FIELD-BASED HIGH-THROUGHPUT PHENOTYPING APPROACH FOR SOYBEAN PLANT IMPROVEMENT

L. Li[1], L.Tian[1], D. Jiang[2], Z.Lu[3], R. P.Campos[4]*

[1]Department of Agriculture and Biological Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA;liujunli@illinois.edu
[2]School of Traffic Science and Engineering, Beihang University, Beijing, China
[3]College of Biosystem Engineering and Food Science, Zhejiang University, Hangzhou, China
[4]Department of Biosystems Engineering, University of São Paulo, Pirassununga, SP, Brazil

ABSTRACT

The continued development of new, high yielding cultivars needed to meet the world's growing food demands will be aided by improving the technology to rapidly phenotype potential cultivars. High-throughput phenotyping (HTP) is essential to maximize the greatest value of genetics analysis and to better understand the plant biology and physiology in view of a "Feed the World in 2050" theme. Field-based high-throughput phenotyping platform including a LiDAR-based proximal sensing field scout and its data processing system were developed and applied for near-real-time remote sensing of soybean 3D canopy structure variation and plant growth condition among different plant and plot scale in this paper.
The proximal sensing field scout consists of an extremely accurate distance measurement sensor, called the SICK LiDAR (Light Detection And Ranging) scanner, with distance measurements over a 180 degree area up to 8 meters away among canopy. The GPS sensor and a 6 DOF (Six Degrees of Freedom) Inertia Measurement Unit were mounted on the tractor. The scanning frequency is 1Hz and the resolution will be 10mm. The canopy height and density will be calculated after the canopy 3D structure is reconstructed. The resolution is 0.16m / pixel, which is much smaller than the individual soybean plant size of 0.762m. The ground reference data of each plot, such as canopy height information, were collected during the growing season, and the yield harvested at the end of the season was measured. These ground-based agronomic traits data will be correlated with the ground proximal sensing data. 4 rows of the soybean were proximal sensed and scouted by LiDAR considering the traveling speed and the sensor resolution. The average maximum height of the plant and the canopy coverage were derived, and the plant canopy volume was calculated based on the canopy 3D reconstruction algorithm. Therefore, a novel canopy volume model algorithm was developed to correlate with the yield considering the canopy height and canopy coverage information based on the entire field and each soybean family evaluated. The results shows that the canopy height by proximal sensing of LiDAR has a good correlation with that of field measurement (R^2=78.51%) and the canopy volume value has a correlation of 69.34% (R^2=69.34%) with soybean yield harvested in 2013. Additionally, some of the families has a positive

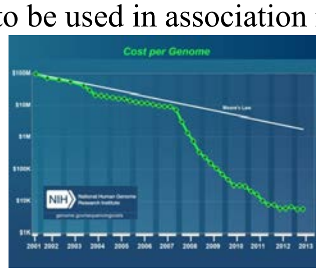
correlation between the yield harvested and the soybean canopy volume across the whole field and the proposed canopy volume model has better stability and robustness than that of the field height only. It is indicated that proposed proximal sensing system is high efficiency and effective to conduct the field-based high-throughput phenotyping and which could be helpful to better understand the plant biology and growth condition for plant improvement. Commercial pre-harvest yield prediction would likely to be made during the early season based on the proximal sensing data. Moreover, the proximal sensing approach for high-throughput phenotyping could be applied in QTL and association mapping with crop genetics in the future.

Keywords: High-throughput phenotyping, plant improvement, proximal sensing, precision agriculture, Light Detection And Ranging (LiDAR)

INTRODUCTION

One of the greatest challenges of the 21st century will be to expand crop production to meet increasing demands for food, clothing, and fuel brought on by both the growing human population and its increasing affluence. This needs to be accomplished in the face of climate change, which is predicted to lead to less favorable environments for crop production. The most environmentally friendly way to meet these demands is through developing and providing highly productive crop cultivars to farmers. The biotech revolution is impacting cultivar development by increasing the capabilities of researchers to analyze the DNA of crop plants. There are now relatively inexpensive methods to analyze plants with tens of thousands of genetic markers and we are on the brink of being able to fully sequence the genome of plants in breeding programs. Genetic analysis of plants has its greatest value when it can be associated with plant phenotypes (Grift et al., 2011). This association gives plant breeders the ability to identify the location of genes that impact economically important traits and this information can be used to guide variety development in plant breeding programs (Cobb et al. 2013). However, the costs of genotyping and phenotyping did not match well so far as indicated in Fig.1. It can be seen from Fig.1 (a) that the cost per genome decreased dramatically down to one tenth since 2007 (Green and Guyer, 2011). While our ability to analyze DNA is increasing at exponential rates, the capacity to phenotype plants in a field setting has not improved nearly as rapidly. As indicated in Fig.1 (b), modern breeders must still walk through fields many times over the course of a growing season and score plants visually in largely the same way they did 30 years ago whereas the tractor/combine and its automation develop faster. To estimate seed yield, breeders still need to harvest plots and measure the weight of seed from the plots. Although harvest technology has improved through the use of self-propelled combines with automatic weighing systems, this still remains a significant bottleneck in research programs. In large breeding programs, tens of thousands of plots are grown annually, making the phenotyping very labor intensive. There is a critical need to improve phenotyping technology to match the improvements in genetic technology. Concurrent with the increased development of the agriculture mechanization, remote sensing and proximal sensing technology, particularly the on-the-go sensing system, have the potential to increase our ability to rapidly phenotype plants in breeding programs and agronomic studies. In addition, proximal sensing and remote sensing can give researchers the ability to measure traits that they have not been able to measure easily in the past, such as photosynthetic efficiency and stress tolerance (Diers, et al., 2013).

