

## ■ Application of sequential statistical data-mining methods to determine tissue nutrient concentrations related to yield of the 'Hass' avocado

S. Campisi-Pinto<sup>1,2</sup>, Y. Zheng<sup>1</sup>, P. Rolshausen<sup>1</sup>, D. Crowley<sup>2</sup>, B. Faber<sup>3</sup>, G. Bender<sup>4</sup>, M. Bianchi<sup>5</sup>, T. Khuong<sup>1</sup> and C. Lovatt<sup>1</sup>

<sup>1</sup>. Department of Botany and Plant Sciences, University of California, Riverside, CA 92521-0124

<sup>2</sup>. Department of Environmental Sciences, University of California, Riverside, CA 92521

<sup>3</sup>. University of California Cooperative Extension Ventura County, 669 County Square Drive, Suite 100, Ventura, CA 93003

<sup>4</sup>. University of California Cooperative Extension San Diego County, 5555 Overland Ave., Suite B4, San Diego, CA 92123

<sup>5</sup>. University of California Cooperative Extension San Luis Obispo County, 2156 Sierra Way, Suite C, San Luis Obispo, CA 93401

Plant tissue analysis completed sufficiently early in a growing season can identify nutritional problems in time to take corrective actions and safeguard commercial yield of perennial fruit crops. Leaf analysis is an established method for precision agriculture, however, with reference to 'Hass' avocado, leaf analysis may not provide adequate guidance for optimal yield, fruit size or fruit quality during the current or following season. For these reasons tissues other than leaves were collected and analyzed. The objective was to estimate the strength of the associations between yield and concentrations of N, P, K, S, Ca, Mg, Fe, B, Zn, Mn and Cu in the following tissues: cauliflower and full bloom stage inflorescences collected in March and April, respectively; pedicels collected monthly from fruit set through harvest; and spring flush leaves collected in September (industry standard). Samples were collected from six commercial 'Hass' avocado orchards located in the northern and southern production areas of California for two consecutive crops. Sequential analysis of the data by statistical methods that included least squares multiple linear regression, principal component analysis and factor analysis provided evidence that cauliflower stage inflorescences may provide the best predictive and descriptive models for expected yield. Within this tissue, yield was significantly related to the concentrations of Mg, Cu, K, S and N ( $R = 0.7075$ ;  $P = 0.0002$ ). Envelope analysis was used to estimate the concentration ranges of each nutrient (% dry mass) in cauliflower stage inflorescences related to yields  $> 40$  kg/tree:  $0.18 < \text{Mg} < 0.27$ ;  $0.76 \times 10^{-3} < \text{Cu} < 2.15 \times 10^{-3}$ ;  $1.70 < \text{K} < 2.15$ ;  $0.25 < \text{S} < 0.32$ ; and  $2.77 < \text{N} < 3.60$ . The utility of these values remain to be tested in the field.

Keywords: Cauliflower and full bloom stage inflorescences, Fruit pedicels, Leaves, Envelope analysis, *Persea americana* Mill., Target nutrient concentration ranges for improved yields, Tree nutrient status.

### INTRODUCTION

Plant tissue analysis completed sufficiently early in a growing season can identify current season nutritional problems in time to take corrective action before yield parameters, including fruit size and quality, are reduced (Snyder, 1998). Plant tissue analysis is especially popular because tissue nutrient concentrations reflect which nutrients present in the soil are actually taken up by the plant. Differences between the results of plant tissue analyses and soil analysis can indicate unfavorable soil characteristics (pH, cation exchange, salinity) or the occurrence of other factors preventing the uptake of nutrients by the roots (low or high temperatures, insufficient or excess water, or root damage caused by disease, insect and nematode pests or mechanical injury).

