were performed using R software (R Development Core Team 2026) and Microsoft Office Excel. For each experiment, trait (plant height and the number of nodes) and sample size simulated (n = 1, 2, 3, ..., 100 plants), 10,000 *m* values, analyses of variance, *CVe*, and *LSD* were calculated. These analyses were carried out using the native functions *aov*() and *sample*(). Next, *m*, *CVe,* and *LSD* values, of each resample, were recorded, and from those, 2.5 and 97.5 percentiles and mean values were obtained, and confidence interval width at 95% (CI) was calculated as the difference between 97.5 and 2.5 percentiles.

For the definition of sample size, the precision gain criterion was used, as established by Cargnelutti Filho et al. (2018) and efficiently used by Souza et al. (2022). Thus, the CI obtained for the lowest sample size (CI_1), that is, the sample size of one plant per experimental unit, was considered as a reference of 100% (maximum CI, and therefore minimum precision of the *m*,

135 *CVe*, and *LSD* estimates). Then, precision gain (PG_i, in %) was calculated by adding the i^{th} plants
136 (i = 2, 3, ..., 100 plants), through the following expression:

137 $$PG_i = 100 - \left(\frac{CI_i}{CI_1}\right) * 100$$

138 where CI_i is the confidence interval width at 95% of the sample sizes from 2 to 100 plants per
139 experimental unit. Posteriorly, PG_i (dependent variable) was parametrized as a function of the
140 number of plants per experimental unit (independent variable) for each statistic calculated
141 through Schumacher's model (Schumacher 1939). The following mathematical model was
142 applied using the *nls*() function:

143 $$PG_i = \alpha \times \exp(\beta \times n^{-1}) + \varepsilon$$

144 where PG_i is the i^{th} precision gain observation per statistic, in each n sample size, α and β are
145 parameters of the model, *exp* is the exponential function and ε_i is the error of random effect.
146 Finally, the maximum curvature point was obtained on the fitted model through the
147 perpendicular distance method, using the *maxcurv*() function, as reported by Silva and Lima
148 (2017), and by Souza et al. (2022) for soybean. This point was considered the optimal sample
149 size per experimental unit for the specific statistics, traits, and sowing dates understudy.
150
151 3. RESULTS
152
153 3.1. Response of experimental precision statistics to sample size scenarios
154
155 The *m* value of the traits was not altered by sample size increase (Figures 1a, d, g, and j)
156 in any of the two experiments. In the first experiment, for instance, the lowest sample size, of
157 one plant per experimental unit, obtained an *m* of 10.63 cm for plant height and of 4.89 units for

158 the number of nodes, which was maintained when sampling 100 plants per experimental unit
159 (Supplementary Tables 1 and 4). Those results indicate the *m* estimator is not biased; therefore,
160 sample size does not interfere with its property, however, it still allows the estimation of its real
161 value to be more precise, which is confirmed by the lower CI and the precision gain obtained at
162 higher sample sizes.

163 Regarding *CVe* and *LSD*, sample size increase from one to 10 plants per experimental unit
164 resulted in a reduction of mean values, in both experiments (Figures 1b, c, e, f, h, i, k, and l).
165 Thus, different than for *m*, these statistics are subjected to higher bias in small sample sizes
166 (Supplementary Tables 2, 3, 5, and 6). However, up from 16 plants per experimental unit, such
167 values tend to stabilize, reducing bias. Also, equally to *m*, the 97.5 percentile tended to a
168 reduction. *CVe*, for instance, presented values of 21.30 % and 19.87 % when sampling one plant
169 per experimental unit for plant height, and 8.73 % and 6.94 % when 100 plants per experimental
170 unit are considered, in the first and second experiments respectively (Figures 1b and h;
171 Supplementary Table 2). At the same time, the 2.5 percentile showed a gradual increase, with
172 values of 11.48 % and 10.06 % for the sampling of 1 plant per experimental unit, and 7.48 % and
173 5.99 % for 100 plants per experimental unit. For the number of nodes, a similar response was
174 identified in both experiments (Supplementary Table 5). Hence, this leads CI to reduce gradually
175 as sample size increases, forming a "funnel" shape, so that estimates get closer to the real values
176 as sample size increases and CI reduces. Therefore, in order to estimate both these statistics
177 reliably, the definition of a minimum sufficient sample size is necessary, considering insufficient
178 samplings may result in non-representative values and misleading results.
179
180 3.2. Precision gain and sample size

