Data Article

A pulse crop dataset of agronomic traits and multispectral images from multiple environments

Kingsley Umani[a], Chongyuan Zhang[a,b], Rebecca J. McGee[c], George J. Vandemark[c], Sindhuja Sankaran[a,*]

[a] *Department of Biological System Engineering, Washington State University, Pullman, WA, United States*
[b] *Department of Botany and Plant Pathology, Purdue University, West Lafayette, Indiana, United States*
[c] *USDA-ARS, Grain Legume Genetics and Physiology Research, Washington State University, Pullman, WA, United States*

ARTICLE INFO

ABSTRACT

Crop yield potential in breeding trials can be captured using unmanned aerial vehicle (UAV) based multispectral imagery. Several digital traits or phenotypes such as vegetation indices can represent canopy crop vigor and overall plant health, which can be used to evaluate differences in performance across varieties in crop breeding programs. This dataset contains agronomic data for named cultivars and breeding lines of spring-sown dry pea and chickpea, and over 275 multispectral images from advanced and preliminary breeding trials. The breeding trials were located at three locations in the "Palouse" region of Eastern Washington and Northern Idaho of the United States across 2017, 2018 and 2019 cropping seasons. The multispectral images were captured using a UAV integrated with a 5-band multispectral camera at multiple time points from early vegetative growth through pod development stages during each cropping season. This dataset details seed yield information from trials of dry peas and chickpea that were obtained from each location, as well as additional agronomic and phenological data recorded at one location (mostly Pullman, WA) for each cropping season. The dataset also includes 20–78 megabytes (MB) Tagged Image Format (TIF) un-

---

* Corresponding author.
  E-mail address: s.sankaran@wsu.edu (S. Sankaran).

calibrated stitched orthomosaic images generated from the photogrammetric software. The images can be processed using any convenient image processing algorithm to obtain vegetation indices and other useful information.

Specification Table

Subject	Agronomy and Crop Science
Specific subject area	Images and Agronomic Dataset
Type of data	Tables and Images (TIF file)
How the data were acquired	Chickpea and pea seed yield data were collected from each location, while other agronomic and phenological traits were only collected at one location (Pullman, WA). Multispectral images were captured using a 5-band multispectral camera (12-bit image, 1.2 MP) mounted on the quadcopter.
Data format	Tag Image File Format (.tif) Microsoft Excel (.xlsx)
Description of data collection	Agronomic data were collected for seven (7) advanced (panels 01, 01B, 02, and 81) and preliminary (panels 03, 04, and 83) trials. The dataset is composed of 275 high-resolution multispectral images and associated data, captured using a multispectral camera. Data were acquired between 10:00 a.m. and 3:00 p.m. local time, and about five time points of data acquisition for each location during each season. Three locations were Fairfield, WA, Genesee, ID, and Pullman, WA and the data were acquired across three planting seasons, years 2017, 2018 and 2019. The time points for data acquisition were selected to capture key growth stages, such as early growth, flowering, and pod development stages, and based on suitable weather conditions for unmanned aerial vehicle flights (e.g., clear sky and low wind). A reflectance panel was placed in the field during image acquisition and used for radiometric correction during image processing. Images from the multispectral camera were pre-processed to generate uncalibrated orthomosaic images covering each experimental site.
Data source location	Institution: Washington State University City/Town/Region: Pullman, Pacific Northwest Country: United States Latitude and longitude coordinates for collected data: Fairfield, WA (47°19'08.0"N, 117°10'05.0"W), Genesee, ID (46°36'40.0"N, 116°57'39.0"W), Pullman, WA (46°41'39.0"N, 117°08'53.0"W).
Data accessibility	Repository name: Zenodo Data identification number: https://doi.org/10.5281/zenodo.8280431.
Related research articles	C. Zhang, R. J. McGee, G. J. Vandemark, S. Sankaran, Crop performance evaluation of chickpea and dry pea breeding lines across seasons and locations using Phenomics data. *Frontiers in Plant Science, 12,* (2021). https://doi.org/10.3389/fpls.2021.640259 A. Marzougui, R.J McGee, S. Van Vleet and S. Sankaran, Remote sensing for field pea yield estimation: A study of multi-scale data fusion approaches in phenomics. *Frontiers in Plant Science, 14,* (2023). https://doi.org/10.3389/fpls.2023.1111575

1. Value of the Data

- This dataset can be utilized to study crop performance across multiple locations (genotype by environment interactions) within the Palouse region of the Pacific Northwest United States. This dataset can be used to study the relationships between spectral information (digital traits) and crop performance (seed yield and other traits) [1].

