Norwegian University
of Life Sciences

Master's Thesis 2021 60 ECTS
Faculty of Biosciences

Wheat Trait Prediction Using UAV

Henrik Lassegård
Bioinformatikk og anvendt statistikk

Acknowledgements

I would like to thank my supervisor Morten Lillemo and co-supervisors Tomasz Mroz and Sahameh Shafiee for interesting field work and for helping me write this thesis

Abstract

Phenotyping is a major bottleneck in breeding programs or crop-related field experiments. Development of high throughput phenotyping (HTP) methodologies holds promise of mitigating this shortcoming and providing applications capable of non-destructive and rapid recording of accurate phenotypes at large scales.

In this work, three field experiments consisting of 300, 24 and 16 spring wheat varieties, respectively, were planted at Vollebekk research station in field season 2021 and phenotyped with DJI Phantom 4 drone across the growing season.

Two sets of images captured by an unmanned aerial vehicle (UAV) at nominal altitudes of 20 and 8 meter above ground were used to estimate plot heights and model heading status of the plots by texture properties of the images. The estimated traits are compared with manually collected ground truth measurements to see whether the traits can be accurately described.

One set of images was captured from at a 75 degree angle and used for generating digital surface models (DSM). The second set of images was captured at nadir and used to investigate how texture properties of the images relate to the heading process of the plots.

Digital surface models (DSM) produced by Pix4D were used to estimate a terrain model which is used to produce estimates of the heights of the wheat plots. The estimated plot heights were compared to manual plot height measurements to assess the accuracy of the estimates. The DSMs were also used to provide altitude values for three-dimensional models of the plot surfaces.

A dataset of uniform images depicting surfaces of known plots at known times was created from drone images calibrated and undistorted by Pix4D. Grey level co-occurrence (GLCM) texture features were extracted from the dataset and a logistic regression model used to assess the features' ability to discriminate heading status of the plots.

Table of Contents

1 Introduction....1
2 Research objectives....4
3 Methods....5
3.1 Field trials....5
3.1.1 Ground truth measurements....5
3.2 Image acquisition....6
3.3 Image processing....7
3.4 Mask generation....8
3.5 Plot height estimation....9
3.5.1 Uniforming of surface models....9
3.5.2 Terrain model generation....9
3.5.3 Value extraction....10
3.6 Texture analysis....10
3.6.1 Extracting plot images....11
3.6.2 Extraction of texture features....13
4 Results....15
4.1 Ground truth....15
4.2 Plot height estimation....16
4.3 Texture analysis....21
5 Discussion....24
5.1 Ground truth measurements....24
5.2 Plot height estimation....24
5.3 Texture analysis....25
5.4 Review of research objectives....27
6 Conclusions....27
7 References....29

1 Introduction

Currently, it takes 10-15 years from the first cross to produce a new variety/cultivar ready to be released on the market. Considering climate change, this delay becomes increasingly problematic. When environmental fluctuations from season to season becomes more severe and unpredictable, long lasting breeding programs will struggle to keep up. On top of this, increased food production is needed for a growing population. The need for accelerated genetic gains in crops is being recognized worldwide and is, alongside reducing the environmental impact of agriculture, one of two main aims of EUs' agricultural policies (*The Common Agricultural Policy at a Glance*, u.å.).

New plant varieties are created by crossing parents with favorable characteristics and screening the offspring for individuals with favorable phenotypes. Progress in breeding relies on several factors, but most importantly on selection intensity, selection accuracy, heritability of the target trait, genetic variance of the target trait in breeding material and the length of generation interval. This relationship is empirically described as the "breeders' equation":

$$\Delta G = h^2 S = \frac{i \cdot r \cdot \sigma_A}{L}$$

Where ΔG is the genetic gain, h^2 is the narrow sense heritability, S is the selection differential between selected parents and parent population, i is the selection intensity, r is the selection accuracy, σ_A is the genetic variance and L is the length of generation interval (Lush, 1937).

Genetic gains can be increased by shortening the generations interval (L) by breeding in multiple climatic zones. Countries spanning several climate zones, for example Mexico, may have access to two wheat growing seasons per year, which would reduce L of spring wheat by half, effectively doubling genetic gains. It is also a common practice for large, international breeding companies to conduct their field trials in multiple locations worldwide. However, as those environments can be fundamentally different, this strategy severely limits the ability to breed for a specific target niche, which is often desired in smaller and local breeding programs.

Another approach to reduce L is "speed-breeding". It uses controlled-conditions chambers with long days (up to 22 hours) and optimal temperature to achieve up to six generations of wheat, barley or chickpea per year, compared to one generation under field conditions or up to three

generations in a traditional glasshouse with 12/24 days. This method holds great promise for drastically reducing L and thereby increase genetic gains. Unfortunately, it is quite costly, and currently only available to organizations with enough resources to invest in such advanced technology. There is also an ongoing debate about the correspondence of the data from speed-breeding nurseries to “real-world” conditions, which are the ultimate target for any crop variety (Watson et al., 2018).

Gene-editing methodologies can be used to override the natural sources of variability in a trait of interest. They replace recombination mechanisms with gene-specific edits, yielding immediate changes in a trait of interest (Jung et al., 2018). However, since gene- edited crops are regulated as genetically modified organisms (GMO) under the current EU legislation, the technology is currently not being used to develop new varieties in Europe (Turnbull et al., 2021).

Phenotyping is a major bottleneck in any breeding program or crop-related field experiment; genetic gains are contingent on the accuracy of phenotyping and the genetic variance of the breeding pool. Phenotyping is traditionally done by human observers taking notes in the field, thus rate limited by the human resources available and prone to human bias and subjectivity.

Recent rapid advancements, cost reduction and popularization of modern data acquisition tools (drones, robots, and cameras) and data analytics (machine learning, deep learning) hold promise to mitigate those shortcomings by developing high-throughput phenotyping (HTP) methodologies.

In recent years, unmanned aerial vehicle (UAV) imagery has successfully been used to predict numerous important crop traits such as grain yield (Duan et al., 2019; Maimaitijiang et al., 2020; Shafiee et al., 2021; Suab & Avtar, 2020; X. Zhou et al., 2017), above-ground biomass (Han et al., 2019; Li et al., 2020; Lu et al., 2019), plant height (Hassan et al., 2019; Tirado et al., 2020), maturity date (Trevisan et al., 2020; J. Zhou et al., 2019), crop emergence (Lu et al., 2019) to name but a few.

Phenotypes obtained using HTP are often more accurate and replicable (yielding higher heritability) than manual ones, for instance in winterkill estimation (Shafiee et al., in preparation). The drastically increased phenotyping capacity and lack of human-generated error introduced by HTP commits to increase accuracy (r) and possible size of breeding

trials/experiments (σ_A), thereby boosting genetic gains significantly while still maintaining affordability. However, HTP methodologies developed for trait estimation/prediction are in many cases still not mature enough to be applied in commercial breeding programs.

Applications of UAVs for phenotyping usually leverage their easy access to multiple viewpoints of some object of interest to create spatial models not easily obtained by other means. These might include reflectance maps and digital surface models (DSM) of a field. The generation of maps from images depends on photogrammetric methods for aligning the imagery, constructing three-dimensional representations of the contents in them and stitching them to a more or less seamless map. For accurate georeferencing, the image alignments can be anchored to ground control points (GCP) of known exact locations.

2 Research objectives

The objectives of this work are: (1) to develop a pipeline for estimating plant height using UAV imagery, (2) to improve the pipeline by estimating a digital terrain model based on multiple phenotyping missions, (3) to investigate the height models relationship to the ground truth with statistical metrics (4) to investigate the effect of crop canopy density on the accuracy of plant height estimates using UAV imagery and (5) to investigate if heading stage can be reliably estimated using RGB imagery. This work has focus on the technical aspects of the UAV phenotyping data analysis.

