ZM009: analysis of data produced at the NPPC, Aberystwyth

Notes on objectives (as per JRA1 discussion on 18 May 2021)

How much discrimination can be seen between the genotypes using locally measured phenotypes, and how many reps would it have taken to identify them (significance level?)?

How transferrable are tools between labs/platforms?

End of June is deadline for providing data to study of cross platform integration; try and measure five traits initially for Emilie for this study.

Identify outlier data.

Build data set to accommodate ISA-Tab format.

Notes on data built in ISA-Tab format

Most data sets built during NPPC experiments is in csv format. These have been converted to .txt files for ISA-Tab entry.

The tables overleaf list the experiment and environment data sets. The main body of this document gives procedure information for derivation of digital phenotypes and environment data analysis.

Data contributors: KSW = Kevin Williams, FC = Fiona Corke, JB = Jason Brook.

Experiment outputs: growth measurements

Entity	Characteristic	Method	Units	Time series	Additional info	Original data file name	ISA-Tab file name	Contributor
Root	Mass	Ground truth	g	No	Measured at harvest Fresh & dry weight	ZM009 harvest.xlsx	ISATab_a_phe_ZM009.txt	FC
Shoot	Mass	Ground truth	g	No	Measured at harvest. Fresh & dry weight; leaf number	ZM009 harvest.xlsx		FC
Shoot	Image	Image collection		Yes	Weekly image collection (LemnaTec)	/venom/ZM009 54 plants, 7 angles, 6 dates = 2268 shoot image files		KSW
Root	Image	Image collection		Yes	Weekly image collection (Dimitrimator)	/_New_2021/SoftwareDevelopment/ImageTransfer/xx 54 plants, 4 faces, 6 dates = 1296 root image files		KSW
Column, weighing	Root column mass	Balance	g	Yes	Weekly weighing (LemnaTec)	ZM009_watering_data.csv		KSW
Root	Area at column face	Image processing	Pixels	Yes	4 column faces; total and by depth			KSW
Root	Mean area of column faces	Computed	Pixels; mm^2	Yes	Mean of 4 column faces			KSW
Root	Max area at column face	Computed	Pixels; mm^2	Yes	Maximum of 4 column faces			KSW
Root	Root depth	Computed	cm	Yes				KSW
Root	Lateral roots	Computed	Pixels; mm^2	Yes	4 column faces; total and by depth	all_dates_all_lateral_pixels.csv 20210701_ZM009_mean_lateral_by_decile.csv	root_pixels_lateral_all.txt root_ pixels_lateral_mean.txt	KSW
Root	Vertical roots	Computed	Pixels; mm^2	Yes	4 column faces; total and by depth	All_dates_all_vertical_pixels.csv 20210701_ZM009_mean_vertical_by_decile.csv	root_ pixels_vertical_all.txt root_ pixels_vertical_mean.txt	KSW
Root	Lateral:total ratio	Computed	Dimensionless	Yes	4 column faces; by depth	20210701_ZM009_mean_lateraltotal_by_decile.csv	root_ratio_lateral_total.txt	KSW
Root	Vertical:total ratio	Computed	Dimensionless	Yes	4 column faces; by depth	20210701_ZM009_mean_verticaltotal_by_decile.csv	root_ratio_vertical_total.txt	KSW
Root	Lateral:vertical ratio	Computed	Dimensionless	Yes	4 column faces; by depth	20210701_ZM009_mean_lateralvertical_by_decile.csv	root_ratio_lateral_vertical.txt	KSW
Shoot	Projected area	Image processing	Pixels	Yes	7 viewing angles			KSW
Shoot	Mean projected area	Computed	Pixels; mm^2	Yes	Mean of 7 viewing angles			KSW
Shoot	Max projected area	Computed	Pixels; mm^2	Yes	Maximum of 7 viewing angles			KSW
Shoot	Senescence deciles	Computed	Index	Yes	Calculated on max projected area image			KSW

Experiment inputs: environment data logging

Entity	Characteristic	Method	Units	Time series	Additional info	Original data file name	ISA-Tab file name	Contributor
All glasshouse data		Cambridge Glasshouse datalogger		Yes	Every 5'; 9 compartments. Air temperature, relative humidity, saturated vapour pressure, vapour pressure deficit	cambridge_data_2021-03-11_to_2021-04-27.csv	---	KSW
Solar radiation	External light level	Datalogger	?	Yes	Every 5'; outside the glasshouse	light_external_2021-03-11_to_2021-04-27.csv	light_external_data.txt	KSW
PAR	Daily PPFD	Datalogger	uE	Yes	Every 5'; compartment 2	ZM009_light_data.csv	light_internal_data.txt	JB
Air	Temperature	Datalogger	°C	Yes	Every 5'; discrete results for compartment 2 reported (c2air) plus degree.day data	c2_data_2021-03-11_to_2021-04-27.csv degree_days_c2_2021-03-11_to_2021-04-27.csv	env.data.txt degree_days.txt	KSW
Air	Relative humidity	Datalogger	%	Yes	Every 5'; discrete results for compartment 2 reported (c2rh)	c2_data_2021-03-11_to_2021-04-27.csv	env.data.txt	KSW
Air	VPD levels	Calculation	Pa	Yes	Every 5'; discrete results for compartment 2 reported (c2svp, c2vpd), as calculated from RH data; built into a periodic data format.	c2_data_2021-03-11_to_2021-04-27.csv vpd_periods_c2_2021-03-11_to_2021-04-27.csv vpd_traffic_light_daytime_compartment_2.csv vpd_traffic_light_night_time_compartment_2.csv	env.data.txt --- --- ---	KSW
Irrigation water	Quantity	Management	n/a	Yes	Applied at start of this experiment only			FC
Fertilizer	Quantity	Management	n/a	Yes	Applied at start of this experiment only			FC
Fertilizer	Type	Management	n/a	Yes	Applied at start of this experiment only			FC

Experiment details

ZM009 is a reference experiment for EPPN, primarily designed to compare root image collection and data analysis between participants.

