

UNIVERSIDADE FEDERAL DE VIÇOSA

RAMON GONÇALVES DE PAULA

**HIGH-TROUGHPUT PHENOTYPING OF PHYSIOLOGICAL AND YIELD RELATED
TRAITS IN SOYBEAN USING HYPERESPECTRAL SENSOR**

**VIÇOSA – MINAS GERAIS
2021**

RAMON GONÇALVES DE PAULA

**HIGH-TROUGHPUT PHENOTYPING OF PHYSIOLOGICAL AND YIELD RELATED
TRAITS IN SOYBEAN USING HYPERESPECTRAL SENSOR**

Tese apresentada à Universidade Federal de Viçosa,
como parte das exigências do Programa de Pós-
Graduação em Genética e Melhoramento, para obtenção
do título de *Doctor Scientiae*.

Orientador: Felipe Lopes da Silva

Coorientadores: Cibele Hummel do Amaral
Pedro Crescêncio Souza Carneiro

**VIÇOSA – MINAS GERAIS
2021**

**Ficha catalográfica elaborada pela Biblioteca Central da Universidade
Federal de Viçosa - Campus Viçosa**

T

P324h
2021 Paula, Ramon Gonçalves de, 1989-
High-throughput phenotyping of physiological and yield
related traits in soybean using hyperspectral sensor / Ramon
Gonçalves de Paula. – Viçosa, MG, 2021.
77 f. : il. (algumas color.) ; 29 cm.

Orientador: Felipe Lopes da Silva.

Tese (doutorado) - Universidade Federal de Viçosa.

Inclui bibliografia.

1. Soja - Melhoramento genético. 2. Folhas. 3. Reflectância.
4. Sensoriamento remoto. 5. Previsão de rendimento.

I. Universidade Federal de Viçosa. Departamento de Agronomia.
Programa de Pós-Graduação em Genética e Melhoramento.

II. Título.

CDD 22. ed. 633.342

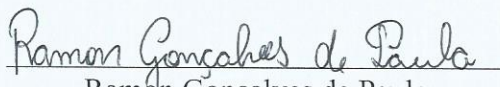
RAMON GONÇALVES DE PAULA

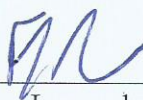
**HIGH-THROUGHPUT PHENOTYPING OF PHYSIOLOGICAL AND YIELD
RELATED TRAITS IN SOYBEAN USING HYPERESPECTRAL SENSOR**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de *Doctor Scientiae*.

APROVADA: 26 de fevereiro de 2021.

Assentimento:


Ramon Gonçalves de Paula
Autor


Felipe Lopes da Silva
Orientador

AGRADECIMENTOS

Agradeço e glorifico a Deus, por ter me dado o dom de viver, e guiar meus passos com sabedoria.

Agradeço também aos meus pais Gildemar e Maria da Penha, por todo o esforço, sacrifício, e apoio incondicional que sempre me deram em toda a minha vida acadêmica. Sem eles não teria chegado até aqui.

À minha futura esposa, Herika, pelo apoio e carinho em todos os momentos da minha vida.

À Universidade Federal de Viçosa e ao Programa de Pós-Graduação em Genética e Melhoramento, pela oportunidade de realização do doutorado.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela concessão da bolsa de Doutorado.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento – 001, para realização desse trabalho.

A todos os professores da Universidade Federal de Viçosa que participaram da minha formação e me ajudaram a adquirir conhecimento. Em especial ao professor Felipe, pela amizade, confiança e orientação durante todo o curso. E aos professores Cibele e Pedro pela coorientação.

Aos secretários e funcionários do departamento de Genética, Marco Tulio e Odilon.

Aos amigos do Programa Soja, que sempre estiveram dispostos a ajudar na condução dos trabalhos realizados nesta tese. Além de proporcionar um ambiente descontraído e alegre no dia a dia de trabalho. Em especial agradeço aos colegas de pós-graduação Isabella, Guilherme, Martha, Bruno, Heloísa, Francisco, Fernanda, Murilo, Maikon, Anunciene.

Aos meus colegas e amigos do GenMelhor, das gestões 2017 e 2018, que contribuíram na minha formação acadêmica e crescimento pessoal.

A todos os amigos que fiz desde a minha infância, que apesar da distância durante esse período, sempre me proporcionaram momentos de alegria e apoio.

Meus sinceros agradecimentos.

ABSTRACT

DE PAULA, Ramon Gonçalves, M.Sc., Universidade Federal de Viçosa, February, 2021. **High-throughput phenotyping of physiological and yield related traits in soybean using hyperspectral sensor.** Adviser: Felipe Lopes da Silva. Co-advisers: Cibelle Hummel do Amaral and Pedro Crescêncio de Souza Carneiro.

Increasing yield still one of the main challenges in soybean breeding programs due to its complexity as a result of the great genetic variability and genotype x environment interaction. Whereas great advantage has been achieved in selecting top yielded cultivars in last decades, breeders have frequently face unfavorable environment situation for selection, such as drought stress caused by lack of rainfall at critical phases of the crop. Better understanding of the phenotype might provide additional information about the genotype and help in a more accurate selection. Assessing physiological, biochemistry, and morphology information of the genotyping is becoming necessary to broken down complex traits and achieve better results in plant breeding. Traditional methods to assess these traits is not feasible in large scale breeding company, so the use of fast, noninvasive and accurate tools such as remote sensing is becoming usual. This study tested the use of leaf hyperspectral reflectance (350 – 2500 nm) in 20 soybean cultivars as a high-throughput phenotyping approach to: (1) estimate physiological trait under drought condition and selected cultivars based on these traits; and (2) predicted yield-related traits and to verify the availability on the selection procedure based on those traits. Partial least square (PLS) regression models accurately estimated seven out of eight physiological traits when the measurements were taken 20 and 28 days after drought imposition. Built PLS discriminant analysis model allowed to select the top five cultivars, having 100% of coincide when comparing to the FAI-BLUP selection index regarding all physiological traits simultaneously. Reliable PLS regression models prediction were also found for six out of ten yield-related traits, including number of seeds per pound and 100 grain weight. Good association was found between the selected wavelengths and FAI-BLUP indexes values, indicating that the model could be used as a selection approach. The results showed in this study indicate the great potential in using hyperspectral reflectance as a feasible, nondestructive, and accurately method to phenotyping physiological and yield-related traits as well as to screening superior genotypes.

Keywords: Leaf hyperspectral reflectance. High-throughput phenotyping. Physiological traits. Yield components. Drought condition. Soybean.

RESUMO

DE PAULA, Ramon Gonçalves, M.Sc., Universidade Federal de Viçosa, fevereiro de 2021. **Fenotipagem de alto rendimento de caracteres fisiológicos e de produtividade em soja usando sensor hiperespectral.** Orientador: Felipe Lopes da Silva. Coorientadores: Cibelle Hummel do Amaral e Pedro Crescêncio de Souza Carneiro.

O aumento da produtividade ainda é um dos principais desafios nos programas de melhoramento da soja devido à sua complexidade em decorrência da grande variabilidade genética e da interação genótipo x ambiente. Embora grande vantagem tenha sido alcançada na seleção de cultivares de alto rendimento nas últimas décadas, os melhoristas frequentemente enfrentam situações ambientais desfavoráveis para a seleção, como o estresse hídrico causado pela falta de chuvas em fases críticas da cultura. Uma melhor compreensão do fenótipo pode fornecer informações adicionais sobre o genótipo e ajudar em uma seleção mais precisa. A avaliação das informações fisiológicas, bioquímicas e morfológicas da genotipagem está se tornando necessária para quebrar características complexas e obter melhores resultados no melhoramento genético. Métodos tradicionais para avaliar essas características não são viáveis em empresas de melhoramento em grande escala, então o uso de ferramentas rápidas, não invasivas e precisas como o sensoriamento remoto está se tornando comum. Este estudo testou o uso de refletância hiperespectral foliar (350 - 2500 nm) em 20 cultivares de soja como uma abordagem de fenotipagem de alto rendimento para: (1) estimar características fisiológicas sob condição de seca e cultivares selecionadas com base nessas características; e (2) características relacionadas à produção prevista e verificar a disponibilidade no procedimento de seleção com base nessas características. Modelos de regressão de mínimos quadrados parciais (PLS) estimaram com precisão sete das oito características fisiológicas quando as medições foram feitas 20 e 28 dias após a imposição da seca. O modelo de análise discriminante PLS construído permitiu selecionar as cinco melhores cultivares, tendo 100% de coincidência na comparação com o índice de seleção FAI-BLUP em relação a todas as características fisiológicas simultaneamente. Predição confiável de modelos de regressão PLS também foi encontrada para seis entre dez características relacionadas ao rendimento, incluindo número de sementes por libra e peso de 100 grãos. Boa associação foi encontrada entre os comprimentos de onda selecionados e os valores dos índices FAI-BLUP, indicando que o modelo pode ser usado como uma abordagem de seleção. Os resultados apresentados neste estudo indicam

o grande potencial no uso da refletância hiperespectral como um método viável, não destrutivo e preciso para fenotipar características fisiológicas e relacionadas à produção, bem como para rastrear genótipos superiores.

Palavras-chave: Refletância hiperespectral foliar. Fenotipagem de alto rendimento. Características fisiológicas. Componentes de produtividade. Seca. Soja.

LISTA DE FIGURAS

Chapter I

- Fig 1. Daily information in the evaluations window regarding air relativity humidity, in percentage, temperature, in Celcius degree, and solar radiation, in Mega Joules per hour. Grey bars highlight the values at the two evaluations dates.28
- Fig 2. Boxplot distribution of physiological traits of cultivars under watered (green boxes) and drought (red boxes) conditions at 20 days after drought (DAD) (A-F) and at 28 DAD (G-L). (A) and (G), chlorophyll a; (B) and (H), chlorophyll b; (C) and (I), carotenoids; (D) and (J), specific leaf area; (E), electrons transport rate; (F), Photosynthetically active radiation; (K), relative water content; (L), temperature.29
- Fig 3. Means of spectral signatures of the cultivars in watered (green) and drought (red) conditions at 20 days after drought impose (DAD), and 28 DAD.....30
- Fig 4. Observed vs predicted values from PLSR models at 20 days after drought (DAD) (A-F) and at 28 DAD (G-L). (A) and (G), chlorophyll a; (B) and (H), chlorophyll b; (C) and (I), carotenoids; (D) and (J), specific leaf area; (E), electrons transport rate; (F), Photosynthetically active radiation; (K), relative water content; (L), temperature. Green dots are cultivars in watered condition and red dots represents cultivars in drought condition. ...31
- Fig 5. Observed vs predicted values from PLSR models at 28 days after drought (DAD) (A-F). (A) and, chlorophyll a; (B), chlorophyll b; (C) carotenoids; (D), specific leaf area; (E), relative water content; (F), leaf temperature. Green dots are cultivars in watered condition and red dots represents cultivars in drought condition32
- Fig 6. Measured values from 28 DAD data versus predicted values from PLSR models built from 20 DAD for (A) chlorophyll a content; (B) chlorophyll b content; (C) carotenoid content; and (D) specific leaf area. Green circles represent cultivars in watered condition while red circles represent cultivars in drought condition.33
- Fig 7. Ranking of the cultivars according to FAI-BLUP index. The class of selected cultivars are set in red while the classes of non-selected cultivars are set in black (A); Classification of the cultivars into the selected and non-selected classes from cross validation results of partial least square regression discriminant model using spectral data of 20 DAD (B), 28 DAD (C), and spectral mean data of both dates (D).34
- Fig 8. Spectral mean signature of the cultivars and the wavelengths selected (blue bars) in the final PLS-DA model using the variable importance in projection (VIP).35

Chapter II

- Fig 1. Scheme showing the workflow of data collection, PLSR model building and selection procedure.59
- Fig 2. - Linear correlation coefficient between seedless pod (SP), one-seed pod (P1), two-seeds pod (P2), three-seeds pod (P3), four-seeds pod (P4), one-hundred seed mass (100SM), and difference between actual and potential seed yield (GAP), at upper triangle matrix. Linear

regression line and 95% confidence interval (red shaded) are shown at lower triangle matrix. * significance at $p<0.05$. ** significance at $p<0.01$	61
Fig 3. Cross-validated (cv) predicted and measured values of a) seedless pod (SP), b) one-seed pod (P1), c) two-seeds pod (P2), d) three-seeds pod (P3), e) four-seeds pod (P4), f) difference between actual and potential yield (GAP), g) one-hundred seed mass (100SM), and h) FAI-BLUP Ideotype indexes	65
Fig 4. Selected wavelengths for PLSR model of the FAI-BLUP index vales (blue cicles) through the leaf hyperspectral reflectance (blue line).	66

LISTA DE TABELAS

Chapter II

Table 1- List of the cultivars evaluated	57
Table 2- Genotypic variance component (V.C.), prediction accuracy of the model and heritability for the traits seedless pod (SP), one seed pod (P1), two seeds pod (P2), three seeds pod (P3), four seeds pod (P4), one-hundred seed mass (100SM), total mass (TM), total seeds (TS), total pods (TP) and the difference between actual and potential seed yield (GAP).....	60
Table 3 - FAI-BLUP index estimated for 20 genotypes ranked from the highest to lowest value, traits associated with each of the selected factor and the genetic gain in selecting the best five cultivars according to FAI-BLUP index	62

MATERIAL SUPPLEMENTAR

Chapter I

Table S1. List of the 20 cultivars used in this study	40
Table S2. Pre processing, number of latent variable (LV), cross validation method, R2 calibration and R2 cross validation, root mean square error of calibration (RMSEC) and root mean square error of cross validation (RMSECV) , and number of wavelengths selected via variable importance in projection (VIP) from the partial least square regression analysis for chlorophyll a (CLa), chlorophyll b (CLb), carotenoids (CAR), specific leaf area (SLA) at 20 days after drought (20 DAD) and 28 DAD, electrons transport rate (ETR), photosynthetic active radiation (PAR) at 20 DAD, and relative water content (RWC), and temperature (TEM) at 28 DAD.	41
Table S3. Pre processing, number of latent variable (LV), cross validation method, R2 calibration and R2 cross validation, and number of wavelengths selected via variable importance in projection (VIP) of the partial least regression discriminant analysis considering the mean wavelengths values of the both 20 DAD and 28 DAD evaluations	42

Chapter II

Table S1. Descriptive statistics of the traits seedless pod (SP), one seed pod (P1), two seeds pod (P2), three seeds pod (P3), four seeds pod (P4), one-hundred seed mass (100SM), total mass (TM), total pods (TP), total seeds (TS) and the difference between real and potential seed yield (GAP)	68
--	----

SUMMARY

GENERAL INTRODUCTION.....	13
REFERENCES	16
Chapter I - ESTIMATION OF PHYSIOLOGICAL TRAITS AND SELECTION OF SOYBEAN CULTIVARS UNDER DROUGHT CONDITION TROUGH LEAF HYPERSPECTRAL REFLECTANCE	20
ABSTRACT.....	20
1. INTRODUCTION	21
2. MATERIAL AND METHODS.....	22
2.1. Physiological traits estimation	23
2.2 Prediction of selected cultivars via spectral signature	26
3. RESULTS.....	27
3.1 Physiological traits estimation using hyperspectral data.....	29
3.2 Cultivars selection through spectral data	33
4. DISCUSSION.....	35
4.1 Drought effects on physiological traits	35
4.2 Physiological traits estimation via spectral signature	37
4.3 Genotype selection	38
REFERENCES	43
Chapter II - POTENTIAL OF LEAF HYPERSPECTRAL PROXIMAL SENSING IN ESTIMATE YIELD-RELATED TRAITS IN SOYBEAN.....	54
ABSTRACT.....	54
1. INTRODUCTION	55
2. MATERIAL AND METHODS.....	56
2.1. Plant material and yield-related traits evaluation	56
2.2. Cultivar selection.....	57
2.3. Leaf hyperspectral data and Partial Least Square Regression (PLSR)	58
3. RESULTS AND DISCUSSION.....	59
REFERENCES	69

GENERAL INTRODUCTION

Considered the most important annual crop in Brazil, soybean (*Glycine max* L.) plays an important role in the economic scenario, in addition for develop technologically growth in several regions in the country in the last decades (SILVA et al., 2017, SEDIYAMA et al., 2015). Economic importance of this leguminous is due to its excellent source of vegetable oil and protein, which contents reach values around 20 and 40%, respectively. In addition, these compounds can be used for other purpose, such as in the biofuel industries (SINGH et al., 2008; Hartman et al., 2011).

The soybean originated from the northeast China and was introduced in Brazil in the second half of the eighteenth century. However, only in the middle of 1970 decade it was spread out around the country. According to USDA (2020), Brazil is the biggest producer of this crop in the world, with reported production of 131.20 million tons in 2019/2020 season. Nowadays, the soybean production has a large share in Brazilian trade market, making this legume the main item in the exporting Brazilian commodities background. The adaptation and expansion in the national territory were due to several governmental investment in agriculture, such as in technology, agriculture frontier expansion, and mainly development of new materials adapted to soil fertility and latitude conditions, thanks to the breeding programs (SEDIYAMA et al., 2015).

In the coming years, the world's population and purchasing power will continue to grow in emerging economies, especially in Asian countries, which concentrates the highest consumption potential. Studies forecast that, by 2050, the earth's population will reach nine billion, which will likely demand 333.674 million tons of food (ALEXANDRATOS and BRUINSMA 2012). Due to the great importance of soybeans in many industrial applications, the demand for its grain will increase in the coming years. In additional to that, stagnation in terms of arable areas and the concern about environment situation has been worrying farmers around the world (CLARK and TILMAN, 2017). Changes in environment patterns has been detected and associated to possible breaks in yield, mostly related to abiotic stresses (MELILLO et al., 2014). In this context, plant breeding can be considered the main mechanism to overcome those issues for generate new genotypes to environmental, economic and social demands (PALANICHAMY e COBB, 2015).

Despite advances in plant breeding, focuses on keeping genotypic variation has remained one of the main alternative for increase genetic gains, which are directly associated to the ability of methodologies to discriminate the potential of genotypes in a given breeding population.

However, identifying the best individuals to recombine is challenging, since the phenotype is determined by the genetic of an individual, the effect of the environment and by the environment interaction over that individual, which allow genotypes behave differently according to environmental conditions (ACQUAAH, 2009).

Advent of genetic markers allowed breeders to visualize the effects of selection on genomes and, with some accuracy, predict recombinants with a higher probability of generating better phenotypes (McCOUCH et al., 2012). With the progress of molecular genetics and the reduction in genotyping costs, it is possible to obtain large amounts of genotype data in a very short time. Thus, plant breeding programs have driven their strategies to predict and select desirable recombinants based on genotypic data. The selection is based on a smaller population (training population), which is carefully genotyped and phenotyped (ASORO et al., 2011). Consequently, great necessity for high-quality phenotypic data is demanded in large scale, since it strengthen the efforts of genetic mapping and contribute to the own direct phenotypic selection in a breeding program (TUBEROSA, 2012).

High-throughput phenotyping is a new area of biology that aims measure and characterize the physical and biochemical characteristics of an individual. It aims to rapidly process the phenotype on a large numbers of individuals for many characters throughout the crop cycle, in a non-destructive manner, with high accuracy and moderate cost (ARAUS et al., 2014). According to Cabrera-Basquet et al. (2012), field phenotyping is currently a challenge for further advance in plant breeding. The quality (accuracy) and intensity (high throughput) of the methodologies implemented in measurements are fundamental in increasing the genetic progress (MASUKA et al., 2012; HAWKESFORD et al., 2013), specially related to abiotic stress (ARAUJO et al., 2015).

Among environment changes and consequently abiotic stresses, the rainfall pattern variation should have bring attention. In soybean, drought stresses can reduce yield especially at reproductive growth stage (BREDEVAN and EGLI, 2003; MUTAVA et al. 2015). Most of breeding programs face very difficult tasks to do a proper phenotyping under drought condition, and better understanding about components affecting yield is becoming crucial. Studies have indicate a strong association between soybean yield to physiological and morphological traits such as leaf chlorophyll content, canopy temperature, net photosynthesis, number of seeds per pod, number of pods per plants, number of branches per plant (KEEP et al., 2016; TODESCHINI et al.,

2016; TEODORO et al., 2019). Therefore, phenotyping of yield components and physiological traits are necessary for a more trustworthy selection.