For research objectives (1), (2), (3) & (4), estimates of wheat plot heights were extracted from drone imagery and compared with ground truth (manual) measurements. A new pipeline for digital terrain model was developed using multiple flight missions throughout the growing season. Several percentiles of the models, ranging from median to max were screened for their correspondence with the ground truth estimates and viewed in light of canopy densities.

For research objective (5), a novel pipeline for extracting plot-level standardized images of crop canopies was developed based on low altitude RGB UAV imagery. Grey-level co-occurrence matrix (GLCM) was used to extract a number of features for each image. Those features were next used to construct a model to predict heading stage in spring wheat.

3 Methods

3.1 Field trials

Wheat trials were conducted during the summer of 2021 at Vollebekk Research Station (Ås, south-eastern Norway, 59º39'N, 10º45'E). Several fields with different compositions of wheat varieties were included in the trials. Data from three of them are used in this thesis: 24 historical cultivars released in Norway between 1972 and 2019 (referred to later as the robot field), 16 cultivars used in the NobalWheat project nitrogen use efficiency trial (referred to later as NobalNUE) and a panel of 300 current and historical spring wheat varieties (referred to later as MASBASIS).

Field trials were sown on the 23rd, 23rd and the 19th of April 2021 for Robot, NobalNUE and MASBASIS trials, respectively.

In the robot and NobalNUE fields, two rates of 75 kgN·ha^{-1} and 150 kgN·ha^{-1} of compound NPK fertilizer (YaraMila 22-3-10) were applied to different parts of the fields before sowing to evaluate the effects of fertilization rates on yield performance and physiological traits. This fertilization treatment has no detectable effect on heading date, a positive effect on canopy density and a small positive effect on plant height (Mroz et al., under review). The extra variability introduced by the fertilization treatment was used to increase the variance of plant height and to evaluate the influence of canopy density on plant height estimation using drone imagery.

In robot and NobalNUE the final plot sizes were 1.5m x 6.5m and in MASBASIS, 1.5m x 5m. The seeding rate was approximately 25g of seeds per m^2. After germination, alleys between the plots were sprayed with glyphosate.

All the trials were managed according to local practice to keep the fields free of weeds and fungal diseases.

3.1.1 Ground truth measurements

The ground truths for plot height and heading date were manually measured for each plot.

The heading date for a plot was defined as the date when heads had emerged in half of the plants in the plot and was judged visually in the field.

Plot heights were recorded shortly after anthesis, on July 9th, when the plants had reached their final heights, with except for parts of MASBASIS, which were measured July 12th. The heights were recorded as the average height of three to five bundles of fertile stems measured from soil bed to the top of the spikes (excluding awns) with a meterstick.

3.2 Image acquisition

All the images used here were captured from a DJI Phantom 4 (original) with a DJI FC330 camera[1] Flight paths used for capturing the images were made in DJI GS Pro, running on an iPad (Figure 1 shows the interface of similar app).

Figure 1: Creation of a flight path.

The images intended for height modelling and texture analysis were captured at nominal altitudes of 20 and 8 meters above the ground, respectively. Images for height modelling were captured with 80% frontal and 85% side overlap. The images for the texture analysis were

1 https://www.dji.com/no/phantom-4/info

captured with 80% frontal and 80% side overlap. An overview of when the two types of flight were done over which fields is shown in Figure 2.

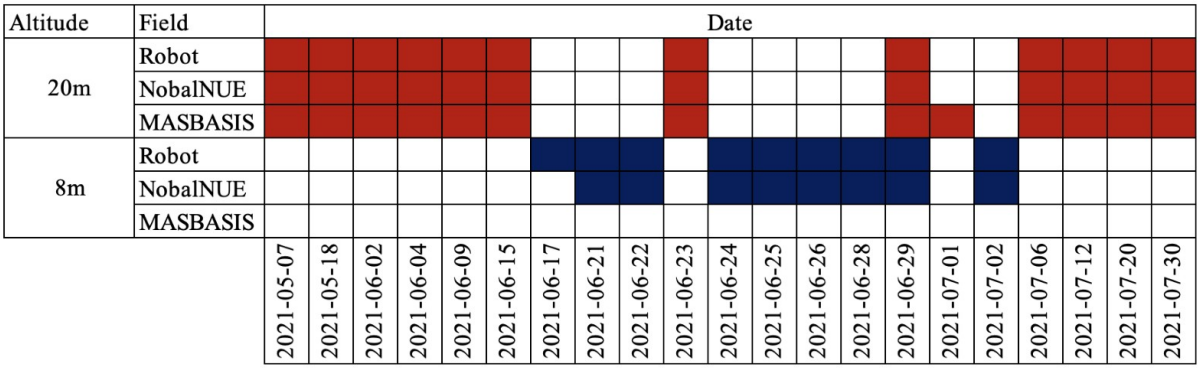

Figure 2: Gantt chart showing when, and over which fields, the two types of flights were done during the summer

Ground control points (GCP) had been placed in the fields and their locations recorded using an RTK GPS with cm precision. The GCPs were located in each of the four corners of the MASBASIS and robot fields. Two additional GCPs were located in the middle of MASBASIS.

3.3 Image processing

All images were grouped by the flight they originated from and processed together in Pix4D. Images of the NobalNUE field were processed together with those of the robot field if they had both been flown over the same day, as no GCPs were present in NobalNUE.

Pix4D was used to create digital surface models (DSM) utilized for height modelling, and to provide undistorted images and the information needed to locate mask coordinates in them. Pix4D saves the DSMs it produces in geotiff format in which the pixel values represent the estimated altitudes of the surfaces of the field.

To locate mask coordinates in undistorted images, Pix4D provides a transformation matrix per calibrated image (pmatrix) and a common offset to be subtracted. Given three-dimensional coordinates in the CRS used in the Pix4D project, pixel coordinates in the undistorted images can be obtained[2]

2 https://support.pix4d.com/hc/en-us/articles/202977149-What-does-the-Output-Params-Folder-contain

$$(X', Y', Z') = (X, Y, Z) - offset$$

$$(x, y, z)^t = \text{PMatrix} * (\text{X}', \text{Y}', \text{Z}', 1)^t; u = \frac{x}{z};\ v = \frac{y}{z}$$

The photogrammetry performed by Pix4D is computationally heavy, necessitating designated computers to handle the map stitching. Two remotely accessed computers running versions 4.4.12 and 4.5.3 of Pix4Dmapper were used to process the data.

3.4 Mask generation

Orthomosaics produced by Pix4D were loaded in QGIS 3.22 (Development Team, 2021) and masks of the fields and the plots within fields were hand drawn on top of them.

Both types of masks were drawn on new temporary scratch layers in the same CRS as the maps and exported as geojson files.

Field masks were drawn as quadrilaterals extending comfortably beyond the GCPs in all corners of the fields. Plot masks were drawn as rectangles aligned with the edges for each plot in the robot and NobalNUE fields and scaled down to 80% of their original size to avoid the edges of the plots and allow some margin for error in map alignment (Figure 3).

