University of São Paulo
"Luiz de Queiroz" College of Agriculture

Towards a new approach to estimate soybean yield at the field level

Marcelo Chan Fu Wei

Dissertation presented to obtain the degree of Master in Science. Area: Agricultural Systems Engineering

Piracicaba
2021

Marcelo Chan Fu Wei
Agonomist

Towards a new approach to estimate soybean yield at the field level

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Advisor:
Prof. Dr. **JOSÉ PAULO MOLIN**

Dissertation presented to obtain the degree of Master in Science. Area: Agricultural Systems Engineering

Piracicaba
2021

International Cataloging Data in Publication
LIBRARY DIVISION – DIBD/ESALQ/USP

Wei, Marcelo Chan Fu

Towards a new approach to estimate soybean yield at the field level / Marcelo Chan Fu Wei. - - versão revisada de acordo com a resolução CoPGr 6018 de 2011. - - Piracicaba, 2021.

73 p.

Dissertação (Mestrado) - - USP / Escola Superior de Agricultura “Luiz de Queiroz”.

1. Agricultura de precisão 2. Componentes de produtividade 3. *Glycine max* 4. Processamento de imagens 5. Regressão linear I. Título

ACKNOWLEDGMENTS

To **my family** that always supported and encouraged me to applied for a graduate program.

To the professor dr. **José Paulo Molin** for being my advisor whom always behaved professionaly and patiently allowing the development of this dissertation more smoothly.

To **all the integrants** of the **Precision Agriculture and Mechanization group (gMAP)** and **Precision Agriculture Laboratory (LAP)** from USP/ESALQ that made this phase less tiring and more productive.

To the **employees** from the **Biosystems Engineering Department** who are always willing to assist us on our activities.

To the **Coordination of Superior Level Staff Improvement (CAPES)** and to the **National Council for Scientific and Technological Development (CNPq)** for the scholarships provided during the master`s degree.

To **Luiz de Queiroz Agrarian Studies Foundation (FEALQ)** for the financial support provided to the publication of an article as a result of this dissertation.

SUMMARY

RESUMO

Uma nova abordagem para estimar a produtividade da soja em condição de campo

A soja é uma das culturas agrícolas mais cultivadas no mundo e para aumentar a sua produção de forma sustentável, os produtores devem aplicar técnicas de manejo mais conscientes considerando não apenas o maior retorno econômico, mas também a redução do impacto negativo ambiental. Para isso, os dados de produtividade apresentam-se como camadas de dados fundamentais que podem auxiliá-los no manejo. A estimativa da produtividade da soja é geralmente obtida por monitores de produtividade e por modelos agrometeorológicos, os quais apresentam limitações relacionados ao uso de seus dados. Por exemplo, a qualidade dos dados provenientes dos monitores de produtividades é afetada pela condição do sistema sensor, tornando obrigatório a realização de uma filtragem de dados. Já, os modelos agrometeorológicos são limitados não apenas pela quantidade de variáveis necessária em seu modelo, mas também pela resolução espacial e temporal. Ciente das limitações dos métodos atuais, torna-se oportuno investigar novos métodos para estimar a produtividade da soja no campo considerando a disponibilidade de acesso a dados de produtividade da soja e suas componentes e dos avanços tecnológicos na agricultura. Os objetivos foram: (a) analisar as relações da produtividade da soja e sua componentes (número de grãos - NG e massa de mil grãos - TGW) a partir de dados publicados e propor um modelo para estimar a produtividade da soja e (b) realizar uma análise exploratória para obtenção de dados das componentes de produtividade da soja a partir de métodos bidimensionais (2*D*) em plantas no estádio fenológico R8. Inicialmente, foi conduzida uma revisão bibliográfica para obtenção de dados de produtividade da soja e suas componentes para compor o conjunto de dados de treinamento. A partir dele, modelos de regressão linear foram ajustados e, posteriormente, testados em um conjunto de dados de validação composto por 58 amostras de campo. A análise exploratória do processamento de imagens foi realizada com imagens de soja obtidas em condição de campo a partir de um sensor de uma câmera comercial. As imagens obtidas foram submetidas ao tratamento básico de imagem seguida da aplicação de um algoritmo booleano para detectar as componentes da produtividade. Como resultado, foram gerados três modelos de regressão linear: o primeiro com base na TGW, o segundo ajustado pelo NG e o terceiro relacionado com TGW e NG. O modelo com base em NG apresentou melhor acurácia, indicando que a variável NG é uma potencial componente de produtividade que pode ser utilizada como variável preditora. A análise exploratória da aplicação do processamento de imagens RGB gerou resultados que auxiliam trabalhos futuros para aprimorar as técnicas de processamento de imagem para obtenção da varíavel NG. Para que a detecção da componente de produtividade seja aprimorada, foram identificadas etapas importantes que devem ser aplicadas antes do uso do algoritmo booleano, como a limiarização. Neste estudo foi proposto um modelo para estimar a produtividade da soja a partir de uma variável preditora (NG) e também foi apresentado um potencial método de processamento de imagem para obtenção desta variável a partir de imagem RGB obtida em campo.

Palavras-chave: Agricultura de precisão, Componentes de produtividade, *Glycine max*, Processamento de imagens, Regressão linear

ABSTRACT

Towards a new approach to estimate soybean yield at the field level

Soybean is one the most cultivated crops in the world. Looking to improve soybean production in a more sustainable way, it is imperative for farmers to make better management practices, improving the economic return and reducing the negative impact on the environment. Thus, yield data becomes an essential data layer that can support farmers to boost the management practices related to the soybean crop. Commonly, soybean yield data are obtained from combine harvester equiped with yield monitor or by agrometeorological models. Both present limitations regarding the data usage. For example, yield monitor data quality are affected by the sensor system that require data filtering process and agrometeorological yield models are limited to the amount of predictor variables required and its spatial and temporal resolution. Thus, aware of these limitations, it becomes an opportunity to investigate new methods to estimate soybean yield at the field level considering soybean yield and yield component data availability and the advances of technological applications in agriculture. The objectives of this study were to: (a) analyze the relationship among soybean yield and its components (number of grains - NG and thousand grains weight -TGW) in a worldwide range from available data and propose a yield estimation model based on yield components and (b) make an exploratory analysis on two dimensions (2*D*) methods to obtain data related to soybean yield from plants at the R8 phenological stage. Initially, it was conducted a literature review to gather soybean yield and its components data to compose the training dataset. Linear regression models based on soybean yield components were fitted on the training dataset and evaluated on a validation dataset composed of 58 samples collected at the field level. To conduct the exploratory analysis of image processing techniques on soybean, images were taken at the field level from a consumer-grade camera, then basic image processing techniques were applied followed by the application of a Boolean-based algorithm to detect the soybean components. As a result, it was generated three linear regression models: the first based on the TGW, the second based on NG and the third based on TGW and NG. The yield model based on NG presented the highest prediction accuracy, indicating that NG can be a potential yield component to be used as predictor variable. The exploratory analysis of the application of image processing techniques on RGB images provided potential results that support further investigation to improve image processing to gather NG data. Aiming to improve the detection of the yield components through image processing, it was found important steps that must be applied before application of the Boolean-based algorithm such as thresholding. In this study, it is proposed a new soybean yield estimation model relying on the use of one predictor variable (NG) and also, it is presented a potential image processing method that allows gathering NG data from soybean RGB images taken at the field level.

Keywords: *Glycine max,* Image processing, Linear regression, Precision agriculture, Yield components

LIST OF FIGURES

LIST OF TABLE

1. INTRODUCTION

The perception of the limited food production to supply the demand of the world population growth is debated and corcerns specialists for decades. One of the main agricultural products cultivated in the world is soybean (*Glycine max* L. Merril), an important oilseed and a valuable agricultural commodity traded around the world. As soybean grain demand increases with population growth, it is required to improve crop production increasing its yield in a sustainable way (Sentelhas et al., 2015).

Aware of soybean importance to the world as a source of food, it is important to obtain data that supports decision makers to enhance soybean production (Batistella et al., 2011; Andrade et al., 2013). In this sense, yield estimation becomes a fundamental parameter that can help farmers to improve their crop management (Gong et al., 2013).

Crop yield is one of the most important data layer used in precision agriculture (PA) to initiate the investigation regarding the existence of spatial variability in the fields (Vega et al., 2019) and it is related to several regulatory interactions that occur at the field level varying spatially and temporally (Wijerwardana et al., 2019). The identification and quantification of these interactions with the PA practices allow to identify possible factors responsible for the uneven development of the crop (Mulla & Scheppers, 1997).

Currently, soybean yield data are obtained from yield monitors (Fulton et al., 2018) and agro-meteorological models (Betbeder et al., 2016). Yield monitors present several limitations that affect the quality of the data such time consuming steps to improve its calibration, regular maintenance and verification of the sensor system (Fulton et al., 2018). Crop yield models are limited to its spatial resolution and the need of several predictor variables as input data to develop accurate yield prediction models such as (a) crop canopy vegetation indices (Al-Gaadi et al., 2016) and (b) historical yield and climate data (Setiyono et al., 2018).

Independently the recent methods used to estimate soybean yield, it is common the application of machine learning techniques. For example, Schwalbert et al. (2020) presented a model to perform in-season soybean yield estimation applying machine learning (statistical) techniques, satellite imagery and weather data. In their research, they were able to forecast soybean yield at the municipality-scale, but relying on the use of multiple predictor variables. Despite their results, the authors highlighted the trade-off that might occur applying these methods in extreme weather conditions, indicating that more error are found in years that weather is different than the normal.

Yu et al. (2016) proposed methods to improve soybean yield estimation and plant maturity prediction applying machine learning techniques into a database composed of data obtained at the field level. In their study, they evaluated the estimated yield with variables such canopy geometric features, plot row length and yield data obtained from a combine harvester. As a result, they found that plot row length and canopy geometric features improved yield estimation accuracy.

Note that despite the potential proposed models from both studies (Yu et al., 2016 and Schwalbert et al. 2020), they relied on the use of several predictor variables to estimate soybean yield, including the use of vegetation indices, and, in some cases, it is not feasible to acquire that amount of predictor variables. In addition to that, the use of vegetation indices to fit soybean yield prediction models is not an acceptable approach among the scientific community as it is for wheat crops that present high correlation among yield and biomass, and, as a consequence, normalized difference vegetation index (NDVI) as shown in Lai et al. (2018), for example.

As machine learning application become more popular in agriculture, studies have been done focusing on crop yield prediction based on its yield components through image processing techniques combined with machine learning statistical approaches. Most of the studies were conducted on crops that yield is accounted by its fruits such as apple (Hung et al., 2015), mango (Qureshi et al., 2017) and peach (Kurtulmus et al., 2014).

There are few studies made to estimate crop yield based on its components related to annual crops. Reza et al. (2019) predicted rice yield at the field level applying a cluster algorithm to delimitate the area of rice grains in an image to estimate its yield. As a result, they were able to acccurately estimate rice yield based on the grain area.

Alharbi et al. (2018), aware that wheat spikes are related to wheat yield, developed an automated counting system based on image processing and machine learning techniques to count wheat spikes from wheat growth images. Their system achieved accuracy rates of up to 90.7% on wheat spikes counting, highlighting the potential to apply image processing and machine learning techniques to acquire yield components data.

Aware of the current limitations to estimate soybean yield at the field scale and that soybean grain is the harvested product of the crop, efforts should be done aiming to improve soybean yield estimation exploring their relation like what was done for rice (Reza et al., 2019) and wheat (Alharbi et al., 2018).

Thus, it is an opportunity to investigate soybean yield and its components relation in a worldwide range considering the theorethical aspects already reported by Fehr & Caviness (1997) and explore possible application of computer vision techniques considering soybean plant images at the field level. The first hypothesis of this study is that soybean yield can be predicted using its yield components as input variable considering that yield is a function of its yield components. The second is that soybean yield components related to yield can be estimated applying computer vision techniques on images obtained at the field level.

The objectives of this study were: (a) analyze the relationship among soybean yield and its components in a worldwide range from available data and propose a yield estimation model based on yield components and (b) make an exploratory analysis on two dimensions (2*D*) and three dimensions (3*D*) methods to obtain data related to soybean yield of plants at the R8 phenological stage.

2. LITERATURE REVIEW

2.1 Precision agriculture

The definition of precision agriculture (PA) is constantly evolving and it varies from author and year. Among some definitions, there are:

(I) Combination of the use and practice of agricultural inputs under specific conditions at the field level aiming to do the right thing, at the right place, at the right moment and at the right manner (Pierce et al., 1994);

(II) Targetting inputs for the production of agricultural crops according to the needs of the crop in a localized manner (Stafford, 1996);

(III) Management strategy that uses information technology to obtain data from several sources that support decision making associated with plant production (The National Research Council, 1997);

(IV) Systematic approach related to soil and crop management to minimize decision uncertainty from a better understanding and management of the spatial and temporal variability (Dobermann et al., 2004);

(V) Agricultural tool set that aims to a site-specific crop management enhancing the use of agricultural inputs and its efficiency and reducing environmental negative impacts (Gebbers & Adamchuk, 2010);

(VI) Application and use of geospatial tools, sensors, remote sensing and global positioning to identify and adopt management strategies according to the field level variability (Zhang & Kovacs, 2012);

(VII) Agricultural management system that considers the field level variability to apply agricultural inputs at the right place, right rate and right time reducing the negative environmental impact and improving farmers profitability (Mulla, 2013);

(VIII) Management strategy based on the use of technology to gather data from different sources to support decision makers (Li & Chung, 2015);

(IX) Management of field spatial and temporal variability from the use of information and communication technologies (Fountas et al., 2016)

(X) ”Management strategy that gathers, processes and analyzes temporal, spatial and individual data and combines it with other information to support management decisions according to estimated variability for improved resource

use efficiency, productivity, quality, profitability and sustainability of agricultural production." (ISPA, 2019);

Despite the variation among PA definitions, in general, they consider the field level spatial and temporal variability aiming to improve crop management, reducing negative environmental impacts and improving net economic return.

To improve the understanding of variability in the agricultural context, it is important to be aware of the variability classification according to Zhang et al. (2002), as shown below:

(I) Topographic variability (elevation, inclination, proximity to boundaries and streams, and others);

(II) Soil variability [physico-chemical properties such as nutrient availability (e.g. nitrogen, phosphorus, potassium and others), soil hidrogen potential (pH), soil water holding capacity and others];

(III) Crop variability related to plant density, height, plant stress caused by biotic and abiotic factors, biomass and others;

(IV) Variability caused by anomalous factors such weed and insect infestation, nematodes, diseases and others;

(V) Variability induced by different management practices such crop rotation, soil amendment application, soil conversation practices and others;

(VI) Yield variability which is the consequence of the interaction among all the others variability.

In view of the theoretical understanding of PA and variability, studies have been carried out to comprehend variability and analyze the feasibility of applying PA to different crops. As an example, there are those that applied PA techniques and obtained yield improvement (Monzon et al., 2018), developed platforms to increase PA adoption as an auxiliary decision tool (Wachowiak et al., 2017), analyzed economical feasibility of applying PA (Silva et al., 2007) and analyzed yield temporal variability (Yost et al., 2017).

Baio et al. (2017) carried out a financial analysis comparing the crop management based on precision agriculture and conventional techniques on a cotton crop. In their case, they found that the application of PA techniques provided higher profitability indicating the feasibility of PA application into their scenario even having higher initial costs with machinery and sensor systems. Note that their study was conducted only for one harvest season.

Colaço et al. (2020) conducted a long-term experiment in citrus to compare a variable and uniform rate fertilisation treatments over five growing seasons considering economic viability, energy and nutrient balance. As a result, in terms of energy, variable rate applications presented lower embodied energy when compared to the uniform application indicating that the PA application resulted not only in a reduced energy input to the system, but also reduced the risk associated to nutrient loss corroborating to minimize the negative impact that agricultural practices can cause to the environment. Considering the profitability, they found areas that variable rate applications yielded higher profits than uniform applications. However, the opposite was also found. Their findings indicate that there are still challenges to overcome the field level variability and to provide the optimal fertilisation recommendations that are still based on standards disregarding the complex spatial interaction among production factors.

Considering Baio et al. (2017) and Colaço et al. (2020), in which both considered yield as one main parameter to be evaluated, we can conclude that there is still a lack of conclusive studies regarding the economical and environmental consequences of the PA applications. Therefore, more efforts on PA studies must be carried out to improve crop management practices.

Aiming to support the development of a more sustainable agriculture relying on the principles of PA, this study focused on the development of a new method to estimate soybean yield at the field level aware that soybean is a major cultivated crop in the world and yield is an essential data that can be used to improve management decisions.

2.2 Soybean crop and yield estimation

Soybean [*Glycine max* (L.) Merrill] is one of the top five oilseed worldwide cultivated. In the 2018/2019 soybean harvest, United States of America was the country that produced the most (120.52 million tons) with a cultivated area of 35.45 million hectares followed by Brazil with 119.00 million tons produced in an area of 35.90 million hectares (USDA, 2020).

Brazilian soybean is mainly destined for three segments: a) 7% for the market stock, b) 44% for the exportation of fresh grain and c) 49% for the processing of the grain which 79% is destined to soybean meal production and 21% to oil production. About 52% of the soybean meal produced is destined for exportation, while the remaining 48% is used for domestic consumption mainly for animal feed (APROSOJA, 2014; CONAB, 2017).

In view of the economic importance of soybean crop, several studies have been conducted to improve soybean crop production applying different management practices and goals: genetic traits (Todeschini et al., 2019), seed composition (Assefa et al., 2019) seed population (Camicia et al., 2018), planting date (Mourtzinis et al., 2017) and seed treatment (Kandel et al., 2016). Despite their different goals, all the cited studies evaluated crop yield as a comparison factor, highlighting the importance of yield in all the cases.

Soybean yield estimation can be conducted by different methods that varies according to the objectives and spatial resolution. At regional scale purposes that aims to forecast crop production, the use of low spatial resolution are suitable as those found in informative bulletins available at National Supply Company (CONAB), Brazilian Agricultural Research Corporation (EMBRAPA), National Agricultural Statistics Service (NASS) and United States Department of Agriculture (USDA), for example. However, if the objective is to support crop management practices at the field level, more detailed data are required, thus, high spatial resolution is needed, like the study of Maestrini & Basso (2018).

Considering low spatial resolution, agrometeorological models can be used to predict crop yield besides its limited spatial resolution and also provide usefull data to support the development of public policies and commodity future contracts. As an example, Schwalbert et al. (2020) developed a soybean yield forecast model at a regional scale, South of Brazil, applying machine learning techniques to a dataset composed of average yield data distinguished by city from the Brazilian Institute of Geography and Statistics (IBGE) database, Rural Environmental Registry (CAR) database, vegetation indices (VIs) obtained from satellite imagery and climatological data such as temperature and rainfall. As a result, the authors found satisfactory prediction accuracy, highlighting the importance of using these predictive models to support the development of public policies.

Other studies considering larger areas to forecast crop yield has been already developed, for example, Sakamoto (2020) proposed a maize and soybean yield estimation model for the United States at the county level. His study was based on the application of different machine learning to a database composed of remote sensing data (VIs), environmental data (daily mean temperature, precipitation, shortwave radiation and soil moisture) and available data from the NASS. As a result, the proposed method yielded satisfactory results.

Fuzzo et al. (2020), conscious that agrometeorological yield models relies on vast number of predictor variables, proposed a method for soybean yield prediction over large areas (Parana State, Brazil) considering data (radiometric temperature and NDVI) from one single satellite (MODIS sensor) and compared to models that relied on the use of more variables. Their findings highlight that the use of data from MODIS sensor alone were suitable to predict soybean yield over large areas.

Note that, Schwalbert et al. (2020), Fuzzo et al. (2020) and Sakamoto (2020) studies were well fitted to the purposes that do not require high spatial resolution. On the other hand, when higher spatial resolution is required to support crop decision making at the field level, it is imperative to obtain data at the field level.

Considering within field, Maestrini & Basso (2018) predicted spatial patterns of different crops (maize, cotton, soybean and wheat) relying on the use of yield data and vegetation indices obtained from harvester with yield monitor and satellite imagery, respectively. In their study, they found that models based on historical yield maps outperformed the models based on VIs, indicanting yield data as the best predictor of spatial patterns and highlighting the importance of using yield maps.

Soybean yield can be estimated from a grain yield monitor system usually coupled with a mass/volume flow sensor, moisture sensor, global navigation satellite system (GNSS) and user-friendly task computer. The system estimates the mass of the harvested grain over an assigned area and the data accuracy is dependent on moisture sensor calibration, mass/volume flow sensor and GNSS (Fulton et al., 2018).

The advantage to use grain yield monitor system is that they can deliver on-the-go data collection, however it is highly dependent on the quality of the sensor system calibration, grain elevator cleanness and wiring harnesses, for example (Fulton et al., 2018). Thus, aware of these limitations, it is required to perform post-harvest yield data processing/cleaning to improve yield accuracy estimation as it is known that poor yield data leads to poor decisions (Fulton et al., 2018).

Aware of the possible errors found in yield monitor data, several authors proposed different methods to cope with different types of errors (Blackmore & Moore, 1999; Menegatti & Molin, 2004; Sudduth & Drummond, 2007; Spekken et al., 2013; Leroux et al., 2018 and Vega et al., 2019). But, despite their efforts, there is no consensus among the scientific community about the best method to filter yield monitor data. Thus, as an alternative to yield monitors,

remote sensing techniques and new sensor technologies that are based on high-throughput phenotyping research are likely to be used in the future to accurately estimate yield (Fulton et al., 2018).

2.3 Remote sensing

Remote sensing (RS) can be described as a combination of technology and techniques for data collection done by sensor systems that are carried on different platforms, positioned at a distance from the target that can vary from a few meters to many thousands of kilometers and operated in various parts of the electromagnetic spectrum (Ferguson & Rundquist, 2018). Within the remote sensing concept, there is the proximal sensing whose definition is similar, except for the distance of the sensor from the target that must be placed within 2 m (Viscarra Rossel et al., 2011).

RS in PA is used to collect and analyze data about crop and soil from sensors systems. Among soil sensing system there are: (a) electrical and electromagnetic sensors, (b) optical and radiometric sensors that cover diffrent parts of the electromagnetic spectrum (gamma ray, x-ray, ultraviolet, visible, infrared, microwaves and radio waves spectrum), (c) sensors that rely on the mechanical interaction among sensors and soil and (d) electrochemical sensors (Adamchuk et al., 2018). Regarding to plant sensing systems, they can be based on: (a) crop canopy reflectance and fluorescence sensors, (b) machine vision, (c) light detection and ranging (LiDAR) and time-of-flight (ToF) cameras, (d) thermal sensors, (e) mechanical sensors and (f) acoustic sensors (Adamchuk et al., 2018).

Despite the different sensor measurement systems, RS applications also depend on platform (proximal, aerial or orbital), region of the electromagnetic spectrum (visible, infrared and microwave), number and length of spectral bands (panchromatic, multispectral and hyperspectral), spatial resolution (high, medium and low), temporal resolution (frequency of revisiting the sensor system), radiometric resolution (shades of gray) and energy source (active or passive sensors) used for data collection (Metternicht, 2008).