Nine maize genotypes were grown in Compartment 2:

NPPC code	EPPN code	NPPC code	EPPN code	NPPC code	EPPN code
ZM009-01	EPPN1_L	ZM009-04	EPPN4_L	ZM009-07	EPPN3_H
ZM009-02	EPPN2_L	ZM009-05	EPPN1_H	ZM009-08	EPPN4_H
ZM009-03	EPPN3_L	ZM009-06	EPPN2_H	ZM009-09	EPPN20_T

Perspex root columns (12 × 12 × 50 cm) were filled with 2500 g of F2 compost, and watered to 3500 g, equivalent to 66.75% field capacity; no other water was added during the course of the experiment. Two seeds were sown on opposite column faces (11 March 2021), and spare plants were removed after germination.

Six replicates of each genotype were arranged in a 10 × 6 rectangle in a randomised order, together with six compost-only columns; this approach was discussed with Emilie Millet (email dated xx). The rectangle was surrounded by a further ring of 36 maize plants/columns, in an attempt to minimise edge effects, and to shroud the faces of columns under test from daylight.

Plants were grown under daylight, with supplementary LED lighting to give 14 hour days. Daytime and night time temperatures were nominally controlled to 25 °C and 20 °C respectively.

A few plants needed a physical support as the soil dried; this was provided by tying them to blue-painted bamboo canes.

Root and shoot images were captured weekly from 19 March to 26 April, and all plants were harvested on 26 April immediately after final image collection.

Dimitimator image collection parameters (roots)

The Dimitrimator rig was housed in a temporary tented structure in Compartment 2, so designed to allow ergonomic image collection for plant columns up to 1.8 m tall without incipient light spoiling the quality of these images.

The plants were grown on a waist-high bench, transferred to the rig for image collection, and then taken to the LemnaTec imaging loop for collection of shoot images (Compartment 1). On return to Compartment 2 they were rearranged in the same order on a bench on the opposite side of the rig.

The technique for handling columns for root image capture was developed over the first two rounds. During the first round it was observed that the column faces exhibited a slow build up of condensation once exposed to the light; accordingly, the open column face in the next row was left vulnerable when removing a plant from the block. In later rounds, a movable barrier was used to successfully attenuate this effect.

Based on an as yet unpublished procedure; Katie Awty-Carrol is (hopefully) writing the paper for this, and we'll get it done sooner or later. References: Dimitra's roots, Mori's 3D stuff.

Images are initially captured and stored on a Linx tablet/laptop thingy (computer details and general camera settings given in procedure: 20190628_imaging_rigs_details_procedures.docx), then downloaded to a personal hard drive. Images are relabelled to the standard NPPC directory structure format (plant_id/yyy-mm-dd/image) and copied to the 'venom' repository (note to self: currently on my PC hard drive, and I need to regain the approval to transfer to venom – somehow lost!).

Example image transfer software:	20210426_ZM009_root_image_transfer.R
Specific image collection parameters:	Nikon D3300 + Nikkor Micro 40 mm f/2.8. 6000 × 4000 pixels; 1/20"; f/8; ISO-100.

LemnaTec image collection parameters (shoots)

Configuration ZM009_00

ZM009_00_VIS_sv_000_A: focus 4000, zoom 1000, exposure 1500.

Side view images also collected at 015, 030, 045, 060, 075 & 090° rotations.

Environmental conditions

NPPC experimental conditions: glasshouse temperature data, thermal time, and VPD

Calculations are conducted using: 20191121_cambridge_sensor_data_post-Oct2015.R

Code written using: R version 3.2.2 (2015-08-14) - - "Fire Safety"

Synopsis: The Cambridge Glasshouse sensor database is interrogated to acquire temperature data between specific dates, and these data converted to thermal time (degree.days, °C.d). Relative Humidity (RH) is also recorded, and Vapour Pressure Deficit (VPD) is derived from these data.

Thermal time calculations

Generalised thermal time calculations are based on "Assessments of wheat growth to support its production and improvement, February 1998, Volume II: How to run a reference crop, Appendix 3". Note: if another model is required for a certain crop type then the software could be readily adapted to support the temperature/growth thresholds of the new model e.g.:

(i) Marshall B, Thompson R (1987) A Model of the influence of air temperature and solar radiation on the time to maturity of calabrese *Brassica oleracea* var *italica*. Annals of Botany 60: 513–519.

(ii) Robertson M, Asseng S, Kirkegaard J, Watkinson A, Holland J, et al. (2002) Environmental and genotypic control of time to flowering in canola and Indian mustard. Crop and Pasture Science 53: 793–809.

For daily calculation of thermal time, the following measurements are required:

- Base temperature (use 0°C for leaf development, and 9°C for grain development)
- Minimum temperature for the period e.g. day, hour
- Maximum temperature for the period e.g. day, hour
- Mean temperature for the period e.g. day, hour, where [mean = (max + min)/2] can be used

To start the calculation process, start from condition 1 in the table below, and go down the list until a successful match is reached. Daily thermal time is calculated as indicated. Accumulated thermal time is the sum of the daily values over the period of interest.

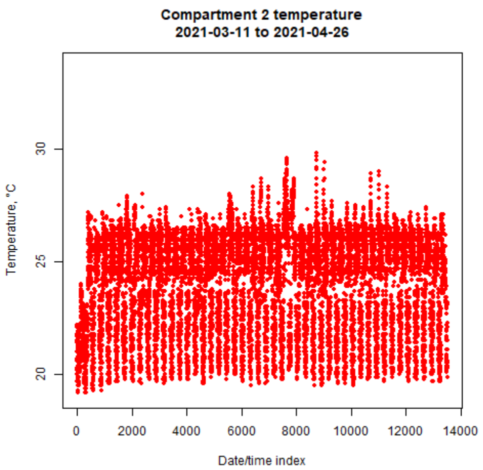
Condition		Daily thermal time
1	Base temperature is below the daily minimum	$(mean - base)$
2	Base temperature is below the daily mean	$\frac{(max - base)}{2} - \frac{(base - min)}{4}$
3	Base temperature is below the daily maximum	$\frac{(max - base)}{4}$
4	Daily maximum and minimum are lower than the base temperature	0

In theory, the NPPC glasshouse environment is designed to never go as low as 9°C (in practice, it has actually done this on one night, when all heating failed during a very frosty night!). There is therefore no advantage in using different base temperatures in calculations if all experiments are self-comparing, so all calculations have used 9°C as the base temperature. If leaf development results are needed then these can be calculated through a simple subtraction.