Physiological and morphological understanding on high quality phenotyping will remain a keystone, and by default, for breeding (CABRERA-BASQUET et al., 2012). Most of the physiological and yield related traits are time consuming, difficult to obtain and, in most case, expensive in large scale evaluations. In this sense, phenotypic performances may infer indirectly about crop physiological conditions in each environment situation. In contrast, spectral data can be efficiently used in breeding programs where thousands of individuals must be screened and help understanding crop physiology. Based on the study of the interaction between electromagnetic radiation and the target, this technique is the most commonly used remote sensing procedure in automated phenotyping systems, since they are fast, non-destructive and cover large areas (ARAUS and CAIRNS, 2014).

Plant spectral reflectance or hyperspectral reflectance spectroscopy is based on the different pattern of light reflectance on leaves at different wavelengths through the photosynthetically active radiation (PAR, 400-700 nm), near infrared radiation (NIR, 700-1 200 nm) and shortwave infrared (up to 2500 nm) regions of the electromagnetic spectrum (ARAUS et al., 2015). The so-called spectral signature is the information extracted through the hyperspectral sensor from a given target (JENSEN, 2009). Through plant spectral, important information of biophysical and biochemical variables of the vegetation is possible to be obtained by building vegetation indices (VIs), which are simple operations (e.g., ratios and differences) between spectral reflectance data at given wavelengths (BARET and GUYOT, 1991). VIs are related to different plant characteristics such as photosynthetic active biomass, pigment content and water status, among other (MARTI et al., 2007; FOMAGGIO and SANCRES, 2017).

It is important to highlight the number of spectral bands of the sensors (multispectral or hyperspectral) that are determinant for the quantification and detailing of the different desired features. Multispectral sensors show information from a relatively smaller number of spectral bands (usually 3 to 20 bands). On the other hand, hyperspectral sensors usually are formed by a greater number of bands, with low wavelength amplitudes (between 1 to 10 nm) and continuously distributed along the spectral band (FORMAGGIO et al., 2017). For creation of new VIs, the use of hyperspectral sensors is recommended, as they extant a greater discrimination of subtle spectral responses (LEE et al., 2004; GOETZ et al., 1985). While the use of spectral reflectance techniques

as a high-throughput phenotyping tool is very promising, its application as a benchmark methodology requires further technical and logistic advances. Technical solutions have been proposed, like the use of spectroradiometers mounted on tractors. Aerial platforms are also very suited for field phenotyping. These include different options from remote controlled helicopters and “policopters”, airplanes to balloons or cranes and conveyers. Those kind of platforms might decrease the manpower and costs and making the acquisition of spectral information in a large number of plots much faster and more precise (CABRERA-BASQUET et al., 2012).

Advances in phenotyping are likely to be essential to capitalize on developments in conventional, molecular, and transgenic breeding and ensure genetic improvement of crops for future food security. The capacity for undertaking precision phenotyping, particularly under repeatable and representative growing conditions in the field, is lagging far behind the capacity to generate genomic information. So, investing in develop phenotyping platforms should become an integral and key component in the breeding pipeline to increase soybeans production in the future.

In this study, hyperspectral reflectance is used to predict physiological content and yield related traits of soybean cultivars, and the possibility of a screening methodology for genotype selection. Specifically, the objectives are to build partial least square regression (PLSR) models from leaf hyperspectral reflectance to predict and select soybean cultivars regarding (1) physiological traits measured under drought and well-watered conditions, and (2) yield related traits.

REFERENCES

- ACQUAAH, H. **Principles of plant genetics and breeding**. Malden, MA, Oxford, p. 584, 2009.
- ALEXANDRATOS, N. et al. **World agriculture towards 2030/2050: the 2012 revision**. FAO, Rome: ESA Working paper, 2012.
- ARAUJO, S. S. et al. Abiotic stress responses in legumes: strategies used to cope with environmental challenges. **Critical Reviews in Plant Sciences**, v. 34, n. 1-3, p. 237-280, 2015.
- ARAUS, J. L.; CAIRNS, J. E. Field high-throughput phenotyping: the new crop breeding frontier. **Trends in Plant Science**, v. 19, n. 1, p. 52-61, 2014.

ARAUS, J. L.; ELAZAB, A.; VERGARA, O.; CABRERA-BOSQUET, L.; SERRET, M. D.; ZAMAN-ALLAH, M.; CAIRNS, J.E. Novas tecnologias para fenotipagem. In: BORÉM, A.; FRITSCHÉ-NETO, R. (Ed.). **Fenômica: como a fenotipagem de próxima geração está revolucionando o melhoramento de plantas**. Viçosa, p. 11-29, 2015.

ASORO, F. G.; NEWELL, M. A.; BEAVIS, W. D.; SCOTT, M. P.; JANINK, J. L. Accuracy and training population design for genomic selection on quantitative traits in elite North American oats. **The Plant Genome**, v. 4, n. 2, p. 132-144, 2011.

BARET, F.; GUYOT, G. Potentials and limits of vegetation indices for LAI and APAR assessment. **Remote Sensing of environment**, v. 35, n. 2-3, p. 161-173, 1991.

BREVEDAN, R.E.; EGLI, D.B. Short periods of water stress during seed filling, leaf senescence, and yield of soybean. **Crop Science**, v. 43, n. 6, p. 2083-2088, 2003.

CABRERA-BOSQUET, L.; CROSSA, J.; ZITZEWITZ, J.; SERRET, M. D.; ARAUS, J. L. High-throughput phenotyping and genomic selection: the frontiers of crop breeding converge. **Journal of integrative plant biology**, v.54, n.5, p. 312-320, 2012.

CLARK, M.; TILMAN, D. Comparative analysis of environmental impacts of agricultural production systems, agricultural input efficiency, and food choice. **Environmental Research Letters**, v. 12, n. 6, p. 064016, 2017.

FORMAGGIO, A. R.; SANCHES, I. D. **Sensoriamento remoto em agricultura**. São Paulo, SP: Oficina de texto, 2017.

GOETZ, A.F.; VANE, G.; SOLOMON, J.E.; ROCK, B.N. Imaging spectrometry for earth remote sensing. **Science**, v. 228, n. 4704, p. 1147-1153, 1985.

HARTMAN, G.L.; WEST, E.D.; HERMAN, T.K. Crops that feed the World 2. Soybean—worldwide production, use, and constraints caused by pathogens and pests. **Food Security**, v.3, n.1, p.5-17, 2011.

HAWKESFORD, M.J.; ARAUS, J. L.; PARK, R.; CALDERINI, D.; MIRALLES, D.; SHEN, T.; PARRY, M. A. 2013. Prospects of doubling global wheat yields. **Food and Energy Security**, v.2, n.1, p. 34-48, 2013. doi: 10.1002/fes3.15

JENSEN, J. R. **Remote sensing of the environment: An earth resource perspective 2/e**. Pearson Education India, 2009.

KEEP, N. R. et al. Changes in physiological traits in soybean with breeding advancements. **Crop Science**, v. 56, n. 1, p. 122-131, 2016.

LEE, K.S.; COHEN, W.B.; KENNEDY, R.E.; MAIERSPERGER, T.K.; GOWER, S.T. Hyperspectral versus multispectral data for estimating leaf area index in four different biomes. **Remote Sensing of Environment**, v. 91, n. 3-4, p. 508-520, 2004.

MARTI J., BORT J., SLAFER G., ARAUS, J.L. Can wheat yield be assessed by early measurements of Normalized Difference Vegetation Index?. **Annals of Applied biology**, v. 150, n. 2, p. 253-257, 2007.

MASUKA, B.; ARAUS, J. L.; SONDER, K.; DAS, B.; CAIRNS, J. E. Deciphering the code: Successful abiotic stress phenotyping for molecular breeding. **Journal of integrative plant biology**, v. 54, p. 238-249, 2012.

McMCOUCH, S.R.; McNALLY, K.L.; WANG, W.; HAMILTON, R.S. Genomics of gene banks: a case study in rice. **American journal of botany**, v. 99, n. 2, p. 407-423, 2012

MELILLO, J.M.; TERESE, T.C.; RICHMOND, G.; YOHE, W. (Ed.). **Climate change impacts in the United States: the third national climate assessment**. Washington, D.C. 2014

MUTAVA, R.N.; PRINCE, S.J.K.; SYED, N.H.; SONG, L.; VALLIYODAN, B.; CHEN, W.; NGUYEN, H.T. Understanding abiotic stress tolerance mechanisms in soybean: a comparative evaluation of soybean response to drought and flooding stress. **Plant Physiology and Biochemistry**, v. 86, p. 109-120, 2015.

PELANICHANY, D; COBB, J, N. Fenômica de caracteres agronômicos a campo. In: BORÉM, A.; FRITSCHÉ-NETO, R (Ed.). **Fenômica: como a fenotipagem de próxima geração está revolucionando o melhoramento de plantas**. Viçosa, p. 11-29, 2015

SEDIYAMA, T.; SILVA, F.; BORÉM, A. (Ed.). **Soja: do plantio à colheita**. Viçosa: UFV, p.333, 2015.

SILVA, F. L.; BORÉM, A.; SEDIYAMA, T; LUDKE, W. **Soybean Breeding**. Springer, p.440, 2017.

SINGH, P.; KUMAR, R.; SABAPATHY, S.N.; BAWA, A.S. Functional and edible uses of soy protein products. **Comprehensive reviews in food science and food safety**, v. 7, n. 1, p. 14-28, 2008.

TEODORO, L. P. R. et al. Understanding the combining ability for physiological traits in soybean. **PloS one**, v. 14, n. 12, p. e0226523, 2019.

TODESCHINI, M. H. et al. Soybean genetic progress in South Brazil: physiological, phenological and agronomic traits. **Euphytica**, v. 215, n. 7, p. 124, 2019.

TUBEROSA, R. Phenotyping for drought tolerance of crops in genomics era. **Frontiers in Physiology**, v.3, p. 347, 2012.

UNITED STATE DEPARTMENT OF AGRICULTURE – USDA. Economic Research Service. Disponível em: <https://www.usda.gov> . accessed in February 2018.

Chapter I - ESTIMATION OF PHYSIOLOGICAL TRAITS AND SELECTION OF SOYBEAN CULTIVARS UNDER DROUGHT CONDITION THROUGH LEAF HYPERSPECTRAL REFLECTANCE

ABSTRACT

Understanding physiological traits would increase efficiency for improving crops toward high-yielded varieties, especially when the selection is made in adverse conditions for abiotic stresses. Traditional methods in access physiological and biochemical information of plants prevent its use on most breeding program pipelines because they are costly and time consuming. Hyperspectral data and statistical approaches such as partial least square regression (PLSR) can be applied to rapidly collect and predict several physiological parameters at canopy level, allowing phenotyping several genotypes in a high-throughput manner. In this study, we investigate the potential of hyperspectral data in phenotyping physiological traits and select best performance soybean cultivars under drought condition. Twenty cultivars were evaluated under well-watered and drought conditions during the reproductive phase in controlled environment conditions. At 20 days after drought establishment, we measured chlorophyll a content (CLa), chlorophyll b content (CLb), carotenoids content (CAR), specific leaf area (SLA), and electrons transfer rate (ETR). At 28 days after drought impose, we measured CLa, CLb, CAR, SLA, relative water content (RWC), and leaf temperature (TEM). Leaf hyperspectral reflectance of the cultivars were collected at each evaluation date over the spectral region between 350 and 2500 nm. PLSR and partial least square discriminant analysis (PLSDA) techniques were used to estimate physiological traits and to select the best cultivars, respectively. Variable importance in projection was used to select the most associate wavelengths with each analysis. PLSR models allowed accurate prediction of CLa, CLb, CAR, SLA, TRE, and TEM (high values of R^2 cross validation and low values of root mean square error of cross validation). Accurate PLSDA model was calibrated using 50 wavelengths for selection of cultivars that are potentially more productive regarding all evaluated physiological traits. The results show the great potential of using hyperspectral reflectance as a feasible, nondestructive, and accurately method to estimate physiological traits and on screening of superior genotypes.

Key-words: remote sensing, multi-trait selection, PLSR

1. INTRODUCTION

Despite the improvement of soybean (*Glycine max* [L.] Merrill) yield in last decades, some concerns arise in terms of the environmental and socio-economic impact of global market and land use trade off (Boerema et al., 2016). Additionally, climate change may disturb soybean production while affecting the capacity to supply oil and protein demand in agriculture (Bhattarai et al., 2017). Drought is currently one of the most important factors influencing edaphoclimatic stress conditions and is responsible for some losses for crop yield in recent years (Fuganti-Pagliarini et al., 2017; Nakashima et al., 2018). Breeding high-yielded cultivars with a good level of drought tolerance should be one of the most efficient ways to cope with the new agriculture challenges regarding water deficit situation.

Understanding how crops deal with drought conditions is a very difficult task in conventional breeding approach, since there are several genes involved in drought tolerance and identifying genotypes with such characteristic in a large-scale screening depends on several factors including plant, soil and atmosphere interaction (Bhatnagar-Mathur et al., 2004). Even though drought tolerance phenotyping has been based on yield performance, minor effect genes may not be considered while introduction of new genetic diversity is often avoided (Eagles et al., 2014; Mutava et al., 2015). Physiological understanding could help breeders better screening for abiotic stress because it could handle the knowledge of well-characterized genotypes that may have favorable alleles which are not taking into account when selection is based only on yield performance (Reynolds & Langridge, 2016).

Phenotyping physiological traits demands time, intense and qualified labor, which leads to either poor resolution characterization or limited number of evaluated lines (Ghanem et al., 2015; Monneveux et al., 2012). Therefore, feasible physiological characterization in breeding programs heavily rely on high-throughput phenotyping approaches (Halperin et al., 2017; Reynolds & Langridge, 2016). Currently considered in most of the high-throughput phenotyping framework, remote sensing is paving the way in agriculture because it is fast, non-invasive and could be used in large scale areas to assess physiological and biochemical traits (Araus et al., 2018). Remote sensing devices used for phenotyping include digital cameras, as well as thermal, fluorescence, multispectral, and hyperspectral sensors, which can be imagers or non-imagers (Araus & Cairns, 2014; Deery et al., 2014; Fahlgren et al., 2015).

Although several physiological traits have been accessed using hyperspectral data, such as chlorophyll content, leaf water potential, stomatal conductance, leaf area index (Kanning et al., 2018; Rapaport et al., 2015; Wu et al., 2008), it is crucial to identify wavelengths related to more than one trait, since high-yielded cultivars associated with drought tolerant likely show a good physiological adaptation in water limited environment. Different from other sensors that have focused in few wavelengths and consequently limit its use for some vegetative indices (visible/multispectral sensors), hyperspectral sensors can measure hundreds of spectral bands between 350 – 2500 nm, which are sample at very high resolution. Such technology could be very useful in detecting several traits in different environment conditions (Mahlein et al., 2012; Wang et al., 2019). Additionally, it needs to be better studied to enable its use as a fast phenotyping method.

We aim to use hyperspectral data to assess physiological traits in soybean cultivars and select those cultivars with better responses in the water-limit situation regarding more than one physiological trait simultaneously. The expected findings of this study may assist on assembly of ground and aerial platform that allow high-throughput phenotyping.

2. MATERIAL AND METHODS

Twenty soybean cultivars were evaluated at Horta Nova experimental station of the Agronomy Department of the Federal University of Viçosa (20° 45' S e 42° 49' W, alt 680 meters). Additional information about them can be found in Table S1 of the supplemental material.

Two trials were conducted simultaneously so that the cultivars in one trial were well-watered, and in the other trial, the cultivars were submitted to drought condition. The trials were planted within a greenhouse compose with galvanized arches and covered with a transparent plastic film of 150 microns. The plastic could be moved up and down so one could prevent from rain. The plots in both trials were consisted of one-meter line, and spaced by 0.50 cm. An independent drip irrigation system was installed so the trials could be irrigated separately. Soil type is classified as dystrophic Red Yellow Latosol (Embrapa, 2013) with a clay texture.

Diary watering necessity was calculated based on the crop evapotranspiration (ET_c), in mm day⁻¹, watering intensity, in mm h⁻¹, and soil/water retention curve obtained from soil sampled

before the trials were planted. Parameters used to ETc calculation were obtained through a weather station (E5000, Irripius, Viçosa, Brazil) installed within the greenhouse. The cultivars in both trials were carried out according to agronomic demand, pests control, and with 100% of soil moisture capacity. When most of the cultivars reach R3 phenological phase (Fehr & Caviness, 1977), one of the trials had its irrigation system stopped while the other trial was kept well-watered. It was done to guarantee the establishment of drought condition when the cultivars reach R5 phenological phase, which is characterized as seed fill and is known as a critical phase for water demand (Fehr & Caviness, 1977). Both trials were equally watered again after 30 days of suspended irrigation from one of the trials.

2.1. Physiological traits estimation

2.1.1 Physiological traits measurements

The cultivars in both trials were evaluated in two dates: the first was made 20 days after drought impose (20 DAD) and the second 28 days after drought impose (28 DAD). Physiological traits assessed at 20 DAD were: chlorophyll a content (CLa), chlorophyll b content (CLb), carotenoids content (CAR), specific leaf area (SLA), and electrons transfer rate (ETR) between photosystem II and photosystem I. Traits assessed at 28 DAD were: chlorophyll a content (CLa), chlorophyll b content (CLb), carotenoids content (CAR), specific leaf area (LSA), relative water content (RWC), and leaf temperature (TEM).

From each plot, one fully expanded leaflet was collected from one leaf located at middle region of a plant. From each leaflet, a total of six discs with 5 mm diameter were sampled and immersed in 5mL of dimethylsulfoxide (DMSO) solution sutured with calcium carbonate stored in dark location for 48 hours. The solution was transferred to 1mL quartz bucket and its absorbance read through Genesys 10UV spectrophotometer (ThermoScientific, EUA). The values of absorbance were used to calculate the content of CLa ($\mu\text{g cm}^{-2}$), CLb ($\mu\text{g cm}^{-2}$), and CAR ($\mu\text{g cm}^{-2}$), according to Wellburn (1994).

SLA was measured by taking another fully expanded leaflet belonging to the same leaf used to quantify the pigments. Four discs (5 mm diameter) were sample and dried at 70° C for 72 hours to measured the dried mass. LSA ($\text{cm}^{-2} \text{ g}^{-1}$) was calculated dividing leaf area by dry leaf mass as described in (Evans & Poorter, 2001). The same discs were used to measure RWC, calculated

according to (Barrs & Weatherley, 1962). The equation is based on fresh leaf mass, dry leaf mass, and full turgor leaf mass, and is given by:

$$\frac{FW - DW}{TW - DW} \times 100 \quad (1)$$

where FW, TW, and DW represent fresh weight (g), turgid weight (g), and dry weight (g), respectively.

A pulse amplitude modulated fluorometer (Mini-PAM, Walz, Germany) was used to measure PAR, and calculated ETR. A leaf clip holder was connected to Mini-PAM and used to collect data from the fourth leaf (soil level) of the plant, avoiding the main vein. The rapid light curve, generate by Mini-PAM, consist of the fluorescence responses to nine different actinic irradiances of 10 s duration, resulting in relevant fluorescence parameters used to calculate ETR.

The ETR values are calculated as follow:

$$\frac{F_m' - F}{F_m} \times PAR \times 0.5 \quad (2)$$

where $F_m' - F / F_m$ is the photochemical yield in the light, PAR is the photosintetically active irradiation, 0.5 is a multiplicative factor used beacuse only half of the absorbed quanta are distributed to the Photosystem II (Figueredo et al., 2009).

Leaf matric potential level was measured at evaluation date (20 DAD) taking one leaflet of one of the plants in each plot. The process was done at predawn and the measurements was done using a pressure chamber (Sholander et al., 1965). Soil matric potential level was measured through three samples collect at 0-20 cm and three at 20-40 cm soil depth, in each of the trials. Soil humidity was estimated from the sampled soil and the matric potential level was calculated throughout a retention curve. Soil matric potential level was measured at both evaluation date (20 DAD and 28 DAD).