RS can assist PA by providing spatial and temporal data on crops, contributing to a better monitoring of biotic and abiotic factors, positively impacting on decision-making such as manpower required for crop treatment, machinery demand and logistics planning before harvest (Andrade et al., 2014). Beyond theses advantages, RS can be a tool used for different purposes such as: to forecast crop yield (Shanahan et al., 2001, Wei et al., 2020), to classify

plant specie (Kamilaris & Prenafeta-Boldú, 2018), to count the number of plants (Gnädinger & Schmidhalter, 2017), to estimate crop phenological stage (Hufkens et al., 2019), to predict crop biomass and height (ten Harkel et al., 2020).

Most of the remote sensing applications related to yield prediction in soybean crops are based on a dataset composed of vegetation indices and agrometeorological data covering large areas as shown in studies from Sakamoto et al. (2020) and Schwalbert et al. (2020). Besides these studies using large areas and low spatial resolution, there are also studies conducted in soybean areas where data was obtained with high spatial resolution such as Eugenio et al. (2020).

Regardless low or high spatial resolution, Sakamoto et al. (2020), Schwalbert et al. (2020) and Eugenio et al. (2020) relied on the use of vegetation indices to fit yield estimation models applying different machine learning techniques and even aware that there is no consensus about the veracity of using VIs to estimate soybean yield, a different approach when it is considered wheat crops. In wheat crops, yield estimation models based on VIs are commonly found in studies (Lai et al., 2018; Aranguren et al. 2020) as there is a positive correlation among VIs and yield.

RS tools can be based on different sensor systems such LiDAR tecnology that can scan the target on two and three dimensions quickly with high accuracy (Rosell & Sanz, 2012), hyperspectral, multispectral, thermal, infrared and red, green, blue (RGB) sensors that provide spectral data of the target (Kamilaris & Prenafeta-Boldú, 2018). Among these sensors, RGB sensors, commonly associated with consumer-grade cameras, can be considered a feasible sensing method to provide crop data with high temporal resolution that can be used to support agriculture, for example, in crop development monitoring (Richardson et al., 2009).

The application of RGB sensors in agriculture has been largely studied to provide high quality data in an easy, efficient and low-cost manner. For example, Nijland et al. (2014) evaluated the use of consumer-grade cameras with RGB and near infrared wavelenghts for vegetation monitoring and phenology studies. As a result, they found that consumer-grade cameras with RGB and NIR for detection of plant phenology and crop monitoring are promising tools under field situations and controlled laboratory experiment. In addition, the authors highlights that consumer-grade RGB cameras are simple and affordable tools for agricultural purposes such plant detection and development. Fernandez-Gallego et al. (2019) found that

consumer-grade cameras can be used to delivery RGB indices with robust statistical approach for durum wheat yield prediction or screening purposes.

Aware that consumer-grade cameras are being used for remote sensing applications in agriculture, Zhang et al. (2017) conducted a study to evaluate the performance of two consumer-grade cameras, a RGB camera and a modified near-infrared (NIR) camera for crop identification and leaf area index (LAI) prediction at different spatial resolutions. As a result, they found that at finer spatial resolutions (0.4 to 4 m) for both purposes, LAI estimation and crop identification, the combined use of consumer-grade RGB camera and modified NIR camera provided results with sufficient accuracy highlighting their suitability to be used in agricultural purposes.

Yang et al. (2019) applied a deep convolution neural network for rice grain yield estimation comparing yield models developed from RGB and multispectral dataset to yield models from vegetation indices dataset. As a result, RGB and multispectral yield models outperformed VI yield models, highlighting the need to study more about the application of RGB and multispectral dataset to forecast crop yield.

Feng et al. (2020) estimated cotton yield using UAV-based multi-sensor imagery composed of RGB camera, multispectral camera and infrared thermal camera. They developed yield models based on VIs, plant height and canopy temperature and found that plant height is an interesting variable to be used to predict cotton yield.

Reza et al. (2019) estimated rice yield applying an image processing method to RGB images obtained from a sensor coupled at an unmanned aerial vehicle. As a result, they sucessfully estimated rice yield achieving a 0.98 of coefficient of determination between the proposed method and ground truth.

Maimaitijiang et al. (2020) predicted soybean yield from UAV using multimodal data fusion and deep learning. Their dataset was composed of canopy spectral, structure, thermal and texture features besides yield. They applied different machine learning (ML) methods such partial least square, support vector machine, random forest regression and deep neural networks (DNN) finding that DNN outperformed the others ML methods. In conclusion, they reported that using a single sensor, spectral features or RGB-based structure information present similar prediction accuracy and their combination improved yield estimation accuracy.

Note that most of the studies have related yield with different vegetation indices and dependending on the crop, yield estimation models were able to accurately estimate the crop

yield. Despite these approaches, there are some researches studying the relation among crop characteristics and its yield. For example, Yu et al. (2016) developed methods to improve soybean yield estimation and predict plant maturity with an UAV based platform. In their study, they applied random forest classifier to measure crop geometric features (canopy area and plot row length) and found high correlations with final yield.

2.4 Crop yield and its components

The investigation of crop yield and its components has been studied in crops such: *Vigna mungo* -black gram (Malik et al., 2008), *Cicer arietinum* L. - chickpea (Lake & Sadras, 2014), *Lupinus mutabilis* - lupin (Sandaña & Calderini, 2012), *Zea mays* - maize (Borghi et al., 2012; Liu et al., 2011), *Oryza sativa* - rice (Hussain et al., 2012; Nascente et al., 2012; Reza et al., 2019), *Carthamus tinctorius* - safflower (Mohamadzadeh et al., 2011; Zafaranieh, 2015), *Sesamum indicum* - sesame (Eskandari et al., 2009), *Helianthus* - sunflower spp (Sedghi et al., 2008), *Triticum aestivum* - wheat (Alharbi et al., 2018; Azimi et al., 2013; Malek-Mohammadi et al., 2013; Rehman & Farooq, 2016) and *Phaseolus* spp - white bean (Kazemi et al., 2012).

The usual yield components found in the literature are: thousand grains weight, hundred grains weight, number of grains per m^2, number of grains per cob, number of grains per panicule and others. The studies varies from different management practices to analyze the effect on yield (Hussain et al., 2012) until genetics approaches (Malik et al., 2008). Despite knowing that yield components are characteristics inherent to yield, these data are not commonly used to developed crop yield models.

In the case of soybean crops, it has been already reported that soybean yield estimation must rely on three main characteristcs: number of pods per plant, number of grains per pod and grain weight. These characteristics are the least sensitive to external factors at the final development stage (R8) and are the result of the interaction among genotype and enviroment (Egli, 1988). In addition to that, according to Fehr & Caviness (1997), at R8, the plant must present more than 95% of their pods at the maturity color, absence of leaves and grain humidity lower than 15%, indicating that it is close to the ideal harvest point (Figure 1), thus, farmers should pay attention to the R8 stage to start harvesting.

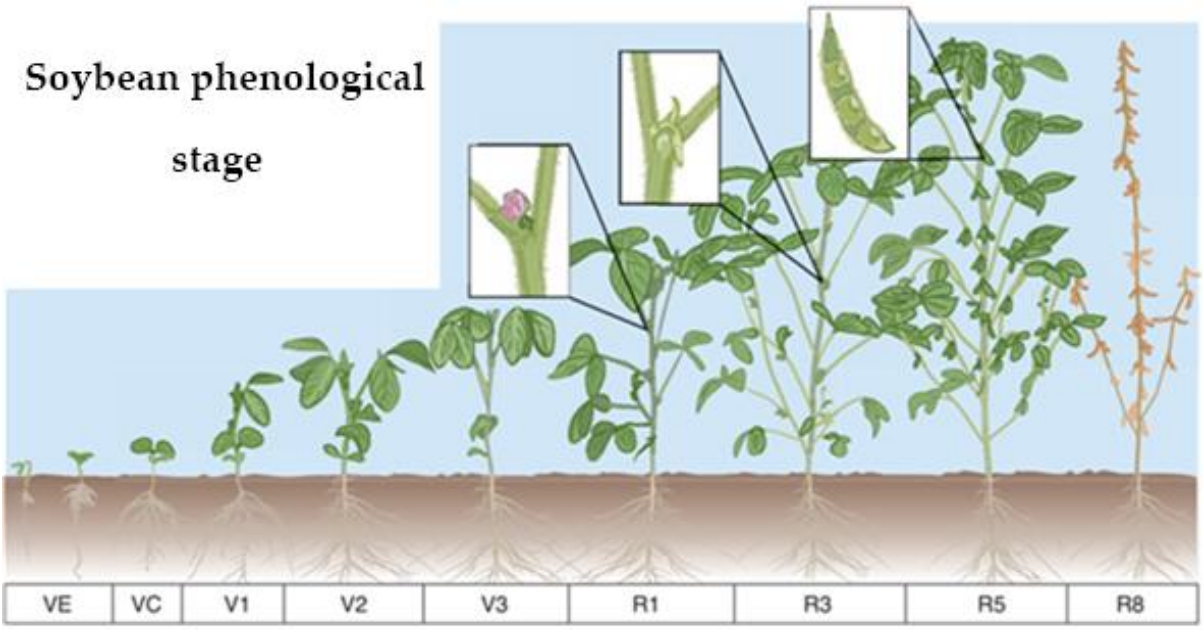

Figure 1. Soybean phenological stage.
V = vegetative stages; E = emergence; C = cotyledon; V(n) = nth node; R = reproductive stages; R1 = beginning bloom; R3 = beginning pod; R5 = beginning seed; R8 = full maturity. Source: adapted from Nordby et al. (2004).

2.5 Data availability and technological advances in agriculture to estimate yield

The access of data and its availability in agricultural research is facilitated because of the advance of technological approaches to a wider public mainly associated to platforms that allow their access on the "open access" policy. Several platforms allow users to obtain free data, most are related to satellite imagery data [EarthExplorer (2020)] and climatological data [National Oceanic and Atmospheric Administration – NOOA (2020)]. Besides these type of platforms, there are those where users can make their dataset available, for example GitHUB (2020) and research challenges supported by private companies that provide extensive database to be explored according to the challenge goal such Sygenta Crop Challenge (2020).

Most of the data available for agricultural purposes are related to indirect factors such as VIs and climatological data. However, as computer processing power increases, technological advances are implemented in agriculture and machine learning techniques become more popular, new methodologies for classification and prediction purposes in agriculture have been developed. For example, Leroux et al. (2019) proposed a method to forecast maize yield based on remote sensing data and application of machine learning methods, in their case, multiple linear regression (MLR) and random forest (RF) regression. As a result, they found that RF regression outperformed MLR, highlighting that the combination

of remote sensing data and RF regression can provide important yield estimation in environments where yield data are scarce.

As the application of machine learning becomes more popular in agriculture, computer vision has also gainned more attention. Computer vision (method to obtain data from images) and machine learning (method that an algorithm learn from a given data) techniques have been implemented together in different purposes making task easier and providing important data that in the past was not feasible to obtain, thus the focus in this study will be the application of these techniques in yield prediction. Their application is more common in fruit crops, where the harvested product is the organ direct related to yield. For example, Qureshi et al. (2017) evaluated different classification methods to count the number of fruits on mango trees canopies finding coefficient of determination values varying from 0.55 to 0.91 when compared to the manual counting. Their results highlight the possibility to apply combined techniques (computer vision and machine learning) in yield prediction purposes.

Other studies related to fruits organs and computer vision were also made. Tran et al. (2017) developed a method to count the number of dragon fruits at the field level. As a result, they were able to accurately count the number of fruits and predict its yield. Rahnemoonfar & Sheppard (2017) proposed a method to count the number of tomatoes fruit from images taken at the field level and found high accuracy rates. Note that despite the suitability of these approaches on counting the number of fruits, it is imperative to relate their value to its real yield, especially knowing that the fruits are traded considering not only the amount in pieces, but also by their unit weight.

A studied applying machine learning and computer vision techniques have been also conducted to predict rice yield different from the common approach based on the counting of fruits. Reza et al. (2019) estimated rice yield based on the grain area information and found results with coefficient of determination reaching up to 0.98 when compared to the ground-truth.

Table 1 shows researches conducted using big data, machine learning and computer vision related to agricultural purposes. Note that there are several original and review papers about these subjects, but few are related to its application on soybean crop and soybean yield estimation.

Table 1. Published papers related to big data, computer vision and machine learning in agriculture.

Author	Focus in agriculture	Type of paper
Coble et al., 2018	Big data	Review
Delgado et al., 2019	Big data	Review
Kamilaris et al., 2017	Big data	Review
Weersink et al., 2018	Big data	Review
Wolfert et al., 2017	Big data	Review
Lin et al., 2020a	Computer vision (object detection)	Original article
Lin et al., 2020b	Computer vision (object detection)	Original article
Mavridou et al., 2019	Computer vision (object detection)	Review
Patrício and Rieder, 2018	Computer vision (object detection)	Review
Tian et al., 2019	Computer vision (object detection)	Review
Gómez et al., 2019	Machine learning (prediction)	Original article
Khan et al., 2020	Machine learning (prediction)	Original article
Koirala et al., 2019	Machine learning (prediction) + Computer vision (object detection)	Review
Liakos et al., 2018	Machine learning (prediction and classification)	Review
Maimaitijiang et al., 2020	Machine learning (prediction)	Original article
Michelon et al., 2018	Machine learning (prediction)	Original article
Nevavuori et al., 2019	Machine learning (prediction)	Original article
Niedbala, 2019	Machine learning (prediction)	Original article
Pantazi et al., 2016	Machine learning (prediction)	Original article
Reza et al., 2019	Machine learning (prediction) + Computer vision (object detection)	Original article
Sabanci et al., 2017	Machine learning (classification) + Computer vision (object detection)	Original article
Verma et al., 2018	Machine learning (prediction)	Original article
Wang et al., 2018	Machine learning (prediction)	Original article
Wei et al., 2020	Machine learning (prediction)	Original article

3. MATERIAL AND METHODS

PART OF MATERIAL AND METHODS SECTION HAS BEEN ALREADY PUBLISHED AT AGRICULTURE (WEI, M. C. F. & MOLIN, J. P. 2020. Soybean yield estimation: a linear regression approach. Agriculture, v. 10, p. 348 – 361, doi:10.3390/agriculture10080348) – ANNEX A

This work has been divided into two: (a) analysis of the relationship among soybean yield and its components and (b) exploratory analysis on two dimensions (2*D*) and three dimensions (3*D*) methods to obtain data related to soybean yield of soybean plants at the R8 phenological stage.

3.1 Analysis of the relationship among soybean yield and its components

Initially, a bibliographic survey was conducted with the following keywords to gather data for the training dataset: i) yield, ii) soybean, iii) estimation, iv) forecast, v) thousand grains weight, and vi) hundred grains weight at the Web of Knowledge, Google Scholar and Scopus research databases from 2010 to 2019. This survey aimed to look for articles, thesis and dissertations that presented published data with at least the variables "yield", "thousand grains weight (TGW)" or "hundred grains weight (HGW)". Despite these variables, it was also, gathered "treatment", "soybean variety", "year of study", "location of study", "number of grains" if available. After gathering the data, the total number of grains (NG), when not available, was estimated from Equation I (adapted from Egli & Yu, 1991; Lindsey, 2020).

$$\mathrm{NG\ (grains\ m^{-2})} = \frac{\mathrm{100\ x\ Yield\ (kg\ ha^{-1})}}{\mathrm{TGW\ (g\ 1000\ grains^{-1})}} \qquad \text{Eq. I,}$$

where, NG is the number of grains (grains m^{-2}) and TGW is thousand grains weight(g 1000 $grains^{-1}$). Note that, when hundred grains weight (HGW) data was available instead of TGW, HGW was converted into TGW multiplying its value by 10.

The validation dataset was composed of 58 randomly selected on-field samples of 1 m^2 from a soybean (cultivar TMG 7062) field collected at Piracicaba, Brazil (22.7356 ºS, 47.6479 ºW) in March of 2019. The variable NG (grains m^{-2}) was carefully and manually counted grain by grain, thousand grains weight (g 1000 $grains^{-1}$) and yield (kg ha^{-1}) were obtained from these samples after the drying process where the counted grains were placed into a forced circulation air oven under 65 ºC for 72 h.

The variables “TGW”, “NG” and “Yield” of the training and validation dataset were submitted to a descriptive analysis (number of observations, minimum, median, mean and maximum values, standard deviation, sample variance and coefficient of variation - CV) and all the previous variables and “Country” were used to apply a scatterplot (“Yield” *versus* “Country”, “Yield” *versus* “NG”, “Yield” *versus* “TGW” and “TGW”*versus* “NG”) and Pearson`s correlation coefficient (r).

The statistical approach used for predicting soybean yield was the linear regression because number of observations were larger than number of variables (Gromping, 2009). All the observations from the variables “Yield”, “TGW” and “NG” were used to fit three linear regression models: (A) Yield as function of TGW and NG, (B) Yield as function of NG and (C) Yield as function of TGW.

The performance of the fitted models were evaluated comparing their root mean squared error (RMSE – Eq. II), mean absolute error (MAE – Eq. III) and coefficient of determination (R^2 – Eq. IV) for both training and validation dataset.

$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathrm{N}}\sum_{i=1}^{\mathrm{N}}(y_{\mathrm{i}} - \hat{\mathrm{y}})^2} \qquad \text{Eq. II,}$$

where RMSE = root mean squared error (kg ha^{-1}), n = number of samples, y = observed variable response and ŷ = predicted variable response.

$$\mathrm{MAE} = \frac{1}{\mathrm{N}}\sum_{i=1}^{\mathrm{N}}|y_{\mathrm{i}} - \hat{\mathrm{y}}| \qquad \text{Eq. III,}$$

where MAE = mean absolute error (kg ha^{-1}), n = number of samples, y = observed variable response and ŷ = predicted variable response.

$$\mathrm{R}^2 = 1 - \frac{\sum(y_{\mathrm{i}} - \hat{\mathrm{y}})^2}{\sum(y_{\mathrm{i}} - \bar{\mathrm{y}})^2} \qquad \text{Eq. IV,}$$

where R^2 = coefficient of determination, n = number of samples, y = observed variable response, ŷ = predicted variable response and ȳ = mean value of the observed response variable.

The statistical analysis was performed on R environment (R Core Team, 2019).

3.2 Exploratory analysis on two dimensional (2*D*) and three dimensional (3*D*) methods to obtain data related to soybean yield on plants at the R8 phenological stage

In this subsection, it was conducted an exploratory analysis on how to obtain structural data of soybean yield components, in this case, number of pods and number of grains applying two methods: (A) 3*D* method (LiDAR sensor) and (B) 2*D* method (RGB consumer-grade sensor).

3.2.1 Three dimensional (3*D*) LiDAR-based application

The sensor systems used to collect 3*D* point cloud data was composed of a LiDAR sensor model LMS 200 (Sick, Waldkirch, Germany) collecting data at 75 Hz with field of view of 180º, a GNSS receiver with Real-Time Kinematic (RTK) model GR-3 RTK (Topcon, Tokyo, Japan), portable Toughbook CF-19 (Panasonic, Osaka, Japan), a height-adjustable support (1 m to 2 m) and a fixed metallic frame (1 m x 1 m).

Before start collecting soybean point of cloud data, the sensor system was tested on a known target, in this case, a plastic chair. To obtain point cloud data and merge it with the GNSS receiver data, it was used an algorithm similar to Colaço et al. (2017). After data acquisition, data processing was conducted using the *Cloud Compare* software. As shown in Figure 2, data acquistion from the sensor system and data processing proved to be a reliable method on acquiring and generating the chair shape as a point cloud.

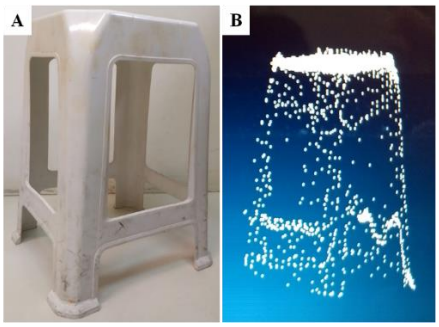

Figure 2. Example of a plastic chair obtained from two sensors. (A) Red, green and blue (RGB) image obtained from a consumer-grade camera and (B) point cloud image obtained from a 3*D* Light Detection and Ranging (LiDAR) sensor system after data post-processing.

Therefore, with the sensor system and data processing working correctly, it was collected soybean data with the plants at the phenological stage R8 in an experimental area

that belongs to the University of São Paulo (USP), in Piracicaba, on March of 2018 (soybean 2017/2018 harvest season).

3.2.2 Two dimensional (2*D*) RGB image acquisition and processing

It was used the rear passive optical sensor from a smartphone model Moto G5 Plus (Lenovo, Sorocaba, Brazil) to capture the 2*D* RGB images. The sensor was set to capture images with resolution of 12 megapixels (width of 4032 pixels and length of 3024 pixels), horizontal and vertical resolution of 72 dpi, 24 bits depth (shades of gray), no-flash mode, focal length of 4 mm and exposure time of 1/747 sec.

Before starting RGB image data acquistion in the field, it was tested image processing techniques containing only soybean pods in a scheme. Image processing was carried out using ImageJ platform (The National Institute of Health, 2018). Basic image processing, such image transformation (reduction of the grayscale of the image), binarization (process to convert a grayscale image to a binary image – black and white), thresholding (method to segment the image based on the histogram) and filtering (noise removal), was carried out using in-built tools from ImageJ. After these processes, it was applied a Boolean-based algorithm (unsupervised method based on the possible values of true or false) developed to ImageJ platform known as "3*D* Object Counter" (Bolte and Cordelières, 2006).

As shown in Figure 3, it is possible to visualize that "3*D* Object Counter" yielded suitable results regarding the identification and accounting the desired object under controlled conditions. Note that, in this scenario, it was used only soybean pods while the desired scenario to be tested should be on-field situations.

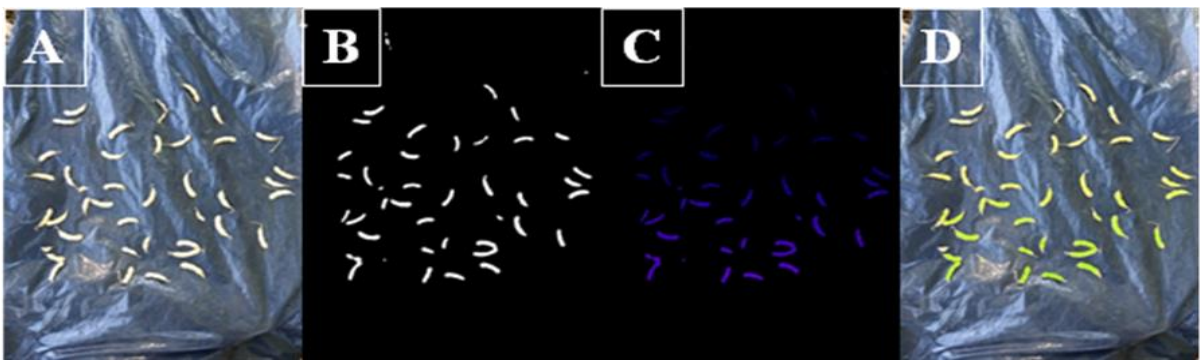

Figure 3. Soybeans pods. (A) RGB image. (B) 8 bits image. (C) Identified and counted objects. (D) Overlaid image (A+C).

Thus, aware of the possibility to obtain yield components data applying image processing techniques on 2*D* RGB images, data collection at field level was carried out in the attempt to obtain the number of pods and number of grains of soybean plants at the R8 stage in a square delimited area of 1 m^2.