The LiDAR sensor is one the most accurate and commonly applied electronic devices for canopy characterization, and the information that is collected while driving a tractor along a crop row could be used to estimate the canopy height and density distribution by evaluating the frequency of LiDAR returns with regards to the driving speed (Llorens et al., 2011). The 3-D canopy structure can be estimated by use of LiDAR scanning and its use to facilitate the location, density and height of plant associated with both the upper and sub-canopy structure, which could be a good auxiliary phenotyping tool for large-scale field phenotyping. Lee et al. (Lee and Lucas, 2007) retrieved the tree and forest structural attributes from LiDAR data and applied the canopy height models to develop a new index, termed the Height-Scaled Crown Openness Index (HSCOI), which provides a quantitative measure of the relative penetration of LiDAR pulses into the canopy for mapping and attributing stems in complex, multi-layered forests. Canopy characterization using LiDAR has been also studied in orchards and vine plantations (Polo et al. 2009, Sanz et al. 2004). Llorens et al. studied the generation of the georeferenced plantation map and correlated with LAI (Polo et al. 2009). Rosell studied the estimation of the 3-D structure and volume of the tree (Polo et al., 2009, Rosell et al., 2009). However, the use of LiDAR for row crops and the correlation between row crop canopy structure information and agronomic traits has not been studied yet, which is critical to understand site-specific crop responses in near-real-time for remote sensing and agronomic precision application as well as for breeding purposes. This research is aimed to develop and test a high-throughput field phenotyping solution combined proximal sensing by use of LiDAR scanner and remote sensing by use of UAV for soybean plant improvement. The soybean 3-D canopy structure information will be used to calculate the canopy height, density and volume based on LiDAR proximal sensing. And the field phentoyping data during the mid-season from proximal sensing will be correlated with agronomic traits such as height, yield, and plant maturity date. Thus a yield-oriented traits model will be established to predict soybean yield and detect the in-field variety of plant stress as well as provide data to be used in association mapping in a related genetics study.



(a) Decreased genotyping cost (b) manual field phenotyping

Fig. 1 Phenotyping capability associated with genotyping

MATERIALS AND METHOD

NAM soybean field for high-throughput phenotyping

As part of high-impact soybean improvement using genetics and genomics, high-throughput field phenotyping was proposed and tested for the NAM soybean field. The research is conducted using a series of 40 mapping populations that each contains 140 experimental lines giving a total of 5,600 experimental lines across all populations as indicated in Fig. 2. These populations were developed by mating the maturity group (MG) III variety IA3023 with 40 other diverse parents (i.e., soybean varieties and experimental lines). Phenotypic data was collected on the 41 parents and 5,600 progeny of the Nested Association Mapping (NAM) populations including yield, plant maturity, height and lodging, and so on. Locations of genes controlling these characters will be determined. The NAM populations will be mapped with 1536 SNPs and the traits will be mapped as QTL (Diers, 2013). The NAM soybean field size is shaping as a reversed "L" covering 15 acre. The field phenotyping is a non-destructive proximal sensing with LiDAR on a tractor. The NAM populations will be field phenotyped and evaluated in 4-row yield plots using specified proximal sensing protocol and data analysis procedures so that breeding values for yield and other traits can be calculated for each of the 5,600 lines as well as that of each particular NAM family.

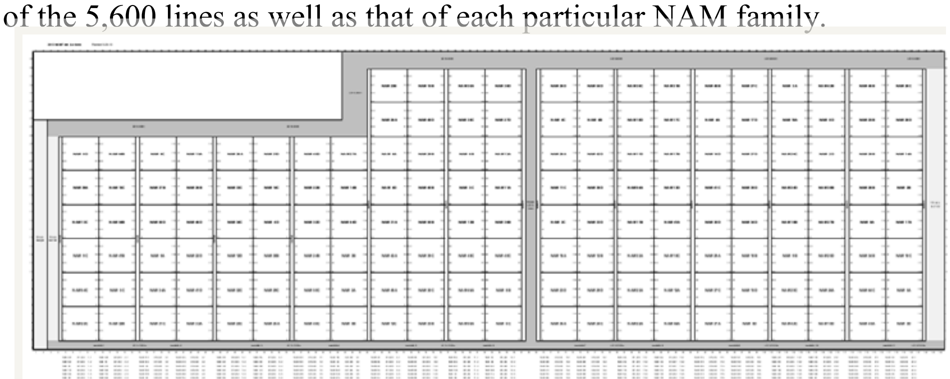
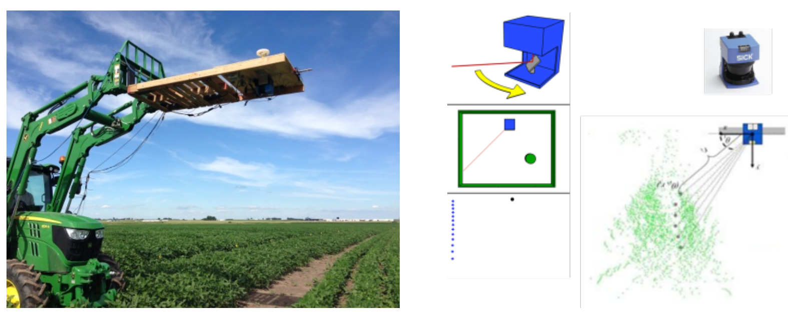