- This dataset can be integrated with other datasets (genetic data, environment data) to establish data mining approaches for pulse crops in the breeding programs.
- The image dataset can be compared with complementary datasets to study the effects of image resolution on predicting crop performances [2].
- This dataset can assist in developing other similar datasets for crop management such as monitoring nutrient applications, pests, weeds, and diseases in pulse crop production [3].
- This dataset can be used to develop and evaluate yield prediction using machine learning approaches [1,2,4]. In addition, dataset can be used to develop and evaluate machine and deep learning algorithms that can be used for various applications (e.g. automated plot segmentation, cloud shadow removal, effect of incident light normalization).

2. Objective

The objective of this dataset is to compile agronomic data and unmanned aerial vehicle (UAV) based multispectral images of dry pea and chickpea breeding evaluation plots across different locations and seasons. This dataset can be used to study the relationship between the digital traits and crop performance data, allow the use of multispectral imagery to assess crop performance at different growth stages, and study genotype-by-environment interaction and yield stability.

3. Data Description

Dry pea and chickpea are pulse crops that are typically grown in rotations with small cereal grains, primarily wheat and barley. Their use as rotational crops provides many benefits to cereal grain production, including a contribution of residual nitrogen produced by their symbiotic association with nitrogen-fixing rhizobacteria, disruption of cyclical diseases and pests affecting cereal grains, and control of grassy weeds. In the United States, pulses are mostly grown in the Palouse region of Washington, Idaho, and Oregon, and in the Northern Great Plains of Montana, North Dakota, and South Dakota [5]. Other diverse datasets on pea and chickpea are available in Data in Brief as well [6–11].

In this study, we present a collection of agronomic (ground truth) data and multispectral images captured on spring-sown pea (*Pisum sativum* L.) and chickpea (*Cicer arietinum* L.) breeding yield trials using a 5-band multispectral camera (RedEdge, MicaSense Inc., Seattle, WA, United States) mounted on a quadcopter unmanned aerial vehicle (AgBot, ATI Inc., Oregon City, OR, United States).

Agronomic data and multispectral images were acquired at Fairfield, WA (47°19'08.0"N, 117°10'05.0"W); Genesee, ID (46°36'40.0"N, 116°57'39.0"W); and Pullman, WA (46°41'39.0"N, 117°08'53.0"W) locations during the 2017, 2018 and 2019 field seasons. The exact locations in these regions varied across years to accommodate crop rotation cycles. The dataset consists of one file (.pdf) 'PulsePlotMaps_2017–2019' that provides the overlay of the uncalibrated orthomosaic UAV imagery with plot maps for each year and location, three .xlsx files (2017PulseDataset, 2018PulseDataset, and 2019PulseDataset) summarizing the agronomic data, and three folders (Images2017, Images2018, and Images2019) containing orthomosaic images of different trials. The entire dataset is less than 12 GB.

3.1. Ground truth data

The ground truth data (three .xlsx files) represent agronomic datasets for the 2017, 2018, and 2019 field seasons. The summary description for advanced and preliminary yield trials of green pea (panels 01 and 03, respectively), yellow pea (panels 02 and 04, respectively), and chickpea (panels 81 and 83, respectively) from Fairfield, Genesee and Pullman in each worksheet within

Table 1

Summary description of data for advanced and preliminary trials of pulse crops for 2017, 2018 and 2019 field seasons at Fairfield, Genesee, and Pullman locations.