At the present time, the California 'Hass' avocado industry relies on nutrient analysis of leaves collected in August to September from the current year's nonfruiting spring vegetative shoots (Embleton & Jones, 1972). The leaves are 4 to 6 months old and collected too late in the season to correct nutrient deficiencies impacting flower and fruit retention and early fruit development (fruit size and quality). In California, 'Hass' avocado leaf analysis was developed to guide replacement fertilization to support the development of the following year's crop (Embleton *et al.*, 1959). However, leaves collected according to this standard protocol have not proven sufficiently sensitive to detect changes in tree nutrient status that cause changes in yield, fruit size or fruit quality during the current or following year (Embleton *et al.*, 1968; Lahav, 1995; Lovatt & Witney, 2001).

For many commercial crops, tissues other than leaves, e.g., flowers, leaf petioles, pedicels (stems of individual flowers and fruit), and young developing fruit, may provide viable alternatives for diagnosing nutrient deficiencies. The results of several studies suggested that nutrient concentrations in some of these tissues might be more explicative than leaf nutrient concentrations (Sanz & Carrera, 1994; Razeto & Granger, 2001; Razeto & Salgado, 2004).

In the context of the current study, the concentration of thirteen plant essential nutrients including nitrogen (N), phosphorus (P), potassium (K), sulfur (S), calcium (Ca), magnesium (Mg), iron (Fe), boron (B), zinc (Zn), manganese (Mn), sodium (Na), chloride (Cl), and copper (Cu) were monitored in a number of plant tissues including: cauliflower and full bloom stage inflorescences collected in March and April, respectively; pedicels collected monthly from fruit set through harvest, and spring flush leaves collected in September, according to the industry standard. Samples were collected in six commercial 'Hass' avocado orchards located in the major northern and southern avocado-growing areas of California during two successive crop years. The objective of the study was to assess the strength of statistical relationships between tissue nutrient concentrations and expected yield (kg/tree). In particular, a first objective was to assess and benchmark the predictive power of nutrient concentrations for the different plant tissues in the study. A second objective was to identify the more explicative nutrients within selected plant tissues.

## MATERIALS AND METHODS

The research was conducted in six commercially producing 'Hass' avocado (*Persea americana* Mill.) orchards (with different but known rootstocks) located in the major southern and northern avocado-growing areas of California: Pauma Valley [lat. 33° 20' 34" N, long. 117° 01' 20" W, elevation 200 m above sea level (asl)], Irvine [lat. 33° 40' 40" N, long. 117° 49' 23" W, elevation 23 m asl], Somis [lat. 34° 15' 26" N, long. 118° 59' 46" W, elevation 93 m asl], Santa Paula (2 sites) [lat. 34° 21' 15" N, long. 119° 03' 29" W, elevation 85 m asl], and San Luis Obispo [lat. 35° 16' 58" N, long. 120° 39' 31" W, elevation 71 m asl]. All trees were in good health with no visible signs of nutrient deficiencies or pest problems. The orchards were managed according to each grower's standard cultural practices. Samples of 'Hass' avocado tissues were collected as follows: entire inflorescence at the cauliflower stage (~March) and at full bloom (~April); pedicels of young fruit in late June-early July (which is just prior to exponential fruit growth and June drop of the setting crop, start of mature fruit drop, and transition from vegetative to reproductive growth), and at the end of the fall vegetative shoot flush (~November); pedicels were collected the following year from the now mature fruit when the trees reached the cauliflower stage of inflorescence development (~March) and at full bloom (~April); and 40 leaves from nonfruiting terminal spring vegetative shoots were collected according to the standard for the industry in September of each year (Embleton *et al.*, 1959). Samples were collected at a height of 1.5 m above ground from the four quadrants (NE, SE, SW and NW) of each data tree. Samples were collected from 16 individual trees (replications) on the diagonal across each orchard for two crop years. In addition, in Somis, the seven tissue samples were collected from trees in two separate fertilization experiments for N (9 treatments replicated on 16 individual trees) and N-P-K (8 treatments replicated on 16 individual trees), respectively. Samples were immediately placed in brown paper bags and stored on ice, taken to the University of California (UC), Riverside, washed thoroughly with dish soap, rinsed three times with distilled deionized H<sub>2</sub>O, oven-dried at 60 °C for 72 h, and ground in a Wiley mill to pass through a 40-mesh (0.635-mm) screen (Embleton *et al.*, 1959). The ground samples were sent to the UC-Division of Agriculture and Natural Resources Laboratory for analysis. Total N was determined after combustion at 1050 °C by thermal conductivity (Leco Corp. St. Joseph, Mich.) (AOAC, 2006). The concentrations of P, K, S, Ca, Mg, Fe, B, Zn, Mn and Cu were determined after nitric acid-hydrogen peroxide microwave digestion by inductively coupled plasma atomic emission spectrometry (ICP-AES) (Meyer & Keliher, 1992).