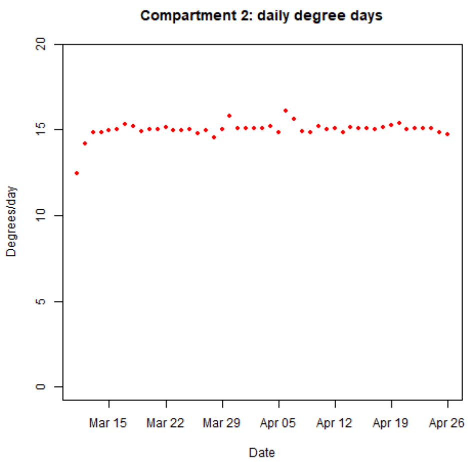
The glasshouse temperature sensors are set to record at 5 minute intervals; this in turn allows finer grain calculations of the maximum, minimum and mean of a period, and a more precise thermal time calculation than that based on a broad daily average. The code is tuned to provide outputs for hourly and daily thermal time, accumulated values for these periods, and for the total accumulated during the course of the experiment; the latter value would become more relevant within an experiment when a wider network of reliable temperature sensors is available throughout both glasshouses. Output data are supplied as csv files and as time series plots.

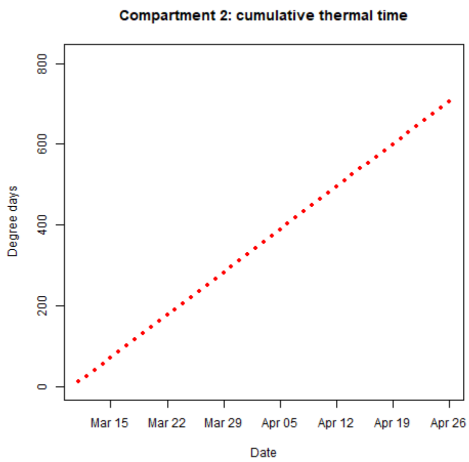
Thermal time (degree.days, °C.d) and temperature curves have been produced for the experiment, and weekly data has also been calculated:

<u>Raw data plots for temperature: Compartment 2, 11 March to 26 April 2021</u>



<u>Time series plots for thermal time: Compartment 2, 11 March to 26 April 2021</u>





VPD calculations

Ideally, humidity and temperature should be measured at the level of the plant canopy. In the NPPC compartments the Cambridge sensors are typically sited above the canopy level to prevent tangling during operation of the conveyor belts. Relative Humidity (and SVP, VPD) data is recorded every 5 minutes, and can either be used as a stream, or averaged over defined periods.

Vapour pressure deficit (VPD) is calculated as follows:

Vapour pressure deficit:	$VPD = e_a - e_s$
Saturation vapour pressure (SVP):	$e_s = 0.6108e^{17.27T/(T+237.3)}$
Actual vapour pressure:	$e_a = RH/100e_s$

So: $VPD = RH/6.108e^{17.27T/(T+237.3)} - 0.6108e^{17.27T/(T+237.3)}$

Where: RH = relative humidity (%); T = ambient temperature (°C).

Primarily, VPD data are used to check that conditions are stable and acceptable, but could also be used to model plant behaviour under environment variations. Conditions in the glasshouse compartments are largely controllable, but strong deviations in VPD can occur and unplanned combinations of heat and drought stress are therefore possible. During the warm weather in late May/early June 2020, the daytime VPD peaked between 3.5 and 4.0 kPa on several days in Compartment 6; this coincided with the drought recovery period for g07 - g12. A brief literature search suggests that VPD levels above ~2.8 kPa leads to stress events in some plants; a more exhaustive search needs to be conducted to better understand this effect. Similarly, a search needs to made to identify an accepted way of quantifying this stress; e.g. would it be acceptable to integrate the period for [VPD>limit] during the daily period? If so, what would each level represent in terms of stress? How does our traffic light system fit with accepted practice?

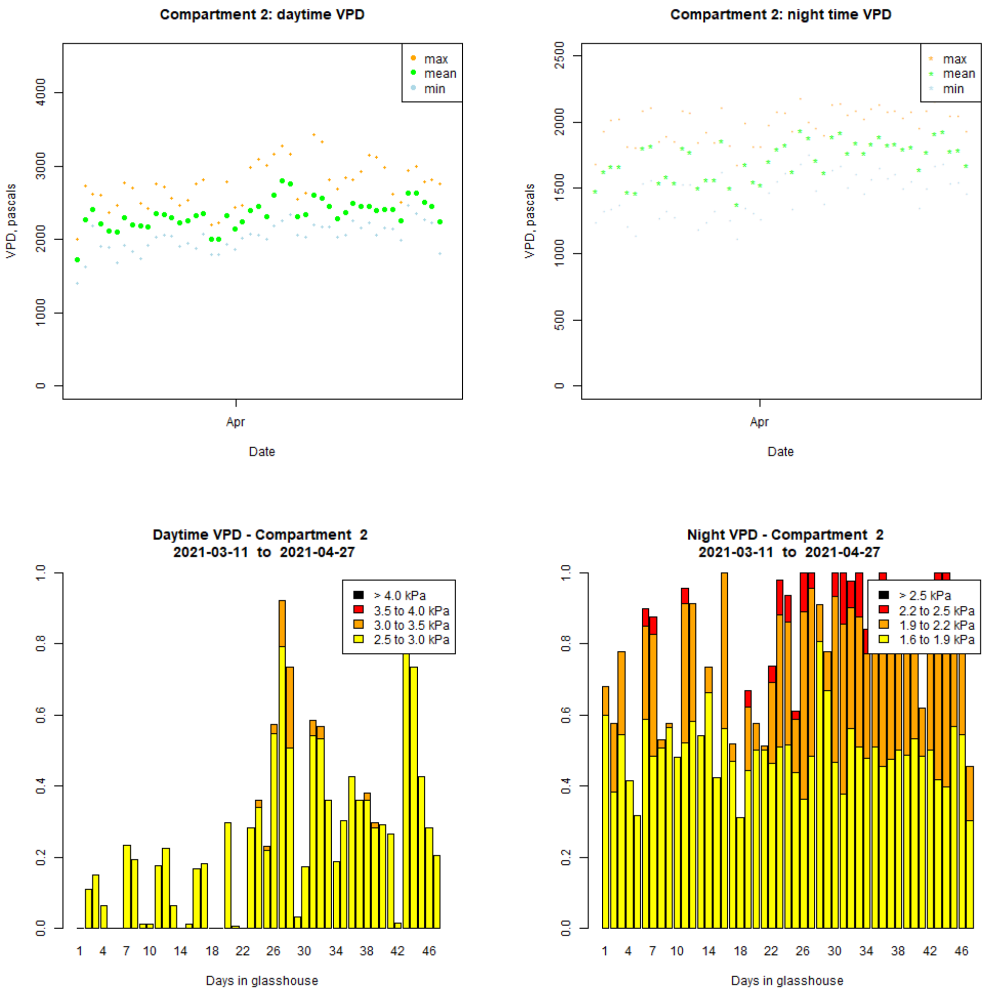
Currently, VPD is reported in two ways:

(i) minimum, maximum and mean values for each stable daytime and night time period (nominally 08:00 to 20:00, and 23:00 to 04:00)

(ii) a traffic light system, flagging up variations in stress level over each day and night time period: the proportion of time above defined thresholds is reported.

Results from method (ii) are also split into daytime (nominally 05:00 to 21:00) and night-time (21:00 to 05:00) stress indicators; thresholds are set at different levels for each of these two periods. The thresholds are arbitrary at present, and can be adjusted once important stress-dependent events are identified in either the diurnal cycle, or within the full growing period.

Time series plots for VPD data: Compartment 2, 11 March to 26 April 2021

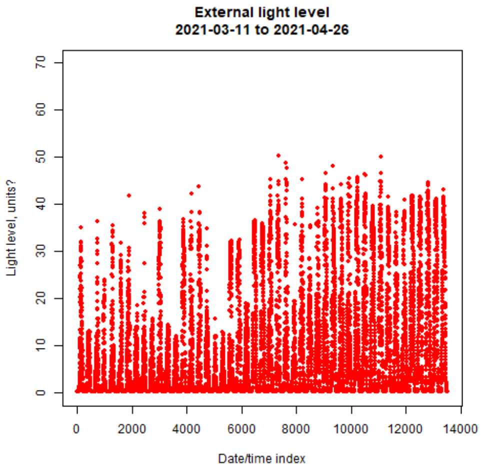


Light levels

In-compartment light meter readings were collected using a LICOR sensor (type??) downloaded using a ??. Initial data was in a .txt file format, with readings taken every second; these were converted to give a mean reading every 5 minutes using R code ('Light meter.R', written by xxx, date/version xxx). The final dataset is given in file xxx.csv.

<Put a plot of the in-compartment light curve here>

External light meter readings are provided through the Cambridge Glasshouse sensors (need to add more details here):



Mass balance of specimens: water, compost, plant materials

Each root column was weighed during the LemnaTec imaging process, enabling comparisons of water use between genotypes.

At the start of the experiment, all watered compost weights were fixed at 3500 g. During the experiment, mass changes occur due to: (i) aggregation of plant material, (ii) water loss through transpiration, (iii) water loss through evaporation from the surface of the compost. Unplanted compost containing columns were used to correct for part (iii) during the initial stages of the experiment before the effects from the other two components rendered these columns unrepresentative.

Put some curves here, with analysis

Link to dataset (csv file)

R code & Excel rendering

References to previous exps, e.g. Michele's wheat

Analysis of side view LemnaTec RGB images (shoots)

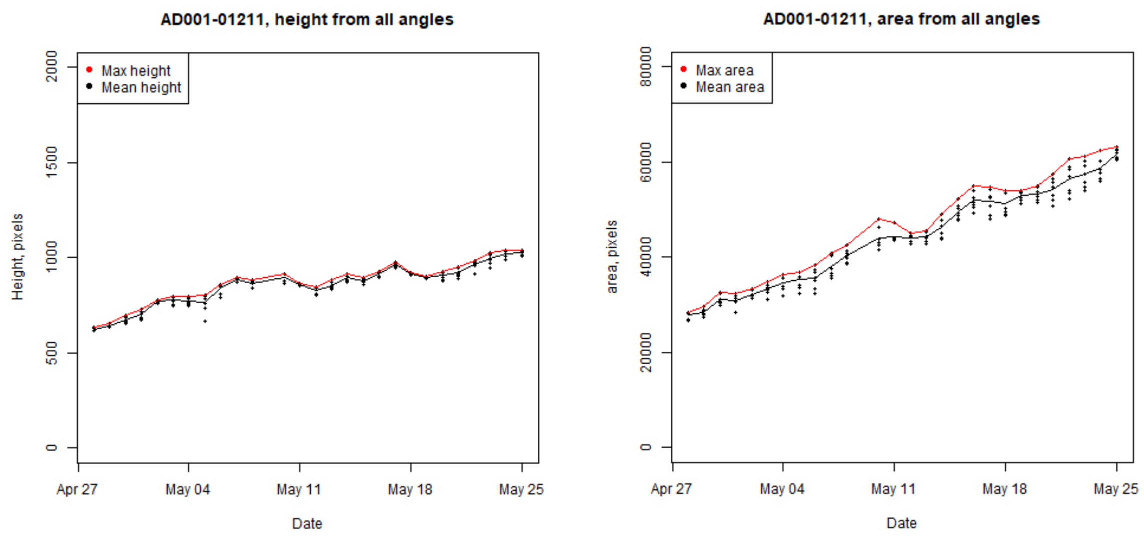
Some brief explanatory notes… camera details, parameters, wibble.

Need to include senescence (refer to the paper on Arundo, used as a starting point for this analysis) & leaf rolling data too (Wanneng's rice/miscanthus paper).