Initially, in the harvest season 2017/2018, soybean images were taken at the field level limited to an area of 1 m^2 with soybean plants at the R8 stage aiming to establish a procedure of image processing focused on the detection of soybean plant characteristics. As an exploratory analysis, images were taken without considering lighting and time conditions, however the distance from the camera to the target was established according to the scene size that must contain at least the target area within 1 m^2. Then, it was tested the application of "3*D* Object Counter" tool.

Besides that, it was collected a sample of 1 m^2 of soybean plants and stuck in the ground in another area without other soybean plants. Four images at different views in relation to the planted linefrom this sample were taken: (a) side view, (b) front view, (c) upper view and (d) upper-side view. The application of "3*D* Object Counter" results the number of detected objects (in this case, soybean pods) and the area of the detected objects. To estimate the number of grains through image processing, it was divided the detected area of the objects by the area value that corresponds to one grain, a similar approach that Reza et al. (2019) applied for rice crops. The comparison among estimated number of pods and number of soybean grains was compared to the manual counting (pod by pod and grain by grain) that was performed after image acquisition.

The statistical parameter used to compare the results from image processing against the manual counting was the absolute relative error in percentage since this work is an initial investigation of the potential to apply computer vision to obtain soybean number of grains. Absolute relative error was calculated applying the module of the value obtained by difference between manual and estimated values and dividing it by the manual value.

4. RESULTS AND DISCUSSION

PART OF THE RESULTS AND DISCUSSION SECTION HAS BEEN ALREADY PUBLISHED AT AGRICULTURE (WEI, M. C. F. & MOLIN, J. P. 2020. Soybean yield estimation: a linear regression approach. Agriculture, v. 10, p. 348 – 361, doi:10.3390/agriculture10080348) – ANNEX A

4.1 Analysis of the relationship among soybean yield and its components

Table 2 presents the descriptive analysis of the soybean training and validation dataset. Looking at the training dataset, it was found 707 observations of soybean data that contained at least "yield" and "thousand grains weight" or "hundred grains weight" from 56 eligible papers. It is shown a large difference between soybean yield and its yield components among countries. Yield presented a range from 167.1 to 10,170 kg ha^{-1}. Different yield values found across the world highlights the regional variability inherent among environment and genotype. Coefficient of variance of yield and NG are higher than 50% and TGW is about 24% indicating that TGW presents less variance than yield and NG variables. Looking at the validation dataset (Table 2), it shows that the CV among samples are higher for both yield and NG variables than TGW, a similar trend found in the training dataset.

Table 2. Descriptive analysis of the soybean training dataset and validation dataset.

Measurement	Training Dataset			Validation Dataset		
	Yield (kg ha^{-1})	TGW (g 1000 $grains^{-1}$)	NG (grains m^{-2})	Yield (kg ha^{-1})	TGW (g 1000 $grains^{-1}$)	NG (grains m^{-2})
n	707.0	707.0	707.0	58.0	58.0	58.0
Minimum	167.1	64.7	100.0	2100.0	301.5	1236.0
Median	2819.0	148.4	1849.0	3460.0	423.3	1888.0
Mean	2764.9	148.0	1826.0	3463.8	420.4	1889.0
Maximum	10,170.0	346.0	5557.0	5400.0	542.9	2975.0
Standard Deviation	1557.6	35.9	932.0	736.6	44.6	444.0
Sample Variance	2,425,981.0	1291.6	868,565.2	542,588.9	1991.8	196,958.0
CV (%)	56.3	24.3	51.1	21.3	10.6	23.5

n = number of observations; CV = coefficient of variation; TGW = thousand grains weight and NG = number of grains. Source: adapted from Wei & Molin (2020).

It is shown in Figure 4 soybean yield variability among countries despite year, genotype and treatments. Figure 4A, worldwide soybean yield, shows soybean yield distribution among countries indicating large yield differences within country level, on other words, representing

yield variability at a country level as expected once we can see in several governmental reports, for example, in the US and Brazil, United States Department of Agriculture—USDA and Brazilian Agricultural Research Corporation—EMBRAPA, respectively, that are responsible for forecasting and presenting harvested yield by location and year.

Soybean yield versus TGW, as shown in Figure 4B, presents a weak linear relationship (r = 0.50) between these two variables which will negatively impact the prediction accuracy of the linear regression model. Whislt, Figure 4C, soybean yield versus NG, presents a strong linear relationship between yield and NG (r = 0.92) indicating that the yield prediction model based on NG variable might present higher prediction accuracy than the model based on TGW. Comparing TGW and NG data (Figure 4D), there is a weak linear relationship (r = 0.19) indicating that there is no collinearity between them, therefore both predictor variables (NG and TGW) can be used in the linear regression yield predictive models.

Aware that yield prediction is one of the most important topics in PA and ML techniques have widely been applied in agriculture, Figure 4 highlights the opportunity to apply ML techniques to estimate soybean yield, but different from the common approach that relies on VIs, for example, this dissertation proposes linear models based on yield components (TGW and NG).

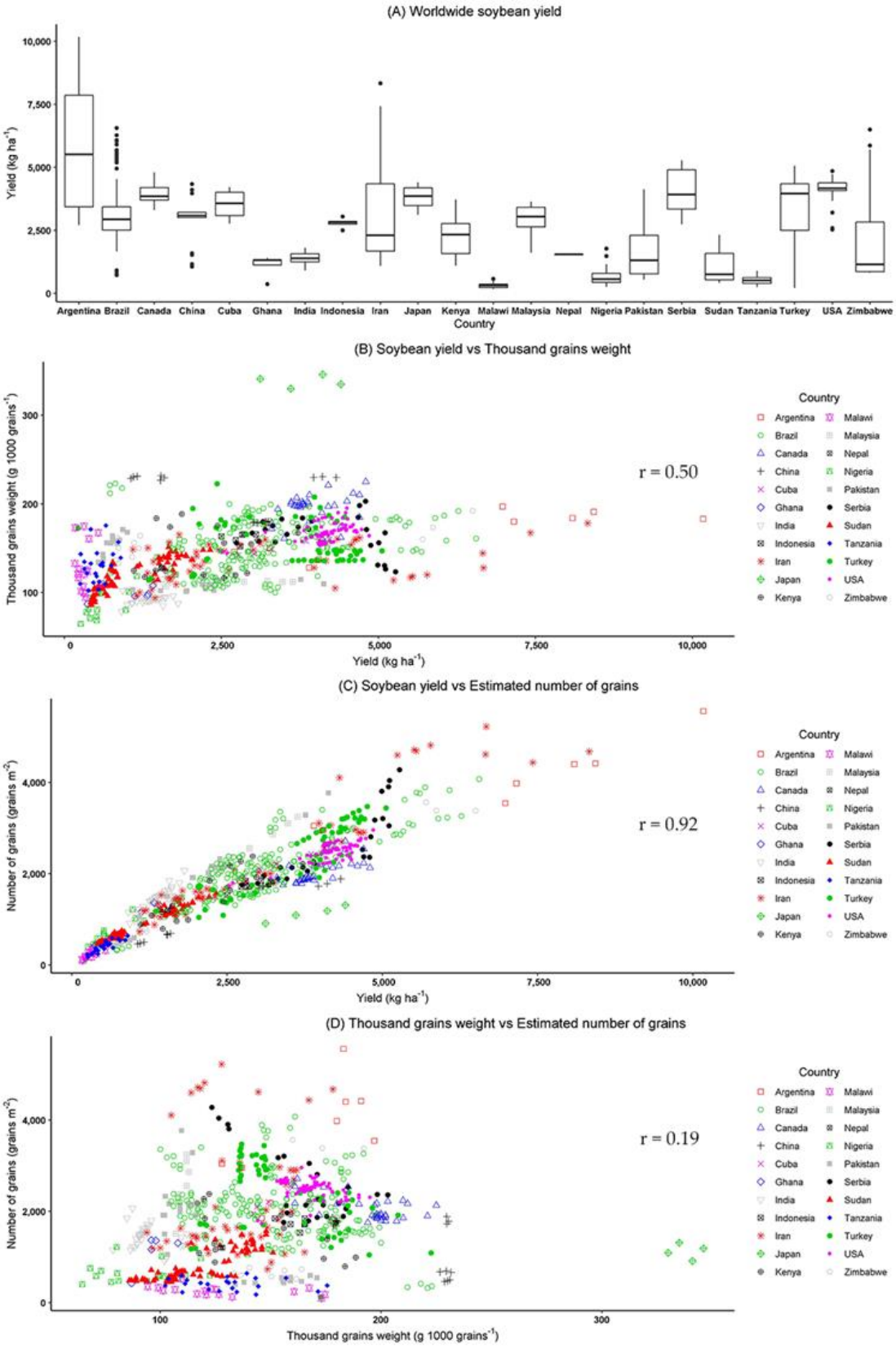

Figure 4. Boxplot and scatterplots of observed soybean yield (kg ha^{-1}) and its yield components by country. (A) Worldwide soybean yield (kg ha^{-1}) by country, (B) thousand grains weight (g 1000 $grains^{-1}$), (C) number of grains (grains m^{-2}) and (D) thousand grains weight (g 1000 $grains^{-1}$) versus number of grains (grains m^{-2}). r = Pearson`s correlation coefficient. Source: adapted from Wei & Molin (2020).

Aware that yield is function of number of pods (np), grain number per pod (gnp) and grain weight (gw) (Eq. V) adapted from Egli & Yu (1997) and Lindsey (2020), it is possible to combine np and gnp, generating a new variable, number of grains (NG), and providing Eq. VI. Thus, it is plausible to affirm that soybean yield is a direct function of NG and gw.

Regarding NG and TGW variables related to yield, NG had more influence on yield compared to gw/TGW as already reported in papers (Board, 1987; De Bruin & Pedersen, 2008; Robinson et al., 2009, Carciochi et al., 2019) and found in Figure 4.

$$\text{Yield (kg ha}^{-1}) = \text{np (unit ha}^{-1}) \text{ x ngp (grain unit}^{-1}) \text{ x gw (kg grain}^{-1}) \quad \text{Eq. V,}$$

where, np, ngp and gw are, respectively, number of pods (unit ha^{-1}), number of grains per pod (grain $unit^{-1}$) and grain weight (kg $grain^{-1}$).

$$\text{Yield (kg ha}^{-1}) = \text{NG (grains ha}^{-1}) \text{ x gw (kg grain}^{-1}) \quad \text{Eq. VI,}$$

where, NG and gw are, respectively, number of grains (grain ha^{-1}) and grain weight (kg $grain^{-1}$).

Figure 5 shows a scatterplot of the observed and predicted soybean yield applied to the training dataset. Linear models were generated from training dataset on which model A (Figure 5A) presented the highest R^2 (0.97) and lowest RMSE (288.78 kg ha^{-1}), results expected because the increment of variables used as predictors, usually yields higher R^2 and lower RMSE. Model B (Figure 5B), yield as function of NG, presented R^2 and RMSE of 0.85 and 455.98 kg ha^{-1}, respectively. As expected due to the weak linear relationship between yield and TGW, model C (Figure 5C) presented the lowest R^2 (0.26) and the highest RMSE (1343.17 kg ha^{-1}). Higher R^2 and lower MAE values were found for models A and B, instead of model C because NG had globally more influence than TGW to predict yield. Hence, it is expected that models A and B present a better prediction accuracy than model C when applied to the validation dataset.

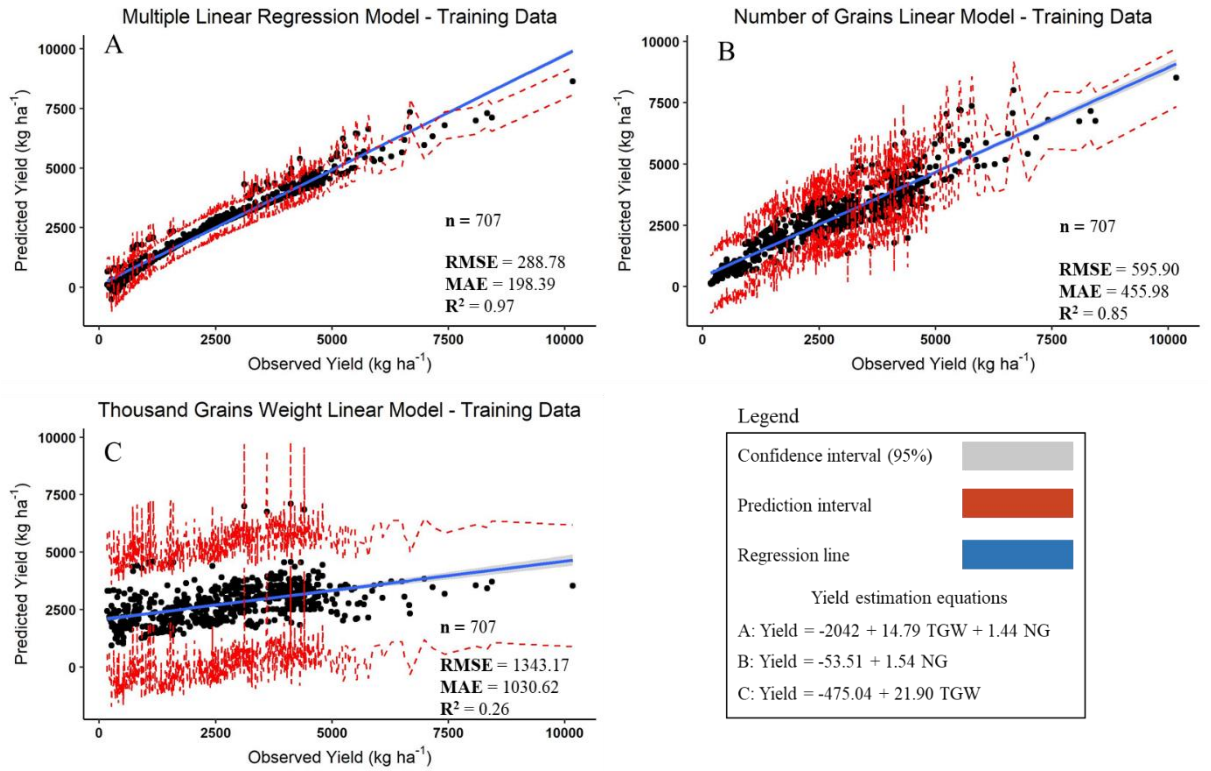

Figure 5. Scatterplot of soybean observed and predicted yield (kg ha^{-1}) applied to training dataset. (A) Yield estimation based on thousand grains weight (TGW in g 1000 grains^{-1}) and number of grains (NG in grain m^{-2}), (B) Yield estimation as function of number of grains (NG in grains m^{-2}) and (C) Yield estimation as functions of thousand grains weight (TGW in g 1000 grains^{-1}).
n = number of observations; MAE = mean absolute error (kg ha^{-1}); RMSE = root mean squared error (kg ha^{-1}); R^2 = multiple R-squared; Yield = soybean yield (kg ha^{-1}); TGW = thousand grains weight (g 1000 grains^{-1}); NG = number of grains (grains m^{-2}). Source: adapted from Wei & Molin (2020).

In Figure 6, models A, B and C generated from the training dataset were applied to the validation dataset containing 58 observations. R^2 values for models A (Figure 5A), B (Figure 5B) and C (Figure 5C) were 0.66, 0.70 and 0.02, respectively. Model B presented the highest R^2, lowest MAE (639.99 kg ha^{-1}) and RMSE (726.67 kg ha^{-1}) followed by model A (MAE – 3420.93 kg ha^{-1} and RMSE – 3449.80 kg ha^{-1}) and model C (MAE – 5267.52 kg ha^{-1} and RMSE – 5391.08 kg ha^{-1}). These results indicated that model B presented a higher accuracy rate to predict soybean yield based on number of grains. Note that, in order to achieve a good accuracy rate, both predicted and observed yield should present, not only high R^2, low RMSE and MAE, but also, the amplitude ratio closest as possible to one.

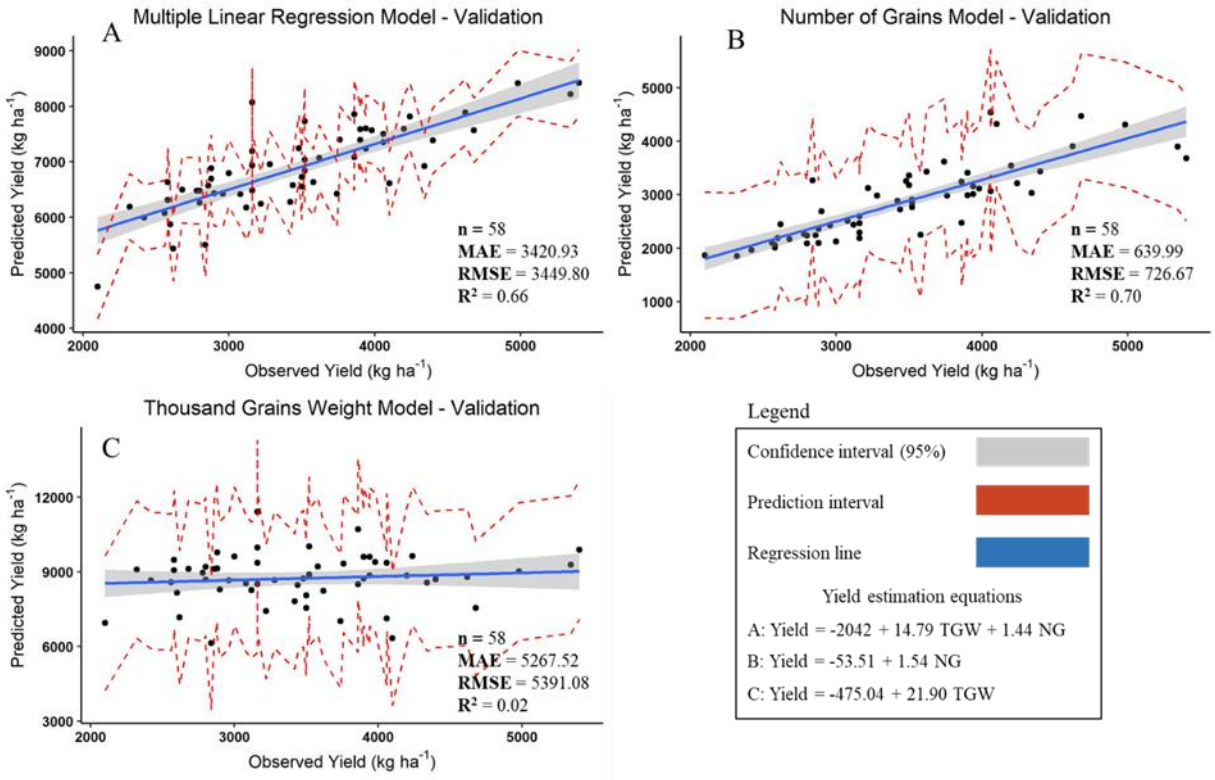

Figure 6. Scatterplot of soybean observed and predicted yield (kg ha^{-1}) applied to validation dataset. (A) Yield estimation based on thousand grains weight (TGW in g 1000 $grains^{-1}$) and number of grains (NG in grains m^{-2}), (B) Yield estimation as function of number of grains (NG in grains m^{-2}) and (C) Yield estimation as functions of thousand grains weight (TGW in g 1000 $grains^{-1}$).
n = number of observations; MAE = mean absolute error (kg ha^{-1}); RMSE = root mean squared error (kg ha^{-1}); R^2 = multiple R-squared; Yield = soybean yield (kg ha^{-1}); TGW = thousand grains weight (g 1000 $grains^{-1}$); NG = number of grains (grains m^{-2}). Source: adapted from Wei & Molin (2020).

This study presented a novel model to predict soybean yield based on its components and data published in a worldwide range applying linear regression models different from the common approach which rely on soil properties (Smidt et al., 2016), historical data about soil, yield and soybean variety (Marko et al., 2016), and vegetation indices (Christenson et al., 2016), for example. The model based on the variable NG (model B) presented the highest R^2 demonstrating its potential to be used as soybean yield predictive model.

The implication to apply this model is to acquire the predictor variable, NG, which is not easy at the moment. But, there are researchers working on high-throughput phenotyping, specifically, estimating the number of grains per pod in soybean plants (Uzal et al., 2018) and number of grains (Li et al., 2019). Despite both studies estimated the number of soybean grains applying computer vision techniques, their research were conducted under controlled environmental conditions and the ideal scenario should be at the field level considering all the variables that can affect the image composition.

Aware that yield monitors present several time consuming steps to calibration to improve its accuracy estimation (Fulton et al., 2018), computer vision techniques are being applied to estimate crop organs related to yield (Liu et al., 2018; Qureshi et al., 2017; Sa et al., 2016) and the accuracy performance of the developed model to predict soybean yield based on number of grains (model B), efforts should be made aiming the application of computer vision techniques to achieve accurately the number of grains at the field level to apply model B to forecast soybean yield as an alternative to the use of yield monitor data.

4.2 Exploratory analysis on two dimensional (2*D*) and three dimensional (3*D*) methods to obtain data related to soybean yield on plants at the R8 phenological stage

4.2.1 Three dimensional (3*D*) LiDAR-based application

The results from the 3*D* LiDAR-based application did not present potential to be used as a tool to obtain soybean plant characteristics at R8 stage, in this case, soybean grains and pods. This occured because it was not possible to acquire point cloud data and generate plant 3*D* modelling like what was made to the plastic chair. The hypothesis about this problem is that the emitted light beam from the LiDAR sensor was not able to return to the sensor making the data acquisition a failure. Hence, assuming that the problem might be related to the plant characteristcs and/or LiDAR sensor resolution and/or sensor system setup, future studies should address the stated problems above.

4.2.2 Two dimensional (2*D*) RGB image acquisition and processing

Due to the failure of the approach using the 3*D* LiDAR-based application similar to Colaço et al. (2017) for acquiring and modelling soybean plant characteristics at R8 stage, efforts have been focused on the exploratory application of 2*D* RGB image acquisition and processing.

Considering that soybean yield can be estimated by the number of grains, RGB image acquisition was focused at the field level. Figure 6 presents an example of image acquistion and processing at the field level with soybean plants at the R8 phenological stage. Figure 7A, a RGB image with 24 bits, contains in the scheme soybean plants within a fixed frame with 1 m^2 and unsidered objects (white object in the background and soybean plants beyond the fixed frame)

that might commit image processing. Figure 7B is the transformation of the RGB image (24 bits) to a 8 bits image, reducing the image size and turning image processing faster. The image results of the application of the tool "3*D* Object Counter" are shown in Figures 7C and 7D. Both Figures 7C and 7D represents the surface object detected and counted, however the first shows the object number labelling and the second does not. In Figures 7E and 7F, it is shown an overlaid image of Figures 7B and 7D and Figures 7A and 7D, respectively. Overlaid images allow the user to visualize the quality of the image processing. In this case, it is shown in Figure 7F that there is still a challenge to limit the target objects within the desired area (the fixed frame with 1 m^2).