Trees were harvested at legal maturity (20.8% dry mass content) between March and July, according to each grower's standard management practice, to determine total kilograms of fruit per tree. The mass of 100 randomly selected individual fruit per tree was used to calculate fruit size distribution per tree. These data were used to calculate the kilograms of fruit of each packing carton size per tree. The following packing carton fruit sizes (grams per fruit) were used: size 84 (99 to 134 g), size 70 (135 to 177 g), size 60 (178 to 212 g), size 48 (213 to 269 g), size 40 (270 to 325 g), size 36 (326 to 354 g), and size 32 (355 to 397 g).

Statistical methods, including quartiles trimming, least-square multiple linear regressions, principle component analysis, factor analysis, and QR decomposition, were used for a series of statistical tests on more than three thousands samples collected in multiple orchards over multiple years. A knowledge discovery approach (KDD) was used to integrate various statistical methods into an organic approach (Maimon & Rokach, 2005; Trevor *et al.*, 2009). In particular, expected yield was used as a criterion of multiple linear regression of nutrient concentrations. In that context, a stepwise strategy was employed and nonsignificant data were removed at each step of the analysis on the basis of statistical tests ( $P > 0.05$ ). The final subset of nutrients was subjected to envelope analysis (Behrenfeld and Falkowski, 1997, Raveh *et al.* 2009).

## RESULTS

Comparison of Pearson correlation coefficients indicated that the cauliflower stage inflorescence provided a greater number of significant correlations between nutrient concentrations and expected total yield and yield of commercially valuable size fruit (kg/tree) than the other tissues included in the research (data not shown). Based on this result, the cauliflower stage inflorescence was selected as the most promising tissue for further analysis. Cauliflower stage inflorescence tissue concentrations of N, P, K, S, Ca, Mg, Fe, B, Zn, Mn, and Cu (11 of the nutrients analyzed) were used to predict yield. There was strong linear correlation between the observed yield (x-axis) and the yield predicted (y-axis) using the 11 nutrients and all trees in the data set (no yield to maximum yield per tree) ( $R = 0.88912$ ;  $P < 0.0001$ ) (Fig. 1, Model-1). The strength of the relationship between the concentrations of the 11 nutrients in cauliflower stage inflorescences and yield was improved by including only trees in the data set that produced 8 to 125 kg per tree ( $R = 0.9377$ ;  $P = 0.0064$ ) (Fig. 1, Model-2).