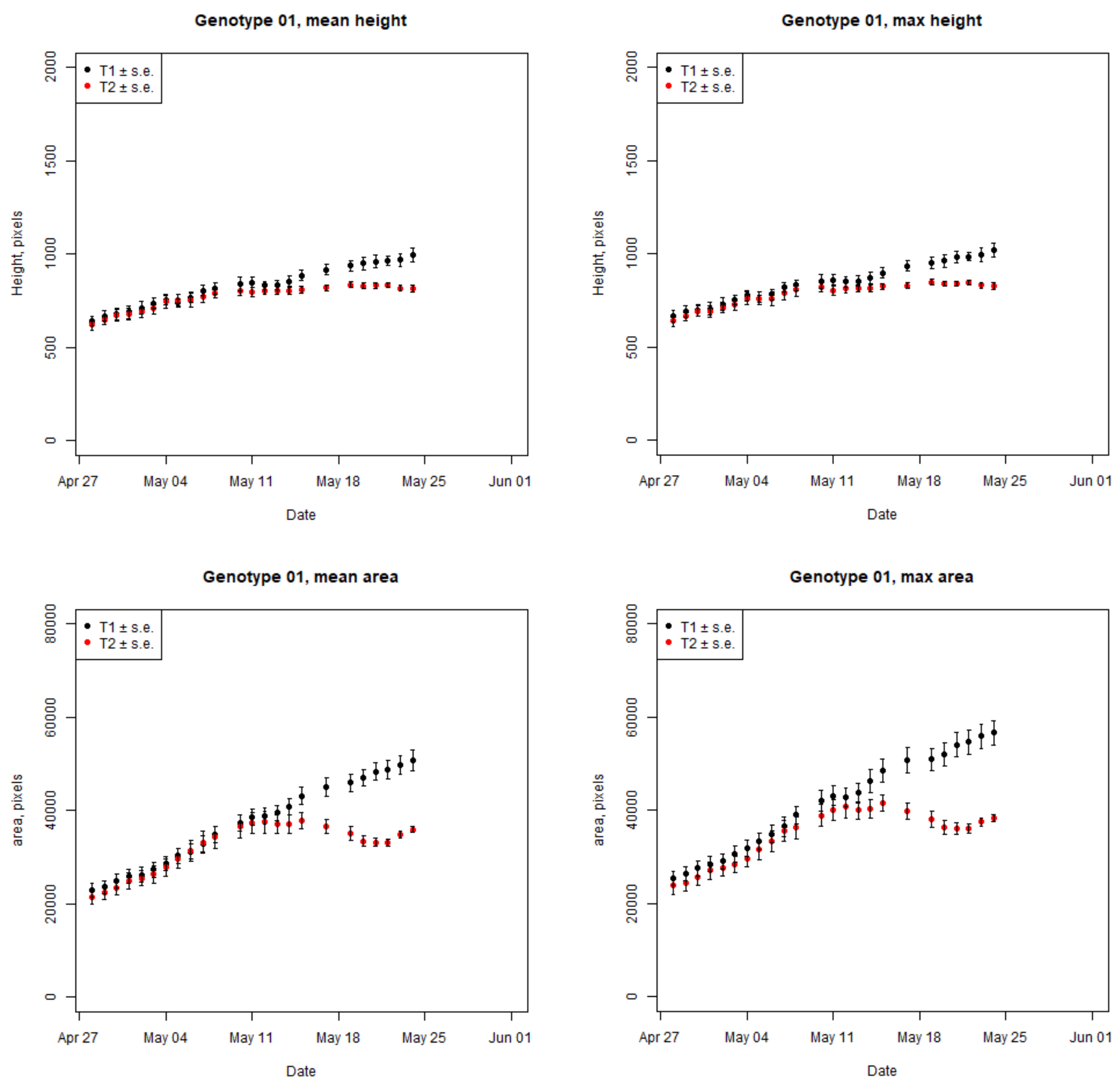
RGB side view images were captured at seven points of rotation (000 to 090 at 15° intervals) on a weekly basis. Each image was segmented to give binary masks from which plant height and projected area could be readily calculated; these were then applied to the original image to derive a separate colour image for identifying the degree of senescence.

A red:blue ratio was calculated from the respective layers in the PNG data; applying a limit to this allowed the blue stick to be identified and removed as necessary.

Plotting dynamic curves for all angles for each plant; derive mean, max values for height, area. Some placeholder curves from last year's Arundo experiment…



Combine by genotype/treatment across all reps; plot means of mean and max values.



Analysis of side view Dimitrimator RGB images (roots)

Based on previous work for Dimitra (analysis of bulk material by depth in standard columns; published) and for Piotr (identification of lateral and vertical root sections, with possible ratiometric variations as time & drought conditions build; unpublished but interesting!).

Put a dynamic image set here.

Build a pseudo 3D image for each plant/date based on the four faces. Not sure how useful it'll be, but some of them will look nice and might be informative. Might be able to calculate through compost root angles (i.e. calculating distance/angle from seed to appearance point on another face, assuming root follows a direct path)? Could pare it down to include larger root sections, or vertical, or horizontals. Don't know until we try…

Region of interest (ROI) selection

In previous root column experiments, several factors had been observed that made subsequent image analysis more difficult:

- Operator error in column selection versus plant ID
- Inconsistent orientation of column faces (labelled A, B, C, D)
- Partial eclipse of the column face from overhanging material (leaves, hands)
- Condensation
- Columns housed off-vertical when on the rig.