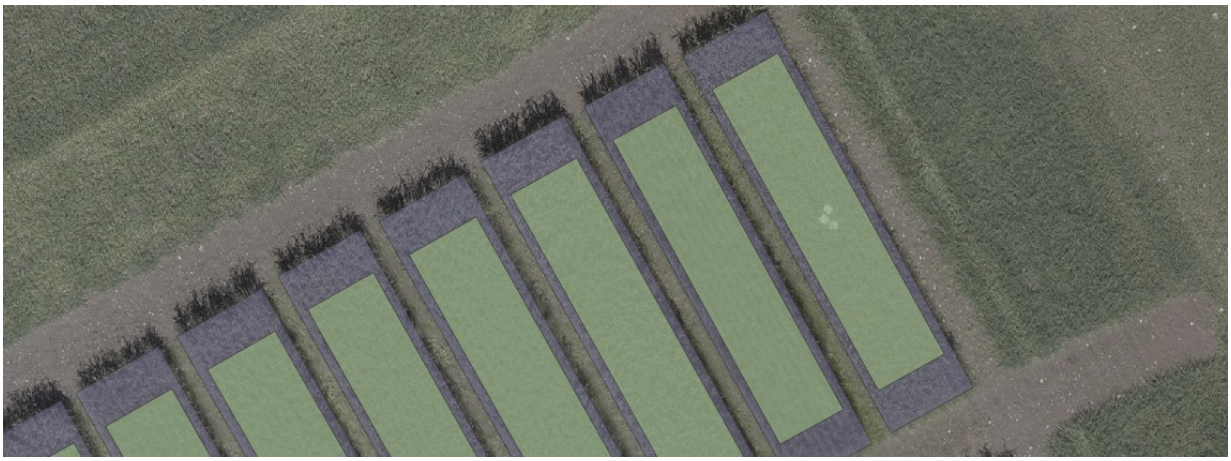

Figure 3: Part of an orthomosaic showing the robot field with full size masks in blue and scaled masks in green overlaid

3.5 Plot height estimation

DSMs were used to create a terrain model (DTM) per field. Height maps were obtained by subtracting a fields DTM from the DSMs. Estimates of plot heights were then extracted from the height maps.

3.5.1 Uniforming of surface models

To facilitate doing computations on the maps, they were cropped, resized and stacked in a single data structure using the rasterio python package.

Target sizes for resizing the maps were found by dividing the extents of the fields by their mean resolutions (and rounding to an integer). The widths and heights of the fields (in meter) were obtained from the field masks described above. The resolutions of the DSMs were read from the geotiffs with rasterio, and the mean resolution calculated per field.

Each field was cropped from the DSMs containing it with rasterio.mask.mask, resized with rasterio.warp.calculate_default_transform and rasterio.warp.reproject and written to a layer in a three-dimensional dataset in a HDF5 file.

3.5.2 Terrain model generation

Areas depicting the ground in the DSMs should reflect its assumed stability throughout the season. The standard deviation was computed along the stacked axis to produce a map of the variation in surface altitudes in the different regions of the DSMs throughout the season. Figure 4 shows the standard deviation map computed for the robot field to the left.

To identify ground areas from the standard deviation map, a few low percentiles of its values were tried as thresholds to produce a mask, considering values lower than the threshold to belong to ground areas. The mask from the lowest percentile yielding interpolatable gaps was chosen.

Median values throughout the season were filled in the mask areas. The median ground map for the robot field is shown in the middle of Figure 4. Finally, the non-ground areas of the median ground map were interpolated with rasterio.fill.fillnodata to create a terrain model.

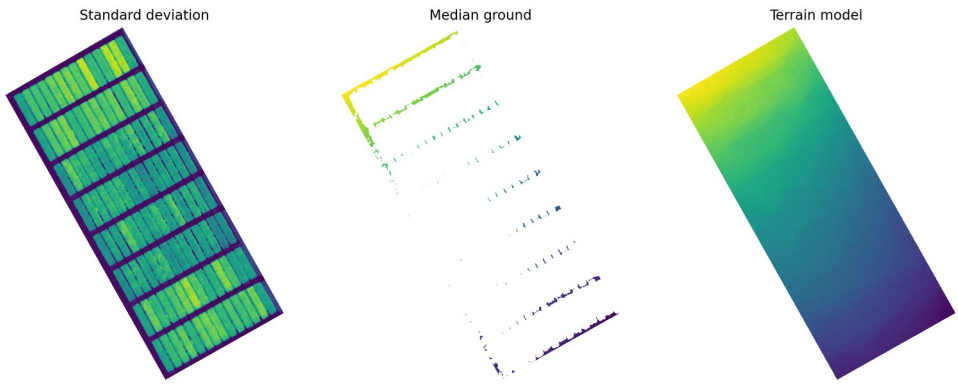

Figure 4: The Standard deviation of the DSMs, the Median ground altitudes where values in the standard deviation map are below a chosen threshold and the terrain model, in which non-ground areas of the median ground map has been interpolated.

3.5.3 Value extraction

Plot surface altitude and height values were extracted from the maps using the plot masks and the rasterstats.zonal_stats function.

Altitude values were extracted directly from the stacked DSMs. Canopy height values were extracted from the height maps resulting from the subtraction of a fields DTM from its DSMs.

For both altitude and height values, the 50^{th} (median), 55^{th}, 65^{th}, 75^{th}, 85^{th}, 95^{th} and 100^{th} (max) percentiles were extracted to investigate which would describe the ground truth better and whether there were any trends in the distributions of the models.

3.6 Texture analysis

A dataset of uniform, square images depicting known plots at known times were extracted from UAV images calibrated and undistorted by Pix4D. Texture features were extracted from the dataset and joined with plot level field data to investigate whether texture features of wheat plot images can describe the plots' heading status.

3.6.1 Extracting plot images

To be most accurately compared, steps were taken to extract images as uniform as possible in the extent of the plots depicted, the angle the plots were viewed at, and the orientation of the sowing rows in them. Squares to crop from the undistorted images were found using their calibrated camera parameters, provided by Pix4D, plot masks and extracted altitude values, and were aided by the use of the shapely python package.

For each of the flights, the extracted surface altitude values, described in section 3.5.3 were interpolated at the hour of the first image per flight for each plot. New sets of three-dimensional plot masks were created by adding the altitude values as a third dimension to the two-dimensional masks. The ground under the plots and the plot canopies were assumed to be flat, therefore only a single altitude value per plot per flight was used.

For each image, the three-dimensional masks were transformed to image coordinates using the pmatrix and offset files generated by Pix4D. The transformed masks were used to check which plots were visible in each image.

A circular mask was created for each plot mask in each image to limit the range of viewing angles. The circle was drawn around the calibrated x, y positions of the cameras in the plane of the estimated plot surfaces with radii found by dividing the distance from the camera to the plot surface altitude by the tangent of 80 degrees. This circle is illustrated in the first column of Figure 5.

If the plot mask and its corresponding circular mask intersected, the centroid of the intersection and the shortest distance from the centroid to the perimeter of the intersection were found. This distance was compared against a desired length of 0.325 m to check whether a square would fit and to find a scaling factor.

If a square would fit, the corresponding distance found in image coordinates was scaled down and a diamond shape with a diagonal of twice the length of the distance constructed. Figure 5 illustrates some of the steps performed after the two masks have been created. To align with the structures of the plot, the orientation of that of the two long edges of the plot mask closest to the center of the circle was found and the diamond shape rotated accordingly.

At most one crop were made per plot per undistorted image. Prepending the plot identifier to the name of the original image ensured unique names for the every image in the dataset.

