PREDICTING BIOMASS AND YIELD AT HARVEST OF SALT-STRESSED TOMATO PLANTS USING UAV IMAGERY

K. Johansen [1,*], M. J. L. Morton [2], Y. Malbeteau [1], B. Aragon [1], S. Al-Mashharawi [1], M. Ziliani [1], Y. Angel [1], G. Fiene [2], S. Negrao [2,3], M. A. A. Mousa [4,5], M. A. Tester [2], M. F. McCabe [1]

[1] Water Desalination and Reuse Center, King Abdullah University of Science and Technology, Thuwal, 23955-6900, Kingdom of Saudi Arabia - (kasper.johansen, yoann.malbeteau, bruno.aragonsolorio, samir.mashharawi, matteo.ziliani, yoseline.angellopez, matthew.mccabe)@kaust.edu.sa
[2] Center for Desert Agriculture, King Abdullah University of Science and Technology, Thuwal, 23955-6900, Kingdom of Saudi Arabia - (mitchell.morton, gabriele.fiene, mark.tester)@kaust.edu.sa
[3] School of Biology and Environmental Science, University College Dublin, Belfield, Dublin 4, Ireland - sonia.negrao@ucd.ie
[4] Department of Arid Land Agriculture, Faculty of Meteorology, Environment & Arid Land Agriculture, King Abdulaziz University, Jeddah, 21589, Kingdom of Saudi Arabia - mamousa@kau.edu.sa
[5] Department of Vegetables, Faculty of Agriculture, Assiut University, 71526, Assiut, Egypt

Commission VI, WG VI/4

ABSTRACT:

Biomass and yield are important variables used for assessing agricultural production. However, these variables are difficult to estimate for individual plants at the farm scale and may be affected by abiotic stressors such as salinity. In this study, the wild tomato species, *Solanum pimpinellifolium*, was evaluated through field and UAV-based assessment of 600 control and 600 salt-treated plants. The aim of this research was to determine, if UAV-based imagery, collected one, two, four, six, seven and eight weeks before harvest could predict fresh shoot mass, tomato fruit numbers, and yield mass at harvest and if predictions varied for control and salt-treated plants. A Random Forest approach was used to model biomass and yield. The results showed that shape features such as plant area, border length, width and length had the highest importance in the random forest models. A week prior to harvest, the explained variance of fresh shoot mass, number of fruits and yield mass were 86.60%, 59.46% and 61.09%, respectively. The explained variance was reduced as a function of time to harvest. Separate models may be required for predicting yield of salt-stressed plants, whereas the prediction of yield for control plants was less affected if the model included salt-stressed plants. This research demonstrates that it is possible to predict biomass and yield of tomato plants up to four weeks prior to harvest, and potentially earlier in the absence of severe weather events.

1. INTRODUCTION

Biomass and yield are two important variables used for informed decision-making and management of agricultural production. Measurements of biomass provide information on a plant's ability to capture sunlight, water and minerals, and turn these into plant material and help determining amounts of fertilizer and irrigation of crops to be applied. Accurate yield forecasting during the growing season provides useful information for growers, allowing application of variable rates of inputs (water, fertilizer, pesticides) and logistical planning of field operations, including harvest scheduling and determining requirements for fruit picking, storage, packaging, and transportation and sales of fruit to wholesalers (Robson et al., 2017). While effects of salinity will generally reduce biomass and yield of plants, it is not well understood how salinity affects the ability to predict biomass and yield.

While field-based collection of data suitable for predicting biomass and yield at the time of harvest is very time-consuming, labor-intensive and subjective, especially for collection of time-series data that demand repetitive collection procedures, Unmanned Aerial Vehicle (UAV) based imagery can be collected efficiently and regularly (Gil-Docampo et al., 2018). Recent developments in UAV technology and miniaturized sensors provide the capability to obtain imagery at high temporal and spatial resolutions suitable for regular assessment of individual tomato plants and their properties. Some research has indicated the potential for using UAV based imagery for modelling biomass of crops. Han et al. (2019) used a number of predictor variables from UAV-based sensing of maize to model above-ground biomass and found the Random Forest model to produce the best results. The aim of this research was to assess the ability of UAV-based imagery, collected one, two, four, six, seven and eight weeks before harvest to predict fresh shoot mass, yield mass and and tomato fruit numbers at harvest and determine how predictions varied for control and salt-treated plants.