Trial Location	Trial No.*	Crop Type	No. of Entries	No. of Replicates	No. of Plots
Fairfield	1701	Advanced Green Pea	40	3	120
	1701B	Advanced Green Yellow New Zealand Pea	19	3	57
	1702	Advanced Yellow Pea	21	3	63
Pullman	1701	Advanced Green Pea	40	3	120
	1701B	Advanced Green Yellow New Zealand Pea	19	3	57
	1702	Advanced Yellow Pea	21	3	63
	1703	Preliminary Green Pea	12	3	36
	1704	Preliminary Yellow Pea	6	3	18
	1781	Advanced Kabuli Chickpea	24	3	72
	1781G	Advanced Green Chickpea	8	3	24
Fairfield	1801	Advanced Green Pea	32	3	96
	1802	Advanced Yellow Pea	23	3	69
	1881	Advanced Kabuli Chickpea	21	3	63
Genesee	1801	Advanced Green Pea	32	3	96
	1881	Advanced Kabuli Chickpea	21	3	63
Pullman	1801	Advanced Green Pea	32	3	96
	1802	Advanced Yellow Pea	23	3	69
	1803	Preliminary Green Pea	30	3	90
	1804	Preliminary Yellow Pea	20	3	60
	1881	Advanced Kabuli Chickpea	21	3	63
	1883	Preliminary Kabuli Chickpea	22	3	66
Fairfield	1901	Advanced Green Pea	29	3	87
	1902	Advanced Yellow Pea	23	3	69
	1981	Advanced Kabuli Chickpea	24	3	72
Genesee	1901	Advanced Green Pea	29	3	87
	1901B	Advanced Green Yellow New Zealand Pea	16	3	48
	1981	Advanced Kabuli Chickpea	24	3	72
Pullman	1901	Advanced Green Pea	29	3	87
	1901B	Advanced Green Yellow New Zealand Pea	16	3	48
	1902	Advanced Yellow Pea	23	3	69
	1903	Preliminary Green Pea	34	3	102
	1904	Preliminary Yellow Pea	23	3	69
	1981	Advanced Kabuli Chickpea	24	3	72
	1983	Preliminary Kabuli Chickpea	24	3	72

* The first two digits of trial number denote the year of the field season.

each file is presented in Table 1. In Table 1, the first two digits (17, 18 and 19) of the trial number denote each year of the growing season. For example, the "1701", "1801" and "1901" represents the green pea advanced trial (01) in the years 2017, 2018 and 2019, respectively. In each of the three data files, the first worksheet describes the trials in that year, with the agronomic data for each panel provided in the following sheets. The established cultivars are named in the first worksheet, while the rest of the lines or entries are coded using the same code name across locations and seasons. The yield distributions of green pea, yellow pea, and chickpea cultivars across each field location (Fairfield, Genesee, and Pullman) and cropping season (2017, 2018 and 2019) are shown in Figs. 1–3, respectively. Multiple pairwise comparisons of yield data of the common plant entries or varieties of green pea, yellow pea, and chickpea across cropping seasons at each location are presented in Figs. 4–6, respectively.

