Development of vehicle-mounted phenotyping and envirotyping mapping platforms

Roberto Mario Buelvas Gómez

Faculty of Agricultural and Environmental Sciences

Department of Bioresource Engineering

Macdonald Campus of McGill University, Ste. Anne de Bellevue

McGill University, Montréal

December 2020

Thesis submitted to McGill University in partial fulfillment of the requirements of the degree of

Doctor of Philosophy

Dedication

This dissertation is dedicated to my wife Yodit, my daughter Eleni, my parents, my brother, and all of my family. This work would not have been possible without them and their support.

Abstract

The ability to measure plant traits in a fast, on-site, and cheap manner would be of benefit to plant breeders, growers, and consumers alike. Reliable on-farm experimentation requires accounting for multiple factors which affect the environment in which plants grow. In this research, two platforms were developed for measuring soil, weather, and crop data in an efficient manner. Methods for data processing, combining the output of these systems, were also implemented. This combination provides a novel framework that facilitates the creation of models to explain plant phenotypical traits. The overall objective of the research was to evaluate the viability of having the platforms work in tandem. The first objective was to develop a High-Throughput Plant Phenotyping (HTPP) platform. The proposed system was designed to be vehicle mounted and it was compared to a manual setup with similar capabilities. The second objective was to develop a Proximal Soil Sensing (PSS) platform. The design of the second system included the ability to open a borehole using a drill and deploying a hyperspectral probe to collect data. The third goal was to combine the data collected from both platforms. A process for predicting soil chemical properties based on soil spectral response was the first component of the combination process. This information could be interpolated across a field where the HTPP platform had measured the phenotypical and weather data for a crop of beans. The effect of the different factors was evaluated during the comparison of the different crop cultivars evaluated in the experiment. Overall, a complete data collection process for on-farm experimentation was based on the proposed platforms. The results of this work were the design of both a HTPP platform and a PSS platform, the implementation of similarity-based regression for soil chemometrics, and the procedure for combining the factors of plant growth. Expanded adoption of these technologies would result in more frequent and convenient data collection surveys for different types of growers, plant breeders, and agronomists. The

availability of these data could be used to improve modelling techniques, allow for better farm management, and to provide automated solutions for the agricultural industry.

Résumé

La capacité de mesurer les caractéristiques des plantes de manière rapide, sur place et économique profiterait aux sélectionneurs, producteurs ainsi qu'aux consommateurs. Afin de réaliser des essais à la ferme fiables, il est nécessaire de tenir en compte plusieurs facteurs qui affectent le milieu dans lequel la plante se développe. Dans la présente étude, deux plateformes ont été développées afin de mesurer les données du sol, des conditions météorologiques et des cultures de manière efficace. Des méthodes pour le traitement des données combinant les résultats de ces systèmes ont également été mises au point. La combinaison de ceux-ci fournit un nouveau cadre qui facilite la création de modèles pour expliquer les traits phénotypiques des plantes. L'objectif général de la présente étude était d'évaluer la viabilité des plateformes travaillant en tandem. Le premier objectif était de développer une plateforme de phénotypage des plantes à haut débit (*High-Throughput Plant Phenotyping*, HTPP). Le système proposé a été conçu pour être monté sur véhicule et a été comparé à une configuration manuelle avec des capacités similaires. Le deuxième objectif était de développer une plateforme de détection proximale des sols (*Proximal Soil Sensing*, PSS). La conception de ce système comprend la capacité de creuser un trou dans le sol pour la collecte des données à l'aide d'une perceuse. Le troisième but était de combiner les données récoltées des deux plateformes. Un processus de prédiction des propriétés chimiques du sol basé sur la réponse spectrale du sol a été le premier composant du processus de combinaison. Cette information peut ensuite être interpolée à travers un terrain où la plateforme HTPP aurait déjà mesuré les données phénotypiques et météorologiques pour des cultures d'haricots. L'effet des facteurs différentes a été évalué pendant la comparaison des différents cultivars de cultures évalués dans l'expérience. Dans l'ensemble, un processus complet de collecte de données pour l'expérimentation à la ferme était basé sur les plates-formes proposées. Les résultats de ce travail ont été la conception d'une

plateforme HTPP, la conception d'une plateforme PSS, la mise en œuvre d'une régression basée sur la similarité pour la chimiométrie du sol et la procédure de combinaison des facteurs de croissance des plantes. L'adoption élargie de ces technologies se traduirait par des enquêtes de collecte de données plus fréquentes et plus pratiques pour différents types de cultivateurs, de sélectionneurs de plantes et d'agronomes. La disponibilité de ces données pourrait être utilisée pour améliorer les techniques de modélisation, parvenir à une meilleure gestion des exploitations agricoles et automatiser des solutions pour l'industrie agricole.

Acknowledgements

I deeply thank my supervisor, Dr. Viacheslav I. Adamchuk, for his support and trust. His guidance was invaluable.

I highly appreciate the contributions of Gabriel Mangeat, Jacques Michiels, and Benjamin de Leener from ChrysaLabs Inc. especially with the materials and information associated with the studies documented in Chapters 4 and 5.

I thank Dr. Valerio Hoyos Villegas for providing testing spaces and materials for the experiments documented in Chapters 3 and 5. I also thank Marc Samoisette and the staff at Emile A. Lods Agronomy Research Centre for facilitating logistics during the data collection process.

I thank Scott Manktelow for all the advice and help I received from him over the years.

I acknowledge the contributions of the past and current Precision Agriculture and Sensor Systems (PASS) research team members Amanda Jacques Boatswain, Dr. Eko Leksono, Dr. Hsin-Hui Huang, Dr. Jaesung Park, Marie-Christine Marmette, Maxime Leclerc, Dr. Md Saifuzzaman, Mohamed Debbagh, Pierce Dias Carlson, among others, for being great friends and colleagues. I especially thank John Lan and Keenan Simpson for their help with fieldwork. I also thank Louis-Jacques Bourdages for his help with manufacturing.

I thank the members of comprehensive exam committee, Dr. Martina Stromvik, Dr. Valerio Hoyos Villegas, Dr. Vijaya Raghavan, and Dr. Zhiming Qi. I also thank Dr. Pierre Dutilleul, Dr. Martina

Stromvik, and Dr. Valerio Hoyos Villegas for being part of the supervisory committee during the academic program.

This work is supported in part by funding provided by MITACS Accelerate program and the Canadian Foundation for Innovation under CFI Grant 36219 – "Accelerating innovations to adapt agro-environments to climate change: The Eastern Canadian Plant Phenotyping Platform".

Contributions of authors

This dissertation consists of three manuscripts, of which the author was fully responsible for designing and building prototypes, synthesizing data, developing the conceptual framework and analytical approaches, including programming and data preparation/exploration, interpreting results, writing this dissertation and manuscripts, and presenting findings. However, this work could not have been achieved without the contribution of Dr. Viacheslav I. Adamchuk, the supervisor of this dissertation and the co-author of all manuscripts; Dr. Adamchuk provided scientific guidance, advice, and support in the development and reviewing of these manuscripts.

Apart from that, Chapter 3 with the planned submission to *Computers and Electronics in Agriculture* was co-authored by Arlene Whitmore, John Lan, Dr. Martina Stromvik, and Dr. Valerio Hoyos Villegas. Arlene Whitmore, Dr. Martina Stromvik, Dr. Valerio Hoyos Villegas provided technical advice and support with materials. John Lan was instrumental in the assembly of the prototype and its field testing. Chapter 4 was co-authored by Benjamin de Leener, Gabriel Mangeat, and John Lan, and it is planned to be submitted to *Biosystems Engineering*. Benjamin de Leener and Gabriel Mangeat provided technical advice and support with materials. John Lan was instrumental in the assembly of the prototype and in its field testing. Chapter 5 with the planned submission to *Precision Agriculture* was co-authored by Benjamin de Leener, Gabriel Mangeat, and Dr. Valerio Hoyos Villegas. Benjamin de Leener and Gabriel Mangeat provided the dataset for training the prediction algorithm. Dr. Valerio Hoyos Villegas provided technical advice and support with materials.

Table of contents

List of tables

List of figures

List of abbreviations and symbols

AI – Artificial Intelligence

ANOVA – ANalysis Of VAriance

ATR – Attenuated Total Reflectance

CCC – Canopy Chlorophyll Content

CEC – Cation Exchange Capacity

CI – Chlorophyll Index

DAS – Days After Seeding

HTPP – High-Throughput Plant Phenotyping

IDW – Inverse Distance Weighting

LAI – Leaf Area Index

LIDAR – LIght Detection And Ranging

NDRE – Normalized Difference Red Edge index

NDVI – Normalized Difference Vegetation Index

NIR – Near InfraRed

PA – Precision Agriculture

PAR – Photosynthetically Active Radiation

PSS – Proximal Soil Sensing

QTL – Quantitative Trait Locus

RGB – Red Green Blue

SVR – Support Vector Regression

UAV – Unmanned Aerial Vehicle

VI – Vegetation Index

Vis-NIR – Visible and Near InfraRed

WSN – Wireless Sensor Network

CHAPTER 1 Introduction

1.1 General introduction

With an ever-increasing world population, crop production must increase both in yield and nutrient use efficiency to meet the world's need for greater food production. . At the same time, climate change poses a threat to food security worldwide, by increasing the likelihood of events like floods and droughts. Many efforts are being made to confront these problems and move the agricultural industry forward. Precision Agriculture (PA), urban farming, and modern plant breeding have been developed to help growers meet current challenges. All of these strategies share a common need for data e.g. measurements of soil spatial variability in PA, concentration of hydroponic nutrients in urban farming, or any of the myriad plant traits considered in breeding programs.

In the present work, Chapter 1 introduces the rationale and goals of the research. Chapter 2 presents a literature review of topics related to each specific objective. Chapter 3 describes the design and testing of an HTPP platform that also can measure atmospheric properties. Chapter 4 describes the design and testing of a PSS platform with drilling capabilities. Chapter 5 illustrates an example procedure to combine the data collected from both platforms by building models that express the plant phenotypical traits in terms of measured factors of plant growth. Chapter 6 summarizes the results and presents the general conclusions of the work. Finally, Chapter 7 highlights contributions to the knowledge evidenced through the research.

1.2 Statement of rationale

On-farm experimentation plays a key role in many of the techniques currently being explored to improve efficiency of agricultural processes. Data collection to power agricultural decision-makers consumes significant time and labour, especially in outdoor conditions. Remote sensing offers the possibility to reduce this work, but not with the same level of spatial resolution as proximal sensing. This is the reason why automating the data collection step, or at least parts of it, could increase efficiency and profits for growing operations. Analysis of plant growth incorporating factors of the plants' environment like soil properties have been developed before, but there are no tools specifically meant to work together for the purpose of combined data collection as a set of mobile platforms. Finally, the fact that many of the products available on the market are based on similar technologies requiring a large investment which poses a key limitation to the adoption of technologies for proximal sensing.

1.3 Objective of the research

The ultimate goal of this research was to evaluate the viability of combining data about the crop and its environment gathered using semi-automated platforms.

The specific objectives of this research are summarized as follows:

1. To develop a High-Throughput Plant Phenotyping (HTPP) platform (Chapter 3).
2. To develop a Proximal Soil Sensing (PSS) platform (Chapter 4).
3. To analyze the data gathered from the developed platforms (Chapter 5).

CHAPTER 2 General literature review

A review of the literature regarding each of the specific objectives is presented in this chapter. The review is not meant to be exhaustive, but rather illustrative of some of the main ideas pertaining to the development of the systems described in this work. A gap in knowledge related to the combined collection of data from mobile platforms for plants, atmosphere and soil was identified.

2.1 High-Throughput Plant Phenotyping

A HTPP platform is a system to measure plant traits. In combination with genomics, plant phenotyping can be used to find QTL (Quantitative Trait Loci) i.e. measured plant traits can be linked with sections of DNA sequences. Properties that are commonly measured by HTPP platforms are the architecture of seeds, roots, storage organs, leaves, fruits, flowers, and entire canopies; processes like biomass productivity, photosynthesis, transpiration or yield formation; specific functions of organs or systems; and tolerance to different types of stress (Pieruschka & Schurr, 2019). To achieve this, different types of sensors are used, either to obtain direct measurements of the desired traits or to produce intermediate values that relate to them. Some of the most prominent types of sensors in HTPP are:

> "•Infrared thermography and imagery to scan temperature profiles/transpiration
>
> •Fluorescent microscopy/spectroscopy to assess photosynthetic rates
>
> •3D reconstruction to assess plant growth rate and structure
>
> •Light detection and ranging (LIDAR) to measure growth rates
>
> •Magnetic resonance imaging and positron emission tomography to measure growth patterns, root/leaf physiology, water relations, and/or assimilate translocation properties
>
> •Canopy spectral reflectance for monitoring dynamic complex traits

•Nuclear magnetic resonance for monitoring the structure of tissues, mapping water movements, and monitoring sucrose allocation

•Digital RGB imaging for recording data on various attributes of roots, shoots, leaves, seeds, and grains"

(Mir *et al.*, 2019)

In most cases, however, HTPP platforms rely on sensor fusion to provide more complex information than any particular individual sensor could. While this increases computational costs, it is generally preferred to overcommitting to one type of sensor. Given the focus on canopy measurements, distance sensors, multi- and hyper-spectral sensors, and cameras combined with computer vision algorithms were highlighted. These sensors are very practical for aboveground measurements.

First, multispectral sensing refers to measuring the absorption or reflection of light at specific wavelengths, providing information about the physical and chemical properties of the plant tissue. Examples of this technology included the use of Vegetation Indexes, like NDVI, related to plant vigor or nitrogen uptake. Hyperspectral sensing takes this concept further by analyzing hundreds of different wavelengths. While hyperspectral sensors could provide more complex data, in general, it is less reliable and durable, as well as more expensive. However, the biggest difference is in terms of their signal-to-noise ratio, with multispectral sensors typically having higher values.

Second, distance sensors like laser rangefinders and ultrasonic time-of-flight sensors are useful for determining the size and shape of the canopy, especially when multiple degrees of freedom are considered. Tracking the morphology of the canopy allows for a greater understanding of the growth processes and more accurate yield monitoring.

Cameras combined with computer vision software extract information from the color of the pixels and from the size and shape of the canopy, as was performed with distance sensors, especially when multiple cameras were used. Cameras are commonly affordable, with most of the cost related to the development of the image processing algorithm. This is also true for configurations with multiple cameras, where the computational cost of stereo-vision algorithms increased considerably.

An important concern for the type of sensor to be used is the need to measure the environment as part of the experimental design of HTPP platforms. This includes measurements of air temperature, pressure and moisture, as well as other measurements such as solar irradiance. This process, sometimes referred to as envirotyping, provides environmental information that maintains the rigor of the experimental statistics, working as metadata for the data collection step. It could also be useful to test for robustness of plant performance under different environmental conditions (Mir *et al.*, 2019).

An additional component in the sensor category is Global Navigation Satellite System (GNSS). These are central to georeferencing the measurements taken, providing the ability to create maps of the field by the measured characteristics and to study the spatial relationship of the traits.

Different institutions and researchers have developed HTPP platforms, some of which are offered commercially e.g. LemnaTec, Phenokey, PhenoSpex, Photon System Instruments, Wiwam, and We Provide Solutions (Gehan & Kellogg, 2017)[1]. Companies may also offer HTPP as a service e.g. KeyGene's PhenoFab. Despite the existence of these solutions, the adoption of HTPP platforms has not achieved its full potential. This is partly because most of the current offerings are focused on controlled environments, like greenhouses. Furthermore, as evidenced in the work of Reynolds *et al.* (2019), there are significant costs related to the implementation of phenotyping platforms. Despite the cost, even the phenotyping systems requiring the highest human/capital investments can have high cost-benefit when genetic gains are significantly boosted or when the achieved understanding of physiology and genetics of the breeding germplasm can be developed into novel, more accessible phenotyping assays, and potentially molecular markers for rapid screening (Reynolds *et al.*, 2020).

For HTPP platforms based on ground-vehicles, full or partial autonomous driving is a feature with great potential. This kind of tool could help reduce the workload of operators, while making the entire process faster and more efficient. Many implementations of autonomous navigation for HTPP platforms leverage local information, like crop rows, to improve the localization estimation of the platform (Schwarz *et al.*, 2013).

2.2 Proximal Soil Sensing

PSS is an emerging area of technologies that enables the determination of physical and chemical

[1] Mention of a trade name, proprietary product, or company name is for presentation clarity and does not imply endorsement by the authors or McGill University, nor the exclusion of other products that may also be suitable.

soil characteristics when sensors are placed in proximity to the soil being tested. “Proximal Soil Sensing provides soil scientists with an effective approach that can be used to learn more about the soil and so improve management in terms of economic benefits to the farmer and reduced environmental impacts from farming activities” (Viscarra-Rossel & Adamchuk, 2013).

The most common tools used in PSS are ground-penetrating radar, electromagnetic induction, and electrical resistivity, according to Adamchuk *et al.* (2017). In Adamchuk *et al.* (2018), some of the limitations of soil spectroscopy regarding the prediction of plant available soil nutrients are noted. Nonetheless, the authors recognize the technology’s potential. According to Viscarra-Rossel *et al.* (2011), in the past couple of decades, the use of spectroscopy in soil is rising due to its speed, low cost, and simplicity. Optical reflectance spectroscopy is mentioned as having potential for expanded use, especially in high-intensity surveys. There is interest from the industry in PSS platforms that use Vis-NIR (Visible and Near InfraRed) spectroscopy for soil mapping, as evidenced by the availability of OpticMapper and P4000 (Veris Technologies, Inc., Salinas, KS, USA), as well as SoilReader (SoilReader, Winnipeg, Canada) and the ChrysaLabs probe (ChrysaLabs, Inc., Montréal, Canada).

Many applications of PSS include accessing below the superficial top layer to characterize the soil at different depths. A typical example would be the insertion of a probe connected to an EC-meter. For this, it is often desirable to have opened a hole before the probe goes in. Drilling is just one of the possible options to open holes in soil (Dainese & Ercoli Finzi, 2006). Though their application is in space, they chose a drill because of the ease of controlling depth and light-weight constraints, both desired attributes shared by a mobile platform for PSS. Ye *et al.* (2015) showed that the geometric parameters of the auger drill bit can be optimized to reduce power consumption. A control loop in the motor driving the drill can be added to control variable rotational and penetration speeds and provide a robust response to changes in soil characteristics, like hardness and compaction (Olsson, Robertsson, & Johansson, 2015). Guaranteeing the straightness and smoothness of the walls of the hole is desirable, so that the stresses on the sensor probe and interruptions in the area of contact between the probe surface and the soil are minimized.

2.3 Data processing

Given the large amount of data that could become available when the technologies mentioned in the previous sections become mainstream, it is important to discuss the potential ways it could be used. In Ranjan *et al.* (2019), yield is predicted from a group of Vegetation Indexes (VIs) in relation to different management practices i.e. strip tillage and irrigation. In Tagarakis *et al.* (2019) crop yield is explained by soil and irrigation data through a crop growth model. This can be considered an example of on-farm experimentation with geo-referenced data (Piepho *et al.*, 2011). The simultaneous evaluation of crop and soil data allows for better insight into the processes of plant growth, as in von Hebel *et al.* (2018), where the combined use of crop and soil data is reported to determine the effect of subsurface soil properties on plant performance by studying the Pearson correlation coefficient.

Because the interpretation of soil spectroscopy is not necessarily straightforward, multivariate methods are frequently used to create prediction models (Carra *et al.*, 2019). Examples of the mentioned methods related with Vis-NIR spectroscopy are found in Xu *et al.* (2018). With the recent advances in Artificial Intelligence and Machine Learning algorithms, more tools are available to extract the information from different data sources. Many of these techniques when used for the prediction of chemical properties are covered by the discipline of chemometrics, which uses mathematical, statistical and computer applications to reveal the hidden information from chemical analyses (Barra *et al.*, 2021). Organic matter, pH, P, K, Fe, Ca, Na, Mg, and CEC are some of the soil properties frequently reported in the literature of soil spectroscopy (Stenberg *et al.*, 2010).

Machine Learning methods are also used to model yield using as explanatory variables soil properties, environmental data, and a crop spectral measurement; multiple methods were compared, among others, counter-propagation artificial neural networks and support-vector machine, which produced accuracies of 0.89 and 0.95, respectively (Nyeki *et al.*, 2019). This difference in performance, given the same datasets, is evidence that the selection of the analysis method is critical. Similarly, soil properties are used as explanatory variables to model the yield of a sugarcane field using a Random Forests technique (Sanches, Graziano Magalhaes, & Junqueira Franco, 2019). The way to move this idea further would be to make more accurate evaluations of treatments or varieties by differentiating accurately the effect of soil and weather on the plant phenotypical traits from the results of the studied treatment and its interactions with other factors.

A model was built to predict yield and other crop traits using canopy cover estimated based on images from a digital camera (Hoyos-Villegas *et al.*, 2014). The predictor variables in this model combine different measurement times, which shows how the relative importance of certain variables changes throughout the growing period. While not being measured, soil data is included in the setup in the form of using different topsoil depths as the treatment that distinguishes different groups to represent variable soil water holding capacities, and thus, influenced crop water deficit stress.

The previously mentioned studies use relatively simple models to describe the relationship between a set of properties of interest and certain predictor variables. These relationships are often limited in terms of exportability and ability to generalize. A more holistic approach is referred to as crop modelling, where the entirety of the plant processes is considered and simulated. APSIM (Holzworth *et al.*, 2014) and DSSAT (Hoogenboom *et al.*, 2019) are some examples of crop models. Soil and environmental data are crucial in this type of modelling where plant development uses information on moisture availability by simulating storage and movement of water in the root zone, based on soil physical properties (Kuang *et al.*, 2012). It is possible to use the HTPP and PSS platforms mentioned in the previous sections to provide data for the crop models and inversely, to use the insight from the crop models to guide the behaviour of the platforms.

Thus far, the reported literature discusses analysis performed as a step separate from the data collection. But it is also possible to consider real-time processing of crop information, for example, in the context of variable rate spraying, where the rate is adjusted based on canopy measurements (Zaman, Schumann, & Miller, 2005). There is the potential to improve the efficacy of this kind of implementation by integrating data about the initial nutrient content in the soil. Finally, during the data collection step, the gathered information can be used as it is being collected to prioritize scouting of regions of the field where more uncertainty has been detected. Examples are found in Liu, Crowe, & Roberge (2009), even if limited to the topography of the field, and Oger, Vismara,

& Tisseyre (2019), where the locations to collect yield samples were selected optimally based on NDVI measurements. A combination of these approaches would be ideal.

References for Chapter 2

Adamchuk, V., Allred, B., Doolittle, J., Grote, K., & Viscarra-Rossel, R. (2017). Tools for Proximal Soil. In C. Ditzler, K. Scheffe, & H. Monger (Eds.), *Soil Survey Manual* (pp. 355-394). Washington: USDA.

Adamchuk, V., Ji, W., Viscarra-Rossel, R., Gebbers, R., & Tremblay, N. (2018). Proximal Soil and Plant Sensing. In D. Shannon, D. Clay, & N. Kitchen (Eds.), *Precision Agriculture Basics* (pp. 123-144). Madison, WI: ASA, CSSA, and SSSA.

Barra, I., Haefele, S. M., Sakrabani, R., & Kebede, F. (2021). Soil spectroscopy with the use of chemometrics, machine learning and pre-processing techniques in soil diagnosis: Recent advances – A review. *Trends in Analytical Chemistry, 135*, 116166.

Carra, J. B., Fabris, M., Dieckow, J., Brito, O. R., Vendrame, P. R. S., & dos Santos, L. M. (2019). Near-infrared spectroscopy coupled with chemometrics tools: a rapid and non-destructive alternative on soil evaluation. *Communications in Soil Science and Plant Analysis*, 50:4, 421-434.

Dainese, C., & Ercoli Finzi, A. (2006). SD2 - A cometary soil drill and sampler device. *Aerotecnica Missili e Spazio, 85*, 174-180.

Gehan, M. A., & Kellogg, E. A. (2017). High-throughput Phenotyping. *American Journal of Botany, 104*(4), 505-508.

Holzworth, D. P., Huth, N. I., deVoil, P. G., Zurcher, E. J., Herrmann, N. I., McLean, G., Chenu, K., van Oosterom, E. J., Snow, V., Murphy, C., Moore, A. D., Brown, H., Whish, J. P. M., Verrall, S., Fainges, J., Bell, L. W., Peake, A. S., Poulton, P. L., Hochman, Z., Thorburn, P. J., Gaydon, D. S., Dalgliesh, N. P., Rodriguez, D., Cox, H., Chapman, S., Doherty, A., Teixeira, E., Sharp, J., Cichota, R., Vogeler, I., Li, F. Y., Wang, E., Hammer, G. L., Robertson, M. J., Dimes, J. P., Whitbread, A. M., Hunt, J., van Rees, H., McClelland, T., Carberry, P. S., Hargreaves, J. N. G., MacLeod, N., McDonald, C., Harsdorf, J., Wedgwood, S., & Keating, B. A. (2014). APSIM – Evolution towards a new generation of agricultural systems simulation. *Environmental Modelling & Software, 62*, 327-350.

Hoogenboom, G., Porter, C. H., Boote, K. J., Shelia, V., Wilkens, P. W., Singh, U., White, J. W., Asseng, S., Lizaso, J. I., Moreno, L. P., Pavan, W., Ogoshi, R., Hunt, L. A., Tsuji, G. Y., & Jones, J. W. (2019). The DSSAT crop modeling ecosystem. In K. J. Boote, *Advances in Crop Modeling for a Sustainable Agriculture* (pp. 173-216). Cambridge, UK: Burleigh Dodds Science Publishing.

Hoyos-Villegas, V., Houx, J., Singh, S., & Fritschi, F. (2014). Ground-Based Digital Imaging as a Tool to Assess Soybean Growth and Yield. *Crop Science, 54*, 1756-1768.

Kuang, B., Mahmood, H. S., Quraishi, M. Z., Hoogmoed, W. B., Mouazen, A. M., & van Henten, E. J. (2012). Chapter four - Sensing Soil Properties in the Laboratory, In Situ, and On-Line: A Review. In D. L. Sparks, *Advances in agronomy* (Vol. 114, pp. 155-223). Academic Press.

Liu, L., Crowe, T. G., & Roberge, M. (2009). Sensor-based scouting algorithms for automatic sampling tasks in rough and large unstructured agricultural fields. *Transactions of the ASABE, 52*(1), 285-294.

Mir, R. R., Reynolds, M., Pinto, F., Khana, M. A., & Bhata, M. A. (2019). High-throughput phenotyping for crop improvement in the genomics era. *Plant Science, 282*, 60-72.

Nyeki, A., Kerepesi, C., Daroczy, B., Benczur, A., Milics, G., Kovacs, A. J., & Nemenyi, M. (2019). Maize yield prediction based on artificial intelligence using spatio-temporal data.

12th European Conference on Precision Agriculture (pp. 1011-1017). Montpellier, France: Wageningen Academic Publishers.

Oger, B., Vismara, P., & Tisseyre, B. (2019). Combining target sampling with route-optimization to optimise yield estimation in viticulture. *12th European Conference on Precision Agriculture* (pp. 173-179). Montpellier, France: Wageningen Academic Publishers.

Olsson, T., Robertsson, A., & Johansson, R. (2015). Flexible force control for accurate low-cost robot drilling. *Proceedings 2007 IEEE International Conference on Robotics and Automation.* Rome, Italy, 2007, pp. 4770-4775.

Piepho, H.-P., Richter, C., Spilke, J., Hartung, K., Kunick, A., & Thole, H. (2011). Statistical aspects of on-farm experimentation. *Crop & Pasture Science, 62*(9), 721-735.

Pieruschka, R., & Schurr, U. (2019). Plant Phenotyping: Past, Present, and Future. *Plant Phenomics*, vol. 2019, Article ID 7507131.

Ranjan, R., Chandel, A. K., Khot, L. R., Bahlol, H. Y., Zhou, J., Boydston, R. A., & Miklas, P. N. (2019). Irrigated pinto bean crop stress and yield assessment using ground based low altitude remote sensing technology. *Information Processing in Agriculture, 6*, 502-514.

Reynolds, D., Baret, F., Welcker, C., Bostrom, A., Ball, J., Cellini, F., Lorence, A., Chawade, A., Khafif, M., Noshita, K., Mueller-Linow, M., Zhou, J., & Tardieuc, F. (2019). What is cost-effective phenotyping? Optimizing costs for different scenarios. *Plant Science, 282*, 14-22.

Reynolds, M., Chapman, S., Crespo-Herrera, L., Molero, G., Mondal, S., Pequeno, D. N. L., Pinto, F., Pinera-Chavez, F. J., Poland, J., Rivera-Amado, C., Saint Pierre, C., & Sukumaran, S. (2020). Breeder friendly phenotyping. *Plant Science, 295*, 110396.

Sanches, G. M., Graziano Magalhaes, P. S., & Junqueira Franco, H. C. (2019). Site-specific assessment of spatial and temporal variability of sugarcane yield related to soil attributes. *Geoderma*(334), 90-98.

Schwarz, H. P., Rüger, P., Kicherer, A., & Töpfer, R. (2013). Development of an autonomous driven robotic platform used for ht-phenotyping in viticulture. *Synergy and Technical Development (Synergy2013).* Gödöllő, Hungary, 13-19 October 2013.

Stenberg, B., Viscarra-Rossel, R. A., Mouazen, A. M., & Wetterlind, J. (2010). Visible and Near Infrared Spectroscopy in Soil Science. *Advances in Agronomy, 107*, 163-215.

Tagarakis, A. C., Mimic, G., Vaessen, H. M., Rodriguez-Moreno, F., van Evert, F. K., & Ciric, V. (2019). Using the WOFOST crop growth model to assess within-field yield variability. *12th European Conference on Precision Agriculture* (pp. 91-97). Montpellier, France: Wageningen Academic Publishers.

Viscarra-Rossel, R. A., Adamchuk, V. I., Sudduth, K. A., McKenzie, N. J., & Lobsey, C. (2011). Proximal soil sensing: An effective approach for soil measurements in space and time. In D. Sparks, *Advances in Agronomy. Vol 113* (pp. 237-282). Academic Press.

Viscarra-Rossel, R., & Adamchuk, V. I. (2013). Proximal soil sensing. In M. Oliver, T. Bishop, & B. Marchant, *Precision agriculture for sustainability and environmental protection* (pp. 99-118). New York: Earthscan.

von Hebel, C., Matveeva, M., Verweij, E., Rademske, P., Kaufmann, M. S., Brogi, C., Vereecken, H., Rascher, U., & van der Kruk, J. (2018). Understanding Soil and Plant Interaction by Combining Ground-Based Quantitative Electromagnetic Induction and Airborne Hyperspectral Data. *Geophysical Research Letter*, 45, 7571-7579.

Xu, S., Zhao, Y., Wang, M., & Shi, X. (2018). Comparison of multivariate methods for estimating selected soil properties from intact soil cores of paddy fields by Vis-NIR spectroscopy. *Geoderma, 310*, 29-43.

Ye, T., Dewei, T., Zongquan, D., Shengyuan, J., & Qiquan, Q. (2015). Drilling power consumption and soil conveying volume performances of lunar sampling auger. *Chinese journal of mechanical engineering, 28*, 451-459.

Zaman, Q. U., Schumann, A. W., & Miller, W. M. (2005). Variable rate nitrogen application in Florida citrus based on ultrasonically-sensed tree size. *Applied Engineering in Agriculture, 21*(3), 331-335.

Connecting text to Chapter 3

From the literature review, a knowledge gap was identified regarding the development of mobile platforms for combined data collection that can facilitate the process of building models of plant phenotypical traits. The necessity to combine phenotypical and environmental information to support on-farm experimentation was described in the previous chapter. Addressing this issue starts with the development of a HTPP platform. Chapter 3 is related to the first objective as listed in Chapter 1. In this chapter, the development of a mobile platform with hardware and software components is reported. Additionally, testing was conducted in comparison with a handheld unit. Initial design considerations were published at the conferences listed below and the final research findings are under preparation for a journal publication.

R. Buelvas, V.I. Adamchuk, J. Lan, V. Hoyos-Villegas, A. Whitmore, & M. Stromvik. (2021). Evaluation of a quick-install rapid phenotyping system. *Computers and Electronics in Agriculture*. (To be submitted)

V.I. Adamchuk, & **R. Buelvas**. (2020). Integration approach to proximal plant sensing. International Conference on Digital Technologies for Sustainable Crop Production. November 1-10, 2020

R. Buelvas, V.I. Adamchuk, A. Pouliot, A. Whitmore, & M. Stromvik. (2019). Development of a quick-install rapid phenotyping system. 2019 ASABE Annual International Meeting. Boston, MA, USA. July 7-10, 2019

CHAPTER 3 Development of High-Throughput Plant Phenotyping platform

Abstract

In recent years, HTPP platforms have been developed and used in greenhouses and other controlled environments. Yet, the need remains for similar systems with the ability to take measurements of plant traits in open fields to become as accessible. This paper presents the design and evaluation for a phenotyping system aiming to address this issue. A combination of ultrasonic and multispectral measurements of the crop canopy with diverse measurements of environmental conditions allows for the collection and processing of field data at a relatively low cost. The combination of these features makes for a complete system with the goal of mapping crop status across a field. In a field experiment, it was found that the system could cover 5,400 m^2/h, almost a 50-fold increase in throughput as compared with a manual setup with similar sensors. Proper use of this technology would support the study of plant responses to different treatments or stresses.

Keywords: Multispectral, Phenotyping, Sensing, Ultrasonic, Vegetation Index

3.1 Introduction

The main advantage of HTPP compared to other phenotyping approaches is the potential to increase the resolution of the data, both in spatial and temporal terms. By acquiring information about plant traits with higher spatial resolutions, it is possible to better understand within-field variability and move the focus to individual plants or even individual plant organs rather than entire canopies. The increased temporal resolution offers the possibility to study time interval-specific QTL and other time-dependant phenomena (Knoch *et al.*, 2020). This concern for spatial and temporal resolution is also evidenced in the selection of the platform where plant phenotyping takes place. While many works like Rossi *et al.* (2020) focus on HTPP platforms for controlled environments, there are also others using the HTPP perspective for outdoor use based on fixed and mobile structures. This includes ground-vehicles, UAVs, cable-mounted systems, or combinations of them (Mir *et al.*, 2019).