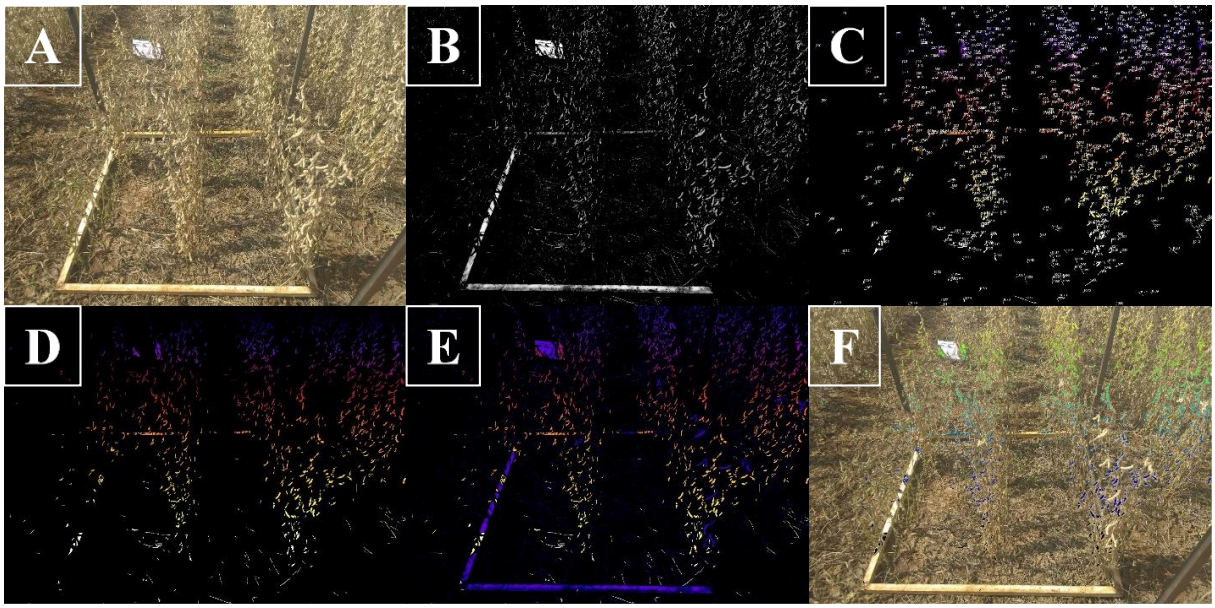

Figure 7. In-field RGB image acquisition of soybean plants at the R8 phenological stage and image post-processing results. (A) Original RGB image, (B) Image transformation. (C) Results of the object dectected and counted with labelling, (D) Results of the object dectected and counted without labelling. (E) Overlaid images B and D and (F) Overlaid images A and D.

As the development of soybean plants are uneven due to the result of the biotic and abiotic factors, a topic widely studied among researchers such as Chen & Wiatrak (2010) and Carverzan et al. (2018), different plant scenarios indicating spatial variability can be found at the field level. For example, Figure 8 shows soybean plants that are within an area considered in the R8 phenological stage where it is expected that they do not present leaves and green stems, however it can be found plants with and without leaves in the same scheme, an indication of variability at the field level. Despite these occurance, image processing was applied in the attempt to explore data acquisition methods. Note that different than Figure 7, in Figure 8, it

is more clear the effectiveness of "3*D* Object Counter" application aiming to detect the targetted object, soybean plants within the 1 m^2.

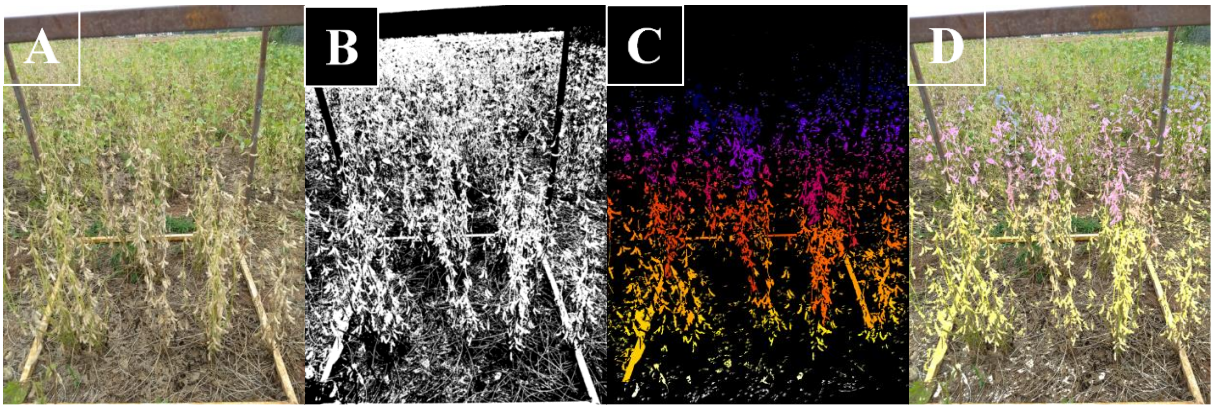

Figure 8. Scheme containing soybean plants with and without leaves. (A) RGB image. (B) 8 bits image. (C) Detection and counting of the detect objects in the scheme. (D) Overlaid image (A+C).

Some difficulties can be found in image processing as it can be seen in Figure 9. Figures 9A, 9B and 9C represents the RGB image, image transformation from 24 bits to 8 bits and image thresholding, respectively. Figures 9D and 9E are the result of different image processing. The first represents the result applying an additional image graph cut tool before the application of "3*D* Object Counter" and the second corresponds to the result without applying graph cut tool. The visual results of Figures 9D and 9E expose the need to apply more image processing techniques before the application of "3*D* Object Counter". Figure 9D shows soybean plants well identified in the scheme applying an image graph-cutting technique. On the other hand, without graph-cutting technique, it is shown in Figure 9E that object identification became negatively affected and more undesired objects such as soil surface and weed are identified instead of the soybean plants. Note that the application of thresholding are capable to remove most of the undesired objects from the scheme, however the application of a graph cut improved to detection as it can be seen in Figure 9D when compared to Figure 9E.

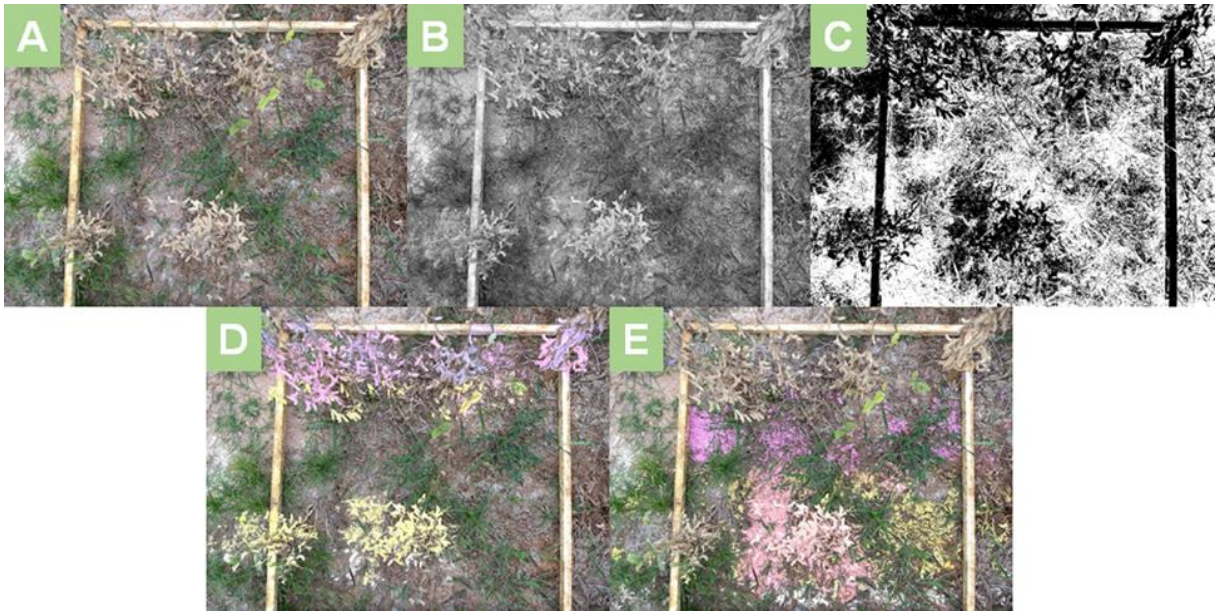

Figure 9. Image containing soybean plants, weeds, fixed frame and soil. (A) RGB image. (B) 8 bits image. (C) Thresholded image. (D) Result of the object detect after applying a graph cut tool. (E) Result of the object detect without the application of a graph cut tool.

In Figure 10, it is shown another example of "3*D* Object Counter" application result considering whether graph cut tool was applied or not. Note that, in this scenario there are more plants present in the same scheme beyong the delimited area different from Figure 9. Figure 9A shows the original image taken at the field level. Figure 9B and 9C represent the 8 bits image and thresholded image, respectively. Note that applying the image processing techniques and the graph cut tool (Figure 9D), it was possible to detect almost all the plants desired in the scheme. While, without the application of a graph cut tool (Figure 9E), more undesired objects were detected and the desired objects were not accounted correctly.

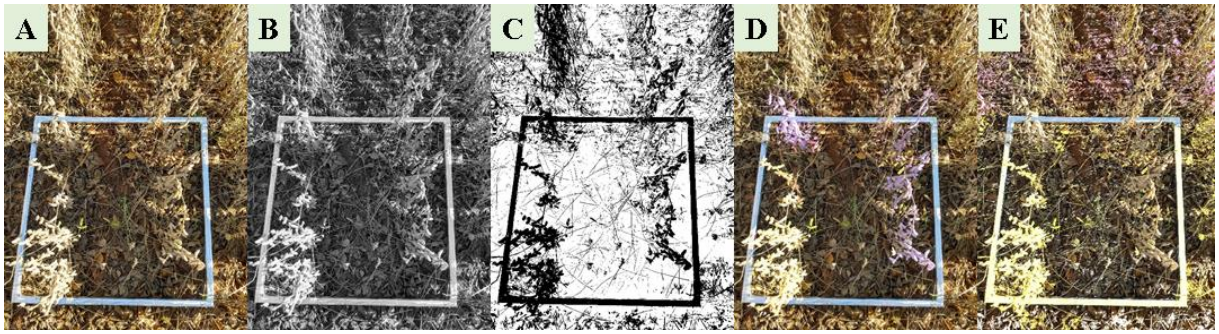

Figure 10. Soybean plants at the R8 phenological stage in-field situation. (A) RGB image. (B) 8 bits image. (C) Thresholded image. (D) Result of the object detect after applying a graph cut tool. (E) Result of the object detect without the application of a graph cut tool.

In this study, it was applied several image processing techniques in-field situations trying to gather the number of grains and number of pods of soybean plants from RGB images. Thus,

as a result, it was identified some processing steps that can improve the detection of the soybean plant characteristcs. Figure 11 represents the procedure containing the steps necessary to be applied before the use of "3*D* Object Counter". The initial step consists on acquiring the RGB image (Figure 11A), then convert the 24 bits image to a 8 bits image (Figure 11B). After image transformation, a thresholding process (Figure 11C) must be applied as it can remove undesired object from the scheme. Depending on the RGB image to be processed, it will be necessary to apply an image filtering (Figure 11D) by removing pixels considering their mean, median, maximum and mininum values within the scheme. In situations that these image filtering do not yield accurate results, an image graph cut tool can be a potential solution to overcome that problem. After the application of basic image processing, the scheme is ready to be submitted to the proccess of "3*D* Object Counter" tool (Figure 11E).

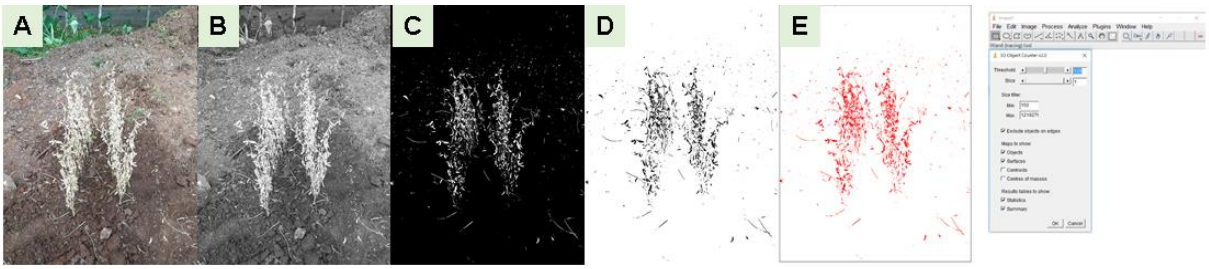

Figure 11. Procedure for soybean image processing before the application of "3*D* Object Counter" tool. (A) RGB image (24 bits). (B) Image transformation (8 bits). (C) Image thresholding. (D) Image filtering (mean, median, maximum and mininum) and/or image cutting. (E) Application of "3*D* Object Counter" tool developed for ImageJ.

Considering the proposed image processing procedure in this study suitable for soybean image analysis, it was tested that process to detect and estimate the number of pods and grains of a soybean sample from images taken at different views.

It is shown in Figure 12 the side view of a soybean sample at R8 stage. Figure 12A represents the original image captured. Figure 12B refers to the transformed image after the application of thresholding, limiarization and binarization. Figure 12C presents the identified objects after applying "3*D* Object Counter" tool. To confirm if the algorithm was correctly identifying the desired objects, it was overlaid the surface object detected (Figure 12C) with the original image (Figure 12A) resulting in the Figure 12D. Note that in Figure 12, even without the presence of other soybean plants next to the sample, there are still undesired object on the scheme such as weeds and grass, but it was not accounted during the image processing step.

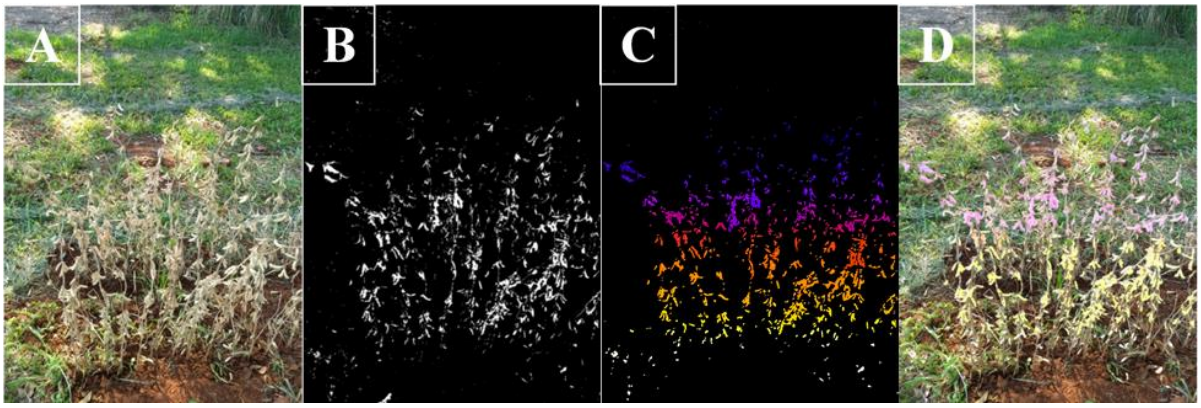

Figure 12. Soybean sample image captured from the side view in relation to the replanted line. (A) Original image from soybean plants at R8 stage without leaves. (B) Transformed image. (C) Surface object detection. (D) Overlaid image (A+C).

Figure 13 shows the front view of the sample in relation to the planted line. Figure 13A represents the original image captured. Figure 13B refers to the transformed image after the application of thresholding, limiarization and binarization. Figure 13C presents the identified objects after applying "3*D* Object Counter". Figure 13D represents the overlaid image from Figure 13A and Figure 13C. In Figure 13, spot lights and background grass are detected by the algorithm and considered as desired object. Note that these detected errors are common in image processing as seen on Gongal et al. (2015).

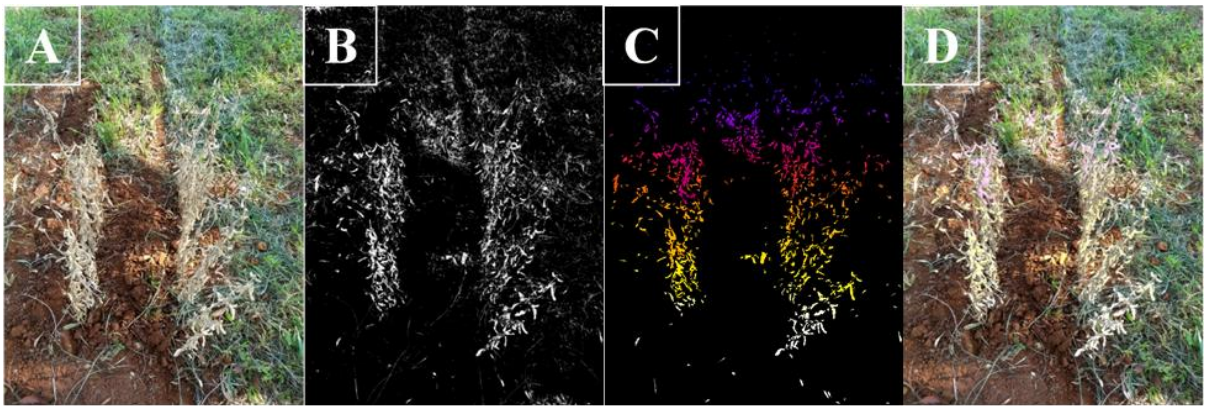

Figure 13. Soybean sample image captured from the front view in relation to the replanted line. (A) Original image from soybean plants at R8 stage without leaves. (B) Transformed image. (C) Surface object detection. (D) Overlaid image (A+C).

The upper view of the sample in relation to the planted line is shown in Figure 14. Figure 14A represents the original image captured. Figure 14B presents the transformed image. Figure 14C presents the identified objects after applying "3*D* Object Counter" tool. Figure 14D is the result of the overlaid image from Figure 14A and Figure 14C. The image scheme in Figure 14 limits the area of the desired target, diminishing the amount of undesired objects in the picture,

despite the presence of weeds and grass. Due to that, it is expected that the results from this view yield a lower absolute error than the views from Figures 12 and 13.

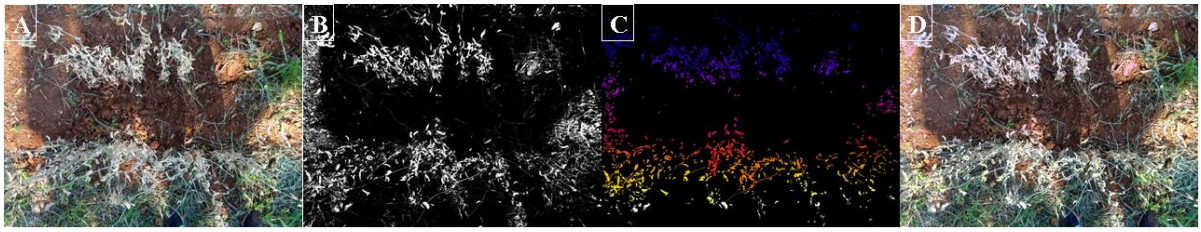

Figure 14. Soybean sample image captured from the upper view in relation to the replanted line. (A) Original image from soybean plants at R8 stage without leaves. (B) Transformed image. (C) Surface object detection. (D) Overlaid image (A+C).

The upper-side view of the sample in relation to the planted line is presented in Figure 15. Figure 15A represents the original image captured. Figure 15B is the transformed image after image basics processing. Figure 15C presents the identified objects after applying "3*D* Object Counter". The result of the overlaid image of Figure 15A and Figure 15C is shown in Figure 15D. As shown in Figure 14, it can be seen that mixing the upper and side view, the scheme becomes wider increasing the presence of more undesired objects in the picture. Aware of that, comparing just the Figures 12, 13, 14 and 15 without looking at the estimated number of pods and grains by view, it is expected that upper view presents the best accurate rate for both number of pods and grains, followed by upper-side view, side view and front view.

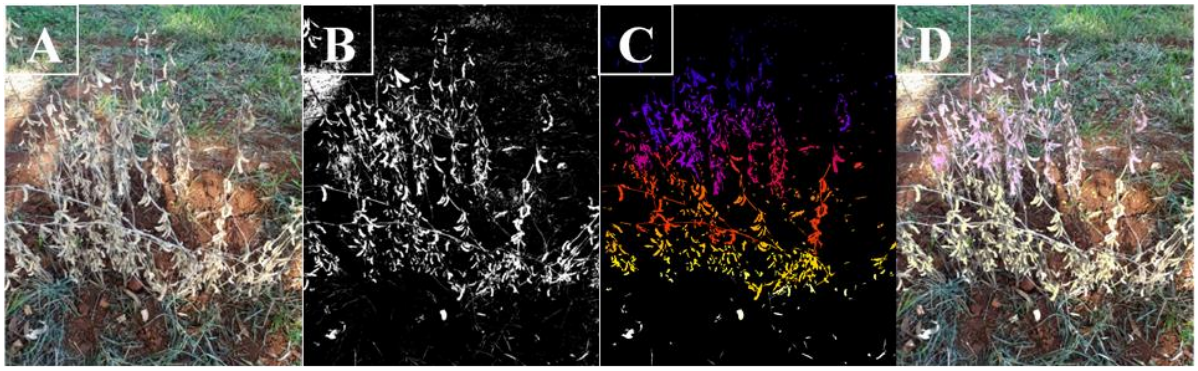

Figure 15. Soybean sample image captured from the upper-side view in relation to the replanted line. (A) Original image from soybean plants at R8 stage without leaves. (B) Transformed image. (C) Surface object detection. (D) Overlaid image (A+C). Source: adapted from Wei & Molin (2020).

The results from manual and estimated number of pods and number of grains per m^2 is shown in Table 3. Reference value is considered the manual value for both number of pod and number of grains. Table 3 represents the absolute relative error among estimated number of

pods and manually counted pods. It shows that there are large error percentages for all the views.

Aware that the algorithm do not perform object division, number of grains was obtained dividing the object surface area by the known area of one grain. A similar approach was made for rice crops (Reza et al., 2019), but in their case they predicted rice yield based on its grain area applying a K-means clustering algorithm and reached to R^2 values of 0.98 when comparing to the ground truth. Note that in this stage, the study proposes a method to obtain the number of soybean grains and not soybean yield.

The estimated number of grains by different views are shown in Table 3 shows that the lowest absolute relative error (2.76%) was found at the upper view as expected. But, it was also expected that side view yields higher accuracy rate followed by upper-side and front view which was not accomplished. Side view presents lower absolute relative error rate (7.15%) than upper-side view (8.83%). Side view might performed better than upper-side view because the last had higher presence of occluded soybean pods. Note that upper-side view also presented the highest absolute relative error for number of pods. Overcome occluded target objects are still a challenge in image processing and it has been largely related in several studies (Aggelopoulou et al., 2011; Dorj et al., 2013; Payne et al., 2013; Roscher et al., 2014, Cortes et al., 2016).

Table 3. Results of the number of pods and number of grains counted and estimated.

View in relation to the soybean cultivated line	Number of pods			Number of grains		
	Manual Counting	Image Processing	Absolute Relative Error (%)	Manual Counting	Image Processing	Absolute Relative Error (%)
Side	1019	573	43,8	2572	2388	7,2
Front	1019	496	51,3	2572	2213	14,0
Upper	1019	534	47,6	2572	2643	2,8
Upper-side	1019	431	57,7	2572	2345	8,8

Different approaches have been used to estimate the number of target crop organs related to its yield, for example, apple (Aggelopoulou et al., 2011), rice (Reza et al., 2019), soybean (Uzal et al, 2018; Li et al., 2019) and wheat (Alharbi et al., 2018).