3.2. UAV multispectral data

The multispectral image data were imported into photogrammetric software, Pix4Dmapper (Pix4D Inc., San Francisco, CA, United States) to be processed into orthomosaic images. The pre-processed multispectral images in three folders (Images2017, Images2018, and Images2019) comprise 275 multispectral images (.tif) where five separate orthomosaic images are generated for
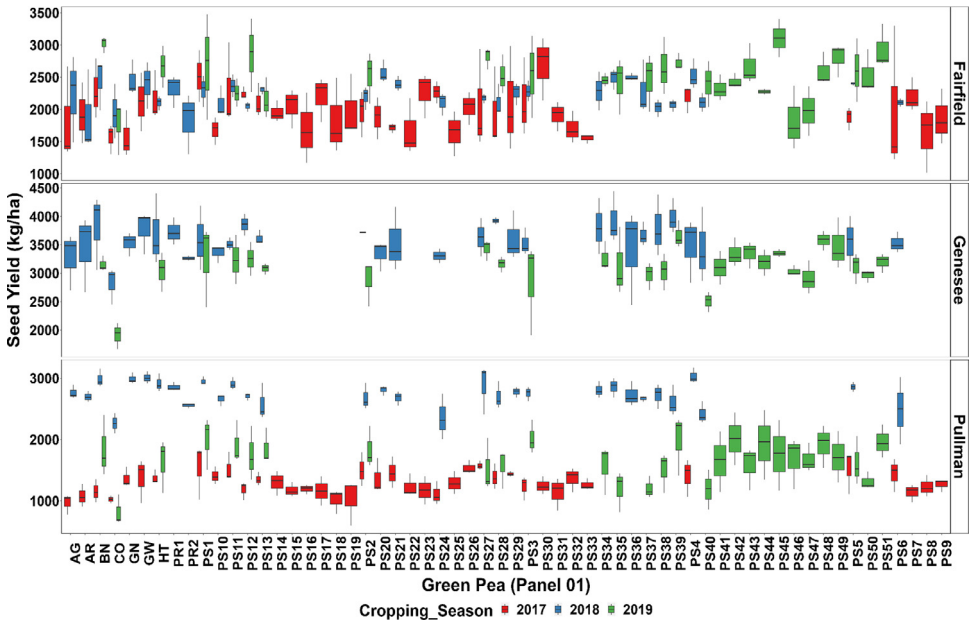
Fig. 1. Boxplot of seed yield data of green pea advanced yield trial (01) across Fairfield, Genesee, and Pullman field locations and 2017, 2018, and 2019 planting seasons.
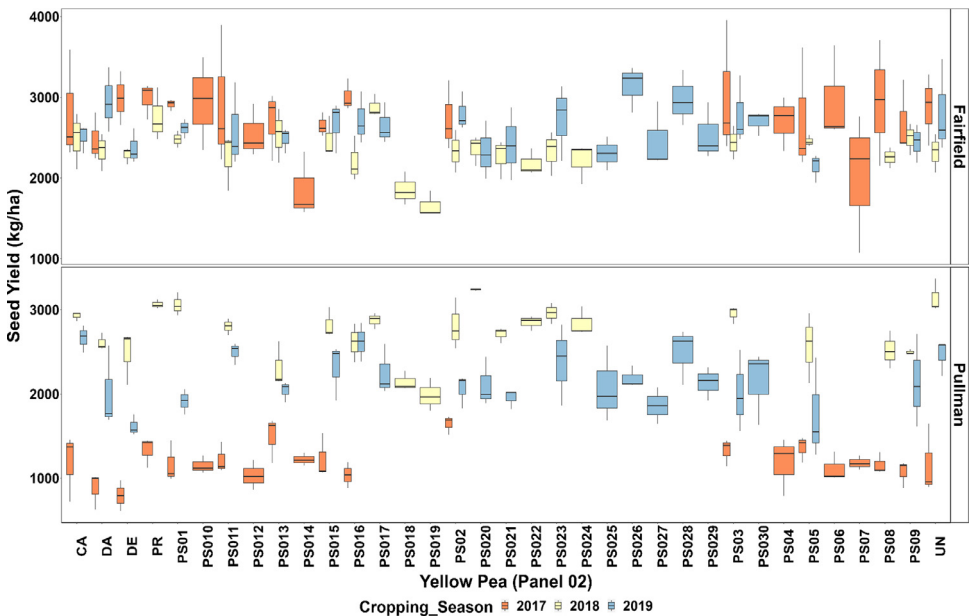

Fig. 2. Boxplot of seed yield data of yellow pea advanced yield trial (02) across Fairfield, Genesee, and Pullman field locations and 2017, 2018, and 2019 cropping seasons.
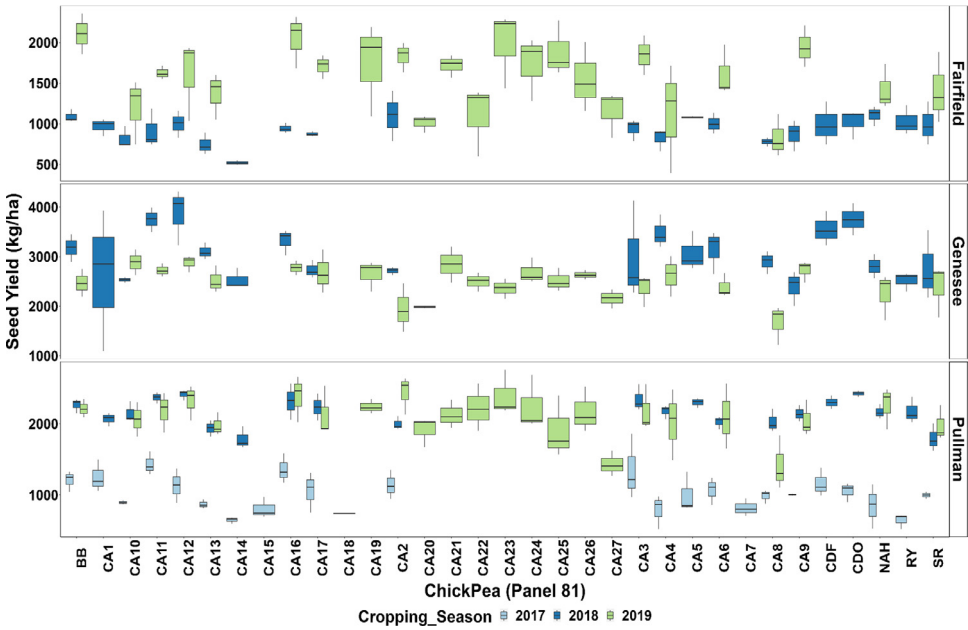
Fig. 3. Boxplot of seed yield data of chickpea advanced yield trial (81) across Fairfield, Genesee, and Pullman field locations and 2017, 2018, and 2019 planting seasons.