Plant breeding is a possible area where a system like the one proposed could provide benefits. This is a context where providing a framework for the screening and selection of different crop varieties is key in plant breeding programs. Developments in HTPP to quantitatively measure key traits will increase accuracy of the selection process while reducing costs. The increment in through-put can also be used to increment the size of the breeding program to enable higher selection intensity, another factor in the genetic gain per time as defined in Equation (3-1), where R_t is the genetic gain per time, i is the selection intensity, r is the selection accuracy, σ_A is the genetic variability, and y is the number of years per cycle (Araus *et al.*, 2018).

$$R_t = \frac{ir\sigma_A}{y} \qquad (3-1)$$

Inclusion of weather data in on-farm experimentation is a step to reduce the effect that factors other than the genome have on measured plant traits, referred to as envirotyping. In the breeding context, this approach could help speed up variety commercialization by increasing selection accuracy, improving multienvironmental trials, and optimizing variety evaluation (Xu *et al.*, 2017). Envirotyping is particularly desirable thanks to its benefits for the comparison of multi-year experiments and helps with the generalization potential of the experimental results. The most common types of sensing systems for this purpose rely on fixed structures, either as a weather station placed in a central location or as Wireless Sensor Networks (WSN) which can cover a wide area (Reynolds *et al.*, 2019). With the former, spatial variability of atmospheric conditions within the field is not taken into account, and with the latter, the installation cost scales up with the area covered as more nodes are required in the network. An economical alternative is to have a mobile platform which can take this type of environmental measurements, potentially complemented by a weather station.

HTPP platforms based on ground vehicles typically mount the sensors directly in the front or rear of the vehicle. This creates a problem whereby the vehicle is meant to be driven through, or above, the planted plots. This could be challenging even at the early stage of crops for several reasons i.e. the soil could be damaged by the vehicle, the status of the rows might not be appropriate for efficient vehicle circulation. Lateral booms reaching over the plots from a vehicle driven all-weather traffic lanes is one option to overcome this (Andrade-Sanchez *et al.*, 2014). In other words, to limit the vehicle to dedicated lanes and use a lateral boom that stretches over the crops.

The goal of this project was to design a HTPP platform for use in field conditions. The system should be able to take canopy measurements from different types of crops at an early stage and is based on the quick-install concept (Pouliot, 2016), so that users could mount the system on a vehicle they already own and unmount it if they need the vehicle for a different operation involving other attachments. The original design considerations were presented in further detail in Buelvas *et al*. (2019).

A system that combines phenotyping and envirotyping allows for the collection of data and provides a context at the same time. This moves the process a step closer to the ultimate goal of equipping all kind of growers with a tool for rapid, non-destructive, reliable, and affordable assessment of their crops. In this work, such a system is proposed and evaluated in a field planted with dry beans (*Phaseolus vulgaris*). The application of HTPP in beans and pulses is not typically found in the literature (Yang *et al.*, 2020).

3.2 Materials and methods

3.2.1 Electronic subsystem

The sensors included in the HTPP platform were:

- 6 multispectral sensors ACS-435 (Holland Scientific, Lincoln, NE, USA)
- 6 ultrasonic sensors ToughSonic 14 (Senix, Hinesburg, VT, USA)
- 2 RGB cameras C525 (Logitech, Lausanne, Switzerland)
- 2 environmental sensors DAS43X (Holland Scientific, Lincoln, NE, USA)
- 1 GNSS unit 19X (Garmin Ltd., Olathe, KS, USA)

Except for the GNSS unit, the set of sensors was divided in two identical halves, one for each side of the HTPP platform. Some of their parameters are described in Table 3-1. Besides the already mentioned sensors, other components in this subsystem were a laptop computer as the control terminal, two 8-Port USB to Serial Hubs (StarTech.com Ltd, London, ON, Canada), and two power banks of 20,000 mAh. The power banks, serial hubs, and their connectors were located inside plastic enclosures for additional protection. The computer ran the Graphical User Interface (GUI) for the operator to monitor and adjust the behavior of the system. The serial hubs are required to facilitate the connections of the previously mentioned sensors to the terminal. Figure 3-1 illustrates the connections between the different electronic components of the HTPP.

Table 3-1. Description of sensors included in platform

Sensor type	Measured variables	Parameters
Multispectral	NDVI, NDRE, CI*, proxy LAI, proxy CCC, proxy distance, Red, Red-Edge, and NIR *: Computed from Red-Edge and NIR	670nm red measurement band 730nm red-edge measurement band 780nm NIR measurement bands ~40° by ~10° field of view
Ultrasonic	Distance	0.1m-4.3m operating range 0.2% of range repeatability
Camera	Video	720p resolution 30fps 68° field of view
Environmental	Upwelling PAR, Downwelling PAR, Canopy temperature, Air temperature, Relative humidity, and Atmospheric pressure	±0.3°C air temperature accuracy ±4% FS relative humidity accuracy ±1.5% FS atmospheric pressure accuracy
GNSS	Longitude, Latitude, Altitude, Speed, and Heading	NMEA 0183 format <15m GPS position accuracy

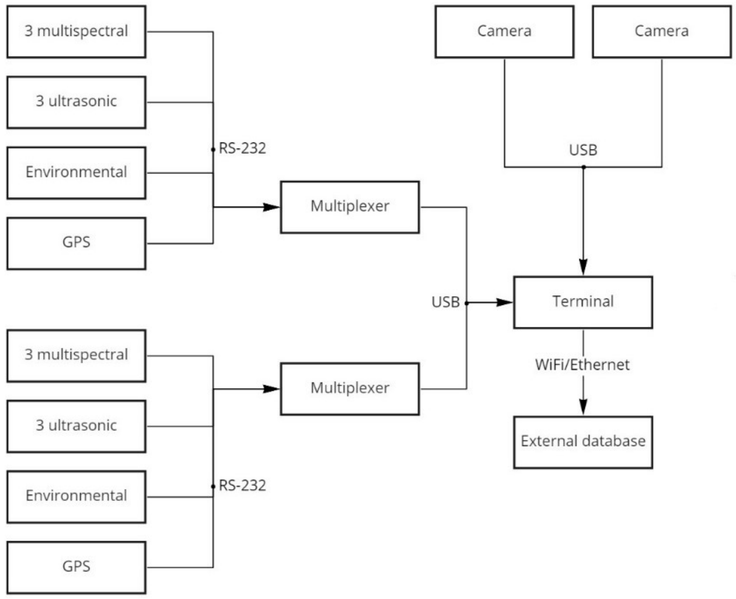


Figure 3-1. Block diagram of the HTPP platform

A handheld version of the system with a subset of its components was made at an earlier stage with a multispectral sensor, an ultrasonic sensor, a power bank, and a tablet Yuma 2 (Trimble Inc., Sunnyvale, CA, USA) with its GPS receiver. This handheld setup was used both to prototype the operation of the HTPP platform and to compare ergonomic performance. One difference between this setup and the vehicle-mounted HTPP platform is that for the multispectral sensor, the older model ACS-430 (Holland Scientific, Lincoln, NE, USA) was used rather than the ACS-435. Both models measure reflectance at 3 bands: NIR, Red-Edge, and Red, as well as two VIs: NDVI and NDRE, shown in Equations (3-2) and (3-3), respectively. The ACS-435 also provides proxy Distance, proxy LAI (Leaf Area Index), and proxy CCC (Canopy Chlorophyll Content). In both cases an additional VI was computed: CI (Chlorophyll Index), as defined by Equation (3-4).

$$NDVI = \frac{\rho_{NIR} - \rho_{Red}}{\rho_{NIR} + \rho_{Red}} \qquad (3-2)$$

$$NDRE = \frac{\rho_{NIR} - \rho_{RedEdge}}{\rho_{NIR} + \rho_{RedEdge}} \qquad (3-3)$$

$$CI = \frac{\rho_{NIR}}{\rho_{RedEdge}} - 1 \qquad (3-4)$$

NDVI is relevant because it is the VI with the most widespread usage. NDRE is a similar measure that helps overcome problems with saturation at late phenological stages, one of the shortcomings of NDVI. While both NDVI and NDRE are popular for their general-purpose applicability, CI was chosen as a more focused VI to complement the other two, which correlates with total chlorophyll content of the leaves.

The properties measured by the environmental sensor DAS43X were air temperature, air humidity, atmospheric pressure, incident Photosynthetically Active Radiation (PAR), and reflected PAR. The ultrasonic sensor measured distance between itself and the target canopy, which can then used to compute plant height.

3.2.2 Mechanical subsystem

A handheld version of the HTPP platform with a subset of the components was built at an earlier stage to compare the ergonomic performance. The basic structure imitates a staff with the sensor near the top. Figure 3-2 illustrates this handheld setup.

Figure 3-2. Picture of handheld setup

Figure 3-3 illustrates the concept of the system mounted on a sprayer boom attached to an agricultural vehicle. The CAD model of the tractor was provided by Marko Voigtländer and that of the sprayer by Shaffic Ssenyimba through GrabCAD (GrabCAD, Inc., Cambridge, MA, USA). The mechanical design was made using Autodesk Inventor 2018 (Autodesk, Inc., San Rafael, CA, USA).

Figure 3-3. CAD of conceptual design

For the vehicle-mounted setup, the main structural component of the mechanical subsystem is a bar clamp with two swivel pads on each jaw pivot to grip nearly any shape. This provides confidence in the mounting capabilities of the brackets on different types of vehicles or on sprayer booms without prior knowledge of their exact geometry. Sensing units can be mounted on each side of the sprayer boom or similar toolbar using aluminum L-brackets bolted to the bar clamp.

The vehicle, a Gator 850D XUV (John Deere, Moline, IL, USA), was equipped with a custom horizontal beam, made by mounting aluminum bars with steel L-brackets to its trunk, so that they could extend 1.5 m to the sides of the vehicle. A steel thread connected to the furthermost end of the beam was used to keep the beam bars straight. A small winch was used to tighten the steel thread. 3D printed brackets were made to attach cameras to the top corners of the frame of the Gator, as well as the GNSS antenna. Figure 3-4 shows a picture of the system.

Figure 3-4. Picture of vehicle-mounted HTPP setup

3.2.3 Software subsystem

A Python (Python Software Foundation, Beaverton, OR, USA) script relying on open-source libraries was made to locally log the measurements from all sensors into text files and display the UI. Within it, there is information for the operator about the sensors that are connected, a map tracking the location of the vehicle, video streaming of the cameras, and real-time plots of the sensor readings. A bar of tabs was used to change between the plots of the several measured variables. Certain settings could be modified, like the spacing between the sensors and it offered the ability to disengage any subset of the sensors. Figure 3-5 shows a screenshot of the main window. The handheld version of the platform ran a preliminary version of the same code with reduced functionalities.

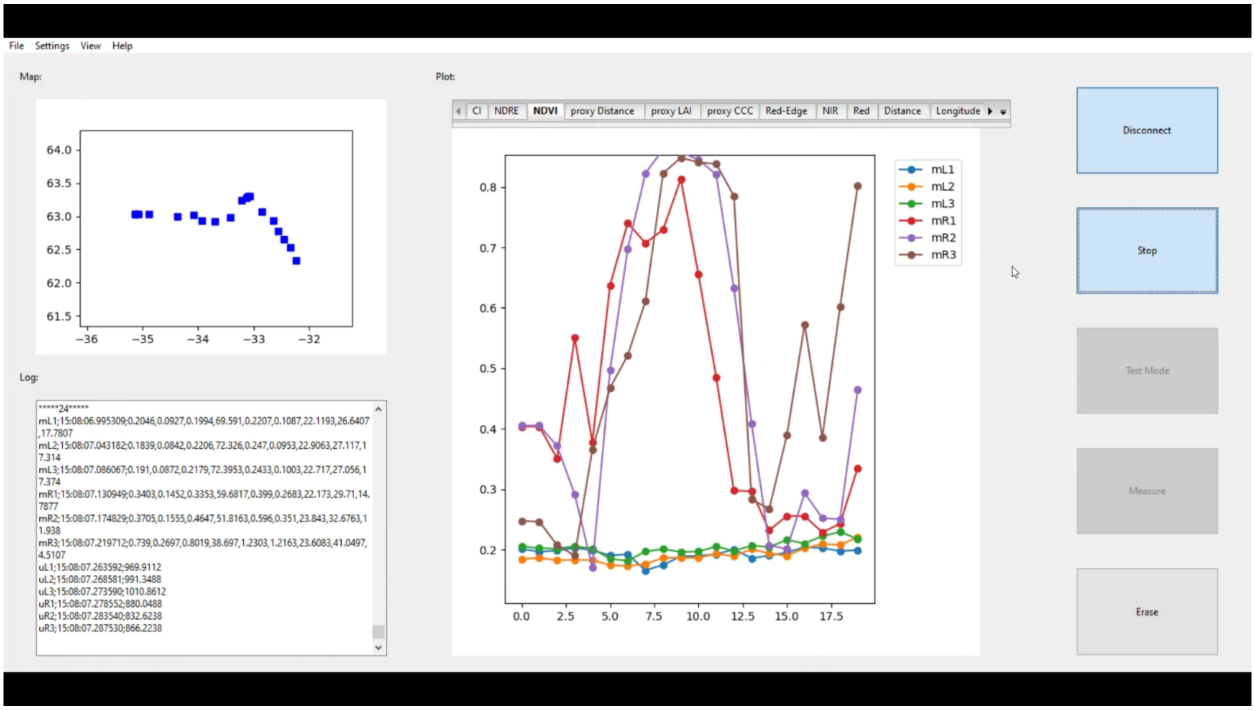


Figure 3-5. Screenshot of GUI (main window)

3.2.4 Experimental design

A first experiment with the handheld system took place during the summer 2019. The plants were sown on June 21st, 2019 at the Emile A. Lods Agronomy Research Center of Macdonald Campus, McGill University, divided in 22 plots, arranged in a grid of 2 by 11. 11 different bean varieties/cultivars (*Phaseolus vulgaris*) were grown, with 2 replicates for each. The varieties were Apex, Argosy, Calmant, Compass, Dresden, Knight Rider, Majesty, Mast, Nautica, Red Rider, and Sheek. In each plot, there were 4 crop rows. Measurements were taken on each plot on the dates listed in Table 3-2, including a geo-reference and a timestamp. For each plot and date, 8 locations were selected as spread out as possible to collect the data. All plots were exposed to the same management practices and were the same size (4.5 m by 1.5 m, for an area of 6.75 m^2). The plants were harvested on October 7th, 2019 and the grain yield was measured. Finally, Figure 3-6 shows a picture of the typical plots used in the experiments.

Table 3-2. Dates of data collection for first experiment

Date	Days After Seeding	Start time	End time
July 5th, 2019	14	12:31	14:12
July 25th, 2019	34	16:07	17:32
August 20th, 2019	60	14:20	16:05
September 18th, 2019	89	13:03	14:43

Figure 3-6. Picture of typical plot

A second experiment with the vehicle-mounted HTPP platform was performed during the summer 2020. This time, plants were sown on June 1st, 2020 divided among 242 plots, arranged in a grid of 11 x 22. This experiment took place in a different field of the Emile A. Lods Agronomy Research Center of Macdonald Campus, McGill University. Thirty-six different bean varieties were grown, and two different treatments were applied: row spacing being either 22 in. (559 mm) or 30 in. (762 mm), and the plant density 10,000 plants/acre (~24,700 plants/ha) or 15,000 plants/acre (~37,100 plants/ha). Again, all plots were the same size (5 m by 3 m, for a total area of 15 m^2) and included 4 crop rows within. While collection of data started on July 28th, 2020, the first set of dates for surveying was dedicated to fine-tuning some of the system parameters. Table 3-3 illustrates the dates of data collection that were used for the analysis. On these dates, the vehicle-mounted HTPP platform would drive through the alleys between plots as close to a constant speed of 0.8 m/s as was possible for the operator and collect measurements on-the-go from each side. The sampling rate was 165 Hz, so a new sensor reading was processed every 6 minutes. Cameras were set to record at about 3 fps. The plants were harvested on October 6th, 2020 and the grain yield was measured. For each plot, only the 2 innermost crop rows were counted for the yield measurement.

Table 3-3. Dates of data collection for second experiment

Date	Days After Seeding	Start time	End time
August 17th, 2020	77	13:04	13:49
August 20th, 2020	80	15:47	16:28
August 24th, 2020	84	12:49	13:29
August 27th, 2020	87	14:27	15:10
August 31st, 2020	91	13:32	14:13
September 3rd, 2020	94	15:35	16:15

Figure 3-7 illustrates the location of each plot in the field. Figures 3-8 and 3-9 show the applied treatments in terms of management practices. Figure 3-10 illustrates the number of plots that were assigned to each crop variety. This histogram shows that most varieties only have 1 or 2 replicates, while a few have more than 10. As shown in the figure, a split was made along that line into two datasets A and B. Only the 9 varieties in dataset B were used to perform statistical analysis, in order not to have big differences in the number of observations.

For both experiments, the data was saved as text files, which were parsed to comma-separated values file format by a Python script and then imported to MATLAB (MathWorks, Natick, MA, USA) for additional processing.

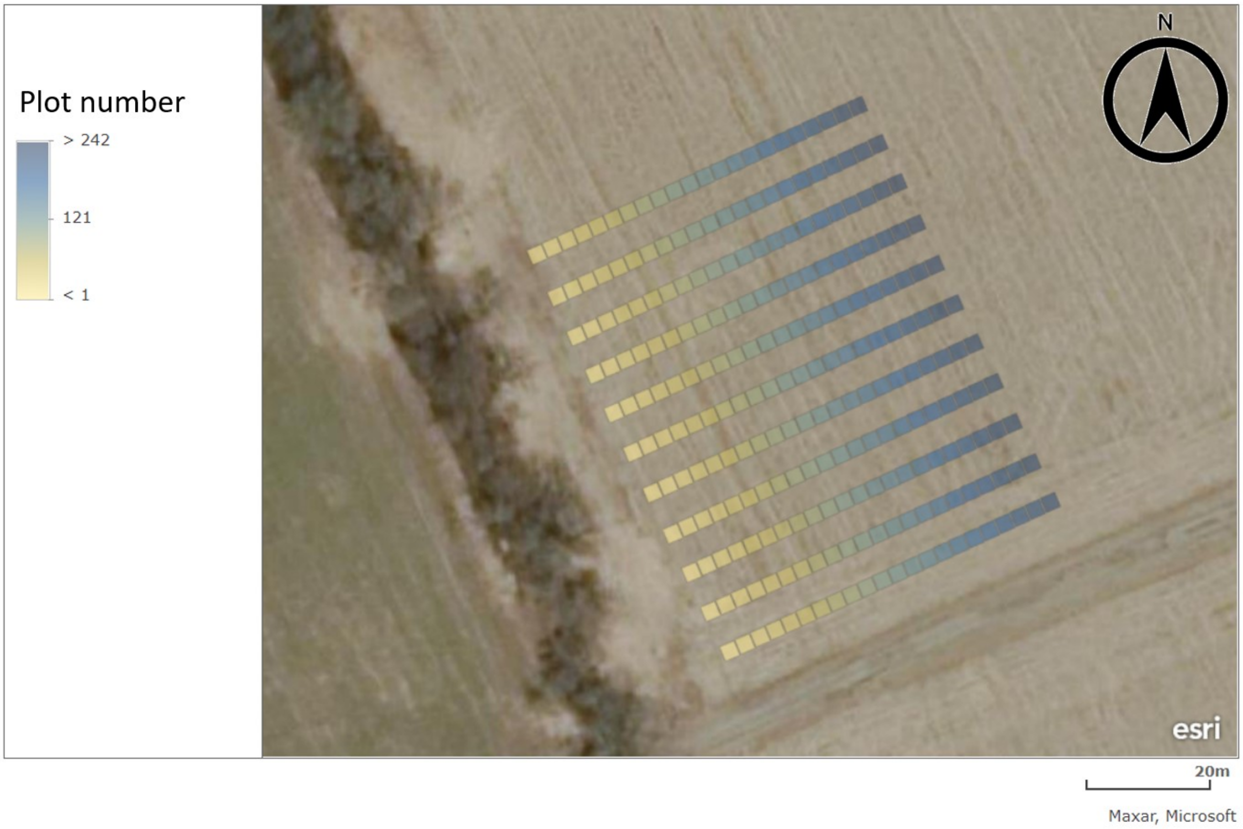


Figure 3-7. Map of plot count

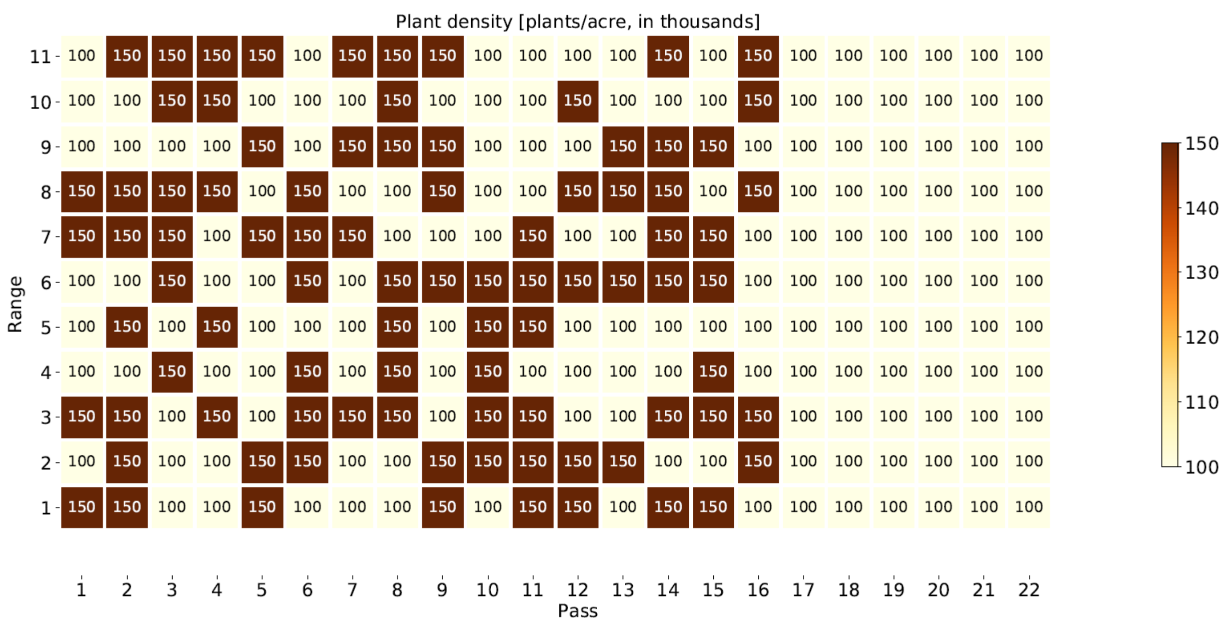


Figure 3-8. Map of the first treatment: plant density

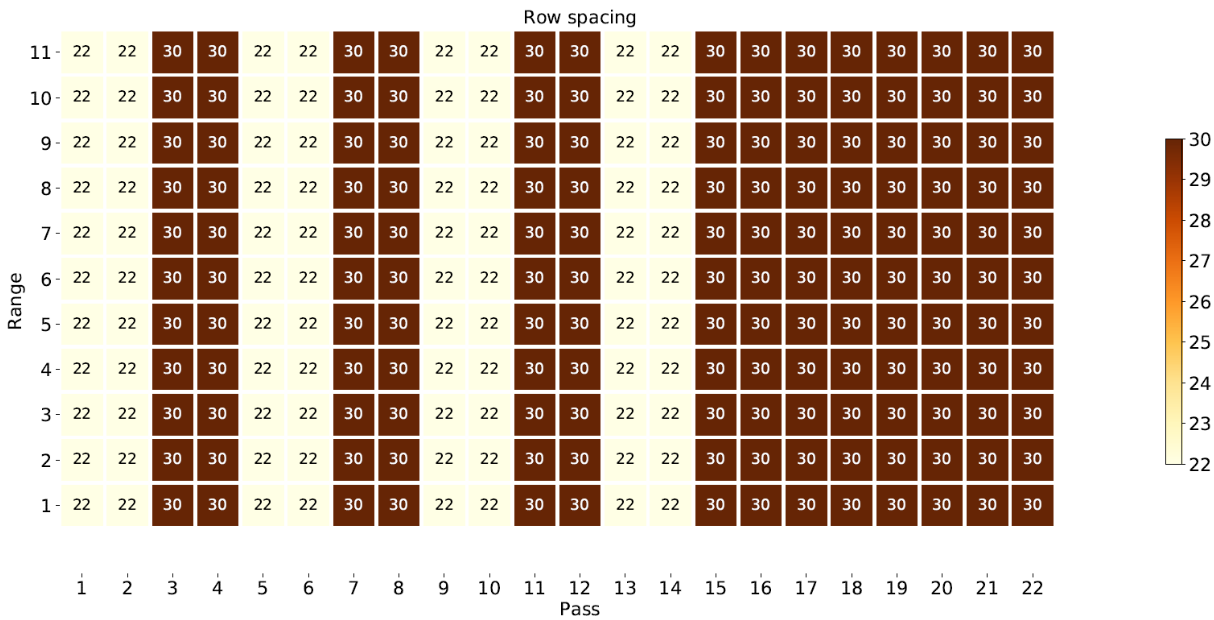


Figure 3-9. Map of the second treatment: row spacing

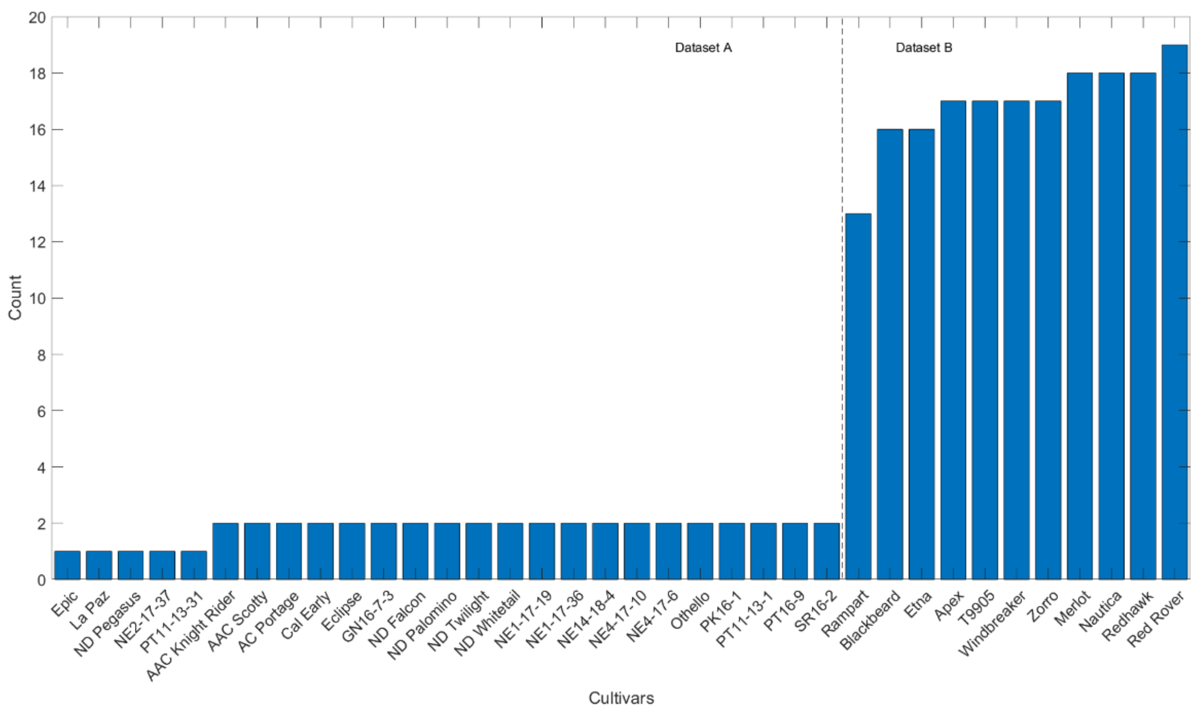


Figure 3-10. Histogram of replicates per bean variety/cultivar

As part of the exploratory data analysis, descriptive statistics were computed, assumption of normal distribution evaluated, and repeated measures ANOVA and Tukey's tests performed. The public records of a nearby weather station operated by Environment and Climate Change Canada at coordinates 45°25'38" N, 73°55'45" W (about 700 m away from the locations of both fields) were referenced. Figure 3-11 shows that both experiments were conducted under similar weather conditions.

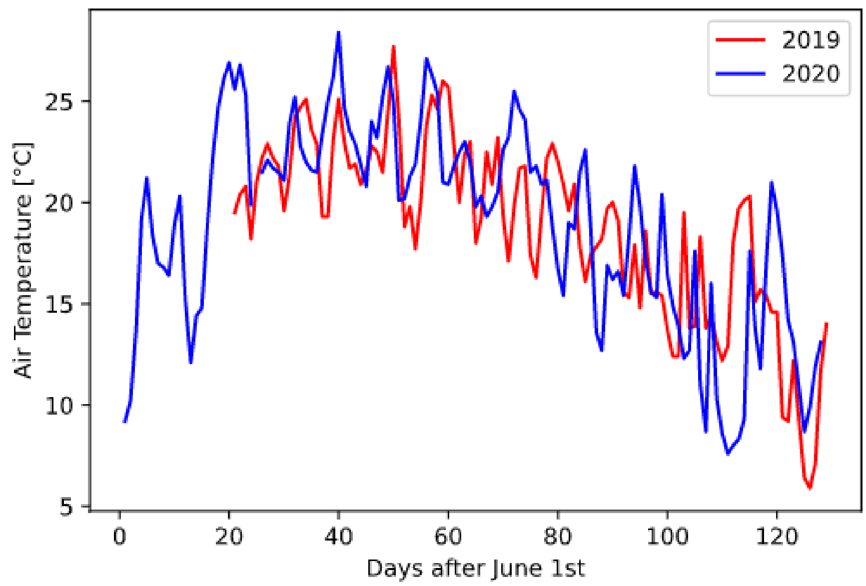


Figure 3-11. Daily average air temperature for dates of both experiments

3.3 Results and discussion

3.3.1 Handheld experiment

Figure 3-12 shows a map of NDRE points over a satellite image of the field as an illustration of the spatial variability of the measurements. Figure 3-13 illustrates the probability distribution of each variety for a given date under the assumption that they are normally distributed. From this plot, it is hard to tell which type of bean is the best variety, but it is possible to say that Nautica, Sheek, and Compass are underperforming compared to the rest. The results of ANOVA with repeated measurements over time are shown in Appendix 4. This latter test shows that both time and the interaction between the variety and time are significant for all the evaluated VIs. When considering the variety as a discrete factor that explains variability in the measured plant traits, its effect varies with time.

Figure 3-12. Map of NDRE for the first date

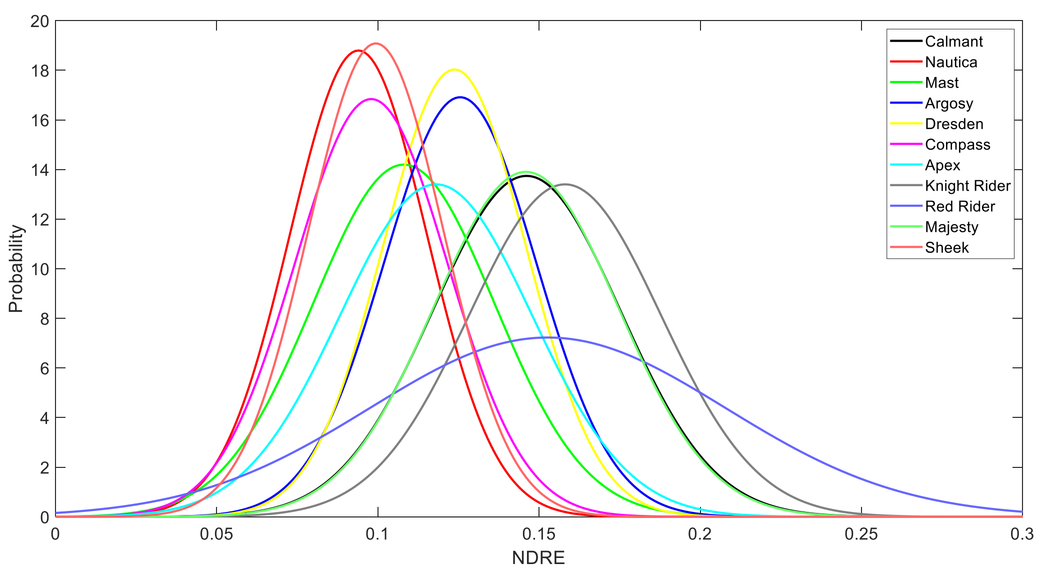


Figure 3-13. Probability density functions of NDRE by crop variety at the last date assuming normal distribution

Figures 3-14 to 3-15 go further by applying Tukey's test for each date to determine the differences between varieties which are statistically significant. This was done separately for each of the measured phenotypical variables, and the plots show NDVI and CI as examples. The error bars indicate the confidence intervals, and when they do not overlap, the differences are statistically significant. Frequently, it is possible to find such differences among the best and worst performing varieties for specific dates. Another important insight is that all VIs will generally agree on ranking the varieties from highest to lowest for the same date.