Fig. 2 NAM soybean field layout

Proximal sensing system development for soybean

As indicated in Fig.3 (a), the John Deere tractor 6125R with rated power of 145hp and wheelbase of 2580mm was utilized as the proximal sensing platform. The sensor board and its LiDAR sensing system was mounted at the center of the front loader bracket. The JD Tractor 6125R has an Autotrac auto-guidance system option and triple-Link Suspension (TLS™) Plus including the vertical linkage, longitudinal linkage and lateral linkage active whenever the tractor exceeds traveling speed of 1.5 km/h (0.9 mph), which ensures driving comfort and sensing stability. The proximal sensing system consists of a sensor board with a LiDAR scanner (SICK LiDAR LMS 200), a GPS antenna/receiver (Model: OEM4, Novatel Co., Calgary, Alberta, Canada), a 6 DOF Inertia Measurement Unit (FOG, Model: DMU-HDX, Crossbow, Milpitas, California, USA) as well as the connected power supply and communication interface with the control computer in the tractor cab. The LabVIEW program in the computer controls the sensors to record and display the readings from the sensor. The SICK LiDAR scanner was mounted at the center and beneath the sensor board for the purpose of drawing a 3D crop height/coverage map based on the driving direction. LiDAR is a fully-automatic divergent laser scanner based on measurement of the time-of-flight (TOF) with an accuracy of ±15 mm in a single shot measurement and a 5 mm standard deviation in a range of up to 8 m. It only works while the vehicle is moving with constant transit speed. The GPS sensor could provide 1Hz position measurement. Due to the vibration of the generator and the uneven off road surface, the result from the SICK LiDAR scanner needed to be corrected based on the vehicle attitude and vibration in the post data processing. A 6 DOF Inertia Measurement Unit (Crossbow) was mounted on the vehicle close to the SICK LiDAR system. The clearance of the senor board could be up to 3607mm. The traveling speed was set at 3mph so that the whole field of 15 acres could be covered within four hours and each plot has 3 scans along the tractor driving direction. The soybean of 8 rows/4 plots was covered in the horizontal direction (perpendicular to the driving direction). Each soybean plot size was 12 feet by 5 feet, so the LiDAR resolution in the driving direction will be



(a) LiDAR based on-the-go proximal sensing (b) Lidar sensing system

Fig. 3 Proximal sensing platform and LiDAR sensing for field phenotyping

4 foot/scan. The time between the transmission and reception of the pulsed near-infrared laser beam is used to measure the distance between the scanner and the reflecting object surface. The laser beam is deflected by a rotating mirror turning at 4,500rpm (75rps), which results in a fan shaped scan pattern in which the maximum scanning angle is 180°. The angular resolution can be set to 0.5, making 361 measurements at full scanning range with a response time of 26ms. The LMS-200 has a standard RS232 serial port for data transfer, which can be set to 9.6, 19.2 or 38.4 Kbaud. The selected configuration during the field tests conducted for this study was: angular resolution 0.5° range of 180° and data transfer 38.4 Kbaud. These characteristics enabled the best resolution to be obtained during the scanning process. Therefore, the resolution of the soybean

plant will be 3 by 70 pixels per plot, which represents the plot adequately; although a lower driving speed is still desirable.

As indicated in Fig. 3 (b), the canopy height will be derived from the LiDAR sensor readings by:

$$Xp = L * \cos\theta \quad (1)$$

$$Yp = Ho - L * sin\theta \quad (2)$$

Where, the Xp and Yp is actual position of the plant canopy, L is the distance reading from the LiDAR sensor, which is the distance between the LiDAR scanner and the soybean plant canopy surface, θ is the angle between the transmission and the horizontal plane.

Table. 1 The proximal sensing system configuration and its application

Sensor	Model	Platform	Application
3D canopy sensing system		Tractor	John Deere 6125R tractor, with wheelbase of 2850mm, with front loader bracket, Autotrac guidance system, and triple-Link Suspension (TLS™) Plus when the driving speed exceeds 0.9mph for driving comfort and sensing stability.
LiDAR	LMS200	Tractor	Scanning plants with maximum width of 8 m to predict canopy 3D structure
RTK-GPS		Tractor	Geo-referencing for data collection

LiDAR-based 3D reconstruction and canopy structure measurement

The LiDAR sensor was controlled and the readings were recorded by the computer LabVIEW program. The 3D profile data of one scan from LiDAR sensor is illustrated as Fig.4. In order to reconstruct the plant 3D structure and calculate the volume information, we proposed to estimate the plant height and its position according to Eq. (1) and Eq. (2). There were at least three pixels (scans)