five spectral bands. The Micasense RedEdge multispectral camera captures reflectance data from five bands: red (663–673 nm), green (550–570 nm), blue (465–485 nm), near-infrared (820–860 nm), and red-edge (712–722 nm).

In general, in each folder, there are subfolders with two levels. In the first level, each image subfolder is labeled as 'Location_TrialName(s)'; in the second level, each subfolder is labeled as 'Location_CropType_DateofDataAcquisition'. Finally, the images are labeled as 'CropTypewithDateofDataAcquisition_ImageType_BandType'. Sometimes, the images have location names in the beginning. The 'PulsePlotMaps_2017–2019' file shows the overlay between the plot maps (plot numbers from each trial) and imagery for all trials.

4. Experimental Design, Materials and Methods

4.1. Agronomic data acquisition

The experimental design for advanced and preliminary yield trials of green pea (01), yellow pea (02), and chickpea (81) breeding lines was a randomized complete block design with three replications. The distance between neighboring plots was about 75 cm and the plots were 6.1 m in length and 1.5 m in width. After the crops emerged, the plots were reduced to a length of 4.9 m, creating corridors of 1.2 m in width [1]. Prior to planting, the pea seeds were treated with a slurry that included the fungicides fludioxonil (0.56 g kg^{-1}; Syngenta, Greensboro, NC, United States), mefenoxam (0.38 g kg^{-1}; Syngenta), and thiabendazole (1.87 g kg^{-1}; Syngenta), the insecticide thiamethoxam (0.66 ml kg^{-1}; Syngenta), and the metal molybdenum (0.35 g kg^{-1}), while each chickpea seed packet was inoculated with 0.5 g *Mesorhizobium ciceri* (1 × 10^8 CFU g^{-1}; Novozyme, Cambridge, MA, United States) a day before planting [1]. Data on seed yield from the pea and chickpea trials were gathered from each location, but agronomic and phenological traits including days to physiological maturity, days to 50 % flowering, pod height, pod
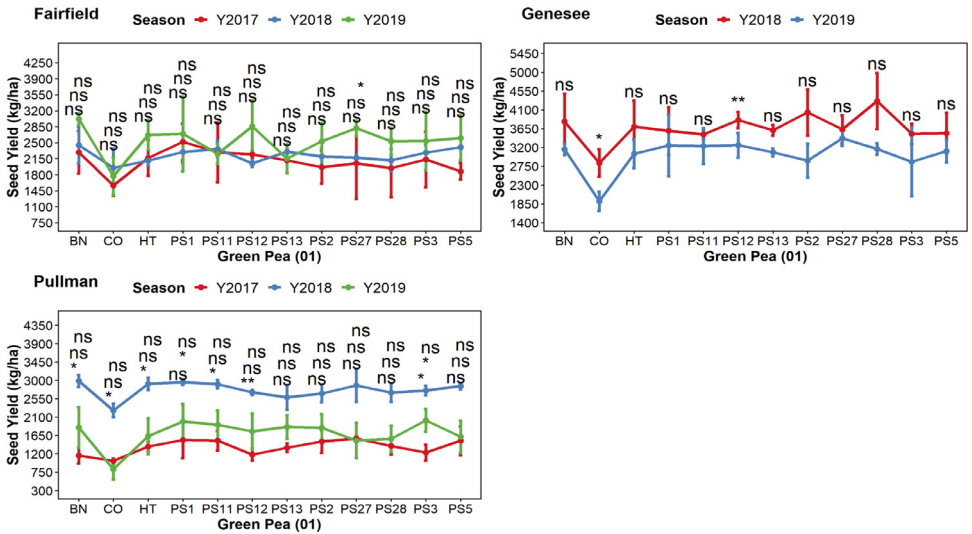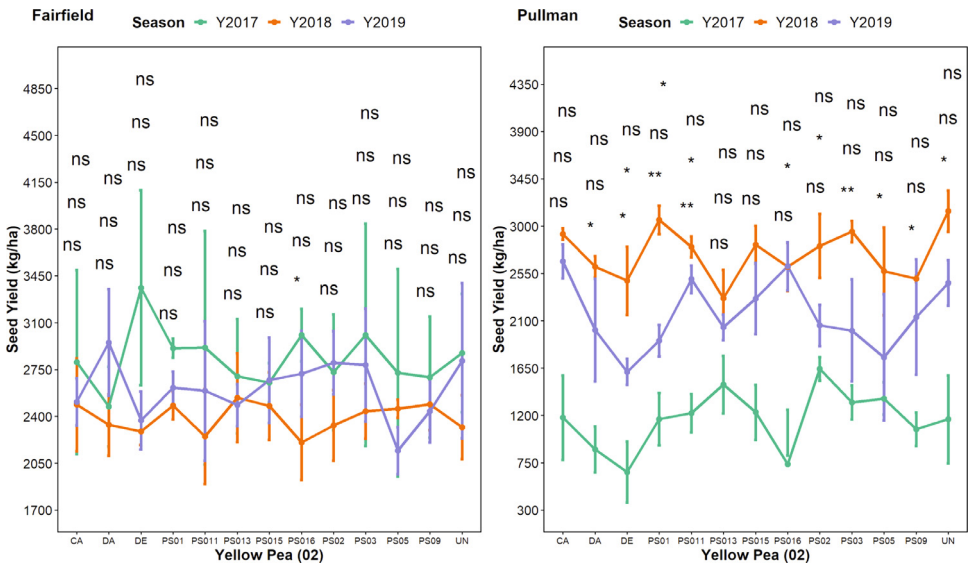
Fig. 4. Multiple pairwise comparisons of seed yield of green pea advanced yield trials (01) across 2017, 2018, and 2019 cropping seasons at Fairfield, Genesee, and Pullman locations. *ns: not significant, (*): significant at p ≤ 0.05, (**): significant at p ≤ 0.001.*