Using a model is a way of addressing the different responses through time. Given the limited number of points, a quadratic polynomial function was chosen. The models were fitted for each pair of variety and VI, in the form described by Equation (3-5), where θ represents the set of parameters a, b, and c, which depend on the variety, and ε is the random error for each observation.

$$VI = f\big(t; \theta(variety)\big) + \varepsilon = a * t^2 + b * t + c + \varepsilon \qquad (3-5)$$

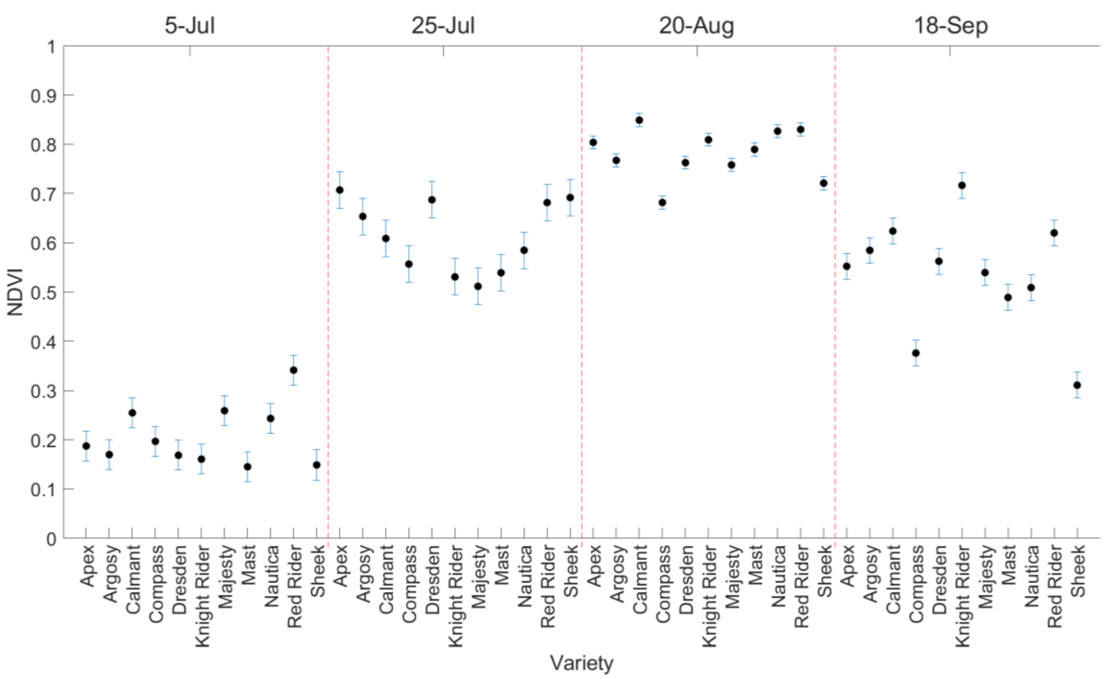


Figure 3-14. Results of Tukey's test for NDVI

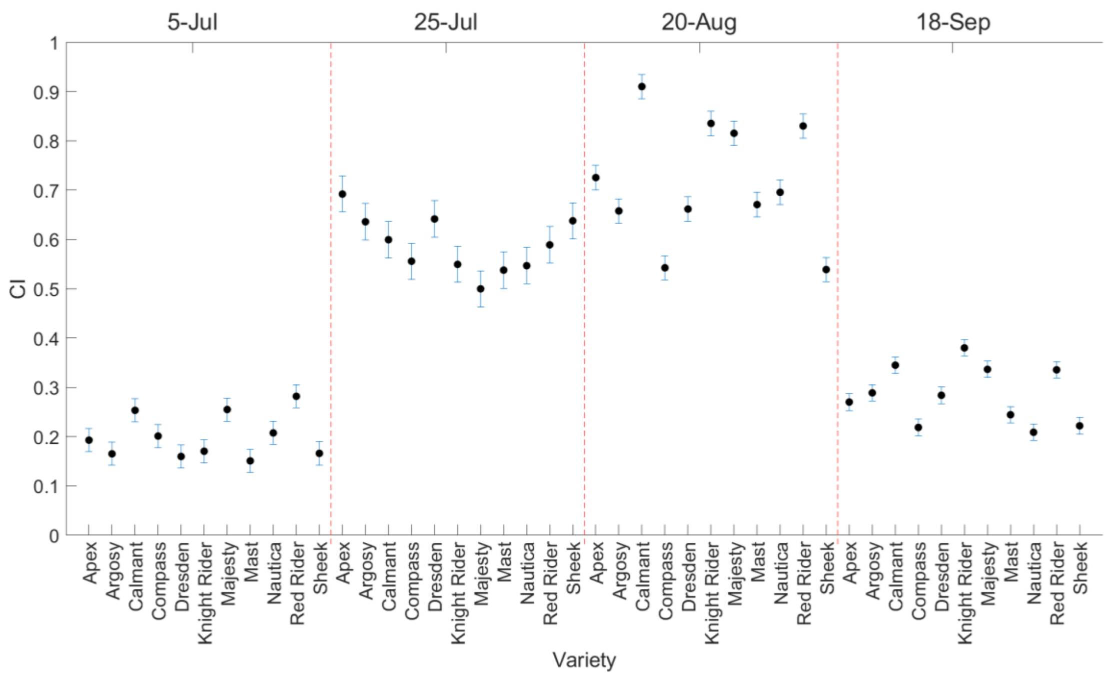


Figure 3-15. Results of Tukey's test for CI

The results of the regressions showed that all but one of the varieties had coefficients of determination above 0.7 when fitted with the previously mentioned model, with the values ranging from 0.62 to 0.90 and average of 0.78. Also, all the parameters were found to be statistically significant with 95% confidence intervals. Figure 3-16 shows an example of one of these, for the case of the NDVI of the Knight Rider variety. Nonetheless, there is large unexplained variance among samples of the same date.

Using the model, it is possible to compare the varieties by their highest value, denoted P, computed in terms of the parameters according to Equation (3-6), given that the function is a concave parabola. Another interesting metric is a growth rate computed as the highest value P divided by the time it takes to reach the peak, as described by Equation (3-7), in this case denoted R. Given that during the regression, each parameter was found along with a 95% confidence interval, the confidence interval values can be used to find the standard error of P and R, using the propagation of error equations (3-8) and (3-9).

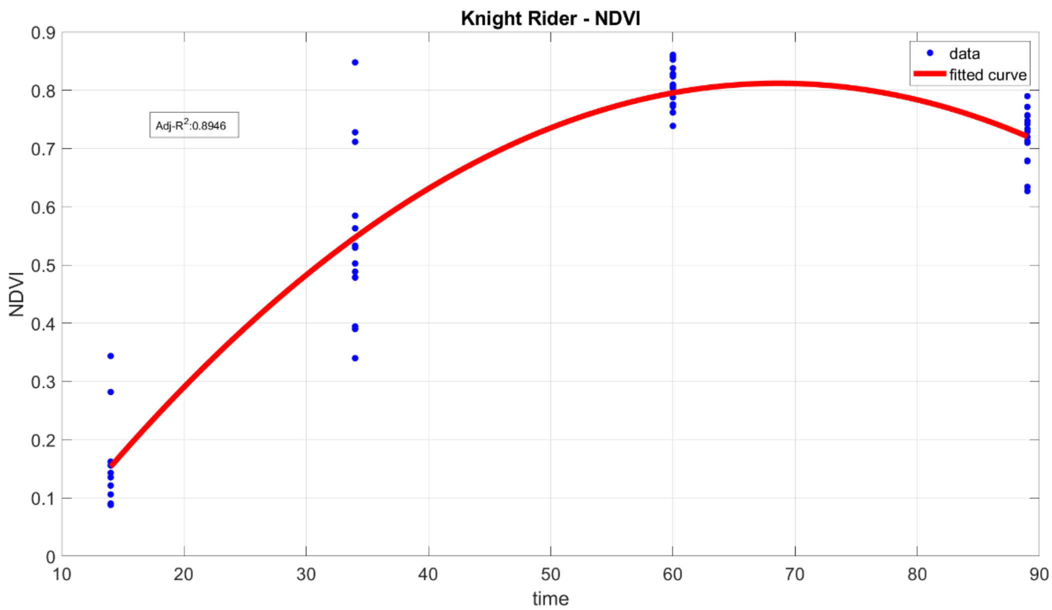


Figure 3-16. Curve fitting of Knight Rider's NDVI as a function of time in DAS

$$P = f(t^*) = c - \frac{b^2}{4a} \qquad (3-6)$$

$$R = \frac{P}{t^*} = \frac{c - \frac{b^2}{4a}}{\frac{-b}{2a}} = \frac{b^2 - 4ac}{2b} \qquad (3-7)$$

$$\Delta P = \sqrt{\left[\left(\frac{\partial P}{\partial a}\right)(\Delta a)\right]^2 + \left[\left(\frac{\partial P}{\partial b}\right)(\Delta b)\right]^2 + \left[\left(\frac{\partial P}{\partial c}\right)(\Delta c)\right]^2} \qquad (3-8)$$

$$\Delta R = \sqrt{\left[\left(\frac{\partial R}{\partial a}\right)(\Delta a)\right]^2 + \left[\left(\frac{\partial R}{\partial b}\right)(\Delta b)\right]^2 + \left[\left(\frac{\partial R}{\partial c}\right)(\Delta c)\right]^2} \qquad (3-9)$$

Appendix 5 summarizes these findings for NDRE, but similar behavior was found for NDVI and CI. While it would be possible to classify Calmant, Knight Rider, Apex, and Red Rider as over-performing; Sheek and Compass as under-performing; and the rest as average, there is no specific variety that has a statistically significant dominance over the others for the entire time domain in this dataset.

In terms of yield, there were no significant differences among varieties, as evidenced by the results of Tukey's test, as shown in Figure 3-17. This test was done after a one-way ANOVA where the variety was the only explanatory factor. The plot only shows 9 out of the 11 varieties because for the other two (Red Rider and Calmant) at least one of the replicates could not be harvested as a result of decay.

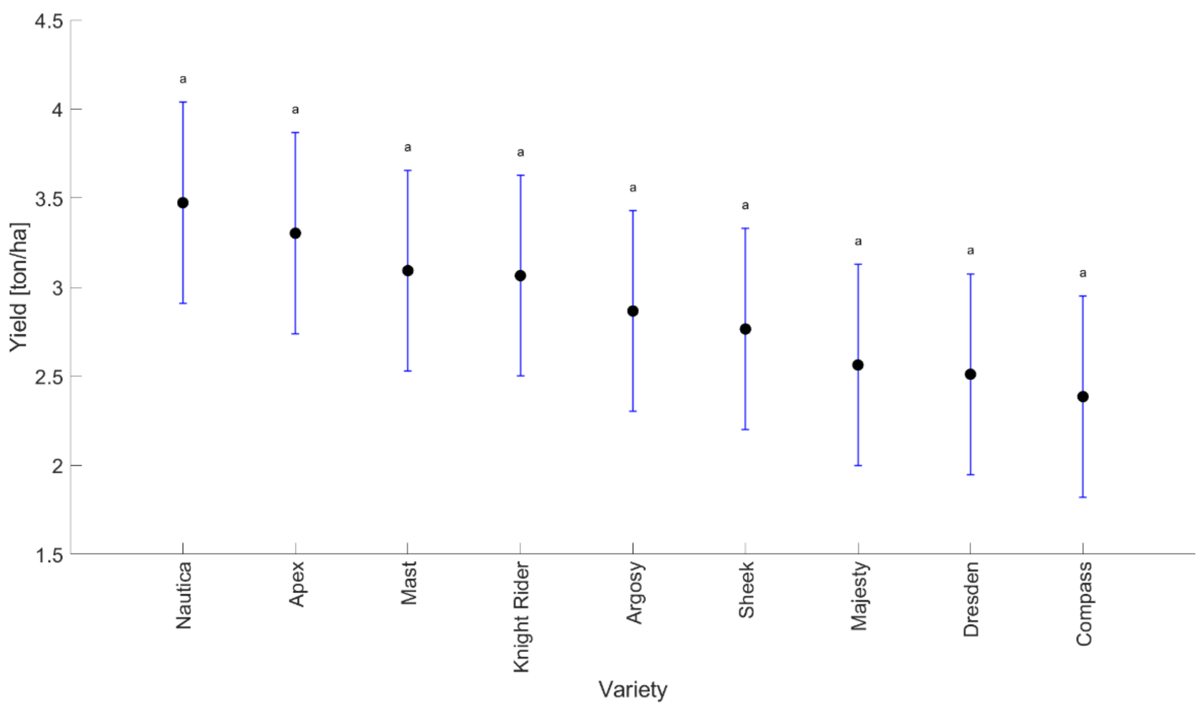


Figure 3-17. Results of Tukey's test for yield among varieties

3.3.2 Vehicle-mounted experiment

As mentioned before, only the 9 varieties with more than 10 replicates were used to perform the statistical analysis. The measurements from the DAS43X sensor were aggregated by averaging over the plots. For the multispectral and ultrasonic sensor, it was important to reduce the effect of soil patches inevitably measured inside the field of view of the sensor given the spacing between the crop rows. Because of this, the phenotypical data was aggregated per plot by two methods: (1) finding the maximum and (2) calculating the mean of the values above a certain threshold, found through analyzing the histograms of the readings and using the principle of Otsu's method to find the optimal threshold. This was a result of the histogram showing a bimodal distribution, assumed to be the sum of two normal distributions: one below the threshold for soil and one above it for plant tissue, as depicted in Figure 3-18. Finally, Figures 3-19 to 3-25 present some of the properties mapped by the system. For a few plots, no phenotypical data was available, either because the plants sown there were removed at an earlier date or because they were inaccessible to the HTPP platform because some structures had to be avoided.

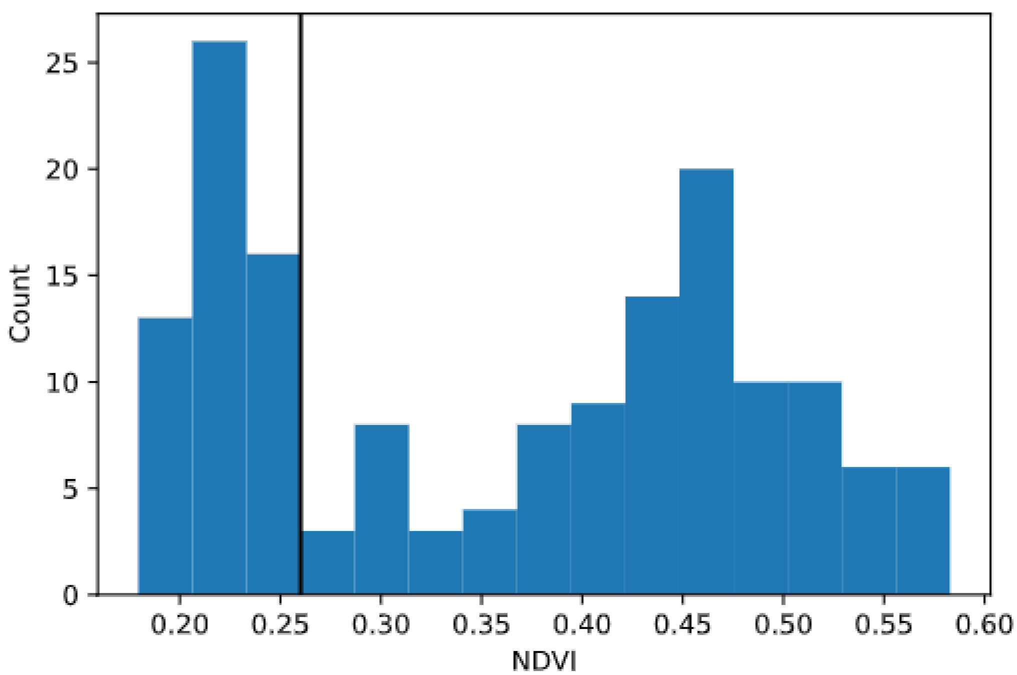


Figure 3-18. Histogram of NDVI at last date for example plot with threshold between measurements assumed to be soil and measurements assumed to be plant tissue

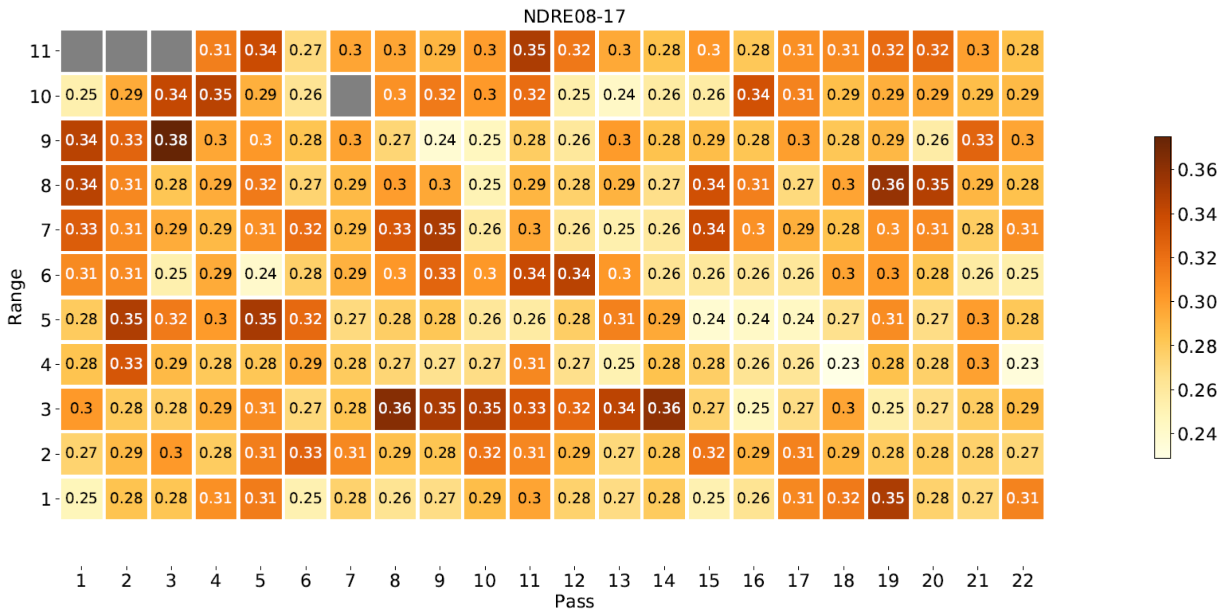


Figure 3-19. Map of NDRE at first date using maximum for aggregation

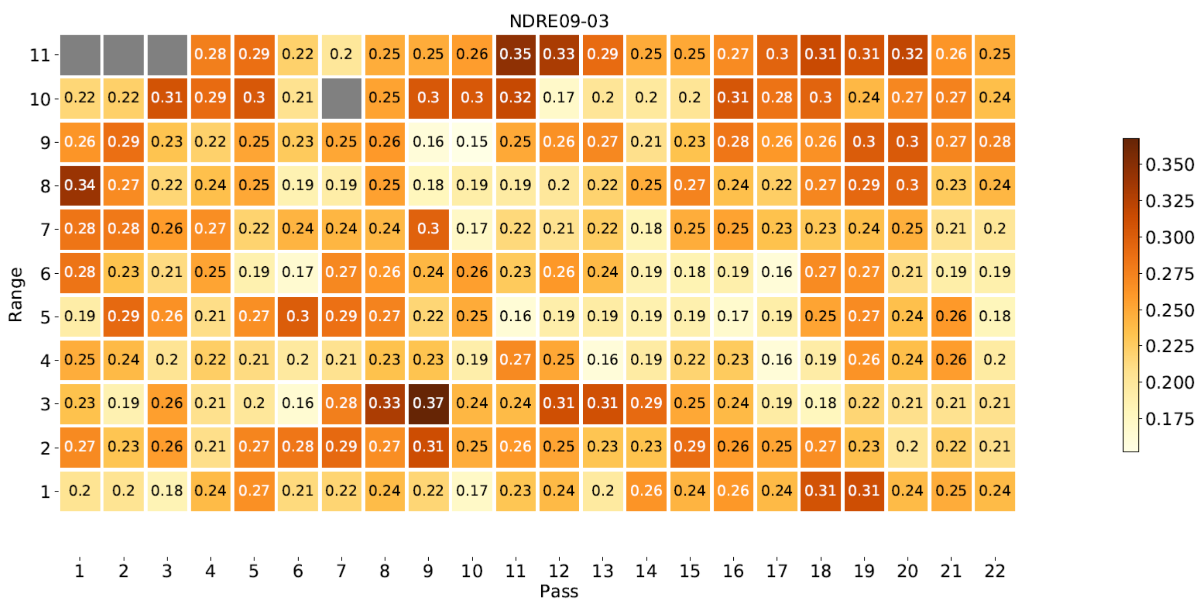


Figure 3-20. Map of NDRE at last date using maximum for aggregation

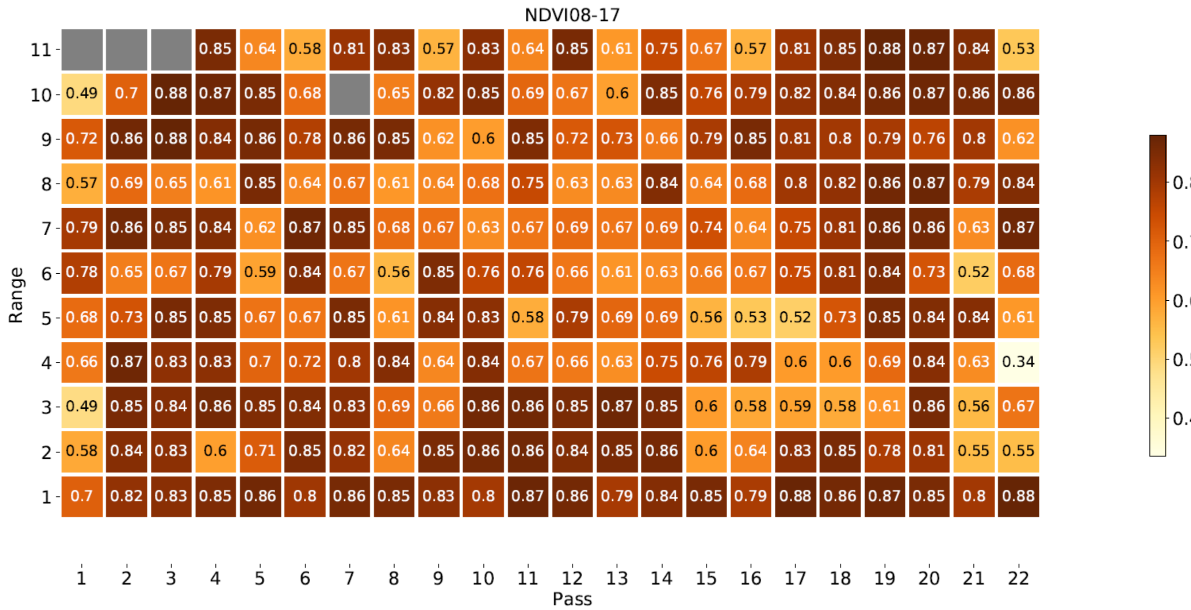


Figure 3-21. Map of NDVI at first date using average for aggregation

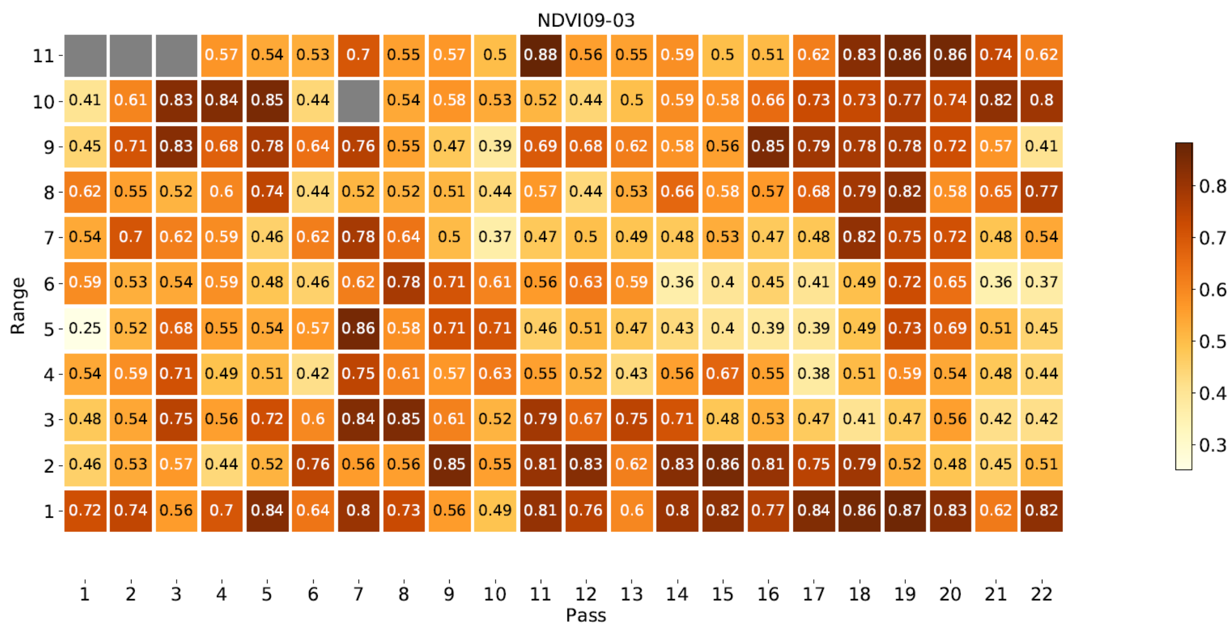


Figure 3-22. Map of NDVI at last date using average for aggregation

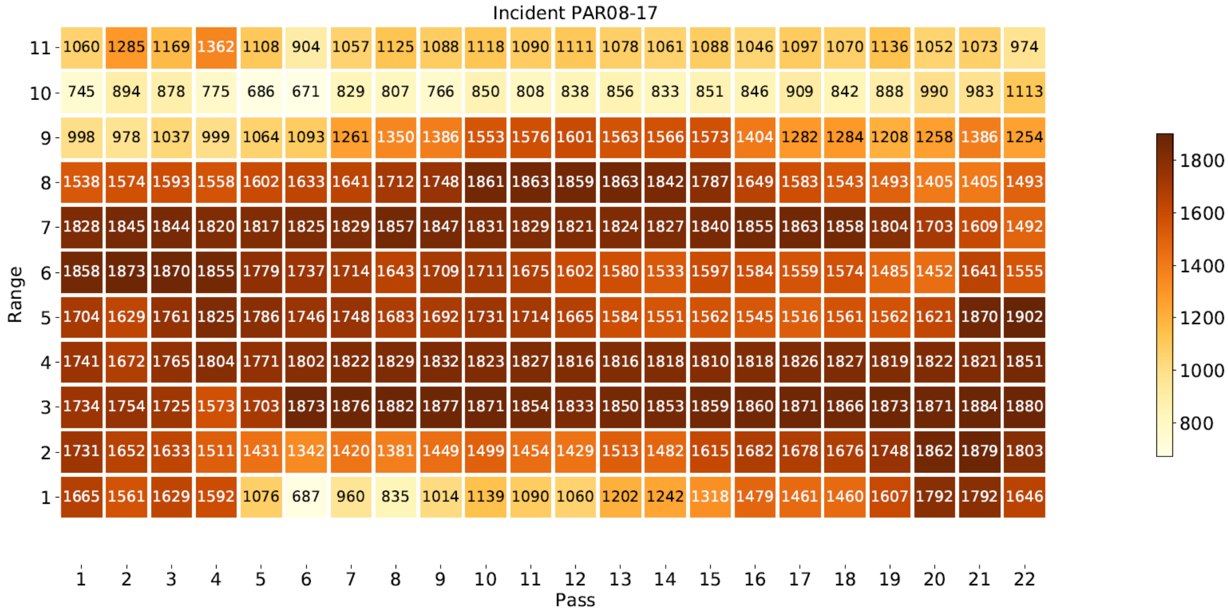


Figure 3-23. Map of incident PAR at first date

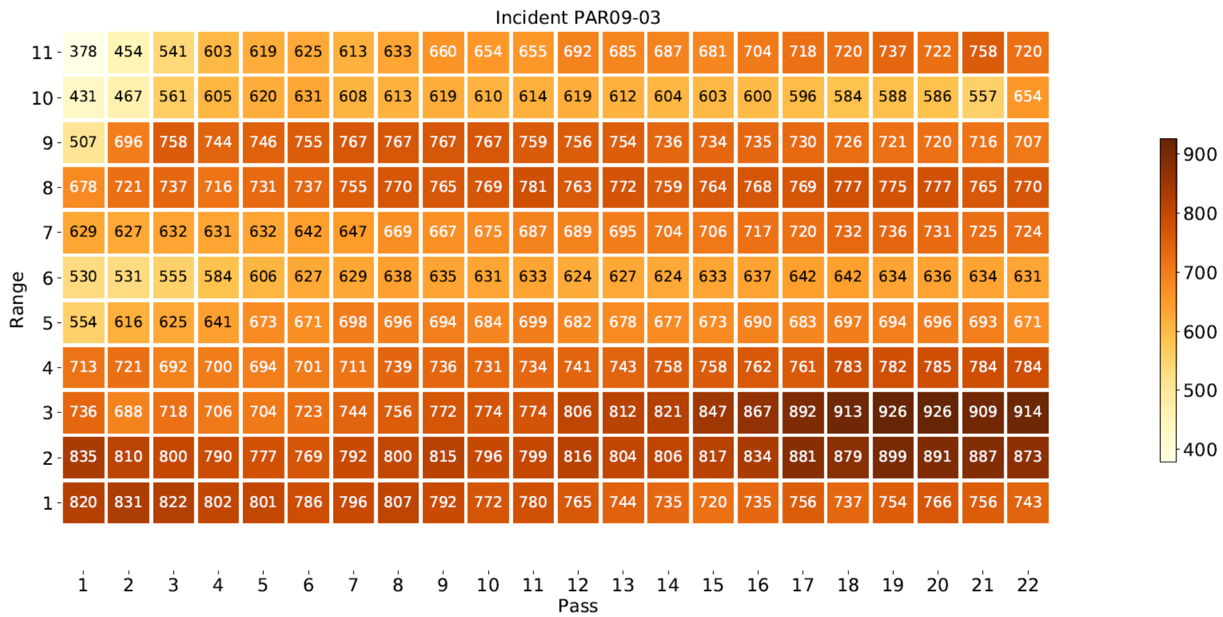


Figure 3-24. Map of incident PAR at last date

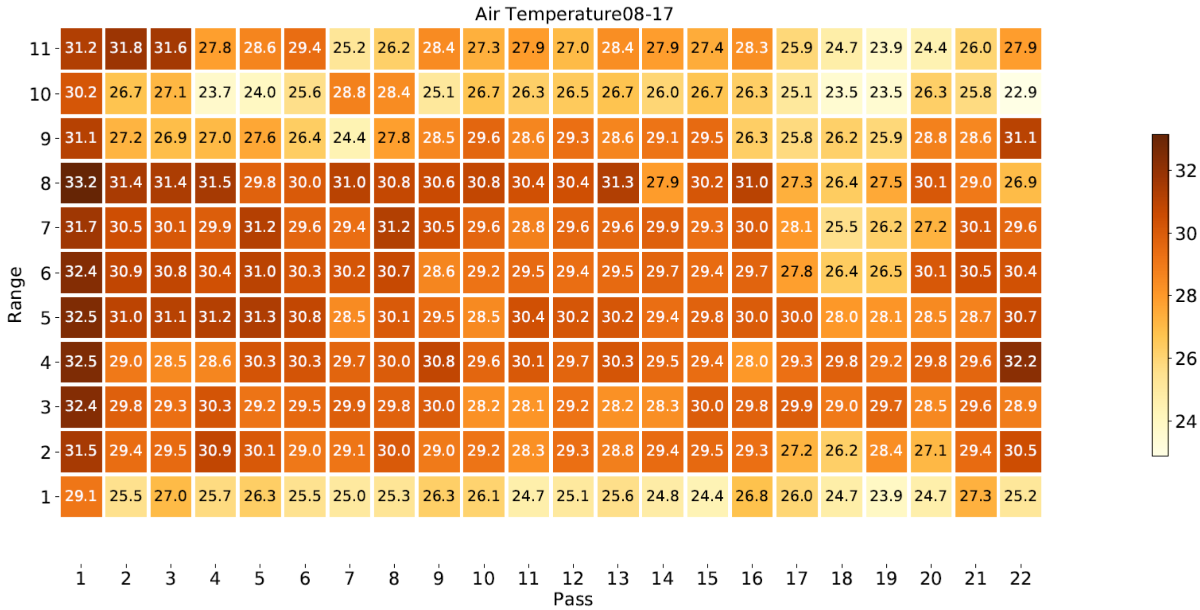


Figure 3-25. Map of air temperature at first date

When calculating the yield, beans that had suffered from mold by harvest time or that were the result of cross-contamination were filtered out. By eliminating those, the values of grain yield were comparatively smaller than those of the first experiment, as shown by the histogram in Figure 26 and the results of the Tukey's test in Figure 27. Similarly, Figures 3-28 and 3-29 show the histograms for NDRE and CI across all the dates. For these figures, only the values aggregated by the average method are shown but follow a similar trend as the maximum values. Finally, Figure 3-30 displays the average behaviour of the variables across time per variety, along with the standard deviation. In general, the values of the measured variables decreased as time passed, as expected given the overall trend seen in the first experiment and the phenological stage of the crop around these dates.