Uzal et al. (2018) applied convolutional neural network and support vector machines to forecast seed per pod from images of soybean seeds under controlled environment, resulting in R^2 of validation and test ranging from 0.90 to 0.95 and 0.50 to 0.86, respectively. Also under

controlled environmental conditions and restricted to soybean pods in the scheme, Li et al. (2019) applied convolutional neural network to estimate soybean seed, reaching MAE and MSE values of 13.21 and 17.62 number of seeds per image, respectively. Both studies are similar to the test that was conducted in the beginning of this study applying "3*D* Object Counter" to detect and count the number of soybean pods under a controlled enviroment (Figure 3). The difference among studies is that they applied advanced computer vision techniques and, in this study, it was used classic image processing. And as a result, it highlights the possibility to use either classic or advanced approaches to cope with image processing under controlled conditions.

However, when considering the field level, a non-controlled environmental condition, it is challenging to apply different image processing techniques to obtain these yield components data. Considering non-controlled environmental conditions and targeting a yield component such as grain, there are no studies conducted at the field level with soybean plants available. Despite that, there are studies conducted with rice crops aiming to forecast its yield based on its yield components. For example, Reza et al. (2019) estimated rice yield based on rice grain area, applying a graph-cut algorithm with K-means clustering in images obtained from an unmanned aerial vehicle and found a R^2 of 0.98 comparing the ground-truth and the proposed method.

This study presents the potential to adapt the application of an available algorithm that was first developed to other applications. Despite the ease of use, the distinction between soybean pods/grains from other soybean characteristics and undesired objects at R8 stages such as plant stem and spot lights are challenging. The presence of similar colors and not well described shapes among soybean characteristics make the application of computer vision techniques more difficult. However, despite its difficulties, after image transformation and application of filters, it was possible to improve the detection of the desired object, in this case, soybean yield components.

This dissertation presents linear models to predict soybean yield based on its components and makes an exploratory analysis on how to gather soybean component data at the field level. More studies should be conducted on the improvement of computer vision applications in agriculture to gather soybean components data at the field level considering that soybean yield can be accurately predicted based on its components.

5. CONCLUSION

This study presented a novel approach to predict soybean yield through linear regression models that was fitted on a database composed of soybean yield and yield components (number of grains and number of pods) data already published. Soybean yield model based on the number of grains presented the highest prediction accuracy rate, highlighting that soybean yield can be accurately predicted through its components.

It was not possible to acquire the point cloud data of soybean plant structure using the 3*D* sensor system (LiDAR-based application) at the field level, making impossible to generate and model the soybean plant at the R8 stage.

The use of soybean RGB image data at R8 phenological stage obtained from a consumer-grade camera and the application of a post-processing method based on free tools ("3*D* Object Counter" and ImageJ platform) proved to be a potential method that can provide soybean yield components data. Besides the suitability of the method, it is highlighted that the results from "3*D* Object Counter" are dependent on the quality of the previous image processing (image thresholding, filtering and cutting).

REFERENCES

Adamchuk, V.; Ji, W.; Viscarra Rossel, R.; Gebbers, R.; Tremblay, N. **Proximal Soil and Plant Sensing**. In:D.K. Shannon, D.E. Clay, N.R. Kitchen (Eds.), Precision Agriculture Basics. Madison: ASA, CSSA, and SSSA. 2018, p. 119 – 140.

Aggelopoulou, A. D.; Bochtis, D.; Fountas, S.; Swain, K. C.; Gemtos, T. A.; Nanos, G. D. (2011). Yield prediction in apple orchards based on image processing. ***Precision Agriculture***, v. 12, p. 448-56.

Al-Gaadi, K. A.; Hassaballa, A. A.; Tola, E.; Kayad, A. G.; Madugundu, R.; Alblewi, B.; Assiri, F. (2016). Prediction of potato crop yield using precision agriculture techniques. ***Plos One***, doi: https://doi.org/10.1371/journal.pone.0162219

Alharbi, N., Zhou, J., Wang, W. **Automatic counting of wheat spikes from wheat growth images**. In *7th international conference on pattern recognition applications and methods*. Funchal, Madeira, Portugal, 2018.

Andrade, R. G.; Rodrigues, C. A. G.; Sanches, I. D.; Torresan, F. E.; Quartaroll, C. F. (2013). Uso de técnicas de sensoriamento remoto na detecção de processos de degradação de pastagens (In Portuguese). ***Engenharia na Agricultura***, v. 21, p. 234-243.

Andrade, R. G.; Vicente, L. E.; Teixeira, A. H. de C.; Leivas, J. F; Victoria, D. de C.; Roggia, S.; Franchini, J. C. **Agricultura de Precisão** (In Portuguese). In: A. C de C. Bernardi; J de M. Naime; A. V. de Resende; L. H. Bassoi; R. Y. Inamasu (Eds). Agricultura de precisão: resultados de um novo olhar (In Portuguese). Brasília, DF: Embrapa, 2014. p. 239 - 245.

Aranguren, M.; Castellón, A.; Aizpurua, A. (2020). Wheat yield estimation with NDVI values using a proximal sensing tool. ***Remote Sensing***, v. 12, doi: 10.3390/rs12172749.

Associação Brasileira dos Produtores de Soja (APROSOJA). Uso da soja (In Portuguese). Available at:<http://aprosojabrasil.com.br/2014/sobre-a-soja/uso-da-soja/>. Accessed on 25 February 2018.

Assefa, Y.; Purcell, L. C.; Salmeron, M.; Naeve, S.; Casteel, S. N.; Kovács, P.; Archontoulis, S.; Licht, M.; Below, F.; Kandel, H.; Lindsey, L. E.; Gaska, J.; Conly, S.; Shapiro, C.; Orlowski, J. M.; Golden, B. R.; Kaur, G.; Singh, M.; Thelen, K.; Laurenz, R.; Davidson D.; Ciampitti, I. A. (2019). Assessing variation in US soybean seed composition (protein and oil). ***Frontiers in Plant Science***, v. 10, doi: 10.3389/fpls.2019.00298.

Azimi, M. S.; Daneshian, J.; Sayfzadeh, S.; Zare, S. (2013). Evaluation of amino acid and salicylic acid application on yield and growth of wheat under water deficit. ***International Journal of Agriculture and Crop Sciences***, v. 5, p. 709 - 712.

Baio, F. H. R.; da Silva, S. P.; Camolese, H. da S.; Neves, D. C. (2017). Financial analysis of the investment in precision agriculture techniques on cotton crop. ***Journal of the Brazilian Assocation of Agricultural Engineering***, v. 37, p. 838 - 847.

Batistella, M.; Andrade, R. G.; Bolfe, E. L.; VicTORIA, D. C.; Silva, G. B. S. (2011). Geotecnologias e gestão territorial da bovinocultura no Brasil (In Portuguese). ***Revista Brasileira de Zootecnia***, v. 40, p. 251- 260.

Betbeder, J.; Fieuzal, R.; Baup, F. (2016). Assimilation of LAI and dry biomass data from optical and SAR images into an agro-meteorological model to estimate soybean yield. ***IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing***, v. 9, p. 2540 - 2553.

Blackmore, B. S. & Moore, M. (1999). Remedial correction of yield map data. ***Precision. Agriculture***, v. 1, p. 53 – 66.

Board, J. E. (1987). Yield components related to seed yield in determinate soybean. ***Crop Science***, v. 27, p. 1296 – 1297.

Bolte, S. & Cordelieres, F. P. (2006). A guided tour into subcellular colocalization analysis in light microscopy. ***Journal of Microscopy***, v. 224, p. 213 – 232.

Borghi, E.; Crusciol, C. A. C.; Nascente, A. S.; Mateus, G. P.; Martins, P. O.; Costa, C. (2012). Effects of row spacing and intercrop on maize grain yield and forage production of palisade grass. ***Crop & Pasture Science***, v. 63, p. 1106 - 1113, doi: http://dx.doi.org/10.1071/CP12344.

Camicia, R. G. da M.; Maggi, M. F.; de Souza, E. G.; Bazzi, C. L.; Konopatzki, E. A.; Michelon, G. K.; Pinheiro, K. B. S. (2018). Productivity of soybean in management zones with application of different sowing densities. ***Ciência Rural***, v. 48, p. 532 - 541.

Carciochi, W.D.; Schwalbert, R.; Andrade, F.H.; Corassa, G.M.; Carter, P.; Gaspar, A.P.; Schmidt, J.; Ciampitti, I. A. (2019). Soybean seed yield response to plant density by yield environment in North America. ***Agronomy Journal***, v. 111, p. 1923 – 1932, doi:10.2134/agronj2018.10.0635.

Chen, G. & Wiatrak, P. (2010). Soybean development and yield are influenced by planting date and environmental conditions in the Southeastern Coastal Plain, United States. ***Agronomy Journal***, v. 102, p. 1731 - 1737.

Christenson, B.S.; Schapaugh, W.T.; An, N.; Price, K.P.; Prasad, V.; Fritz, A. K. (2016). Predicting soybean relative maturity and seed yield using canopy reflectance. ***Crop Science***, v. 56, p. 625 - 643. doi: 10.2135/cropsci2015.04.0237.

Coble, K. H.; Mishra, A. K.; Ferrell, S.; Griffin, T. 2018. Big data in agriculture: a challenge for the future. ***Applied Economic Perspective and Policy***, v. 40, p. 79 - 96.

Colaço, A. F.; Trevisan, R.G.; Molin, J.P.; Rossell-Polo, J.R.; Escolà, A. (2017). A method to obtain orange crop geometry information using a mobile terrestrial laser scanner and 3D modeling. ***Remote Sensing***, v. 9, n. 8, p. 763.

Colaço, A. F.; Pagliuca, L. G.; Romanelli, T. L.; Molin, J. P. 2020. Economic viability, energy and nutrient balances of site-specific fertilisation for citrus. ***Biosystems Engineering***, v. 200, p. 138 - 156.

Companhia Nacional de Abastecimento (CONAB). **A produtividade da soja: análise e perspectivas** (In Portuguese). Brasília: CONAB, 2017. (Compêndio de estudos CONAB, v10).

Cortes, D. F. M.; Catarina, R. S.; Barros, G. B. de A.; Arêdes, F. A. S.; Silveira, S. F.; Ferreguetti, G. A.; Ramos, H. C. C.; Viana, A. P.; Pereira, M. G. (2017). Model-assisted phenotyping by digital images in papaya breeding program. ***Scientia Agricola***, v. 74, p. 294 – 302, doi: https://doi.org/10.1590/1678-992x-2016-0134.

De Bruin, J. L. & Pedersen, P. (2008). Soybean seed yield response to planting date and seeding rate in the upper Midwest. ***Agronomy Journal***, v. 100, p. 696 – 703, doi:10.2134/agronj2007.0115.

Delgado, J. A.; Short Jr, N. M.; Roberts, D. P.; Vandenberg, B. (2019). Big data analysis for sustainable agriculture on a geospatial cloud framework. ***Frontiers in Sustainable Food Systems***, v. 3, doi: 10.3389/fsufs.2019.00054.

Dobermann, A.; Blackmore, B. S.; Cook, S.; Adamchuk, V. I. **Precision farming: challenges and future directions**. In *New Directions for a Diverse Planet. Proceeding of 4th International Crop Science*, Brisbrane, Australia. 2004, p. 1 – 19.

Dorj, U.; Malrey, L.; Sangsub, H. (2013). A comparative study on tangerine detection, counting and yield estimation algorithm. ***International Journal of Security and Its Applications***, v. 7, p. 405 - 412.

Earth Explorer. USGS science for a changing world. Available at: < https://earthexplorer.usgs.gov>. Accessed on 20 Jun 2020.

Egli, D.B. & Yu, Z. (1991). Crop growth rate and seeds per unit area in soybean. ***Crop Science***, v. 31, p. 439 – 442.

Egli, D. B. (1988). Plant density and soybean yield. ***Crop Science***, v. 28, p. 977 - 981.

Eskandari, H.; Zehtab-Salmasi, S.; Ghassemi-Golezani, K.; Gharineh, M. H. (2009). Effects of water limitation on grain and oil yields of sesame cultivars. ***Journal of Agriculture & Environment***, v. 7, p. 339 - 342.

Eugenio, F. C.; Gorhs, M.; Venancio, L. P.; Schuh, M.; Bottega, E. L.; Ruoso, R.; Schons, C.; Mallmann, C. L.; Badin, T. L.; Fernandes, P. (2020). Estimation of soybean yield from machine learning techniques and multispectral RPAS imagery. ***Remote Sensing Applications: Society and Environment***, v. 20, doi: https://doi.org/10.1016/j.rsase.2020.100397.

Fehr, W. R. & Caviness, C. E. (1977). Stages of soybean development. ***Special Report 87***. Available at: <http://lib.dr.iastate.edu/specialreports/87>. Accessed on 25 February 2018.

Feng, A.; Zhou, J.; Vories, E. D.; Sudduth, K. A.; Zhang, M. (2020). Yield estimation in cotton using UAV-based multi-sensor imagery. ***Biosystems Engineering***, v. 193, p. 101 - 114.

Ferguson, R. & Rundquist, D. **Remote Sensing for Site-Specific Crop Management**. In:D.K. Shannon, D.E. Clay, N.R. Kitchen (Eds.), Precision Agriculture Basics. Madison: ASA, CSSA, and SSSA. 2018, p. 103 – 118.

Fernandez-Gallego, J.A.; Kefauver, S. C.; Vatter, T.; Gutiérrez, N. A.; Nieto-Taladriz, M. T.; Araus, J. L. (2019). Low-cost assessment of grain yield in durum wheat using RGB images. ***European Journal of Agronomy***, v. 105, p. 146 - 156.

Fountas, S.; Aggelopoulou, K.; Gemtos, T. A. **Precision Agriculture**. In Supply Chain Management for Sustainable Food Networks. Chichester, UK: John Wiley & Sons, Ltd. 2016, p. 41 – 65, doi: http://doi.org/10.1002/9781118937495.ch2

Fulton, J.; Hawkins, E.; Taylor, R.; Franzen, A. **Yield Monitoring and Mapping**. In:D.K. Shannon, D.E. Clay, N.R. Kitchen (Eds.), Precision Agriculture Basics. Madison: ASA, CSSA, and SSSA. 2018, p. 63 – 78.

Fuzzo, D. F. S.; Carlson, T. N.; Kourgialas, N. N.; Petropoulos, G. P. (2020). Coupling remote sensing with a water balance model for soybean yield predictions over large areas. ***Earth Science Informatics***, v. 13, p. 345 - 359.

Gebbers, R. & Adamchuk, V. I. (2010). Precision agriculture and food security. ***Science***, v. 327, p. 828 - 831.

GitHUB. Built for developers. Available at: <https://github.com>. Accessed on 20 Jun 2020.

Gnädinger, F. & Schmidhalter, U. (2017). Digital Counts of Maize Plants by Unmanned Aerial Vehicles (UAVs). ***Remote Sensing***, v. 9, p. 544.

Gómez, D.; Salvador, P.; Sanz, J.; Casanova, J.L. (2019). Potato yield prediction using machine learning techniques and Sentinel 2 data. ***Remote Sensing***, v. 11, p. 1745 - 1762, doi: 10.3390/rs11151745.

Gong, A.; Yu, J.; He, Y.; Qiu, Z. (2013). Citrus yield estimation based on images processed by an Android mobile phone. ***Biosystems Engineering***, v. 115, p. 162 – 170.

Gongal, A.; Amatya, S.; Karkee, M.; Zhang, Q.; & Lewis, K. (2015). Sensors and systems for fruit detection and localization: A review. ***Computers and Electronics in Agriculture***, v. 116, p. 8–19.

Hufkens, K.; Melaas, E. K.; Mann, M. L.; Foster, T.; Ceballos, F.; Robles, M.; Kramer, B. (2019). Monitoring crop phenology using a smartphone based near-surface remote sensing approach. ***Agricultural and Forest Meteorology***, v. 265, p. 327 - 337.

Hung, C.; Underwood, J.; Nieto, J.; Sukkarieh, S. **A Feature Learning Based Approach for Automated Fruit Yield Estimation**. In: L. Mejias; P. Corke; J. Roberts (Eds.), Field and Service Robotics. Cham: Springer. 2015, p. 485 – 498.

Hussain, M.; Khan, M. A.; Khan, M. B.; Farooq, M.; Farooq, S. (2012). Boron application improves growth, yield and net economic return of rice. ***Rice Science***, v. 19, p. 259 - 262.

International Society of Precision Agriculture (ISPA). Available at: <https://www.ispag.org>. Accessed on 24 March 2020).

Kamilaris, A. & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: a survey. ***Computers and Electronics in Agriculture***, v. 147, p. 70 - 90.

Kamiaris, A.; Kartakoullis, A.; Prenafeta-Boldú, F. X. (2017). A review on the practice of big data analysis in agriculture. ***Computers and Electronics in Agriculture***, v. 143, p. 23 - 37.

Kandel, Y. R.; Wise, K. A.; Bradley, C. A.; Tenuta, A. U.; Mueller, D. S. (2016). Effect of planting date, seed treatment, cultivar on plant population, sudden death syndrome and yield of soybean. ***Plant Disease***, v. 100, p. 1735 - 1743.

Kazemi, E.; Naseri, R.; Karimi, Z.; Emami, T. (2012). Variability of grain yield and yield components of white bean (*Phaseolus vulgaris* L.) cultivars as affected by different plant density in Western Iran. ***American-Eurasian Journal of Agricultural & Environmental Sciences***, v. 12, p. 17 - 22.

Khan, M. S.; Semwal, M.; Sharma, A.; Verma, R. K. (2020). An artificial neural network model for estimating mentha crop biomass yield using Landsat 8 OLI. ***Precision Agriculture***, v. 21, p. 18 - 33, doi: https://doi.org/10.1007/s11110-019-09655-9.

Koirala, A.; Walsh, K. B.; Wang, Z.; McCarthy, C. (2019). Deep learning - method overview and review of use for fruit detection and yield estimation. ***Computers and Electronics in Agriculture***, v. 162, p. 219 - 234.

Kurtulmus, F.; Lee, W. S.; & Vardar, A. (2014). Immature peach detection in colour images acquired in natural illumination conditions using statistical classifiers and neural network. ***Precision Agriculture***, v. 15, p. 57 – 79.

Lai, Y. R.; Pringle, M. J.; Kopittke, P. M.; Menzies, N. W.; Orton, T. G.; Dang, Y. P. (2018). An empirical model for prediction of wheat yield, using time-integrated Landsat NDVI. ***International Journal of Applied Earth Observation and Geoinformation***, v. 72, p. 99 – 108.

Lake, L. Sadras, V. O. (2014). The critical period for yield determination in chickpea (*Cicer arietinum* L.). ***Field Crops Research***, v. 168, p. 1 - 7.

Leroux, L.; Castets, M.; Baron, C.; Escorihuela, M.; Bégué, A.; Seen, D. L. (2019). Maize yield estimation in West Africa from crop process-induced combinations of multi-domain remote sensing indices. ***European Journal of Agronomy***, v. 108, p. 11 - 26.

Leroux, C.; Jones, H.; Clenet, A.; Dreux, B.; Becu, M.; Tisseyre, B. (2018). A general method to filter out defective spatial observations from yield mapping datasets. ***Precision Agriculture***, v. 19, p. 789 – 808.

Li, M. & Chung, S. (2015). Special issue on precision agriculture. ***Computers and Electronics in Agriculture***, v. 112, doi: https://doi.org/10.1016/j.compag.2015.03.014.

Li, Y.; Jia, J.; Khattak, A.M.; Sun, S.; Gao, W.; Wang, M. (2019). Soybean seed counting based on pod image using two-column convolution neural network. ***IEEE Access***, doi: 10.1109/ACCESS.2019.2916931

Liakos, K.G.; Busato, P.; Moshou, D.; Pearson, S.; Bochtis, D. (2018). Machine learning in agriculture: a review. ***Sensors*** , v. 18, doi:10.3390/s18082674.

Lin, G.; Tang, Y.; Zou, X.; Cheng, J.; Xiong, J. (2020). Fruit detection in natural environment using partial shape matching and probabilistic Hough transform. ***Precision Agriculture***, v. 21, p. 160 - 177, doi: https://doi.org/10.100/s11119-019-09662-w.

Lin, G.; Tang, Y.; Zou, X.; Xiong, J.; Fang, Y. 2020. Color-, depth-, and shape-based 3D fruit detection. Precision Agriculture, v. 21, p. 1 - 17, doi: https://doi.org.10.1007/s11119-019-09654-w.

Lindsey, L. **Agronomic Crops Network: Estimating Soybean Yield**. 2020. Available at: <https://agcrops.osu.edu/newsletter/corn-newsletter/2015-26/estimating-soybean-yield>. Accessed on 28 Jan 2020.

Liu, K.; Ma, B. L.; Limin, L.; Chaohai, L. (2011). Nitrogen, phosphorus, and potassium nutrient effects on grain filling and yield of high-yielding summer corn. ***Journal of Plant Nutrition***, v. 34, p. 1516 - 1531.

Liu, X.; Jia, W.; Ruan, C.; Zhao, D.; Gu, Y.; Chen, W. (2018). The recognition of apple fruits in plastic bags based on block classification. ***Precision Agriculture***, v. 19, p. 735 – 749. doi:10.1007/s11119-017-9553-2

Maestrini, B. & Basso, B. (2018). Predicting spatial patterns of within-field crop yield variability. ***Field Crops Research***, v. 219, p. 106 – 112, doi: https://doi.org/10.1016/j.fcr.2018.01.028.

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***, v. 237, p. 11599 - 111619.

Malek-Mohammadi, M.; Maleki, A.; Siaddat, S. A.; Beigzade, M. (2013). The effect of zinc and potassium on the quality yield of wheat under drought stress conditions. International ***Journal of Agriculture and Crop Sciences***, v. 6, p. 1164 - 1170.

Malik, M. F. A.; Awan, S. I.; Niaz, S. (2008). Comparative study of quantitative traits and association of yield and its components in black gram (*Vigna mungo*) genotypes. ***Asian Journal of Plant Sciences***, v. 7, p. 26 - 29.

Marko, O.; Brdar, S.; Panic, M.; Lugonja, P.; Crnojevic, V. (2016). Soybean varieties portfolio optimisation based on yield prediction. ***Computers and Electronics in Agriculture***, v. 127, p. 467 – 474. doi: 10.1016/j.compag.2016.07.009.

Mavridou, E.; Vrochidou, E.; Papakostas, G. A.; Pachidis, T.; Kaburlasos, V. G. (2019). Machine vision systems in precision agriculture for crop farming. ***Journal of Imaging***, v. 5, p. 89 - 121, doi: 10.3390/jimaging5120089.

Menegatti, L. A. A. & Molin, J. P. (2004). Remoção de erros em mapas de produtividade via filtragem de dados brutos (in Portuguese). ***Revista Brasileira de Engenharia Agrícola e Ambiental***, v. 8, p. 126 – 134.

Metternicht, G. **Remote sensing**. In: KEMP, K. (Ed.), Encyclopedia of Geographic Information Science. Sage Publication, New York, 2008, p. 365 – 368.

Michelon, G. K.; de Menezes, P. L.; Bazzi, C. L.; Jasse, E. P.; Magalhães, P. S. G.; Borges, L. F. (2018). Artificial neural networks to estimate the productivity of soybeans and corn by chlorophyll readings. ***Journal of Plant Nutrition***, v. 41, p. 1285 - 1292, doi: 10.1080/01904167.2018.1447579.