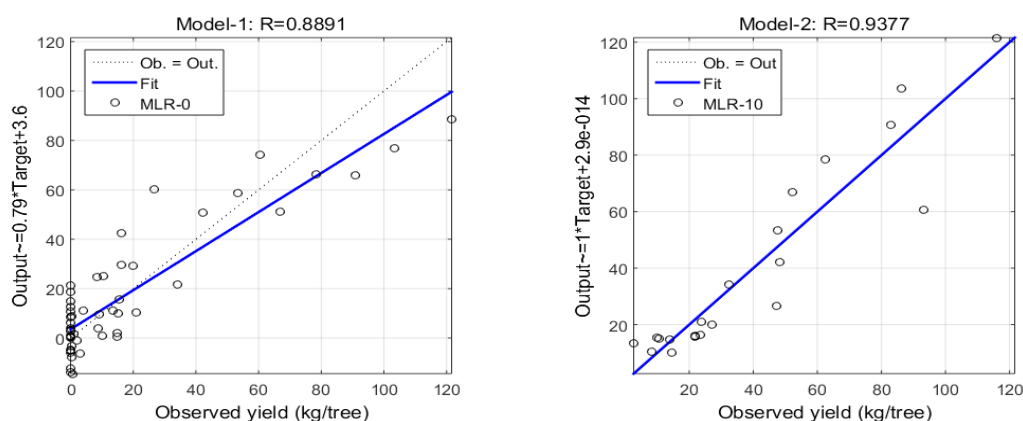
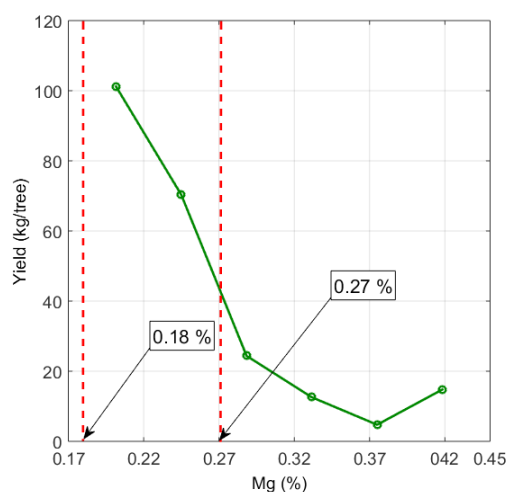


Fig. 1. The linear correlation between the observed yield (x-axis) and the yield predicted (y-axis) using the concentrations of 11 nutrients in the cauliflower stage inflorescence by Model-1 (left panel), which used all trees in the data set (no yield-maximum yield per tree), and by Model-2 (right panel), which used only trees that produced 8 to 125 kg/tree.

Subsequent analyses provided evidence that within the cauliflower stage inflorescence, linear correlations with expected yield were higher for the subset N, P, K, Mg, S, Zn, than Cu, Ca, B, Mn and Fe. Thus, least square multiple linear regression analysis with N, P, K, Mg, S, Zn, and Cu was used to predict expected yield (Model-3). As anticipated, by reducing the number of predictors from eleven to seven, some degree of predictive accuracy was lost. In particular, R decreased from 0.9377 for Model-2, using 11 nutrients (Fig. 1), to 0.7576 for Model-3, using only 7 nutrients ( $P = 0.0006$ ); both models used only trees with yields from 8 to 125 kg per tree. The strength of the estimated coefficients for each nutrient included in Model-3, i.e., the magnitude of its effect, was defined by the absolute value of its  $\beta$ -score and t-statistic and the corresponding P-value. Thus, the seven cauliflower stage nutrients included in Model-3 were ranked in decreasing order of their effect on yield (strongest to weakest):  $Mg > Cu > K > S > N > P > Zn$ . Model-4 was developed using the five nutrients with the strongest relationships to yield (Mg, Cu, K, S and N) ( $R = 0.7075$ ;  $P = 0.0002$ ).