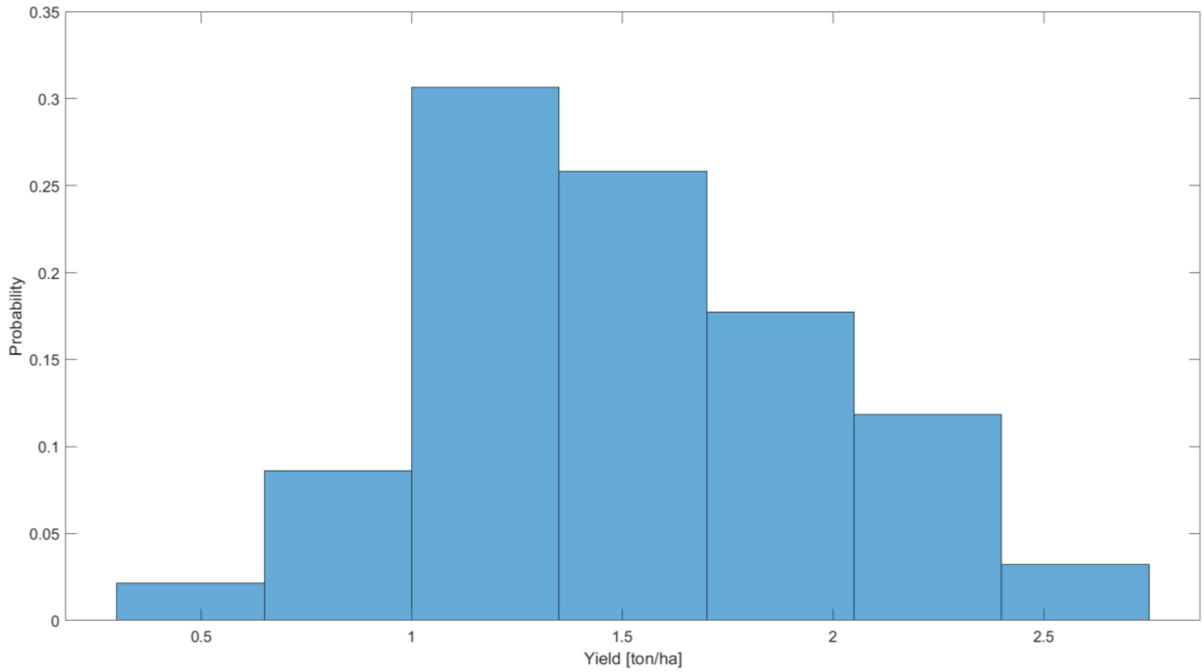


Figure 3-26. Histogram of yield

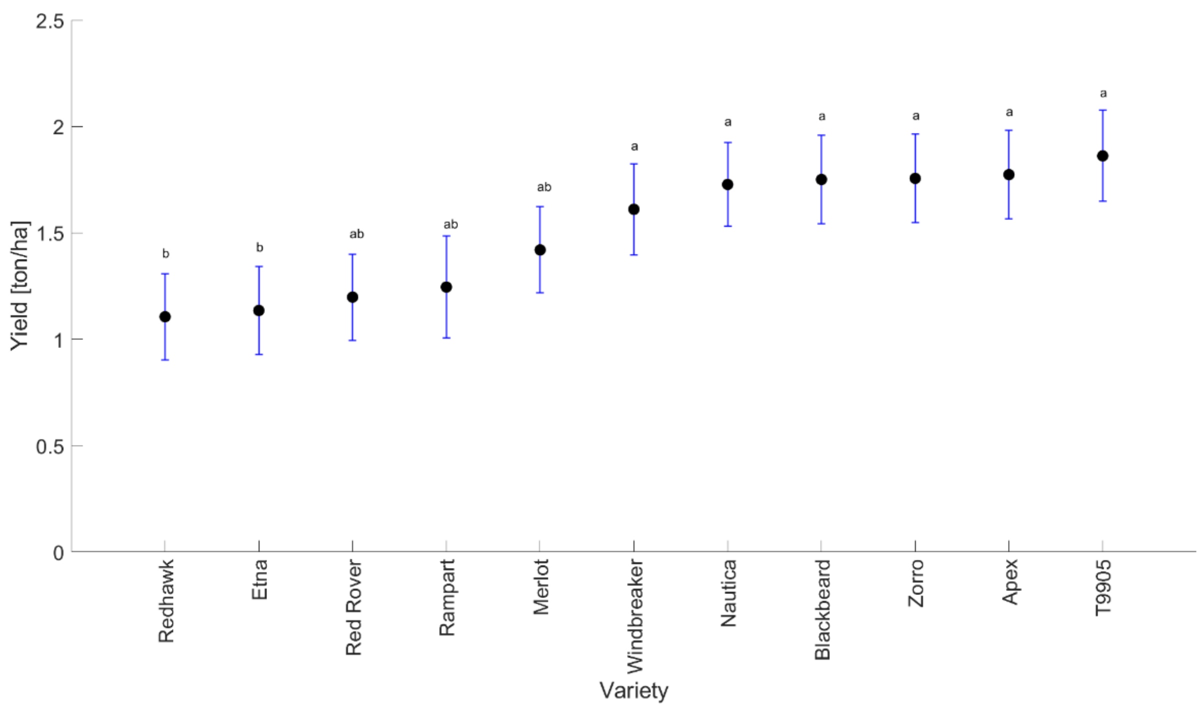


Figure 3-27. Result of Tukey's test for yield

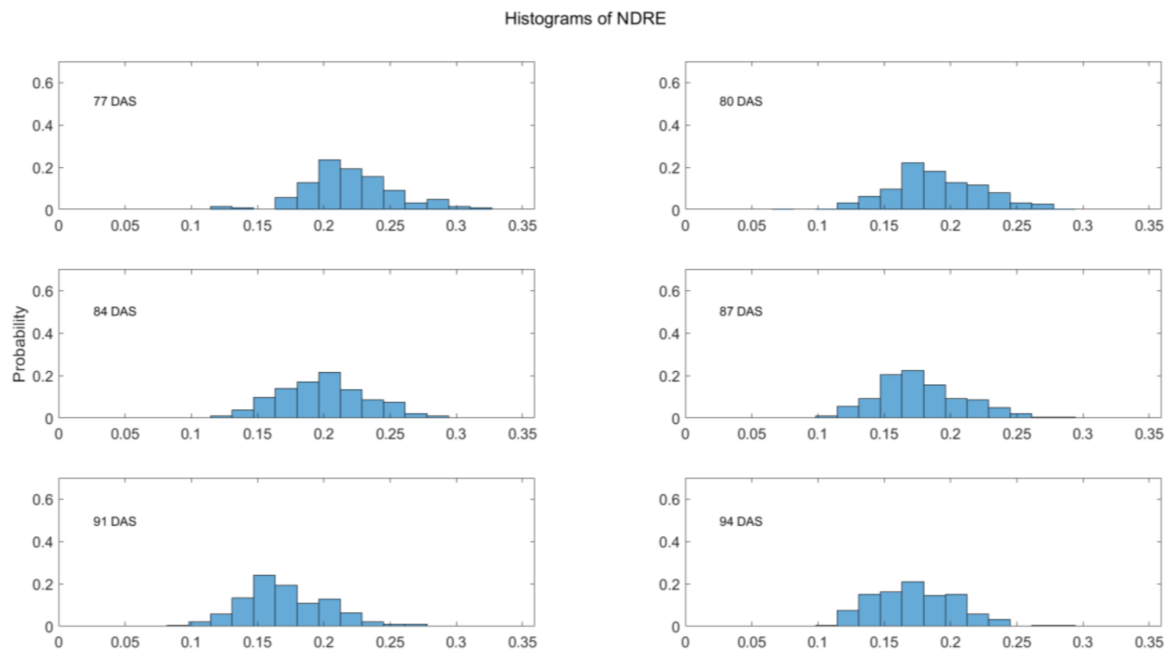


Figure 3-28. Histograms of NDRE measurements for each date

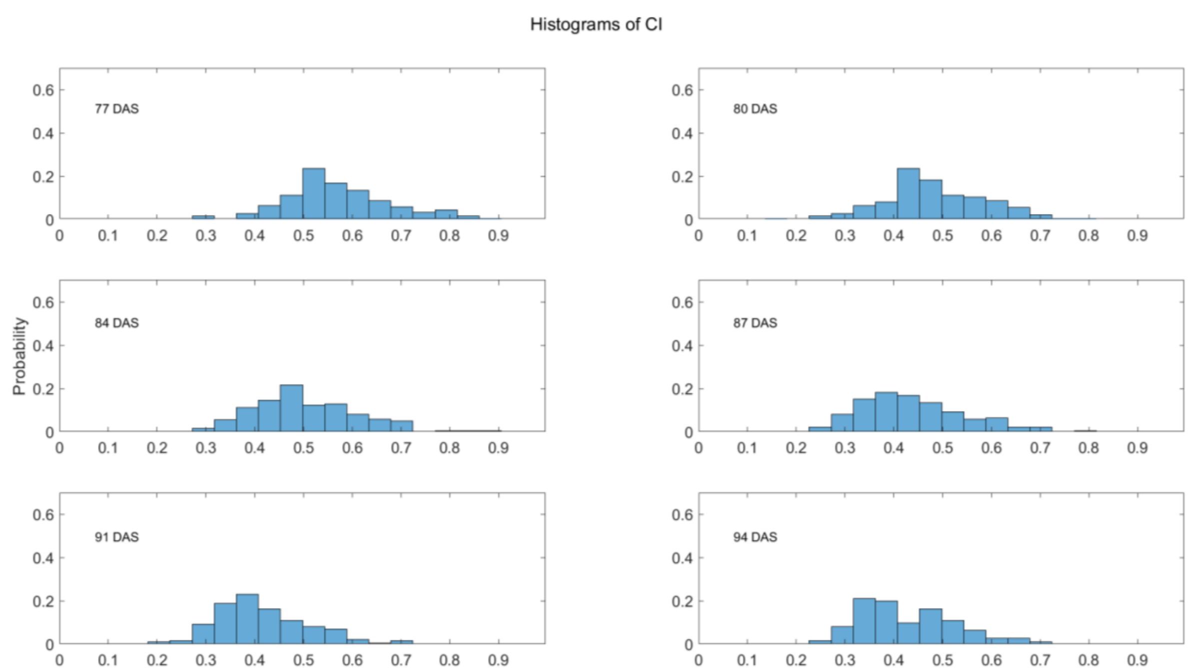


Figure 3-29. Histograms of CI measurements for each date

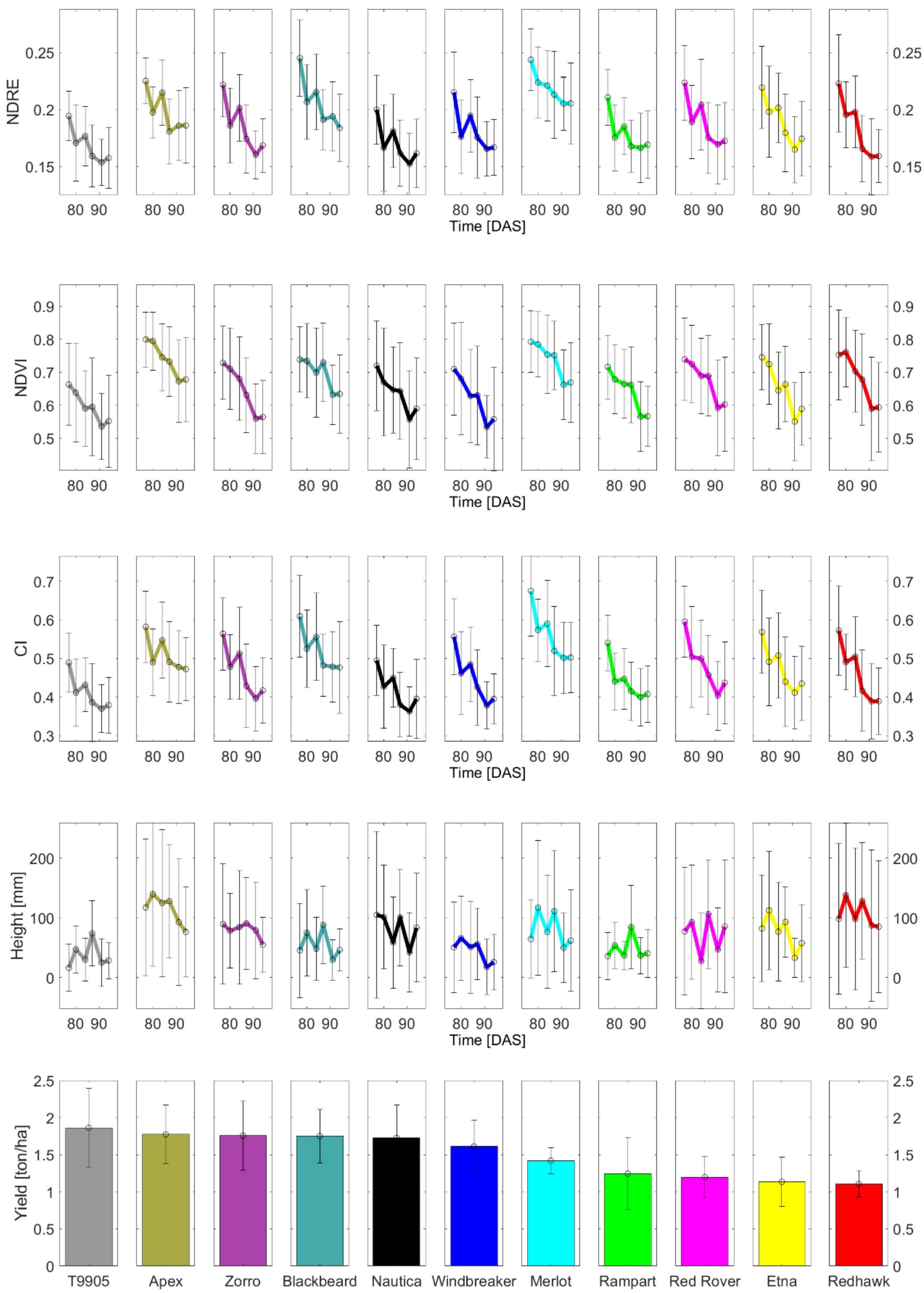


Figure 3-30. Evolution of phenotypical data per variety

3.3.3 Comparison

The average duration of the data collection for the first experiment with the handheld setup was just above 1.5 hours, for a field size of 150 m^2. The rate being 100 m^2/h. With the vehicle-mounted HTPP platform, the average time was slightly more than 45 min, for a field of size 3,600 m^2. Consequently, the rate was 5,400 m^2/h, more than 50 times faster, even with the low travel speed of the vehicle as it scanned the field.

The ability to include the environmental sensors and the cameras, as well as increased comfort for the operator is also an advantage. On the other hand, it is easier for an operator with the handheld setup to reach inner parts of the plots, as they can walk more safely between crop rows when that would be too risky for the vehicle. As a result of this, while the 8 measured locations within a plot represent about 10% of the area of the plot, they are distributed all around the possible surfaces and as such, could be a more representative sample of the population. Instead, for the vehicle-mounted HTPP platform, the covered area represents almost 40% of the plot's surface, with the caveat that it is concentrated near the edges of the plot closest to the alleys where driving is possible. The vehicle-mounted platform offers more consistent distancing between the sensors and the target i.e. the top of the canopy, at the cost of increased vibrations from the vehicle's engine.

While the system was tested in a light-weight utility vehicle, an interesting approach to consider would be to mount it in a high-clearance vehicle, which would allow access to more areas of the field. There is a trend towards this type of vehicle in similar scenarios. An additional consideration is the potential for this system to be mounted on an autonomous vehicle that can be driven without an operator. In such a scenario, the system could monitor the field continuously.

3.4 Conclusions

Both the ultrasonic and multispectral sensors provide information about crop status. The height measured from the ultrasonic sensor is an important parameter of crop architectonics by itself. Some breeding techniques have been used to adjust the height of certain crops to better match with the range covered by combine harvesters. Height can be related with canopy volume. The different VIs measured are less frequently used as traits by themselves, but they can relate with multiple other properties like plant vigor and canopy coverage. The combination of both types of sensors could be used as an example to improve the prediction of biomass, where the height relates with the volume of the canopy and the VIs with the density of the canopy.

The proposed vehicle-mounted HTPP platform provides the ability to map phenotypical and environmental data across a field in an efficient manner. Crop modelling techniques can benefit from the availability of such data at multiple scales. Comparison against the handheld setup shows that both options have strengths and drawbacks, as mentioned in Section 3.3.3, and ultimately it will depend on many factors such as field size and the goal of the experiment to determine if the investment in a HTPP platform can be justified. Nonetheless, the ideas presented in this work serve as progress towards low-cost HTPP platforms. Further research is required to leverage the video recorded from the cameras, a potential resource underutilized in this work, to extract additional plant traits.

References for Chapter 3

Andrade-Sanchez, P., Gore, M. A., Heun, J. T., Thorp, K. R., Carmo-Silva, A. E., French, A. N., Salvucci, M. E., & White, J. W. (2014). Development and evaluation of a field-based high-throughput phenotyping platform. *Functional Plant Biology, 41*, 68-79.

Araus, J. L., Kefauver, S. C., Zaman-Allah, M., Olsen, M. S., & Cairns, J. E. (2018). Translating High-Throughput Phenotyping into Genetic Gain. *Trends in Plant Science*, 23(5), 451-466.

Buelvas, R., Adamchuk, V., Pouliot, A., Stromvik, M., & Whitmore, A. (2019). Development of a quick-install rapid phenotyping system. *2019 ASABE Annual International Meeting.* Boston, MA, USA.

Knoch, D., Abbadi, A., Grandke, F., Meyer, R. C., Samans, B., Werner, C. R., Snowdon, R. J., & Altmann, T. (2020). Strong temporal dynamics of QTL action on plant growth progression revealed through high-throughput phenotyping in canola. *Plant Biotechnology Journal, 18*, 68-82.

Mir, R. R., Reynolds, M., Pinto, F., Khan, M. A., & Bhat, M. A. (2019). High-throughput phenotyping for crop improvement in the genomics era. *Plant Science, 282*, 60-72.

Pouliot, A. (2016). Development of a Quick-Install Auto-Steering System for Agricultural Vehicles. McGill University. *MS Thesis*.

Reynolds, D., Baret, F., Welcker, C., Bostrom, A., Ball, J., Cellini, F., Lorence, A., Chawade, A., Khafif, M., Noshita, K., Mueller-Linow, M., Zhou, J., & Tardieuc, F. (2019). What is cost-effective phenotyping? Optimizing costs for different scenarios. *Plant Science, 282*, 14-22.

Rossi, R., Leolini, C., Costafreda-Aumedes, S., Leolini, L., Bindi, M., Zaldei, A., & Moriondo, M. (2020). Performances Evaluation of a Low-Cost Platform for High-Resolution Plant Phenotyping. *Sensors, 20*(11), 3150-3166.

Xu, Y., Li, P., Zou, C., Lu, Y., Xie, C., Zhang, X., Prasanna, B. M., & Olsen, M. S. (2017). Enhancing genetic gain in the era of molecular breeding. *Journal of Experimental Botany*, 68(11), 2641-2666.

Yang, W., Feng, H., Zhang, X., Zhang, J., Doonan, J. H., Batchelor, W. D., Xiong, L., & Yan, J. (2020). Crop Phenomics and High-Throughput Phenotyping: Past Decades, Current Challenges, and Future Perspectives. *Molecular Plant, 13*, 187-214.

Connecting text to Chapter 4

Chapter 4 is related to the second objective of this study listed in Chapter 1. In the previous chapter, the development of an HTPP platform which also allowed for envirotyping was evaluated. A natural extension of this approach would be to include soil information to get as close as possible to completely characterizing the environment in which the plant expresses its phenotype. The following chapter aims to address that by developing a PSS platform based on Vis-NIR spectroscopy, a technique that allows soil spectra to relate to a wide array of soil physical and chemical properties. Results obtained from this project were partially reported and published as a conference meeting paper listed below and the final research findings are under preparation for a journal publication.

R. Buelvas, V.I. Adamchuk, J. Lan, B. de Leener, & G. Mangeat. (2021). Evaluation of a semi-automated *in-situ* soil sensor using Vis-NIR spectroscopy. *Biosystems Engineering*. (To be submitted)

R. Buelvas, V.I. Adamchuk, B. de Leener, & G. Mangeat. (2020). Development of a semi-automated *in-situ* soil sensor using Vis-NIR spectroscopy. 2020 ASABE Annual International Meeting. Omaha, MA, USA. July 12-15, 2020

CHAPTER 4 Development of Proximal Soil Sensing platform

Abstract

Soil spectroscopy technology has great potential for PSS. In particular when combined with the automation required for on-the-spot sampling and a platform capable of controlling measurement depth. The present work evaluates the development of a soil sensor system, equipped with a vertical drill to facilitate the collection of depth-specific measurements. The ChrysaLabs probe performs Vis-NIR spectroscopy *in situ* with 360° optical integration. The ability of the user to command complete measurement cycles from inside the vehicle's cabin improves the ergonomics and logistics of data collection surveys. The results showed the system was able to collect measurements at 4 depths with an average time of 3 min per measurement. Proper use of this technology would support the adoption of technologies for real-time determination of spatially variable soil characteristics.

Keywords: Automation, On-the-spot sampling, Proximal Soil Sensing, Vis-NIR spectroscopy.

4.1 Introduction

PSS offers an alternative to traditional soil lab analysis which allows for *in situ* and real-time measurements. According to (Nocita *et al.*, 2015), the overall accuracy of soil spectroscopy applications might be improved by obtaining soil data in more locations, at different times, and at various depths. The last component of multi-depth measurements is a key difference with remote sensing, a complementary approach which has also seen growth in recent times. Currently, acquiring soil measurements over large areas can be a long and tedious task. In this context, automating data collection has the potential to increase efficiency and affordability of high-density soil sensing. Many other challenges make *in-situ* soil analysis impracticable for non-experienced users e.g. inserting a soil probe into the soil can be difficult under certain conditions.

ChrysaLabs (ChrysaLabs Inc., Montréal, Canada) has developed and patented a portable optical probe that improves the signal-to-noise ratio (SNR) of *in situ* soil spectroscopy. The technology consists of a tubular light collector combined with a Vis-NIR spectrometer and LEDs, which readily provides instantaneous analysis of soil minerals and characteristics. This sensor provides hyperspectral readings that can be related with the soil physical or chemical properties like organic matter or nitrogen content. Given its fragile structure, however, it requires a hole to be opened in the soil beforehand.

The system should be able to take on-the-spot soil measurements at controlled depths by inserting the ChrysaLabs Probe, depicted in Figure 4-1. The design of this system would be similar to that of the soil analyzer in Adamchuk, Dhawale, & Rene-Laforest (2014). Low vibrations are particularly desired to guarantee the straightness of the hole, thus, ensuring that the stresses on the sensor probe and interruptions in the area of contact between the probe surface and the soil are minimized. Finally, the system can be quickly mounted and dismounted on many suitable, commercially available, agricultural vehicles.

Figure 4-2 shows the graph of example soil spectra collected with the ChrysaLabs probe from 240 locations. Multiple processing methods use this kind of data as input to predict soil chemical properties, like pH or the concentrations of P, K, or Ca, as well as physical properties like soil texture (Stenberg *et al.*, 2010).

Figure 4-1. Picture of ChrysaLabs probe in prototype stage

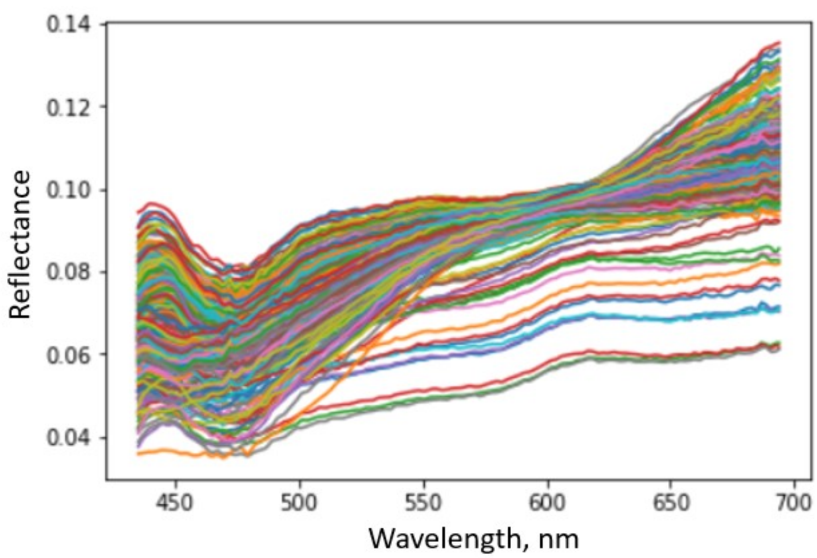


Figure 4-2. Example of a portion of soil spectral response measured by ChrysaLabs probe

The objective of this research was to design, build, and evaluate a system with the characteristics mentioned above. The combination of these features makes for a soil mapping system that could benefit growers and agronomists.

4.2 Materials and methods

4.2.1 Platform development

The ChrysaLabs probe has a built-in hyperspectral sensor which measures light reflected from the soil through a circular window. The measurement bands cover visible and near-infrared segments of the light spectrum. The sensor is encased in a metallic probe. One way to open the required borehole is to use a vertical drill with a soil auger, which can be automated and mounted on a vehicle. A linear actuator can be used to push both the drill and the probe into the soil, allowing for precise depth control. To use only one linear actuator for both operations, a tool-change mechanism is required, so that the drill or the probe to link and unlink from the linear actuator. Part of this mechanism requires a mobile tool holder which can select which tool to use next and to place it close to the linear actuator. A third tool was added, a smooth rod working as a press, to expand the hole opened by the drill to its final dimension.

The CAD design of the system is presented in Figure 4-3. The parts in orange were custom manufactured by laser cutting and metal bending, while the other components were commercially available. The main frame was welded, and the rest of the components were bolted together into the assembly. A simplified version was built first to test the drilling capabilities without having to worry about the probe itself. The hole should have at least a 1 in. (25.4 mm) diameter and 100 mm depth, though greater depths are desirable. There is a guide to ensure the alignment of the vertical axis near the bottom of the structure. A solenoid was used to hold the tool in place once inserted into the connector. The electronics were placed inside an enclosure at the middle height of the central beam. The frame was arranged in an L-shape with a foot that fits into a standard 2 in. (50.8 mm) vehicle hitch and is assembled from square steel beams.

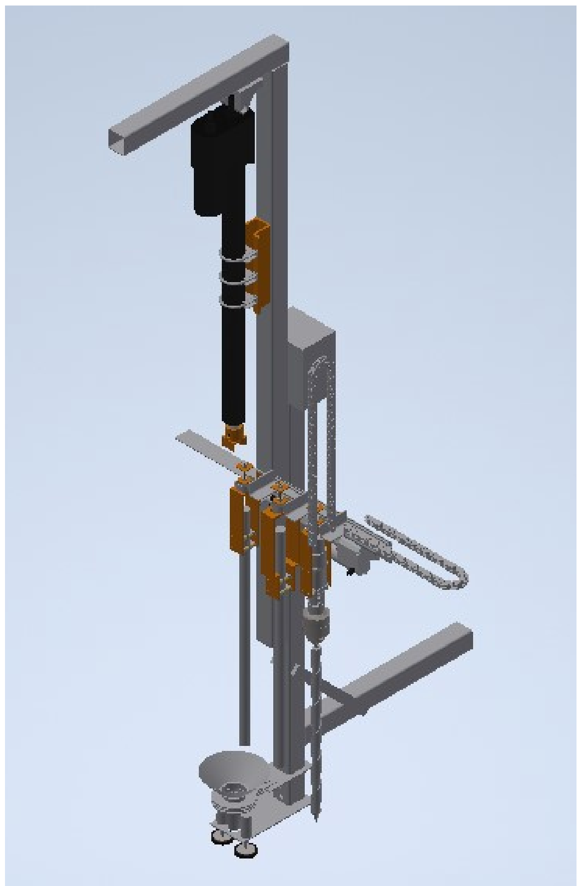


Figure 4-3. Screenshot of the finalized CAD of the PSS platform

The linear actuator used in this system was a PA-17 (Progressive Automations, Inc., Richmond, Canada) with a stroke of 24 in. (610 mm), with an integrated absolute encoder. The motor and gearbox for the drill were RS550 and P61S-4444 (BaneBots LLC, Loveland, CO, USA). The vehicle used with the prototype was a Suzuki Grand Vitara (Suzuki Motor Corporation, Hamamatsu, Japan). The slider mechanism was built around a T-slotted rail (Servocity LLC, Winfield, KA, USA), with an outer U-channel surrounding the rail and sliding through it with the support of mini V-wheels. This slider mechanism was driven by a PA-14P linear actuator (Progressive Automations, Inc., Richmond, Canada) with a stroke of 18 in. (457 mm). Tool holders were added with fixed spacing.

The prototype was connected to the probe by USB and could transmit the sensor readings by Wi-Fi using the smartphone app of ChrysaLabs. This communication does not require an Internet connection. The prototype was controlled by commands sent through Bluetooth from the same smartphone using the Serial Bluetooth Terminal v1.12 app (Kai Morich, Hockenheim, Germany). A Raspberry Pi 3 Model B (Raspberry Pi Foundation, Cambridge, UK) received the Bluetooth commands and responded by performing the associated task as defined by the Python script. In most cases, tasks required communication with the motor drivers, Roboclaw 2x15A Motor Controller (Basic Micro, Temecula, CA, USA) (e.g., adjusting motor speeds). One of the motor drivers was connected to both the horizontal and vertical linear actuator, the other motor driver controlled the drill and the solenoid. The linear actuators have an integrated absolute encoder for position feedback and the drill motor has an external current sensor for feedback. The motor drivers themselves have current sensors for all its ports, but the external current sensor was added to the drill because the higher resolution was desired. The current sensor interfaced with the Raspberry Pi by an Analog-to-Digital Converter ADS1115 (Zhiwei Robotics Corp, Shanghai, China). All the electronics were powered from the vehicle's battery via a cigarette lighter adapter. A DC-DC converter was required to step down the voltage to 5V for the power supply of the smaller electronic components. An emergency stop button was added in case of an emergency and two relays are used to safely disengage the power supply when it is pressed. Figures 4-4 and 4-5 show pictures of the finalized prototype.

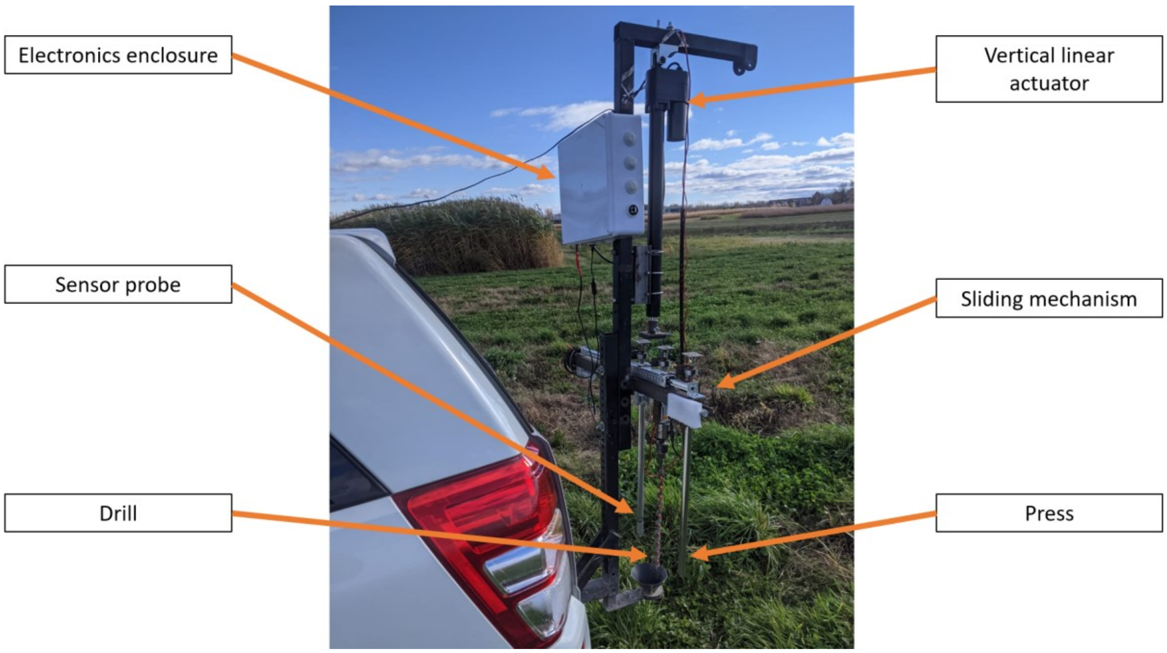


Figure 4-4. Picture of the finalized prototype

Figure 4-5. Picture of vehicle with prototype mounted

The sliding mechanism for tool change followed these steps for connecting the tool to the vertical linear actuator:

(1) The vertical linear actuator was extended so that the slot of the connector and the top end of the tool were at the same level.

(2) The horizontal linear actuator was extended and inserted the tool into the slot of the connector.

(3) The solenoid was activated, pushing a pin inside a notch at the top end of the tool, to hold it inside the connector.

(4) The vertical linear actuator was retracted and pulled the tool up, so the hook that attached it to the sliding mechanism could now make contact on a segment with a smaller radius, loosening the spring.

(5) The horizontal linear actuator was retracted, and the tool detached from the hook and remained connected to the vertical linear actuator.

(6) The solenoid was deactivated to prevent overheating, while the tool remained connected given the geometry of the slot.

This process is illustrated in Figure 4-6. After this, the vertical linear actuator could extend or retract freely to insert the tool in the ground and use it. The sliding mechanism was constrained to a small movement range while the vertical linear actuator was extended with a tool attached, to prevent self-collision. All the tools were mounted by standard top ends and holders to facilitate this process.

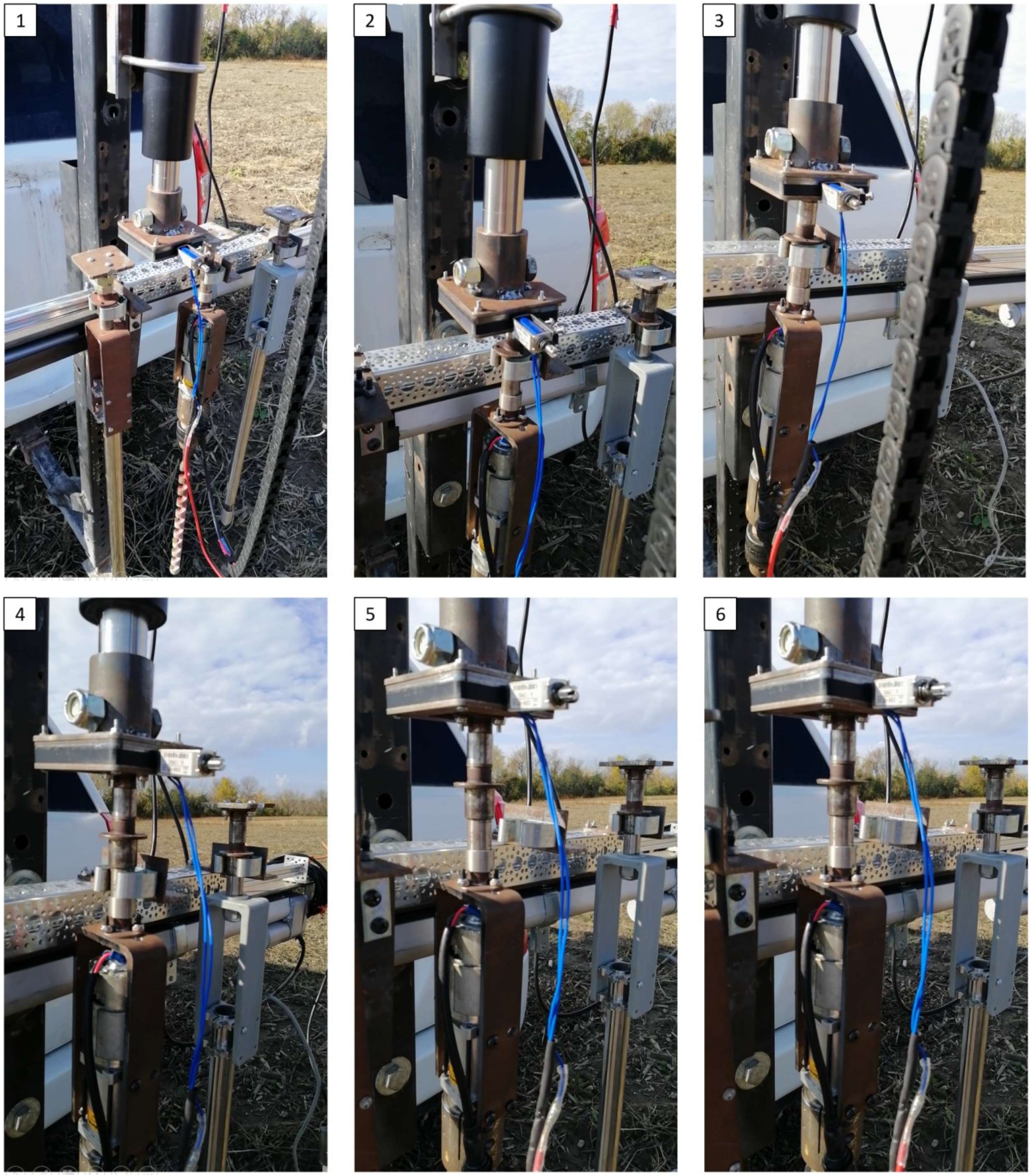


Figure 4-6. Pictures depicting steps for tool connection

Once the system finished using a certain tool and needed to return it to its support, the process was:

(1) The vertical linear actuator was retracted so that the tool was at the same level as the hook.

(2) The horizontal linear actuator was extended and pushed into the receiving end of the tool until it clicked into place surrounded by the hook.

(3) The horizontal linear actuator was retracted and because the tool was held by the hook, it remained attached to it.

The linear actuator could then be retracted and the sliding mechanism moved to the next tool and the process was repeated for each tool, at every location. All the tools can be easily replaced, if necessary.