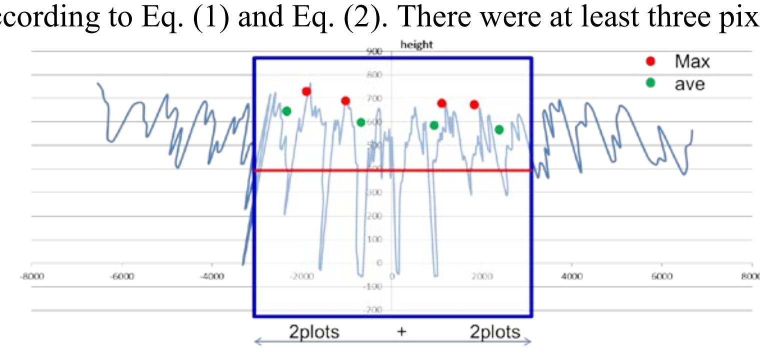


Fig. 4 Plant 3D structure reconstruction for quantitative estimation

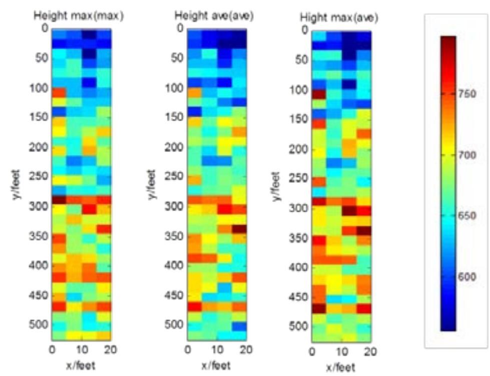


Fig. 5 Canopy height calculation option based on plots

for each plot in the driving direction, and about 68 pixels (points) per scan for each plot perpendicular to the driving direction. There are at least three types of height model as indicated in Fig.5 that could be considered to the height value for each plot: (1) The max of max height value of each scan; (2) the average of max height value of each scan; (3) the average of average height value of each scan. The canopy height is estimated by the average of the max height values among the three scans in the tractor's driving direction as calculated as follows:

$$\mathrm{H} = \frac{\sum_{i=1-3}^{j=1-68} \boldsymbol{max(h_{ij})}}{\boldsymbol{ij}} \tag{3}$$

Where, i is the pixel number in the direction along the tractor's driving direction for each plot ranging from 1 to 3(scans), and j is the pixel number perpendicular to the driving direction range from 1-68.

The canopy height option based on plots were calculated and illustrated. The average of the maximum height value of each scan was chosen for the following analysis. The resolution of each plot will be 3 by 68 pixels, that is to say, the height value of each plot will be the average or max of the 3 by 68, for a total of 204 pixels. The preliminary results show that the resolution is high enough to estimate the plot-based canopy structure information.

Therefore, the canopy volume could be determined once the canopy height and coverage calculated.

$$\boldsymbol{V} = \boldsymbol{HCL_0W_0} \tag{4}$$

Where, V is the estimated canopy volume in m^3, H is the canopy height, C is the canopy coverage, L_0=12 feet is the plot size along the driving direction and W_0 =5 feet is the plot size perpendicular to the driving direction.

The canopy volume-yield model based on LiDAR proximal sensing system was put forward as follows:

$$\mathbf{Y} = \boldsymbol{a_o}{*}\mathbf{V} + \mathbf{b_0} \tag{5}$$

Where, the canopy volume V was calculated and applied to be correlated with the yield Y. a_0 and b_0 is the constant related to the crop stage during the season from planting to harvesting.
Therefore, the canopy volume model based on LiDAR could be conducted using high-throughput phenotyping and these results could be correlated with yield.

RESULTS AND DISCUSSION

The LiDAR based proximal sensing system was designed, completed and tested in the NAM soybean field, and the soybean field phenotyping data was collected during the mid-season, analyzed and correlated (based on the soybean family) with the field-measured agronomic data such as the yield and height, maturity, lodging. The details are discussed as follows.

Soybean canopy height proximal sensing and field measurement
As mentioned earlier, the soybean height has three calculation options, and the average of the maximum height value for each scan was chosen for the canopy height calculation for each plot in this research. The in-field canopy height distributed was plotted as indicated in Fig.6. The canopy height was field sampled and measured as indicated in Fig. 7. Furthermore, the canopy height proximal sensed by LiDAR was correlated with the canopy height by field measurement as indicated in Fig.8. The results shows that the accuracy of the proximal sensing system is as high as 78.51% (R^2=78.51%), and there are two aspects need to be considered for the further experimental protocol design and analysis: one is the system error from the proximal sensing system, which might be improved by more stable and lower driving speed; another will be real-time proximal sensing and ground referencing data collection, which could be further improve the phenotyping data correlation efficiency. The proximal sensing was conducted in the middle of the growing season (July) and the field measurement of height was collected almost at the end of the season (September).