2.1.2 Spectral signature collection

The spectral signatures of the cultivars were collected using another fully expanded leaflet during the sampling days (20 DAD and 28 DAD). The collected leaflets from each cultivar was placed within plastic bag which was placed within a iced box to avoid further physiological degradation. The platics bags did not have direct contact with the ice, avoinding cold stresses. The iced box was transported to the lab within 50 minutes to collect the spectral signature from each

leaflet, using a hyperspectral ASD FieldSpec® 4 Hi-Res (Analytical Spectral Devices Inc., Boulder, CO, USA) spectroradiometer, ranging from 350-2500 nm, with resample spectral resolution of 1.4 nm at 350-1000 nm and 1.1 nm at 1001-2500nm. Leaf spectral reflectance was measured using a contact prob connected to spectroradiometer, with 10 measurements taken on each leaflet. Each cultivar's Spectral signature at each evaluation date was recorded as the mean of the 10 spectral measured from each leaflet. The spectral signatures were corrected using the splicing procedure to avoid spectra noises caused by different detectors arrays in the spectroradiometer.

2.1.3 Partial Least Square Regression Analysis

For each evaluation date, partial least square regression (PLSR) was performed to estimate the physiological traits using cultivars spectral signature. This technique is useful when there is a large set of independent variables (i.e., wavelengths) predicting a small number of dependent variables, searching for a set of components (latent variable) that explain the maximum covariance between response variables and predictor variables (Abdi, 2003; Carrascal et al., 2009). To determine the optimal number of latent variable (LV), the lowest value of cross validation root mean square error (RMSECV) based on k-fold venetian blind cross validation was selected. In this cross validation, each fold is built taking samples from dataset of a k-multiple position until the end of the dataset. A tradition k-fold cross validation is trained with k-1 fold, and tested with the remained fold, rotating the latter until all of them have been trained. The average performance of the k cross validation models is computed, obtained a final RMSECV, which is calculated as follows:

$$\text{RMSECV} = \sqrt{\frac{\sum_{i=1}^n (y_i - y_p)^2}{n}} \quad (2)$$

where y_i and y_p represent the measured and predicted values of the traits for sample i ; n is the number of samples from the dataset.

After building the full spectrum PLSR model, the variable importance in projection (VIP) was calculated to optimize the number of wavelengths that better predict the response variables. Wavelengths with VIP greater than one were selected to build a new PLSR model (Wold et al., 2001). The selected wavelengths were used to build a new PLSR model, choosing a new set of LV

based on the lowest value of RMSECV. A new VIP was calculated and wavelengths selection criteria was maintain the same as before ($VIP > 1$). This iterative process of building a new PLSR using selecting wavelengths from latest model was repeated until we obtain the highest value of coefficient of determination of cross validation ($R^2 CV$) and the lowest value of RMSECV.

PLSR models built at 20 DAD were cross validated with data of Cla, Clb, CAR, and SLA, measured at 28 DAD. This external validation aims to verify model robustness and predictive capacity of content from some of the physiological traits evaluated.

2.2 Prediction of selected cultivars via spectral signature

2.2.1 Cultivars selection using phenotypic data

The cultivar performance considering phenotypic data of all physiological traits simultaneously was estimated using the FAI-BLUP selection index (Rocha et al., 2018). This index uses a genotype-ideotype distance to rank the cultivars regardless of collinearity among the traits and without requirement to assign weights. Each cultivars' genotypic mean is used to create an ideotype based on desirable and undesirable values from each physiological trait. In order to ranking robust and highly performance cultivars, the ideotype was built considering the maximum value of CLa, CLb, CAR, ETR, RWC, and PAR, and minimum value of SLA and TEM in drought condition dataset, at the two evaluation dates. The top five cultivars were selected using this index, corresponding to a selection intensity of 25%.

The genotypic mean of the cultivars used in the FAI-BLUP index was through the following mixed-model:

$$y = X\tau + Z_g\mu_g + Z_i\mu_i + e \quad (2)$$

where y is phenotypic data vector; τ is the fixed effects vector (overall mean, experiments, and experiments blocks); μ_g is the cultivars effects vector (random), with $\mu_g \sim N(0, \sigma_g^2)$; μ_i is the cultivars by experiment interaction (random), with $\mu_i \sim N(0, \sigma_i^2)$; e is the residual effects vector (random), $e \sim N(0, R)$; X , Z_g , and Z_i are incidence matrix for τ , μ_g , and μ_i , respectively. R is a diagonal matrix with the residual variances, that is, $R = \bigotimes_{i=1}^t \sigma_{e_i}^2$ where $\sigma_{e_i}^2$ is the residual variance in the i experiment and t is the number of experiment, which in this case are two.

Cultivars and cultivars by experiment interaction effects were tested via maximum likelihood ratio, considering 1% of probability. Analysis were made using R software (R Development Core Team, 2005), using package '*sommer*'.

2.2.2 *Partial Least Squared Discriminant Analysis (PLSDA)*

We used the spectral signature of the cultivars to verify the efficiency in predicting the selected cultivars from the FAI-BLUP analysis. For that purpose, a PLSDA was carried taking two different classes as categorical variable, the selected cultivars and non-selected cultivars. This method is used for classification purposes in multivariate discriminant analysis (Barker & Rayens, 2003), so a discrimination between classes can be found and the model parameters can be used to discriminate cultivars.

The suitable number of LV was chosen based on the highest value or R^2 CV after performing a venetian blind cross validation, taking one sample out from the dataset. VIP was used to select the best wavelengths and a new PLSDA was performed again. These steps were proceeded until obtain the minimum number of wavelengths associated with the highest R^2 CV. PLSDA models were built for 20 DAD data, 28 DAD data, and for mean data from both dates.

Discrimination capacity from the cross validation results were measured using specificity (capacity to correctly identify the absence of a class when the class is not present) and sensitivity (capacity to correctly identify a class when the class is present). In other word, specificity is computed as a ratio of true negatives to the total of true negatives and false positives, and sensitivity is computed as a ratio of true positives to the total of true positives and false negatives.

3. RESULTS

The matric potential level in soil at 20 DAS was -134 kPa in the well-watered condition and -436 kPa in the drought condition. In the second evaluation date (28 DAS), the matric potential water in soil was -152 kPa and -363 kPa in the well-watered and drought conditions, respectively (Figure S1). The matric potential on the cultivars measured on leaves before dawn at 20 DAS, were -392 kPa on well-watered cultivars and -596 kPa on drought condition. Those values were -304 kPa on well-watered condition and 650 kPa on drought condition at 28 DAS. The results of matric potential water in either soil or leaf indicate difference between the water conditions in which the cultivars were evaluated in both dates.

Daily information about air relative humidity, temperature, and solar radiation during the evaluations can be found at Fig. 1. The relative air humidity was around 80% at the two evaluated dates, while there was an increase more than one Celsius degree and decrease in radiation of 265 MJ h⁻¹ between the two evaluation dates.

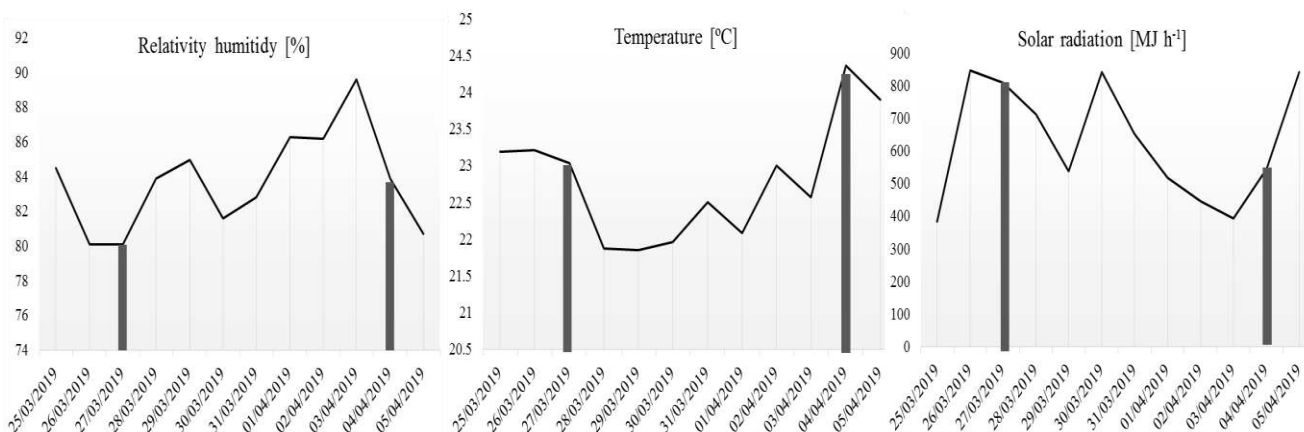


Fig 1. Daily information in the evaluations window regarding air relative humidity, in percentage, temperature, in Celcius degree, and solar radiation, in Mega Joules per hour. Grey bars highlight the values at the two evaluations dates.

Comparison of mean values of the cultivars in watered and drought conditions showed greater mean values in watered condition for CLa (Fig 2A and Fig 2F), CLb (Fig 2B and Fig 2G), and CAR (Fig 2C and Fig 2H). Also, cultivars at drought condition showed significant increase in TEM at 28 DAD (Figure 2J). No significant difference was observed for SLA (Fig 2D and Fig 2I), RWC (Fig 2K), and ETR (Fig 2E).

With exception for RWC, significant genotypic effect was found for CLa, CLb, CAR, SLA, and TRE at the 20 DAD evaluation. Yet for 28 DAD evaluation, significant genotypic effect was found for CLa, CLb, CAR, and TEM, and no significant genotypic effect for ELA and RWC. Those results point out to the presence of genetic variability among the cultivars and therefore feasibility to select the best ones in terms of physiological traits. The genotypic x condition interaction was not significant for the evaluated traits in both dates, which indicates that the drought condition affects all the cultivars similarly and consequently cultivars ranking do not change between conditions.

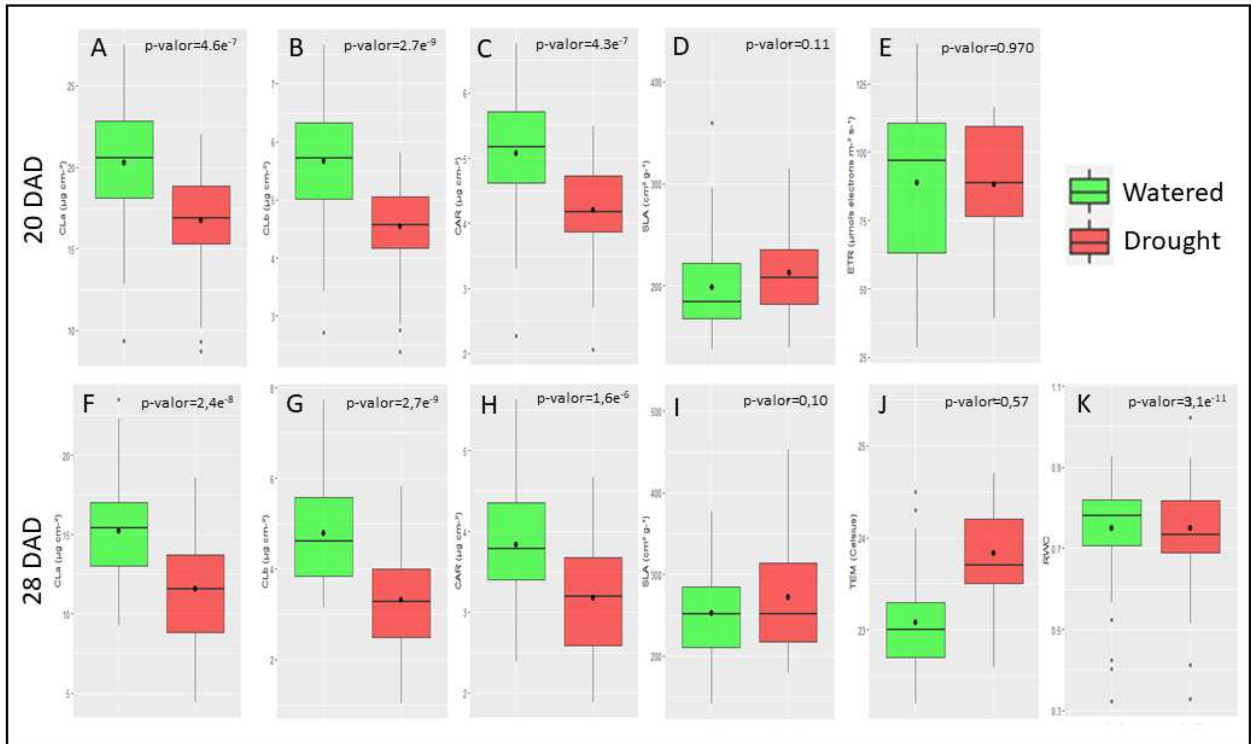


Fig 2. Boxplot distribution of physiological traits of cultivars under watered (green boxes) and drought (red boxes) conditions at 20 days after drought (DAD) (A-E), and at 28 DAD (F-K). (A) and (F), chlorophyll a; (B) and (G), chlorophyll b; (C) and (H), carotenoids; (D) and (I), specific leaf area; (E), electrons transport rate; (J), temperature; (K), relative water content

3.1 Physiological traits estimation using hyperspectral data

Spectral mean reflectance for the cultivars in well-watered and drought conditions are found in Fig 3. Higher reflectance values were usually observed for cultivars in drought conditions, especially when the drought was extended (28 DAD). Those difference between the means were more evident at the visible region (400- 700 nm), mostly in near infrared region (700 -1300 nm), and some wavelengths around the 1400 nm, 1600-1800 nm and 2200nm, which belong to the short-wave infrared (1300-2500 nm).

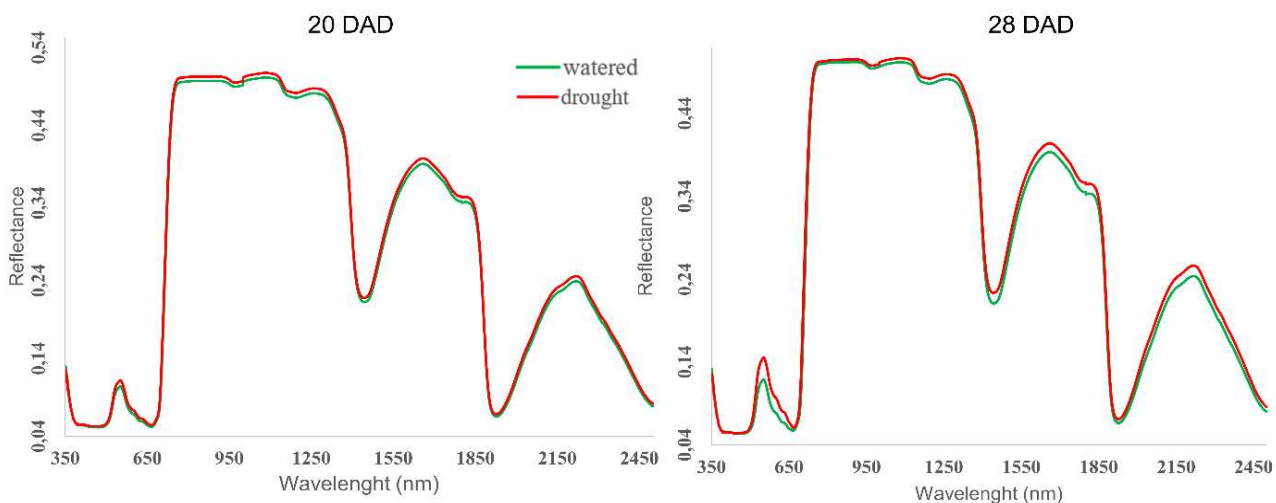


Fig 3. Means of spectral signatures of the cultivars in watered (green) and drought (red) conditions at 20 days after drought impose (DAD), and 28 DAD.

PLSR models presented decent R^2 calibration and cross validation values for CLa, CAR, and SLA at 20 DAD (Fig. 4A, 4C, and 4D, respectively). More robust models parameters for these three traits were found at 28 DAD, having R^2 cross validation values higher than 80%, associated with lower values of RMSECV (Fig. 5A, 5C, and 5D, respectively). On the other hand, better result for PLRS model was found at 20 DAD for estimating CLb when comparing with values of 28 DAD (Fig. 4B and 5B). Reasonable result was also found for predicting RWC at 28 DAD, having values of R^2 calibration higher than 88% and R^2 cross validation higher than 61% (Figure 5E). No such good results was possible to be found for ETR (R^2 cross validation = 38.4%) (Fig. 4E). Strong PLSR model was built to estimate TEM, with value of R^2 cross validation of 79.9% associated with a relative low value of RMSECV = 0.19 Celcius degree (Fig 5F). Models adjustment can also be checked taking the one-by-one line as a reference. Values near this line indicate better fitness of the models. In general, a segregation between estimated cultivars values in well-watered and drought condition can be visualized for leaf pigments traits (Fig. 4A, 4B, 4C, 5A, 5B, and 5C) and for TEM (Fig. 5F).

Information about pre-processing parameters, number of LV, cross-validation methods, R^2 calibration, R^2 validation, RMSE, RMSECV values, and the final number of wavelengths used from each PLRS model build in both evaluation dates can be checked at Table S2.

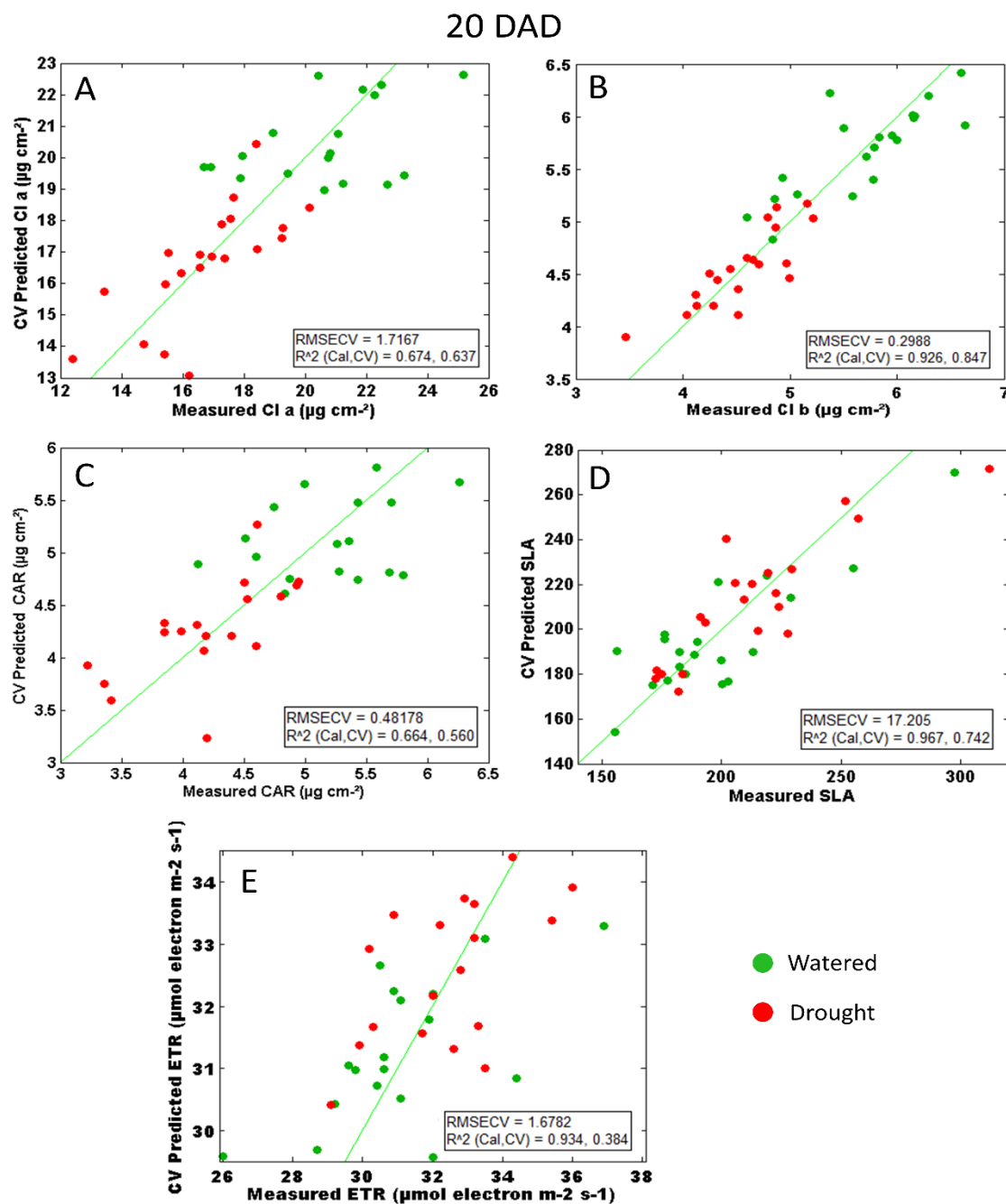


Fig 4. Observed vs predicted values from PLSR models at 20 days after drought (DAD) (A-E). (A) and, chlorophyll a; (B), chlorophyll b; (C) carotenoids; (D), specific leaf area; (E), electrons transport rate. Green dots are cultivars in watered condition and red dots represents cultivars in drought condition.