Mohamadzadeh, M.; Siadat, S. A.; Norof, M. S.; Naseri, R. (2011). The effects of planting date and row spacing on yield, yield components and associated traits in winter safflower under rain fed conditions. **American-Eurasian Journal of Agricultural & Environmental Sciences**, v. 10, p. 200 - 206.

Monzon, J. P.; Calviño, P. A.; Sadras, V. O.; Zubiaurre, J. B.; Andrade, F. H. (2018). Precision agriculture based on crop physiological principles improves whole-farm yield and profit: a case study. ***European Journal of Agronomy***, v. 99, p. 62 - 71.

Mourtzinis, S.; Gaspar, A. P.; Naeve, S. L.; Conley, S. P. (2017). Planting data, maturity, and temperature effects on soybean seed yield and composition. ***Agronomy Journal***, v. 109, doi: 10.2134/agronj2017.05.0247.

Mulla, D. J. & Schepers, J. S. **Key process and properties for site-specific soil and crop management**. In: F. J. Pierce & E. J. Sadler (Ed.) The state of site-specific management for agriculture. Madison: ASA: CSSA: SSSA, 1997. p. 1-18.

Mulla, D. J. (2013). Twenty five years of remote sensing in precision agriculture: key advances and remaining knowledge gaps. ***Biosystems Engineering***, v. 114, p. 358 - 371.

Nascente, A. S. Crusciol, C. A. C.; Stone, L. F.; Cobucci, T. (2013). Upland rice yield as affected by previous summer crop rotation (soybean or upland rice) and glyphosate management on cover crops. ***Planta Daninhas***, v. 31, p. 147 - 155.

Nevavuori, P.; Narra, N.; Lipping, T. (2019). Crop yield prediction with deep convolutional neural networks. ***Computers and Electronics in Agriculture***, v. 163, p. 104859 - 104868.

Niedbala, G. (2019). Application of artificial neural networks for multi-criteria yield prediction of winter rapeseed. ***Sustainability***, v. 11, p. 533 - 545, doi:10.3390/su11020533.

Nijland, W.; De Jong, R.; De Jong, S. M.; Wulder, M. A.; Bater, C. W.; Coops, N. C. (2014). Monitoring plant condition and phenology using infrared sensitive consumer grade digital cameras. ***Agricultural and Forest Meteorology***, v. 184, p. 98 – 106, doi:10.1016/j.agrformet.2013.09.007.

National Oceanic and Atmospheric Administration (NOOA). Climate data online. Available at: <https://www.ncdc.noaa.gov/cdo-web/>. Accessed on 20 Jun 2020.

Nordby, D.; Bentsen, M.; Sunderland, K.; Greenberg, M. R. (2014). Pocket Guide to Crop Development: Illustrated Growth Timelines for Corn, Sorghum, Soybean, and Wheat. University of Illinois Extension. Available at: < http://weeds.cropsci.illinois.edu/extension/Other/POCKETcrop.pdf>. Accessed on 25 February 2018.

Pantazi, X. E.; Moshou, D.; Alexandridis, T.; Whetton, R. L.; Mouazen, A. M. (2016). Wheat yield prediction using machine learning and advanced sensing techniques. ***Computers and Electronics in Agriculture***, v. 121, p. 57 - 65.

Patrício, D.I.; Rieder, R. (2018). Computer vision and artificial intelligence in precision agriculture for grain crops: A systematic review. ***Computers and Electronics in Agriculture***, v. 153, p. 69 – 81.

Payne, A.B.; Walsh, K.B.; Subedi, P.P.; Darvis, P.P. (2013). Estimation of mango crop yield using image analysis: segmentation method. ***Computers and Electronics in Agriculture***, v. 91, p. 57 - 64.

Pierce, F. J.; Robert, P. C.; Mangold, G. **Site-specific management: The pros, the cons, and the realities**. In *Proceedings of the International Crop Management Conference*, Iowa State University, Iowa. Press, Ames. 1994, p. 17 – 21.

Qureshi, W. S.; Payne, A.; Walsh, K. B.; Linker, R.; Cohen, O.; Dailey, M. N. (2017). Machine vision for counting fruit on mango tree canopies. ***Precision Agriculture***, v. 18, p. 224 – 244, doi: https://doi.org/10.1007/s11119-016-9458-5

R Core Team. 2019. R: A language and environment for statistical computing (Version 3.6.0). Vienna, Austria: R Foundation for Statistical Computing.

Rahnemoonfar, M. & Sheppard, C. (2017). Deep count: fruit counting based on deep simulated learning. ***Sensors***, v. 17, https://doi.org/10.3390/s17040905.

Rehman, A. & Farooq, M. (2016). Zinc seed coating improves the growth, grain yield and grain biofortification of bread wheat. ***Acta Physiologiae Plantarum***, v. 38, doi: 10.1007/s11738-016-2250-3.

Reza, M. N.; Na, I. S.; Baek, S. W.; Lee, K.H. (2019). Rice yield estimation based on K-means clustering with graph-cut segmentation using low-altitude UAV images. ***Biosystems Engineering***, v. 177, p. 109 – 121.

Richardson, A.; Braswell, B.; Hollinger, D.; Jenkins, J.; Ollinger, S. (2009). Near-Surface Remote Sensing of Spatial and Temporal Variation in Canopy Phenology. ***Ecological Applications***, v. 19, p. 1417 – 1428.

Robinson, A. P.; Conley, S. P.; Volenec, J. J.; Santini, J.B. (2009). Analysis of high yielding, early-planted soybean in Indiana. ***Agronomy Journal***, v. 101, p. 131 – 139. doi:10.2134/agronj2008.0014x.

Roscher, R.; Herzog, K.; Kunkel, A.; Kicherer, A.; Töpfer, R.; Förstner, W. (2014). Automated image analysis framework for high-throughput determination of grapevine berry sizes using conditional random fields. ***Computers and Electronics in Agriculture***, v. 100, p. 148 - 158.

Rosell, J. R. & Sanz, R. (2012). A review of methods and applications of the geomtric characterization of tree crops in agricultural activities. ***Computers and Electronics in Agriculture***, v. 81, p. 124 - 141.

Sa, I.; Ge, Z.; Dayoub, F.; Upcroft, B.; Perez, T.; Mccool, C. (2016). DeepFruits: A Fruit Detection System Using Deep Neural Networks. ***Sensors***, v. 16, doi:10.3390/s16081222.

Sabanci, K.; Toktas, A.; Kayabasi, A. (2017). Grain classifier with computer vision using adaptive neuro-fuzzy inference system. ***Journal of the Science of Food and Agriculture***, v. 97, p. 3994 - 4000.

Sakamoto, T. (2020). Incorporating environmental variables into a MODIS-based crop yield estimation method for United States corn and soybeans through the use of a random forest regression algorithm. ***ISPRS Journal of Photogrammetry and Remote Sensing***, v. 160, p. 208 - 228.

Sandaña, P. & Calderini. (2012). Comparative assessment of the critical period for grain yield determination of narrow-leafed lupin and pea. ***European Journal of Agronomy***, v. 40, p. 94 - 101.

Schwalbert, R. A.; Amado, T.; Corassa, G.; Pott, L. P.; Prasad, P. V. V.; Ciampitti, I. A. (2020). Satellite-based soybean yield forecast: integrating machine learning and weather data for improving crop yield prediction in southern Brazil. ***Agricultural and Forest Meteorology***, v. 284, p. 107886 - 107895.

Sedghi, M.; Sharifi, R. S.; Namvar, A.; Khandan-e-Bejandi, T,; Molaei, P. (2008). Responses of sunflower yield and grain filling period to plant density and weed interference. ***Research Journal of Biological Sciences***, v. 3, p. 1048 - 1053.

Sentelhas, P.C.; Battisti, R.; Câmara, G. M. S.; Farias, J. R. B.; Hampf, A. C.; Nendel, C. (2015). The soybean yield gap in Brazil - magnitude, causes and possible solutions for sustainable production. ***Journal of Agricultural Science***, v. 153, p. 1394 - 1411.

Setiyono, T.D.; Quicho, E.D.; Gatti, L.; Campos-Taberner, M.; Busetto, L.; Collivignarelli, F.; Garcia-Haro, F. X.; Boschetti, M.; Khan, N. I.; Holecz, F. (2018). Spatial rice yield estimation based on MODIS and Sentinel-1 SAR data and ORYZA crop growth model. ***Remote Sensing***, v. 10, doi: https://doi.org/10.3390/rs10020293.

Shanahan, J. F.; Schepers, J. S.; Francis, D. D.; Varvel, G. E.; Wilhelm, W. W.; Tringe, J. S.; Schlemmer, M. R.; Major, D. J. (2001). Use of remote sensing imagery to estimate corn yield. ***Agronomy Journal***, v. 93, p. 583 - 589.

Silva, C. B.; do Vale, S. M. L. R.; Pinto, F. A. C.; Muller, C. A. S.; Moura, A. D. (2007). The economic feasibility of precision agriculture in Mato Grosso do Sul State, Brazil: a case study. ***Precision Agriculture***, v. 8, p. 255 - 265.

Smidt, E.R.; Conley, S.P.; Zhu, J.; Arriage, F.J. (2016). Identifying field attributes that predict soybean yield using random forest analysis. ***Agronomy Journal***, v. 108, p. 637 - 646. doi: 10.2134/agronj2015.0222

Stafford, J. V. **Essential technology for precision agriculture**. In P. C. Robert; R. H. Rust; W. E. Larson (Eds.), Precision Agriculture. Proceedings of the third International Conference on Precision Agriculture. Minneapolis, USA: American Society of Agronomy, Crop Science Society or America, Soil Science Society of America. 1996, p. 595–604

Sudduth, K. A. & Drummond, S. T. (2007). Yield editor: software for removing errors from crop yield maps. ***Agronomy Journal***, v. 99, p. 1471 - 1482.

Syngenta (2020). Syngenta Crop Challenge In Analytics. Available online at: <https://www.ideaconnection.com/syngenta-crop-challenge/challenge.php>. Accessed on 25 Jun 2020.

ten Harkel, J.; Bartholomeus, H.; Kooistra, L. (2019). Biomass and crop height estimation of different crops using UAV-Based LiDAR. ***Remote Sensing***, v. 12, doi: 10.3390/rs12010017.

The National Institute of Health (2018). Image Processing and Analysis in Java. Available at: <https://imagej.nih.gov/ij/>. Accessed on 20 March 2018.

The National Research Council. 1997. **Precision Agriculture in the 21st Century**, Washington, DC: National Academy Press. 150p

Tian, H.; Wang, T.; Liu, Y.; Qiao, X.; Li, Y. (2019). Computer vision technology in agricultural automation - a review. ***Information Processing in Agriculture***, v. 7, p. 1 - 19.

Todeschini, M. H.; Milioli, A. S.; Rosa, A. C.; Dallacorte, L. V.; Panho, M. C.; Marchese, J. A.; Benin, G. (2019). Soybean genetic progress in South Brazil: physiological, phenological and agronomic traits. ***Euphytica***, v. 215, doi: https://doi.org/10.1007/s10681-019-2439-9.

Tran, C. C.; Nguyen, D. T.; Le, H. D.; Truong, Q. B. **Automatic dragon fruit counting using adaptive threshold for image segmentation and shape analysis**. In *4th NAFOSTED Conference on Information and Computer Science*, Hanoi, Vietnam.2017, p. 132 – 137, doi: 10.1109/NAFOSTED.2017.8108052.

United States Department of Agriculture (USDA). Foreign Agricultural Service: Soybeans. Available at: <https://www.fas.usda.gov/commodities/soybeans>. . Accessed on 5 July 2020.

Uzal, L.C.; Grinblat, G.L.; Namías, R.; Larese, M.G.; Bianchi, J.S.; Morandi, E.N.; et al. (2018). Seed-per-pod estimation for plant breeding using deep learning. ***Computers and Electronics in Agriculture***, v. 150, p. 196 – 204.

Vega, A.; Córdoba, M.; Castro-Franco, M.; & Balzarini, M. (2019). Protocol for automating error removal from yield maps. ***Precision Agriculture***, v. 19, p. 1 – 15, doi: https://doi.org/10.1007/s11119-018-09632-8.

Verma, A.; Jatain, A.; Bajaj, S. (2018). Crop yield prediction of wheat using fuzzy C Means clustering and neural network. ***International Journal of Applied Engineering Research***, v. 13, p. 9816 - 9821.

Viscarra Rossel, R.; Adamchuk, V.; Sudduth, K. A.; McKenzie, N. J.; Lobsey, C. **Proximal Soil Sensing: An Effective Approach for Soil Measurements in Space and Time**. In: D.L. Sparks, Advances in Agronomy. Academic Press, 2011, v. 113, p. 237 – 283.

Wachowiak, M. P.; Walters, D. F.; Kovacs, J. M.; Wachowiak-Smolíková, R.; James, A. L. (2017). Visual analytics and remote sensing imagery to support community-based research for precision agriculture in emerging areas. ***Computers and Electronics in Agriculture***, v. 143, p. 149 - 164.

Wang, A. X.; Tran, C.; Desai, N.; Lobell, D.; Ermon, S. **Deep transfer learning for crop yield prediction with remote sensing data**. In *COMPASS '18: ACM SIGCAS Conference on Computing and Sustainable Socieites (COMPASS)*, Menlo Park and San Jose, CA, USA, ACM, New York, USA, June 20 - 22, 2018.

Weersink, A.; Fraser, E.; Pannell, D.; Duncan, E.; Rotz, S. (2018). Opportunities and challenges for big data in agricultural and environmental analysis. ***Annual review of resource economics***, v. 10, p. 19 – 37, doi: https://doi.org/10.1146/annurev-resource-100516-053654.

Wei, M. C. F. & Molin, J. P. (2020). Soybean yield estimation and its components: a linear regression approach. ***Agriculture***, v. 10, p. 348 - 360, doi: 10.3390/agriculture10080348.

Wei, M. C. F.; Maldaner, L. F.; Ottoni, P. M. N.; Molin, J. P. (2020). Carrot yield mapping: a precision agriculture approach based on machine learning. ***Artificial Intelligence***, v. 1, p. 229 - 241, doi: 10.3390/ai1020015.

Wijewardana, C.; Reddy, K. R.; Bellaloui. 2019. Soybean seed physiology, quality and chemical composition under soil moisture stress. Food Chemistry, v. 278, p. 92 - 100.

Wolfert, S.l Ge, L.; Verdouw, C.; Bogaardt, M. (2017). Big data in smart farming - a review. ***Agricultural Systems***, v. 153, p. 69 - 80.

Yang, Q.; Shi, L.; Han, J.; Zha, Y.; Zhu, P. (2019). Deep convolutional neural networks for rice grain yield estimation at the ripening stage using UAV-based remotely sensed images. ***Field Crops Research***, v. 235, p. 142 - 153.

Yost, M. A.; Kitchen, N. R.; Sudduth, K. A.; Sadler, E. J.; Drummond, S. T.; Volkmann, M. R. (2017). Long-term impact of a precision agriculture system on grain crop production. ***Precision Agriculture***, v. 18, p. 823 - 842, doi: 10.1007/s11119-016-9490-5.

Yu, N.; Li, L.; Schmitz, N.; Tian, L. F.; Greenberg, J. A.; Diers, B. W. (2016). Development of methods to improve soybean yield estimation and predict plant maturity with an unmanned aerial vehicle based platform. ***Remote Sensing of Environment***, v. 187, p. 91 - 101.

Zafaranieh, M. (2015). Effect of various combinations of safflower and chickpea intercropping on yield and yield components of safflower. ***Agriculture Science Developments***, v. 4, p. 31 - 34.

Zhang, C. & Kovacs, J. M. (2012). The application of small unmanned aerial systems for precision agriculture: A review. ***Precision Agriculture***, v. 13, p. 693 – 712, doi:10.1007/s11119-012-9274-5.

Zhang, N.; Wang, M.; Wang, N. (2002). Precision agriculture - a worldwide overview. ***Computers and Electronics in Agriculture***, v. 36, p. 113 - 132.

Zhang, J.; Yang, C.; Zhao, B.; Song, H.; Hoffmann, W. C.; Shi, Y. (2017). Crop classification and LAI estimation using original and resolution-reduced images from two consumer-grade cameras. ***Remote Sensing***, v. 9, doi:10.3390/rs9101054.

ANNEX

Annex A. PUBLISHED PAPER AT THE JOURNAL "AGRICULTURE" (WEI, M. C. F. & MOLIN, J. P. 2020. Soybean yield estimation: a linear regression approach. Agriculture, v. 10, p. 348 – 361, doi:10.3390/agriculture1008034

agriculture

MDPI

Article

Soybean Yield Estimation and Its Components: A Linear Regression Approach

Marcelo Chan Fu Wei and José Paulo Molin *

College of Agriculture "Luiz de Queiroz", University of Sao Paulo, 11 Padua Dias Avenue, Piracicaba 13418-900, Brazil; marcelochan@usp.br

* Correspondence: jpmolin@usp.br

Received: 21 June 2020; Accepted: 16 July 2020; Published: 11 August 2020

check for updates

Abstract: Soybean yield estimation is either based on yield monitors or agro-meteorological and satellite imagery data, but they present several limiting factors regarding on-farm decision level. Aware that machine learning approaches have been largely applied to estimate soybean yield and the availability of data regarding soybean yield and its components (number of grains (NG) and thousand grains weight (TGW)), there is an opportunity to study their relationships. The objective was to explore the relationships between soybean yield and its components, generate equations to estimate yield and evaluate its prediction accuracy. The training dataset was composed of soybean yield and its components' data from 2010 to 2019. Linear regression models based on NG, TGW and yield were fitted on the training dataset and applied to a validation dataset composed of 58 on-field collected samples. It was found that globally TGW and NG presented weak (r = 0.50) and strong (r = 0.92) linear relationships with yield, respectively. In addition to that, applying the fitted models to the validation dataset, model based on NG presented the highest accuracy, coefficient of determination (R^2) of 0.70, mean absolute error (MAE) of 639.99 kg ha^{-1} and root mean squared error (RMSE) of 726.67 kg ha^{-1}.

Keywords: hundred grains weight; machine learning; number of grains; precision agriculture; thousand grains weight

1. Introduction

Yield is a quantitative measurement of the crop and it is an important feature that can benefit decision makers by supporting and improving their crop management [1]. The common approach to estimate grain yield with higher spatial resolution providing farm level data is using harvesters coupled to yield monitors [2]. Yield is computed through the convergence of yield monitor flow, global navigation satellite system (GNSS) receiver and moisture sensor data [3]. Grain yield monitor estimation accuracy is affected by many factors such as: (i) mass flow sensor, (ii) cleanness of the system, (iii) wiring harnesses and (iv) moisture sensor. Hence, constant inspection of the sensor system and calibration is required before the start of harvesting to achieve higher accuracy rates and obtain reliable data of estimated yield [3].

Other approaches have also been developed to estimate grain yield, but they have a lower spatial resolution, limiting its application on a farm level. Among these approaches, the most common is the application of remote sensing techniques based on agro-meteorological [4] and satellite imagery data [5,6]. Regarding this kind of data, efforts have also been made applying advanced algorithms to estimate yield [7].

Worldwide, researchers are making efforts to apply machine learning (ML) techniques in different databases to estimate crop yield. Most of these databases are composed of spectral band/vegetation indices from crop canopies [8,9], chlorophyll index [9], fusion of chemical and physical soil properties,

historical weather conditions and historical crop yield [10]. For example, Syngenta Crop Challenge [11] supplied a database with 2267 experimental corn hybrids planted in different locations across Canada and the United States between 2008 and 2016. Using that database to feed a deep learning neural network, the corn hybrid yield of 2017 was successfully predicted with a 12% root mean squared error compared to the average yield [12].

However, to support on-farm decision making, it is necessary to apply methods to gather higher spatial resolution data and the application of new sensor technologies and remote sensing techniques relying on the concepts of high-throughput phenotyping research could be suitable for this situation [3]. In this sense, some researchers have already applied ML techniques to analyze yield and its components. For example, Romero et al. [13] predicted durum wheat yield based on its plant height, peduncle length, spikelet number per ear, grain number per ear and grain weight per ear applying different ML decision trees. Aggelopoulou et al. [14] predicted apple tree yield based on its yield components, in their case, full bloom. The prediction was made by applying an image processing-based algorithm which learned texture patterns and correlated them to its yield.

Soybean yield components can be affected by several factors such: shading and temperature [15], genotype and environmental conditions [16], seed vigor [17] and seeding date, environment and cultivar [18]. Despite that, yield prediction through its components is becoming more popular in crops whose yield component is the product to be marketed, for example, mango [19] and rice [20], highlighting the possibility of applying it to soybean crops since the final product is grain.

Therefore, due to the ability to predict crop yield through its components (cotton [21] and rice [20]), large application of ML techniques in agriculture and soybean yield and its components' data availability, there is an opportunity to investigate the possibility to forecast soybean yield based on its components in a global scale, unlike the common approach which relies on vegetation indices. In this sense, this paper aimed to (i) explore soybean yield and yield components' relations from data published around the world, (ii) generate a global equation to estimate soybean yield and (iii) evaluate linear regressions prediction accuracy.

2. Materials and Methods

The training dataset was composed of published data from 2010 to 2019 that contained soybean yield and thousand grains weight (TGW) or hundred grains weight (HGW). In cases where HGW data was available instead of TGW, HGW values were multiplied by 10 to convert them to TGW. It was decided to use the training dataset relying on available data because it allowed the analysis of soybean yield and its components' relations on a global perspective since the dataset came from different locations of study, treatments, soybean varieties and year of study.

Based on literature reviews [22,23], Equation (1) was used to estimate the number of soybean grains (grains m^{-2}) because normally this data is not available and it is known that soybean yield (kg ha^{-1}) is a function of the number of grains in a certain area (grains m^{-2}) and the average weight per grain [22], in this case, it was adapted to thousand grains weight (g 1000 $grains^{-1}$).

$$\mathrm{NG}\left(\text{grains m}^{-2}\right) = \frac{100 \times \mathrm{Yield}\left(\text{kg ha}^{-1}\right)}{\mathrm{TGW}\left(\text{g 1000 grains}^{-1}\right)} \quad (1)$$

where, NG is the number of grains (grains m^{-2}) and TGW is the weight of a thousand grains (g 1000 $grains^{-1}$).

A descriptive analysis (number of observations, minimum, median, mean and maximum values, standard deviation, sample variance and coefficient of variation) and scatterplot (yield versus country, yield versus TGW and yield versus NG) were applied to the training dataset to provide better data visualization and inference. In addition to the scatterplots, Pearson's correlation (r) between soybean yield and its components was also calculated.

Linear regression models were used to predict yield because the number of observations was larger than the number of variables [24]. All yield observations, TGW and NG from the training dataset were used to fit three linear regression models: (A) Yield as function of TGW and NG, (B) yield as function of NG and (C) yield as function of TGW.

As linear regression allows formulation of the equation to predict the desired parameter, in our case, yield, it was also gathered from each model using the equation. Linear regression models generated from the training dataset were further applied to the validation dataset to predict soybean yield.

The validation dataset was composed of 58 randomly selected on-field samples of 1 m^2 from a soybean field (cultivar TMG 7062) collected at Piracicaba, Brazil (22.7356° S, 47.6479° W) in March of 2019. Number of grains (grains m^{-2}) was carefully and manually counted grain by grain, thousand grains weight (g 1000 grains^{-1}) and yield (kg ha^{-1}) were obtained from these samples after the drying process (105 °C for 72 h in a forced circulation air oven).