In both experiments, all experimental precision statistics for both early-cycle traits obtained a precision gain ≥ 85% when sampling 100 plants per experimental unit compared to the sampling of only one plant. Initially, that is, at smaller sample sizes, precision gains were higher, however, as the sampled number of plants per experimental unit increased, it tended to reduce gradually (Figure 2). For instance, when increasing the sample size from one to two plants per experimental unit, the precision gain is up to 30.27%. Nevertheless, this rate decreases when 17 plants per experimental unit are selected. Such a reduction is an indicator that sampling a high number of plants, such as 100 plants per experimental unit, would not guarantee a much higher precision gain than sampling only 17 plants.

Furthermore, the parametrization of mathematical models through Schumacher's method (1939) allowed predicting precision gain as a function of the number of sampled plants per experimental unit satisfactorily, according to the quality indicators (Table 1). Root mean square error (RMSE) was ≤ 2.00, and the coefficient of determination (R^2), and Willmott's agreement index were ≥ 0.98 and ≥ 0.99, respectively. When modeling this response, a point was defined up from which precision gain stabilized, corresponding to a certain sample size, which was defined as optimal. The optimal sample size presented a slight variation, of only one plant, among experimental precision statistics and experiments. Thus, the maximum value among those (17 plants per experimental unit) was considered sufficient to estimate both early-cycle traits in soybean (Table 2).

4. DISCUSSION

Recently, Souza et al. (2022) highlighted the choice for empirical sample numbers in studies with soybean and suggested an adequate sample size to estimate experimental precision statistics for this crop. The authors defined the sample size for soybean grain yield and recommended that further research should be carried out considering other traits. In this sense, our study focuses on soybean early-cycle phenotypic traits to find the optimal sample size for experimental precision statistics by understanding the response of those statistics to the number of sampled plants per experimental unit.

When analyzing the precisions statistics, *m* showed a distinct response concerning *CVe* and *LSD*. As already mentioned, the *m* estimator is not biased, which was also observed by Cargnelutti Filho et al. (2009b), Toebe et al. (2015), Bittencourt et al. (2022), and Souza et al. (2022). However, similarly to the *CVe* and *LSD*, its CI is higher for a low number of samples, suggesting that this estimator becomes more distant from its populational parameter in those scenarios (Siegel 2012; Anderson et al. 2017) so that finding an adequate sample size to estimate *m* is still necessary (Alvez-Silva et al. 2018). Souza et al. (2022) found a sample size of 16 plants per experimental unit for *m* when studying soybean grain yield under subtropical conditions in highlands. In this study, under similar edaphoclimatic conditions, slight variations were observed for the plant height and number of nodes, with values up to 17 plants per experimental unit. Such results may be useful for future studies, once they promote the obtention of reliable estimates of this base statistic (*m*), which can be used to make other mathematical inferences (Cruz et al. 2012).

CVe and *LSD* statistics are subjected to quite inconsistent estimates in small sample sizes. Similar responses have been observed for other crops and phenotypic traits (Piñera-Chavez et al. 2020; Bittencourt et al. 2022), as well as for soybean grain yield (Souza et al. 2022). One of the

possible reasons for this may be *EMS* estimative through the least squares method, which is biased in very small sample sizes, as reported by Souza et al. (2022). Hence, when using this statistic, *CVe* and *LSD* become extremely sensitive to sample size so the need for the optimal sample size for these precision statistics is even stronger than for *m*. Nevertheless, the three tended to have similar precision gains when 17 plants per experimental unit were sampled.

Once a precision of ~70% is gained considering the sampling of only one plant per experimental unit, precision gain reaches stabilization, similarly to the results obtained by Souza et al. (2022) for soybean grain yield. This condition can be explained by the reduction of extreme values resulting from sample size increase, once the observed values get closer to the real values. Thus, with each increase in the number of samples, the precision gain becomes lower, given CI will reach a stabilization point that will lead to similar CI values at high sample sizes. Because of this, the use of extremely high sample sizes does not compensate for precision, considering that the higher the number of samples, the higher the costs and the greater the need for resources and labor (Toebe et al. 2015; Bittencourt et al. 2022). Therefore, 17 plants per experimental unit, which reached a precision gain of ~70% are sufficient for the reliable estimate of *m*, *CVe,* and *LSD* for the plant height and number of nodes of early-cycle soybean.