28 DAD

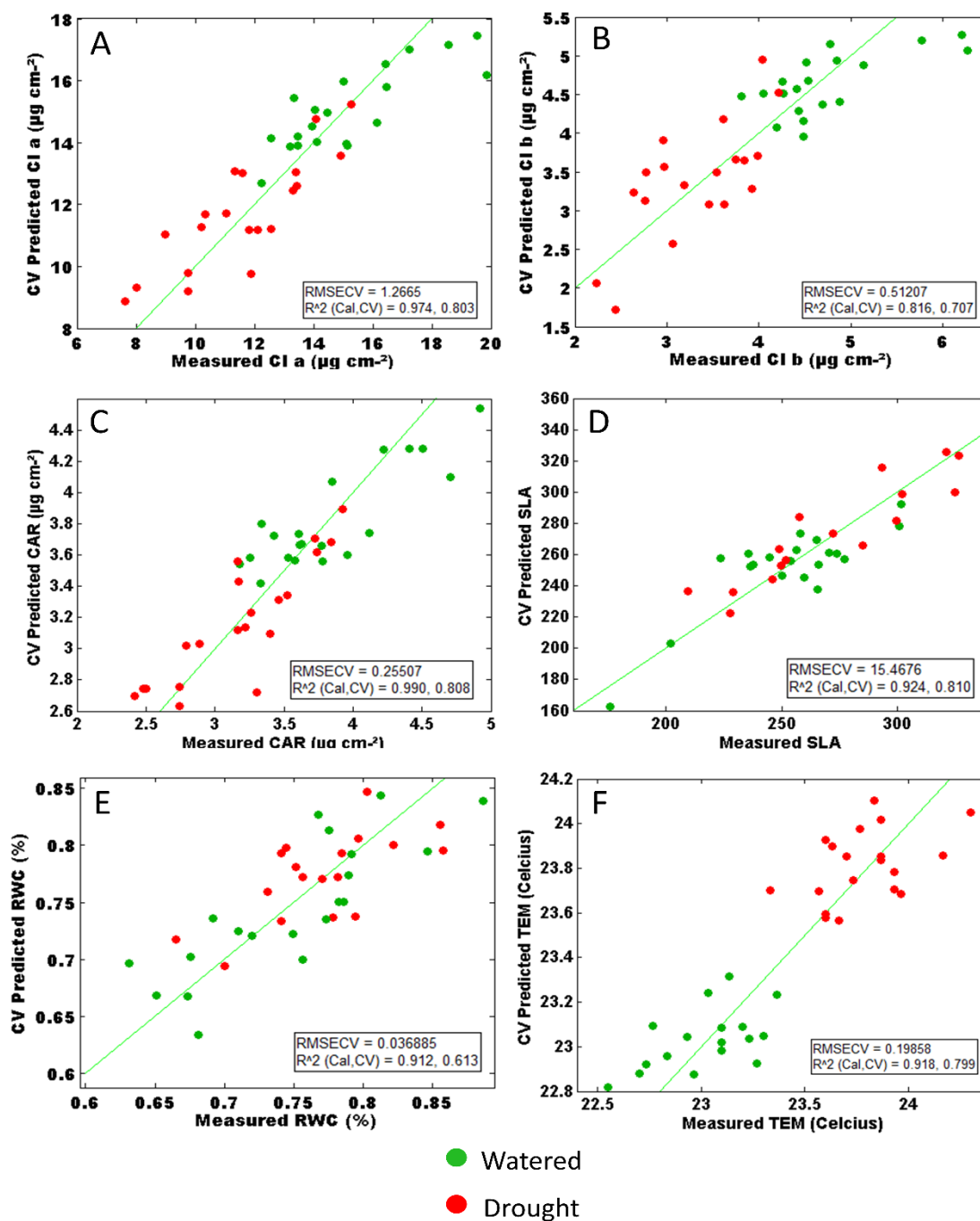


Fig 5. Observed vs predicted values from PLSR models at 28 days after drought (DAD) (A-F). (A) and, chlorophyll a; (B), chlorophyll b; (C) carotenoids; (D), specific leaf area; (E), relative water content; (F), leaf temperature. Green dots are cultivars in watered condition and red dots represents cultivars in drought condition

Results from external cross validation of the models built from 20 DAD data indicated reasonable capacity to predict CLa, CLb, and CAR values from 28 DAD data, having R^2 values ranging from 61.4% for CAR to 70.7% for CLa. Those relatively high R^2 values were also associated with low root mean square error of prediction (RMSEP). On the other hand, no such good result was found for SLA, having a R^2 value of just 26.6% (Fig 6). These results show that the built models are consistent to predict leaf pigment content and that ground data and spectral data can be measured in different dates.

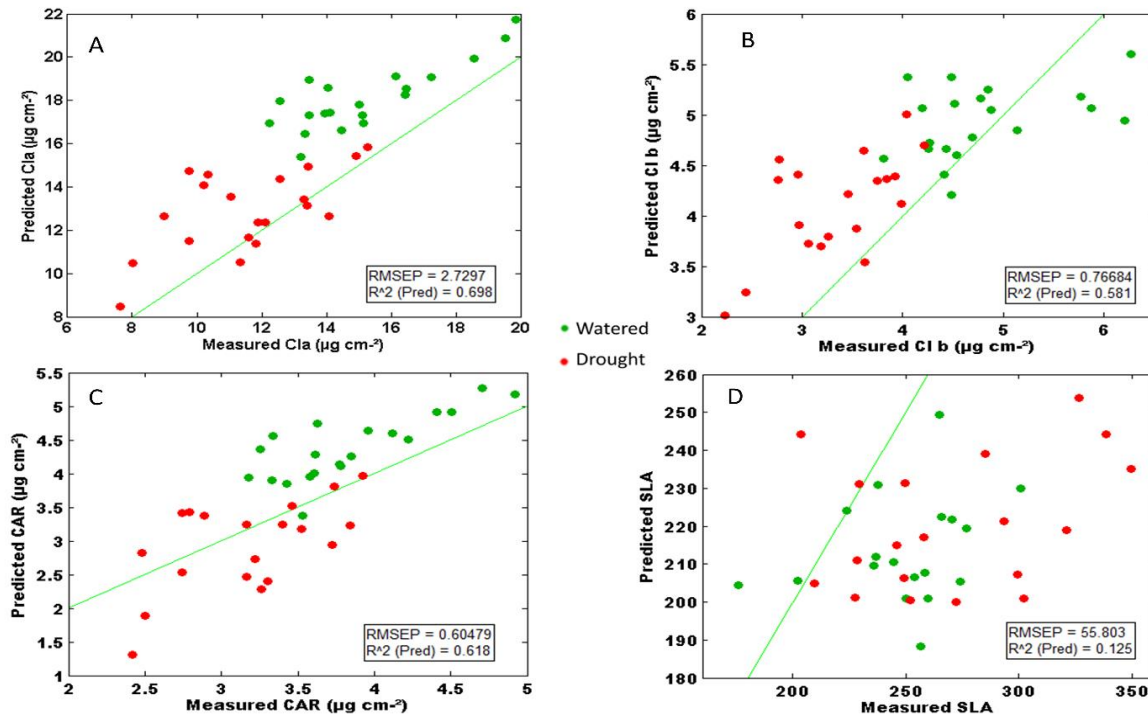


Fig 6. Measured values from 28 DAD data versus predicted values from PLSR models built from 20 DAD for (A) chlorophyll a content; (B) chlorophyll b content; (C) carotenoid content; and (D) specific leaf area. Green circles represent cultivars in watered condition while red circles represent cultivars in drought condition.

3.2 Cultivars selection through spectral data

FAI-BLUP cultivars classification according to physiological traits ideotype are showed at Figure 7A. The five best cultivars (CONQUISTA, NK 7059, CD 244, FUNDACEP 57, and BRS 185) were set as a selected cultivar while the rest of the cultivars were set as a unselect cultivars regarding desirable physiological traits in drought and watered conditions. Model built using data from 28 DAD result in one cultivar misclassified between the classes from the model built using

data from 28 DAD (Figure 7C), which result in 93% of model sensibility and 100% of specificity from selected cultivar class. Model built from 20 DAD data correctly set all cultivars according to their respectively class (Figure 7B), corresponding to 100% of both specificity and sensibility. Same results were found when the mean of the spectral signature from both data analysis was used to build the PLS-DA (Figure 7D). These results show that the selected wavelengths can be used as criteria for selection soybean cultivars remotely regarding physiological traits.

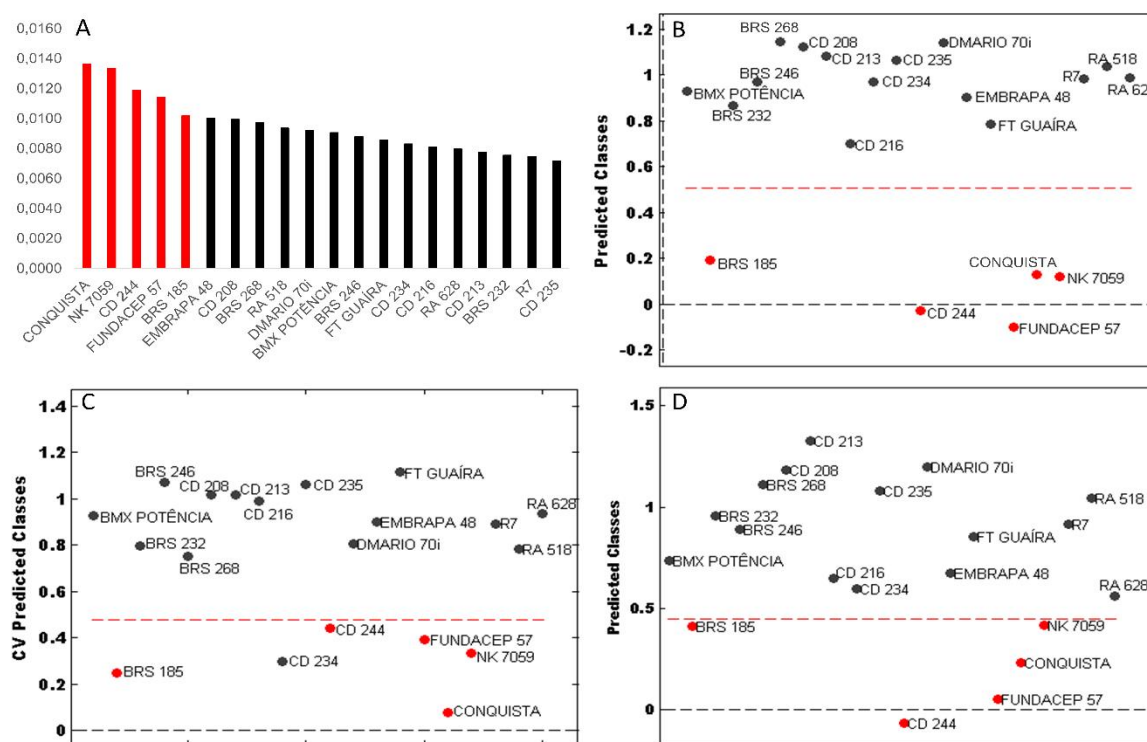


Fig 7. Ranking of the cultivars according to FAI-BLUP index. The class of selected cultivars are set in red while the classes of non-selected cultivars are set in black (A); Classification of the cultivars into the selected and non-selected classes from cross validation of partial least square regression discriminant model using spectral data of 20 DAD (B), 28 DAD (C), and spectral mean data of both dates (D).

The selection norms of PLS-DA models were similar to those of PLSR, which means that they were based on the lowest value of RMSECV and the highest value of cross-validation R^2 . From the model built using spectral mean of both evaluation data, 50 were selected through the

VIP parameter in the final iterative PLS-DA model. Most of these wavelengths were found at the visible range of the spectrum, especially at the blue region. Additionally, wavelengths in NIR and SWIR were also useful in the discriminant analysis (Fig 8). Autoscaling, multiway center, and second order smoothing were used as preprocessing, three latent variables were chosen to build the model and a 3-split venetian blind with 1-fold was used as cross-validation method. Information about PLS-DA parameter and the wavelengths selected for the final model can be found also at Table S3.

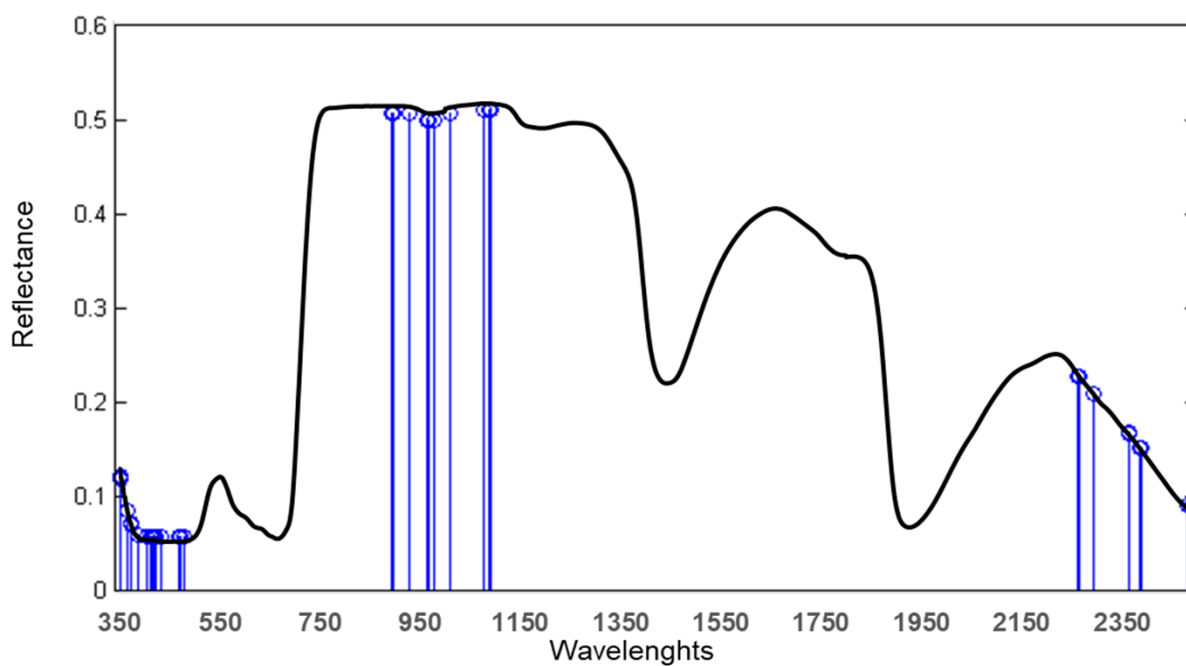


Fig 8. Spectral mean signature of the cultivars and the wavelengths selected (blue bars) in the final PLS-DA model using the variable importance in projection (VIP).

4. DISCUSSION

4.1 Drought effects on physiological traits

Matric potential values of the soil and of the plants evidence a substantial difference between the cultivars in well-watered and drought conditions in terms of water availability in the

filling-pod stage of soybean (R5 stage). This fact could be confirmed through the difference between means from both conditions from majority physiological traits (Fig. 2). Drought can considerably affect yield while it limits prime physiological functions to soybeans achieve their real potential, especially when it happens at R5 stage (Manavalan et al., 2009). Despite all the cultivars have survived, the non-significant effect of cultivars x conditions interactions could be related to the duration of the drought (30 days), which have likely affected all cultivars similarly.

The reduction in leaf pigments traits of the evaluated cultivars indicates that they had their photosynthetic rates reduced due to the imposition of water deficit. In general, plants under water deficit face reductions in the photosynthetic rate because the drought stress damages the photosynthetic apparatus components (Reddy et al., 2004). Reductions in chlorophyll and carotenoid concentrations of soybeans plants under drought conditions have been reported by (Makbul et al., 2011), and similar results were observed for other crops such as sorghum (Nxele et al., 2017), peanuts (Meher et al., 2018), and barley (LI et al., 2006).

Moreover, the lower photosynthetic rate under water deficit can be attributed to the reductions in RWC, and ETR, associated with the increase in TEM and SLA (Crusiol et al., 2020; Fioreze et al., 2015; Iqbal et al., 2019). These traits might be useful for selecting genotypes with higher photosynthetic efficiency, both under normal and water stress conditions. Although we did not find any significant difference between the means of the experiments for SLA, ETR, and RWC, the results show that, in general, they were essential to discriminate genotypes in watered and drought condition. Therefore, they might be used to select genetic resources candidates for tolerance to water deficit.

Having access to physiological traits in a breeding program can help breeders in identified elements of success that genotypes use to further enhance adaptation in several environment situations and consequently helping target genetic variation in the germplasm to improve crop yields. Crossing parents using additional information from physiological traits might increase the number of favorable alleles in further generetaions as it increase the chance of finding complementary traits (Reynolds & Langridge, 2016; Richards, 2006). Physiological breeding is very important to continuing the improvement crops efficiently once it contribute to better genetic understanding and drives the selection of improved cultivars.

4.2 Physiological traits estimation via spectral signature

PLSR allows the use of full spectral information, which contributes to improve traits estimation based on biochemical and biophysical attributes (Hansen & Schjoerring, 2003; Yendrek et al., 2017). This technique focuses on finding factors (latent variables) that maximize the amount of variation explained by the wavelengths that are relevant to predict the analyzed traits (Wise et al., 2006). The wrong choice of the LV number can lead to an over-adjustment or under-adjustment, both undesirable. In our study, we chose the number of latent variables that resulted in lower values of RMSECV and high values of R^2 calibration and R^2 cross validation, resulting in models with adequate predictive capacity associated with high accuracy.

The significance of the wavelengths used for calibrating the model can be set using the VIP (Gomez et al., 2008; Wold et al., 2001). This parameter allows one to considerably reduce the number of wavelengths without losing the quality of the model by discarding the most redundant ones (Chong & Jun, 2005). The relatively high number of wavelengths selected for the evaluated traits shows how complex is to reduce the number of predictor variables drastically. On the other hand, PLSR model performance has shown great capacity to improve prediction of physiological traits compared with short-band vegetation indices (Hansen & Schjoerring, 2003; Yendrek et al., 2017). One other advantage is that PLSR models are built for specific crops and conditions, such as in this case for soybean in drought condition.

In general, cross validation results showed consistent estimation from most of the physiological traits considering the two evaluation dates. In addition, good predictive ability for leaf pigments traits are feasible, since we use PLSR model built from data of the first evaluation (20 DAD) as dataset for external validation from physiological traits measured at 28 DAD. These results corroborate the capacity and solidity of the models to predict physiological traits in different dates.

Improving plants in terms of physiology parameters is required to increase productivity while getting a better understanding of plant-environment interactions (Großkinsky et al., 2015; Reynolds & Langridge, 2016). Traditional methods to measure most physiological traits are costly, destructive, and time-consuming, preventing phenotyping these kinds of traits in a large-scale breeding program (Ghanem et al., 2015; Moshelion & Altman, 2015). However, spectral signature couple with PLSR modeling approach have shown great capability to rapidly and simultaneously phenotype numerous physiological traits in a cost-effective and non-invasively procedure.