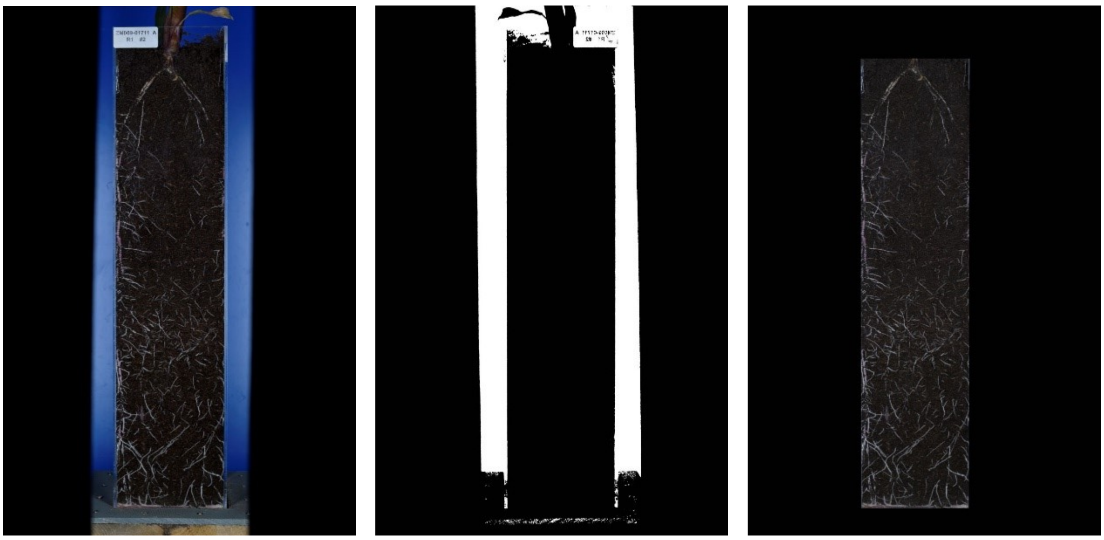
The first three of these factors are largely due to operator error, and can be improved through training and attention to detail.

The appearance of condensation is a physical problem, encountered when the local environment is conducive; this can also be attenuated somewhat (see earlier comments).

The problem of off-vertical columns is also through operational problems, but is less readily eliminated; this renders a simple ROI selection difficult, as it would need to be smaller than the true column face in order to accommodate problem images. To this end, the first step in the root segmentation process was to identify the ROI even for columns leaning more than 1-2°.

A piece of code was built in R [*20210610a_ZM009_agravitropic_roots.R*] that allowed manual selection for the top of the root structure (arbitrarily selected at the same point for all columns) and the base; this typically gave a ROI height of 5090 pixels. The column sides were identified by constructing a binary (blue-red) mask, then finding the level at which the edge was readily detected. 99 points along this edge were selected equidistantly through the middle range of the column height (500 – 4500 px), and a linear function built to describe the full edge; this function was applied to the full height of the column, with a 5 px edge reduction to reduce noise (usually caused by column distortions or scratches). The resultant ROI was typically 5090 × 1215 ± 5 px (46.3 × 11.1 cm).

For each column face, the area of ROI was calculated. There was good consistency between these data (curves shown in Appendix wibble), with the exception of two plants on one date (09 April, ZM009-01712 and ZM009-09712) where the lighting seems to have changed (possibly the curtains weren't pulled, allowing light from the glasshouse to interfere with the selection).

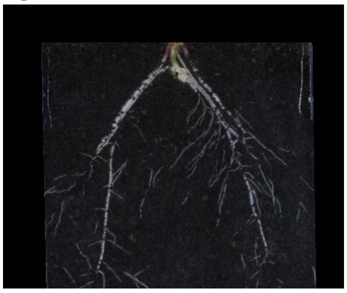


Root segmentation

Several steps were developed to identify the root structures:

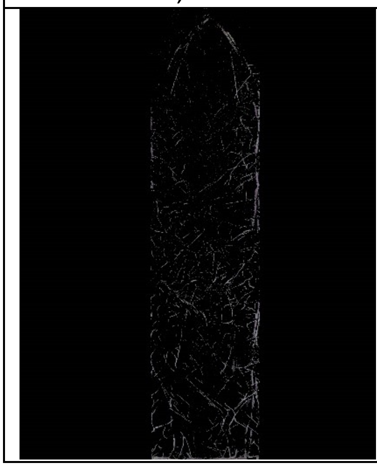
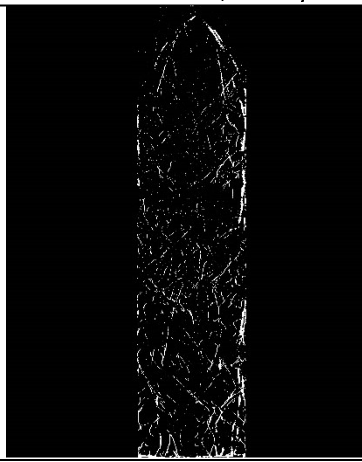
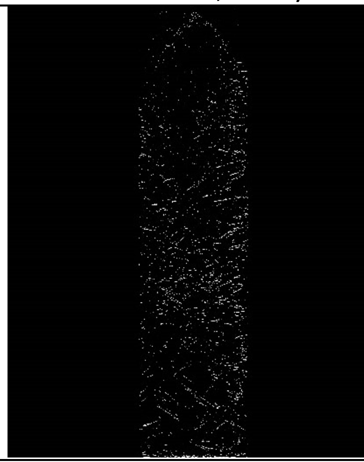
(i) A filter was developed to attenuate the appearance of condensation in the first week's images [*20210611_ZM009_root_segmentation_condensation.R*]. Comparison between image processing for week 1 and week 3 [*20210615d_ZM009_root_segmentation.R*] is given overleaf, and shows that, whilst the filter can successfully remove condensation from week 1 images, there is a reduction in visible roots if applied to images without condensation. It is accepted that data from week 1 will duly be slightly compromised.

(ii) Processed results of images for weeks 2 to 6 [*20210615d_ZM009_root_segmentation.R*] included the soil image along with the root image, to provide evidence that satisfactory segmentation had been conducted, e.g. for root structures at the top of the columns:



(iii) Processed images were subsequently trimmed to show only the region between seed and column base; these images were used to calculate the total root density on a decile depth basis [*20210621_ZM009_analysis.R*].

(iv) Trimmed images were further processed to split into vertical and lateral components [*20210628b_ZM009_classifier_mask.R*]. Three variants of this code were initially evaluated: using a 9×9 classifier, a 25×25 classifier, and using a double-pass 9×9 classifier. Both of the latter options worked well, showing individual strengths and weaknesses; the double-pass classifier was chosen for the main body of work.

All roots, true colour	Vertical roots, binary	Lateral roots, binary
		

(v) Quantification of the vertical and lateral components was conducted in a similar fashion to step (iii) [*20210629_ZM009_lateral_vertical_data.R*], and ratiometric and dynamic difference calculations were also conducted [*20210629_ZM009_curves.R*].