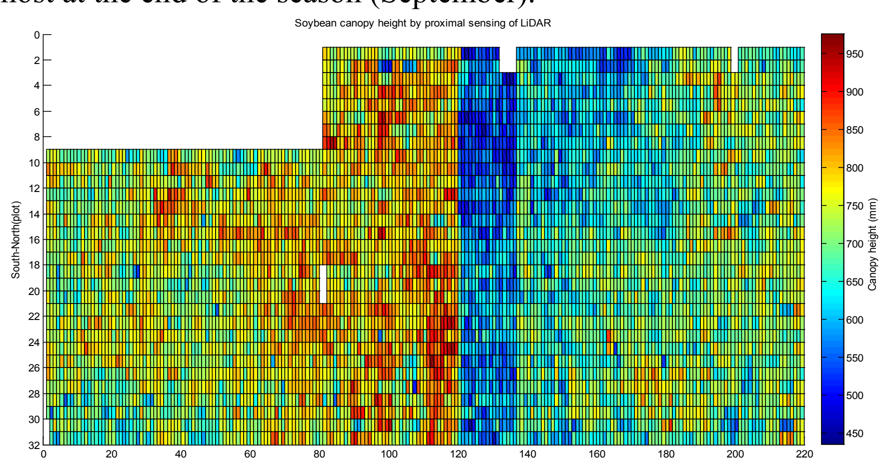


Fig. 6 Canopy height distribution proximal sensed by proposed LiDAR system

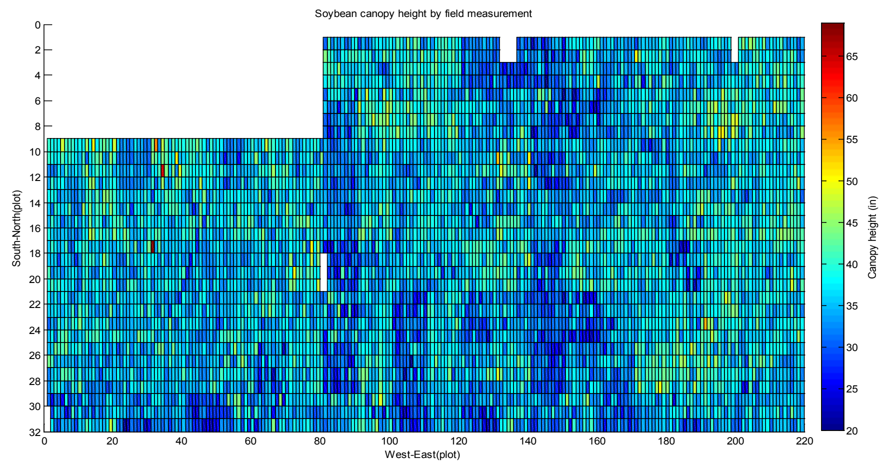


Fig. 7 Canopy height distribution by field measurement

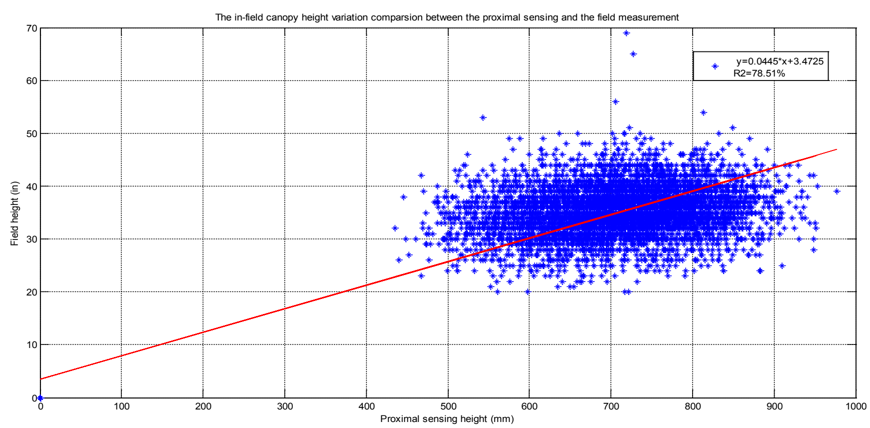


Fig. 8 Correlation between the canopy height between proximal sensing and field measurement

Soybean canopy coverage sensing and estimation

Canopy coverage estimates of the area of influence of the plant, and is more highly related to biomass than plot density, meaning canopy coverage reflects the amount of CO_2 and light that the plant captures and turns into phytomass (above-ground plant biomass) (Daubenmire, 1959). Therefore, the canopy coverage is critical along with the plant canopy height for correlation with yield. Due to the proximal sensing at mid and late season, the canopy nearly closed. In this case, we calculated the percentage of the canopy 400mm above the ground over the whole canopy as the canopy coverage. The canopy coverage was calculated and the result was mapped in Fig.9. It can be seen that there are still obvious in-field variability among the canopy although the canopy coverage is almost 1, that is to say, the canopy is fully closed at the late season.