4.3 Genotype selection

The FAI-BLUP selection index allows one to select genotypes based on a desirable ideotype built from the dataset itself (Rocha et al., 2018). Through factor analysis, this selection index deals with multicollinearity, so that it is possible to carry out the selection of genotypes even in the presence of correlated traits, as is the case of leaf pigments traits, for example. Furthermore, FAI-BLUP takes into account the genetic diversity between genotypes, which can be interesting in breeding pipeline aiming complementary crosses (Bernardo & Woodbury, 2020). The five cultivars selected in this study showed the most likely values of a desirable ideotype regarding physiological traits considering both drought and watered conditions, and can be used as possible parents in breeding programs to obtain more physiologically efficient cultivars.

The 50 wavelengths selected via PLS-DA analysis were able to select the same cultivars that would be selected by FAI-BLUP analysis. Sensitivity values highlight the probability that a cultivar will not be classified in a class that it does not belong. Simultaneously, specificity provides the probability that a cultivar will be classified correctly in the class to which it belongs. The high values for these two parameters highlight the robustness of the model generated. However, additional experiments still necessary to further explore validate the model, mainly considering cultivars with different genetic background (Meacham-Hensold et al., 2019)

Among the 50 wavelengths selected by the VIP in the construction of the PLS-DA model, those from the visible region are responsible for identifying basic differences in the concentration of leaf pigments (Yoder & Pettigrew-Crosby, 1995; J. Zhang et al., 2016). A large amount of wavelengths is highlighted in the ultraviolet and blue region (350-500 nm). Despite relevant information is often neglected due to technical sensors limitations, the blue region of the spectrum would be very useful once it is associated to maxima absorption of chlorophyll a and b, and β -carotene (Mahlein et al., 2012). Therefore, pattern changes of these pigments in leaf can be very sensitive in this region of spectral signature. The wavelengths found in the near-infrared region are associated with the structural part of the leaf tissue (Carley et al., 2018; Xu et al., 2009), while those present in the long-wave infrared may be related with cellulose and water (Clevers et al., 2010; Curran, 1989).

Studies using PLS-DA from spectral data to select traits related to biotic and abiotic stresses have already been successfully reported (Lee et al., 2014; Marquetti et al., 2016; K. Yu et al., 2018). Araus et al. (2018) highlight that methodologies based on hyperspectral data are very prominent to

be used as a high-throughput screening in a breeding program as it can address which genotypes are having better performance regarding several hard-to-measure traits such as physiological traits. This procedure differs from conventional breeding by spreading the phenotyping into a larger range of traits and allowing breeders to dissect trait and select plants with improved adaptation, besides of improve the chance to identify useful genes (Reynolds & Langridge, 2016; Shunmugam et al., 2018).

While hyperspectral reflectance is advancing to a new level of phenotyping requirement, its use still linked to high cost instruments which are difficult to handle at field base experiments (Roitsch et al., 2019). However, new high-resolutions sensors are rapidly coming up, letting to more cost-efficient products as they get more widely commercialized. Alternatively, with the advent of new software and hardware, low-cost yet efficient sensors can be manufactured using regular instruments (Lucieer et al., 2014; Uto et al., 2016). Such approach can be adopted to a next step in this study, using the selected wavelengths as a proxy to building new sensors.

SUPPLEMENTARY MATERIAL

Table S1. List of the 20 cultivars used in this study

Cultivar	Flower color	Pubescence color	Growth habit	Maturity group
BMX POTÊNCIA	White	Grey	Indeterminate	6.6
BRS 185	Purple	Grey	Determinate	6.8
BRS 232	Purple	Grey	Determinate	6.9
BRS 246	White	Brown	Determinate	7.2
BRS 268	White	Grey	Determinate	6.9
CD 208	White	Grey	Determinate	6.9
CD 213	Purple	Grey	Determinate	6.1
CD 216	White	Light brown	Indeterminate	5.5
CD 234 RR	Purple	Grey	Determinate	8.0
CD 235 RR	White	Brown	Semi-Determinate	6.4
CD 244 RR	White	Grey	Determinate	8.0
DMARIO 70i	White	Grey	Indeterminate	6.2
EMBRAPA 48	White	Grey	Determinate	6.8
FT GUAÍRA	NI*	NI	NI	NI
FUNDACEP 57 RR	White	Brown	Determinate	6.5
CONQUISTA	Purple	Brown	Determinate	8.1
NK 7059 RR	White	Grey	Indeterminate	5.9
R7	NI	NI	NI	SI
RA 518	Purple	Light brown	SemiDerteminate	6.4
RA 628	White	Grey	Indeterminate	6.4

*NI: No information

Table S2. Pre processing, number of latent variable (LV), cross validation method, R2 calibration and R2 cross validation, root mean square error of calibration (RMSEC) and root mean square error of cross validation (RMSECV) , and number of wavelengths selected via variable importance in projection (VIP) from the partial least square regression analysis for chlorophyll a (CLa), chlorophyll b (CLb), carotenoids (CAR), specific leaf area (SLA) at 20 days after drought (20 DAD) and 28 DAD, electrons transport rate (ETR), photosynthetic active radiation (PAR) at 20 DAD, and relative water content (RWC), and temperature (TEM) at 28 DAD.

20 DAD						28 DAD					
Pre processing	LV	Cross Validation	R ² Cal R ² Val	RMSEC RMSECV	VIP	Pre processing	LV	Cross Validation	R ² Cal R ² Val	RMSEC RMSECV	VIP
CLa 2 nd order smoothing Multiway center Autoscale	3	Venetian blind with 3 splits and blind thickness =1	0.674 0.637	1.607 1.716	146	CLa 2 nd detivative 2 nd order Autoscale	3	Venetian blind with 6 splits and blind thickness =2	0.974 0.803	0.452 1.266	874
CLb 2 nd detivative 2 nd order Autoscale	2	Venetian blind with 10 splits and blind thickness =1	0.992 0.857	0.074 0.289	165	CLb 1 st detivative 2 nd order Autoscale	2	Venetian blind with 10 splits and blind thickness =3	0.816 0.707	0.403 0.512	470
CAR 2 nd order smoothing Multiway center Autoscale	2	Venetian blind with 14 splits and blind thickness =1	0.664 0.560	0.417 0.481	159	CAR 2 nd detivative 2 nd order Autoscale	4	Venetian blind with 15 splits and blind thickness =1	0.989 0.808	0.057 0.255	796
SLA 2 nd detivative 2 nd order Autoscale	4	Venetian blind with 10 splits and blind thickness =2	0.967 0.741	6.07 17.205	269	SLA 1 st detivative 2 nd order Autoscale	4	Venetian blind with 12 splits and blind thickness =2	0.923 0.809	9.746 15.467	447
ETR 3 rd order 2 nd derivative Multyway scale	4	Venetian blind with 6 splits and blind thickness =2	0.933 0.384	0.549 1.678	333	RWC 2 nd detivative 2 nd order Autoscale	3	Venetian blind with 11 splits and blind thickness =1	0.911 0.647	0.017 0.035	180
PAR 2 nd order 1 st detivative Autoscale	3	Venetian blind with 11 splits and blind thickness =1	0.859 0.518	19.516 36.648	211	TEM Group scale 1 st detivative 2 nd order Autoscale	3	Venetian blind with 10 splits and blind thickness =2	0.915 0.783	0.127 0.205	151

Table S3. Pre processing, number of latent variable (LV), cross validation method, R² calibration and R² cross validation, and number of wavelengths selected via variable importance in projection (VIP) of the partial least regression discriminant analysis considering the mean wavelengths values of the both 20 DAD and 28 DAD evaluations

Pre processing	LV	Cross Validation	R ² Cal R ² Val	Selected wavelengths via VIP							
2 nd order smoothing	3	Venetian blind with 3 splits and blind thickness =1	0.969	353-355	367	374-375	389	408	414-420	424	435
Multiway center				473-474	482-483	896-898	930	968-969	981		
Autoscale				1012-1014	1080	1090-1094	2267-2269	2297-2298	2367-2369	2391-2392	2485-2486 2494

REFERENCES

- Abdi, H. (2003). Partial least square regression (PLS regression). *Encyclopedia for research methods for the social sciences*, 6(4), 792-795.
- Araus, J. L., & Cairns, J. E. (2014). Field high-throughput phenotyping: The new crop breeding frontier. In *Trends in Plant Science* (Vol. 19, Issue 1, pp. 52–61). Elsevier Current Trends. <https://doi.org/10.1016/j.tplants.2013.09.008>
- Araus, J. L., Kefauver, S. C., Zaman-Allah, M., Olsen, M. S., & Cairns, J. E. (2018). Translating High-Throughput Phenotyping into Genetic Gain. In *Trends in Plant Science* (Vol. 23, Issue 5, pp. 451–466). Elsevier Ltd. <https://doi.org/10.1016/j.tplants.2018.02.001>
- Barker, M., & Rayens, W. (2003). Partial least squares for discrimination. *Journal of Chemometrics*, 17(3), 166–173. <https://doi.org/10.1002/cem.785>
- Barrs, H., & Weatherley, P. (1962). A Re-Examination of the Relative Turgidity Technique for Estimating Water Deficits in Leaves. *Australian Journal of Biological Sciences*, 15(3), 413. <https://doi.org/10.1071/bi9620413>
- Bernardo, R., & Woodbury, S. P. (2020). BREEDING for QUANTITATIVE TRAITS in PLANTS Third Edition. <http://stemmapress.com>
- Bhatnagar-Mathur, P., Sharma, K. K., Devi, M. J., Serraj, R., Yamaguchi-Shinozaki, K., & Vadez, V. (2004). Evaluation of transgenic groundnut lines under water limited conditions.
- Bhattarai, M. D., Secchi, S., & Schoof, J. (2017). Projecting corn and soybeans yields under climate change in a Corn Belt watershed. *Agricultural Systems*, 152, 90–99. <https://doi.org/10.1016/j.agry.2016.12.013>

- Boerema, A., Peeters, A., Swolfs, S., Vandevenne, F., Jacobs, S., Staes, J., & Meire, P. (2016). Soybean trade: Balancing environmental and socio-economic impacts of an intercontinental market. *PLoS ONE*, 11(5). <https://doi.org/10.1371/journal.pone.0155222>
- Carley, D. S., Jordan, D. L., Dharmasri, C. L., Shew, B. B., Sutton, T. B., & Brandenburg, R. L. (2018). Examples of Differences in Red Edge Reflectance and Normalized Difference Vegetative Index Caused by Stresses in Peanut. *Crop, Forage & Turfgrass Management*, 4(1), 180042. <https://doi.org/10.2134/cftm2018.06.0042>
- Carrascal, L. M., Galván, I., & Gordo, O. (2009). Partial least squares regression as an alternative to current regression methods used in ecology. *Oikos*, 118(5), 681-690.
- Chong, I. G., & Jun, C. H. (2005). Performance of some variable selection methods when multicollinearity is present. *Chemometrics and Intelligent Laboratory Systems*, 78(1), 103–112. <https://doi.org/10.1016/j.chemolab.2004.12.011>
- Clevers, J. G. P. W., Kooistra, L., & Schaepman, M. E. (2010). Estimating canopy water content using hyperspectral remote sensing data. *International Journal of Applied Earth Observation and Geoinformation*, 12(2), 119–125. <https://doi.org/10.1016/j.jag.2010.01.007>
- Crusiol, L. G. T., Nanni, M. R., Furlanetto, R. H., Sibaldelli, R. N. R., Cezar, E., Mertz-Henning, L. M., Nepomuceno, A. L., Neumaier, N., & Farias, J. R. B. (2020). UAV-based thermal imaging in the assessment of water status of soybean plants. *International Journal of Remote Sensing*, 41(9), 3243–3265. <https://doi.org/10.1080/01431161.2019.1673914>
- Curran, P. J. (1989). Remote sensing of foliar chemistry. In *Remote Sensing of Environment* (Vol. 30, Issue 3, pp. 271–278). Elsevier. [https://doi.org/10.1016/0034-4257\(89\)90069-2](https://doi.org/10.1016/0034-4257(89)90069-2)

- Deery, D., Jimenez-Berni, J., Jones, H., Sirault, X., & Furbank, R. (2014). Proximal Remote Sensing Buggies and Potential Applications for Field-Based Phenotyping. *Agronomy*, 4(3), 349–379. <https://doi.org/10.3390/agronomy4030349>
- Eagles, H. A., Cane, K., Trevaskis, B., Vallance, N., Eastwood, R. F., Gororo, N. N., Kuchel, H., & Martin, P. J. (2014). Ppd1, Vrn1, ALMT1 and Rht genes and their effects on grain yield in lower rainfall environments in southern Australia. *Crop and Pasture Science*, 65(2), 159. <https://doi.org/10.1071/CP13374>
- EMBRAPA (Empresa Brasileira de Pesquisa Agropecuaria) (2013). Sistema brasileiro de classificação de solos. 3 ed. Brasília, DF, Embrapa, p. 353.
- Evans, J. R., & Poorter, H. (2001). Photosynthetic acclimation of plants to growth irradiance: The relative importance of specific leaf area and nitrogen partitioning in maximizing carbon gain. *Plant, Cell and Environment*, 24(8), 755–767. <https://doi.org/10.1046/j.1365-3040.2001.00724.x>
- Fahlgren, N., Gehan, M. A., & Baxter, I. (2015). Lights, camera, action: High-throughput plant phenotyping is ready for a close-up. In *Current Opinion in Plant Biology* (Vol. 24, pp. 93–99). Elsevier Ltd. <https://doi.org/10.1016/j.pbi.2015.02.006>
- Fehr, W., & Caviness, C. (1977). Stages of soybean development. Special Report. <https://lib.dr.iastate.edu/specialreports/87>
- Figueredo, C. C., Giani, A., & Filho, J. P. L. (2009). Photosynthetic capacity of three phytoplanktonic species measured by a pulse amplitude fluorometric method. *Brazilian Journal of Plant Physiology*, 21(3), 167–174. <https://doi.org/10.1590/s1677-04202009000300001>

- Fioreze, S. L., Vandeir, :, & Guimarães, F. (2015). MANUTENÇÃO DO STATUS HÍDRICO DE PLANTAS DE SOJA INDUZIDA POR ETIL-TRINEXAPAC. *Scientia Agraria Paranaensis*, 14(3), 166–172. <https://doi.org/10.18188/1983-1471/sap.v14n3p166-172>
- Fuganti-Pagliarini, R., Ferreira, L. C., Rodrigues, F. A., Molinari, H. B. C., Marin, S. R. R., Molinari, M. D. C., Marcolino-Gomes, J., Mertz-Henning, L. M., Farias, J. R. B., De Oliveira, M. C. N., Neumaier, N., Kanamori, N., Fujita, Y., Mizoi, J., Nakashima, K., Yamaguchi-Shinozaki, K., & Nepomuceno, A. L. (2017). Characterization of soybean genetically modified for drought tolerance in field conditions. *Frontiers in Plant Science*, 8, 448. <https://doi.org/10.3389/fpls.2017.00448>
- Ghanem, M. E., Marrou, H., & Sinclair, T. R. (2015). Physiological phenotyping of plants for crop improvement. In *Trends in Plant Science* (Vol. 20, Issue 3, pp. 139–144). Elsevier Ltd. <https://doi.org/10.1016/j.tplants.2014.11.006>
- Gomez, C., Lagacherie, P., & Coulouma, G. (2008). Continuum removal versus PLSR method for clay and calcium carbonate content estimation from laboratory and airborne hyperspectral measurements. *Geoderma*, 148(2), 141–148. <https://doi.org/10.1016/j.geoderma.2008.09.016>
- Großkinsky, D. K., Svendsgaard, J., Christensen, S., & Roitsch, T. (2015). Plant phenomics and the need for physiological phenotyping across scales to narrow the genotype-to-phenotype knowledge gap. *Journal of Experimental Botany*, 66(18), 5429–5440. <https://doi.org/10.1093/jxb/erv345>
- Halperin, O., Gebremedhin, A., Wallach, R., & Moshelion, M. (2017). High-throughput physiological phenotyping and screening system for the characterization of plant-

environment interactions. *The Plant Journal*, 89(4), 839–850.

<https://doi.org/10.1111/tpj.13425>

Hansen, P. M., & Schjoerring, J. K. (2003). Reflectance measurement of canopy biomass and nitrogen status in wheat crops using normalized difference vegetation indices and partial least squares regression. *Remote Sensing of Environment*, 86(4), 542–553.

[https://doi.org/10.1016/S0034-4257\(03\)00131-7](https://doi.org/10.1016/S0034-4257(03)00131-7)

Iqbal, N., Hussain, S., Raza, M. A., Yang, C.-Q., Safdar, M. E., Brestic, M., Aziz, A., Hayyat, M. S., Asghar, M. A., Wang, X. C., Zhang, J., Yang, W., & Liu, J. (2019). Drought Tolerance of Soybean (*Glycine max* L. Merr.) by Improved Photosynthetic Characteristics and an Efficient Antioxidant Enzyme Activities Under a Split-Root System. *Frontiers in Physiology*, 10(2), 786. <https://doi.org/10.3389/fphys.2019.00786>

Kanning, M., Kühling, I., Trautz, D., & Jarmer, T. (2018). High-Resolution UAV-Based Hyperspectral Imagery for LAI and Chlorophyll Estimations from Wheat for Yield Prediction. *Remote Sensing*, 10(12), 2000. <https://doi.org/10.3390/rs10122000>

Lee, H., Kim, M., Jeong, D., Delwiche, S., Chao, K., & Cho, B.-K. (2014). Detection of Cracks on Tomatoes Using a Hyperspectral Near-Infrared Reflectance Imaging System. *Sensors*, 14(10), 18837–18850. <https://doi.org/10.3390/s141018837>

LI, R. hua, GUO, P. guo, Michael, B., Stefania, G., & Salvatore, C. (2006). Evaluation of Chlorophyll Content and Fluorescence Parameters as Indicators of Drought Tolerance in Barley. *Agricultural Sciences in China*, 5(10), 751–757. [https://doi.org/10.1016/S1671-2927\(06\)60120-X](https://doi.org/10.1016/S1671-2927(06)60120-X)

Lucieer, A., Malenovsky, Z., Veness, T., & Wallace, L. (2014). HyperUAS - Imaging

- spectroscopy from a multirotor unmanned aircraft system. *Journal of Field Robotics*, 31(4), 571–590. <https://doi.org/10.1002/rob.21508>
- Mahlein, A. K., Oerke, E. C., Steiner, U., & Dehne, H. W. (2012). Recent advances in sensing plant diseases for precision crop protection. In *European Journal of Plant Pathology* (Vol. 133, Issue 1, pp. 197–209). Springer. <https://doi.org/10.1007/s10658-011-9878-z>
- Makbul, S., Saruhan Güler, N., & Durmuş, N. (2011). Changes in anatomical and physiological parameters of soybean under drought stress. *Turk J Bot*, 35, 369–377. <https://doi.org/10.3906/bot-1002-7>
- Manavalan, L. P., Guttikonda, S. K., Phan Tran, L. S., & Nguyen, H. T. (2009). Physiological and molecular approaches to improve drought resistance in soybean. In *Plant and Cell Physiology* (Vol. 50, Issue 7, pp. 1260–1276). Oxford Academic. <https://doi.org/10.1093/pcp/pcp082>
- Marquetti, I., Link, J. V., Lemes, A. L. G., Scholz, M. B. dos S., Valderrama, P., & Bona, E. (2016). Partial least square with discriminant analysis and near infrared spectroscopy for evaluation of geographic and genotypic origin of arabica coffee. *Computers and Electronics in Agriculture*, 121, 313–319. <https://doi.org/10.1016/j.compag.2015.12.018>
- Meacham-Hensold, K., Montes, C. M., Wu, J., Guan, K., Fu, P., Ainsworth, E. A., Pederson, T., Moore, C. E., Brown, K. L., Raines, C., & Bernacchi, C. J. (2019). High-throughput field phenotyping using hyperspectral reflectance and partial least squares regression (PLSR) reveals genetic modifications to photosynthetic capacity. *Remote Sensing of Environment*, 231, 111176. <https://doi.org/10.1016/j.rse.2019.04.029>
- Meher, Shivakrishna, P., Ashok Reddy, K., & Manohar Rao, D. (2018). Effect of PEG-6000

imposed drought stress on RNA content, relative water content (RWC), and chlorophyll content in peanut leaves and roots. *Saudi Journal of Biological Sciences*, 25(2), 285–289. <https://doi.org/10.1016/j.sjbs.2017.04.008>