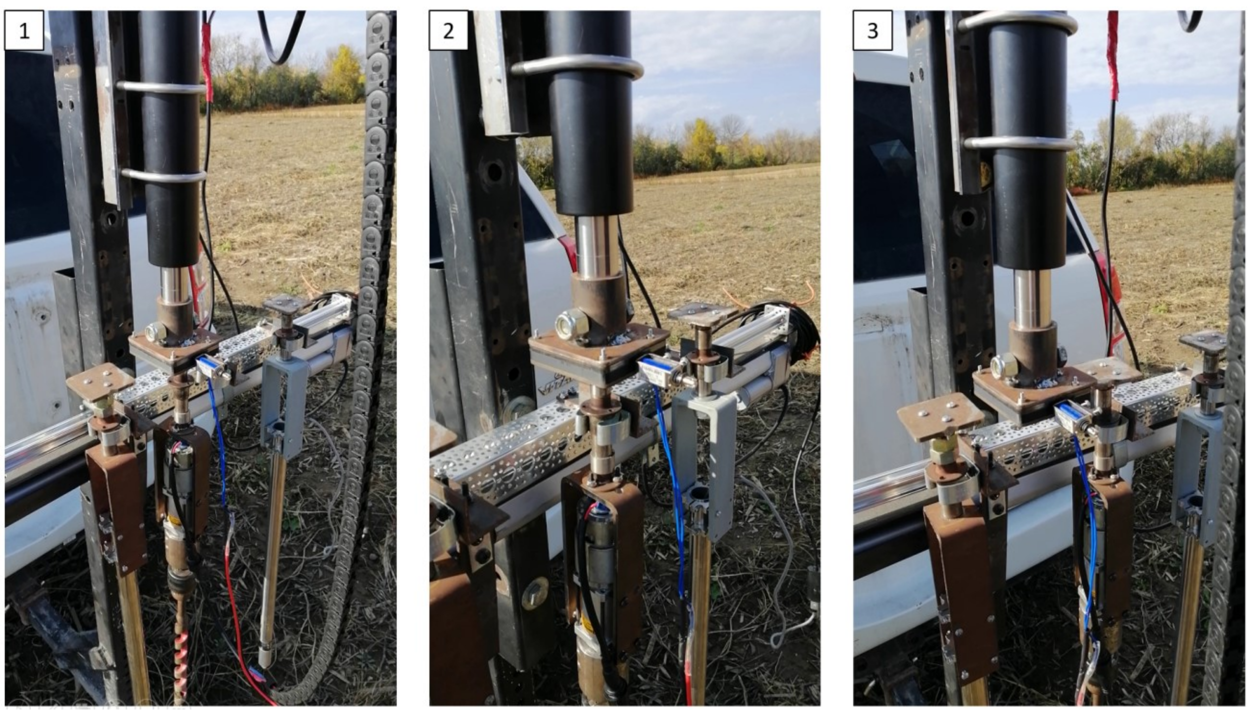


Figure 4-7. Pictures depicting steps for tool disconnection

A first test was done with multiple types of drill bits mounted on a hand drill. The candidate drill bits were: Twist Drill, SDS Bit, Flat Bit, Ship Auger, and Soil Auger. The criteria used was the cleanliness of the hole and the relative ease of use with the hand drill. From the evaluated drill bits, the one that seemed to work best was a Ship Auger type, normally used for wood applications, depicted in Figure 4-8. This type of drill bit produced the cleanest hole. The voltage and current ratings of the hand drill were later used to select the motor.

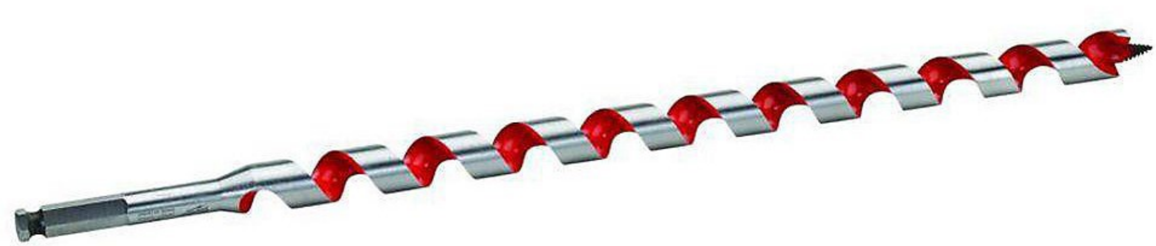


Figure 4-8. Picture of the best-performing drill bit

Finally, Figure 4-9 illustrates the measurement cycle timeline. The total measurement time per location was about 12 minutes. In that time, 4 measurements were performed at 4 different depths of 50, 70, 90, and 110 mm. The maximum current required was 8 A, but only when the drill motor and the vertical linear actuator were operating at the same time.

4.2.2 Experimental design

Soil samples were collected with a generic soil sampler (JMC, Clements Associates Inc., Newton, Indiana, USA) and with the prototype PSS platform in three fields of the Emile A. Lods Agronomy Research Center of Macdonald Campus, McGill University. In each of the fields, six locations were measured, as shown in Figure 4-10. At each location, the prototype PSS platform collected the spectral response of the soil at four different depths, by placing the middle of the probe's window at 50, 70, 90, and 110 mm below the surface level. Then, four more holes were made with the soil sampler to the sides of the original hole made by the PSS platform, as illustrated by Figure 4-11. The soil cores collected by the soil sampler were divided into shallow (between surface level and 152 mm) and deep (between 152 mm and 305 mm), stored in soil bags and labelled for later standard laboratory analysis. The soil sampling was done on October 23rd and 24th, 2020. The laboratory analysis was performed on October 29th, 2020 by A & L Labs (A & L Canada Laboratories, Inc., London, ON, Canada). Maps were made with ArcGIS Online (ESRI, Redlands, CA, USA).

Time (s)	Vertical linear actuator	Horizontal linear actuator & Solenoid	Drill motor	Sensor probe	Raspberry Pi	Current (A)
45		Select drill				2
65	Extend					3
85			Drill			8
135	Retract					4
205		Select press				2
225	Extend					3
245						5
285	Retract					3.5
355		Select probe				2
375	Extend				Process	3
440				Measure		1
445	Extend					3
510				Measure		1
515	Extend					3
580				Measure		1
585	Extend					3
650				Measure		1
690	Retract					3
715		Return probe				2

Figure 4-9. Diagram of complete measurement cycle

Figure 4-10. Map of soil sampling locations

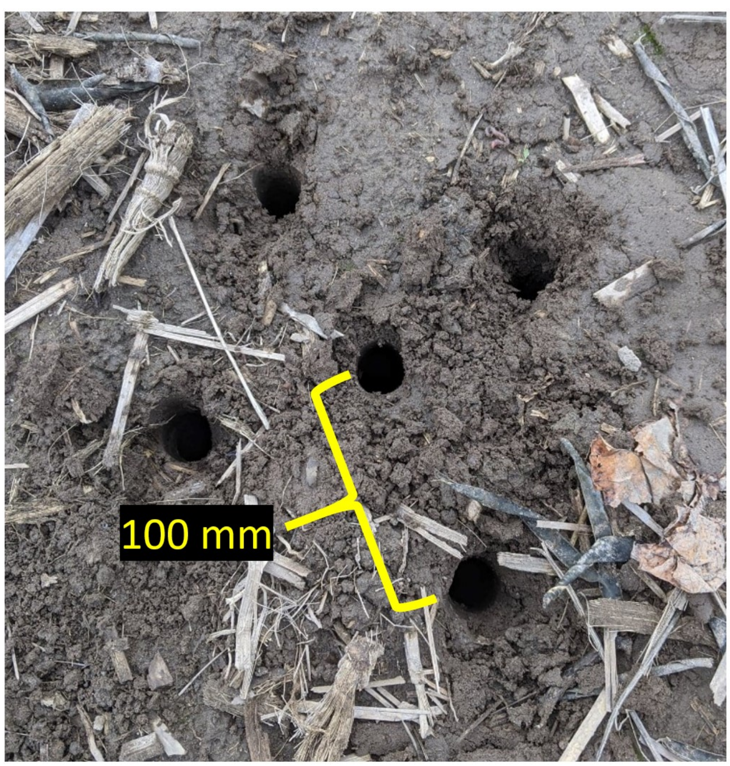


Figure 4-11. Picture of hole arrangement

4.3 Results and discussion

Mounting and dismounting the system from the vehicle was found to be an easy operation for two people. All the measurements were georeferenced and Figures 4-12 to 4-15 show maps of the soil attributes measured by the lab analysis at those locations, which confirms soil variability both between and within the fields. The properties measured were organic matter content, P, K, Mg, Ca, Na, Al, K/Mg ratio, water pH, buffer pH, Cation Exchange Capacity (CEC), and percentages of saturation for P, K, Mg, Ca, Na, Al, and H. The previously mentioned variability is reflected in the spectral response of the soil, illustrated by Figure 4-16. For each of the depths, a subset of the raw spectra collected from the visible channel of the probe in the range from 435 nm to 694 nm is shown. Following the previously described procedures, it is possible to build a dataset of soil spectra and target attributes to predict soil attributes. Different processing and modelling techniques can then be used to find the appropriate relationships between them.

Figure 4-12. Map of Ca content in kg/ha for the shallow soil

Figure 4-13. Map of Ca content in kg/ha for the deep soil

Figure 4-14. Map of organic matter content in % for the shallow soil

Figure 4-15. Map of Al content in kg/ha for the deep soil

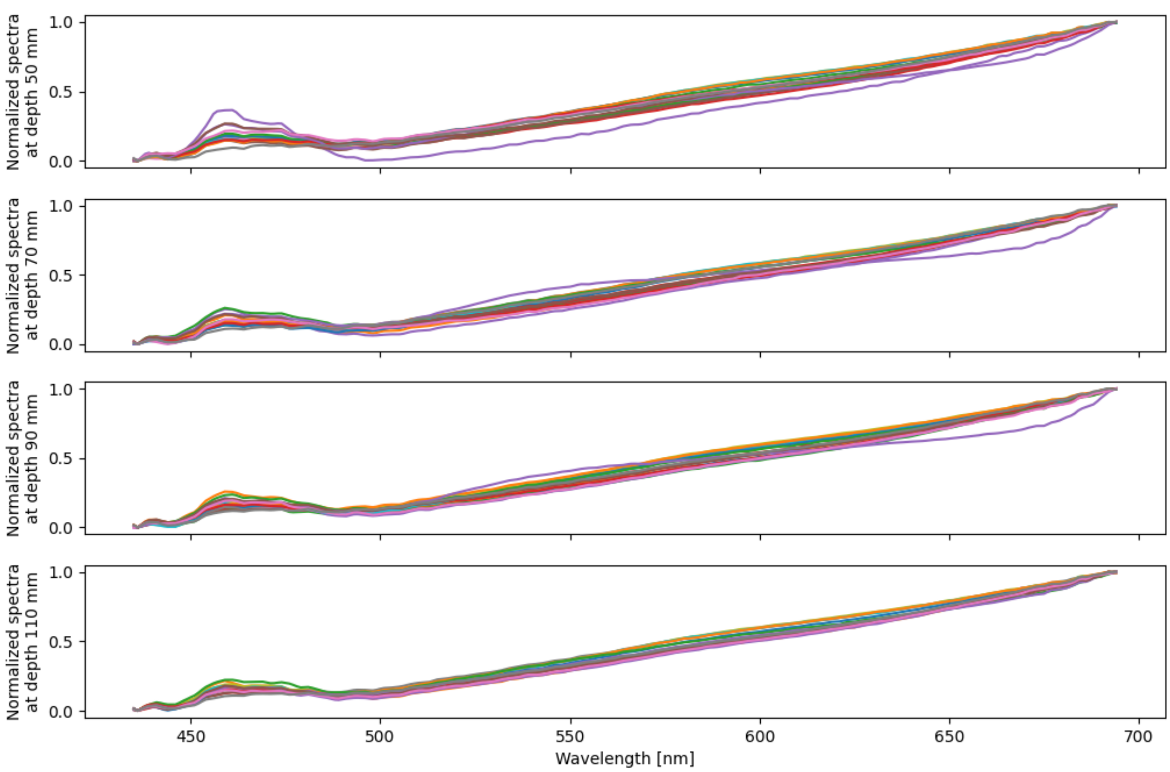


Figure 4-16. Subset of dataset showing normalized spectra in the visible range per depth

Considering the nature of the system as a prototype, the sensor probe of the PSS platform was missing certain features i.e. complete sealing from moisture. Because of this, the sensor readings had biases. The maximum depth for a hole opened with the PSS platform was 230 mm. While the duration of the measurement cycle was 12 minutes per location, it is no longer than what would be required for a traditional soil sampling setup. However, it is worth considering that measurements are being taken at four different depths, which would average about 3 minutes per sensor reading including all the operations of opening the borehole. That is without including the fact that once proper prediction models have been trained for the target soil properties, it is possible to obtain the information about the soil attributes in real-time, without the need to send the samples to a laboratory. The system could be further sped up by replacing some of the linear actuators with similar overall power but different gear ratios, making them faster at the price of a reduced maximum load. The linear actuators chosen in this design were deliberately over-dimensioned as a safety factor in case it was found more force was required, but for the vertical linear actuator, it was found that only about 300 lbs (1,300 N) were ever needed, even at peak load. As a reference, the vertical linear actuator was rated for a maximum load of 850 lbs (3,800 N). With that adjustment, the time for the complete measurement cycle could easily be reduced to less than 10 min. The same would hold true for the horizontal linear actuator.

The increased spatial resolution followed by precise control of the depth at which measurements are taken is also an advantage, as 3D soil mapping could be realized with this technology. Another consideration is the comfort of the operator, who could collect all the data from inside the cabin of the vehicle without being exposed to excessive sun when it is hot or wind chill when it is cold. Finally, it is possible to mount this system on an autonomous vehicle to minimize downtime.

4.4 Conclusions

The proposed PSS platform provides the ability to map soil data across fields at controlled depths. Successful completion of drilling and probe insertion was achieved. The system collected measurements with an average time of 3 min per measurement at multiple depths. The design of the platform had considerations of weight, power, and size constraints. Soil modelling techniques can benefit from the ability to acquire this data in an efficient manner. The advantages of this system in terms of ergonomics and multi-depth positioning provide an alternative to promote the adoption of PSS technologies, especially in the context of low-cost platforms. Further research is required to integrate prediction models into the measurement cycle to provide complete *in situ* and real-time soil analysis.

References for Chapter 4

Adamchuk, V. I., Dhawale, N., & Rene-Laforest, F. (2014). Development of an on-the-spot analyzer for measuring soil chemical properties. *Proceedings of 12th International Conference on Precision Agriculture*, July 20-23, 2014, Sacramento, California, USA.

Nocita, M., Stevens A., van Wesemael, B., Aitkenhead, M., Bachmann, M., Barthes, B., Ben Dor, E., Brown, D. J., Clairotte, M., Csorba, A., Dardenne, P., Dematte, J., Genot, V., Guerrero, C., Knadel, M., Montanarella, L., Noon, C., Ramirez-Lopez, L., Robertson, J., Sakai, H., Soriano-Disla, J. M., Shepherd, K. D., Stenberg, B., Towett, E., Vargas, R., & Wetterlind, J. (2015). Soil Spectroscopy: An alternative to wet chemistry for soil monitoring. In D. Sparks (Ed.), *Advances in Agronomy*, Vol. 132 (pp. 139-159). Academic Press.

Stenberg, B., Viscarra-Rossel, R. A., Mouazen, A. M., & Wetterlind, J. (2010). Visible and Near Infrared Spectroscopy in Soil Science. In D. Sparks (Ed.), *Advances in Agronomy*, Vol. 107 (pp. 163-215). Academic Press.

Connecting text to Chapter 5

Chapter 5 is related to the third objective of this study listed in Chapter 1. The previous chapter presented the development of a PSS platform to complement the HTPP platform described in Chapter 3. The following chapter proposes a technique for relating the soil spectral response with the soil chemical properties and then discusses the combination of the data from both platforms with the aim of providing benefits with on-farm experimentation. The results of this project were prepared in a manuscript for the journal listed below.

R. Buelvas, V.I. Adamchuk, V. Hoyos-Villegas, G. Mangeat, & B. de Leener. (2021). Combination of phenotyping and envirotyping as a tool to assess bean growth and yield. *Precision Agriculture*. (To be submitted)

CHAPTER 5 Combined analysis of data collected from HTPP and PSS platforms

Abstract

The implementation of a novel similarity-based prediction algorithm was used to find models describing the relationship between soil spectroscopy and the soil chemical properties. Coefficients of determination around 0.5 were achieved. A dataset was assembled from the measurements of HTPP and PSS platforms. The effects of soil and weather factors were considered during the evaluation of bean varieties by physical traits. The proposed framework for on-farm experimentation could be beneficial for growers, plant breeders, and the agricultural industry as a whole.

Keywords: Chemometrics, Data analysis, Modelling, Plants, Spectroscopy, Soil

5.1 Introduction

Plant phenotype is produced as the result of the combined effect of the plant's genome, the management practices used, and the environment to which the plant must adapt (Kanso *et al.*, 2020). When designing on-farm experiments, the cultivars of the plants and the treatments to be applied are known beforehand. However, spatial and temporal variability in weather and soil conditions have to be accounted for during the experiment. Randomized designs are used to minimize the effect such factors could play in the analyzed variables, but as the complexity of experimental design increases, the number of replicates required scales up as well. An alternative approach is to measure and characterize the environmental variables across the experiment and subtract their effect from that of other factors of plant growth, allowing for better comparisons

(Reynolds *et al.*, 2020). This approach is better suited for compounding data collected from multiple experiments, in different years and geographical locations, to build plant models that generalize under diverse scenarios. Collecting the data required to feed the training of the algorithms to build models in the most efficient way is a current topic of research, as well as finding the best techniques for extracting information from the raw data (Coppens *et al*., 2017). Multiple techniques are used for analyzing data in the agricultural context, including modern and complex methods applying Artificial Intelligence (AI). A method of this type was used in Shinde (2017) to optimize fertilizer applications.

In this research, a data workflow was proposed for comparing different cultivars in terms of certain physical traits while considering environmental factors. The starting point is the type of data collected from the HTPP and PSS platforms developed in Chapters 3 and 4. Finding models that relate soil spectral response with its chemical properties is key to this process.

5.2 Materials and methods

5.2.1 Prediction of soil chemical properties

The similarity method presented in Shinde (2017) is based on Equations (5-1) to (5-3).

$$\lambda_{j,u} = \prod_{k=1}^{K} \left(1 - \delta\lambda_k \left|\frac{x_{k,j} - x_{k,u}}{x_{k,max} - x_{k,min}}\right|\right)^q \qquad (5-1)$$

$$\varepsilon_{i,j} = (Y_i - Y_j)\lambda_j \qquad (5-2)$$

$$p(Y_i) = \frac{\frac{e^{-\frac{(0-avg(\varepsilon_i))^2}{2*std(\varepsilon_i)}}}{\sqrt{2\pi}*std(\varepsilon_i)}}{\sum_{i=1}^{N}\frac{e^{-\frac{(0-avg(\varepsilon_i))^2}{2*std(\varepsilon_i)}}}{\sqrt{2\pi}*std(\varepsilon_i)}} \quad (5-3)$$

In the above equations, λ are the similarity coefficients among the observations and ε is the estimated error for a possible value Y_i. Bringing these general equations to the context of chemometrics, the feature x_k would correspond to the spectral response of the soil at specific wavelengths and K the total number of such features in the bandwidth. The values of Y_i would be all the possible values that the chemical property in question could take e.g. 1 to 14 in the case of pH. $p(Y_i)$ is the probability that Y_i has 0 error i.e. meaning it is the correct prediction. The result of this process is a probability distribution which covers all possible values of the target variable. To make this computation practical, a predefined range and resolution must be chosen for candidate values, Y_i. Table 5-1 lists the ranges considered for the soil properties in the dataset. For all properties, the resolution was chosen so that 400 values were evaluated in the range. Finally, an individual prediction can be chosen by finding the center of mass of the probability density function or by choosing the value at which the peak probability is found.

Table 5-1. Ranges of evaluated values for each soil property

Soil property	Minimum value	Maximum value
Water pH	3	11
Buffer pH	3	11
Organic matter	0	8
P	3	500
K	30	1800
Ca	100	17500
Mg	30	2500
Al	200	1800
CEC	5	50
P-Al	0.0002	1

Because different types of transformations and pre-processing could be applied to the soil spectral response, Equation (5-1) was replaced by Equation (5-4), where the Kth root is evaluated to make the values of λ independent of the number of features evaluated. This makes the formula akin to a geometrical mean. The values of q and $\delta\lambda_k$ are considered hyperparameters of the method.

$$\lambda_{j,u} = \sqrt[K]{\prod_{k=1}^{K}\left(1 - \delta\lambda_k \left|\frac{x_{k,j} - x_{k,u}}{x_{k,max} - x_{k,min}}\right|\right)^q} = \prod_{k=1}^{K}\left(1 - \delta\lambda_k \left|\frac{x_{k,j} - x_{k,u}}{x_{k,max} - x_{k,min}}\right|\right)^{q/K} \qquad (5-4)$$

An implementation of this algorithm was done by creating a class whose parents are BaseEstimator and RegressorMixin from scikit-learn (Pedregosa *et al.*, 2011). This allows compatibility with other functions of the library to facilitate its use. A tool to draw different profiles of $\delta\lambda_k$ was also created. Multiple types of transformations were applied and evaluated with the regression method, as indicated by Table 5-2. The parameters q and $\delta\lambda_k$ were optimized by grid search among a list of predefined values.

The proposed regression method was trained with data collected in 2019 using the ChrysaLabs probe (ChrysaLabs, Inc., Montréal, QC, Canada) for the soil spectral response and traditional laboratory analysis for the target chemical properties. The dataset included measurements of water pH, buffer pH, organic matter, P, K, Ca, Mg, Al, P/Al ratio, and CEC for 240 locations from 7 different fields. The data was collected by manually opening a borehole and inserting the probe. 10-fold cross-validation was used and the coefficients of determination and RMSE values were used as performance metrics. Other methods, including linear regression, Support Vector Regression (SVR), and random forests, were also used with this dataset for comparison..

Table 5-2. Transformations applied to soil spectra

Transformation name	Description
Normalization	Apply linear rescaling to the range [0,1]
Standardization	Subtract average and divide by standard deviation
Power transformation	Apply Yeo-Johnson method (Yeo & Johnson, 2000)
Filtering	Apply Savitzky-Golay filter
Differentiation	Compute numeric derivation
Differentiation & Division	Compute numeric derivation and divide by original signal
Detrend	Remove linear trend along axis from data
Log	Apply natural logarithm
Root	Apply square root
FFT	Apply 1-D discrete Fourier transform
Square	Elevate to the power of 2
Inverse	Elevate to the power of -1
PCA	Reduce to first n principal components. Evaluated values of n included 5, 7, 9, and 11
Function concatenation	Serial combination of other transformation methods. For example, normalize the original spectra and then detrend the normalized spectra
Vector concatenation	Parallel combination of other transformation methods. For example, normalize the original spectra, detrend the original spectra, and then create a new vector with both

5.2.2 Comparison of crop varieties

Plant data was collected with the vehicle-mounted HTPP platform developed in Chapter 3 during the summer 2020. Thirty-six different bean (*Phaseolus vulgaris*) varieties/cultivars were sown on June 1st, 2020 divided among 242 plots, arranged in a grid of 11 x 22. This experiment took place in a field of the Emile A. Lods Agronomy Research Center of Macdonald Campus, McGill University. All plots were the same size (5 m by 3 m, for a total area of 15 m^2) and included 4 crop rows. Data started to be collected on July 28th, 2020, the first set of dates for surveying was dedicated to fine-tune some of the system parameters. Table 5-3 illustrates the dates of data collection that were used for the analysis. On these dates, the vehicle-mounted HTPP platform would drive through the alleys in between plots as close to a constant speed of 0.8 m/s as possible for the operator to collect measurements on-the-go from each side. The variables measured included NDRE, NDVI, CI, and height. The plants were harvested on October 6th, 2020 and the grain yield was measured. For each plot, only the 2 innermost crop rows were counted for the yield measurement.

Table 5-3. Dates of data collection for crop experiment

Date	Days After Seeding
August 17th, 2020	77
August 20th, 2020	80
August 24th, 2020	84
August 27th, 2020	87
August 31st, 2020	91
September 3rd, 2020	94

Soil samples were collected with a generic soil sampler (JMC, Clements Associates Inc., Newton, Indiana, USA) and with the prototype PSS platform developed in Chapter 4 in the same field of the Emile A. Lods Agronomy Research Center of Macdonald Campus, McGill University where the beans had been planted. Six locations were measured, as shown in Figure 5-1. At each location, the prototype PSS platform collected the spectral response of the soil at four different depths, by placing the middle of the probe's window at 50, 70, 90, and 110 mm below the surface level. Then, four more holes were made with the soil sampler to the sides of the original hole made by the PSS platform, as illustrated by Figure 5-2. The soil cores, collected by the soil sampler, were divided into shallow (between surface level and 152 mm) and deep (between 152 mm and 305 mm), stored in soil bags and labelled for later, standard laboratory analysis. The soil sampling was done on October 23rd and 24th, 2020. The laboratory analysis was performed on October 29th, 2020 by A & L Labs (A & L Canada Laboratories Inc., London, ON, Canada) and measured organic matter content, P, K, Mg, Ca, Na, Al, K/Mg ratio, water pH, buffer pH, Cation Exchange Capacity (CEC), and percentages of saturation for P, K, Mg, Ca, Na, Al, and H. Maps were made with ArcGIS Online (ESRI, Redlands, CA, USA). The measurements from those six locations were interpolated using Inverse Distance Weighting (IDW). The values were then evaluated for all the center locations of the plots where beans were planted.

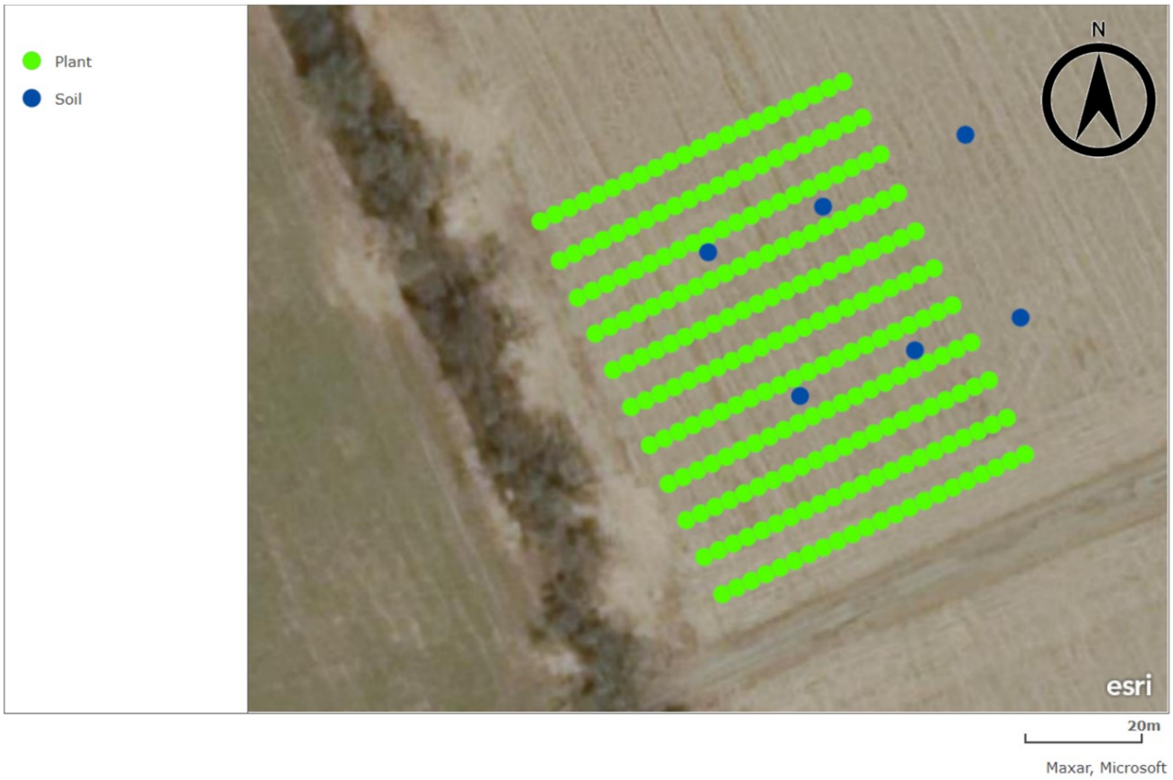


Figure 5-1. Map of locations where plant and soil measurements were taken

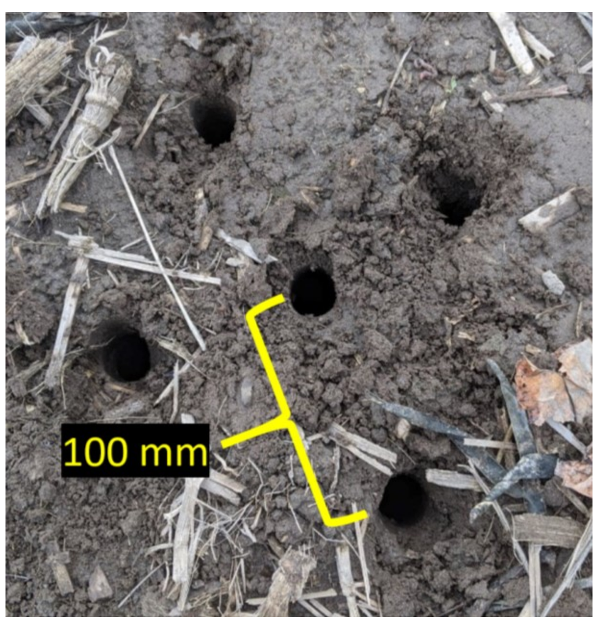


Figure 5-2. Picture of hole arrangement

For the plots where beans were planted, two different treatments were applied: row spacing being either 22 in. (559 mm) or 30 in. (762 mm), and the plant density 10,000 plants/acre (~24,700 plants/ha) or 15,000 plants/acre (~37,100 plants/ha). The distribution of the treatments is shown in Figure 5-3 and 5-4.

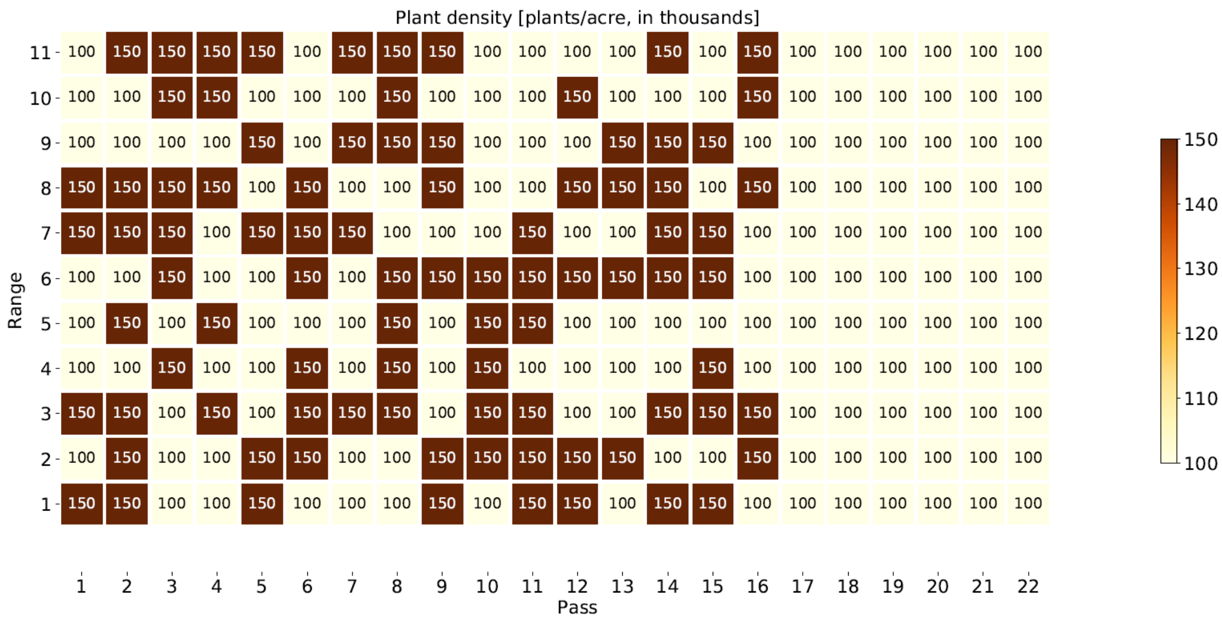


Figure 5-3. Map of the first treatment: plant density

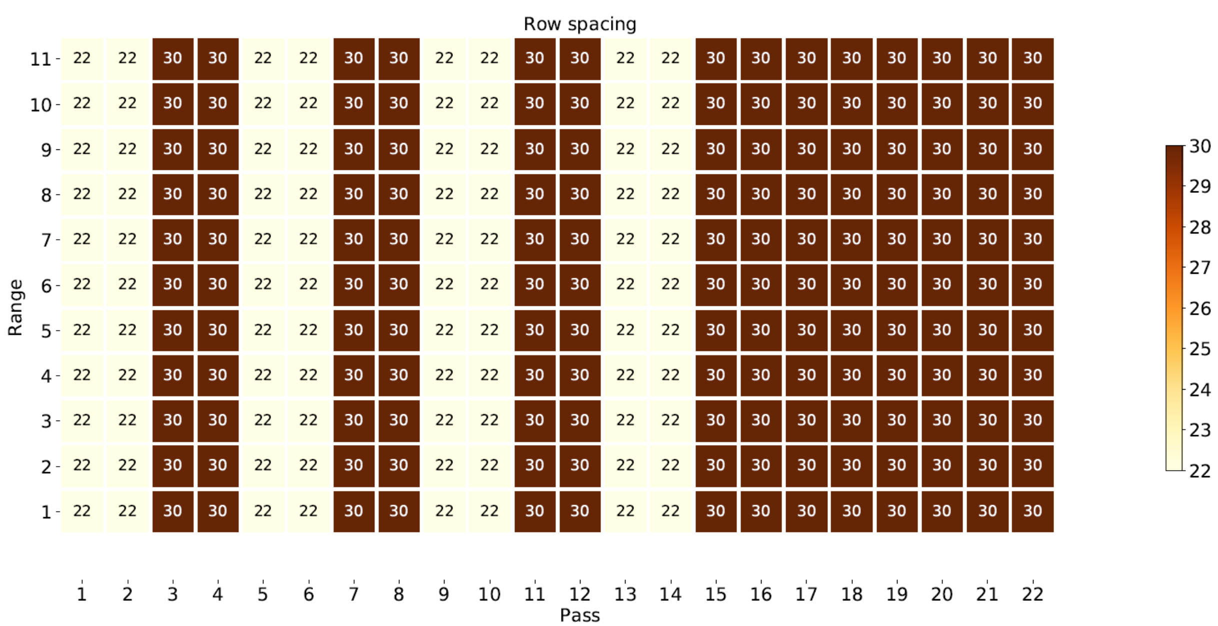


Figure 5-4. Map of the second treatment: row spacing

For processing, linear models were built where the phenotypical data for each of the measurement dates plus the yield were the dependent variable and the management practices, soil, and measured weather factors, as shown in Figure 5-5, were independent variables. Each linear model was iteratively evaluated by adding new variables, one at a time out of the set of predictors. The coefficient of determination was evaluated for each date and then averaged across all dates. The models which maximized the coefficient of determination for all six dates were chosen as a way to prevent overfitting for specific dates. The weather variables were considered in four ways:

(a) Only the weather data of the same date as the phenotypical data was used.