Fig. 5. Multiple pairwise comparisons of seed yield of yellow pea advanced yield trials (02) across 2017, 2018, and 2019 cropping seasons at Fairfield, and Pullman locations. *ns: not significant, (*): significant at p ≤ 0.05, (**): significant at p ≤ 0.001.*

height at maturity, length of the overall vine, canopy height at maturity, node of first flower, and number of reproductive nodes were obtained only at Pullman for each cropping season.

Description of field agronomic data characters is as follows:

- **Harvest date (*har_date*)**: Date the trial was harvested.
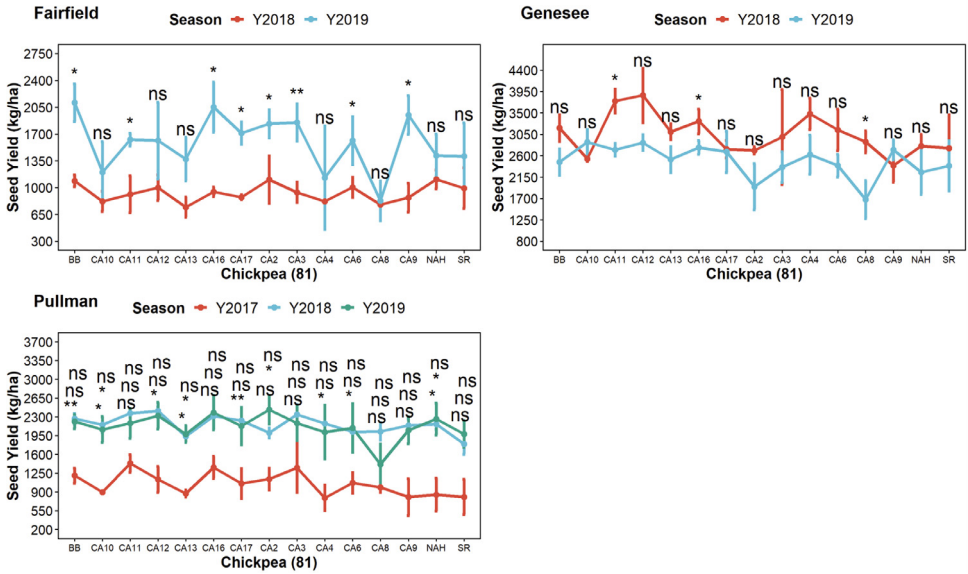- **Hundred seed weight (*hundseedwt*):** Weight of 100 seeds measured in grams.