Monneveux, P., Jing, R., & Misra, S. C. (2012). Phenotyping for drought adaptation in wheat using physiological traits. *Frontiers in Physiology*, 3, 429. <https://doi.org/10.3389/fphys.2012.00429>

Moshelion, M., & Altman, A. (2015). Current challenges and future perspectives of plant and agricultural biotechnology. In *Trends in Biotechnology* (Vol. 33, Issue 6, pp. 337–342). Elsevier Ltd. <https://doi.org/10.1016/j.tibtech.2015.03.001>

Mutava, R. N., Prince, S. J. K., Syed, N. H., Song, L., Valliyodan, B., Chen, W., & Nguyen, H. T. (2015). Understanding abiotic stress tolerance mechanisms in soybean: A comparative evaluation of soybean response to drought and flooding stress. *Plant Physiology and Biochemistry*, 86, 109–120. <https://doi.org/10.1016/j.plaphy.2014.11.010>

Nakashima, K., Kanamori, N., Nagatoshi, Y., Fujita, Y., Takasaki, H., Urano, K., Mogami, J., Mizoi, J., Mertz-Henning, L. M., Neumaier, N., Farias, J. R. B., Fuganti-Pagliarini, R., Marin, S. R. R., Shinozaki, K., Yamaguchi-Shinozaki, K., & Nepomuceno, A. L. (2018). Application of biotechnology to generate drought-tolerant soybean plants in Brazil: Development of genetic engineering technology of crops with stress tolerance against degradation of global environment. In *Crop Production under Stressful Conditions: Application of Cutting-edge Science and Technology in Developing Countries* (pp. 111–130). Springer Singapore. https://doi.org/10.1007/978-981-10-7308-3_7

Nxele, X., Klein, A., & Ndimba, B. K. (2017). Drought and salinity stress alters ROS

- accumulation, water retention, and osmolyte content in sorghum plants. *South African Journal of Botany*, 108, 261–266. <https://doi.org/10.1016/j.sajb.2016.11.003>
- R Development Core Team. (2005). R: a language and environment for statistical computing. R Foundation for Statistical Computing. <https://doi.org/10.1016/j.dendro.2008.01.002>
- Rapaport, T., Hochberg, U., Shoshany, M., Karnieli, A., & Rachmilevitch, S. (2015). Combining leaf physiology, hyperspectral imaging and partial least squares-regression (PLS-R) for grapevine water status assessment. *ISPRS Journal of Photogrammetry and Remote Sensing*, 109, 88–97. <https://doi.org/10.1016/j.isprsjprs.2015.09.003>
- Reddy, A. R., Chaitanya, K. V., & Vivekanandan, M. (2004). Drought-induced responses of photosynthesis and antioxidant metabolism in higher plants. In *Journal of Plant Physiology* (Vol. 161, Issue 11, pp. 1189–1202). Elsevier GmbH. <https://doi.org/10.1016/j.jplph.2004.01.013>
- Reynolds, M., & Langridge, P. (2016). Physiological breeding. In *Current Opinion in Plant Biology* (Vol. 31, pp. 162–171). Elsevier Ltd. <https://doi.org/10.1016/j.pbi.2016.04.005>
- Richards, R. A. (2006). Physiological traits used in the breeding of new cultivars for water-scarce environments. *Agricultural Water Management*, 80(1-3 SPEC. ISS.), 197–211. <https://doi.org/10.1016/j.agwat.2005.07.013>
- Rocha, J. R. do A. S. de C., Machado, J. C., & Carneiro, P. C. S. (2018). Multitrait index based on factor analysis and ideotype-design: proposal and application on elephant grass breeding for bioenergy. *GCB Bioenergy*, 10(1), 52–60. <https://doi.org/10.1111/gcbb.12443>
- Roitsch, T., Cabrera-Bosquet, L., Fournier, A., Ghamkhar, K., Jiménez-Berni, J., Pinto, F., &

- Ober, E. S. (2019). Review: New sensors and data-driven approaches—A path to next generation phenomics. *Plant Science*, 282, 2–10.
<https://doi.org/10.1016/J.PLANTSCI.2019.01.011>
- Scholander, P. F., Bradstreet, E. D., Hemmingsen, E. A., & Hammel, H. T. (1965). Sap pressure in vascular plants: negative hydrostatic pressure can be measured in plants. *Science*, 148(3668), 339-346.
- Shunmugam, A., Kannan, U., Jiang, Y., Daba, K., & Gorim, L. (2018). Physiology Based Approaches for Breeding of Next-Generation Food Legumes. *Plants*, 7(3), 72.
<https://doi.org/10.3390/plants7030072>
- Uto, K., Seki, H., Saito, G., Kosugi, Y., & Komatsu, T. (2016). Development of a Low-Cost, Lightweight Hyperspectral Imaging System Based on a Polygon Mirror and Compact Spectrometers. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(2), 861–875. <https://doi.org/10.1109/JSTARS.2015.2472293>
- Wang, D., Vinson, R., Holmes, M., Seibel, G., Bechar, A., Nof, S., & Tao, Y. (2019). Early Detection of Tomato Spotted Wilt Virus by Hyperspectral Imaging and Outlier Removal Auxiliary Classifier Generative Adversarial Nets (OR-AC-GAN). *Scientific Reports*, 9(1), 1–14. <https://doi.org/10.1038/s41598-019-40066-y>
- Wellburn, A. R. (1994). The Spectral Determination of Chlorophylls a and b, as well as Total Carotenoids, Using Various Solvents with Spectrophotometers of Different Resolution. *Journal of Plant Physiology*, 144(3), 307–313. [https://doi.org/10.1016/S0176-1617\(11\)81192-2](https://doi.org/10.1016/S0176-1617(11)81192-2)
- Wise, B. M., Shaver, J. M., Gallagher, N. B., Rasmus, W. W., & Koch, B. R. S. (n.d.).

Continuum Parameter Number of LVs PRESS PLS_Toolbox 4.0 for use with
MATLAB™. Retrieved January 17, 2021, from www.eigenvector.com

Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-regression: A basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58(2), 109–130.
[https://doi.org/10.1016/S0169-7439\(01\)00155-1](https://doi.org/10.1016/S0169-7439(01)00155-1)

Wu, C., Niu, Z., Tang, Q., & Huang, W. (2008). Estimating chlorophyll content from
hyperspectral vegetation indices: Modeling and validation. *Agricultural and Forest
Meteorology*, 148(8–9), 1230–1241. <https://doi.org/10.1016/j.agrformet.2008.03.005>

Xu, H. R., Yu, P., Fu, X. P., & Ying, Y. Bin. (2009). On-site variety discrimination of tomato
plant using visible-near infrared reflectance spectroscopy. *Journal of Zhejiang University:
Science B*, 10(2), 126–132. <https://doi.org/10.1631/jzus.B0820200>

Yendrek, C. R., Tomaz, T., Montes, C. M., Cao, Y., Morse, A. M., Brown, P. J., McIntyre, L. M.,
Leakey, A. D. B., & Ainsworth, E. A. (2017). High-throughput phenotyping of maize leaf
physiological and biochemical traits using hyperspectral reflectance. *Plant Physiology*,
173(1), 614–626. <https://doi.org/10.1104/pp.16.01447>

Yoder, B. J., & Pettigrew-Crosby, R. E. (1995). Predicting nitrogen and chlorophyll content and
concentrations from reflectance spectra (400–2500 nm) at leaf and canopy scales. *Remote
Sensing of Environment*, 53(3), 199–211. [https://doi.org/10.1016/0034-4257\(95\)00135-N](https://doi.org/10.1016/0034-4257(95)00135-N)

Yu, K., Anderegg, J., Mikaberidze, A., Karisto, P., Mascher, F., McDonald, B. A., Walter, A., &
Hund, A. (2018). Hyperspectral Canopy Sensing of Wheat Septoria Tritici Blotch Disease.
Frontiers in Plant Science, 9, 1195. <https://doi.org/10.3389/fpls.2018.01195>

Zhang, J., Han, W., Huang, L., Zhang, Z., Ma, Y., & Hu, Y. (2016). Leaf Chlorophyll Content Estimation of Winter Wheat Based on Visible and Near-Infrared Sensors. *Sensors*, 16(4), 437. <https://doi.org/10.3390/s16040437>

Chapter II - POTENTIAL OF LEAF HYPERSPECTRAL PROXIMAL SENSING IN ESTIMATE YIELD-RELATED TRAITS IN SOYBEAN

ABSTRACT

Increasing yield is of the most priority in a soybean breeding program. Understanding yield related traits would help breeders in having more accurate decision making in several breeding strategies. Despite manually assess yield components traits is not feasible in a large scale program, the usage of tools based on remote sensing can provide several information about yield related traits in a faster and noninvasive way. In this study, we investigate the capacity to predict multiple yield related traits through hyperspectral data. Field trials was conducted to evaluate twenty soybeans cultivars regarding number of seedless pods (SP), one-seed pods (P1), two-seed pods (P2), three-seed pods (P3), four-seed pods (P4), one-hundred seed weight (100SW), total seed weight (TW), total number of seeds (TS), total number of pods (TP), and the difference between total seed potential considering if all cavity of the pods had been filled and the actual total number of seeds (GAP). FAI-BLUP index was used as a standard methodology for multi-trait selection. Leaf hyperspectral data was collected using a portable spectrometer (350-2500 nm), and the partial least square regression (PLSR) models were built in order to associate spectral wavelengths to measured traits. No genotypic variance were found for TW, TS, and TP, thus no PLSR models were built for these traits. Results show significant predictive capacity of seed-per-pod traits (SP, P1, P2, P3, and P4), with R^2 cross validation values ranging from 0.54 to 0.81, associated with relative low root mean square error (RMSE). Consistent results were also found for 100SW and GAP, which show R^2 values higher than 0.91 and 0.82, respectively. Association between FAI-BLUP indexes values and spectral information show significant results (R^2 cross validation = 0.60 and RMSE = 0.029), indicating that is possible to use hyperspectral data to predict multiple yield related traits in soybean.

Key-words: yield prediction; PLSR; leaf reflectance

1. INTRODUCTION

Soybean (*Glycine max* (L.) Merr.) demand in the food industry will still increasing in the coming years since this grain is the main source of both vegetable oil and animal protein feed, and there is precognition in global population growth (Balboa et al., 2018; Ray et al., 2013). On the other hand, stagnation in terms of arable areas and the concern about climate situation has been worrying farmers around the world (Clark & Tilman, 2017). In this context, plant breeding can be considered the main strategy to overcome those issues in generating new genotypes that attend environmental, economic, and social demands.

With the goal of continuously improving soybean varieties to achieve better production every year, breeder face several challenges regarding environmental conditions and crop management, such as location adaptation, sowing time, biotic and abiotic stresses (Bruce et al., 2019; Sinclair, 2011). Understanding the mechanism from improving yield would contribute to provide detailed information about developing new cultivars. Thus, association between yield to its components is crucial as it can turn out for selection criteria in a breeding program pipeline and can help to identify different genotype patterns among environments (Herbert & Litchfield, 1982; Pandey & Torrie, 1973). On the other hand, phenotyping yield components on a large scale demands time and intense labor, and therefore makes it unpractical.

Despite the advance in genomic association and genomic prediction for yield and yield-related traits, the lack of accurately high-throughput phenotyping approaches remains a bottleneck to increasing the success of such methodology, since accurately phenotyping is key information for establishing good statistical models (Araus et al., 2018; Desta & Ortiz, 2014). Additionally, fast and precise phenotyping can contribute to increasing early decision-making while reducing input resource and environmental issues (Maimaitijiang et al., 2020; Panda et al., 2010). In this sense, non-destructive yield-related traits prediction in plot-level would allow the identification of promising genotypes in large scale breeding programs (Zhang et al., 2019).

Developing phenotyping platforms based in remote sensing approach have allowing measure crop traits in the field on a large scale, such as chlorophyll content, plant height, leaf area index, yield, etc. (Friedli et al., 2016; Harbinson et al., 2012; Shakoore et al., 2017). Most of these researches focus on few band indices based on multispectral or even RGB information, while the

influence of environmental effects could divert for a more robust interpretation of the results (Araus & Cairns, 2014; Tattaris et al., 2016). Full spectrum information can improve the effectiveness of remote sensing once it captures crop reflectance in narrow bands and consequently provide better information in the phenotyping process, especially for complex traits such as grain yield (Montesinos-López et al., 2017; Weber et al., 2012; Zarco-Tejada et al., 2016).

Selecting more than one yield-related trait simultaneously would be a key to achieve high-yield genotypes since grain yield is affected by several traits. Thus, using selection indexes that account for those traits are very useful, especially when the traits are genetically correlated (Rocha et al., 2018). Therefore, we aim to use hyperspectral remote sensing for predicting yield-related traits and check the usefulness of the predicted models to select the cultivars based on the yield-related traits simultaneously.

2. MATERIAL AND METHODS

2.1. Plant material and yield-related traits evaluation

Twenty soybean cultivars (Table 1) were sowing in the 2018-2019 season at the horta nova experimental site (UEPE) of the Federal University of Viçosa (20° 45' S and 42° 49' W, 680 m). These cultivars were evaluated in a field experiment in a randomized complete block design with six replications. Each plot consisted of one row 1 m long with an inter-row spacing of 0.50 m, thinned to a uniform stand of 10 plants m⁻¹. The plants were carried out regarding all agronomic management until they reach full maturity, when all the plants of each plot were manually harvested.

From all the plants of each plot, we assessed the following yield-related traits: number of seedless pods (SP), one-seed pods (P1), two-seed pods (P2), three-seed pods (P3), four-seed pods (P4), one-hundred seed weight (100SW), total seed weight (TW), total number of seeds (TS), total number of pods (TP), and the difference between total seed potential considering if all cavity of the pods had been filled and the actual total number of seeds (GAP). For each plot, the mean of the individually values from all the plants was computed for the traits values, which were used in further analysis.

Table 1- List of the cultivars evaluated

Cultivar	Flower color	Pubescence color	Growth habit	Maturity group
BMX POTÊNCIA	White	Grey	Indeterminate	6.6
BRS 185	Purple	Grey	Determinate	6.8
BRS 232	Purple	Grey	Determinate	6.9
BRS 246	White	Brown	Determinate	7.2
BRS 268	White	Grey	Determinate	6.9
CD 208	White	Grey	Determinate	6.9
CD 213	Purple	Grey	Determinate	6.1
CD 216	White	Light brown	Indeterminate	5.5
CD 234 RR	Purple	Grey	Determinate	8.0
CD 235 RR	White	Brown	Semi-Determinate	6.4
CD 244 RR	White	Grey	Determinate	8.0
DMARIO 70i	White	Grey	Indeterminate	6.2
EMBRAPA 48	White	Grey	Determinate	6.8
FT GUAÍRA	NI*	NI	NI	NI
FUNDACEP 57 RR	White	Brown	Determinate	6.5
CONQUISTA	Purple	Brown	Determinate	8.1
NK 7059 RR	White	Grey	Indeterminate	5.9
R7	NI	NI	NI	SI
RA 518	Purple	Light brown	SemiDerteminate	6.4
RA 628	White	Grey	Indeterminate	6.4

*NI: No information

2.2.Cultivar selection

Best Linear Unbiased Prediction (BLUP) from the cultivar genotypic mean was estimated through the following mixed linear model:

$$y = \mathbf{X}\tau + \mathbf{Z}_g\mu_g + e$$

which: y is the vector of phenotypic mean; τ is the vector of blocks added by general mean, assumed as fixed; μ_g is the vector of random effects of cultivars, with $\mu_g \sim N(0, \sigma_g^2)$; and e is the random effect of residuals, with $e \sim N(0, \sigma)$; $\mathbf{X}$ and $\mathbf{Z}_g$ are indexes matrix associated with τ and μ_g effects, respectively. Variance components were estimated through the restricted maximum likelihood (REML). The effect of the genotypic variance component of cultivars was tested through

the likelihood ratio test. We used the *sommer* package present in R software (R Development Core Team, 2005) to do the REML-BLUP procedure.

The FAI-BLUP selection index was used to select the best cultivars regarding all traits simultaneously. This index uses the factor analysis and genotype-ideotype distance as the selection approach, allowing the correlation among traits being considered (Rocha et al., 2018). Only the traits with significant cultivar effect were used in the FAI-BLUP index and the ideotype design used for the desired cultivar was: minimum values from SP and GAP, and maximum values from P1, P2, P3, P4, 100SW, TW, TP, and TS. This analysis was processed in R software.

2.3. Leaf hyperspectral data and Partial Least Square Regression (PLSR)

At the R₅ phenological stage, which is characterized by the seed feeling period (Fehr & Caviness, 1977), spectral signatures were collected using a spectroradiometer ASD FieldSpec® 4 Hi-Res (ASD Inc., Boulder, CO, USA) with a spectral range over 350-2500 nm and a resampled resolution of 1.4 nm at 350-1000 nm and 1.1 nm at 1001-2500 nm. The spectral signature was measured in a completed opened leaf located in the middle of a random plant sampled within each plot. Leaf hyperspectral reflectance was measured using a contact prob connected to spectroradiometer, with 10 measurements taken on each leaflet, and the mean value of these measurements was record. The spectral signatures were corrected using the splicing procedure to avoid spectra noises caused by different detectors arrays in the spectroradiometer. This procedure was repeated within an interval of 7 days, and the mean hyperspectral of each plot considering the two evaluations dates was used for the analysis.

We used the PLSR to associate the cultivars hyperspectral signature with the yield-related traits and the FAI-BLUP score indexes, and to build models to predict these traits. Autoscale and first derivative of second order were used as preprocessing methods on the spectral data. To selecting the optimal number of latent variable (LV), the lowest value of cross-validation root mean square error (RMSECV) based on venetian blind cross-validation was determined. The lowest value of Root Mean Square Error Cross-Validation (RMSECV), and the highest values of calibration and cross-validation R² were used to fit the PLRS model.

After selecting the calibrated PLSR model, we applied the Variable Importance in Projection (VIP) to identify the contribution of each wavelength in fitting the model. The selected

wavelengths were those above the VIP threshold score set to one (Wold et al., 2001). A new model was built with the selected wavelengths to maintain the same selection criteria as the previous step. Selecting wavelengths through VIP and building a new PLSR model was repeated until we get the best-fitted model with the highest possible calibration and cross-validation R^2 and the lowest RMSECV. The PLSR analysis was processed in the SOLO (Eingenvector Inc.) software.

A scheme of data collection, PLRS analysis and model selection procedure is showed at Figure 1.

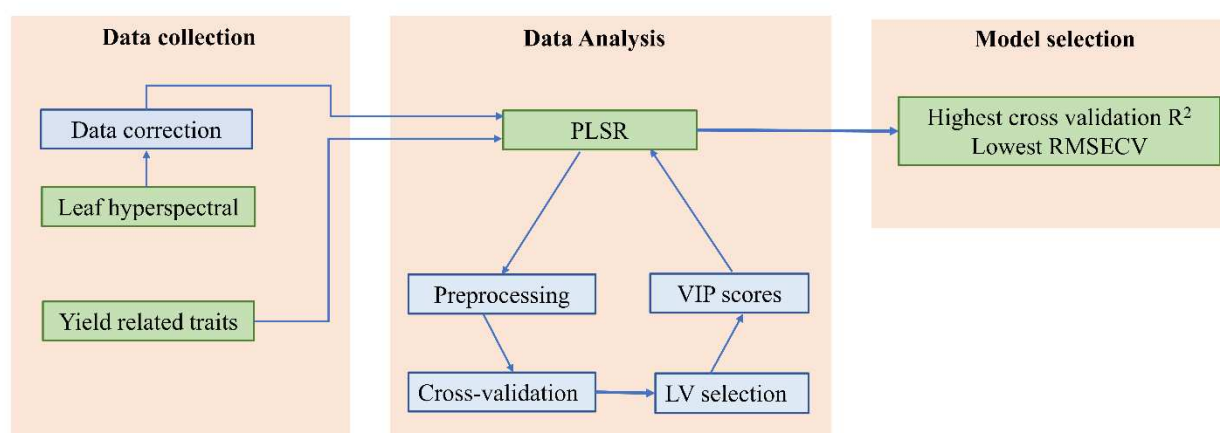


Fig 1. Scheme showing the workflow of data collection, PLSR model building and selection procedure.