Overall performance of the fitted equations was evaluated by comparing their root mean squared error (RMSE), mean absolute error (MAE) and coefficient of determination (R^2) for both the training and validation datasets. All the statistical analysis was performed in R environment [25].

3. Results and Discussion

In 707 observations of soybean data (training dataset) it was found that they contained at least "yield" and "thousand grains weight" or "hundred grains weight" from 56 eligible papers [26–81] across the world, as presented in Table 1. A large difference between soybean yield and its yield components between countries is observed. Yield presented a range from 167.1 to 10,170 kg ha^{-1}. Different yield values found across the world highlight the regional variability inherent among environment and genotype. Furthermore, note that the coefficient of variation of yield and number of grains are higher than 50% and thousand grains weight is approximately 24%. These values indicate that TGW presents less variance than yield and NG in a worldwide range. The validation dataset, also shown in Table 1, presents variability among samples which is higher for both yield and NG than TGW when looking at the coefficient of variation.

Table 1. Descriptive analysis of the soybean training dataset obtained from [26–81] and validation dataset.

Measurement	Training Dataset			Validation Dataset		
	Yield (kg ha^{-1})	TGW (g 1000 Grains^{-1})	NG (Grains m^{-2})	Yield (kg ha^{-1})	TGW (g 1000 Grains^{-1})	NG (Grains m^{-2})
n	707.0	707.0	707.0	58.0	58.0	58.0
Minimum	167.1	64.7	100.0	2100.0	301.5	1236.0
Median	2819.0	148.4	1849.0	3460.0	423.3	1888.0
Mean	2764.9	148.0	1826.0	3463.8	420.4	1889.0
Maximum	10,170.0	346.0	5557.0	5400.0	542.9	2975.0
Standard Deviation	1557.6	35.9	932.0	736.6	44.6	444.0
Sample Variance	2,425,981.0	1291.6	868,565.2	542,588.9	1991.8	196,958.0
CV (%)	56.3	24.3	51.1	21.3	10.6	23.5

n = number of observations; CV = coefficient of variation; TGW = thousand grains weight and NG = number of grains.

Soybean yield variability among countries despite year, genotype and treatments can be seen in Figure 1. Figure 1A (worldwide soybean yield) shows soybean yield distribution among countries. Note that within country there are large yield differences representing yield variability at a country level which is expected, as we can see in several governmental reports, for example, in the US and Brazil, United States Department of Agriculture—USDA [82] and Brazilian Agricultural Research Corporation—EMBRAPA [83], respectively, that are responsible for forecasting and presenting harvested yield by location.

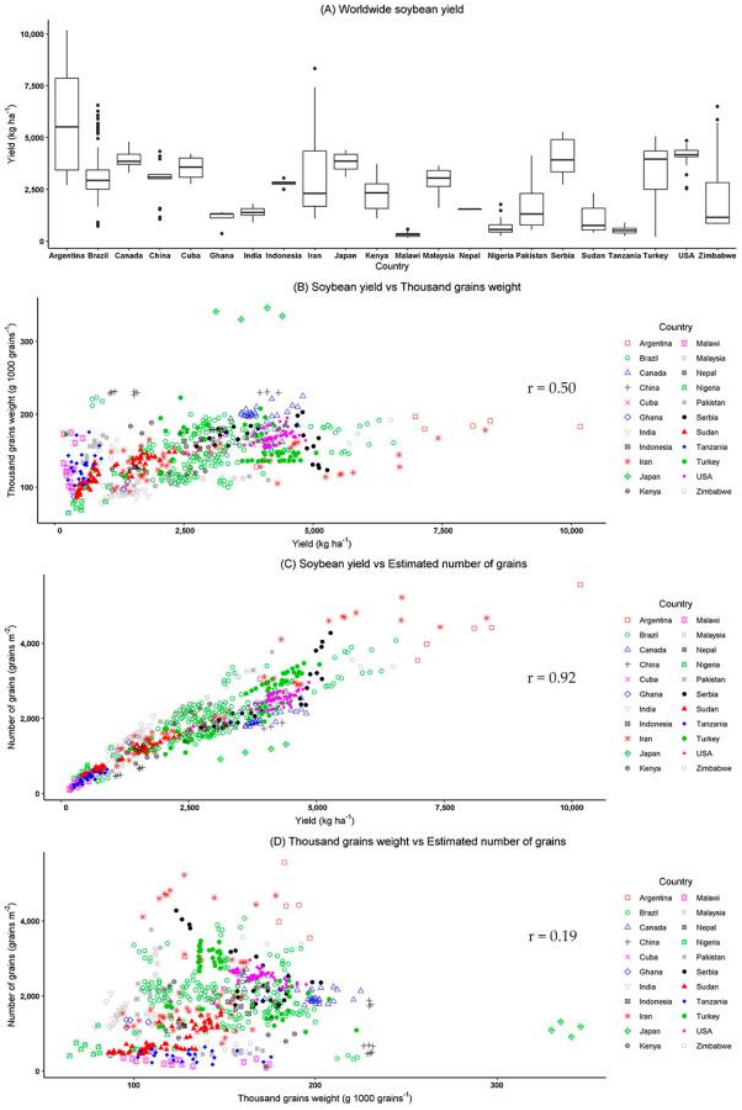

Figure 1. Boxplot and scatterplots of observed soybean yield (kg ha^{-1}) and its yield components by country. (**A**) Worldwide soybean yield (kg ha^{-1}) by country, (**B**) thousand grains weight (g 1000 grains^{-1}), (**C**) number of grains (grains m^{-2}) and (**D**) thousand grains weight (g 1000 grains^{-1}) versus number of grains (grains m^{-2}). r = correlation coefficient.

The graphic of Figure 1B (soybean yield versus TGW) shows that there is weak linear relationship (r = 0.50) between these two variables which will further negatively impact the linear regression model prediction accuracy. On the other hand, Figure 1C (soybean yield versus NG) presents a strong linear relationship between yield and NG (r = 0.92). Comparing TGW and NG data (Figure 1D), there is a weak linear relationship (r = 0.19) indicating that there is no collinearity between them.

Based on Figure 1, it shows an opportunity to further investigate the possibility of applying machine learning techniques to estimate soybean yield based on its components. Corroborating the idea that yield prediction is one of the most important topics in precision agriculture and that ML techniques have widely been applied in agriculture, this paper also applied ML techniques to forecast yield [84], but relying on the use of its components as predictor variables.

Yield is a function of the number of pods (np), grain number per pod (gnp) and grain weight (gw), as presented in Equation (2) adapted from Egli and Zhen-wen [22] and Lindsey [23]. Combining np and gnp, a new variable is generated, which is the number of grains (NG), providing Equation (3). Based on that, soybean yield is a direct function of NG and gw. Regarding these two variables related to yield, number of grains (grains m^{-2}) has more influence on yield than grain weight/thousand grains weight [34,85–87] which corroborates with the findings in Figure 1.

$$\text{Yield}\left(\text{kg ha}^{-1}\right) = \text{np}\left(\text{unit ha}^{-1}\right) \times \text{ngp}\left(\text{grain unit}^{-1}\right) \times \text{gw}\left(\text{kg grain}^{-1}\right) \quad (2)$$

where, np, ngp and gw are number of pods (unit ha^{-1}), number of grains per pod (grain $unit^{-1}$) and grain weight (kg $grain^{-1}$), respectively.

$$\text{Yield}\left(\text{kg ha}^{-1}\right) = \text{NG}\left(\text{grains ha}^{-1}\right) \times \text{gw}\left(\text{kg grain}^{-1}\right) \quad (3)$$

where, NG and gw are, respectively, number of grains (grain ha^{-1}) and grain weight (kg $grain^{-1}$).

Figure 2 presents the relation between observed and predicted soybean yields. The linear models were generated from the training dataset. Model A (Figure 2A) presents the highest R^2 (0.97) and lowest RMSE (288.78 kg ha^{-1}), model B (Figure 2B) presents R^2 of 0.85 and RMSE of 455.98 kg ha^{-1}, and model C (Figure 2C) presents the lowest R^2 (0.26) and the highest RMSE (1343.17 kg ha^{-1}), as expected due to the weak linear relationship between yield and TGW. Higher R^2 and lower MAE values are found for models A and B and not for model C because NG has more global influence than TGW to predict yield, as seen previously in Figure 1. Hence, it is expected that models A and B present a better overall performance compared to model C when applied to the validation dataset.

Results of the fitted models A, B and C application to the validation dataset is shown in Figure 3. R^2 values for models A (Figure 3A), B (Figure 3B) and C (Figure 3C) are 0.66, 0.70 and 0.02, respectively. Model B presents the highest R^2, lowest MAE (639.99 kg ha^{-1}) and RMSE (726.67 kg ha^{-1}) followed by model A (MAE—3420.93 kg ha^{-1} and RMSE—3449.80 kg ha^{-1}) and model C (MAE—5267.52 kg ha^{-1} and RMSE—5391.08 kg ha^{-1}). These results indicate that model B presents higher accuracy rate to predict soybean yield based on number of grains from data collected in a single field. Therefore, in this case, yield prediction based on NG can support on-farm decision making which is the scale of precision agriculture actions that deals with the variability within field levels [88]. Note that, to achieve a good accuracy rate, both predicted and observed yields should present, not only high R^2, low RMSE and MAE, but also, the amplitude ratio between predicted and observed yield values as close as possible to one (R^2 = 1.00), meaning that there was a perfect prediction without error and unexplained variance [89].

This paper presents a novel model to predict soybean yield based on its components and data published worldwide by applying linear regression models different from those published based on soil properties [90], historical data about soil, yield and soybean variety [91] and vegetation indices [92]. Model B, based on the number of grains per m^2 which presented the highest R^2, demonstrated its potential to predict soybean yield.

Currently, our soybean yield estimations are either based on agro-meteorological/crop models or based on yield monitors. Agro-meteorological/crop models are restricted by their spatial resolution and, most of the time, they do not represent farm-level support for on-farm decision making. On the other hand, yield monitors have a higher spatial resolution compared to agro-meteorological/crop models, but they are limited due to operational limitations and several error factors. These factors can be caused by different sources such miss calibration, sensor system errors and many others [3].

As stated by Patrício and Rieder [93], the application of computer vision, machine learning and high-performance computing together are considered as one of the main drivers to solving different problems in agriculture; we have to look for options considering these applications that will deliver better data and information to farmers to support and improve their decision making regarding their crop at farm level.

Efforts have been made in digital image processing techniques related to plant diseases [94] and reviews about computer vision systems to evaluate grain quality [95]. In this sense, there are several researchers working on grain quality and plant disease diagnosis by applying computer vision, which is expected since its evaluations are based on the visualization. Hence, being aware that there is a strong linear relation of number of grains per m^2 with soybean yield (r = 0.92) in a worldwide range and that computational processing techniques are becoming more efficient, it is highlighted that there is an opportunity to apply computer vision techniques to gather these data to estimate soybean yield instead of looking for indirect factors to estimate its yield.

The proposed model in this paper can be suitable for different approaches from on-farm yield sampling for genetic plot trials to punctual yield scouting. The implication of applying this model is to acquire the predictor variable (NG) which is not easy at the moment. However, there are researchers working on high-throughput phenotyping, specifically, estimating the number of grains per pod of soybean plants [96] and number of grains [97], both under controlled environmental conditions.

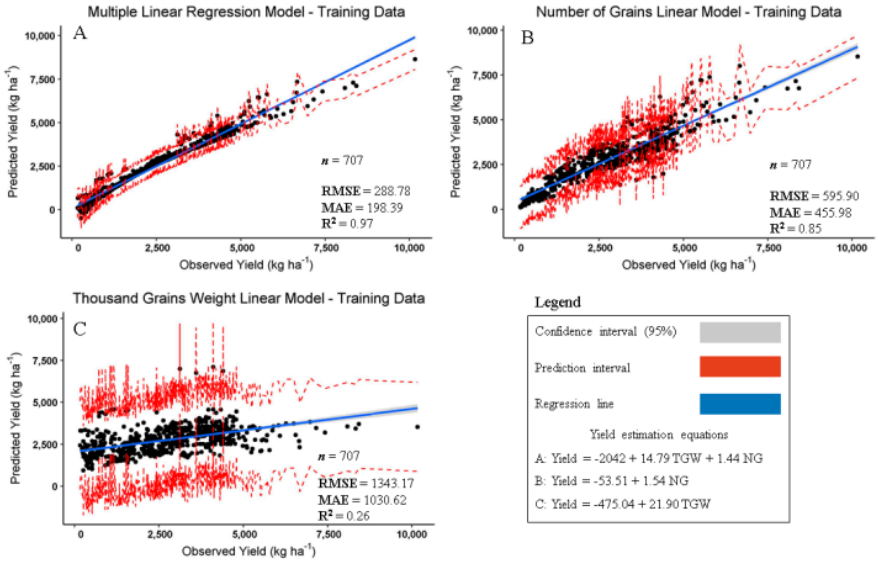

Figure 2. Scatterplot of soybean observed and predicted yields (kg ha^{-1}) applied to training dataset. (**A**) Yield estimation based on thousand grains weight (TGW in g 1000 $grains^{-1}$) and number of grains (NG in grain m^{-2}), (**B**) yield estimation as function of number of grains (NG in grains m^{-2}) and (**C**) yield estimation as functions of thousand grains weight (TGW in g 1000 $grains^{-1}$). *n* = number of observations; MAE = mean absolute error (kg ha^{-1}); RMSE = root mean squared error (kg ha^{-1}); R^2 = multiple R-squared; Yield = soybean yield (kg ha^{-1}); TGW = thousand grains weight (g 1000 $grains^{-1}$); NG = number of grains (grains m^{-2}).

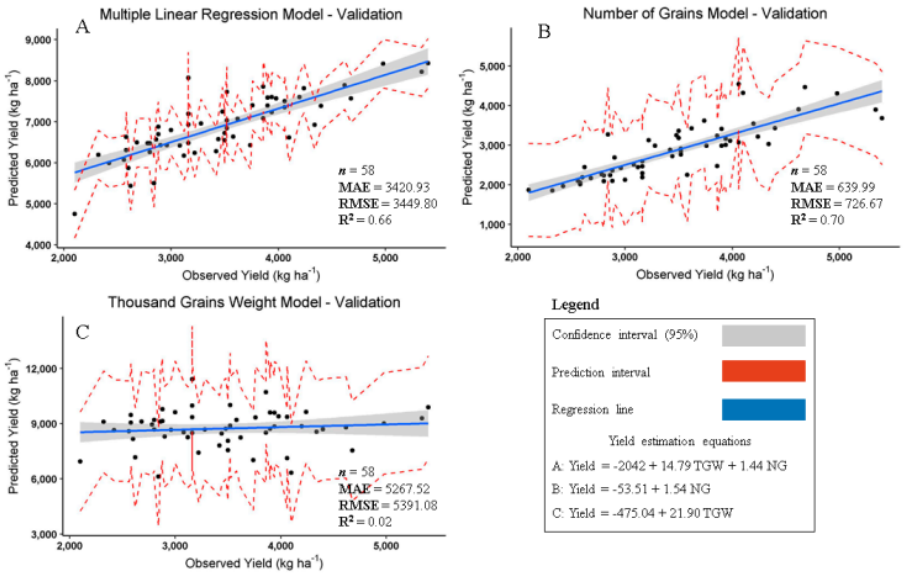

Figure 3. Scatterplot of soybean observed and predicted yields (kg ha^{-1}) applied to validation dataset. (**A**) Yield estimation based on thousand grains weight (TGW in g 1000 grains^{-1}) and number of grains (NG in grains m^{-2}), (**B**) yield estimation as function of number of grains (NG in grains m^{-2}) and (**C**) yield estimation as functions of thousand grains weight (TGW in g 1000 grains^{-1}). *n* = number of observations; MAE = mean absolute error (kg ha^{-1}); RMSE = root mean squared error (kg ha^{-1}); R^2 = multiple R-squared; Yield = soybean yield (kg ha^{-1}); TGW = thousand grains weight (g 1000 grains^{-1}); NG = number of grains (grains m^{-2}).

Uzal et al. [96] developed an approach based on convolutional neural networks and support vector machines to estimate seed per pod for soybean plants through image processing resulting in R^2 of validation and test ranging from 0.90 to 0.95 and 0.50 to 0.86, respectively. Li et al. [97] estimated soybean seed based on deep learning techniques, in their case, applying two convolution neural networks reaching MAE and MSE values of 13.21 and 17.62 number of seeds per image, respectively.

Considering non-controlled environmental conditions and targeting a yield component such as grain, Reza et al. [20] looking to estimate rice yield based on its components, in their case rice grain area, they applied a graph-cut algorithm with K-means clustering in images obtained from an unmanned aerial vehicle and found a coefficient of determination of 0.98 comparing the ground-truth and the proposed method. Thus, computer vision application in agriculture presents the potential to obtain yield components data, like rice grain which can be extrapolated to soybean grain.

Figure 4 represents an example of computer vision technique application to gather the number of grains of a sample. It shows the potential of applying computer vision techniques to obtain soybean number of grains. Note that in this paper our focus is not to demonstrate how to implement computer vision, but highlight that soybean yield can be accurately predicted through its components, in our suggested model by the number of grains per m^2, and one possible method to reach these components data is by applying computer vision techniques. Indeed, more research applying computer vision to acquire soybean number of grains is required, but we must rely on technological and computational advances in agriculture that have been improving in large steps.

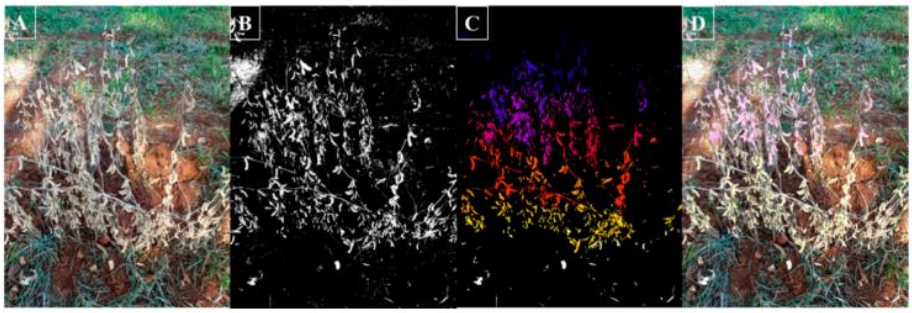

Figure 4. Example of computer vision technique application to gather soybean number of grains on-field situations. (**A**) Original soybean image, (**B**) soybean image transformation, (**C**) surface object detection image and (**D**) original soybean image overlaid with surface object detection image.

Being aware that yield monitors present several time consuming steps for calibration to improve their accuracy estimation [3] and that soybean yield can be accurately estimated by model B (based on the number of grains per m^2), efforts should be made aiming at achieving accurate numbers of grains in on-field situations without harvesting the crop to apply model B to forecast soybean yield. Therefore, in accordance that computer vision techniques are being applied to estimate crop components related to yield [98–101] and the example of Figure 4, it is highlighted that computer vision is a promising method that will support how soybean number of grains are gathered on-field in a fast, reliable and accurate way.

4. Conclusions

In this work, considering the dataset used, it can be concluded that globally soybean yield presents a strong linear relationship with number of grains even for different varieties, countries and treatments and presents a weak linear relationship with thousand grains weight. It was possible to estimate soybean yield based on two equations, one dependent on thousand grains weight and number of grains (model A) and another dependent only on the number of grains (model B). Model B, yield as function of number of grains, applied to the validation dataset presented the best prediction accuracy compared to models A and C, highest R^2 (0.70), lowest MAE (639.99 kg ha^{-1}) and RMSE (726.67 kg ha^{-1}).

Despite the suitability of fit of model B to predict soybean yield, it is worth noting that, currently, it is difficult and time-consuming to access NG data to use the proposed model. However, as the implementation of technology in agriculture advances, the proposed method corroborates towards the use and development of computer vision application to estimate crop yield, in this case, soybean yield.

Author Contributions: Conceptualization, M.C.F.W. and J.P.M.; methodology, M.C.F.W. and J.P.M.; validation, M.C.F.W. and J.P.M., investigation, M.C.F.W. and J.P.M.; writing—original draft preparation, M.C.F.W. and J.P.M.; writing—review and editing, M.C.F.W. and J.P.M.; visualization, M.C.F.W. and J.P.M.; supervision, J.P.M. All authors have read and agreed to the published version of the manuscript.

Funding: The authors are grateful to CAPES and CNPq (process number: 130968/2019-6) for the scholarship support to M.W. (Masters student, University of Sao Paulo) and to Fundação de Estudos Agrários Luiz de Queiroz (FEALQ) for supporting the publication fee.