2. METHODS

2.1 Field Design and Data

Our study area covered 75 m x 75 m and was located in Hada Al-Sham, 60 km east of Jeddah, Saudi Arabia. The site is located in a tropical arid climate that receives less than 100 mm

---
* Corresponding author

of rainfall annually, and has a sandy loam soil type. A total of 1200 tomato plants were planted, consisting of 200 genotypes that included 199 *S. pimpinellifolium* accessions and one *S. lycopersicum* accession (the commercial tomato, Heinz 1706). The 1200 tomato plants were sown a month before transplantation in the field on 1-2 November 2017 and divided into two control and two salt-treated plots each with 300 plants (Figure 1). Harvest occurred between 16 and 22 January 2018 (Johansen et al., 2019).

During harvest, all fruits on each plant were manually counted and weighed. For small plants (< 1 kg shoot mass), all fruits > 3 mm in diameter were counted and weighed. For large plants (> 1 kg shoot mass), a representative subset of the whole shoot was selected, and all fruits > 3 mm in diameter were counted and weighed and this subset data was used to extrapolate overall yield by multiplying the measured yield by the ratio of the whole shoot mass and the shoot mass of the selected subset. The number of fruits ranged from 1 to 3349 per plant with an average number of 528 fruits/plant. Yield ranged from 0.1 to 1433 g per plant with an average yield of 227 g/plant. Fresh shoot mass was weighed directly at harvest for all plants and varied from 17 to 5402 g per plant with an average mass of 713 g per plant (Johansen et al., 2019).
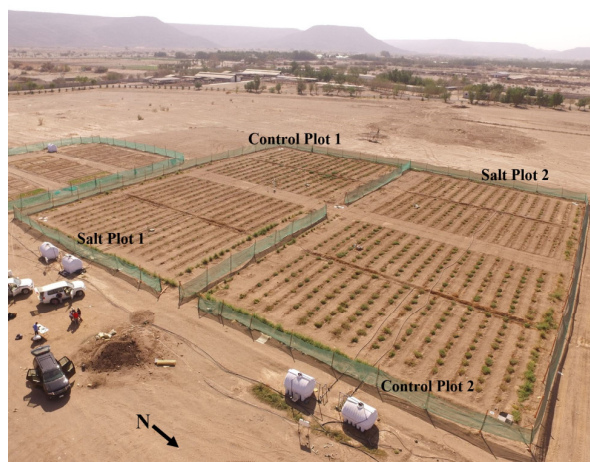

Figure 1. Study area showing the field design of the 1200 tomato plants, divided into two control and two salt-treated plots each consisting of 300 plants.

2.2 UAV Image Collection

UAV imagery was collected with a Red-Green-Blue (RGB) Zenmuse X3 sensor mounted to a DJI Matrice 100 quadcopter on 23 and 30 November, 6 and 20 December, and 7 and 14 January. All UAV image data were collected close to solar noon under clear sky conditions at a speed of 2 m/s, a flight duration of 15 minutes, a height of 13 m, and with an 82% sidelap and 78% along-track overlap. This produced a pixel size of 0.50 m. UAV-based orthomosaics, Digital Surface Models (DSM), and a Digital Terrain Models (DTM) were produced in Agisoft PhotoScan. Based on the relationship between field-derived spectrometer measurements and the orthomosaics' digital numbers of six deployed radiometric calibration panels within the mapped area, the digital numbers were converted to at-surface reflectance for the RGB imagery, using an empirical line correction (Johansen et al. 2018). A canopy height model was produced by subtracting the DTM from the DSM. The orthomosaics and canopy height models were used for object-based image analysis to delineate each individual tomato plant in each of the six UAV image data sets.

2.3 Object-Based Image Analysis

The eCognition Developer 9.3 software was used to develop a rule set for automatic delineation of the individual tomato plants. After a fine scale multiresolution segmentation, the Green-Blue vegetation index was used to merge pixels with high index values together and classify these as tomato plants. The outlines of the plants were refined using a number of loops to grow and shrink initially delineated tomato plants. All plants in the six data sets were manually inspected, and if required, manual editing of the delineation results were used to refine the outline of the plants (Johansen et al., 2019).