3. RESULTS AND DISCUSSION

There was genotypic variance among the cultivars for the traits SP, P1, P2, P3, P4, 100SW, and GAP at 1% of probability. The accuracy values for estimating the genotypic variance components for those traits were high, especially for P2 (0.92) and 100SW (0.91). Similarly, the highest value of heritability for P2 (0.84) and 100SW (0.80) indicated that most of the variability founded for these traits was due to genetic source (Table 2). Faisal et al. (2006) also recorded high heritability for 100SW, showing that this trait is very important for yield-related phenotyping. Seeds-per-pod traits (SP, P1, P2, P3, and P4) are directed associated with grain yield and can be individually studied as possible selection criteria for grain yield improvement (Korte et al., 1983; Xianzhi et al., 2010). GAP expresses how close the actual yield is to its yield potential, given an

idea on how efficiently each cultivar can filling pods. In other words, GAP can be associated with stresses situations since it measures the difference between the actual total seeds and the expected total seeds, considering if all pods (P1, P2, P3, and P4) had reached the maximum number of seed possible.

Table 2- Genotypic variance component (V.C.), prediction accuracy of the model and heritability for the traits seedless pod (SP), one seed pod (P1), two seeds pod (P2), three seeds pod (P3), four seeds pod (P4), one-hundred seed mass (100SM), total mass (TM), total seeds (TS), total pods (TP) and the difference between actual and potential seed yield (GAP)

Trait	Genotypic V.C. [†]	Prediction Accuracy	Heritability	Trait	Genotypic V.C. [†]	Prediction Accuracy	Heritability
SP	5.974**	0.852	0.64 ±0.11	100SW	2.808**	0.911	0.80 ±0.06
P1	2.252**	0.885	0.68 ±0.09	TW	3.715 ^{ns}	0.557	0.24 ±0.19
P2	156.806**	0.927	0.84 ±0.05	TS	0.000 ^{ns}	0.000	0.00 ±0.22
P3	69.140**	0.881	0.67 ±0.09	TP	0.000 ^{ns}	0.000	0.00 ±0.21
P4	0.307**	0.775	0.41 ±0.13	GAP	13.890**	0.823	0.52 0.12

[†]: ** significant at 1% probability ; ^{ns} non-significant

There were no significant genotypic effects for TW, TP, and TS. These results are associated with the high environmental influence on the genotypic mean for those trait, which can be evidenced through the low accuracies values (Table 2). Since data from individual plants was measurement, it is possible that most of the variance among plants within plots was due to within-individual plasticity as a change in a particular phenotype concerning variation in the environment (Westneat et al., 2015). Soybean is known for its high biomass and pod production plasticity (Carpenter & Board, 1997; Kumagai et al., 2015), thus identifying genotypic variance for TW, TP and TS could be difficult when evaluating individual plants.

Correlation information among the traits are presented in Figure 2. Strong and negative correlation were identified between P2 and P3, meaning that cultivars likely have preference in developing two or three seeds per pod. Despite less strongly correlated, similar result was also found when comparing P2 and P4. Positive correlation values were observed when comparing SP to and P2, indicating that plants investing in less seeds per pod are more common to have pod without any seeds. GAP also has a small but significative correlation with SP. This would be explain because all values of SP were used to computed the derived trait GAP. No correlation was

observed between 100SW and the other traits, which suggest that the seed size varies regardless the number of seeds per pods and GAP.

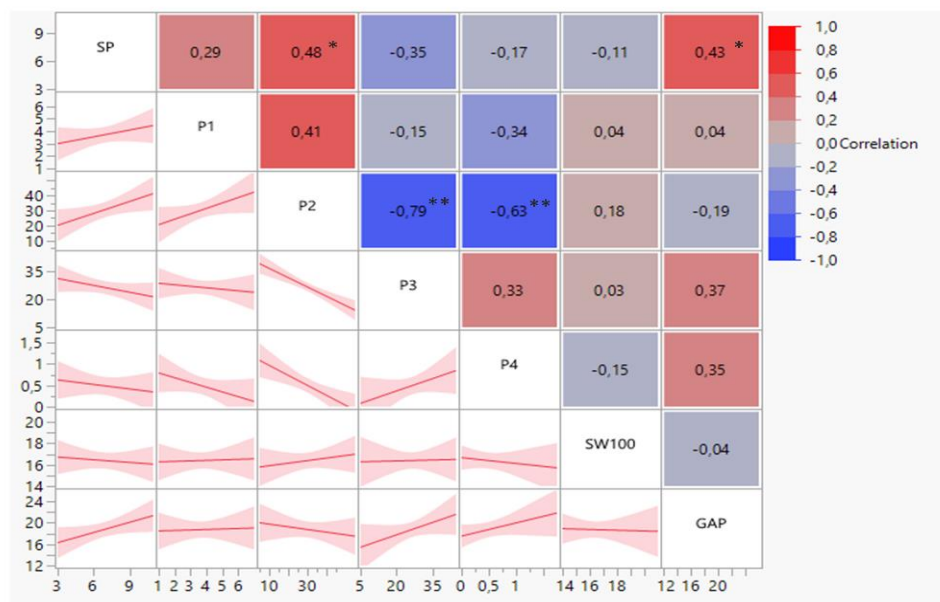


Fig 2. - Linear correlation coefficient between seedless pod (SP), one-seed pod (P1), two-seeds pod (P2), three-seeds pod (P3), four-seeds pod (P4), one-hundred seed mass (100SM), and difference between actual and potential seed yield (GAP), at upper triangle matrix. Linear regression line and 95% confidence interval (red shaded) are shown at lower triangle matrix. * significance at $p < 0.05$. ** significance at $p < 0.01$.

The five best cultivars selected by the FAI-BLUP analysis were NK 7059, DMARIO 70i, CONQUISTA, BRX POTÊNCIA, and BRS 268 (Table 3). The selection is based on an index, which is estimated based on factor analysis, and the factorial scores of an ideotype are designed for the desirable factors, so the genotypes are ranked according to a spatial probability of genotype-ideotype distance. Besides, the FAI-BLUP index is free from multicollinearity, taking into account the genetic correlation between traits within each factor and gives biological meaning to the ideotypes (Rocha et al., 2018). Because there was no genetic effect for TW, TP, and TS, these traits were not included in this analysis.

Three factors were selected through the factor analysis used in the FAI-BLUP procedure (Table 3). Traits clustered in the first factor can be related to yield fallback (SP, P1, and GAP). GAP value indicates that plants produce pods and did not have enough energy to fill them out, SP

is strongly related to abiotic stress (Giordani et al., 2019), and P1 is associated with decreasing total seed number (Carter & Tegeder, 2016). Traits in the second factor (P2, P3, and P4) are more associated to yield increase than the first, since increasing number of seeds per pod would also increase the final seed yield (Pandey and Torrie, 1973). 100SW was allocated in one separated factor, which is emphasized by the low correlation of this trait with the other traits. Therefore, the ideotype was build considering decreasing the values on the first factor and increasing the ones on the second and third factors.

Table 3 - FAI-BLUP index estimated for 20 genotypes ranked from the highest to lowest value, traits associated with each of the selected factor and the genetic gain in selecting the best five cultivars according to FAI-BLUP index

FAI-BLUP index				Factor	Trait ¹	Genetic gain (%)
Cultivar	Index	Cultivar	Index			
NK 7059	0.193	CD 234	0.110	1	SP	-25.11
DMARIO 70i	0.189	BRS 246	0.108		P1	-17.03
CONQUISTA	0.179	CD 244	0.103		GAP	-8.15
BRX POTÊNCIA	0.174	CD 216	0.101			
BRS 268	0.149	BRS 185	0.095	2	P2	-14.84
FT GUAIRA	0.131	CD 208	0.094		P3	13.16
BRS 232	0.118	R7	0.090		P4	-8.82
EMBRAPA 48	0.115	CD 213	0.088	3		
RA 268	0.114	CD 235	0.087		100SW	9.04
FUNDACEP 57	0.113	RA 518	0.069			

¹ SP= seedless pod, P1= one seed pod, P2= two seeds pod, P3= three seeds pod, P4= four seeds pod, 100SM= one-hundred seed mass, (GAP) difference between actual and potential seed yield (GAP)

The five selected cultivars have the potential to improve more than one trait at the same time. According to the expected genetic gain, the cultivars should increase 100SW and P3, while decreasing SP, P1, and GAP (Table 3). Additionally, no desirable gains was computed for P2 and P4. Selecting cultivars with more seeds per pods are beneficial to increase yield, thus the FAI-BLUP index tried to capture the cultivars with as much as possible of seeds per pods. The selected cultivars tends to increase the number of P3, and since P2 and P3 are high negatively correlated, the number of P2 would decrease. On the other hand, the number of P4 would decrease due to low frequency of this trait among the cultivars and therefore it did not have enough weight on the

selection procedure. Moreover, the results suggest that the set of cultivars used in this work is not enough to improve the P4 trait. Descriptive analysis of the evaluated traits can be found at Table S1.

The prediction of yield-related traits through the cultivars spectral signature are shown in Figure 3. Very good R^2 calibration values for pod-related traits were found, and similar results were found for R^2 cross-validation prediction values, with exception for P1, which got a R^2 cross validation values of 54% (Figure 3a, 3b, 3c, and 3d). They all show R^2 calibration values higher than 90%. For R^2 cross validation, we found values around 80% for SP, P2, and P3, and almost 69% for P4. Besides the high values of R^2 for P4, the result associated with this trait gives a false sense of prediction due to the presence of a great amount of zero values on the data set. Four-seed pod is found in very low frequency in soybean cultivars, so it is natural to have zero values associated with some of the evaluated cultivars. Besides been related to seed yield (Pandey and Torrie 1973), increasing seed number per pod may improve yield stability across several environments (Jin et al., 2010). Assessing less environment-influenced traits would be required to achieve a better understanding of top-high yield cultivars (Aditya et al., 2011; Li et al., 2019). Consequently, improving these traits can give a good perspective on select cultivars genetically superior in terms of yield.

Decent R^2 values from train and cross-validation data were also found for 100SW, couple with a relative low RMSECV (Figure 3g). Such result highlights that 100SW can be assessed remotely and before harvest, given a certain prediction of an important trait in terms of yield performance. Seed weight and seed size have been focused in soybean domestication due to its influence on germination ability, seedling vigor, nutrient resource, and stress tolerance (Coomes & Grubb, 2003; C. Yu et al., 2017). Even though candidate genes have been identified by Karikari et al. (2019) and Li et al. (2019), no spectral information has been reported for this trait in the phenotyping framework. Consequently, this finding suggests that high-throughput phenotyping can be applying for early estimation of seed-weight in soybean genotypes.

High training and cross-validation R^2 values were also found for GAP (Figure 3f). Once this trait can be associated with a stress-related trait, the result shows the potential to identifying cultivars affected in some sort of biotic or abiotic stresses. Evaluating of those cultivars under drought condition, Paula (2020, not yet published) was able to find difference on phenotypic for

chlorophyll a, chlorophyll b, and carotenoids contents, specific foliar area, and canopy temperature, as well as prediction those traits using spectral signatures. Works showing the usage of wavelengths and vegetation indexes in prediction soybean physiological traits in stress conditions can be found in the literature (Carvalho et al., 2015; Kovar et al., 2019; Wijewardana et al., 2019), and can also be used to associate with complex traits such as yield.

A relative good result was found for cross-validated FAI-BLUP Ideotype indexes with spectral data by PLSR analysis, which gives a R^2 cross validation of 60% and RMSECV value of 0.029 (Figure 1h). Higher estimated Ideotype indexes also gives higher predicted Ideotype indexes as well, which is emphasized by the five best cultivars selected by the FAI-BLUP analyses, which is highlighted in red in Figure 1h. The ranked cultivars by the ideotype indexes represent an approximation of the best possible genotype regarding all the evaluated yield-related traits simultaneously. Therefore, predicting ideotype indexes by spectral information might be useful for early identification of soybean genotypes that gather several yield component traits, and consequently gives an idea about yield prediction.

The selected wavelength via VIP methodology for the PLSR model of FAI-BLUP index was distributed in the spectrum fraction of 65% in the visible region, 19% in near-infrared, and 16% in the short-wave infrared region (Figure 4). This result indicates that wavelengths from the visible region of the spectra was very important in building a prediction model of a multi-trait selection index, specially at blue region of the spectra. Reflectance of the visible-region is associated with leaf pigments absorption, which is directly related to the photosynthesis rate and therefore on the number of seed per square meter (Jiang and Egli, 1995; Zhang et al., 2016). This result corroborate with recent works that have shown the feasibility of using RGB sensors to be used as a proxy to predict soybean yield (Randelović et al., 2020; Yuan et al., 2019), and therefore are bringing valuable insights to make such approach useful in the breeding pipeline.

Relevant wavelengths were also identified in the red edge region, which is a transition region between the red and near infrared region. This region has proven to be sensitive to leaf pigment content and nitrogen content, and therefore would be associate to plant stresses including water stress (Ballester et al., 2019; Ju et al., 2010). Also, the results indicate that the wavelengths in the red edge region are correlated with those at the visible region and are also important to select the cultivars in terms of yield-related traits.

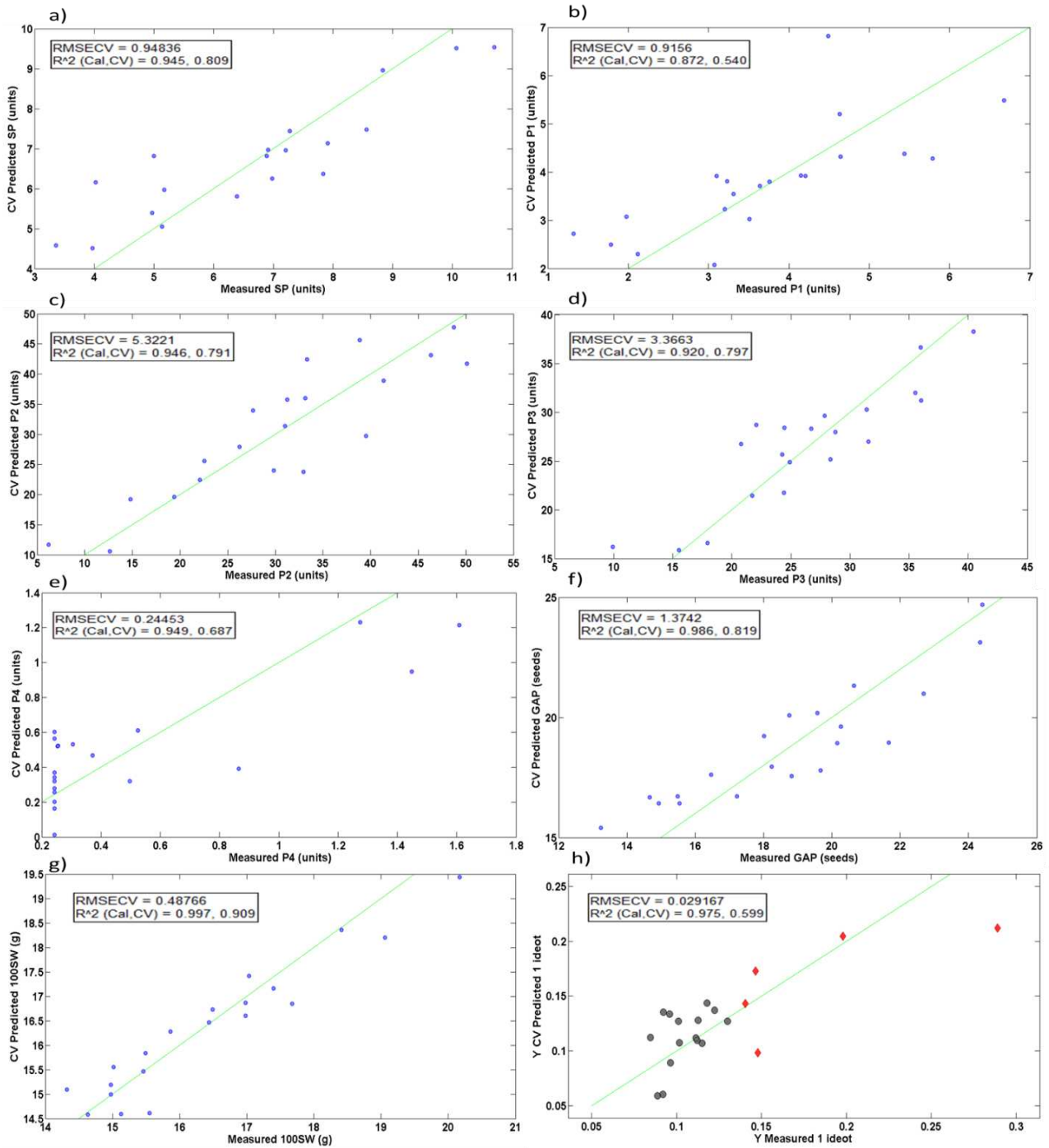


Fig 3. Cross-validated (cv) predicted and measured values of a) seedless pod (SP), b) one-seed pod (P1), c) two-seeds pod (P2), d) three-seeds pod (P3), e) four-seeds pod (P4), f) difference between actual and potential yield (GAP), g) one-hundred seed mass (100SM), and h) FAI-BLUP Ideotype indexes

The wavelengths selected at near infrared region also show reliable contribution on the built model. Spectral information of these regions are basically associated to internal leaf structure, and water content (Carley et al., 2018; Benabdelouahab et al., 2016; Hunt et al., 2016). Some indexes based on the near-infrared information, such as NDVI, can be used as a proxy of biotic and abiotic stress such as disease and water status (Zhou et al., 2020), and, consequently, can help identifying most adapted and improved genotypes.

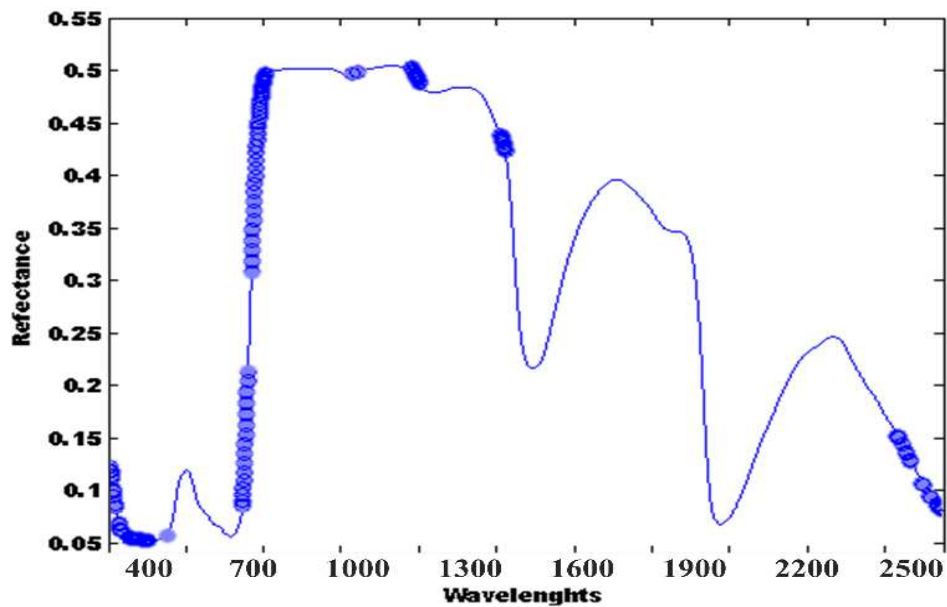


Fig 4. Selected wavelengths for PLSR model of the FAI-BLUP index vales (blue circles) through the leaf hyperspectral reflectance (blue line).

Although soybean-breeding companies efforts are to select high-yielded and adapted genotypes, in most cases, the yield is counted only in terms of grain per hectare (Balboa et al., 2018; Specht et al., 2015). Focusing only on yield is useful to define many breeding strategies and relatively less labor to be evaluated. However, important information may not be considered for a better selection procedure since yield is highly influenced by environment changes (Matiu et al., 2017; Zanon et al., 2016). Such information can be assessed taking into account yield related traits and also associate the selection procedure to high yield cultivars (Nouri et al., 2011; Hu et al., 2020).