Moreover, based on our results, other relevant soybean traits deserve attention and can be targeted in sample-size-dimensioning studies. Also, as recommended by Souza et al. (2022), we highly encourage the definition of optimal sample sizes for the traits and statistics here studied under different edaphoclimatic conditions and expect that the sample size recommendations proposed in this study serve as a reference for future research that measures these traits in soybean. Still, we recommend that the sample sizes found here are used strictly for soybean once

different crops, even with similar plant architecture, may require a different number of samples to estimate the precision statistics here studied.

5. CONCLUSION

Precision statistics *m*, *CVe,* and *LSD* were influenced by sample size for soybean early-cycle traits plant height and number of nodes, where sampling increase up to 17 plants per experimental unit provided an increase in precision, for which this number was recommended as the optimal sample size to estimate such experimental precision statistics reliably for those traits.

CONFLICT OF INTEREST

The authors have no conflicts of interest to declare.

SUPPORTING INFORMATION

Supporting information for this article is available online.

REFERENCES

Alvez-Silva E, Santos JC, Cornelissen TG. 2018. How many leaves are enough? The influence of sample size on estimates of plant developmental instability and leaf asymmetry. Ecol. Indic. 89: 912-924.

271 Anderson SF, Kelley K, Maxwell SE. 2017. Sample-size planning for more accurate statistical
272 power: a method adjusting sample effect sizes for publication bias and uncertainty. Psychol. Sci.
273 28: 1547-1562.
274 Bittencourt KC, Souza RRS, Pazetto SB, Toebe M, Toebe ICD. 2022. What is the best way to
275 define sample size for cauliflower seedlings? Cienc. Rural 52.
276 Cargnelutti Filho A, Evangelista DHR, Gonçalvez ECP, Storck L. 2009a. Sample size of the
277 characters in soybeans genotypes. Cienc. Rural 39: 983-991.
278 Cargnelutti Filho A, Storck L, Ribeiro ND. 2009b. Measures of experimental precision in
279 common bean and soybean genotype trials. Pesqui. Agropecu. Bras. 44: 1225-1231.
280 Cargnelutti Filho A, Toebe M, Burin C, Neu IMM, Alves BM. 2018. Number of leaves for
281 modelling the leaf area of velvet bean according to leaf dimensions. Rev. Ciênc. Agrovet. 17:
282 571-578.
283 Confalonieri R, Stroppiana D, Boschetti M, Gusberti D, Bocchi S, Acutis M. 2006. Analysis of
284 rice sample size variability due to development stage, nitrogen fertilization, sowing technique
285 and variety using the visual jackknife. Field Crops Res. 97: 135-141.
286 Confalonieri R, Perego A, Chiodini ME, Scaglia B, Rosenmund AS, Acutis M. 2009. Analysis of
287 sample size for variables related to plant, soil, and soil microbial respiration in a paddy rice field.
288 Field Crops Res. 113: 125-130.
289 Cruz CD, Carneiro PCS, Regazzi AJ. 2012. Biometric models applied to genetical improvement,
290 4th ed. UFV, Viçosa, Brazil.
291 Gordón-Mendoza R, Camargo-Buitrago I. 2015. Statistical selection for estimating the accuracy
292 in experimental corn trials. Agron. Mesoam. 26: 55-63.