2.4 Extraction of Image Variables

For the delineated plants, a number of variables were selected from the RGB UAV imagery to assess their correlation with biomass and yield. The extracted variables included the three spectral bands, the Green-Red Vegetation Index, nine shape features, four gray-level co-occurrence textural measures (homogeneity, contrast, entropy, and dissimilarity) for all spectral bands and the Greed-Red Vegetation Index as well as plant area and maximum height. Linear, exponential, logarithmic and 2nd order polynomial relationships were assessed between the image derived variables and both biomass and yield. Those variables with a coefficient of determination (R^2) < 0.1 in all of the four assessed relationships were omitted from further analysis.

2.5 Random Forest Modeling and Analysis

The Random Forest machine learning approach has been applied in many ecological studies (e.g. Cutler et al., 2007; Prasad et al., 2006) and has proved to be capable of modeling complex interactions between variables and prevent overfitting. The remaining layers with an R^2 value < 0.1 were used as predictor variables in the Random Forest algorithm (mtry = 10, trees = 1000) to predict fresh shoot mass and yield in terms of number of tomatoes and their total weight per plant for both control and salt-treated plants and determine the capability of forecasting biomass and yield at harvest based on the six individual UAV data sets. The default 0.632 bootstrap approach in random forest was used for selection of the training sets at each individual tree.

Each of the six RGB image data sets were used to develop Random Forest models to predict fresh shoot mass, total fruit number and total yield mass, using (1) all plant observations, (2) only salt-treated observations, and (3) only control observations. The models for all plant observations were also applied to the subset of salt-treated plant observations and the subset of control observations to assess how separate models, adapted to either salt-treated or control plants, performed against those for all plant observations. The selected accuracy measures for each model were the percentage explained variance, the model importance of each of the predictor variables, and the root mean square error (RMSE) between the input values, i.e. field-derived observations of fresh shoot mass, total fruit numbers and total yield mass per plant, and the corresponding predicted values based on the UAV-derived information.

3. RESULTS AND DISCUSSION

3.1 Tomato Plant Delineation

For the six UAV image data sets, 94.6 - 99.1% of all plants were automatically detected, with 7-16% of the plants requiring manual adjustment of the delineation results. The need for manual adjustment of the delineation increased towards the time of harvest due to the presence of dead and senescent plants, exhibiting reflectance characteristics similar to neighbouring bare ground. Using field measured plant length and width for comparison with the automatically delineated plant area, an R^2 value of 0.85 (n = 132) with an RMSE of 0.052 m was achieved, with smaller plants slightly overestimated and larger plants slightly underestimated in length (Johansen et al., 2019). An example of the delineation results is provided in Figure 2.
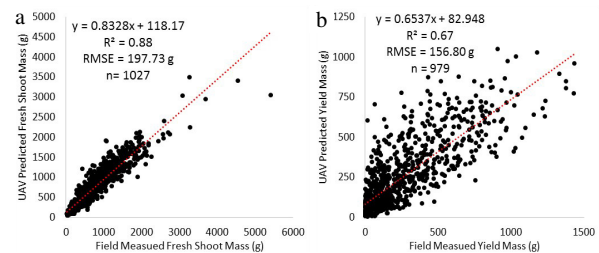

Figure 3. Scatterplots showing the correlation between field and UAV measured (a) fresh shoot mass and (b) yield mass based on UAV imagery collected on 7 January 2018.

Figure 2. Yellow outlines show the delineation results for the UAV imagery collected on 7 January 2018 for (a) the whole field site and (b) a subset of 20 tomato plants.