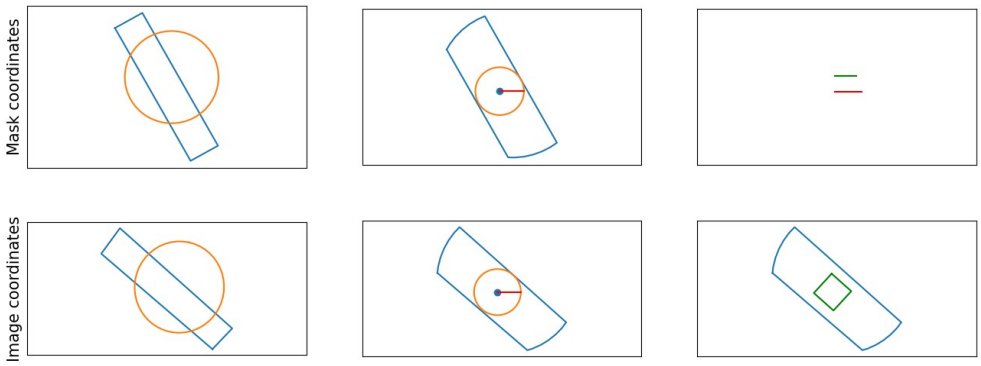

Figure 5: Shows the masks and shapes used when extracting a square from the images. The top row shows shapes defined in the coordinate system of the masks, with units in meter. The bottom row shows shapes transformed to image coordinates. In the first column, the intersection of the two masks are found. In the second column, the centroid of their intersection and its distance to the intersection perimeter is shown. A circle has been drawn using the distance as radius for illustration purposes. In the third column, second row, the same distance as in the second column is shown in red. It is compared to a desired length (in green) to get a scaling factor which is applied when constructing the final square in image coordinates (third column, second row).

Figure 6: Example of an image captured by the drone and the shapes used to extract parts of the plots. Pink shapes are plot masks, the circle in the center is used to limit the viewing angle, and the olive colored squares are the parts to be cropped from the image.

3.6.2 Extraction of texture features

Before computing any texture features, the images were resized to 120x120 pixels, and histogram equalized with skimage.exposure.equalize_adapthist, to correct for varying lighting conditions on the different dates.

A time series of irradiation measurements (W/m^2) was obtained from NIBIO[3] («Q0 – Globalstråling», the sum of direct and diffuse sun irradiation) to investigate its effect on the images and whether histogram equal.

A grey level co-occurrence matrix (GLCM) is a square matrix counting the times two greyscale values occur together in an image at a given offset (Haralick et al., 1973). Several features can be calculated from a GLCM that characterizes the texture in the image

3 https://lmt.nibio.no/agrometbase/getweatherdata_new.php?weatherStationId=5

No assumptions were made about which features and combinations of distances and angles were likely to carry information about heading state. GLCMs were therefore collected using the combinations of a range of offsets (1, 2, 4, 8, 16, 32 and 64 pixels) and three angles (0, 45 and 90 degrees) for each layer in the RGB images, using skimage.feature.graycomatrix. For each GLCM, the contrast, dissimilarity, homogeneity, angular second moment (ASM), energy and correlation (Table 1) were calculated with skimage.feature.greycoprops.

Table 1: Formulas for the GLCM features. P is the GLCM.

Contrast	$\sum_{i,j=0}^{levels-1} P_{i,j}(i-j)^2$
Dissimilarity	$\sum_{i,j=0}^{levels-1} P_{i,j}\lvert i-j\rvert$
Homogeneity	$\sum_{i,j=0}^{levels-1} \frac{P_{i,j}}{1+(i-j)^2}$
ASM	$\sum_{i,j=0}^{levels-1} P_{i,j}^2$
Energy	$\sqrt{ASM}$
Correlation	$\sum_{i,j=0}^{levels-1} P_{i,j}\left[\frac{(i-\mu_i)(j-\mu_j)}{\sqrt{(\sigma_i^2)(\sigma_j^2)}}\right]$

4 Results

4.1 Ground truth

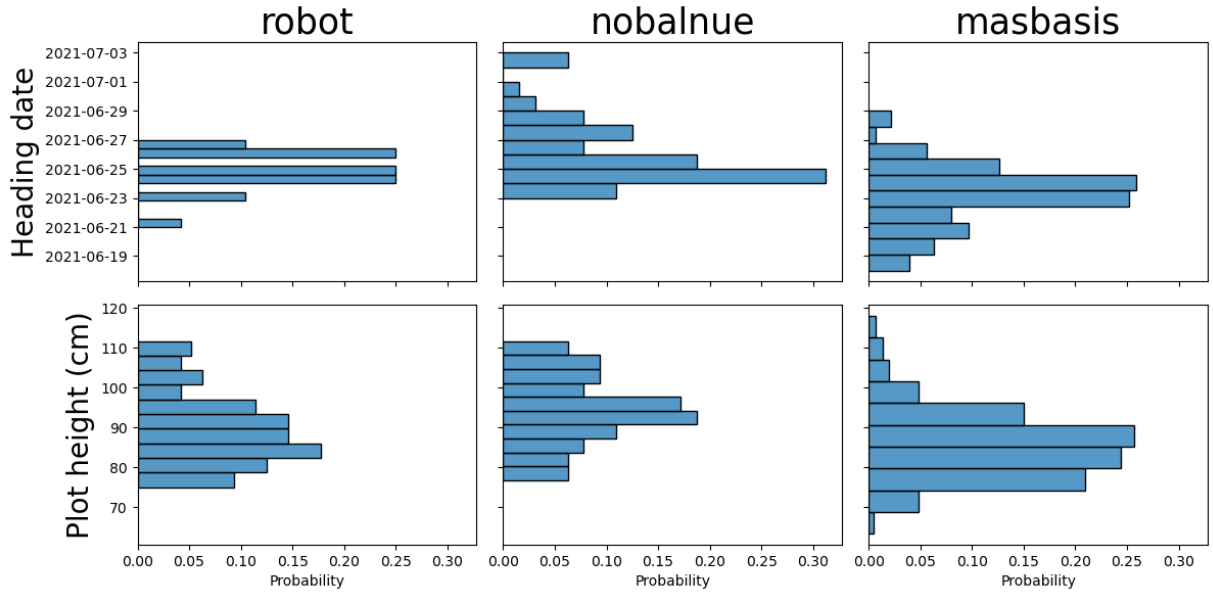

Figure 7: Histograms of heading dates and heights for plots in the three fields. The bars sum to 1.

The distributions of heading dates and heights of the plots are shown for each field in Figure 7. MASBASIS, containing the greatest genetic diversity had the greatest variation in phenotypes for both traits, with heights ranging from 63 cm to 118 cm. Plots in the NobalNUE field headed on average later than those in the other fields.

4.2 Plot height estimation

The flights closest in time to the manual height measurements were done July 6th and July 12th, three days before and after the manual measurements. Plot height estimates from both these dates were compared to the ground truth.