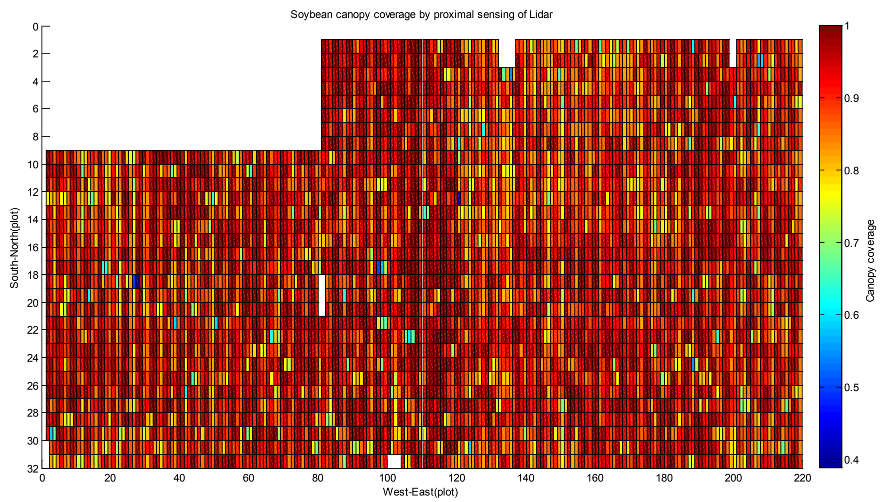


Fig. 9 Soybean canopy coverage distribution based on LiDAR

Soybean canopy volume calculation based on LiDAR system

Based on the Eq. (4), the canopy volume representing the canopy 3D structure information could be calculated as indicated in Fig. 10. The canopy volume could be converted to m^3, which is more highly related to the biomass or yield. Beside correlation with the canopy height, the canopy volume during the mid-season will be correlated with yield harvested in the end of the season.

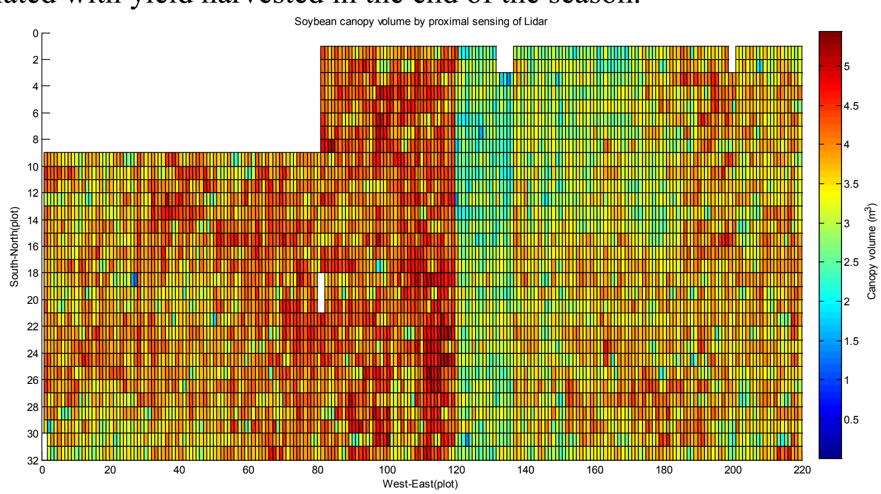


Fig. 10 Soybean canopy volume information mapped from LiDAR system

Canopy volume yield model based on proximal sensing

To enhance the estimation of field agronomic traits such as the plant height, the canopy coverage and the canopy vegetation status for plant improvement as well as to verify the feasibility of soybean yield prediction based on the proximal sensing data, the canopy volume model based on the canopy 3-D structure information from the LiDAR proximal sensing data was built and will be correlated with the soybean yield harvested in the field at the end of the season.

The yield data based on each plot from the yield monitor of the combine was indicated in Fig. 11. The canopy volume information in the entire field was correlated with the yield harvested by the combine in the end of the season as indicated in Fig. 12. The results show that the canopy volume has a good correlation of 69.34% (R^2=69.34%) with the soybean yield in 2013. And the proposed canopy volume-yield model and the proximal sensing accuracy and the coefficient between the canopy structure information and the yield data were verified.