(b) Only the weather data of the previous measured date as the phenotypical data was used.

(c) The average of all weather data until the same date as the phenotypical data was used.

(d) The average of all weather data until the previous measured date as the phenotypical data was used.

For the first date of measurements, since there are no previous days, only method (a) was used. Once the best models were found, the residuals were computed by subtracting the predicted value from the actual value. Then a one-way ANOVA was performed with the variety as the explanatory factor. Tukey's test was performed to evaluate the effect of the varieties on the residuals. Only those varieties with more than 10 replicates, labelled as Dataset B in Figure 5-6, were considered in the statistical analysis.

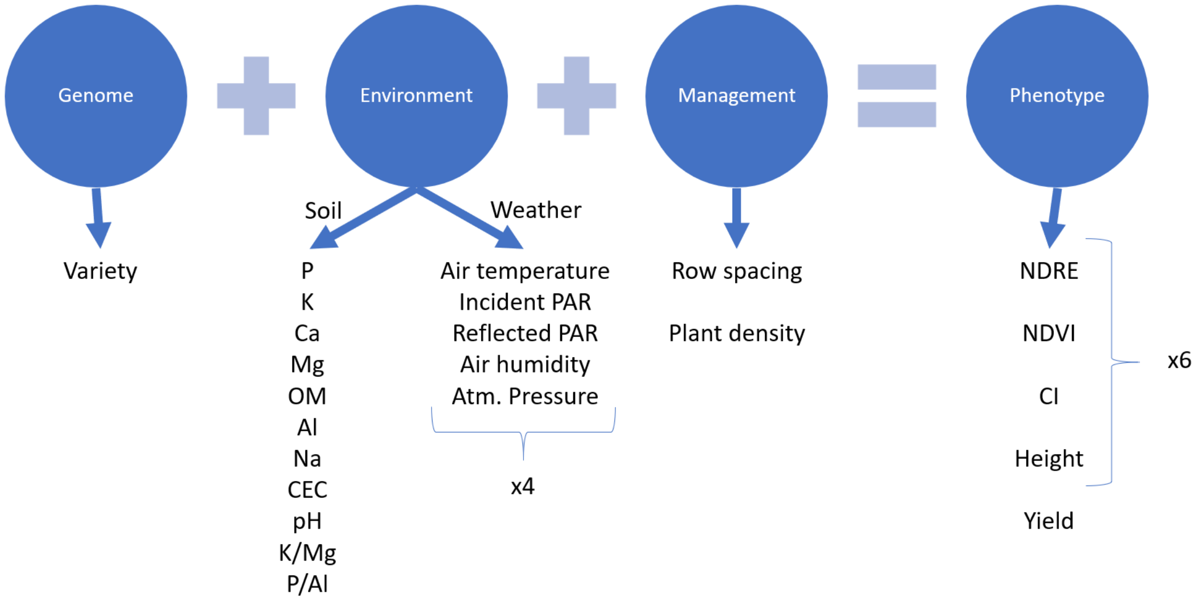


Figure 5-5. Diagram of variables used in models

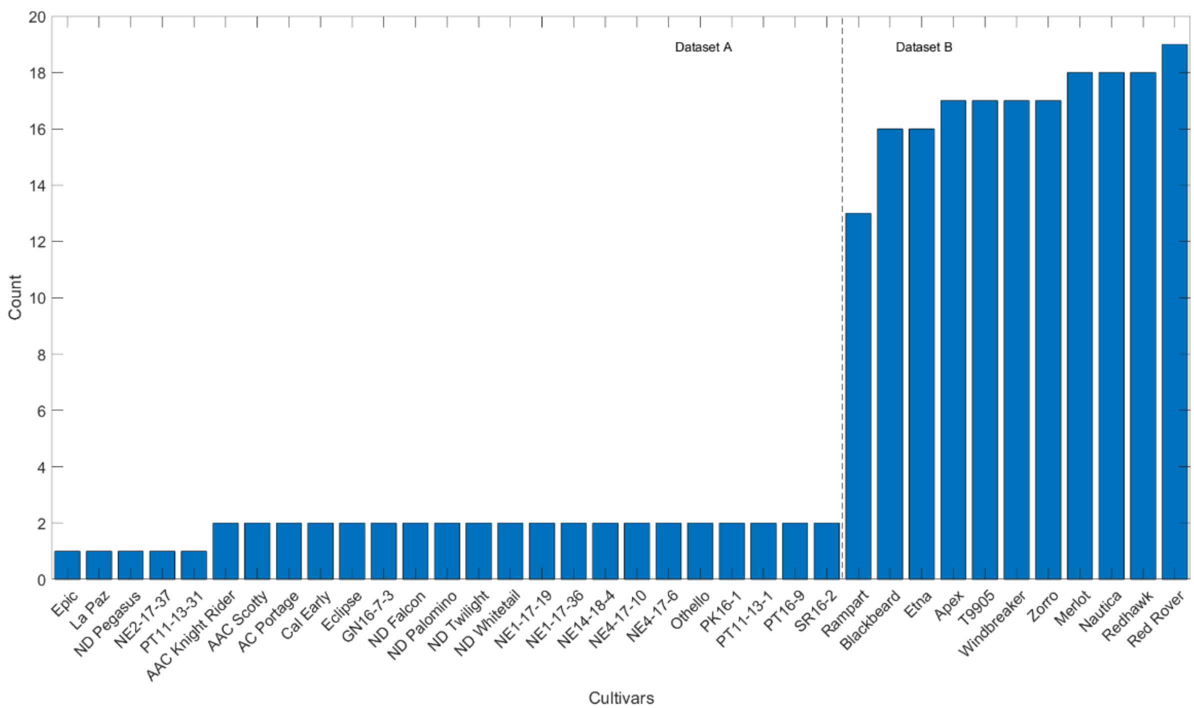


Figure 5-6. Histogram of replicates per bean variety/cultivar

5.3 Results and discussion

5.3.1 Prediction of soil chemical properties

Figure 5-7 shows the process that a single point goes through in the prediction process when the similarity-based procedure is used. While a single prediction is produced at the end, the probability distribution itself contains other information which might be relevant for decision-making.

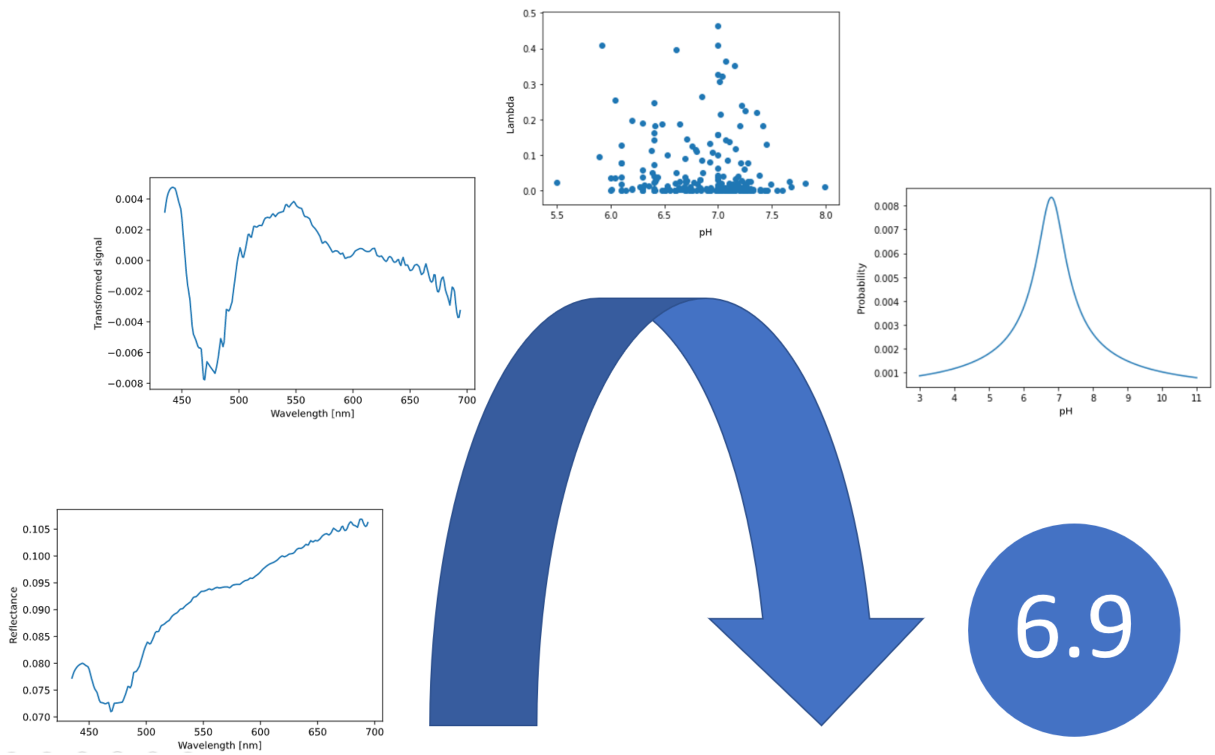


Figure 5-7. Diagram describing data stages through the similarity-based procedure

Figure 5-8 shows, in the case of CEC prediction, the RMSE of different supervised learning methods, including the similarity-based procedure proposed. For each, the best-performing feature transformations were used. Table 5-4 illustrates the best results for each of the measured soil properties by using both the similarity-based procedure and the best of the other methods.

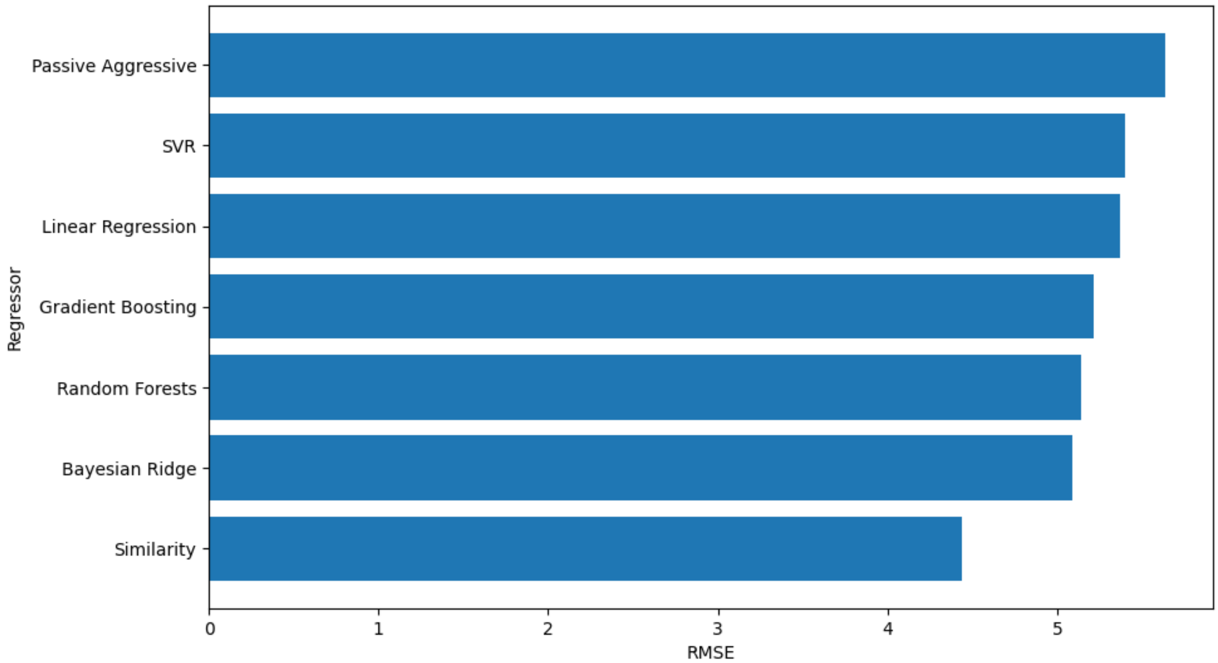


Figure 5-8. Comparison of RMSE values of CEC prediction for different methods

Table 5-4. Prediction performance metrics for each chemical property

	Similarity		Other methods		
Property	**R^2**	**RMSE**	**Method**	**R^2**	**RMSE**
Water pH	0.43	0.42	Bayesian Ridge	0.50	0.39
Buffer pH	0.41	0.32	SVR	0.40	0.32
organic matter, %	0.45	0.85	Bayesian Ridge	0.50	0.81
P, ppm	0.59	50.1	Bayesian Ridge	0.50	55.5
K, ppm	0.37	199	Bayesian Ridge	0.49	178
Ca, ppm	0.49	2170	Bayesian Ridge	0.63	1845
Mg, ppm	0.61	332	SVR	0.63	322
Al, ppm	0.20	148	Bayesian Ridge	0.36	132
P-Al	0.49	0.074	Bayesian Ridge	0.40	0.081
CEC, mg eq./100 g	0.60	4.44	Bayesian Ridge	0.48	5.09

While in some cases, as for P and CEC, the proposed algorithm performs better than the other methods evaluated, in others it underperforms, as for K and Ca. Overall, it is in the same order of magnitude and might be considered a valuable addition to the toolset of supervised learning algorithms. A limiting factor to the performance of the similarity-based procedure is the performance when few points with similar features are in the dataset. This results in the algorithm are especially susceptible to unbalanced training data.

5.3.2 Comparison of crop varieties

Figure 5-9 shows an example of the interpolated soil properties for the case of P content. Using NDRE as an example, Figures 5-10 illustrates the regression fit with the environmental and managerial data. Figure 5-11 shows the residuals in terms of the NDRE model built from environmental and managerial data. The residuals would later be used as the input for the ANOVA and Tukey's tests. As this is still an intermediate step, it will not be necessarily randomly distributed.

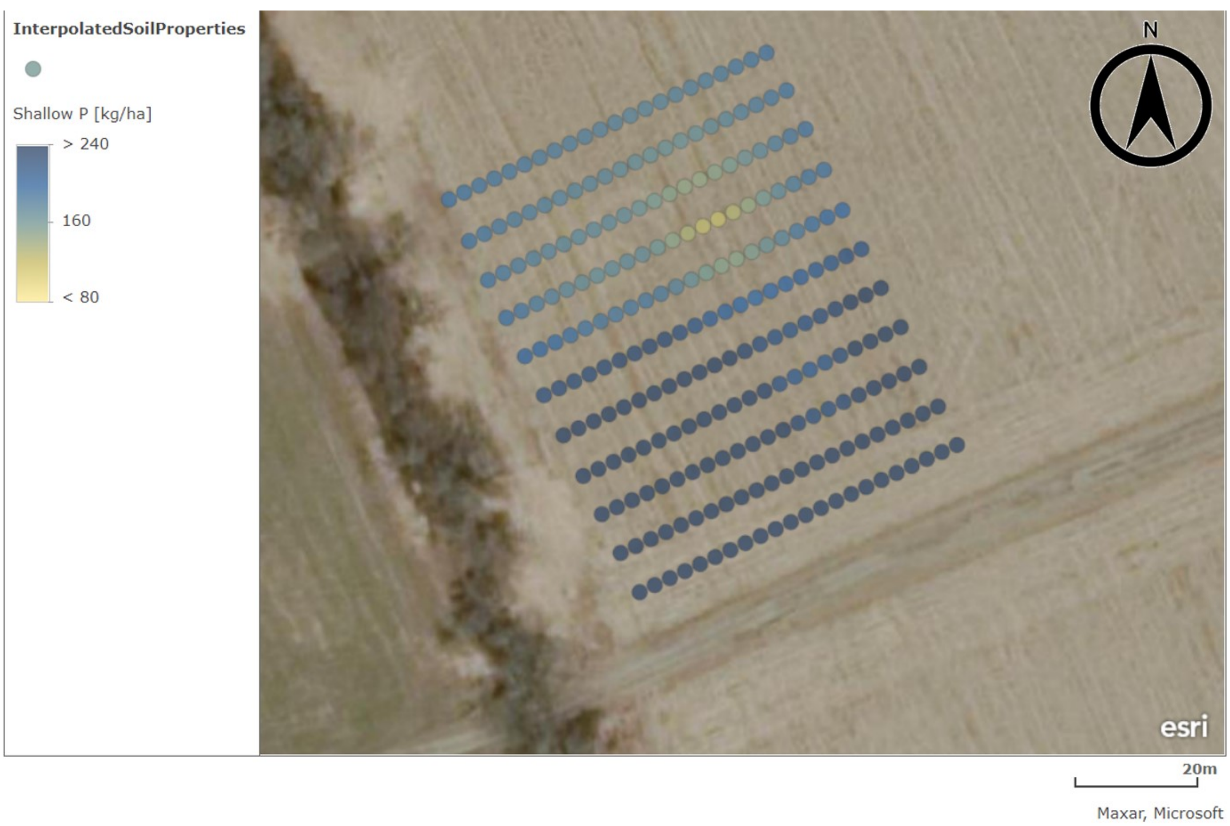


Figure 5-9. Map of interpolated P content for shallow soil evaluated at locations of bean plots

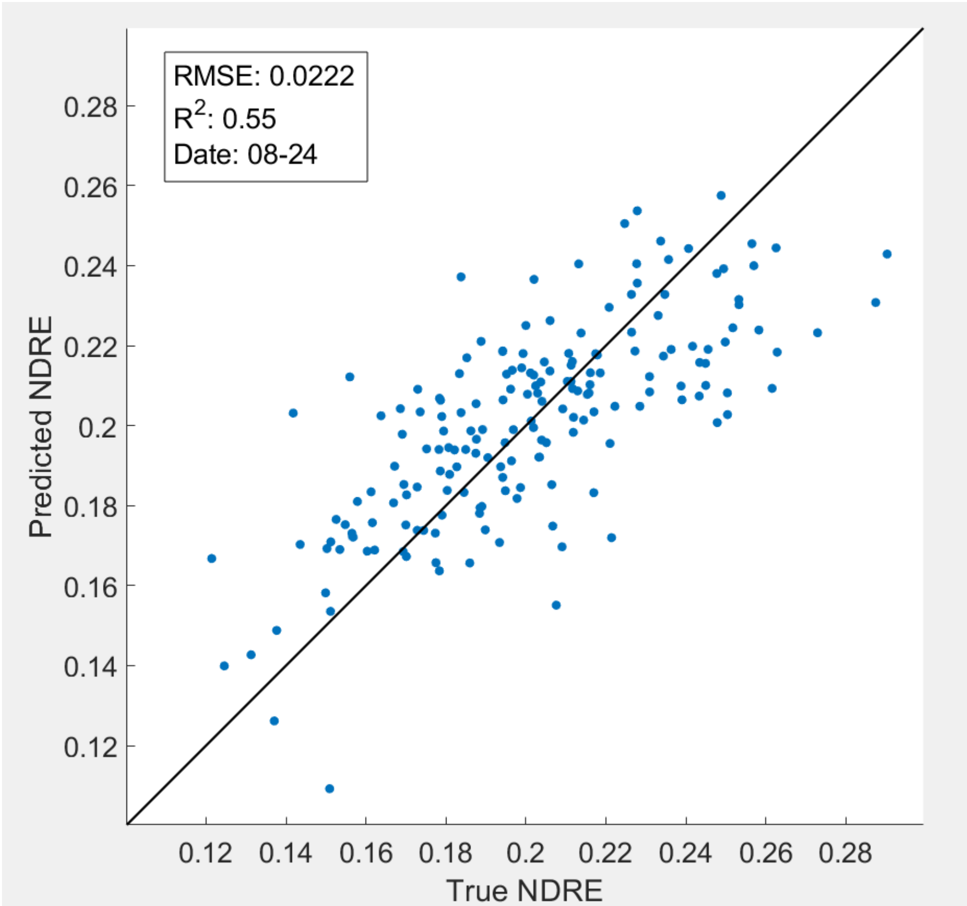


Figure 5-10. Function fit for NDRE model on third date of data collection (August 24th)

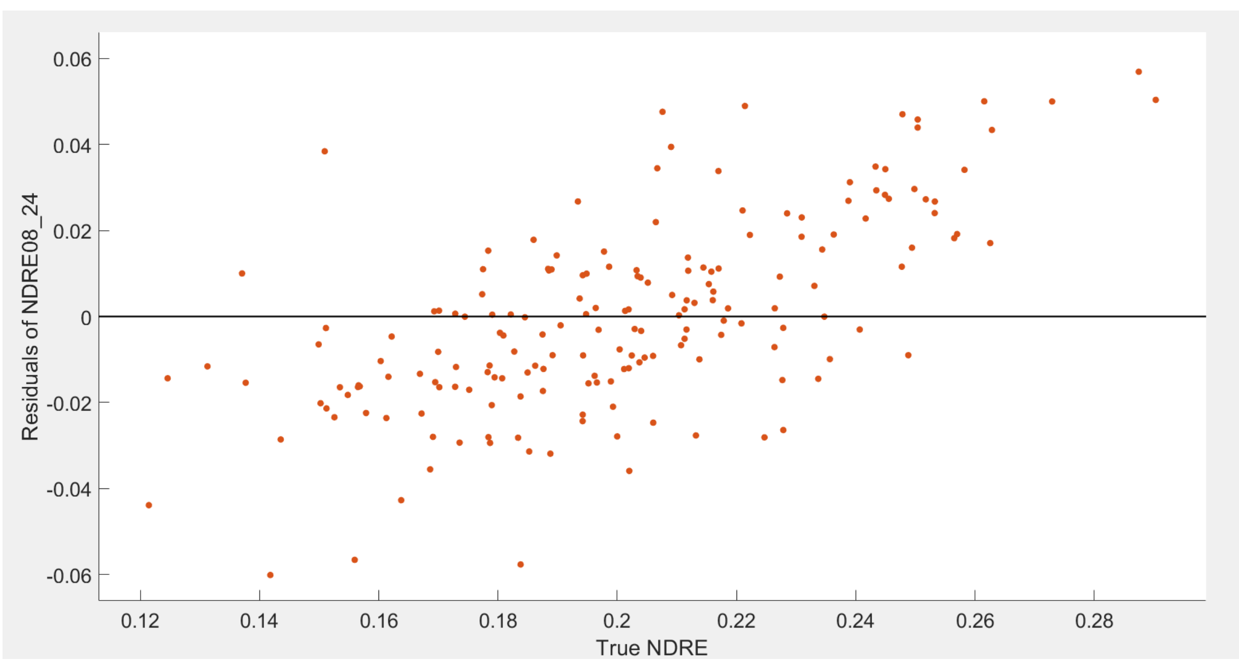


Figure 5-11. Residuals of NDRE model for third date of data collection (August 24th)

Figures 5-12 and 5-13 compare the effect of the variety on the yield before and after considering the effect of environmental and managerial factors. While there are many similarities between the two cases, a key difference is that the Blackbeard variety went from 4th to 2nd place after considering the additional factors. Following a similar procedure, Figures 5-14 and 5-15 show the results of the Tukey's test for NDRE for the first and last dates of data collection. Additional dates are shown in Appendices 6 to 9. Other properties followed a similar trend. When comparing the results of the effect of the variety on yield against the effect of the variety on other phenotypical traits, like the VIs, Blackbeard performed above average in general, but it was a source of discrepancy between the yield and VIs. Only by evaluating with a holistic view that includes the additional factors is it possible to realize that there is more agreement between both perspectives.

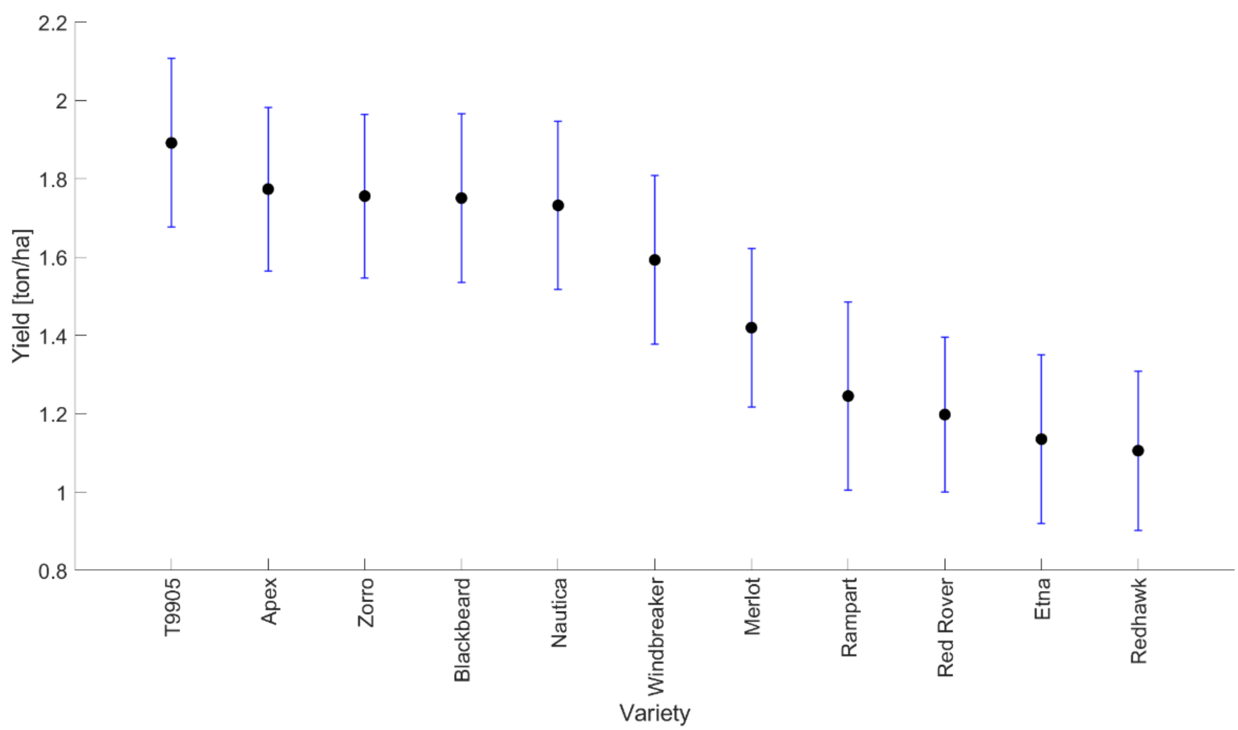


Figure 5-12. Result of Tukey's test for yield without considering other factors

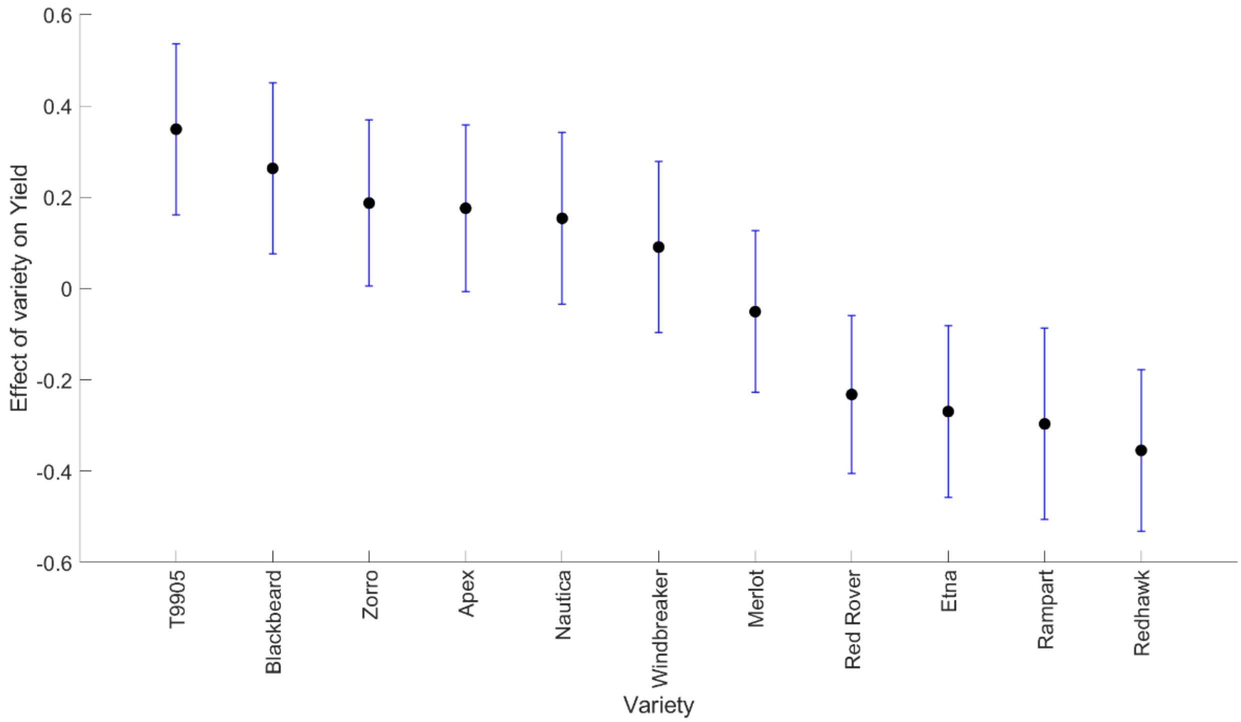


Figure 5-13. Result of Tukey's test for yield residuals after considering other factors

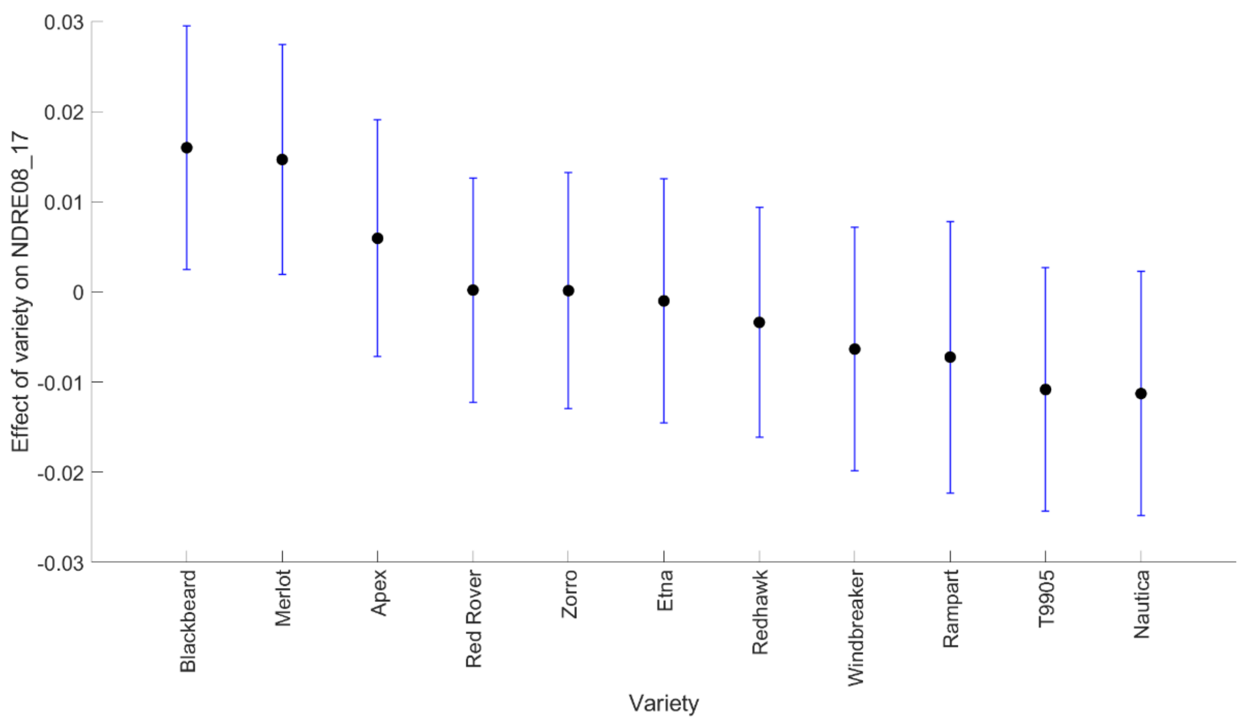


Figure 5-14. Result of Tukey's test for NDRE in the first date

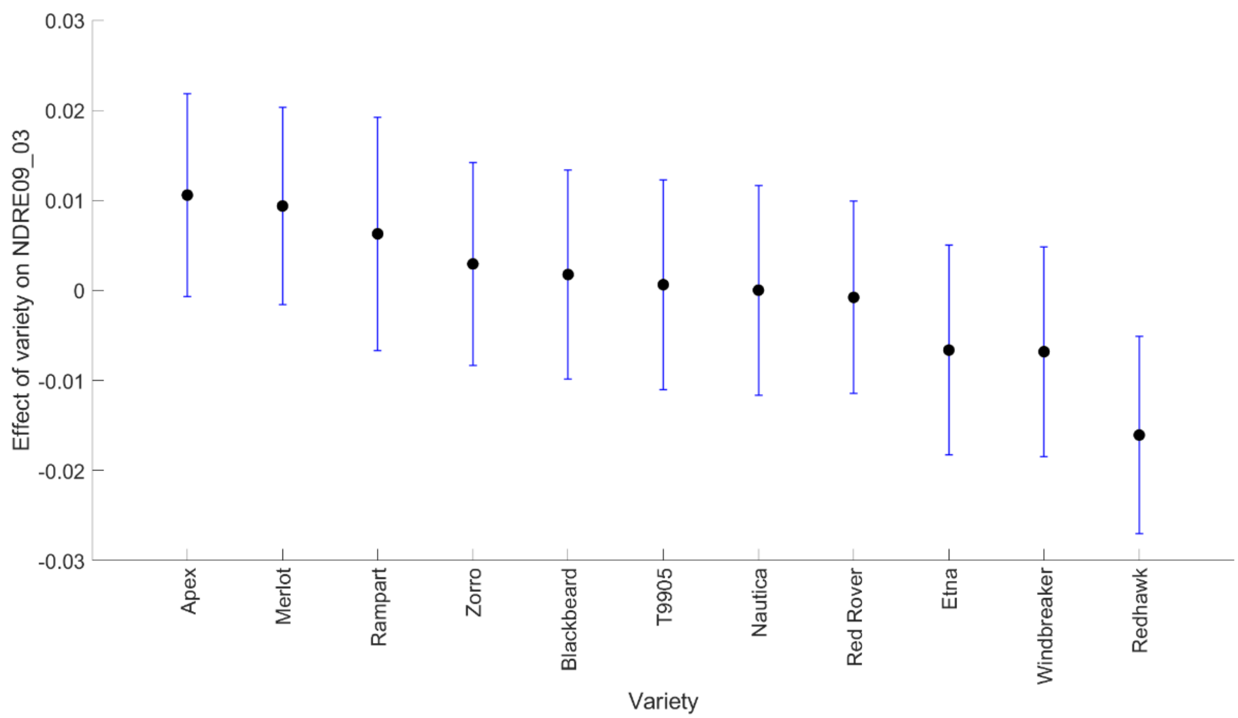


Figure 5-15. Result of Tukey's test for NDRE in the last date

Table 5-5 recounts which variables were selected in the iterative process to build the linear models to provide the best fit, following the procedure described in Section 5.2.2. For the weather factors, the number in between brackets identifies the approach used to leverage the fact that the variables had been measured multiple times. The first approach of just using the information from the same date as the target phenotypical variable was the best performing option in most cases. For certain dates, the previous measurement date provided an even better prediction, but as mentioned before, only the best performing option across all dates was selected. Only in the case of Height was it found that some of the other approaches worked best for all dates. Shallow organic matter was found in all models, confirming it as a variable that consistently has explanatory power. The same can be said about air temperature and PAR, which are key atmospheric variables, included in all the built models. More than these results, though, is the procedure to obtain this type of information that must be highlighted as a tool for future experiments in the field.