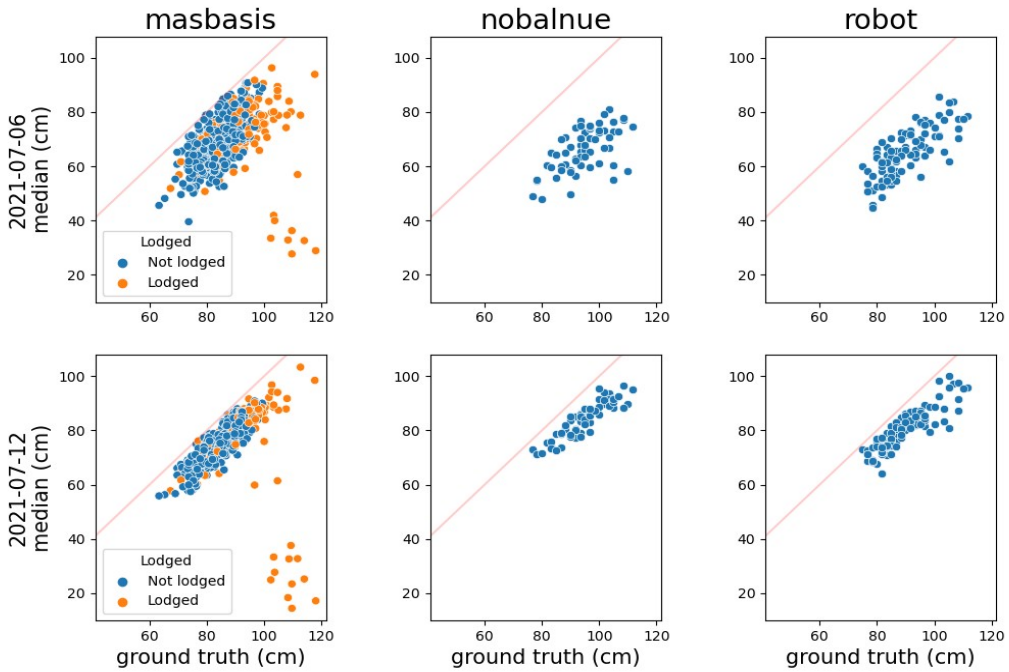

Figure 8: Scatterplots of the measured plot heights on the x-axes against the median value of the modelled plot heights on the y-axes. The top row shows the heights modelled from flights done three days prior to the manual measurements and the bottom row from a flight done three days after (except for parts of MASBASIS, which were measured the same day). Plots that had lodged by the end of the season are colored orange. They had not necessarily lodged at the time of the flights. 1:1 line in red.

The median height values estimated from the imagery consistently underestimates the ground truth, by 18.5 and 10.8 cm on average across all the fields for July 6th and July 12th respectively. Values estimated on July 6th show a bigger spread around the ground truth compared to the estimates from July 12th (Figure 8).

The relationship between the maximum value estimated from the imagery and the ground truth is subject to strong variability among the trials and flight days (Table 2). However, the maximum values get closest to the true plot height, shown by the lowest mean absolute error values (Table 3).

Median, as the default metric used in plant height estimation, shows high and consistent correlation with the ground truth through all the trials and flight missions, on average outperforming the maximum value but still underestimating the true height (Table 3, Figure 8). The correlations tend to increase, and the errors tend to decrease with higher percentile values of the height models, up to and including the 95th percentile. (Table 2, Table 3).

Table 2: The correlations between the statistics extracted from the models and the manual measurements for each of the fields for the two dates. Plots that had lodged by the end of the season are not included.

	MASBASIS		NobalNUE		robot	
Percentile	2021-07-06	2021-07-12	2021-07-06	2021-07-12	2021-07-06	2021-07-12
median	0.69	0.87	0.60	0.90	0.76	0.89
55	0.69	0.87	0.61	0.91	0.76	0.89
65	0.69	0.88	0.64	0.91	0.77	0.90
75	0.70	0.89	0.66	0.91	0.78	0.91
85	0.71	0.89	0.68	0.91	0.78	0.92
95	0.73	0.90	0.70	0.91	0.79	0.93
max	0.78	0.89	0.51	0.89	0.58	0.89

Table 3: The mean absolute errors of the statistics extracted from the models and the manual measurements for each of the fields for the two dates. Plots that had lodged by the end of the season are not included.

	MASBASIS		NobalNUE		robot	
Percentile	2021-07-06	2021-07-12	2021-07-06	2021-07-12	2021-07-06	2021-07-12
median	14.81	9.50	28.16	10.49	24.97	9.24
55	14.22	9.13	27.60	10.02	24.42	8.87
65	12.98	8.41	26.51	9.07	23.33	8.13
75	11.63	7.67	25.34	8.17	22.13	7.33
85	9.97	6.78	24.02	7.16	20.63	6.36
95	7.53	5.46	21.96	5.72	18.07	4.97
max	4.26	3.57	17.50	3.90	12.79	3.25

Table 4: The median difference in cm between the DSMs and the DTM for the areas within the ground mask.

	MASBASIS	NobalNUE	robot
2021-07-06	0.01	0.81	0.74
2021-07-12	-2.85	-0.20	1.52

The ground areas found in the DSMs from July 6th were on average 0.01, 0.81 and 0.74 cm higher than those in the DTMs (the median of the DSMs) for the MASBASIS, NobalNUE and robot fields, respectively (Table 4). The same differences calculated for July 12th showed that the ground areas in the DSMs for MASBASIS and NobalNUE on average were 2.85 and 0.20 cm lower than those in the DTM. In the robot field they were 1.52 cm higher on July 12th.

Comparing values between the two nitrogen treatments in robot and NobalNUE reveals that the low nitrogen group on average was underestimated more than the high nitrogen group (Table 6). This underestimation is also reflected in Figure 9 in which the 95th percentile of the models are plotted against days to heading. The differences in errors were larger on July 6th than on July 12th. The correlations between the models and the ground truth are similar for the nitrogen levels in the upper percentiles (excluding max) but diverges closer to the median (Table 5).

Table 5: The mean correlations between the statistics extracted from the models and the manual measurements for each of the two dates for each of the two nitrogen treatments for the robot and nobalnue fields combined.

	2021-07-06		2021-07-12	
Percentile	Low N	High N	Low N	High N
median	0.73	0.81	0.88	0.92
55	0.74	0.81	0.88	0.93
65	0.76	0.82	0.90	0.93
75	0.78	0.82	0.91	0.93
85	0.79	0.81	0.92	0.93
95	0.78	0.81	0.93	0.93
max	0.37	0.80	0.93	0.87

Table 6: The mean absolute errors of the statistics extracted from the models and the manual measurements for each of the two dates for each of the two nitrogen treatments for the robot and nobalnue fields combined.

	2021-07-06		2021-07-12	
Percentile	Low N	High N	Low N	High N
median	30.09	22.40	10.38	9.10
55	29.48	21.91	9.98	8.68
65	28.28	20.92	9.19	7.83
75	26.97	19.86	8.34	6.99
85	25.42	18.55	7.33	6.03
95	22.80	16.45	5.84	4.69
max	17.39	11.95	3.58	3.44