Many studies have indicated number of seeds per pod as one of the most important trait associates to yield components (Machikow and Laosuwan, 2011; Jin et al., 2010; Mahbuh et al.,

2015; Bianchi et al., 2020). Information about yield components would be very useful to decision making at the selection stage. However, assessing such information in a breeding pipeline where thousands of varieties are tested every year seems almost unfeasible. With the advance of new technologies and the possibility to implement multispectral and even hyperspectral sensors at phenotyping platforms, new frontiers can be open in terms of breeding since difficult to measure traits such as yield components could be estimated and help on better understanding and selection of varieties.

Supplementary material

Table S1. Descriptive statistics of the traits seedless pod (SP), one seed pod (P1), two seeds pod (P2), three seeds pod (P3), four seeds pod (P4), one-hundred seed mass (100SM), total mass (TM), total pods (TP), total seeds (TS) and the difference between real and potential seed yield (GAP)

Trait	Mean	Std Dev	Sum	Minimum	Maximum
SP	6.821	1.757	136.4	3.18	10.09
P1	3.438	1.049	68.76	1.525	5.76
P2	29.9	7.715	598.1	13.42	42.81
P3	26.54	5.981	530.8	13.76	40.11
P4	0.528	0.575	10.56	0	1.825
100SW	16.42	1.354	328.3	14.31	19.05
TM	21.32	2.419	426.4	16.47	25.16
TP	62.46	3.712	1249.2	54.79	67.86
TS	132.3	7.588	2646.2	114.4	146.1
GAP	18.72	2.698	374.4	13.88	23.73

REFERENCES

- Aditya, J. P., Bhartiya, P., & Bhartiya, A. (2011). Genetic variability, heritability and character association for yield and component characters in soybean (*G. max* (L.) Merrill). *Journal of Central European Agriculture*, 12(1), 27-34.
- Araus, J. L., & Cairns, J. E. (2014). Field high-throughput phenotyping: The new crop breeding frontier. In *Trends in Plant Science* (Vol. 19, Issue 1, pp. 52–61). Elsevier Current Trends. <https://doi.org/10.1016/j.tplants.2013.09.008>
- Araus, J. L., Kefauver, S. C., Zaman-Allah, M., Olsen, M. S., & Cairns, J. E. (2018). Translating High-Throughput Phenotyping into Genetic Gain. In *Trends in Plant Science* (Vol. 23, Issue 5, pp. 451–466). Elsevier Ltd. <https://doi.org/10.1016/j.tplants.2018.02.001>
- Balboa, G. R., Sadras, V. O., & Ciampitti, I. A. (2018). Shifts in Soybean Yield, Nutrient Uptake, and Nutrient Stoichiometry: A Historical Synthesis-Analysis. *Crop Science*, 58(1), 43–54. <https://doi.org/10.2135/cropsci2017.06.0349>
- Ballester, C., Brinkhoff, J., Quayle, W. C., & Hornbuckle, J. (2019). Monitoring the effects of water stress in cotton using the green red vegetation index and red edge ratio. *Remote Sensing*, 11(7), 873. <https://doi.org/10.3390/RS11070873>
- Benabdelouahab, T., Balaghi, R., Hadria, R., Lionboui, H., & Tychon, B. (2016). Assessment of vegetation water content in wheat using near and shortwave infrared SPOT-5 Data in an irrigated area. *Revue Des Sciences de l'Eau*, 29(2), 97–107. <https://doi.org/10.7202/1036542ar>
- Bianchi, J. S., Quijano, A., Gosparini, C. O., & Morandi, E. N. (2020). Changes in leaflet shape and seeds per pod modify crop growth parameters, canopy light environment, and yield components in soybean. *The Crop Journal*, 8(2), 351-364.
- Bruce, R. W., Grainger, C. M., Ficht, A., Eskandari, M., & Rajcan, I. (2019). Trends in Soybean Trait Improvement over Generations of Selective Breeding. *Crop Science*, 59(5), 1870–1879. <https://doi.org/10.2135/cropsci2018.11.0664>

- Carley, D. S., Jordan, D. L., Dharmasri, C. L., Shew, B. B., Sutton, T. B., & Brandenburg, R. L. (2018). Examples of Differences in Red Edge Reflectance and Normalized Difference Vegetative Index Caused by Stresses in Peanut. *Crop, Forage & Turfgrass Management*, 4(1), 180042. <https://doi.org/10.2134/cftm2018.06.0042>
- Carpenter, A. C., & Board, J. E. (1997). Branch yield components controlling soybean yield stability across plant populations. *Crop Science*, 37(3), 885–891. <https://doi.org/10.2135/cropsci1997.0011183X003700030031x>
- Carter, A. M., & Tegeder, M. (2016). Increasing Nitrogen Fixation and Seed Development in Soybean Requires Complex Adjustments of Nodule Nitrogen Metabolism and Partitioning Processes. *Current Biology*, 26(15), 2044–2051. <https://doi.org/10.1016/j.cub.2016.06.003>
- Clark, M., & Tilman, D. (2017). Comparative analysis of environmental impacts of agricultural production systems, agricultural input efficiency, and food choice. *Environmental Research Letters*. <https://doi.org/10.1088/1748-9326/aa6cd5>
- Coomes, D. A., & Grubb, P. J. (2003). Colonization, tolerance, competition and seed-size variation within functional groups. In *Trends in Ecology and Evolution* (Vol. 18, Issue 6, pp. 283–291). Elsevier Ltd. [https://doi.org/10.1016/S0169-5347\(03\)00072-7](https://doi.org/10.1016/S0169-5347(03)00072-7)
- Desta, Z. A., & Ortiz, R. (2014). Genomic selection: Genome-wide prediction in plant improvement. In *Trends in Plant Science* (Vol. 19, Issue 9, pp. 592–601). Elsevier Ltd. <https://doi.org/10.1016/j.tplants.2014.05.006>
- Faisal, M., Malik, A., & Ghafoor, A. (2006). Genetic Variability of the Main Yield Related Characters in Soybean Proteomic evaluation of *Nigella sativa* L. for seed protein peptides and multivariate analysis View project. <http://www.fspublishers.org>
- Fehr, W., & Caviness, C. (1977). Stages of soybean development. Special Report. <https://lib.dr.iastate.edu/specialreports/87>
- Friedli, M., Kirchgessner, N., Grieder, C., Liebisch, F., Mannale, M., & Walter, A. (2016). Terrestrial 3D laser scanning to track the increase in canopy height of both monocot and dicot

crop species under field conditions. *Plant Methods*, 12(1), 9. <https://doi.org/10.1186/s13007-016-0109-7>

Giordani, W., Goncalves, L. S. A., Moraes, L. A. C., Ferreira, L. C., Neumaier, N., Farias, J. R. B., Nepomuceno, A. L., Oliveira, M. C. N., & Mertz-Henning, L. M. (2019). Identification of agronomical and morphological traits contributing to drought stress tolerance in soybean (Humanities & Social Sciences Collection) - Informit. *Australian Journal of Crop Science*, 13(1), 35. <https://search.informit.com.au/documentSummary;dn=337951904526560;res=IELHSS>

Harbinson, J., Prinzenberg, A. E., Kruijer, W., & Aarts, M. G. M. (2012). High throughput screening with chlorophyll fluorescence imaging and its use in crop improvement. In *Current Opinion in Biotechnology* (Vol. 23, Issue 2, pp. 221–226). Elsevier Current Trends. <https://doi.org/10.1016/j.copbio.2011.10.006>

Herbert, S. J., & Litchfield, G. V. (1982). Partitioning Soybean Seed Yield Components 1. *Crop Science*, 22(5), 1074–1079. <https://doi.org/10.2135/cropsci1982.0011183x002200050044x>

Hu, D., Zhang, H., Du, Q., Hu, Z., Yang, Z., Li, X., ... & Kan, G. (2020). Genetic dissection of yield-related traits via genome-wide association analysis across multiple environments in wild soybean (*Glycine soja* Sieb. and Zucc.). *Planta*, 251(2), 39.

Hunt, E. R., Daughtry, C. S. T., & Li, L. (2016). Feasibility of estimating leaf water content using spectral indices from WorldView-3's near-infrared and shortwave infrared bands. *International Journal of Remote Sensing*, 37(2), 388–402. <https://doi.org/10.1080/01431161.2015.1128575>

Jiang, H., & Egli, D. B. (1995). Soybean seed number and crop growth rate during flowering. *Agronomy Journal*, 87(2), 264-267.

Jin, J., Liu, X., Wang, G., Mi, L., Shen, Z., Chen, X., & Herbert, S. J. (2010). Agronomic and physiological contributions to the yield improvement of soybean cultivars released from 1950 to 2006 in Northeast China. *Field Crops Research*, 115(1), 116–123. <https://doi.org/10.1016/j.fcr.2009.10.016>

Carvalho, J. F. C, Crusiol, L. G. T., Perini, L. J., Sibaldelli, R. N. R, Ferreira, L. C., Marcelino-Guimarães, F. C., Neumaier, N. & Farias, J. R. B. (2015). Phenotyping soybeans for drought

responses using remote sensing techniques and non-destructive physiological analysis | carvalho | global science and technology. *Global Science and Technology*, 1–16.

<https://rv.ifgoiano.edu.br/periodicos/index.php/gst/article/view/808/459>

Ju, C. H., Tian, Y. C., Yao, X., Cao, W. X., Zhu, Y., & Hannaway, D. (2010). Estimating leaf chlorophyll content using red edge parameters. *Pedosphere*, 20(5), 633–644.

[https://doi.org/10.1016/S1002-0160\(10\)60053-7](https://doi.org/10.1016/S1002-0160(10)60053-7)

Karikari, B., Chen, S., Xiao, Y., Chang, F., Zhou, Y., Kong, J., Bhat, J. A., & Zhao, T. (2019). Utilization of Interspecific High-Density Genetic Map of RIL Population for the QTL Detection and Candidate Gene Mining for 100-Seed Weight in Soybean. *Frontiers in Plant Science*, 10, 1001. <https://doi.org/10.3389/fpls.2019.01001>

Korte, L. L., Specht, J. E., Williams, J. H., & Sorensen, R. C. (1983). Irrigation of Soybean Genotypes During Reproductive Ontogeny II. Yield Component Responses 1. *Crop Science*, 23(3), 528–533. <https://doi.org/10.2135/cropsci1983.0011183x002300030020x>

Kovar, M., Brestic, M., Sytar, O., Barek, V., Hauptvogel, P., & Zivcak, M. (2019). Evaluation of Hyperspectral Reflectance Parameters to Assess the Leaf Water Content in Soybean. *Water*, 11(3), 443. <https://doi.org/10.3390/w11030443>

Kumagai, E., Aoki, N., Masuya, Y., & Shimono, H. (2015). Phenotypic plasticity conditions the response of soybean seed yield to elevated atmospheric CO₂ concentration. *Plant Physiology*, 169(3), 2021–2029. <https://doi.org/10.1104/pp.15.00980>

Li, X., Zhang, X., Zhu, L., Bu, Y., Wang, X., Zhang, X., ... & Xing, H. (2019). Genome-wide association study of four yield-related traits at the R6 stage in soybean. *BMC genetics*, 20(1), 39.

Machikowa, T., & Laosuwan, P. (2011). Path coefficient analysis for yield of early maturing soybean. *Songklanakarin Journal of Science & Technology*, 33(4).

Mahbub, M. M., Rahman, M. M., Hossain, M. S., Mahmud, F., & Kabir, M. M. (2015). Genetic variability, correlation and path analysis for yield and yield components in soybean (No. RESEARCH).

- Maimaitijiang, M., Sagan, V., Sidike, P., Hartling, S., Esposito, F., & Fritschi, F. B. (2020). Soybean yield prediction from UAV using multimodal data fusion and deep learning. *Remote Sensing of Environment*, 237, 111599. <https://doi.org/10.1016/j.rse.2019.111599>
- Matiu, M., Ankerst, D. P., & Menzel, A. (2017). Interactions between temperature and drought in global and regional crop yield variability during 1961-2014. In J. L. Gonzalez-Andujar (Ed.), *PLoS ONE* (Vol. 12, Issue 5, p. e0178339). Public Library of Science. <https://doi.org/10.1371/journal.pone.0178339>
- Montesinos-López, A., Montesinos-López, O. A., Cuevas, J., Mata-López, W. A., Burgueño, J., Mondal, S., Huerta, J., Singh, R., Autrique, E., González-Pérez, L., & Crossa, J. (2017). Genomic Bayesian functional regression models with interactions for predicting wheat grain yield using hyper-spectral image data. *Plant Methods*, 13(1), 1–29. <https://doi.org/10.1186/s13007-017-0212-4>
- Nouri, A., Etminan, A., Teixeira da Silva, J. A., & Mohammadi, R. (2011). Assessment of yield, yield-related traits and drought tolerance of durum wheat genotypes (*Triticum turgidum* var. durum Desf.). *Australian Journal of Crop Science*, 5(1), 8.
- Panda, S. S., Ames, D. P., & Panigrahi, S. (2010). Application of Vegetation Indices for Agricultural Crop Yield Prediction Using Neural Network Techniques. *Remote Sensing*, 2(3), 673–696. <https://doi.org/10.3390/rs2030673>
- Pandey, J. P., & Torrie, J. H. (1973). Path Coefficient Analysis of Seed Yield Components in Soybeans (*Glycine max* (L.) Merr.) 1 . *Crop Science*, 13(5), 505–507. <https://doi.org/10.2135/cropsci1973.0011183x001300050004x>
- R Development Core Team. (2005). R: a language and environment for statistical computing. R Foundation for Statistical Computing. <https://doi.org/10.1016/j.dendro.2008.01.002>
- Randelović, P., Đorđević, V., Milić, S., Bale ević-Tubić, S., Petrović, K., Miladinović, J., & Đukić, V. (2020). Prediction of soybean plant density using a machine learning model and vegetation indices extracted from RGB images taken with a UAV. *Agronomy*, 10(8), 1108. <https://doi.org/10.3390/agronomy10081108>

Ray, D. K., Mueller, N. D., West, P. C., & Foley, J. A. (2013). Yield Trends Are Insufficient to Double Global Crop Production by 2050. *PLoS ONE*, 8(6), e66428.

<https://doi.org/10.1371/journal.pone.0066428>

Rocha, J. R. do A. S. de C., Machado, J. C., & Carneiro, P. C. S. (2018). Multitrait index based on factor analysis and ideotype-design: proposal and application on elephant grass breeding for bioenergy. *GCB Bioenergy*, 10(1), 52–60. <https://doi.org/10.1111/gcbb.12443>

Shakoor, N., Lee, S., & Mockler, T. C. (2017). High throughput phenotyping to accelerate crop breeding and monitoring of diseases in the field. In *Current Opinion in Plant Biology* (Vol. 38, pp. 184–192). Elsevier Ltd. <https://doi.org/10.1016/j.pbi.2017.05.006>

Sinclair, T. R. (2011). Challenges in breeding for yield increase for drought. In *Trends in Plant Science* (Vol. 16, Issue 6, pp. 289–293). Elsevier Current Trends.

<https://doi.org/10.1016/j.tplants.2011.02.008>

Specht, J. E., Diers, B. W., Nelson, R. L., de Toledo, J. F. F., Torrion, J. A., & Grassini, P. (2015). Soybean (Vol. 59901, pp. 311–355). John Wiley & Sons, Ltd.

<https://doi.org/10.2135/cssaspecpub33.c12>

Tattaris, M., Reynolds, M. P., & Chapman, S. C. (2016). A Direct Comparison of Remote Sensing Approaches for High-Throughput Phenotyping in Plant Breeding. *Frontiers in Plant Science*, 7(AUG2016), 1131. <https://doi.org/10.3389/fpls.2016.01131>

Weber, V. S., Araus, J. L., Cairns, J. E., Sanchez, C., Melchinger, A. E., & Orsini, E. (2012). Prediction of grain yield using reflectance spectra of canopy and leaves in maize plants grown under different water regimes. *Field Crops Research*, 128, 82–90.

<https://doi.org/10.1016/j.fcr.2011.12.016>

Westneat, D. F., Wright, J., & Dingemanse, N. J. (2015). The biology hidden inside residual within-individual phenotypic variation. *Biological Reviews*, 90(3), 729–743.

<https://doi.org/10.1111/brv.12131>

Wijewardana, C., Alsajri, F. A., Irby, J. T., Krutz, L. J., Golden, B., Henry, W. B., Gao, W., & Reddy, K. R. (2019). Physiological assessment of water deficit in soybean using midday leaf

water potential and spectral features. *Journal of Plant Interactions*, 14(1), 533–543.

<https://doi.org/10.1080/17429145.2019.1662499>

Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-regression: A basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58(2), 109–130.

[https://doi.org/10.1016/S0169-7439\(01\)00155-1](https://doi.org/10.1016/S0169-7439(01)00155-1)

Xianzhi, W., Xiaojuan, Z., Rong, Z., Aihua, S., Xuejun, W., Shuping, C., Dezhen, Q., & Xinan, Z. (2010). QTL analysis of seed and pod traits in soybean RIL population (pp. 441–448).

Yu, C., Qu, Z., Zhang, Y., Zhang, X., Lan, T., Adelson, D. L., Wang, D., & Zhu, Y. (2017). Seed weight differences between wild and domesticated soybeans are associated with specific changes in gene expression. *Plant Cell Reports*, 36(9), 1417–1426. <https://doi.org/10.1007/s00299-017-2165-5>

Yuan, W., Wijewardane, N. K., Jenkins, S., Bai, G., Ge, Y., & Graef, G. L. (2019). Early Prediction of Soybean Traits through Color and Texture Features of Canopy RGB Imagery. *Scientific Reports*, 9(1), 1–17. <https://doi.org/10.1038/s41598-019-50480-x>

Zanon, A. J., Streck, N. A., & Grassini, P. (2016). Climate and Management Factors Influence Soybean Yield Potential in a Subtropical Environment. *Agronomy Journal*, 108(4), 1447–1454. <https://doi.org/10.2134/agronj2015.0535>

Zarco-Tejada, P. J., González-Dugo, M. V., & Fereres, E. (2016). Seasonal stability of chlorophyll fluorescence quantified from airborne hyperspectral imagery as an indicator of net photosynthesis in the context of precision agriculture. *Remote Sensing of Environment*, 179, 89–103. <https://doi.org/10.1016/j.rse.2016.03.024>

Zhang, J., Han, W., Huang, L., Zhang, Z., Ma, Y., & Hu, Y. (2016). Leaf Chlorophyll Content Estimation of Winter Wheat Based on Visible and Near-Infrared Sensors. *Sensors*, 16(4), 437. <https://doi.org/10.3390/s16040437>

Zhang, X., Zhao, J., Yang, G., Liu, J., Cao, J., Li, C., Zhao, X., & Gai, J. (2019). Establishment of Plot-Yield Prediction Models in Soybean Breeding Programs Using UAV-Based

Hyperspectral Remote Sensing. *Remote Sensing*, 11(23), 2752.

<https://doi.org/10.3390/rs11232752>

Zhou, J., Zhou, J., Ye, H., Ali, M. L., Nguyen, H. T., & Chen, P. (2020). Classification of soybean leaf wilting due to drought stress using UAV-based imagery. *Computers and Electronics in Agriculture*, 175, 105576. <https://doi.org/10.1016/j.compag.2020.105576>.

CONCLUSIONS

In the first chapter consistent results were found in terms of estimating physiological traits in soybean under drought condition using leaf hyperspectral reflectance. Additionally, it was possible to using few selected wavelengths to select cultivars considering multiple physiological traits simultaneously. These results might be useful to support new high-throughput phenotyping platforms and allows on screening a great number of genotypes in a breeding program.

On the second chapter, yield-related traits were predicted also using leaf hyperspectral reflectance. Out of the traits with good prediction capacity, highlight can be given for seeds per pod and one-hundred grain weight as they are highly associated to grain yield in soybean. Moreover, it was show that a good association can be made between spectral data and selection indexes based on phenotypic information considering yield components traits.

Forecasting yield is one of the most critical aims in soybean breeding since it can save time and labor for harvesting only the most prominent genotypes. Efforts in prediction yield using remote sensing approach are at the top-edge phenotyping, and therefore, using physiological and yield-related traits can help paving the way to help to achieve such challenging goal.