Table 5-5. Variables selected for linear models of each phenotypical trait that maximized coefficient of determination across all dates

Dependant variable	Independent variables
NDRE	Shallow organic matter Deep Na Air temperature [1] Incident PAR [1] Reflected PAR [1] Variety
NDVI	Shallow organic matter Deep Na Deep Mg Air temperature [1] Incident PAR [1] Reflected PAR [1] Variety
CI	Shallow organic matter Deep Na Deep Mg Air temperature [1] Incident PAR [1] Reflected PAR [1] Row spacing Variety
Height	Shallow organic matter Deep Ca Air temperature [3] Incident PAR [1] Reflected PAR [2] Row spacing Variety
Yield	Shallow organic matter Shallow Ca Air temperature [1] Incident PAR [1] Reflected PAR [1] Row spacing Variety

Figures 5-16 and 5-17 compare the effect sizes of the groups of explanatory variables for CI and Height on the second date. The component labelled as Error could be further reduced by the inclusion of some of the properties not directly considered, like soil moisture, which was not included in this analysis as it was not included in the standard laboratory analysis for reference.

While interaction factors among the current variables were evaluated, none were found to be significant in this dataset, so they were removed. Not in all cases, the management practices were significant. In general, the trend was that the error diminished with time, as it was possible to explain a larger portion of the variability in the plant traits through the evaluated factors. However, it was accompanied by a less significant portion of the total effect size due to the variety component.

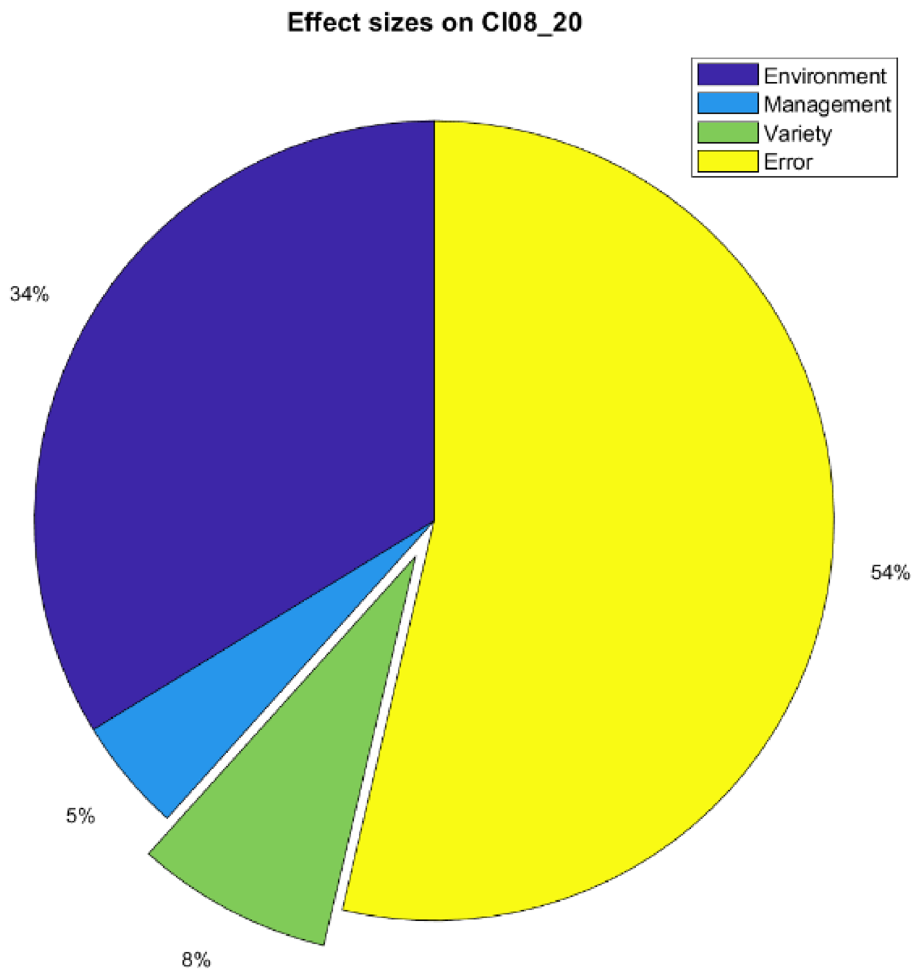


Figure 5-16. Effect sizes for CI on second date

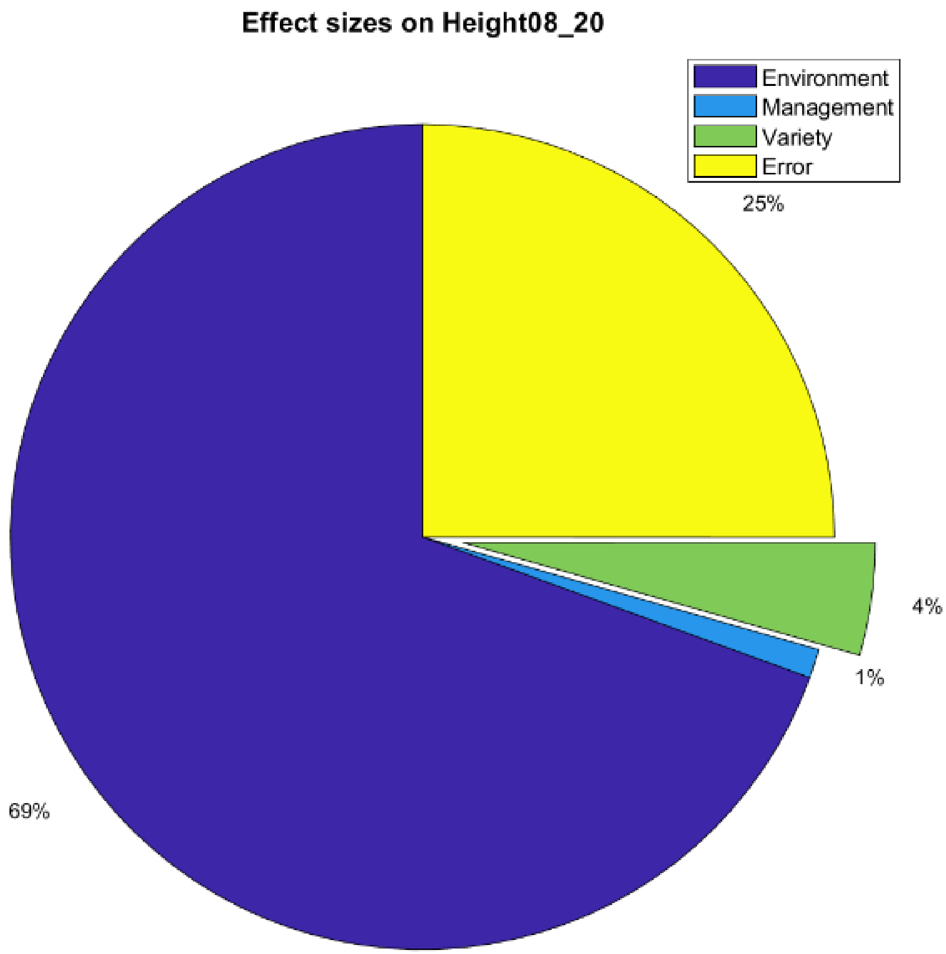


Figure 5-17. Effect sizes for Height on second date

5.4 Conclusions

Data processing methods were evaluated to combine soil spectral information, weather and plant phenotypical data to inform a comparison among a selection of cultivars. The application of chemometrics for predicting soil chemical properties from the soil spectral information as an intermediate step in the process is useful to provide insight into properties of interest for growers and to increase interpretability of analysis. The proposed methods had a decent performance, achieving coefficients of determination comparable to other regressors evaluated in the same dataset (around 0.5). It is expected that the evaluation of the algorithm with larger datasets would provide even better results.

More complex techniques could benefit as well from the general workflow discussed in this work, enabled by the availability of data gathered with mobile platforms. The holistic analysis allows for models to be built to serve as tools to evaluate effect sizes or relative importance of different factors of plant growth. Combining shallow and deep measurements of soil properties proved effective in explaining variability of the phenotypical traits in terms of environmental conditions. The central role of organic matter, air temperature, and incident PAR in the plant processes is evidenced by their relevance in the feature selection step of the model building. Between 50% and 75% of the variability in the phenotypical variables was explained by the linear models built including between 6 and 8 explanatory variables related to the environmental, managerial, and genetic factors of plant growth. Further research is required to evaluate this framework with more types of crops and in different geographical locations.

References for Chapter 5

Coppens, F., Wuyts, N., Inzé, D., & Dhondt, S. (2017). Unlocking the potential of plant phenotyping through integration and data-driven approaches. *Current Opinion in Systems Biology, 4*, 58-63.

Kanso, H., Quilot-Turion, B., Memah, M. M., Bernard, O., Gouzé, J. L., & Baldazzi, V. (2020). Reducing a model of sugar metabolism in peach to catch different patterns among genotypes. Mathematical Biosciences, 321, 108321.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research, 12*, 2825-2830.

Reynolds, M., Chapman, S., Crespo-Herrera, L., Molero, G., Mondal, S., Pequeno, D. N. L., Pinto, F., Pinera-Chavez, F. J., Poland, J., Rivera-Amado, C., Saint Pierre, C., & Sukumaran, S. (2020). Breeder friendly phenotyping. *Plant Science, 295*, 110396.

Shinde, S. (2017). Development of a numeric online decision support system for crop fertilizer optimization. McGill University. *MS Thesis*.

CHAPTER 6 Summary and general conclusions

6.1 Summary

This research was conducted to improve the efficiency of the data collection process for on-farm experimentation. There is a need to address this issue as the collection of plant phenotypical data is considered a bottleneck in scenarios of crop modelling, breeding, and on-farm experimentation. Progress in the understanding of genetic and environmental interactions depends on having access to large amounts of data, which would be too costly or time-consuming to compile using traditional methods. Novel platforms were developed for HTPP and PSS intended to operate in tandem, featuring modularity and designs that aim to facilitate the adoption of these kinds of tools. Additionally, a workflow of the data acquired from these platforms was evaluated to highlight the potential benefits of the technology.

In the first study, an HTPP platform was designed and evaluated. The system was capable of collecting on-the-go measurements from the aboveground plant canopy, as well as weather data. The sampling rate allowed for dense mapping of plant traits across a field and the GUI was easy to operate.

In the second study, a PSS platform was designed and evaluated. The prototype allowed for a drill to open a borehole and then insert a sensor probe inside the opened hole. The probe would then collect readings of the soil spectral response. The user is then able to collect on-the-spot measurements operating the system from the vehicle’s cabin.

In the third study, the data collected from both platforms was combined. The process started by using a similarity-based method to convert the spectral data from the PSS platform to the soil chemical properties. Later, linear models were built using the environmental variables to predict phenotypical data and then the residuals were tested to determine the effect the variety/cultivar had on them.

6.2 General conclusions

The combination of both platforms developed in this work along with the data processing techniques that were evaluated are an important research phase for extracting greater benefits from on-farm experimentation. The features of those systems show potential to improve the efficiency of data collection and represent a step towards automated field scouting. For example, the HTPP platform offers the ability to map fields covering 5400 m^2/h. The PSS platform can collect depth-specific measurements with an average measurement time of 3 min per depth at each location. With both systems improving the ergonomics for the operator, there is potential for more ambitious experimental designs when using these tools. The methodology presented in this thesis successfully collected and delivered valuable information for additional agronomical studies. Further research is required to extend these concepts to other crops and types of experiments.

CHAPTER 7 Contributions to knowledge and suggestions for future research

7.1 Contributions to knowledge

This dissertation contributes to the development of mapping sensor systems for crops, weather, and soil. The research resulted in the following contributions to knowledge:

1. A design for a quick-install HTPP platform for on-the-go measurements of plant canopy and weather in outdoor environments.
2. A design for a PSS platform for on-the-spot depth-specific measurements of soil spectral response directly in the field with drilling and tool-change capabilities.
3. An implementation of a similarity-based regression algorithm for the prediction of soil chemical properties based on the soil spectral response, with explicit probability distribution output for improved interpretability of results.
4. A framework for combining genetic, phenotypical, managerial and environmental factors of plant growth in the context of on-farm experimentation, combining and processing the data from both mobile platforms developed throughout this work.

7.2 Suggestions for future research

Future work for both sensing platforms should focus on collecting more evaluation data in different field conditions. The following areas of further research were identified throughout this dissertation work:

1. Multi-year experiments where the same crop varieties are evaluated. This would allow for the computation of validation errors for predictions made from the models built from the data collected by the platforms.

2. More crop species need to be examined with different dynamics, like leafy greens and potatoes. This would confirm or negate the applicability of the developed systems for different types of growing operations.
3. Non-linear and interaction models should be included in the comparison of crop varieties. This would allow for models to be built with better performance in terms of a larger portion of the variability in the phenotypical traits to be explained.
4. Manual measurements of plant biomass by early harvesting of sections of the field as an additional ground truth measurement should be performed. This would allow an extension of the models to additional plant phenotypical traits that are of interest for growers.

References

Adamchuk, V. I., Dhawale, N., & Rene-Laforest, F. (2014). Development of an on-the-spot analyzer for measuring soil chemical properties. *Proceedings of 12th International Conference on Precision Agriculture*, July 20-23, 2014, Sacramento, California, USA.Adamchuk, V., Allred, B., Doolittle, J., Grote, K., & Viscarra-Rossel, R. (2017). Tools for Proximal Soil. In C. Ditzler, K. Scheffe, & H. Monger, *Soil Survey Manual* (pp. 355-394). Washington: USDA.

Adamchuk, V., Ji, W., Viscarra-Rossel, R., Gebbers, R., & Tremblay, N. (2018). Proximal Soil and Plant Sensing. In D. Shannon, D. Clay, & N. Kitchen, *Precision Agriculture Basics* (pp. 123-144). Madison, WI: ASA, CSSA, and SSSA.

Andrade-Sanchez, P., Gore, M. A., Heun, J. T., Thorp, K. R., Carmo-Silva, A. E., French, A. N., Salvucci, M. E., & White, J. W. (2014). Development and evaluation of a field-based high-throughput phenotyping platform. *Functional Plant Biology, 41*, 68-79.

Barra, I., Haefele, S. M., Sakrabani, R., & Kebede, F. (2021). Soil spectroscopy with the use of chemometrics, machine learning and pre-processing techniques in soil diagnosis: Recent advances – A review. *Trends in Analytical Chemistry, 135*, 116166.

Buelvas, R. (2018). Crop canopy measurements using a low-cost laser for biomass estimation. McGill University. *MS Thesis*.

Buelvas, R., Adamchuk, V., Pouliot, A., Stromvik, M., & Whitmore, A. (2019). Development of a quick-install rapid phenotyping system. *2019 ASABE Annual International Meeting.* Boston, MA, USA.

Carra, J. B., Fabris, M., Dieckow, J., Brito, O. R., Vendrame, P. R. S., & dos Santos, L. M. (2019). Near-infrared spectroscopy coupled with chemometrics tools: a rapid and non-destructive alternative on soil evaluation. *Communications in Soil Science and Plant Analysis*, 50:4, 421-434.

Coppens, F., Wuyts, N., Inzé, D., & Dhondt, S. (2017). Unlocking the potential of plant phenotyping through integration and data-driven approaches. *Current Opinion in Systems Biology, 4*, 58-63.

Dainese, C., & Ercoli Finzi, A. (2006). SD2 - A cometary soil drill and sampler device. *Aerotecnica Missili e Spazio, 85*, 174-180.

Gehan, M. A., & Kellogg, E. A. (2017). High-throughput Phenotyping. *American Journal of Botany, 104*(4), 505-508.

Holzworth, D. P., Huth, N. I., deVoil, P. G., Zurcher, E. J., Herrmann, N. I., McLean, G., Chenu, K., van Oosterom, E. J., Snow, V., Murphy, C., Moore, A. D., Brown, H., Whish, J. P. M., Verrall, S., Fainges, J., Bell, L. W., Peake, A. S., Poulton, P. L., Hochman, Z., Thorburn, P. J., Gaydon, D. S., Dalgliesh, N. P., Rodriguez, D., Cox, H., Chapman, S., Doherty, A., Teixeira, E., Sharp, J., Cichota, R., Vogeler, I., Li, F. Y., Wang, E., Hammer, G. L., Robertson, M. J., Dimes, J. P., Whitbread, A. M., Hunt, J., van Rees, H., McClelland, T., Carberry, P. S., Hargreaves, J. N. G., MacLeod, N., McDonald, C., Harsdorf, J., Wedgwood, S., & Keating, B. A. (2014). APSIM – Evolution towards a new generation of agricultural systems simulation. *Environmental Modelling & Software, 62*, 327-350.

Hoogenboom, G., Porter, C. H., Boote, K. J., Shelia, V., Wilkens, P. W., Singh, U., White, J. W., Asseng, S., Lizaso, J. I., Moreno, L. P., Pavan, W., Ogoshi, R., Hunt, L. A., Tsuji, G. Y.,

& Jones, J. W. (2019). The DSSAT crop modeling ecosystem. In K. J. Boote, *Advances in Crop Modeling for a Sustainable Agriculture* (pp. 173-216). Cambridge, UK: Burleigh Dodds Science Publishing.

Hoyos-Villegas, V., Houx, J., Singh, S., & Fritschi, F. (2014). Ground-Based Digital Imaging as a Tool to Assess Soybean Growth and Yield. *Crop Science, 54*, 1756-1768.

Kanso, H., Quilot-Turion, B., Memah, M. M., Bernard, O., Gouzé, J. L., & Baldazzi, V. (2020). Reducing a model of sugar metabolism in peach to catch different patterns among genotypes. *Mathematical Biosciences*, 321, 108321.Knoch, D., Abbadi, A., Grandke, F., Meyer, R. C., Samans, B., Werner, C. R., Snowdon, R. J., & Altmann, T. (2020). Strong temporal dynamics of QTL action on plant growth progression revealed through high-throughput phenotyping in canola. *Plant Biotechnology Journal, 18*, 68-82.

Kuang, B., Mahmood, H. S., Quraishi, M. Z., Hoogmoed, W. B., Mouazen, A. M., & van Henten, E. J. (2012). Chapter four - Sensing Soil Properties in the Laboratory, In Situ, and On-Line: A Review. In D. L. Sparks, *Advances in agronomy* (Vol. 114, pp. 155-223). Academic Press.

Liu, L., Crowe, T. G., & Roberge, M. (2009). Sensor-based scouting algorithms for automatic sampling tasks in rough and large unstructured agricultural fields. *Transactions of the ASABE, 52*(1), 285-294.

Mir, R. R., Reynolds, M., Pinto, F., Khana, M. A., & Bhata, M. A. (2019). High-throughput phenotyping for crop improvement in the genomics era. *Plant Science, 282*, 60-72.

Nocita, M., Stevens A., van Wesemael, B., Aitkenhead, M., Bachmann, M., Barthes, B., Ben Dor, E., Brown, D. J., Clairotte, M., Csorba, A., Dardenne, P., Dematte, J., Genot, V., Guerrero, C., Knadel, M., Montanarella, L., Noon, C., Ramirez-Lopez, L., Robertson, J., Sakai, H., Soriano-Disla, J. M., Shepherd, K. D., Stenberg, B., Towett, E., Vargas, R., & Wetterlind, J. (2015). Soil Spectroscopy: An alternative to wet chemistry for soil monitoring. In D. Sparks (Ed.), *Advances in Agronomy*, Vol. 132 (pp. 139-159). Academic Press.

Nyeki, A., Kerepesi, C., Daroczy, B., Benczur, A., Milics, G., Kovacs, A. J., & Nemenyi, M. (2019). Maize yield prediction based on artificial intelligence using spatio-temporal data. *12th European Conference on Precision Agriculture* (pp. 1011-1017). Montpellier, France: Wageningen Academic Publishers.

Oger, B., Vismara, P., & Tisseyre, B. (2019). Combining target sampling with route-optimization to optimise yield estimation in viticulture. *12th European Conference on Precision Agriculture* (pp. 173-179). Montpellier, France: Wageningen Academic Publishers.

Olsson, T., Robertsson, A., & Johansson, R. (2015). Flexible force control for accurate low-cost robot drilling. *Proceedings 2007 IEEE International Conference on Robotics and Automation.* Rome, Italy, 2007, pp. 4770-4775.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research, 12*, 2825-2830.

Piepho, H.-P., Richter, C., Spilke, J., Hartung, K., Kunick, A., & Thole, H. (2011). Statistical aspects of on-farm experimentation. *Crop & Pasture Science, 62*(9), 721-735.

Pieruschka, R., & Schurr, U. (2019). Plant Phenotyping: Past, Present, and Future. *Plant Phenomics*, vol. 2019, Article ID 7507131.

Pouliot, A. (2016). Development of a Quick-Install Auto-Steering System for Agricultural Vehicles. McGill University. *MS Thesis*.

Ranjan, R., Chandel, A. K., Khot, L. R., Bahlol, H. Y., Zhou, J., Boydston, R. A., & Miklas, P. N. (2019). Irrigated pinto bean crop stress and yield assessment using ground based low altitude remote sensing technology. *Information Processing in Agriculture, 6*, 502-514.

Reynolds, D., Baret, F., Welcker, C., Bostrom, A., Ball, J., Cellini, F., Lorence, A., Chawade, A., Khafif, M., Noshita, K., Mueller-Linow, M., Zhou, J., & Tardieuc, F. (2019). What is cost-effective phenotyping? Optimizing costs for different scenarios. *Plant Science, 282*, 14-22.

Reynolds, M., Chapman, S., Crespo-Herrera, L., Molero, G., Mondal, S., Pequeno, D. N. L., Pinto, F., Pinera-Chavez, F. J., Poland, J., Rivera-Amado, C., Saint Pierre, C., & Sukumaran, S. (2020). Breeder friendly phenotyping. *Plant Science, 295*, 110396.

Rossi, R., Leolini, C., Costafreda-Aumedes, S., Leolini, L., Bindi, M., Zaldei, A., & Moriondo, M. (2020). Performances Evaluation of a Low-Cost Platform for High-Resolution Plant Phenotyping. *Sensors, 20*(11), 3150-3166.

Sanches, G. M., Graziano Magalhaes, P. S., & Junqueira Franco, H. C. (2019). Site-specific assessment of spatial and temporal variability of sugarcane yield related to soil attributes. *Geoderma*, 334, 90-98.

Schwarz, H. P., Rüger, P., Kicherer, A., & Töpfer, R. (2013). Development of an autonomous driven robotic platform used for ht-phenotyping in viticulture. *Synergy and Technical Development (Synergy2013).* October 13-19, 2013, Gödöllő, Hungary.

Shinde, S. (2017). Development of a numeric online decision support system for crop fertilizer optimization. McGill University. *MS Thesis*.

Stenberg, B., Viscarra-Rossel, R. A., Mouazen, A. M., & Wetterlind, J. (2010). Visible and Near Infrared Spectroscopy in Soil Science. *Advances in Agronomy, 107*, 163-215.

Tagarakis, A. C., Mimic, G., Vaessen, H. M., Rodriguez-Moreno, F., van Evert, F. K., & Ciric, V. (2019). Using the WOFOST crop growth model to assess within-field yield variability. *12th European Conference on Precision Agriculture* (pp. 91-97). Montpellier, France: Wageningen Academic Publishers.

Viscarra-Rossel, R. A., Adamchuk, V. I., Sudduth, K. A., McKenzie, N. J., & Lobsey, C. (2011). Proximal soil sensing: An effective approach for soil measurements in space and time. In D. Sparks, *Advances in Agronomy. Vol 113* (pp. 237-282). Academic Press.

Viscarra-Rossel, R., & Adamchuk, V. I. (2013). Proximal soil sensing. In M. Oliver, T. Bishop, & B. Marchant, *Precision agriculture for sustainability and environmental protection* (pp. 99-118). New York: Earthscan.

von Hebel, C., Matveeva, M., Verweij, E., Rademske, P., Kaufmann, M. S., Brogi, C., Vereecken, H., Rascher, U., & van der Kruk, J. (2018). Understanding Soil and Plant Interaction by Combining Ground-Based Quantitative Electromagnetic Induction and Airborne Hyperspectral Data. *Geophysical Research Letter*, 45, 7571-7579.

Xu, Y., Li, P., Zou, C., Lu, Y., Xie, C., Zhang, X., Prasanna, B. M., & Olsen, M. S. (2017). Enhancing genetic gain in the era of molecular breeding. *Journal of Experimental Botany, 68*(11), 2641-2666.

Xu, S., Zhao, Y., Wang, M., & Shi, X. (2018). Comparison of multivariate methods for estimating selected soil properties from intact soil cores of paddy fields by Vis-NIR spectroscopy. *Geoderma, 310*, 29-43.

Yang, W., Feng, H., Zhang, X., Zhang, J., Doonan, J. H., Batchelor, W. D., Xiong, L., & Yan, J. (2020). Crop Phenomics and High-Throughput Phenotyping: Past Decades, Current Challenges, and Future Perspectives. *Molecular Plant, 13*, 187-214.

Ye, T., Dewei, T., Zongquan, D., Shengyuan, J., & Qiquan, Q. (2015). Drilling power consumption and soil conveying volume performances of lunar sampling auger. *Chinese journal of mechanical engineering, 28*, 451-459.

Yeo, I. K., & Johnson, R. A. (2000). A new family of power transformations to improve normality or symmetry. *Biometrika, 87*(4), 954-959.

Zaman, Q. U., Schumann, A. W., & Miller, W. M. (2005). Variable rate nitrogen application in Florida citrus based on ultrasonically-sensed tree size. *Applied Engineering in Agriculture, 21*(3), 331-335.