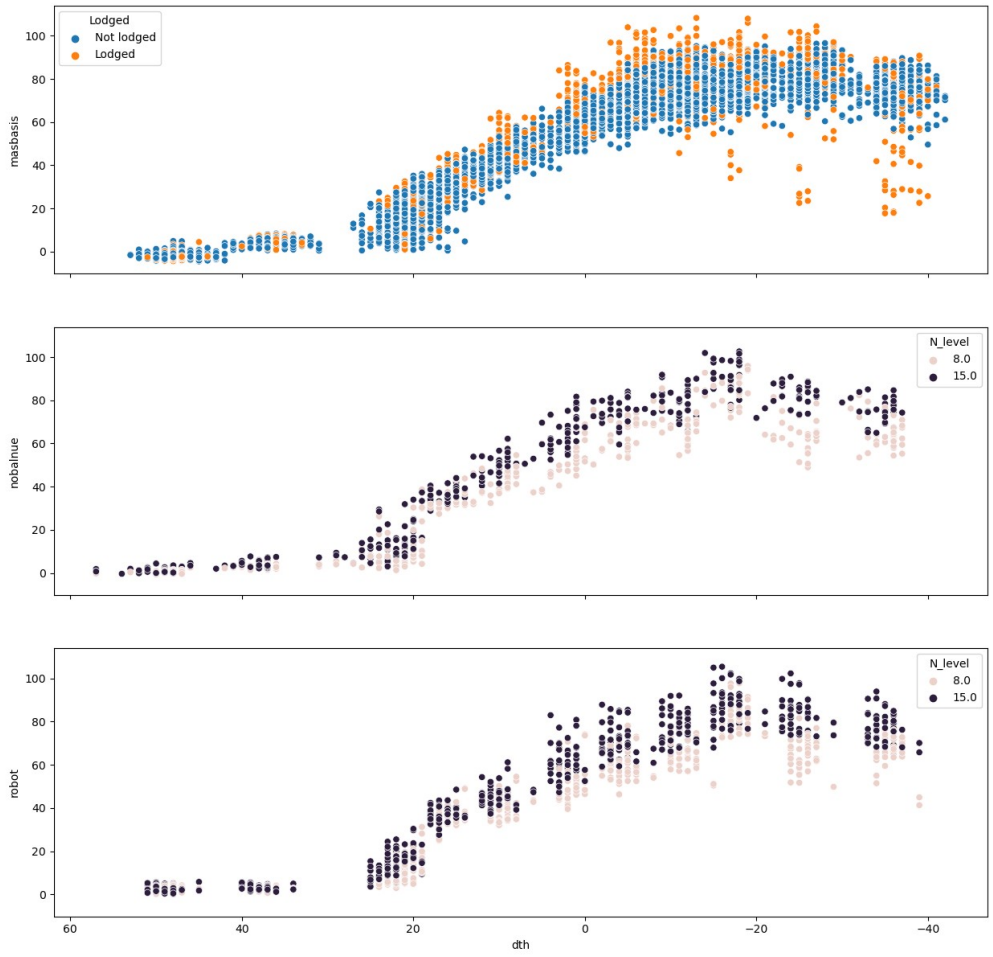

Figure 9: Scatterplots of the 95th percentile of the height models per plot on the y-axis against days to heading on the x-axis for each field. For MASBASIS the points are colored by whether the plots were judged to be lodged at the end of the season or not. For the robot and NobalNUE fields the points are colored by the nitrogen treatment the plots were subject to.

From Figure 9 the maximum estimated heights are on average reached around 15 days after heading for plots in the robot and NobalNUE fields and approximately 10 days after heading for the MASBASIS field. A decline in the estimated heights after reaching their max is also visible.

4.3 Texture analysis

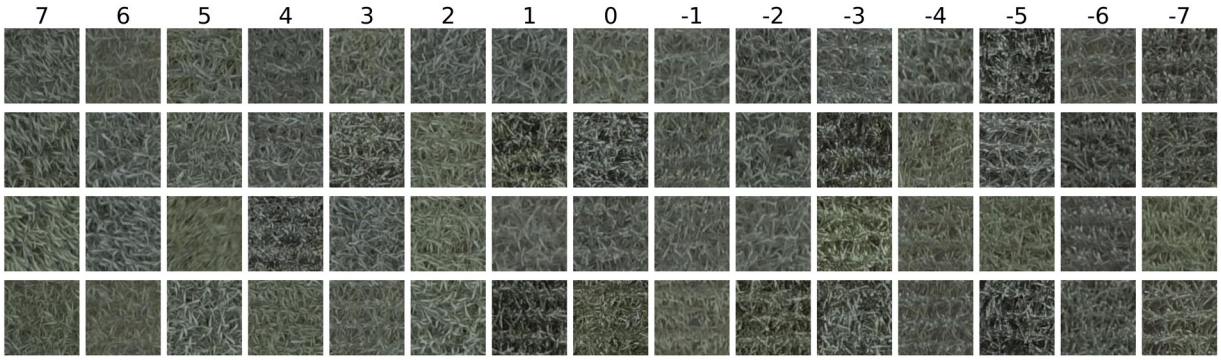

Figure 10: Four randomly drawn images from the dataset for each of the days from seven days before heading to seven days after heading.

A total of 4201 images depicting plot canopies within one week of their heading date were cropped from the UAV images. When compared visually, the cropped images showed signs of increasingly distinguished sowing rows with the development of the plants. Heads were visible in many of the post-heading images. Variations in contrast due to varying light conditions were also visible. Some example images are shown in Figure 10.

Logistic regression models were fitted on the GLCM features, using the heading status (headed/not headed) as target variable, with a leave-one-date-out approach; data from each of the flight dates were in turn used as the test-set and the model fitted on the rest of the data. The results are shown in Table 7.

Images taken the same date as a plot was heading were excluded when fitting the models.

Using these splits, the ratios of headed plots included in the training sets were fairly balanced for all dates. The ratios of headed plots in the test sets increased for each date (Table 7) as heading progressed in the fields.

Figure 11 shows a histogram of interpolated irradiation values at the time of each image. From the principal component plot in Figure 12 it is clear that there is some effect of irradiation in the GLCM features, even after histogram equalization.

Table 7: The ratios of headed images in the training sets (Excl. Date) and test sets (Date) used, the number of images in each set for each split and the resulting confusion matrix and accuracies when predicting on the test set.

	Ratio headed		N images			Confusion matrix			
	Excl. Date	Date	Excl. Date	Date	Accuracy	TP	FN	FP	TN
2021-06-17	0.56	0	3670	120					
2021-06-21	0.62	0	3330	460					
2021-06-22	0.62	0.03	3272	518	0.61	9	7	194	308
2021-06-24	0.58	0.18	3464	326	0.32	52	8	214	52
2021-06-25	0.54	0.55	3260	530	0.58	218	75	145	92
2021-06-26	0.51	0.76	3364	426	0.68	250	72	65	39
2021-06-28	0.47	0.96	3213	577	0.53	282	273	1	21
2021-06-29	0.48	0.96	3321	469	0.89	409	43	10	7
2021-07-02	0.50	0.97	3426	364	0.39	133	221	0	10

With the exception of June 28th and July 2nd, the classification results show a high number of false positives; plots predicted to have headed when in fact they had not. June 24th was the day with the highest irradiation (Figure 11) and achieves a low accuracy due to the model erroneously predicting 214 of the 266 not headed plots as headed. The two next brightest days, June 26th and 29th, achieves higher accuracies when their headed plots are correctly predicted.

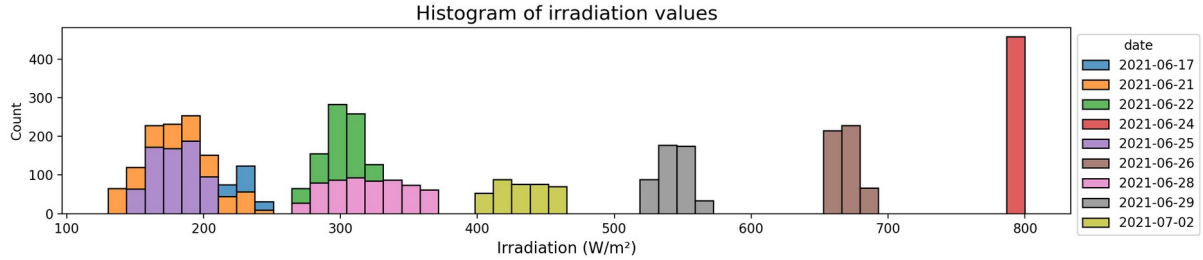

Figure 11: Histogram of interpolated irradiation values at the time of each image. Colored by date of the flight