Fig. 6. Multiple pairwise comparisons of seed yield of chickpea advanced yield trials (81) across 2017, 2018, and 2019 cropping seasons at Fairfield, Genesee, and Pullman locations. *ns: not significant, (*): significant at $p \leq 0.05$, (**): significant at $p \leq 0.001$.*

- **Plot seed weight (*plotseedwt*):** Weight of the seeds from entire plot measured in grams.
- **Seed yield (*seedyield*):** Seed yield of the plot in kg/hectare.
- **Bloom date (*datefl50*):** Date at which 50 % of the plants that had an open flower.
- **Final bloom date:** Date at which only 10 % of the plants that had an open bloom remaining on the vine.
- **Node of first flower (*flowrnode*):** Number of nodes at which the first flower/pod occurred. This includes the node where a pod may have aborted and includes the scale nodes (number 2) that are generally below the soil surface. Most often the first node that is visible above the soil surface is identified as number 3. Data were recorded for two plants randomly selected from each plot (one from each end of the plot). **NOTE:** The same two plants were used to record data for node to first flower, pods per peduncle, number of reproductive nodes, pod height (green) and vine length.
- **Pods per peduncle (*podspedun*):** The number of pods per peduncle (node). Some peduncles on a plant may have one or two pods but record the number that is most representative of the plant. Record data for two plants randomly selected from each plot.
- **Number of reproductive nodes (*reprnodes*):** The number of nodes with a developing pod. Disregard nodes where flowers have aborted and no pods will develop. Recorded data for two plants randomly selected from each plot (one from each end of the plot).
- **Pod height (green)(*podht*):** Distance from the soil surface to the lowest tip of the lowest pod as it hangs from the plant. Data taken at full pod stage on two plants randomly selected from each plot (recorded in cm).
- **Pod height at maturity (*podhtmat*):** Distance from the soil surface to the lowest pod at harvest maturity. Data taken at harvest by carefully placing a meter stick into the plant canopy. Measure in two places (recorded in cm).
- **Pod Height Index (*podhi*):** Lodging score of pod height. Calculated by dividing *podhtmat* by *podht*.
- **Overall vine length (*vinelength*):** Vine length was determined by measuring the distance from the soil surface to the apical meristem on the main stem of randomly select plants

at the full pod stage in the interior rows of the plot. It is necessary to stretch the plant to achieve maximum length. Measure in two places (recorded in cm).

- **Canopy height at maturity (*canopyht*):** (Important to know how high the pods were expected to be off the ground) Placing a meter stick upright in the plot in at least two places. Measurements were taken within a couple days before harvest and care was taken not to compress the plants and artificially reduce height (recorded in cm).
- **Plant Height Index (*planthi*):** Plant lodging index. Calculated by dividing *canopyht* by *vinelength*.
- **Maturity date (physiological) (*datephymat*):** Date at which 95% of pods tan in color and all pods flexible (leathery).

4.2. Multispectral data acquisition

Mission Planner ((https://ardupilot.org/planner/) was used to program the UAV, which flied at a speed of 2–3 m/s and at 25, 30, or 45 m above ground level (AGL). This results in a ground sampling distance of 1.7, 2.0, or 3.1 cm/pixel, and it allowed for the acquisition of images with about 80 % horizontal and 70 % vertical overlaps [1]. During the process of image acquisition and processing, a reflectance panel, which was either a MicaSense reflectance panel (RedEdge, MicaSense Inc.) in 2017 or a Spectralon reflectance panel (99 % reflectance; Spectralon, SRS-99-120, Labsphere Inc., North Sutton, NH, United States), in 2018 and 2019, was placed in the field and used for radiometric calibration.

The UAV data collection process was conducted between 10:00 am and 3:00 pm local time at five time points for each season. The selection of time points for data collection was based on the critical growth stages of the plant, including early growth, flowering, and pod and seed development stages. Additionally, the chosen time points were based on favorable weather conditions for UAV flights, including clear skies and low wind speeds. The Pix4Dmapper software (Pix4D Inc., San Francisco, CA, United States) was used to generate orthomosaic images from the preprocessing of multispectral camera images obtained from each field location and season. The fundamental framework of Pix4Dmapper relies on Ag Multispectral, and its initial processing was carried out using the "Alternative" calibration technique [1].