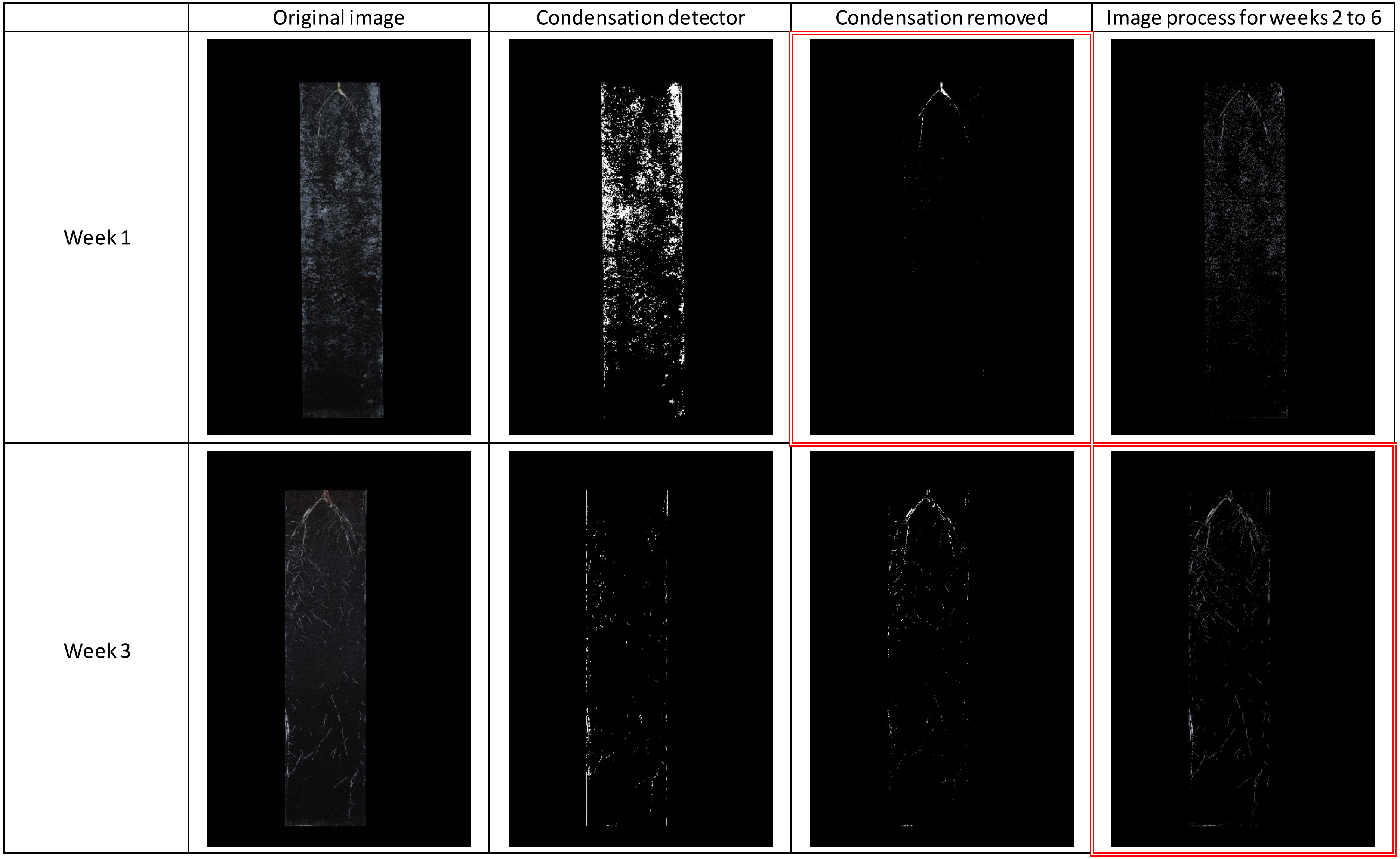


Figure: Comparison of segmentation with and without condensation filters

(vi) Results across all replicates were combined to give overall total, lateral and vertical root pixel growth curves. A csv file was produced for each column face of each set, with mean, SD & rep number data for each decile:

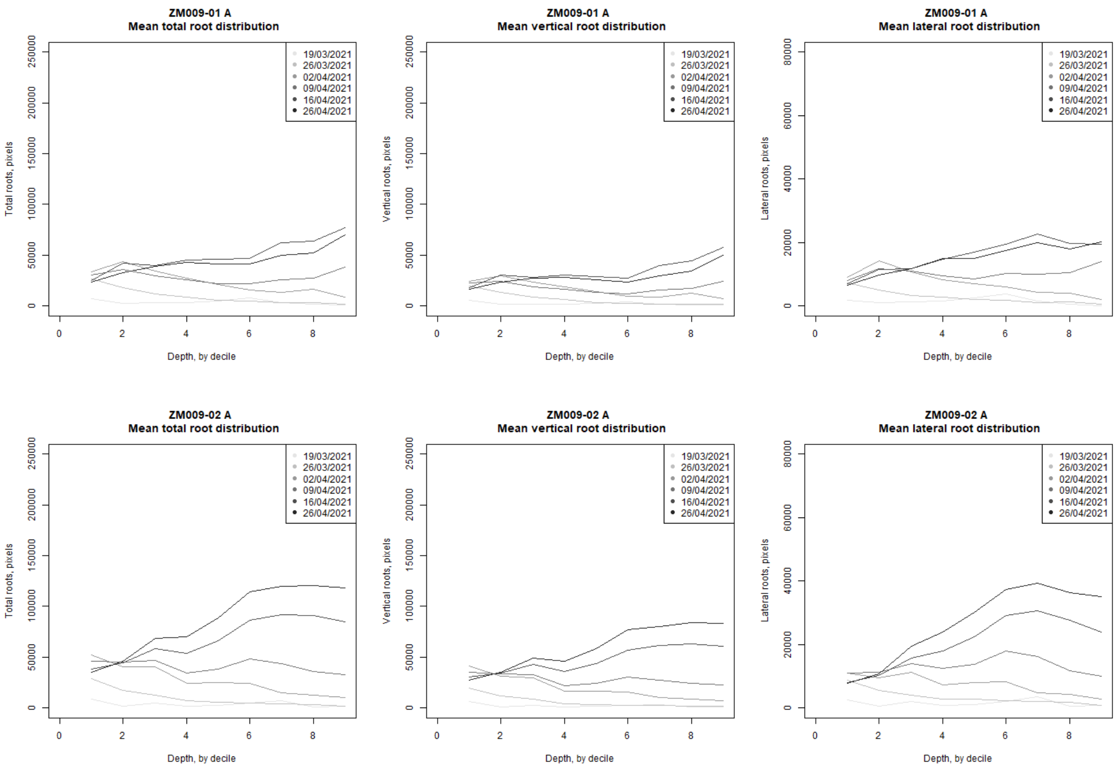
20210701_ZM009_mean_total_by_decile.csv
20210701_ZM009_mean_lateral_by_decile.csv
20210701_ZM009_mean_vertical_by_decile.csv

e.g.