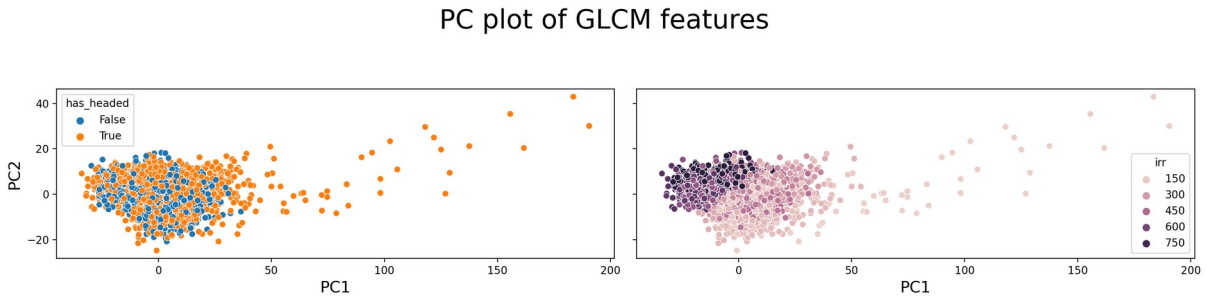

Figure 12: Scatterplots of the two first principal components of the GLCM features. Colored by the heading status of the depicted plot to the left and by the irradiation at the time of capturing the image to the right

5 Discussion

5.1 Ground truth measurements

The heading date is defined as the day heads have emerged from half of the plants for a plot. Visual judgement of heading dates in the field is complicated by both morphological variation between the cultivars and uneven development within a plot, caused by either soil or genetic conditions. Variation within plots also complicates measurement of plot heights. Gradients in height across the plot will be averaged when taking multiple measurements, but high local variation can make it difficult to decide on a representative height for an area. Scoring of heading date and plot height will both to some degree be influenced by the subjective judgement of the observer.

Another aspect to manual scores is their purpose. The scoring for plant height and heading date in this work were performed the same way a plant breeder would. For breeders or scientists, the absolute values for a trait may be less important than the relative differences between varieties. This because most traits are a function of the environment and their absolute values will vary more with the environment than their relative differences, or rankings, between varieties.

A phenomenon describing the situation when the rankings of accessions differ from environment to environment is called genotype by environment interaction (GxE) (Hill, 1975).

The actual method of scoring a trait is therefore in many cases secondary, as long as it is executed consistently over the whole experiment.

5.2 Plot height estimation

The plot height estimates generated from UAV imagery here from July 12th showed high correlation with the ground truth, which means that the two agree on the relative height differences between the plots. The low correlation with the ground truth of the models from July 6th (Table 2) may suggest that not all plots had reached their full heights yet at that date. This is also reflected in Figure 9, in which the plots appear to reach their maximum heights some days after heading.

The heights were however consistently underestimated. Manual measurements consider the total height of the plants, from soil bed to the top of the spikes. The UAV method cannot

benefit from such precise distinctions and its estimations will incorporate a portion of the leaf canopy heights, which will result in lower than ground truth estimates.

In the fields, plots in the two nitrogen treatments had visibly different canopy densities. It may be harder for Pix4D to find and match keypoints in the top of the sparse canopies of low nitrogen plots than in the top of the denser plots in the high nitrogen treatment, which could be why Table 5 and Table 6 show larger differences between the nitrogen treatments in the lower than in the upper percentiles (excluding max).

When drawing the plot masks used to extract the height and altitude values, they were scaled in size around their center, which excessively reduced the lengths of the long edges. More representative measurements from the maps, and more, or larger, plot surface images, could have been cropped from undistorted images had the plot masks been scaled more appropriately.

5.3 Texture analysis

Before capturing images for texture analysis, a balance had to be found between the altitude of the flight and the necessary coverage for successful camera calibration in Pix4D, as the flight time and number of images captured increases substantially when lowering the altitude while keeping the coverage constant.

When extracting the images of plot surfaces, steps were taken to ensure they were as uniform as possible to allow more confident results than could otherwise be obtained. This was assumed to be important for single statistics like the GLCM properties to be comparable between images.

For limiting the range of viewing angles in the undistorted images, as described in above, it was assumed that the ground was flat and that the straws were standing straight up, reaching their estimated heights. In reality the plots have a more chaotic structure.

The texture in an image of a plot canopy will be more sensitive to the camera angle the further the heads are from the background. As the heads emerge and rise above the leaves, their stems will make up greater portions of the image unless viewed precisely from above, thus influencing the texture. Figure 13 shows this with images of a plot in the robot field from three dates. In the first column, the canopy only consists of leaves, and it is hard to perceive camera

or straw angles, in the second and third columns the heads appear and the leaves wilt, making it easier to get a sense of their angles.

An important factor influencing the texture perception of GLCM is dierct solar irradiation at the time of capturing the images.

When the three brightest days, June 24th, 26th and 29th, are left out, there is a tendency for the model to shift towards predicting that the plots in them have headed (high FP for June 24th and high TP for June 26th and 29th Table 7). This could suggest that some effect of the light is not corrected for by the histogram equalization and captured by the GLCMs. A shift towards predicting that plots have headed is seen also in June 25th however, which was one of the dimmest days (Figure 11). More variation in the data would be desirable.

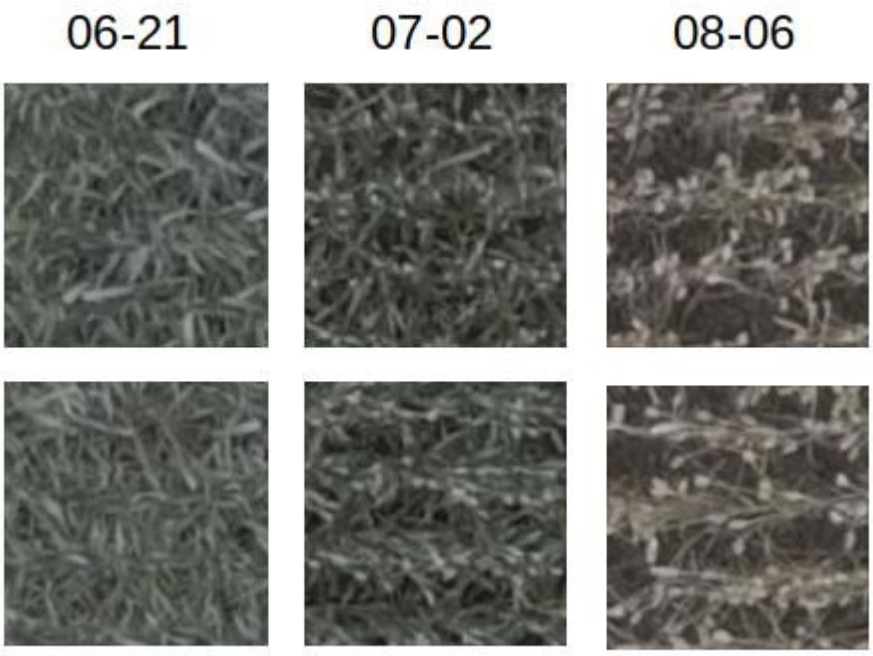

Figure 13: Two cropped images of plot 1210 from the robot field from three different dates. In the first column, the canopy consists of leaves, in the second column heads have emerged and in the third column the leaves have wilted leaving the ground visible

GLCM is a relatively simple and straight forward technique for describing texture, but GLCM texture values cannot easily be transferred from one situation to another (Hall-Beyer, 2017), and may not be the most appropriate method for representing image texture in different applications. Therefore, one should be cautious applying a model developed in on one dataset on images captured in different conditions.

New approaches for representing texture are usually based on Convolutional Neural Networks (CNN) (Liu et al., 2019). CNNs are both more flexible and powerful and capable of handling more variation in the input data, which could make them useful on less uniform images than are presented here.