293 Li M, Liu Y, Wang C, Yang X, Li D, Zhang X, Xu C, Zhang Y, Li W, Zhao L. 2020.
294 Identification of traits contributing to high and stable yields in different soybean varieties across
295 three Chinese latitudes. Front. Plant Sci. 10: 1642.
296 Liu C, Tian Y, Liu Z, Gu Y, Zhang B, Li Y, Na J, Qiu L. 2022. Identification and
297 characterization of long-InDels through whole genome resequencing to facilitate fine-mapping of
298 a QTL for plant height in soybean (Glycine max L. Merr.). J. Integr. Agric. 21: 1903-1912.
299 Maxwell SE, Kelley K, Rausch JR. 2008. Sample size planning for statistical power and
300 accuracy in parameter estimation. Annu. Rev. Psychol. 59: 537-563.
301 Mwiinga B, Sibiya J, Kondwakwenda A, Musvosvi C, Chigeza G. 2020. Genotype ×
302 environment interaction analysis of soybean (Glycine max (L.) Merrill) grain yield across
303 production environments in Southern Africa. Field Crops Res. 256: 107922.
304 Oikawa S, Ainsworth EA. 2016. Changes in leaf area, nitrogen content and canopy
305 photosynthesis in soybean exposed to an ozone concentration gradient. Environ. Pollut. 215:
306 347-355.
307 Piñera-Chavez FJ, Berry PM, Foulkes MJ, Molero G, Reynolds MP. 2020. Optimizing
308 phenotyping methods to evaluate lodging risk for wheat. Field Crops Res. 258: 107933.
309 R Development Core Team. 2022. R: A language and environment for statistical computing. R
310 Foundation for Statistical Computing, Vienna, Austria.
311 Santos HG dos, Jacomine PKT, Anjos LHC dos, Oliveira VA de, Lumbreras JF, Coelho MR,
312 Almeida JA de, Araújo Filho JC de, Oliveira JB de, Cunha TJF. 2018. Brazilian soil
313 classification system. EMBRAPA, Brasília, Brazil.
314 Schumacher FX. 1939. A new growth curve and its application to timber yield studies. J. For. 37:
315 819-820.

316 Siegel AF. 2016. Practical business statistics. Academic Press, London, United Kingdom.
317 Silva AR da, Lima RP. 2017. Determination of maximum curvature point with the R package
318 soilphysics. Int. J. Curr. Res. 9: 45241-45245.
319 Souza RR de S, Toebe M, Marchioro VS, Cargnelutti Filho A, Lúcio AD, Benin G, Mello AC,
320 Tartaglia FL de, Manfio GL. 2022. Soybean yield variability per plant in subtropical climate:
321 sample size definition and prediction models for precision statistics. Eur. J. Agron. 136: 126489.
322 Souza RRS, Toebe M, Marchioro VS, Cargnelutti Filho A, Mello AC, Manfio GL, Soldateli FJ,
323 Soares FS dos, Martins V, Junges DL. 2021. Soybean grain yield in Highland and lowland
324 cultivation systems: a genotype by environment interaction approach. Ann. Appl. Biol. 179: 302-
325 318.
326 Storck L, Garcia DC, Lopes SJ, Estefanel V. 2016. Plant experimentation. UFSM, Santa Maria,
327 Brazil.
328 Toebe M, Cargnelutti Filho A, Lopes SJ, Burin C, Silveira TR da, Casarotto G. 2015. Sample
329 dimensioning for estimating coefficients of correlation in maize hybrids, harvests and precision
330 levels. Bragantia 74: 16-24.
331 Wrege MS, Steinmetz S, Reisser Júnior C, Almeida IR de. 2012. Climatic atlas of the South
332 Region of Brazil: States of Paraná, Santa Catarina and Rio Grande do Sul. EMBRAPA, Brasília,
333 Brazil.
334 Zanon AJ, Streck NA, Grassini P. 2016. Climate and management factors influence soybean
335 yield potential in a subtropical environment. Agron. J. 108: 1447-1454.
336 Zhang D, Wang X, Li S, Wang C, Gosney MJ, Mickelbart MV, Ma J. 2019. A post-
337 domestication mutation, Dt2, triggers systemic modification of divergent and convergent
338 pathways modulating multiple agronomic traits in soybean. Mol. Plant 12: 1366-1382.

Zhang H, Hai-yan W, Rui T, You-bin K, Jia-hao C, Xin-zhu X, Hui D, Yuan J, Xi-huan L, Cai-ying Z. 2022. Genome-wide association and linkage mapping strategies reveal genetic loci and candidate genes of phosphorus utilization in soybean. J. Integr. Agric. 2022. In press.

TABLES

Table 1 – Schumacher models and quality indicators [coefficient of determination (R^2), root mean square error (RMSE), and Willmott's agreement index (d index)] for soybean early-cycle traits plant height and number of nodes.