Envelope plotting and analysis was the last step in the knowledge discovery process. In this analysis, yields were plotted as a function of the nutrient concentrations of cauliflower stage inflorescences. The results provided boundaries related to optimal nutrient concentrations in cauliflower stage inflorescences related to yield. Envelope plots for Mg, Cu, K, S, N, P, and Zn indicated that greater yields, as well as lower yields, were associated with lower nutrient concentrations in this tissue, whereas greater nutrient concentrations in cauliflower stage inflorescences were always and only associated with lower yields. For example, using 40 kg per tree as a reasonable threshold between satisfactory yield (any yield above 40 kg/tree) and unsatisfactory yield (any yield below 40 kg/tree), Mg concentrations in cauliflower stage inflorescences greater than 0.18% but less than 0.27% were likely to be supportive of above average yield, though low yielding trees were also found within this range (Fig. 2). In contrast, Mg concentrations greater than 0.27% were only associated with yields below 40 kg/tree. Similar interpretations of the envelope analysis for the other nutrients included in Model-4 provided an estimate of the nutrient concentration ranges (% dry mass) in cauliflower stage inflorescences related to yields  $> 40$  kg/tree:  $0.18 < Mg < 0.27$ ;  $0.76 \times 10^{-3} < Cu < 2.15 \times 10^{-3}$ ;  $1.70 < K < 2.15$ ;  $0.25 < S < 0.32$ ; and  $2.77 < N < 3.60$ . The ranking order reported for Model-3 that identified the nutrients used in Model-4 combined with the information derived from envelope analysis provides some statistical support for decision making in terms of target nutrient concentrations in cauliflower stage inflorescences as indicators of tree nutrient status associated with greater yields.



**Fig. 2. Envelope analysis of yield plotted as a function of the Mg concentration (% dry mass) of cauliflower stage inflorescences. The red dotted lines represent the lower and upper boundaries for the range of nutrient concentration associated with yields  $> 40$  kg/tree; the lower and upper nutrient concentrations for the optimum range are reported in the black box. The green line is the envelope, which depicts the relationship between yield and the concentration of Mg in cauliflower stage inflorescence tissue.**

## DISCUSSION

Avocado orchards are heterogeneous entities, in which biomass and commercial fruit production vary greatly depending on a number of factors including, but not limited to, soil pH, soil depth and density, organic matter content, water quality, irrigation and drainage patterns, fertilization, pest management and pruning. Climatic conditions, with particular reference to maximum and minimum temperatures in irrigated orchards, and in addition precipitation under rain fed conditions, can be overriding factors affecting production. Avocado orchards also experience pronounced temporal fluctuations in yield reflected in irregular or alternate bearing. Despite the great amount of environmentally and managerially induced variation at the orchard level, significant relationships between expected yield and the concentrations of essential nutrients in cauliflower stage inflorescences were demonstrated. The combination of statistical tools used in this research provided evidence that the nutrient concentrations of cauliflower stage inflorescences functioned well in predictive and descriptive models for expected yields. Within this tissue, yield was significantly related to the concentrations of Mg, Cu, K, S, and N ( $R = 0.7075$ ;  $P = 0.0002$ ) (Model-4), with the following ranges (% dry mass) associated with yields  $> 40$  kg/tree:  $0.18 < Mg < 0.27$ ;  $0.76 \times 10^{-3} < Cu < 2.15 \times 10^{-3}$ ;  $1.70 < K < 2.15$ ;  $0.25 < S < 0.32$ ; and  $2.77 < N < 3.60$ . The utility of these values in California, or in other avocado-growing areas of the world, remain to be tested in the field. Envelope analysis provides a prescription necessary but not sufficient for improvement of avocado yield under all conditions. “Prescription” refers to those concentrations that have the potential to bring about the desired yield, but the large number of unpredictable variables affecting any particular orchard in any given year (for instance an adverse climatic event) could potentially override the model’s predictive power.

## CONCLUSION

The results of this research provided evidence that for the 'Hass' avocado, nutrient concentrations of cauliflower stage inflorescences might prove to be a valuable tool for assessing tree nutrient status sufficiently early in the season that corrective action could be taken to improve flower and fruit retention and fruit growth to increase total yield and yield of commercially valuable size fruit (178-325 g/fruit). The cauliflower stage inflorescence provides the advantage that concentrations of specific nutrients in this tissue are strongly related to yield, which was not the case for other tissues analyzed in this research or for the standard leaf analysis used in California. For commercial 'Hass' avocado orchards in California, the nutrient concentration ranges (% dry mass) in cauliflower stage inflorescences more likely to be associated with greater yields (> 40kg/tree) are:  $0.18 < \text{Mg} < 0.27$ ;  $0.76 \times 10^{-3} < \text{Cu} < 2.15 \times 10^{-3}$ ;  $1.70 < \text{K} < 2.15$ ;  $0.25 < \text{S} < 0.32$ ;  $2.77 < \text{N} < 3.60$ . ( $R = 0.7075$ ;  $P = 0.0002$ ).