3.2 Predicting Biomass and Yield

Biomass and yield was predicted for each of the delineated tomato plants based on the extracted variables using the Random Forest algorithm. The week prior to harvest, the percentages of explained variance of fresh shoot mass, number of fruits and yield mass were 86.60% (Root Mean Square Error (RMSE) = 208.4 g, range = 22-5402 g, n = 1018), 59.46% (RMSE = 379.7, range = 1-3349, n = 974) and 61.09% (RMSE = 168.9 g, range 0.1-1434 g, n = 973), respectively. Two weeks prior to harvest, the RGB imagery explained a slightly higher percentage of the variance for fresh shoot mass, fruit numbers and yield mass (Figure 3), which was attributed to the removal of dead plants prior to the UAV data collection on 7 January, whereas some dead plants were present on the 14 January. Example maps of predicted fresh shoot mass and yield mass are provided in Figure 4.
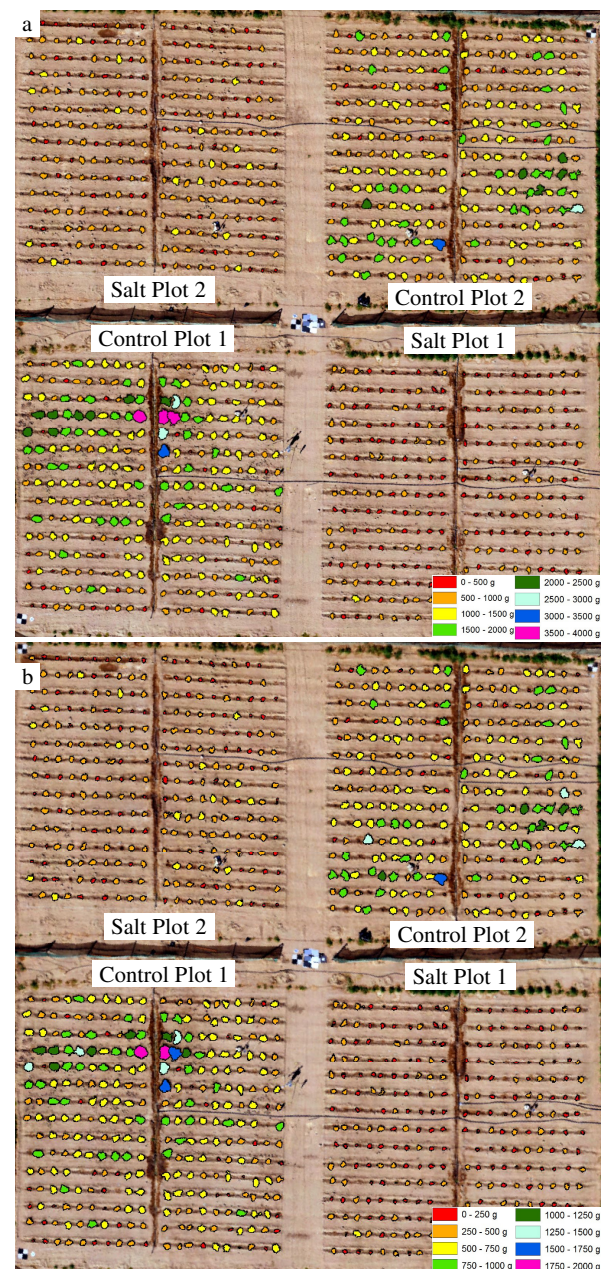

Figure 4. Prediction results of (a) fresh shoot mass and (b) yield mass based on the UAV imagery collected on 7 January 2018.

Figure 4 clearly displays the difference in fresh shoot mass and yield mass between the control and salt-treated plots. The average fresh shoot mass was 1072 and 355 g/plant and the average yield mass was 362 and 87 g/plant for the control and salt-treated plots, respectively. However, for the salt-treated plants the maximum fresh shoot mass was 746.8 g and the maximum yield mass was 1497 g. Hence, some of the tomato plant accessions may have been more salt tolerant than others. Further analysis of individual accessions in terms of biomass and yield is still to be undertaken.

On 20 December, the explained variance of fresh shoot mass, fruit number and yield mass was 79.20 (RMSE = 259.8 g), 55.90 (RMSE = 395.5) and 57.73% (RMSE = 175.7 g). A significant reduction (> 20%) in explained variance and associated increase in RMSE occurred on 6 December for all three variables, which was attributed to two destructive sandstorms, occurring on 8 and 16 December and causing damage to many plants (≈ 9%). Hence, the imagery collected after the sandstorm events provided more representative information on biomass and yield at harvest. The data collected on 30 and 23 November were found unfeasible for predicting biomass and yield at harvest.

Johansen et al. (2019) found that phenotyping individual tomato plants in terms of plant area, growth rate, condition, and plant projective cover allowed the identification of the best-performing accessions of terms of yield, with plant area being best correlated with yield. Similar findings were identified in this research, where shape features such as plant area, border length, width and length had the highest importance in the random forest models, followed by the Green-Red Vegetation Index, and the entropy texture feature extracted from a gray level co-occurrence matrix. While crop height, which is generally measurable from UAV imagery, has been identified as an important parameter for predicting crop yield (Ziliani et al., 2018), it was found to be of little importance in the Random Forest models for the six dates assessed in this research. This was attributed to the flattening of tomato plants caused by sandstorms and the fact that once tomato fruits become larger and heavier, the weight is likely to bend branches downwards, potentially reducing plant height.