5.4 Review of research objectives

A pipeline for estimating plant height using UAV imagery was developed, viewing multiple flights in comparison to enable creation of a DTM necessary for plot height estimation. The height models' relation to the ground truth were asessed by comparing several extracted model statistics to the ground truths, The effect of canopy densities on the height models was checked by comparing the results from plots in two different nitrogen treatments. A pipeline for extracting plot level images from UAV imagery was also developed and heading status attempted classified from their texture features.

6 Conclusions

In this thesis, en extensive dataset of aerial imagery was analyzed and compared against ground truth data for plant height and heading date.

Plant height estimation based on DSMs and a DTM estimated from the DSMs achieved high correlations with ground truth (0.51 to 0.93, depending on the metric used). The median metric, commonly used for UAV plant height estimation, tends to underestimate the heights compared with manual measurements. However, this underestimation is consistent and does not significantly affect the rankings of the varieties. Among other metrics investigated, the 95th percentile is the one that resembles the ground truth the most and the best alternative to common practice.

A dataset of plot surface images was created, using extracted altitude values from DSMs and UAV images captured at low altitudes. Results from classification of heading states of plots from GLCM features of the images was not conclusive, but suggests sun irradiation is an influencing factor on the features.

7 References

Duan, B., Fang, S., Zhu, R., Wu, X., Wang, S., Gong, Y., & Peng, Y. (2019). Remote Estimation of Rice Yield With Unmanned Aerial Vehicle (UAV) Data and Spectral Mixture Analysis. *Frontiers in Plant Science*, *10*, 204. https://doi.org/10.3389/fpls.2019.00204

Han, L., Yang, G., Dai, H., Xu, B., Yang, H., Feng, H., Li, Z., & Yang, X. (2019). Modeling maize above-ground biomass based on machine learning approaches using UAV remote-sensing data. *Plant Methods*, *15*(1), 10. https://doi.org/10.1186/s13007-019-0394-z

Haralick, R. M., Shanmugam, K., & Dinstein, I. (1973). Textural Features for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics*, *SMC-3*(6), 610–621. https://doi.org/10.1109/TSMC.1973.4309314

Hassan, M. A., Yang, M., Fu, L., Rasheed, A., Zheng, B., Xia, X., Xiao, Y., & He, Z. (2019). Accuracy assessment of plant height using an unmanned aerial vehicle for quantitative genomic analysis in bread wheat. *Plant Methods*, *15*(1), 37. https://doi.org/10.1186/s13007-019-0419-7

Jung, C., Capistrano-Gossmann, G., Braatz, J., Sashidhar, N., & Melzer, S. (2018). Recent developments in genome editing and applications in plant breeding. *Plant Breeding*, *137*(1), 1–9. https://doi.org/10.1111/pbr.12526

Li, B., Xu, X., Zhang, L., Han, J., Bian, C., Li, G., Liu, J., & Jin, L. (2020). Above-ground biomass estimation and yield prediction in potato by using UAV-based RGB and hyperspectral imaging. *ISPRS Journal of Photogrammetry and Remote Sensing*, *162*, 161–172. https://doi.org/10.1016/j.isprsjprs.2020.02.013

Lu, N., Zhou, J., Han, Z., Li, D., Cao, Q., Yao, X., Tian, Y., Zhu, Y., Cao, W., & Cheng, T. (2019). Improved estimation of aboveground biomass in wheat from RGB imagery and

point cloud data acquired with a low-cost unmanned aerial vehicle system. *Plant Methods*, *15*(1), 17. https://doi.org/10.1186/s13007-019-0402-3

Lush, J. L. (1937). *Animal breeding plans,*. Collegiate Press, Inc.

Maimaitijiang, M., Sagan, V., Sidike, P., Hartling, S., Esposito, F., & Fritschi, F. B. (2020). Soybean yield prediction from UAV using multimodal data fusion and deep learning. *Remote Sensing of Environment*, *237*, 111599. https://doi.org/10.1016/j.rse.2019.111599

Mroz, T., Dieseth, J. A., & Lillemo, M. (under review). *Historical Grain Yield Genetic Gains in Norwegian Spring Wheat Under Contrasting Fertilization Regimes*.

QGIS Development Team. (2021). *QGIS Geographic Information System*. QGIS Association. https://www.qgis.org

Shafiee, S., Lied, L. M., Burud, I., Dieseth, J. A., Alsheikh, M., & Lillemo, M. (2021). Sequential forward selection and support vector regression in comparison to LASSO regression for spring wheat yield prediction based on UAV imagery. *Computers and Electronics in Agriculture*, *183*, 106036. https://doi.org/10.1016/j.compag.2021.106036

Shafiee, S., Mroz, T., & Lillemo, M. (in preparation). *Winterkill estimation in winter wheat using UAV multispectral imagery outperforms manual scoring in accuracy*.

Suab, S. A., & Avtar, R. (2020). Unmanned Aerial Vehicle System (UAVS) Applications in Forestry and Plantation Operations: Experiences in Sabah and Sarawak, Malaysian Borneo. I R. Avtar & T. Watanabe (Red.), *Unmanned Aerial Vehicle: Applications in Agriculture and Environment* (s. 101–118). Springer International Publishing. https://doi.org/10.1007/978-3-030-27157-2_8

The common agricultural policy at a glance. (u.å.). [Text]. European Commission - European Commission. Hentet 10. desember 2021, fra https://ec.europa.eu/info/food-farming-fisheries/key-policies/common-agricultural-policy/cap-glance_en

Tirado, S. B., Hirsch, C. N., & Springer, N. M. (2020). UAV-based imaging platform for monitoring maize growth throughout development. *Plant Direct*, *4*(6). https://doi.org/10.1002/pld3.230

Trevisan, R., Pérez, O., Schmitz, N., Diers, B., & Martin, N. (2020). High-Throughput Phenotyping of Soybean Maturity Using Time Series UAV Imagery and Convolutional Neural Networks. *Remote Sensing*, *12*(21), 3617. https://doi.org/10.3390/rs12213617

Turnbull, C., Lillemo, M., & Hvoslef-Eide, T. A. K. (2021). Global Regulation of Genetically Modified Crops Amid the Gene Edited Crop Boom – A Review. *Frontiers in Plant Science*, *12*, 630396. https://doi.org/10.3389/fpls.2021.630396

Watson, A., Ghosh, S., Williams, M. J., Cuddy, W. S., Simmonds, J., Rey, M.-D., Asyraf Md Hatta, M., Hinchliffe, A., Steed, A., Reynolds, D., Adamski, N. M., Breakspear, A., Korolev, A., Rayner, T., Dixon, L. E., Riaz, A., Martin, W., Ryan, M., Edwards, D., … Hickey, L. T. (2018). Speed breeding is a powerful tool to accelerate crop research and breeding. *Nature Plants*, *4*(1), 23–29. https://doi.org/10.1038/s41477-017-0083-8

Zhou, J., Yungbluth, D. C., Vong, C. N., Scaboo, A. M., & Zhou, J. (2019). <i>Estimation of maturity date of soybean breeding lines using UAV-based imagery</i> *2019 Boston, Massachusetts July 7- July 10, 2019*. 2019 Boston, Massachusetts July 7- July 10, 2019. https://doi.org/10.13031/aim.201900427

Zhou, X., Zheng, H. B., Xu, X. Q., He, J. Y., Ge, X. K., Yao, X., Cheng, T., Zhu, Y., Cao, W. X., & Tian, Y. C. (2017). Predicting grain yield in rice using multi-temporal vegetation indices from UAV-based multispectral and digital imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, *130*, 246–255. https://doi.org/10.1016/j.isprsjprs.2017.05.003