Trait	Precision statistic	Schumacher's model	R^2	RMSE	d index
	First sowing date (10/24/2017)				
PH	Plant height (cm^{-1})	$PG_i = 90.8201 \times exp(-2.6372 \times n^{-1})$	0.99	1.52	0.99
PH	*CVe*[*]	$PG_i = 88.2822 \times exp(-3.1405 \times n^{-1})$	0.98	1.85	0.99
PH	*LSD*	$PG_i = 87.9993 \times exp(-3.2718 \times n^{-1})$	0.98	1.86	0.99
NN	Number of nodes	$PG_i = 90.8900 \times exp(-2.6748 \times n^{-1})$	0.99	1.47	0.99
NN	*CVe*	$PG_i = 88.2822 \times exp(-3.0582 \times n^{-1})$	0.98	1.81	0.99
NN	*LSD*	$PG_i = 88.0339 \times exp(-3.1375 \times n^{-1})$	0.98	1.86	0.99
	Second sowing date (11/15/2017)				
PH	Plant height (cm^{-1})	$PG_i = 90.8101 \times exp(-2.6522 \times n^{-1})$	0.99	1.52	0.99
PH	*CVe*	$PG_i = 90.1934 \times exp(-2.6427 \times n^{-1})$	0.99	1.61	0.99
PH	*LSD*	$PG_i = 89.8914 \times exp(-2.7429 \times n^{-1})$	0.99	1.61	0.99
NN	Number of nodes	$PG_i = 90.6354 \times exp(-2.7690 \times n^{-1})$	0.99	1.47	0.99
NN	*CVe*	$PG_i = 88.0292 \times exp(-3.1013 \times n^{-1})$	0.98	2.00	0.99
NN	*LSD*	$PG_i = 87.6505 \times exp(-3.2882 \times n^{-1})$	0.98	1.98	0.99

[*]PH: Plant height (cm^{-1}); NN: Number of nodes (units); CVe: coefficient of variation; LSD: Least significant difference at 5% error probability; PG_i: Precision gain (%) calculated for sample sizes ranging from 2 to 100 plants.

Table 2 – Maximum curvature points and maximum confidence interval width at 95% (CI) for the precision statistics overall experimental mean (*m*), coefficient of variation (*CVe*), and least significant difference at 5% error probability (LSD) for the soybean early-cycle traits plant height and number of nodes.

	m	*CVe*	*LSD*
	First sowing date (10/24/2017)		
	Plant height (cm^{-1})		
Maximum curvature	15.8530	16.6847	16.9422
Maximum CI	76.0373	73.1354	72.5455
Sample size	16	17	17
	Number of nodes		
Maximum curvature	15.7143	16.5065	16.6649
Maximum CI	76.6643	73.4252	72.9265
Sample size	16	17	17
	Second sowing date (11/15/2017)		
	Plant height (cm^{-1})		
Maximum curvature	15.6747	15.6549	15.8530
Maximum CI	76.6744	76.1837	75.6094
Sample size	16	16	16
	Number of nodes		
Maximum curvature	15.9124	16.6055	16.9818
Maximum CI	76.1595	73.0326	72.2207
Sample size	16	17	17