In this research, the inclusion of image texture was found to improve the explained variance of fresh shoot mass, total fruit numbers and yield mass. Several of the gray-level co-occurrence texture measures derived at the individual plant level were consistently found to have high model importance as predictor variables, with R^2 values against fresh shoot mass, yield mass and fruit numbers as high as 0.67, 0.44 and 0.41, respectively. While UAV-based image textural information has been used for classification purposes of land-cover classes (Kwak and Park, 2019), our findings emphasize that future research may benefit from including image textural information for predicting crop biophysical parameters.

3.3 Control Versus Salt-Treated Plants

The Random Forest models developed just for the salt-treated plants explained a similar percentage of variance but had a lower RMSE compared with the models based on both control and salt-treated plants for fresh shoot mass. However, for fruit numbers and yield mass between 6 December and 14 January, the explained variance was higher (up to 5.28%) and the RMSE lower when using the Random Forest models specifically developed for only the salt-treated plants. For the control plants,

the explained variance was similar (within 1.1%) when using the models for only the control plants compared with the ones for all plants, emphasizing that separate models may be required for predicting yield of salt-stressed plants, whereas the prediction of yield for control plants is less affected if the model includes salt-stressed plants.

4. CONCLUSIONS

This research demonstrates that it is possible to predict biomass and yield of tomato plants up to four weeks prior to harvest, and potentially earlier in the absence of severe weather events. This type of information is valuable input for growers seeking to optimize yield, manage plant growth and prepare for harvesting, sale and distribution of tomatoes. Subsequent to UAV image data collection, the processing steps, including the orthomosaic and canopy height model generation, the object-based plant delineation (with some manual editing required for some plants), the extraction of image based information used as predictor variables, and the Random Forest prediction of biomass and yield, were automated. Hence, predictions can be provided to growers throughout the growing season and well ahead of harvest. Future work should assess if the developed prediction models can be applied to different growing seasons and for different varieties of tomato plants.

REFERENCES

Cutler, D.R., Edwards, T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., Lawler, J.J. 2007. Random forests for classification in ecology. *Ecology*, 88(11), 2783-2792. doi.org/10.1890/07-0539.1.

Gil-Docampo, M.L., Arza-Garcia, M., Ortiz-Sanz, J., Martinez-Rodriguez, S., Marcos-Robles, J.L., Sanchez-Sastre, L.F. 2018. Above-ground biomass estimation of arable crops using UAV-based SfM photogrammetry. *Geocarto International*, doi.org/10.1080/10106049.2018.1552322.

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, 10. doi.org/10.1186/s13007-019-0394-z.

Johansen, K., Raharjo, T., McCabe, M.F. 2018. Using multi-spectral UAV imagery to extract tree crop structural properties and assess pruning effects. *Remote Sensing*, 10(6), 854. doi.org/10.3390/rs10060854.

Johansen, K., Morton, M.J.L., Malbeteau, Y.M., Aragon, B., Al-Mashharawi, S.K., Ziliani, M.G., Angel, Y., Fiene, G.M., Negrao, S.S.C., Mousa, M.A.A., Tester, M.A., McCabe, M.F. 2019. Unmanned Aerial Vehicle-based phenotyping using morphometric and spectral analysis can quantify responses of wild tomato plants to salinity stress. *Frontiers in Plant Science*, 10, 370. doi.org/10.3389/fpls.2019.00370.

Kwak, G.-H., Park, N.-W. 2019. Impact of texture information on crop classification with machine learning and UAV images. *Applied Sciences*, 9, 643. doi.org/10.3390/app9040643.

Prasad, A.M., Iverson, L.R., Liaw, A. 2006. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. *Ecosystems*, 9, 181-199. doi.org/10.1007/s10021-005-0054-1.

Robson, A., Rahman, M.M., Muir, J. 2017. Using WorldView satellite imagery to map yield in avocado (*Persea americana*): A case study in Bundaberg, Australia. Remote Sensing, 9(12), 1223. doi.org/10.3390/rs9121223.

Ziliani, M.G., Parkes, S.D., Hoteit, I., McCabe, M.F. 2018. Intre-season crop height variability at commercial farm scales using a fixed-wing UAV. *Remote Sensing*, 10(12), 2007. doi.org/10.3390/rs10122007.