Appendices

Appendix 1. Code snippet of MainWindow class for HTPP UI

```
""" Display main window """

# Author: Roberto Buelvas

from datetime import datetime
import time
import math
import os

import numpy as np
import wx

from .sensors import SensorHandler, setupGPSProjection, variables
from .ports_dialog import PortsDialog, devices
from .plot_notebook import Map, Plot, PlotNotebook
from .layout_dialog import LayoutDialog
from .cameras import CameraFrame


class MainWindow(wx.Frame):
    """ Class to define main window

    This class both creates the main window of the UI and controls the
    operations that are done within it.

    Attr:
        cfg (wx.ConfigBase): Settings are saved in here, which creates a file
            in a hidden folder to store values. These values are kept if the
            program stops running and even if the computers turns off. It uses
            a key-value system similar to dictionaries
        btn_connect (wx.ToggleButton): Button to toggle connect/disconnect
        btn_start (wx.ToggleButton): Button to toggle start/stop reading every second
        btn_test (wx.ToggleButton): Button to toggle in and out of 'Test Mode'
        btn_measure (wx.Button): Button to take a single set of readings
        logText (wx.TextCtrl): Control where information is logged
        timer (wx.Timer): Object to call a function periodically
        sensor_handler (SensorHandler): Object to control multiple sensors at once
        camera_frame (CameraFrame): Secondary frame to display video from cameras
        mapPanel (Plot): Panel containing the Figure where the map is drawn
        axes (dict): Dict to hold the Axes for each measured variables. Keys
            are of the format 'mL1/NDVI' or 'gR/Latitude'. Used to create the
            plots
        labels (list): List of all possible labels of the style mL1 or gR given
            the number of scaling sensors
        last_record (list): List showing positions in the text log where each
            set of measurements ends. Used to erase the last set of values from
            the log text
        num_readings (int): Stores how many sets of measurements have been taken
            in the current survey. Set back to 0 when log text is cleared or
            exported to file
    """

    def __init__(self, *args, **kwargs):
```

```
        """ Create new window """
        super(MainWindow, self).__init__(*args, **kwargs)
        self.cfg = wx.Config("HTPPconfig")
        self.cfg.WriteBool("notEmpty", True)
        self.labels = self.updateLabels()
        self.axes = {}
        self.reset()
        self.initUI()
        self.sensor_handler = None
        self.updateSensorHandler()
        self.camera_frame = None
        self.updateCameraFrame()
        self.camera_frame.Bind(wx.EVT_CLOSE, self.OnCameraClose)
        self.timer = wx.Timer(self, wx.Window.NewControlId())
        self.Bind(wx.EVT_TIMER, self.OnUpdate, id=self.timer.GetId())
        self.Bind(wx.EVT_CLOSE, self.OnClose)

    def initUI(self):
        """ Define window elements """
        # Toolbar
        menubar = wx.MenuBar()

        fileMenu = wx.Menu()
        newmi = wx.MenuItem(fileMenu, wx.ID_NEW, "&New")
        fileMenu.Append(newmi)
        self.Bind(wx.EVT_MENU, self.OnNew, newmi)
        savemi = wx.MenuItem(fileMenu, wx.ID_SAVE, "&Save")
        fileMenu.Append(savemi)
        self.Bind(wx.EVT_MENU, self.OnSave, savemi)
        fileMenu.AppendSeparator()
        qmi = wx.MenuItem(fileMenu, wx.ID_EXIT, "&Quit")
        fileMenu.Append(qmi)
        self.Bind(wx.EVT_MENU, self.OnQuit, qmi)
        menubar.Append(fileMenu, "&File")

        settingsMenu = wx.Menu()
        portsmi = wx.MenuItem(settingsMenu, wx.ID_PREFERENCES, "&Ports")
        settingsMenu.Append(portsmi)
        self.Bind(wx.EVT_MENU, self.OnPorts, portsmi)
        layoutmi = wx.MenuItem(settingsMenu, wx.ID_ANY, "&Layout")
        settingsMenu.Append(layoutmi)
        self.Bind(wx.EVT_MENU, self.OnLayout, layoutmi)
        clearmi = wx.MenuItem(settingsMenu, wx.ID_ANY, "&Clear")
        settingsMenu.Append(clearmi)
        self.Bind(wx.EVT_MENU, self.OnClear, clearmi)
        menubar.Append(settingsMenu, "&Settings")

        viewMenu = wx.Menu()
        self.camerami = viewMenu.AppendCheckItem(wx.ID_ANY, "&Camera")
        self.Bind(wx.EVT_MENU, self.OnCamera, self.camerami)
        menubar.Append(viewMenu, "&View")

        helpMenu = wx.Menu()
        aboutmi = wx.MenuItem(helpMenu, wx.ID_ABOUT, "&About")
        helpMenu.Append(aboutmi)
        self.Bind(wx.EVT_MENU, self.OnAbout, aboutmi)
        menubar.Append(helpMenu, "&Help")

        self.SetMenuBar(menubar)

        # Window
```

```
        backgroundPanel = wx.Panel(self)
        backgroundPanel.SetBackgroundColour("#ededed")

        font = wx.SystemSettings.GetFont(wx.SYS_SYSTEM_FONT)
        font.SetPointSize(9)

        outerBox = wx.BoxSizer(wx.HORIZONTAL)

        leftBox = wx.BoxSizer(wx.VERTICAL)
        st1 = wx.StaticText(backgroundPanel, label="Map:")
        self.mapPanel = Map(backgroundPanel)
        st2 = wx.StaticText(backgroundPanel, label="Log:")
        self.logText = wx.TextCtrl(
            backgroundPanel, style=wx.TE_MULTILINE | wx.TE_READONLY
        )
        self.logSettings()
        leftBox.Add(st1, proportion=0, flag=wx.ALL)
        leftBox.Add(self.mapPanel, wx.ID_ANY, wx.EXPAND | wx.ALL, 20)
        leftBox.Add(st2, proportion=0, flag=wx.ALL)
        leftBox.Add(self.logText, wx.ID_ANY, wx.EXPAND | wx.ALL, 20)

        middleBox = wx.BoxSizer(wx.VERTICAL)
        st3 = wx.StaticText(backgroundPanel, label="Plot:")
        self.plotter = PlotNotebook(backgroundPanel)
        num_sensors = self.cfg.ReadInt("numSensors", 1)
        for device_name in variables.keys():
            variable_names = variables[device_name]
            scaling = devices[device_name][1]
            for name in variable_names:
                self.plotter.add(name, device_name, scaling, num_sensors)

        middleBox.Add(st3, proportion=0, flag=wx.ALL)
        middleBox.Add(self.plotter, proportion=7, flag=wx.EXPAND | wx.ALL, border=20)

        rightBox = wx.BoxSizer(wx.VERTICAL)
        self.btn_connect = wx.ToggleButton(backgroundPanel, label="Connect")
        self.btn_start = wx.ToggleButton(backgroundPanel, label="Start")
        self.btn_start.Disable()
        self.btn_test = wx.ToggleButton(backgroundPanel, label="Test Mode")
        self.btn_measure = wx.Button(backgroundPanel, label="Measure")
        self.btn_measure.Disable()
        btn_erase = wx.Button(backgroundPanel, label="Erase")
        self.btn_connect.Bind(wx.EVT_TOGGLEBUTTON, self.OnConnect)
        self.btn_start.Bind(wx.EVT_TOGGLEBUTTON, self.OnStart)
        self.btn_measure.Bind(wx.EVT_BUTTON, self.OnUpdate)
        btn_erase.Bind(wx.EVT_BUTTON, self.OnErase)
        rightBox.Add(self.btn_connect, proportion=1, flag=wx.EXPAND | wx.ALL, border=20)
        rightBox.Add(self.btn_start, proportion=1, flag=wx.EXPAND | wx.ALL, border=20)
        rightBox.Add(self.btn_test, proportion=1, flag=wx.EXPAND | wx.ALL, border=20)
        rightBox.Add(self.btn_measure, proportion=1, flag=wx.EXPAND | wx.ALL, border=20)
        rightBox.Add(btn_erase, proportion=1, flag=wx.EXPAND | wx.ALL, border=20)

        outerBox.Add(leftBox, proportion=2, flag=wx.EXPAND | wx.ALL, border=20)
```

```
        outerBox.Add(middleBox, proportion=3, flag=wx.EXPAND | wx.ALL, border=20)
        outerBox.Add(rightBox, proportion=1, flag=wx.EXPAND | wx.ALL, border=20)
        backgroundPanel.SetSizer(outerBox)

        self.Maximize()
        self.SetTitle("High-Throughput Plant Phenotyping Platform")
        self.Centre()

    def OnClose(self, e):
        """ Response to close event

        Make sure to disconnect from all sensors and camera before exiting
        """
        self.sensor_handler.closeAll()
        self.camera_frame.close()
        self.DestroyLater()

    def OnNew(self, e):
        """ Toolbar option to reset log without saving """
        confirmDiag = wx.MessageDialog(
            None,
            ("Are you sure you want to clear " + "the log?"),
            "Question",
            (wx.YES_NO | wx.NO_DEFAULT | wx.ICON_QUESTION),
        )
        dialogFlag = confirmDiag.ShowModal()
        if dialogFlag == wx.ID_YES:
            self.logText.SetValue("")
            self.logSettings()
            num_sensors = self.cfg.ReadInt("numSensors", 1)
            self.plotter.reset(num_sensors)
            self.mapPanel.clear()
            self.mapPanel.refresh([])
            self.reset()

    def OnSave(self, e):
        """ Toolbar option to save and reset log """
        rootName = "data/HTPPLogFile" + datetime.now().strftime("%Y-%m-%d") + "X"
        i = 1
        while os.path.isfile(rootName + str(i) + ".txt"):
            i += 1
        finalFilename = rootName + str(i) + ".txt"
        self.logText.SaveFile(finalFilename)
        self.logText.SetValue("")
        self.logSettings()
        num_sensors = self.cfg.ReadInt("numSensors", 1)
        self.plotter.reset(num_sensors)
        self.mapPanel.clear()
        self.mapPanel.refresh([])
        self.reset()

    def OnQuit(self, e):
        """ Toolbar option to exit application """
        self.Close()

    def OnCamera(self, e):
        """ Show or hide camera frame depeding on checkable menu item """
        self.camera_frame.Show(self.camerami.IsChecked())
```

```python

        def OnCameraClose(self, e):
            """ Safely close camera frame """
            self.camerami.Check(False)
            self.updateCameraFrame()

        def OnAbout(self, e):
            """ Toolbar option to show About dialog """
            wx.MessageBox(
                (
                    "High-Throughput Plant Phenotyping Platform \n"
                    "Made by Roberto Buelvas\n"
                    "McGill University, 2020\n"
                    "Version 0.1\n"
                ),
                "About",
                wx.OK | wx.ICON_INFORMATION,
            )

        def OnPorts(self, e):
            """ Toolbar option to open ports dialog window """
            pDialog = PortsDialog(self, self.cfg)
            dialogFlag = pDialog.ShowModal()
            if dialogFlag == wx.ID_OK:
                results = pDialog.getSettings()
                num_sensors = results.ReadInt("numSensors", 1)
                self.cfg.WriteInt("numSensors", num_sensors)
                self.labels = self.updateLabels()
                for label in self.labels:
                    self.cfg.WriteBool(
                        "connected" + label, results.ReadBool("connected" + label)
                    )
                    self.cfg.Write("port" + label, results.Read("port" + label))
                self.updateSensorHandler()
                self.updateCameraFrame()
                self.logSettings()
                self.plotter.reset(num_sensors)
            pDialog.Destroy()

        def OnLayout(self, e):
            """ Toolbar option to open ports dialog window """
            lDialog = LayoutDialog(self, self.cfg)
            dialogFlag = lDialog.ShowModal()
            if dialogFlag == wx.ID_OK:
                results = lDialog.getSettings()
                settings_list = lDialog.getSettingsList()
                for setting_key in settings_list:
                    if setting_key[0] == "D":
                        self.cfg.WriteFloat(setting_key, results.ReadFloat(setting_key))
                    if setting_key[0] == "I":
                        self.cfg.WriteInt(setting_key, results.ReadInt(setting_key))
                self.logSettings()
            condition = False
            if condition:
                wx.MessageBox(
                    "The dimensions in Layout have not been properly set",
                    "Empty port",
                    wx.OK | wx.ICON_WARNING,
                )
```

```python
            lDialog.Destroy()

    def OnClear(self, e):
        """ Toolbar option to clear all settings

        This removes every entry from the attribute cfg, except for a dummy
        entry 'notEmpty' to prevent the ConfigBase of being entirerly empty,
        which causes it to crash.
        Not to be confounded with the method reset(), which resets
        some other attributes whose values are only relevant for the current
        survey
        """
        confirmDiag = wx.MessageDialog(
            None,
            ("Are you sure you want to clear " + "the settings?"),
            "Question",
            (wx.YES_NO | wx.NO_DEFAULT | wx.ICON_QUESTION),
        )
        dialogFlag = confirmDiag.ShowModal()
        if dialogFlag == wx.ID_YES:
            all_config_keys = []
            more, value, index = self.cfg.GetFirstEntry()
            while more:
                all_config_keys.append(value)
                more, value, index = self.cfg.GetNextEntry(index)
            all_config_keys.remove("notEmpty")
            for key in all_config_keys:
                self.cfg.DeleteEntry(key)

    def OnConnect(self, e):
        """ Toggle button action to connect/disconnect from sensors

        When in 'Test Mode', it simulates a first GPS reading to compute the projection
        constants
        Start and Measure buttons are only enabled if connected
        Test button is only enabled if disconnected
        """
        btn = e.GetEventObject()
        is_pressed = btn.GetValue()
        is_test_mode = self.btn_test.GetValue()
        if is_test_mode:
            if is_pressed:
                for label in self.labels:
                    if (label[0] == "g") and self.cfg.ReadBool(
                        "connected" + label, False
                    ):
                        reading = [-73.939830, 45.423804, 45, 1, 10]
                        self.sensor_handler.GPS_constants = setupGPSProjection(reading)
                btn.SetLabelText("Disconnect")
            else:
                btn.SetLabelText("Connect")
        else:
            if is_pressed:
                success = self.sensor_handler.openAll()
                if success:
                    btn.SetLabelText("Disconnect")
                    self.btn_start.Enable(is_pressed)
                    self.btn_measure.Enable(is_pressed)
                    self.btn_test.Enable(not is_pressed)
```

```python
                else:
                    wx.MessageBox(
                        "At least one port has not been properly set up",
                        "Empty port",
                        wx.OK | wx.ICON_WARNING,
                    )
                    btn.SetValue(False)
            else:
                self.sensor_handler.closeAll()
                btn.SetLabelText("Connect")
        self.btn_start.Enable(is_pressed)
        self.btn_measure.Enable(is_pressed)
        self.btn_test.Enable(not is_pressed)

    def OnStart(self, e):
        """ Button action to take measurements periodically

        When clicked for the first time, it will start the timer. When clicked again,
        it will stop it.
        """
        btn = e.GetEventObject()
        is_pressed = btn.GetValue()
        if is_pressed:
            self.timer.Start(300.0)
            btn.SetLabelText("Stop")
        else:
            self.timer.Stop()
            btn.SetLabelText("Start")
        self.btn_connect.Enable(not is_pressed)
        self.btn_measure.Enable(not is_pressed)
        self.btn_test.Enable(not is_pressed)

    def OnErase(self, e):
        """ Button action to delete last measurement from log text """
        if self.logText.GetValue() != "":
            lastPosition = self.logText.GetLastPosition()
            self.logText.Remove(self.last_record[-1], lastPosition)
            if len(self.last_record) > 1:
                del self.last_record[-1]

    def OnUpdate(self, e):
        """ Updates UI by getting new sensor data

        This is a general method that calls the more specific ones if necessary
        """
        is_test_mode = self.btn_test.GetValue()
        self.last_record.append(self.logText.GetLastPosition())
        self.logText.AppendText("*****" + str(self.num_readings) + "*****\n")
        if is_test_mode:
            for label in self.labels:
                if self.cfg.ReadBool("connected" + label, False):
                    reading = self.sensor_handler.simulate(
                        label, self.num_readings, self.cfg
                    )
                    if label[0] == "g":
                        self.updateMap(reading, label)
                    self.updateLog(reading, label)
                    self.updatePlot(reading, label)
        else:
            for label in self.labels:
```

```
                reading = self.sensor_handler.read(label, self.num_readings, self.cfg)
                if reading is not None:
                    if label[0] == "g":
                        self.updateMap(reading, label)
                    self.updateLog(reading, label)
                    self.updatePlot(reading, label)
        self.num_readings += 1

    def updateLog(self, some_value, label):
        """ Update log text after receiving new sensor data """
        if some_value is not None:
            ts = datetime.now().strftime("%H:%M:%S.%f")
            value_text = []
            if label[0] == "g":
                for value in some_value:
                    value_text.append(str(value))
            else:
                for value in some_value:
                    value_text.append(str(np.round(value, 4)))
            self.logText.AppendText(
                (label + ";" + ts + ";" + ",".join(value_text) + "\n")
            )

    def updatePlot(self, some_value, label):
        """ Updates plots after receiving new sensor data

        Instead of appending new points to a pre-existing plot, everytime a new
        point arrives, a new plot is created next to it. Because it is made to
        match in color and style, it looks as if everything was connected.
        However, because of this way of updating the plots, the legends need to
        be created manually.
        Besides creating new plots, this method updates the attribute
        previous_measurements
        """
        sensor_type = label[0]
        measured_properties = variables[sensor_type]
        for i, measured_property in enumerate(measured_properties):
            self.plotter.update(some_value[i], label, measured_property)

    def updateMap(self, some_value, label):
        """ Update map after receiving new sensor data

        Place a marker in the locations of the map where the vehicle and
        the sensors are (except the GPS receiver itself).
        It uses the Settings from the Layout dialog to calculate relative
        positions and convert from the world coordinate system to that of the
        vehicle (or vice-versa)
        The first set of readings doesn't produce changes in the plot because
        at least two measurements are required to compute the heading, which
        in turn is required to know how to orient the sensor markers
        """
        if (self.num_readings > 0) and (not any(np.isnan(some_value[[2, 7, 8]]))):
            vehicle_x = some_value[7]
            vehicle_y = some_value[8]
            heading_radians = math.pi * some_value[2] / 180
            line_list = []
            line_list.append(self.mapPanel.ax.plot(vehicle_x, vehicle_y, "bs")[0])
            for label in self.labels:
```

```
                if (label[0] != "g") and self.cfg.ReadBool("connected" + label, False):
                    if label[0] == "u":
                        color = "y"
                        db = self.cfg.ReadFloat("DB1") / 100
                    else:
                        color = "r"
                        db = (self.cfg.ReadFloat("DB1") + self.cfg.ReadFloat("DB2")) / 100
                    if label[0] == "e":
                        color = "g"
                        index = self.cfg.ReadInt("IE" + label[1])
                        accumulator = 0
                        for i in range(index):
                            accumulator += self.cfg.ReadFloat("D" + label[1] + str(i + 1))
                        if label[1] == "L":
                            ds = -1 * accumulator / 100
                        else:
                            ds = accumulator / 100
                        ds += self.cfg.ReadFloat("DE" + label[1]) / 100
                    else:
                        accumulator = 0
                        for i in range(int(label[2])):
                            accumulator += self.cfg.ReadFloat("D" + label[1] + str(i + 1))
                        if label[1] == "L":
                            ds = -1 * accumulator / 100
                        else:
                            ds = accumulator / 100
                    sensor_x = (
                        vehicle_x
                        + ds * math.sin(heading_radians)
                        - db * math.cos(heading_radians)
                    )
                    sensor_y = (
                        vehicle_y
                        - ds * math.cos(heading_radians)
                        - db * math.sin(heading_radians)
                    )
                    # line_list.append(
                    #     self.mapPanel.ax.plot(
                    #         sensor_x,
                    #         sensor_y,
                    #         marker="P",
                    #         color=color,
                    #         markerfacecolor=color,
                    #     )[0]
                    # )
            self.mapPanel.refresh(line_list)
            line_list = []

    def logSettings(self):
        """ Append settings to log """
        self.logText.AppendText("**************Settings-Start**************\n")
        more, value, index = self.cfg.GetFirstEntry()
        while more:
            initial = value[0]
            if value != "notEmpty":
                if (initial == "I") or (initial == "n"):
```

```
                    property = str(self.cfg.ReadInt(value))
                if initial == "D":
                    property = str(self.cfg.ReadFloat(value))
                if initial == "c":
                    property = str(self.cfg.ReadBool(value))
                if initial == "p":
                    property = self.cfg.Read(value)
                self.logText.AppendText("{" + value + ": " + property + "}\n")
            more, value, index = self.cfg.GetNextEntry(index)
        self.logText.AppendText("*************Settings-End***************\n")

    def updateLabels(self):
        """ Produces list of sensor labels

        Since many methods need to iterate over all the labels, it is handy to
        have the attribute labels to keep them available. This method is used
        to update the value of labels whenever the number of sensors changes
        """
        num_sensors = self.cfg.ReadInt("numSensors", 1)
        labels = []
        device_tuples = list(devices.values())
        for device_tuple in device_tuples:
            name = device_tuple[0]
            scaling = device_tuple[1]
            initial = name[0].lower()
            if scaling:
                for i in range(num_sensors):
                    labels.append(initial + "L" + str(i + 1))
                for i in range(num_sensors):
                    labels.append(initial + "R" + str(i + 1))
            else:
                labels.append(initial + "L")
                labels.append(initial + "R")
        return labels

    def updateSensorHandler(self):
        """ Create sensor handler and populate it by adding serial sensors """
        if self.sensor_handler is not None:
            self.sensor_handler.closeAll()
        self.sensor_handler = SensorHandler()
        for label in self.labels:
            is_connected = self.cfg.ReadBool("connected" + label, False)
            port = self.cfg.Read("port" + label, "")
            if is_connected and (port != ""):
                self.sensor_handler.add(port, label)

    def updateCameraFrame(self):
        """ Format camera ports as a list ready to be used as CameraFrame's input """
        if self.camera_frame is not None:
            self.camera_frame.close()
        camera_ports = [None, None]
        if self.cfg.ReadBool("connectedcL", False):
            camera_ports[0] = int(self.cfg.Read("portcL", ""))
        if self.cfg.ReadBool("connectedcR", False):
            camera_ports[1] = int(self.cfg.Read("portcR", ""))
        self.camera_frame = CameraFrame(self, camera_ports)
        self.camera_frame.Show(self.camerami.IsChecked())

    def reset(self):
```

```
        """ Reset the values of attributes any time a new survey starts """
        self.labels = self.updateLabels()
        self.num_readings = 0
        self.last_record = [0]
```

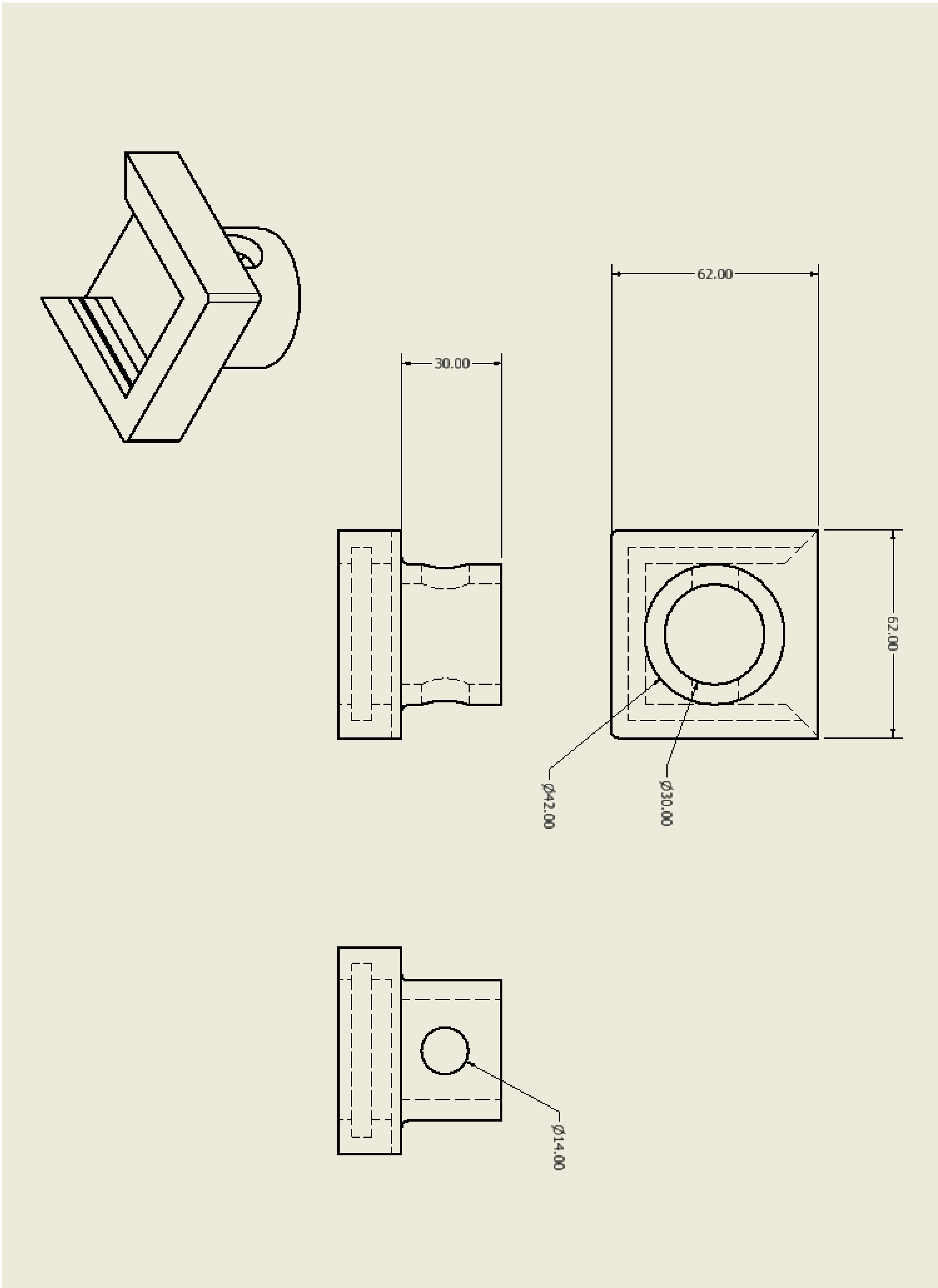
Appendix 2. Code snippet of statistical analysis of residuals

```
%%
close all hidden;
clearvars;
clc;
%% Get the data
total_traits = load('../data/totalTraits.mat').total_traits;
yieldModel = load('../data/yieldModels.mat').YieldMEModel;
phenoModels = load('../data/RegressionModels.mat');
pheno_properties = phenoModels.pheno_properties;
model_results = phenoModels.results;
dates = {'08_17', '08_20', '08_24', '08_27', '08_31', '09_03'};
DAS = [77, 80, 84, 87, 91, 94];

%% Do Tukey's with residuals for yield
% Yield ~ Variety
variety = total_traits.Variety;
row_spacing = total_traits.RowSpacing;
density = total_traits.Density_plants_acre_;
yield = total_traits.Yield_ton_ha_;
[p,tbl,stats] = anova1(yield,variety, 'off');
fig = figure();
[tk, m, ~, gnames] = multcompare(stats);
%%
ax=gca;
means = zeros(11,1);
confidence_intervals = zeros(11,1);
for j=1:11
    means(j) = (ax.Children(2*j-1).XData(2)+ax.Children(2*j-1).XData(1))/2;
    confidence_intervals(j) = (ax.Children(2*j-1).XData(2)-ax.Children(2*j-
1).XData(1))/2;
end
close(fig);
tukey_table = table(means);
tukey_table.ci = confidence_intervals;
tukey_table.variety = gnames(11:-1:1);
tukey_table = sortrows(tukey_table, 'means', 'descend');
fig = makeTukeyPlot(tukey_table, 'Yield [ton/ha]');
saveas(fig, '../data/final_plots/FinalTukeyYield.png');
close(fig);
%%

% Yield - Environment - Management ~ Variety
estimated_yield = yieldModel.predictFcn(total_traits);
residual_yield = yield - estimated_yield;
[p,tbl,stats] = anova1(residual_yield,variety);
fig = figure();
[tk, m, ~, gnames] = multcompare(stats);
ax=gca;
means = zeros(11,1);
confidence_intervals = zeros(11,1);
```

```
for j=1:11
    means(j) = (ax.Children(2*j-1).XData(2)+ax.Children(2*j-1).XData(1))/2;
    confidence_intervals(j) = (ax.Children(2*j-1).XData(2)-ax.Children(2*j-
1).XData(1))/2;
end
close(fig);
tukey_table = table(means);
tukey_table.ci = confidence_intervals;
tukey_table.variety = gnames(11:-1:1);
tukey_table = sortrows(tukey_table, 'means', 'descend');
fig = makeTukeyPlot(tukey_table, 'Effect of variety on Yield');
saveas(fig, '../data/final_plots/FinalTukeyYieldFull.png');
close(fig);
es_environment = 0.14;
es_management = 0.05;
es_genome = (tbl{2,2}/tbl{4,2})*(1-es_environment-es_management);
fig = makeEffectSizePie(es_genome, es_management, es_environment, 'Effect
sizes on Yield');
saveas(fig, ['../data/final_plots/FinalPieYield.png']);
close(fig);

% Yield ~ Variety + Management
% varnames = {'Variety';'RowSpacing';'Density'};
% [tbl,chi,p,factorvals] = crosstab(variety, row_spacing, density);
% [p2,tbl2,stats2,terms2] = anovan(yield,{variety row_spacing,
density},3,3,varnames);
% fig = figure();
% tk2 = multcompare(stats2);

% Yield - Environment ~ Variety + Management
% [p3,tbl3,stats3,terms3] = anovan(residual_yield,{variety row_spacing,
density},3,3,varnames);
% fig = figure();
% tk3 = multcompare(stats3, 'Dimension', 1);
% fig = figure();
% tk3 = multcompare(stats3, 'Dimension', 2);
% fig = figure();
% tk3 = multcompare(stats3, 'Dimension', 3);

%% Phenotyping data
% Phenotype - Env - Management* ~ Variety
for property_index=1:length(pheno_properties)
    ranova_tbl = table();
    for date_index=1:6
        model = model_results{property_index}.Models{2, date_index};
        predictors = model_results{property_index}.Predictors{2, date_index};
        estimated = predict(model, total_traits(:, predictors));
        residual =
total_traits.([pheno_properties{property_index},dates{date_index}]) -
estimated;
        ranova_tbl.(['m',num2str(date_index)]) = residual;
        %[p,tbl,stats,terms] = anovan(residual,{variety row_spacing
density},1,3,varnames);
```

```
        [p,tbl,stats] = anova1(residual,variety, 'off');
        fig = figure();
        [tk, m, ~, gnames] = multcompare(stats);
        ax=gca;
        means = zeros(11,1);
        confidence_intervals = zeros(11,1);
        for j=1:11
            means(j) = (ax.Children(2*j-1).XData(2)+ax.Children(2*j-
1).XData(1))/2;
            confidence_intervals(j) = (ax.Children(2*j-1).XData(2)-
ax.Children(2*j-1).XData(1))/2;
        end
        close(fig);
        tukey_table = table(means);
        tukey_table.ci = confidence_intervals;
        tukey_table.variety = gnames(11:-1:1);
        tukey_table = sortrows(tukey_table, 'means', 'descend');
        fig = makeTukeyPlot(tukey_table, ['Effect of variety on
',pheno_properties{property_index},dates{date_index}]);
        saveas(fig,
['../data/final_plots/FinalTukey',pheno_properties{property_index},dates{date
_index},'.png']);
        close(fig);
        es_environment = model_results{property_index}.Rsquared(118,
date_index);
        [~, best_management_index] =
max(sum(model_results{property_index}.Rsquared(118:121, :), 2));
        es_management = model_results{property_index}.Rsquared(117 +
best_management_index, date_index) - es_environment;
        if es_management < 0
            es_management = 0;
        end
        es_genome = (tbl{2,2}/tbl{4,2})*(1-es_environment-es_management);
        fig = makeEffectSizePie(es_genome, es_management, es_environment,
['Effect sizes on ',pheno_properties{property_index},dates{date_index}]);
        saveas(fig,
['../data/final_plots/FinalPie',pheno_properties{property_index},dates{date_i
ndex},'.png']);
        close(fig);
    end
    ranova_tbl.Variety = variety;
    rm = fitrm(ranova_tbl, 'm1-m6 ~ Variety','WithinDesign',DAS);
    ranova_result=ranova(rm);
    disp([pheno_properties{property_index},dates{date_index}]);
    disp(ranova_result);
end
```

Appendix 3. Drawing of tool connector part



*: Units in mm

Appendix 4. Results of repeated measurements ANOVA

NDRE

	SumSq	DF	MeanSq	F	pValue	pValueGG
(Intercept):Time	2.334	3	0.77799	644.53	1.6289e-171	4.3044e-153
Variety:Time	0.26324	33	0.0079769	6.6086	2.5336e-23	5.285e-21
Error(Time)	0.60474	501	0.0012071			

NDVI

	SumSq	DF	MeanSq	F	pValue	pValueGG
(Intercept):Time	20.365	3	6.7884	538.27	2.9612e-156	3.273e-124
Variety:Time	2.9886	33	0.090562	7.181	9.1618e-26	8.0338e-21
Error(Time)	6.3183	501	0.012611			

CI

	SumSq	DF	MeanSq	F	pValue	pValueGG
(Intercept):Time	21.182	3	7.0605	644.46	1.6648e-171	7.7333e-143
Variety:Time	2.7538	33	0.083449	7.617	1.332e-27	2.8519e-23
Error(Time)	5.4888	501	0.010956			

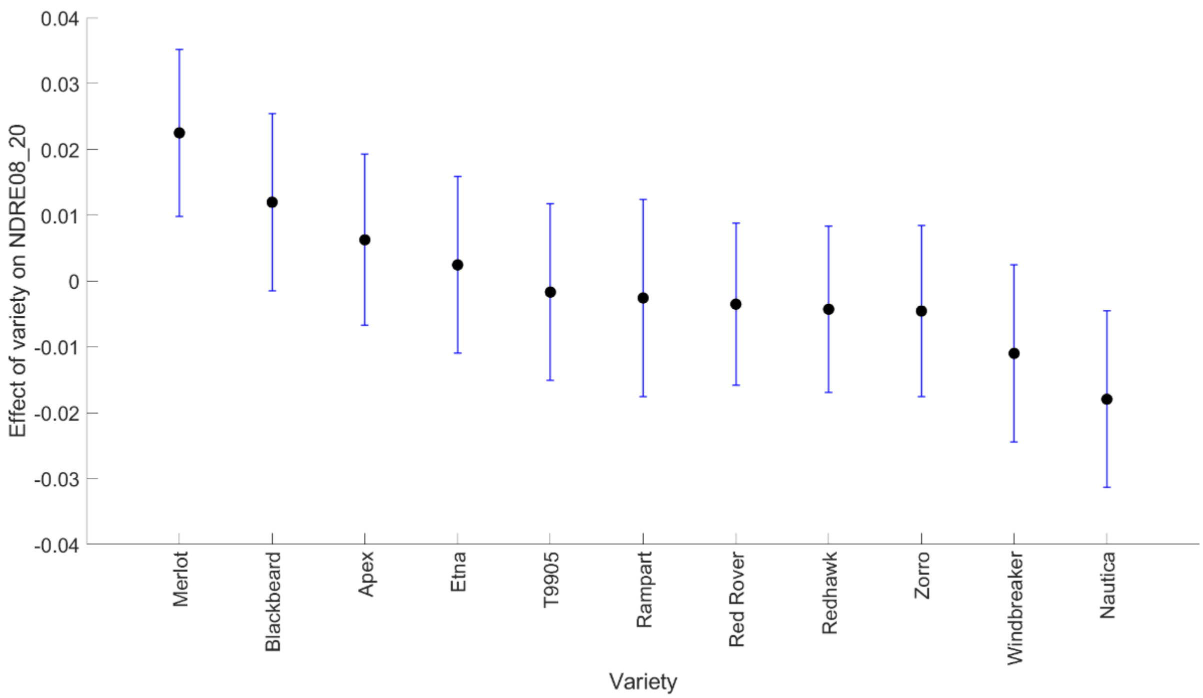
Height

	SumSq	DF	MeanSq	F	pValue	pValueGG
(Intercept):Time	1.2373e+07	3	4.1243e+06	241.97	5.0162e-97	1.1653e-75
Variety:Time	1.9869e+06	33	60208	3.5324	6.8022e-10	4.5764e-08
Error(Time)	8.5394e+06	501	17045			

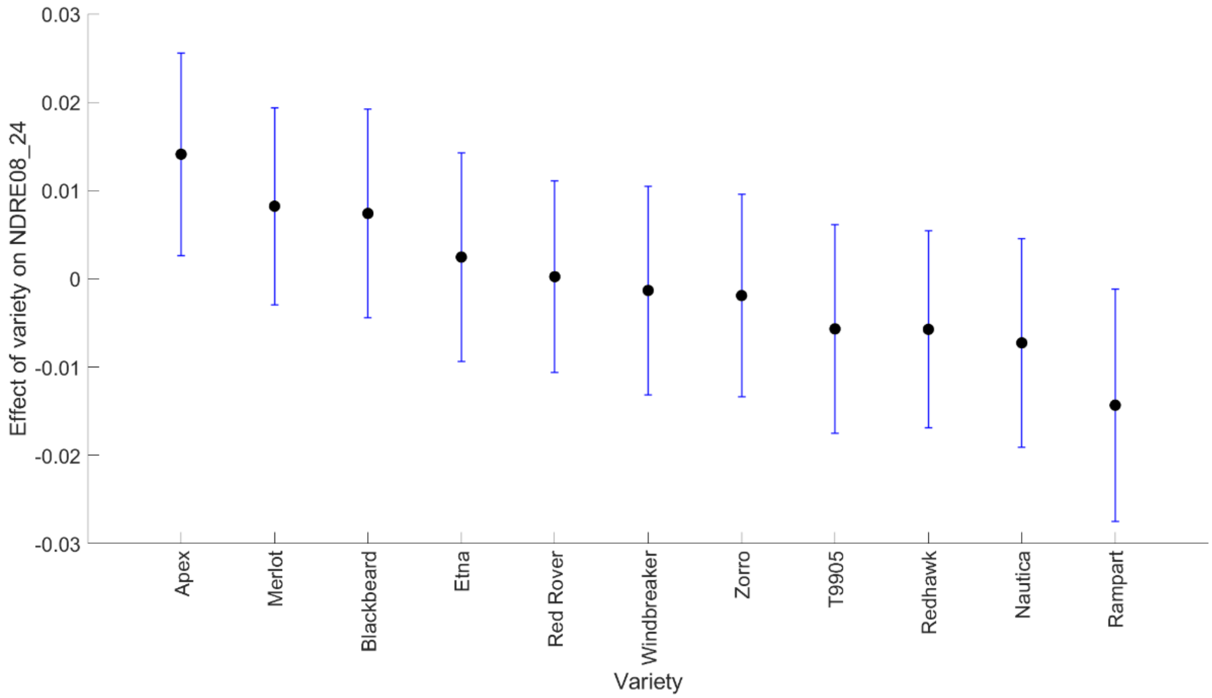
Appendix 5. Ranking of varieties according to NDRE's highest value in the model

Variety	P	Lower confidence bound	Upper confidence bound	R	Lower confidence bound	Upper confidence bound
Calmant	0.300	0.176	0.424	0.00552	0.00415	0.00689
Knight Rider	0.286	0.196	0.375	0.00505	0.00407	0.00603
Apex	0.285	0.195	0.375	0.00541	0.00436	0.00646
Red Rider	0.283	0.166	0.401	0.00522	0.00401	0.00643
Dresden	0.268	0.178	0.359	0.00500	0.00395	0.00604
Majesty	0.268	0.138	0.397	0.00488	0.00353	0.00624
Argosy	0.265	0.158	0.373	0.00494	0.00371	0.00617
Nautica	0.258	0.177	0.339	0.00498	0.00402	0.00594
Mast	0.254	0.150	0.358	0.00472	0.00350	0.00595
Sheek	0.244	0.155	0.332	0.00472	0.00368	0.00575
Compass	0.231	0.144	0.318	0.00449	0.00351	0.00547

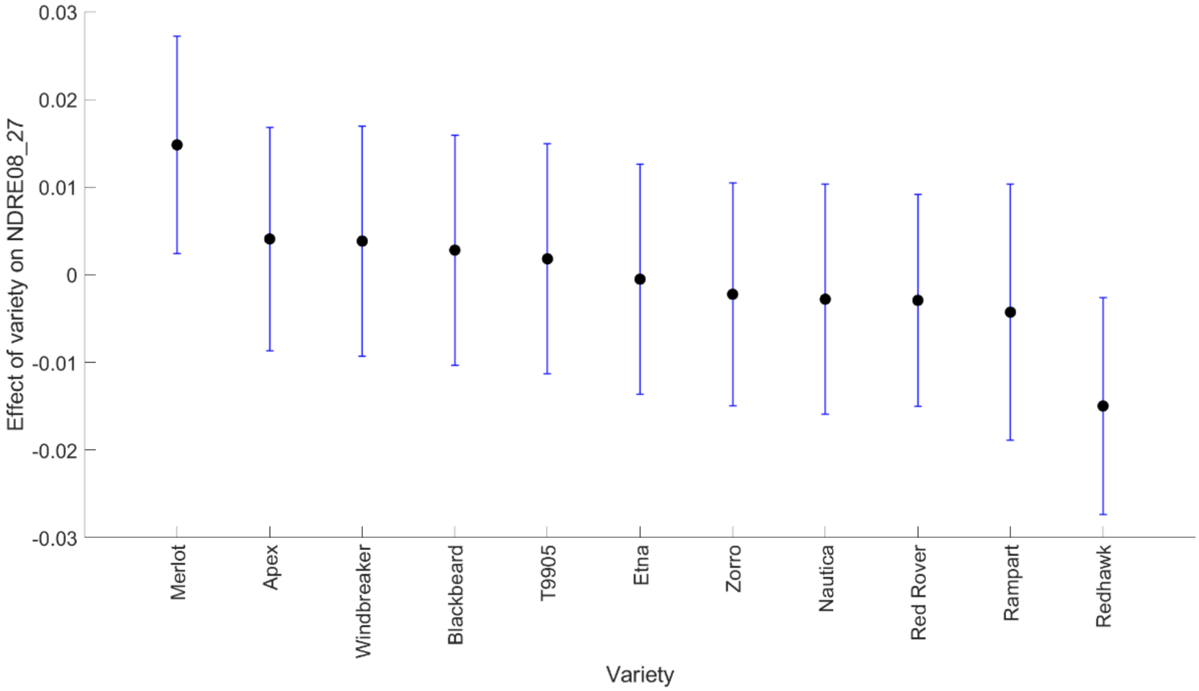
Appendix 6. Result of Tukey's test for NDRE in the second date



Appendix 7. Result of Tukey's test for NDRE in the third date



Appendix 8. Result of Tukey's test for NDRE in the fourth date



Appendix 9. Result of Tukey's test for NDRE in the fifth date