365 FIGURES LEGENDS

366

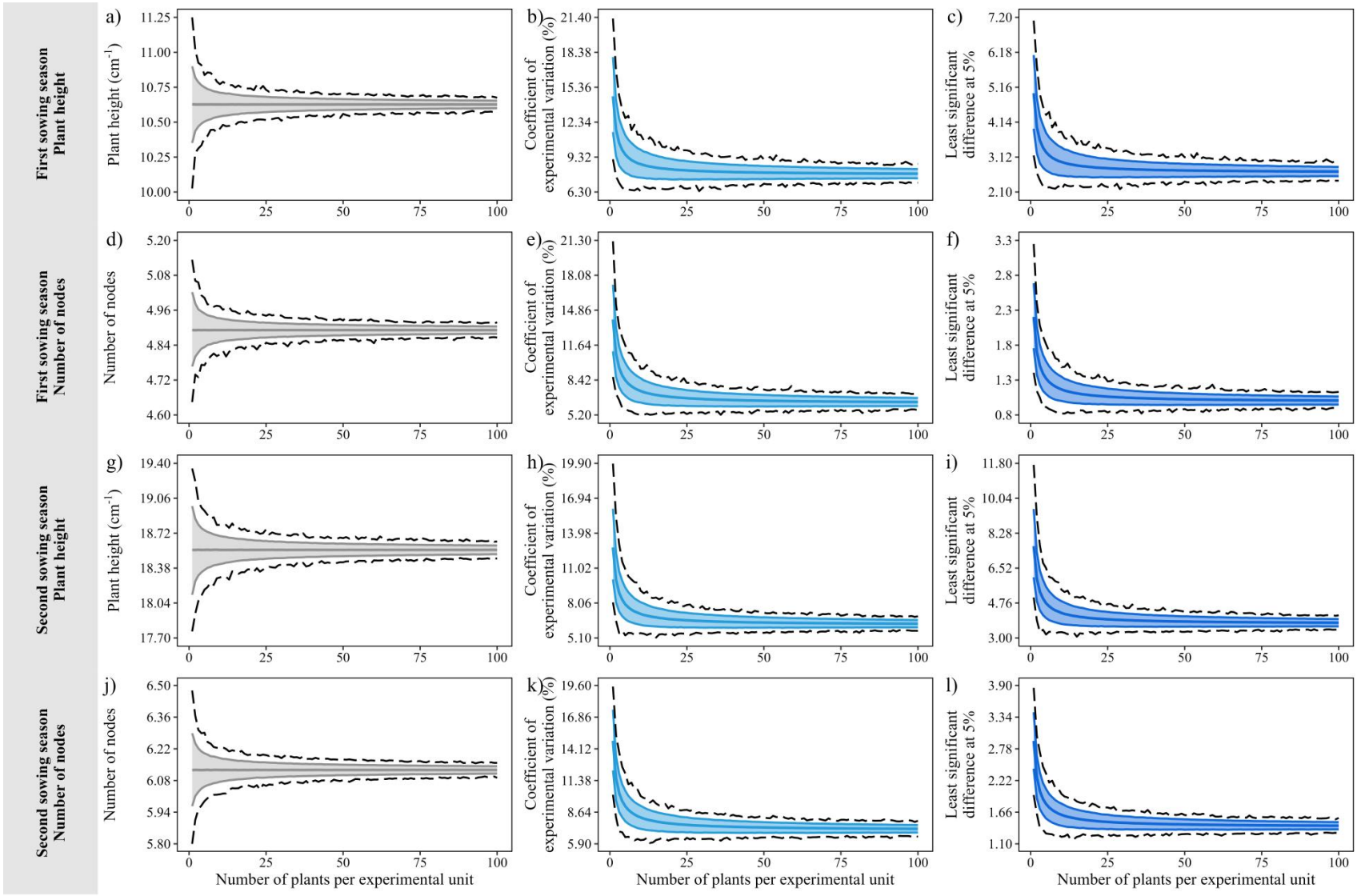
367 FIGURE 1. Descriptive statistics of the bootstrap resamplings per simulated sample size and
368 their respective minimum, 2.5 percentile, mean, 97.5 percentile, and maximum values for a)
369 overall experimental mean (in cm), b) coefficient of variation (in %), and c) least significant
370 difference of Tukey's test at 5% (in cm) of the plant height measured in 30 soybean cultivars
371 during the first sowing season (October 24, 2017); d) overall experimental mean (in units), e)
372 coefficient of variation (in %), and f) least significant difference of Tukey's test at 5% (in units)
373 of the number of nodes in the main stem measured in 30 soybean cultivars during the first
374 sowing season (October 24, 2017); g) overall experimental mean (in cm), h) coefficient of
375 variation (in %), and i) least significant difference of Tukey's test at 5% (in cm) of the plant
376 height measured in 30 soybean cultivars during the second sowing season (November 15, 2017);
377 and j) overall experimental mean (in units), k) coefficient of variation (in %), and l) least
378 significant difference of Tukey's test at 5% (in units) of the number of nodes in the main stem
379 measured in 30 soybean cultivars during the second sowing season (November 15, 2017).