## REFERENCES

- AOAC Official Method 972.43, Microchemical determination of carbon, hydrogen, and nitrogen, Automated method, in Official Methods of Analysis of AOAC International, 18th edition, Revision 1, 2006. Chapter 12, pp. 5-6, AOAC International, Gaithersburg, MD.
- Behrenfeld, MJ & Falkowski PG. 1997. A consumer's guide to phytoplankton primary productivity models. *Limnology and Oceanography* Vol. 42, no. 7, pp. 1479-1491.
- Embleton, TW & Jones WW. 1972. Development of nitrogen fertilizer programs for California avocados. *Calif. Avocado Soc. Yrbk.* Vol. 56, pp. 90-96.
- Embleton, TW, Jones, WW & Garber MJ. 1959. Leaf analysis as a guide to nitrogen fertilization of the Hass avocado. *Calif. Avocado Soc. Yrbk.* Vol. 43, pp. 94-95.
- Embleton, TW, Jones, WW, Garber MJ, & Boswell SB. 1968. Nitrogen fertilization of the Hass avocado. *Calif. Avocado Soc. Yrbk.* Vol.52, pp. 131-134.
- Lahav, E 1995, Avocado nutrition – A review. *Proc. III World Avocado Congr.* Vol. 1, pp. 143-159.
- Lovatt, CJ & Witney, G. 2001. Nitrogen fertilization of the 'Hass' avocado in California. *AvoRes.* Vol. 1, no. 3, pp. 1-4,11.
- Maimon, O & Rokach, L. 2005. Decomposition methodology for knowledge discovery and data mining: Theory and applications. Series in Machine Perception and Artificial Intelligence. Vol. 61, 323 p. World Scientific Publishing.
- Meyer, GA & Keliher PN. 1992. An overview of analysis by inductively coupled plasma-atomic emission spectrometry. pp. 473-516. In: A. Montaser and D.W. Golightly (ed.). *Inductively coupled plasmas in analytical atomic spectrometry.* VCH Publishers, New York, NY.
- Razeto, B. & Granger C. 2001. Análisis químico del pedúnculo del fruto y la inflorescencia, posibles herramientas de diagnóstico nutricional en palto (*Persea americana* Mill.), p. 73. In: Resúmenes. 52° Congreso Agronómico de Chile, Quillota, Chile.
- Razeto, B & Salgado, J. 2004. The influence of fruit peduncle as indicators of nitrogen status of the avocado tree. *HortScience* Vol. 39, no. 6, pp. 1173-1174.
- Raveh, E., Hilman, Z., Aharon, M., and Peres, M. 2009. Optimal leaf mineral values for citrus. *Et-Hadar*, 64; 24-27.
- Sanz, M & Carrera, M. 1994. The possibility of using floral analysis to diagnose the nutritional status of pear trees. *Acta Hort.* Vol. 367, pp. 290-295.
- Snyder, CS. 1998. Plant tissue analysis – A valuable nutrient management tool. *News & Views* pp. 1-3.
- Trevor, H, Tibshirani, R, & Friedman J. 2009. The elements of statistical learning: Data mining, inference, and prediction. Series in Statistics, Springer-Verlag, 763 p.





# ACTAS • PROCEEDINGS

## VIII CONGRESO MUNDIAL DE LA PALTA 2015

del 13 al 18 de Septiembre. Lima, Perú 2015

[www.wacperu2015.com](http://www.wacperu2015.com)