4.3. Statistical analysis

Data obtained from field experiments were subjected to statistical analysis using R programming software [12]. Multiple comparative analyses were conducted using yield data of the common entries across seasons for each field location.

Data Availability

A pulse crop dataset of agronomic traits and multispectral images from multiple environments (Original data) (Zenodo)

CRediT Author Statement

Kingsley Umani: Data curation, Methodology, Investigation, Visualization, Writing – original draft; **Chongyuan Zhang:** Data collection, Methodology, Investigation, Writing – review & editing; **Rebecca J. McGee:** Conceptualization, Writing – review & editing; **George J. Vandemark:** Conceptualization, Writing – review & editing; **Sindhuja Sankaran:** Conceptualization, Funding acquisition, Investigation, Methodology, Supervision, Project administration, Writing – review & editing.

Ethics Statements

The work does not involve experiments with humans and animals.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

References

[1] C. Zhang, R.J. McGee, G.J. Vandemark, S. Sankaran, Crop performance evaluation of chickpea and dry pea breeding lines across seasons and locations using Phenomics data, Front. Plant Sci. 12 (2021), doi:10.3389/fpls.2021.640259.

[2] A. Marzougui, R.J. McGee, S. Van Vleet, S. Sankaran, Remote sensing for field pea yield estimation: a study of multiscale data fusion approaches in phenomics, Front. Plant Sci. 14 (2023), doi:10.3389/fpls.2023.1111575.

[3] H. Kandel, G. Endres, in: Pulse Crop Production Field Guide for North Dakota (A1922), Northern Pulse Growers Association, 2019, pp. 1–133.

[4] Y. Ji, Z. Chen, Q. Cheng, R. Liu, M. Li, X. Yan, G. Li, D. Wang, L. Fu, Y. Ma, X. Jin, X. Zong, T. Yang, Estimation of plant height and yield based on UAV imagery in faba bean (*Vicia faba* L.), Plant Methods 18 (2022) 1–13, doi:10.1186/s13007-022-00861-7.

[5] J.J. Knodel, G. Shrestha, Pulse crops: pest management of wireworms and cutworms in the northern great plains of United States and Canada, Ann. Entomol. Soc. Am. 111 (2018) (2018) 195–204 4, doi:10.1093/aesa/say018.

[6] C. Summo, D. De Angelis, L. Ricciardi, F. Caponio, C. Lotti, S. Pavan, A. Pasqualone, Data on the chemical composition, bioactive compounds, fatty acid composition, physico-chemical and functional properties of a global chickpea collection, Data Brief 27 (2019) 104612, doi:10.1016/j.dib.2019.104612.

[7] A. Carmona, P. Castro, A. Perez-Rial, J.V. Die, Genomic data of two chickpea populations sharing a potential Ascochyta blight resistance region, Data Brief 50 (2023) 109624, doi:10.1016/j.dib.2023.109624.

[8] B. Kutasy, K. Decsi, G. Hegedűs, G.E. Virág, Dataset of conditioning effect of herbal extract-based plant biostimulants in pea (*Pisum sativum*), Data Brief 46 (2023) 108800, doi:10.1016/j.dib.2022.108800.

[9] J. Huguet, C. Chassard, R. Lavigne, F. Irlinger, I. Souchon, S. Marette, A. Saint-Eve, C. Pénicaud, Dataset about the life cycle assessment of new fermented food products mixing cow milk and pea protein sources, Data Brief 48 (2023) 109263, doi:10.1016/j.dib.2023.109263.

[10] J.R. Lamichhane, H. Boizard, C. Dürr, G. Richard, J. Boiffin, Effect of cropping systems and climate on soil physical characteristics, field crop emergence and yield: a dataset from a 19-year field experiment, Data Brief 39 (2021) 107581, doi:10.1016/j.dib.2021.107581.

[11] C. Hamel, Y. Gan, D. Messer, L.D. Bainard, Soil 16S DNA sequence data and corresponding soil property and wheat yield data from a 72-plot field experiment involving pulses and wheat crops grown in rotations in the semiarid prairie, Data Brief 23 (2019) 103790, doi:10.1016/j.dib.2019.103790.

[12] R Programming Version 4.2.2, R Core Group, Auckland, New Zealand, 2022.