380

381 FIGURE 2. Precision gain and maximum curvature point through the perpendicular distances
382 method for the definition of the optimal sample size for the overall experimental mean (in cm –
383 a, d, m and p); coefficient of variation (in % – b, e, n and q); and least significant difference of
384 Tukey's test at 5% (in cm – c, f, o and r) of the plant height measured in 30 soybean cultivars
385 during the first sowing season (October 24, 2017) (a–f) and second sowing season (November
386 15, 2017) (m–r); and for the overall experimental mean (in units – g, j, s and v); coefficient of
387 variation (in % – h, k, t and x); and least significant difference of Tukey's test at 5% (in units – i,

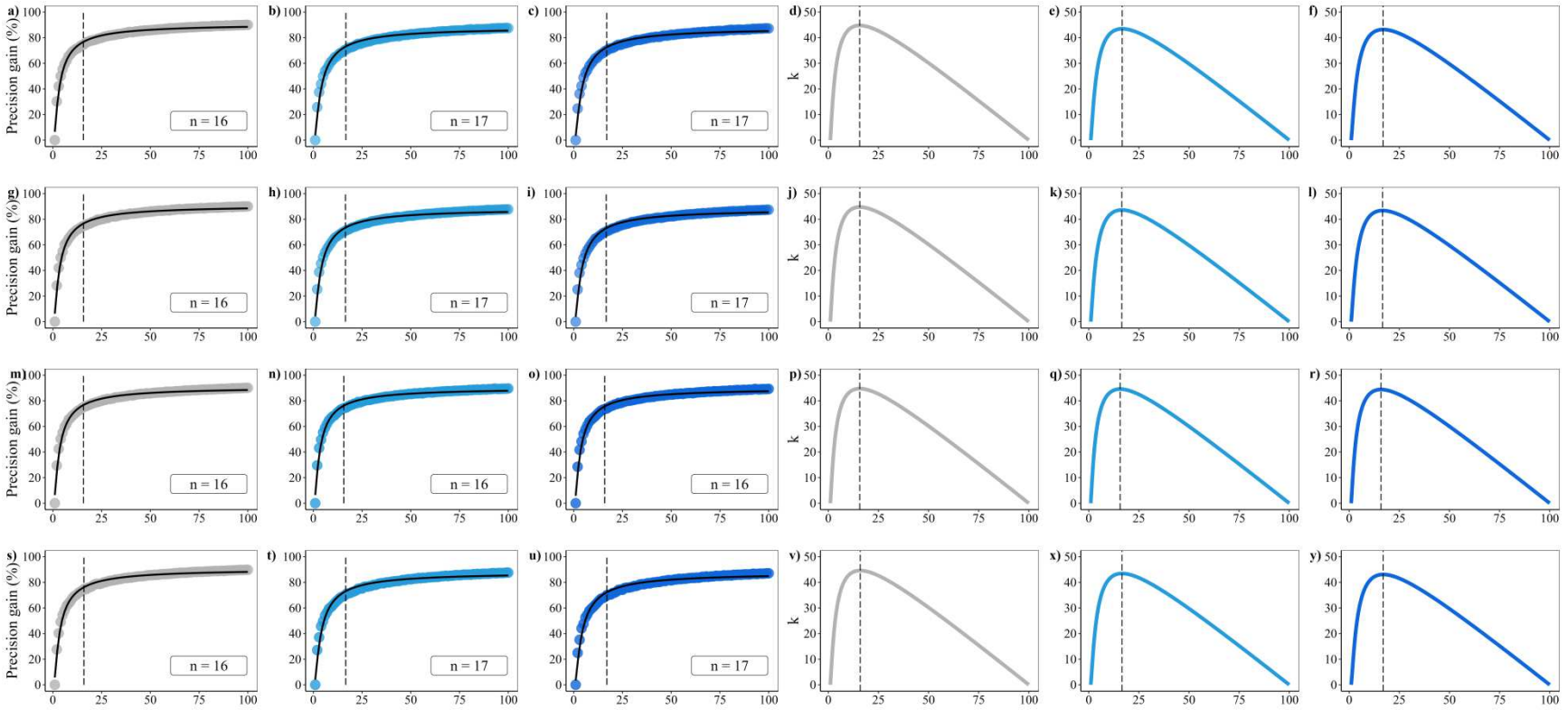
388 l, u and y) of the number of nodes in the main stem measured in 30 soybean cultivars during the
389 first sowing season (October 24, 2017) (g–l) and second sowing season (November 15, 2017) (s–
390 y).*k: values of 5,000 empirical distances assumed from the parametrization of Shumacher's model as a function of
391 the precision gain and number of plants per experimental unit.

392 FIGURES



393

394



395 Number of plants per experimental unit

396

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- JCSBSupplementarymaterial.docx