plant_id	face	date	type	mean01	mean02	mean03
ZM009-01	A	19/03/2021	total	7225.167	2721.333	3009.667
ZM009-01	A	26/03/2021	total	26912.83	18053.5	11950.33
ZM009-01	A	02/04/2021	total	33494.17	44013.67	34394.17
213 lines, 34 columns						

The data for decile 10 (the base of the column) is not considered useful, as once the roots reach this depth they accumulate fairly rapidly (typically 2 – 3 weeks after sowing).

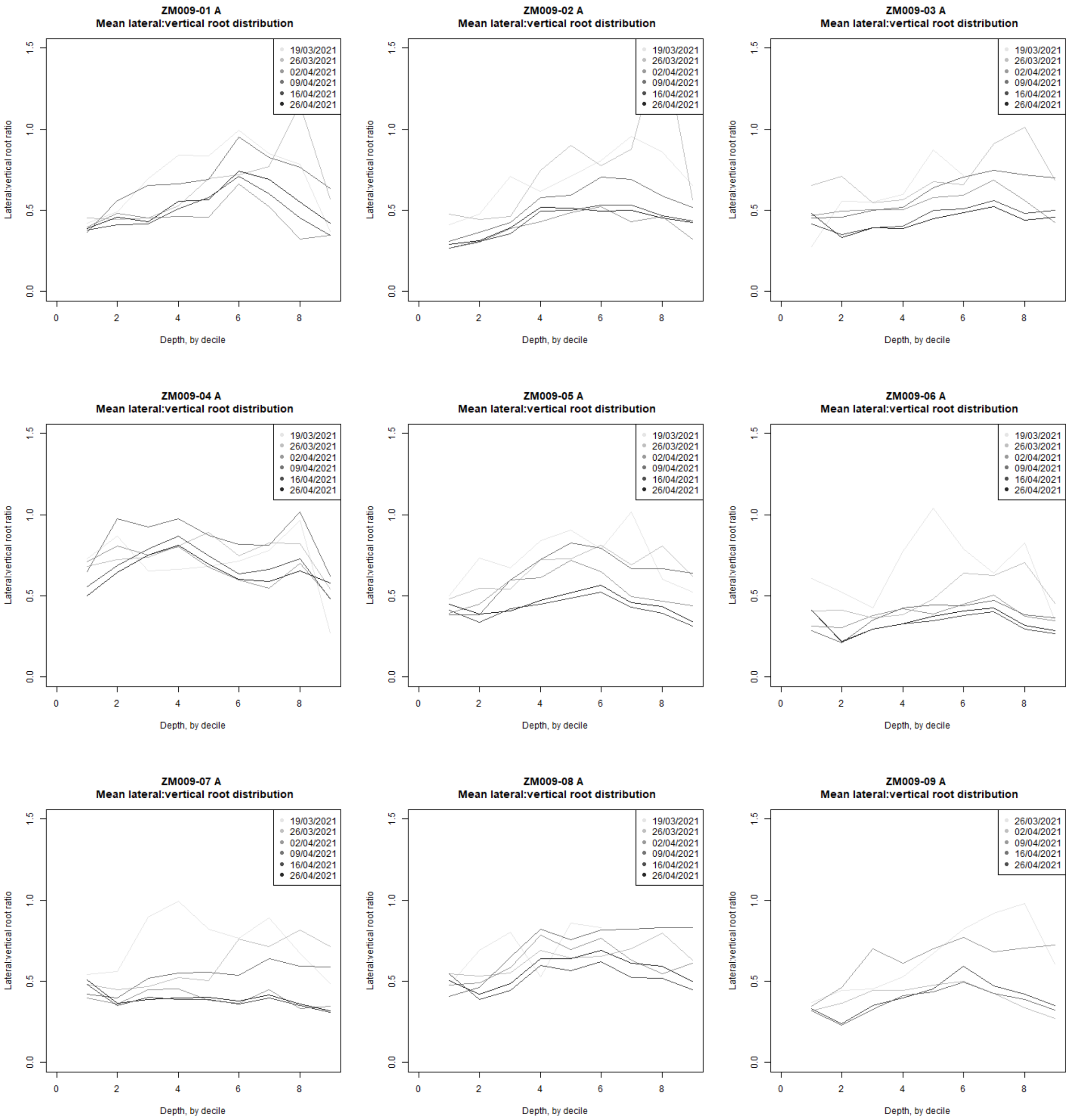
Examples for genotypes 01 & 02:



(vii) Ratio calculations between lateral, vertical, and total root pixels were similarly made for each genotype for deciles 1 to 9 (mean, sd, rep number). The examples below are of lateral:vertical root pixels count, and appear to show strong behavioural differences across the genotypes.

A csv file was produced for each ratio, genotype, face combination:

20210701_ZM009_mean_lateraltotal_by_decile.csv
20210701_ZM009_mean_lateralvertical_by_decile.csv
20210701_ZM009_mean_verticaltotal_by_decile.csv



Data presentation in ISA-TAB format

Appendices

1	Weekly mass balance curves, genotypes 01 to 09
2	Mean of mean height, by genotype
3	Mean of max height, by genotype
4	Mean of mean area, by genotype
5	Mean of max area, by genotype
6	Senescence stuff
7	Roooooooooots
8	
9	
10	
11	

Appendix wibble