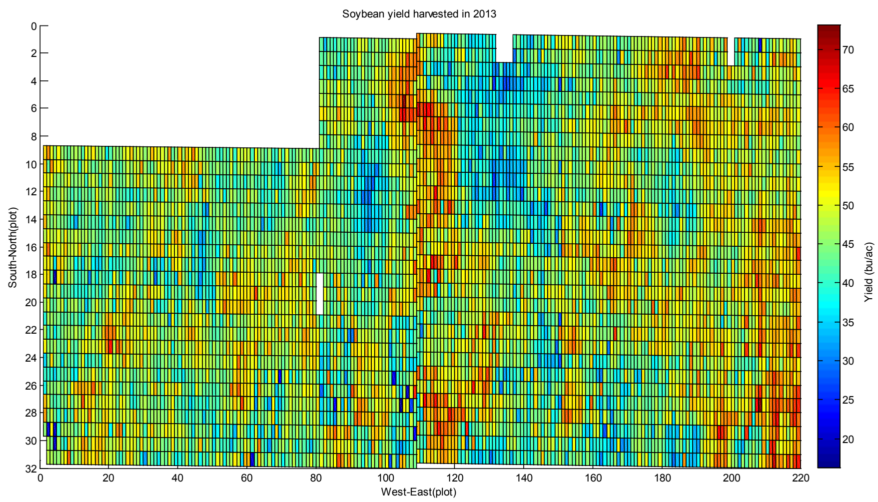


Fig. 11 Soybean yield harvested data in 2013 by combine monitor

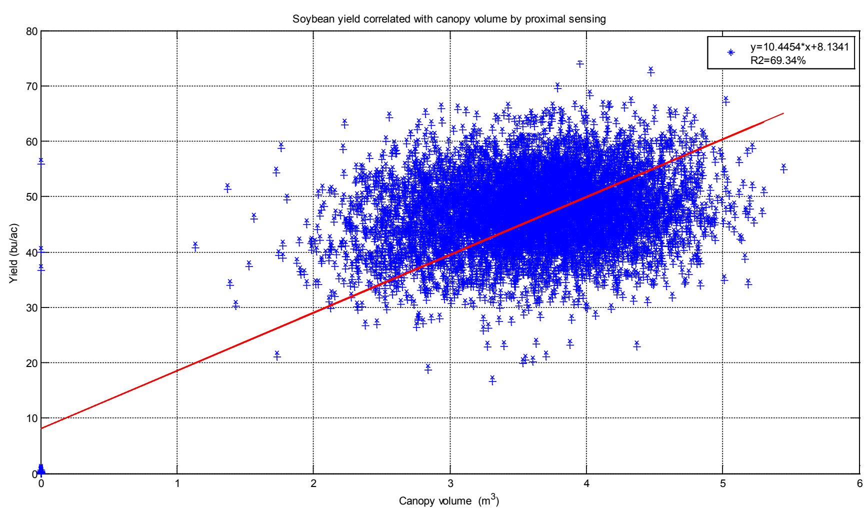


Fig. 12 Yield correlation based on canopy volume by proximal sensing

The canopy volume information of two families, NAM 28 and NAM 4, were chosen randomly to be correlated with the yield and to verify the prediction accuracy of the proposed canopy volume model for the yield prediction. As indicated in Fig. 13, the canopy height by field measurement and the proposed canopy volume values were correlated with the soybean yield and compared with each other. It could be seen that the correlation between the field height and the proximal sensing height vs the yield is 30.45% and 29.28% for NAM 28, and 6.86% and 29.14% for NAM 4. The results show that the genetics analysis need to be further conducted to better understand the plant biology and correlate with the trait such as yield, stress resistance and so on. It is either indicated that the correlation between the canopy volume value and the yield is more stable than that of the canopy height only by field measurement when compared Fig. 13(b) and (d) (both of them is around 30%). However, the canopy volume model needs to be further improved in a second season of field phenotyping data so that the phenotyping information as well as the canopy volume yield model could be better used for family-based associate mapping and breeding study in the future.

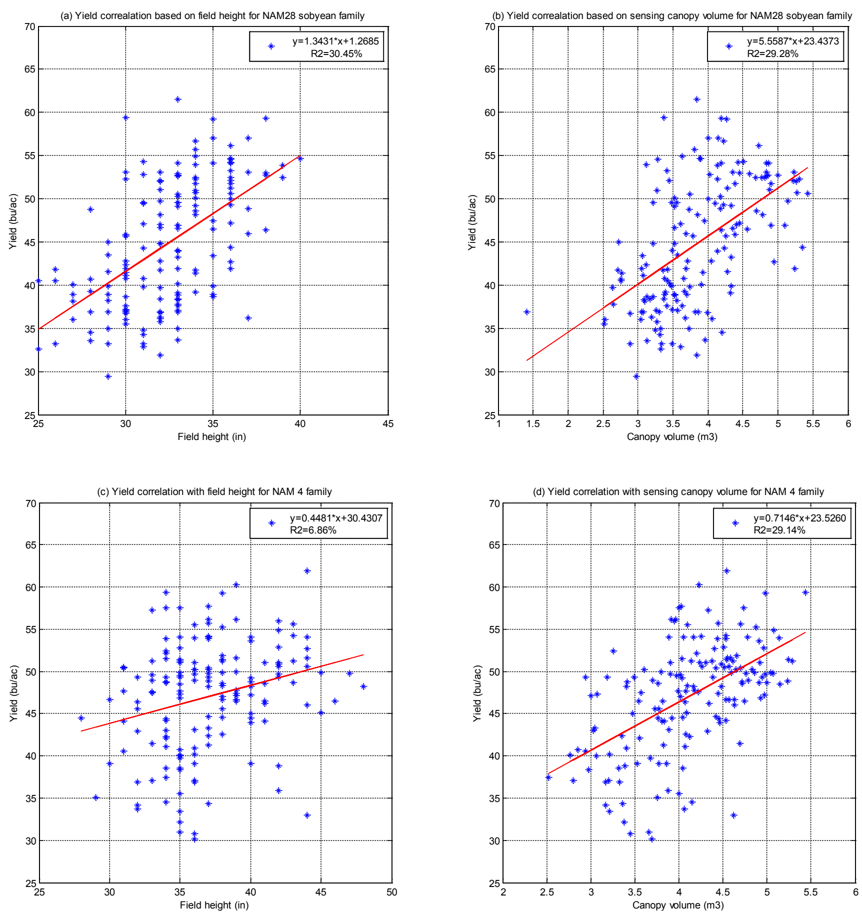


Fig. 13 Family based NAM soybean yield correlation with field height and proximal sensing canopy volume information