Acknowledgments: We thank Horst Bremer Neto that allowed us to collect the soybean samples on-field and Ricardo Chan for proofreading the English manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Fritz, S.; See, L.; Bayas, J.C.L.; Waldner, F.; Jacques, D.; Becker-Reshef, I.; Rembold, F. A comparison of global agricultural monitoring systems and current gaps. *Agric. Syst.* **2019**, *168*, 258–272. [CrossRef]

2. Fulton, J.; Hawkins, E.; Taylor, R.; Franzen, A. Yield monitor data: Collection, management, and usage. *Crops Soils* **2018**, *51*, 4–51. [CrossRef]
3. Fulton, J.; Hawkins, E.; Taylor, R.; Franzen, A. Yield monitoring and mapping. In *Precision Agriculture Basics*; Shannon, D.K., Clay, D.E., Kitchen, N.R., Eds.; ASA, CSSA, and SSSA: Madison, WI, USA, 2018; pp. 63–78.
4. Betbeder, J.; Fieuzal, R.; Baup, F. Assimilation of LAI and dry biomass data from optical and SAR images into an agro-meteorological model to estimate soybean yield. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 2540–2553. [CrossRef]
5. Figueiredo, G.K.; Brunsell, N.A.; Higa, B.H.; Rocha, J.V.; Augusto, R.; Lamparelli, C. Correlation maps to assess soybean yield from EVI data in Paraná State, Brazil. *Sci. Agric.* **2016**, *73*, 462–470. [CrossRef]
6. Maestrini, B.; Basso, B. Predicting spatial patterns of within-field crop yield variability. *Field Crops Res.* **2018**, *219*, 106–112. [CrossRef]
7. You, J.; Li, X.; Low, M.; Lobell, D.; Ermon, S. Deep gaussian process for crop yield prediction based on remote sensing data. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017.
8. Al-Gaadi, K.A.; Hassaballa, A.A.; Tola, E.; Kayad, A.G.; Madugundu, R.; Alblewi, B.; Assiri, F. Prediction of potato crop yield using precision agriculture techniques. *PLoS ONE* **2016**, *11*, e0162219. [CrossRef]
9. Coelho, A.P.; De Faria, R.T.; Leal, F.T.; de Arruda Barbosa, J.; Dalri, A.B.; Rosalen, D.L. Estimation of irrigated oats yield using spectral indices. *Agri. Water Manag.* **2019**. [CrossRef]
10. Setiyono, T.D.; Quicho, E.D.; Gatti, L.; Campos-Taberner, M.; Busetto, L.; Collivignarelli, F.; García-Haro, F.J.; Boschetti, M.; Khan, N.I.; Holecz, F. Spatial rice yield estimation based on MODIS and Sentinel-1 SAR data and ORYZA crop growth model. *Remote Sens.* **2018**, *10*, 293. [CrossRef]
11. Syngenta. Syngenta Crop Challenge in Analytics. 2018. Available online: https://www.ideaconnection.com/syngenta-crop-challenge/challenge.php (accessed on 20 January 2020).
12. Khaki, S.; Wang, L. Crop yield prediction using deep neural networks. *Front. Plant Sci.* **2019**, *10*. [CrossRef]
13. Romero, J.R.; Roncallo, P.F.; Akkiraju, P.C.; Ponzoni, I.; Echenique, V.C.; Carballido, J.S. Using classification algorithms for predicting durum wheat yield in the province of Buenos Aires. *Comput. Electron. Agric.* **2013**, *96*, 173–179. [CrossRef]
14. Aggelopoulou, A.; Bochtis, D.; Fountas, S.; Swain, K.C.; Gemtos, T.; Nanos, G.D. Yield prediction in apple orchards based on image processing. *Precis. Agric.* **2011**, *12*, 448–456. [CrossRef]
15. Kurosaki, H.; Yumoto, S. Effects of low temperature and shading during flowering on the yield components in soybeans. *Plant Prod. Sci.* **2003**, *6*, 17–23. [CrossRef]
16. Hao, D.; Cheng, H.; Yin, Z.; Cui, S.; Zhang, D.; Wang, H.; Yu, D. Identification of single nucleotide polymorphisms and haplotypes associated with yield and yield components in soybean (*Glycine max*) landraces across multiple environments. *Theor. Appl. Genet.* **2012**, *124*, 447–458. [CrossRef]
17. Caverzan, A.; Giacomin, R.; Muller, M.; Biazus, C.; Langaro, N.D.; Chavarria, G. How does seed vigor affect soybean yield components? *Agron. J.* **2018**, *110*, 1318–1327. [CrossRef]
18. MacMillan, K.P.; Guiden, R.H. Effect of seeding date, environment and cultivar on soybean seed yield, yield components, and seed quality in the Northern Great Plains. *Agron. J.* **2020**, *112*, 1666–1678. [CrossRef]
19. Stein, M.; Bargoti, S.; Underwood, J. Image based mango fruit detection, localisation and yield estimation using multiple view geometry. *Sensors* **2016**, *16*, 1915. [CrossRef]
20. Reza, M.N.; Na, I.S.; Baek, S.W.; Lee, K.-H. Rice yield estimation based on K-means clustering with graph-cut segmentation using low-altitude UAV images. *Biosyst. Eng.* **2018**, *177*, 109–121. [CrossRef]
21. Tedesco-Oliveira, D.; da Silva, R.P.; Maldonado, W., Jr.; Zerbato, C. Convolutional neural networks in predicting cotton yield from images of commercial fields. *Comput. Electron. Agric.* **2020**, *171*, 105307. [CrossRef]
22. Egli, D.B.; Yu, Z. Crop growth rate and seeds per unit area in soybean. *Crop Sci.* **1991**, *31*, 439–442. [CrossRef]
23. Lindsey, L. Agronomic Crops Network: Estimating Soybean Yield. 2020. Available online: https://agcrops.osu.edu/newsletter/corn-newsletter/2015-26/estimating-soybean-yield (accessed on 28 January 2020).
24. Gromping, U. Variable importance assessment in regression: Linear regression versus random forest. *Am. Stat.* **2009**, *63*, 309–319. [CrossRef]
25. R Foundation for Statistical Computing. *R: A Language and Environment for Statistical Computing*, version 3.6.0; R Core Team: Vienna, Austria, 2019.

26. Abbasi, M.K.; Majeed, A.; Sadiq, A.; Khan, S.R. Application of *Bradyrhizobium japonicum* and phosphorus fertilization improved growth, yield and nodulation of soybean in the sub-humid hilly region of Azad Jammu and Kashmir, Pakistan. *Plant Prod. Sci.* **2008**, *11*, 368–376. [CrossRef]
27. Afzal, A.; Bano, A.; Fatima, M. Higher soybean yield by inoculation with N-fixing and P-solubilizing bacteria. *Agron. Sustain. Dev.* **2010**, *30*, 487–495. [CrossRef]
28. Almaz, M.G.; Halim, R.A.; Martini, M.Y. Effect of combined application of poultry manure and inorganic fertilizer on yield and yield components of maize intercropped with soybean. *Pertanika J. Trop. Agric. Sci.* **2017**, *40*, 173–184.
29. Ambrosini, V.G.; Fontoura, S.M.V.; De Moraes, R.P.; Tamagno, S.; Ciampitti, I.A.; Bayer, C. Soybean yield response to *Bradyrhizobium* strains in fields with inoculation history in Southern Brazil. *J. Plant Nutr.* **2019**, *42*, 1941–1951. [CrossRef]
30. Arslan, H.; Karakus, M.; Hatipoglu, H.; Arslan, D.; Bayraktar, O.V. Assessment of performances of yield and factors affecting the yield in some soybean varieties/lines grown under semi-arid climate conditions. *Appl. Ecol. Env. Res.* **2018**, *16*, 4289–4298. [CrossRef]
31. Belkheir, A.M.; Zhou, X.; Smith, D.L. Variability in yield and yield component responses to genistein pre-incubated *Bradyrhizobium japonicum* by soybean [*Glycine max* (L.) Merr] cultivars. *Plant Soil* **2001**, *229*, 41–46. [CrossRef]
32. Bertham, R.R.; Arifin, Z.; Nusantara, A.D. The improvement of yield and quality of soybeans in a coastal area using low input technology based on biofertilizers. *Int. J. Adv. Sci. Eng. Inf. Technol.* **2019**, *9*, 787–791. [CrossRef]
33. Bryant, C.J.; Krutz, L.J.; Nuti, R.C.; Truman, C.C.; Locke, M.A.; Falconer, R.; Atwill, L.; Wood, C.W.; Spencer, G.D. Furrow diking as Mid-Southern USA irrigation strategy: Soybean grain yield, irrigation water use efficiency, and net returns above furrow diking costs. *Crop. Forage Turfgrass Manag.* **2019**, *5*, 180076. [CrossRef]
34. Carciochi, W.D.; Schwalbert, R.; Andrade, F.H.; Corassa, G.M.; Carter, P.; Gaspar, A.P.; Schmidt, J.; Ciampitti, I.A. Soybean seed yield response to plant density by yield environment in North America. *Agron. J.* **2019**, *111*, 1923–1932. [CrossRef]
35. Chirchir, G.J. Seed Quality of Soybean (*Glycine max* [L.] Merrill) Genotypes under Varying Storage and Priming Methods, Mother Plant Nutrient Profiles and Agro-Ecologies in Keyna. Master's Thesis, Kenyatta University, Nairobi, Keyna, 2011.
36. Crusciol, A.A.C.; Nascente, A.S.; Borghi, E.; Soratto, R.P.; Martins, P.O. Improving soil fertility and crop yield in a tropical region with palisadegrass cover crops. *Agron. J.* **2015**, *107*, 2271–2280. [CrossRef]
37. De Freitas, R.M.S.; De Lima, L.E.; Silva, R.S.; Campos, H.D.; Perin, A. Fluxapyroxad in the asian soybean rust control in the Cerrado biome. *Rev. Caatinga* **2016**, *29*, 619–628. [CrossRef]
38. De Luca, M.J.; Hungría, M. Plant densities and modulation of symbiotic nitrogen fixation in soybean. *Sci. Agric.* **2014**, *71*, 181–187. [CrossRef]
39. Dos Santos, H.P.; Fontaneli, R.S.; Pires, J.; Lampert, E.A.; Vargas, A.M.; Verdi, A.C. Grain yield and agronomic traits in soybean according to crop rotation systems. *Bragantia* **2014**, *73*, 263–273. [CrossRef]
40. Ekhtiari, S.; Kobraee, S.; Shamsi, K. Soybean yield under water deficit conditions. *J. Biodivers. Env. Sci.* **2013**, *3*, 46–52.
41. Felisberto, G.; Bruzi, A.T.; Zuffo, A.M.; Zambiazzi, E.V.; Soares, I.O.; De Rezende, P.M.; Botelho, F.B.S. Agronomic performance of RR® soybean cultivars using different pre-sowing desiccation periods and distinct post-emergence herbicides. *Afr. J. Agric. Res.* **2015**, *10*, 3445–3452. [CrossRef]
42. Ferreira, A.S.; Balbinot, A.A., Jr.; Werner, F.; Zucareli, C.; Franchini, J.C.; Debiasi, H. Plant density and mineral nitrogen fertilization influencing yield, yield components and concentration of oil and protein in soybean grains. *Bragantia* **2016**, *75*, 362–370. [CrossRef]
43. Franchini, J.C.; Balbinot, A.A., Jr.; Debiasi, H.; Procópio, S.O. Intercropping of soybean cultivars with Urochloa. *Pesq. Agropec. Trop.* **2014**, *44*, 119–126. [CrossRef]
44. Gai, Z.; Zhang, J.; Li, C. Effects of starter nitrogen fertilizer on soybean root activity, leaf photosynthesis and grain yield. *PLoS ONE* **2017**, *12*. [CrossRef]
45. Galindo, F.S.; Filho, M.C.T.; Buzetti, S.; Santini, J.M.K.; Ludkiewicz, M.G.Z.; Baggio, G. Modes of application of cobalt, molybdenum and *Azospirillum brasilense* on soybean yield and profitability. *Rev. Bras. Eng. Agric. Ambient.* **2017**, *21*, 180–185. [CrossRef]

46. Gaspar, A.P.; Conley, S.P. Responses of canopy reflectance, light interception, and soybean seed yield to replanting suboptimal stands. *Crop Sci.* **2015**, *55*, 377–385. [CrossRef]
47. Gerçek, S.; Boydak, E.; Okant, M.; Dikilita, S. Water pillow irrigation compared to furrow irrigation for soybean production in a semi-arid area. *Agric. Water Manag.* **2008**, *96*, 87–92. [CrossRef]
48. Ghassemi-Golezani, K.; Lotfi, R. Response of soybean cultivars to water stress at reproductive stages. *Int. J. Plant. Anim. Environ. Sci.* **2012**, *2*, 198–202.
49. Gulluoglu, L.; Bakal, H.; Arioglu, H. The effects of twin-row planting pattern and plant population on seed yield and yield components of soybean at late double-cropped planting in Cukurova region. *Turk. J. Field Crop* **2016**, *21*, 59–65. [CrossRef]
50. Hayder, G.; Mumtaz, S.S.; Khan, A.; Khan, S. Maize and soybean intercropping under various levels of soybean seed rates. *Asian J. Plant Sci.* **2003**, *2*, 339–341.
51. Hernández, M.; Cuevas, F. The effect of inoculating with arbuscular *Mycorrhiza* and *Bradyrhizobium* strains on soybean (*Glycine max* (L) Merrill) crop development. *Cult. Trop.* **2003**, *24*, 19–21.
52. Ibrahim, S.E. Agronomic studies on irrigated soybeans in central Sudan: II. Effect of sowing date on grain yield and yield components. *Int. J. Agric. Sci.* **2012**, *2*, 766–773.
53. Jephter, B.F.M.; Mumba, P.; Bokosi, J.M. Assessment of the agronomic productivity and protein content in 16 soybean genotypes. *Afr. J. Food Agric. Nutr. Dev.* **2017**, *17*, 12600–12613. [CrossRef]
54. Kaschuk, G.; Nogueira, M.A.; De Luca, M.J.; Hungria, M. Response of determinate and indeterminate soybean cultivars to basal and topdressing N fertilization compared to sole inoculation with *Bradyrhizobium*. *Field Crops Res.* **2016**, *195*, 21–27. [CrossRef]
55. Keramati, S.; Pirdashti, H.; Esmaili, M.A.; Abbasian, A.; Habibi, M. The critical period of weed control in soybean (*Glycine max* (L.) Merr.) in north of Iran conditions. *Pak. J. Biol. Sci.* **2008**, *11*, 463–467.
56. Kumar, M.; Das, T.K. Integrated weed management for system productivity and economics in soybean (*Glycine max*)—Wheat (*Triticum aestivum*) system. *Indian J. Agron.* **2008**, *53*, 189–194.
57. Lamptey, S.; Yeboah, S.; Sakodie, K.; Berdjour, A. Growth and yield response of soybean under different weeding regimes. *Asian J. Agric. Food Sci.* **2015**, *3*, 155–163.
58. Leblanc, M.L.; Cloutier, D.C. Susceptibility of row-planted soybean (*Glycine max*) to the rotary hoe. *J. Sustain. Agric.* **2011**, *18*, 53–61. [CrossRef]
59. Liu, X.; Rahman, T.; Song, C.; Su, B.; Yang, F.; Yong, T.; Wu, Y.; Zhang, C.; Yang, W. Changes in light environment, morphology, growth and yield of soybean in maize-soybean intercropping systems. *Field Crops Res.* **2017**, *200*, 38–46. [CrossRef]
60. Lyimo, L.D.; Tamba, M.R.; Madege, R.R. Effects of genotype on yield and yield components of soybean (*Glycine max* (L) Merrill). *Afr. J. Agric. Res.* **2016**, *10*, 1930–1936. [CrossRef]
61. Maleki, A.; Naderi, A.; Naseri, R.; Fathi, A.; Bahamin, S.; Maleki, R. Physiological performance of soybean cultivars under drought stress. *Bull. Env. Pharmacol. Life Sci.* **2013**, *2*, 38–44.
62. Mandal, K.G.; Hati, K.M.; Misra, A.K. Biomass yield and energy analysis of soybean production in relation to fertilizer-NPK and organic manure. *Biomass Bioenergy* **2009**, *33*, 1670–1679. [CrossRef]
63. Madanzi, T.; Chiduza, C.; Kageler, S.J.R.; Muziri, T. Effects of different plant populations on yield of different soybean (*Glycine Max* (L.) Merrill) varieties in a smallholder sector of Zimbabwe. *J. Agron.* **2012**, *11*, 9–16. [CrossRef]
64. Mandić, V.; Simić, A.; Krnjaja, V.; Bijelić, Z.; Tomić, Z.; Stanojković, A.; Muslić, D.R. Effect of foliar fertilization on soybean grain yield. *Biotech. Anim. Husb.* **2015**, *31*, 133–143. [CrossRef]
65. Mbah, E.U.; Muoneke, C.O.; Okpara, D.A. Effect of compound fertilizer on the yield and productivity of soybean and maize in soybean/maize intercrop in southeastern Nigeria. *Trop. Subtrop. Agroecosyst.* **2007**, *7*, 87–95.
66. Moosavi, S.S.; Mirhadi, S.M.J.; Imani, A.A.; Khaneghah, A.M.; Moghanlou, B.S. Study of effect of planting date on vegetative traits, reproductive traits and grain yield of soybean cultivars in cold region of Ardabil (Iran). *Afr. J. Agric. Res.* **2011**, *6*, 4879–4883.
67. Mostafavi, K. Grain yield and yield components of soybean upon application of different micronutrient foliar fertilizers at different growth stages. *Int. J. Agric. Res. Rev.* **2012**, *2*, 389–394.
68. Muoneke, C.O.; Ogwuche, M.A.O.; Kalu, B.A. Effect of maize planting density on the performance of maize/soybean intercropping system in a guinea savannah agroecosystem. *Afr. J. Agric. Res.* **2007**, *2*, 667–677.

69. Nico, M.; Miralles, D.J.; Kantolic, A.G. Post-flowering photoperiod and radiation interaction in soybean yield determination: Direct and indirect photoperiodic effects. *Field Crops Res.* **2015**, *176*, 45–55. [CrossRef]
70. Orlowski, J.M.; Haverkamp, B.J.; Laurenz, R.G.; Marburger, D.A.; Wilson, E.W.; Casteel, S.N.; Conley, S.P.; Naeve, S.L.; Nafziger, E.D.; Roozeboom, K.L.; et al. High-input management systems effect on soybean seed yield, yield components and economic break-even probabilities. *Crop Sci.* **2016**, *56*, 1988–2004. [CrossRef]
71. Paudel, B.; Karki, T.B.; Shah, S.C.; Chaudhary, N.K. Yield and economics of maize (*Zea mays*) + soybean (*Glycin max* L. Merrill) intercropping system under different tillage methods. *World J. Agric. Res.* **2015**, *3*, 74–77. [CrossRef]
72. Rosa, C.B.; Marchetti, M.E.; Serra, A.P.; De Souza, L.C.F.; Ensinas, S.C.; Da Silva, E.F.; Lourente, E.R.P.; Dupas, E.; De Moraes, E.M.; Mattos, F.A.; et al. Soybean agronomic performance in narrow and wide row spacing associated with NPK fertilizer under no-tillage. *Afr. J. Agric. Res.* **2016**, *11*, 2947–2956. [CrossRef]
73. Salehi, M. Effect of foliar application of methanol on the growth and yield of soybean. *Environ. Treat. Tech.* **2013**, *1*, 122–125.
74. Sharma, O.P.; Tiwari, S.C.; Raghuwanshi, R.K. Effect of doses and sources of sulphur on nodulation, yield, oil and protein content of soybean and soil properties. *Soybean Res.* **2004**, *2*, 35–40.
75. Souza, R.; Teixeira, I.; Reis, E.; Silva, A. Soybean morphophysiology and yield response to seeding systems and plant populations. *Chil. J. Agric. Res.* **2016**, *76*. [CrossRef]
76. Popović, V.; Vidić, M.; Jocković, D.; Ikanović, J.; Jakšić, S.; Cvijanović, G. Variability and correlations between yield components of soybean [*Glycine max* (L.) Merr.]. *Genetika* **2012**, *44*, 33–45. [CrossRef]
77. Taheri, N.; Zarghami, R.; Oveysi, M.; Tarighaleslami, M. The effect of source limitations on yield and yield components of soybean (*Glycine max* L.) under drought stress. *World Appl. Sci. J.* **2012**, *18*, 788–795. [CrossRef]
78. Tonello, E.S.; Fabbian, N.L.; Sacon, D.; Netto, A.; Silva, V.N.; Milanesi, P.M. Soybean seed origin effects on physiological and sanitary quality and crop yield. *Semin. Ciências Agrárias* **2019**, *40*, 1789–1804. [CrossRef]
79. Uchino, H.; Iwama, K.; Jitsuyama, Y.; Yudate, T.; Nakamura, S. Yield losses of soybean and maize by competition with interseeded cover crops and weeds in organic-based cropping systems. *Field Crops Res.* **2009**, *113*, 342–351. [CrossRef]
80. Yari, V.; Frnia, A.; Maleki, A.; Moradi, M.; Naseri, R.; Ghasemi, M.; Lotfi, A. Yield and yield components of soybean cultivars as affected by planting date. *Bull. Env. Pharmacol. Life Sci.* **2013**, *2*, 85–90.
81. Zuffo, A.M.; Bruzi, A.T.; De Rezende, P.M.; Bianchi, M.C.; Zambiazzi, E.V.; Soares, I.O.; Ribeiro, A.B.M.; Vilela, G.L.D. Morphoagronomic and productive traits of RR® soybean due to inoculation via *Azospirillum brasilense* grove. *Afr. J. Micr. Res.* **2016**, *10*, 438–444. [CrossRef]
82. United States Department of Agriculture (USDA). Foreign Agricultural Service: Soybeans. Available online: https://www.fas.usda.gov/commodities/soybeans (accessed on 5 July 2020).
83. Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA). Embrapa Soja. 2020. Available online: https://www.embrapa.br/soja/cultivos/soja1 (accessed on 5 July 2020). (In Portuguese).
84. Liakos, K.G.; Busato, P.; Moshou, D.; Pearson, S.; Bochtis, D. Machine learning in agriculture: A review. *Sensors* **2018**, *18*, 2674. [CrossRef]
85. Board, J.E. Yield components related to seed yield in determinate soybean. *Crop Sci.* **1987**, *27*, 1296–1297. [CrossRef]
86. De Bruin, J.L.; Pedersen, P. Soybean seed yield response to planting date and seeding rate in the upper Midwest. *Agron. J.* **2008**, *100*, 696–703. [CrossRef]
87. Robinson, A.P.; Conley, S.P.; Volenec, J.J.; Santini, J.B. Analysis of high yielding, early-planted soybean in Indiana. *Agron. J.* **2009**, *101*, 131–139. [CrossRef]
88. Zhang, C.; Kovacs, J.M. The application of small unmanned aerial systems for precision agriculture: A review. *Precis. Agric.* **2012**, *13*, 693–712. [CrossRef]
89. Palmer, P.B.; O'Connell, D.G. Regression analysis for prediction: Understanding the process. *Cardiopulm. Phys. Ther. J.* **2009**, *20*, 23–26. [CrossRef] [PubMed]
90. Smidt, E.R.; Conley, S.P.; Zhu, J.; Arriage, F.J. Identifying field attributes that predict soybean yield using random forest analysis. *Agron. J.* **2016**, *108*, 637–646. [CrossRef]
91. Marko, O.; Brdar, S.; Panic, M.; Lugonja, P.; Crnojevic, V. Soybean varieties portfolio optimisation based on yield prediction. *Comput. Electron. Agric.* **2016**, *127*, 467–474. [CrossRef]
92. Christenson, B.S.; Schapaugh, W.T.; An, N.; Price, K.P.; Prasad, V.; Fritz, A.K. Predicting soybean relative maturity and seed yield using canopy reflectance. *Crop Sci.* **2016**, *56*, 625–643. [CrossRef]

93. Patrício, D.I.; Rieder, R. Computer vision and artificial intelligence in precision agriculture for grain crops: A systematic review. *Comput. Electron. Agric.* **2018**, *153*, 69–81. [CrossRef]
94. Barbedo, J.G.A. Digital image processing techniques for detecting, quantifying and classifying plant diseases. *SpringerPlus* **2013**, *2*, 660. [CrossRef]
95. Vithu, P.; Moses, J.A. Machine vision system for food grain quality evaluation: A review. *Trends Food Sci. Technol.* **2016**, *56*, 13–20. [CrossRef]
96. Uzal, L.C.; Grinblat, G.L.; Namías, R.; Larese, M.G.; Bianchi, J.S.; Morandi, E.N.; Granitto, P.M. Seed-per-pod estimation for plant breeding using deep learning. *Comput. Electron. Agric.* **2018**, *150*, 196–204. [CrossRef]
97. Li, Y.; Jia, J.; Khattak, A.M.; Sun, S.; Gao, W.; Wang, M. Soybean seed counting based on pod image using two-column convolution neural network. *IEEE Access* **2019**, *7*, 64177–64185. [CrossRef]
98. Liu, X.; Jia, W.; Ruan, C.; Zhao, D.; Gu, Y.; Chen, W. The recognition of apple fruits in plastic bags based on block classification. *Precis. Agric.* **2018**, *19*, 735–749. [CrossRef]
99. Qureshi, W.S.; Payne, A.; Walsh, K.B.; Linker, R.; Cohen, O.; Dailey, M.N. Machine vision for counting fruit on mango tree canopies. *Precis. Agric.* **2017**, *18*, 224–244. [CrossRef]
100. Ramos, P.J.; Prieto, F.A.; Montoya, E.C.; Oliveros, C.E. Automatic fruit count on coffee branches using computer vision. *Comput. Electron. Agric.* **2017**, *137*, 9–22. [CrossRef]
101. Sa, I.; Ge, Z.; Dayoub, F.; Upcroft, B.; Perez, T.; Mccool, C. Deepfruits: A fruit detection system using deep neural networks. *Sensors* **2016**, *16*, 1222. [CrossRef] [PubMed]