CONCLUSION

To maximize the value of the soybean genetics research, a field phenotyping system including a field scout with a LiDAR based proximal sensing system and its data processing system was proposed for high-throughput phenotyping and plant improvement in the paper. And the NAM populations was field phenotyped and evaluated in 2-row yield plots using specified proximal sensing protocol and data analysis procedures so that breeding values for yield and other traits will be calculated for each of the 5,600 lines as well as that of the particular population/family. The field phenotyping is to non-destructive proximal sense the soybean canopy 3D structure and vegetation status so that the phenotyping data could be correlated with the field agronomic data such as yield, height, maturity and associated mapping with the genetics as well. The experimental field phenotyping for soybean for the whole 15 acre NAM field was conducted for the first time. The 3-D canopy structure from the LiDAR scanner readings was reconstructed and then the canopy height, canopy coverage information was derived based on the average of the max height algorithm and canopy coverage calculation algorithm. The canopy volume was derived and considered to be correlated with the soybean yield. The ground truth data of each plot such as canopy height, yield information was collected during the growing season, which was correlated with the ground proximal sensing data. Therefore, a novel canopy volume model was developed to correlate as well as predict the yield considered the canopy 3D volume information based on the soybean family. The results shows that the canopy height by proximal sensing of LiDAR has a good correlation with that of field measurement (R^2=78.51%) and the proposed canopy volume has a correlation of 69.34% with the soybean yield in 2013. Furthermore, some of the families have a positive correlation between the yield harvested and the soybean canopy volume across the whole field and the proposed canopy volume model has better stability and robustness than that of the field height only. It is indicated that the synergy of proximal sensing is high efficiency and effective to conduct the field-based high-throughput phenotyping and which could be helpful to better understand the plant biology and growth condition for plant improvement. Therefore, the proposed proximal sensing system and its methodology as well as the canopy volume-yield model was verified to be a promising tool to correlate the yield and detect the in-field variety of plant stress. And the canopy volume model need to be further improved in another season's field phenotyping data so that the phenotyping information as well as the canopy volume yield model could be used for family-based associate mapping and breeding study in the future.

ACKNOWLEDGEMENTS

The authors would like to thank Campus research project (ID # 13238), and the ACES Future Interdisciplinary Research Explorations (FIRE) project of University of Illinois at Urbana-Champaign to support the program "High-throughput phenotyping for plant improvement".

REFERENCES

Cobb, J.N., Declerck, G., Greenberg A., Clark R., McCouch, S. 2013. Next-generation phenotyping: requirements and strategies for enhancing our understanding of genotype-phenotype relationships and its relevance to crop improvement. Theor. Appl. Genet. 126(4): p. 867-87. doi: 10.1007/s00122-013-2066-0.

Diers, B. USDA research project report. 2013. High-impact research for soybean improvement using genetics and genomics Illinois. http://www.ncaur.usda.gov/research/projects/projects.htm?ACCN_NO=422335&fy=2013

Daubenmire, R. F. 1959. A canopy-coverage method of vegetation analysis. Northwest Science. 33: p. 43-64

Grift, T.E., M., Bohn, and J. Novais .2011. High-Throughput phenotyping technology for maize roots. Biosystems Engineering. 110: p. 40-48, doi:10.1016/j.biosystemseng.2011.06.004.

Green, E. D., Guyer, M. S.. 2011. Charting a course for genomic medicine from base pairs to bedside. Nature. 470: p. 204-213. doi:10.1038/nature09764

Llorens, J.; Gil, E.; Llop, J.; Escolà, A. 2011. Ultrasonic and LiDAR sensors for electronic canopy characterization in vineyards: advances to improve pesticide application Methods. Sensors. 11: p. 2177-2194

Llorens J., Gil, E., Llop, J. and Queraltó M.. 2011. Georeferenced LiDAR 3D vine plantation map generation. Sensors. 11(6): p. 6237-6256. doi:10.3390/s110606237

Lee, A.C.; Lucas, R.M.. 2007. A LiDAR-derived canopy density model for tree stem and crown mapping in Australian forests. Remot. Sens. Environ. 111(4): p. 493-518.

Milton E.J.. 1987. Principles of field spectroscopy, Int. J. Remote Sensing. 8(12): p. 1807-1827

Polo, R., Sanz, R., Llorens, J., Arnó, J., et al. et al. 2009. A Tractor-Mounted Scanning LIDAR for the Non-Destructive Measurement of Vegetative Volume and Surface Area of Tree-Row Plantations: A Comparison with Conventional Destructive Measurements. Biosyst. Eng. 102: p128-134.

Rosell, J.R., Llorens, J.; Sanz, R.; Arnó, J. et al. 2009. Obtaining the three-dimensional structure of tree orchards from remote 2D terrestrial LIDAR scanning. Agric. Forest Meteorol. 149(9): p. 1505-1515.

M, Rosell JR. 2004. Advances in the measurement of structural characteristics of plants with a LIDAR scanner. In: Eur AgEng Conference. Leuven, Belgium: p.12–16.