Yield mapping methods for manually harvested crops

A. F. Colaço[1], R. G. Trevisan[1], F. H. S. Karp[1], J. P. Molin[1]
[1]*Biosystems Engineering Department, University of São Paulo. Av. Pádua Dias 11, 13418-900, Piracicaba, São Paulo, Brazil.*
andre.colaco@usp.br

Abstract

During the harvest of citrus and other fruit crops, fruits are placed into bags across the field before they are loaded into trucks. The location of bags can be georeferenced for yield mapping purposes. Several alternatives are possible for processing bag location data to produce the final yield map. The objective of this study was to demonstrate and test the accuracy of different data processing methods for yield mapping in manually harvested crops. Two main types of data processing and variations of these were studied. The first type calculates yield at each point by dividing the mass of the bag by its coverage area in the field. The second type is based on the distribution and density of points across the field. The proposed methods were tested over orange bags location data and also over a modeled yield map. All methods showed similar yield variation patterns, but with different level of detail and accuracy. Methods that calculate yield at every bag location got the highest correlation (0.7 of R^2) and lowest average error (15 %) among the evaluated methods. These methods were considered suitable to produce yield maps and support further site-specific management actions.

Introduction

The yield map is probably the most important information in a successful precision agriculture system. Yet, for manually harvested crops like many of the horticultural species, yield mapping is still not much adopted. This might be one of the reasons why precision agriculture is still at a low rate of adoption in this kind of crop. Yield mapping is more common in crops that are mechanically harvested because yield monitors are usually available. On the other hand, crops that are not mechanically harvested demand yield mapping methods that suit the manually harvest procedure. This is the case of the citrus crop in Brazil, which is the largest orange juice producer in the world (around 730,000 ha grown – FAO, 2012). Citrus growers and the machinery industry are still struggling to find the ways to harvest mechanically. Precision agriculture is not yet much adopted and yield mapping is not common among growers. Most of fruit and vegetable crops are still harvested manually and also have difficulties to produce yield maps.

There are many methods available for yield mapping in manually harvested crops. Ye et al. (2007) presented a method to estimate citrus yield based on hyperspectral aerial imagery which relates canopy reflectance to plant yield (R^2 of 0.84 between predicted and actual yield was achieved). Okamoto and Lee (2009), Bansal et al. (2012) and Gong et al. (2013) used ground images of orange trees to count the number of fruits using different image processing algorithms (up to 90 % accuracy was achieved, Gong et al. 2013). Zaman et al. (2006) also predicted orange yield by measuring canopy volume and its relation with yield (over 90 % accuracy was achieved). These methods can be carried out prior to the harvest providing not only the yield map but also a harvest forecast. Besides, they are suited for either, manual or mechanical harvesting. Although

high accuracy might be achieved, these types of technology are often expensive or not available for growers.
A simpler type of yield mapping for orange groves, exclusively for manual harvest, was first presented by Whitney et al. (1999). It is based on georeferencing the bins that are used during harvest. In a typical manual harvest of citrus, fruits are briefly stored into bins, bags, boxes or containers which are placed close to the harvested trees across the field. Afterwards, a truck mounted crane picks each bin and discharges it into a trailer. For yield mapping, the bins must be georeferenced with a GNSS receiver. Yield is calculated based on either the distribution of these points in the field or on the coverage area of each bin. This type of method has been reported in Colaço and Molin (2014). Coordinates can be registered automatically with a yield monitor at the loading truck (Schueller et al., 1999; Whitney et al., 2001; Tumbo et al., 2002), by post processing the loading truck GPS track (Colaço et al. 2013), or manually with a common handheld GNSS receiver. Many fruit and vegetable crops that are harvested manually adopt bins or containers to briefly store the harvested product in the field before they are loaded. Therefore, georeferencing these bins are a viable solution for yield mapping in many different crops.
Different solutions have been used for collecting the position of bags throughout the field, but little research has been carried out regarding the data processing steps to obtain the final yield map from the bag position data. Several alternatives are possible for processing this data, but the accuracy of each method is still unknown. Therefore, the objective of this work was to demonstrate and compare different data processing methods for yield mapping in manually harvested crops, as well as to verify the accuracy of each method.

Material and Methods

<u>Citrus manual harvest</u>

During the harvest, the oranges are picked manually and placed into “big bags”. The pickers usually harvest a strip of orange trees ranging from two up to six rows wide. The bags are alligned at the center row of the harvested strip (Figure 1). The mass of each bag is estimated by the harvest team leader before it is discharged into the loading truck. The total mass of each bag ranges around 540 kg.

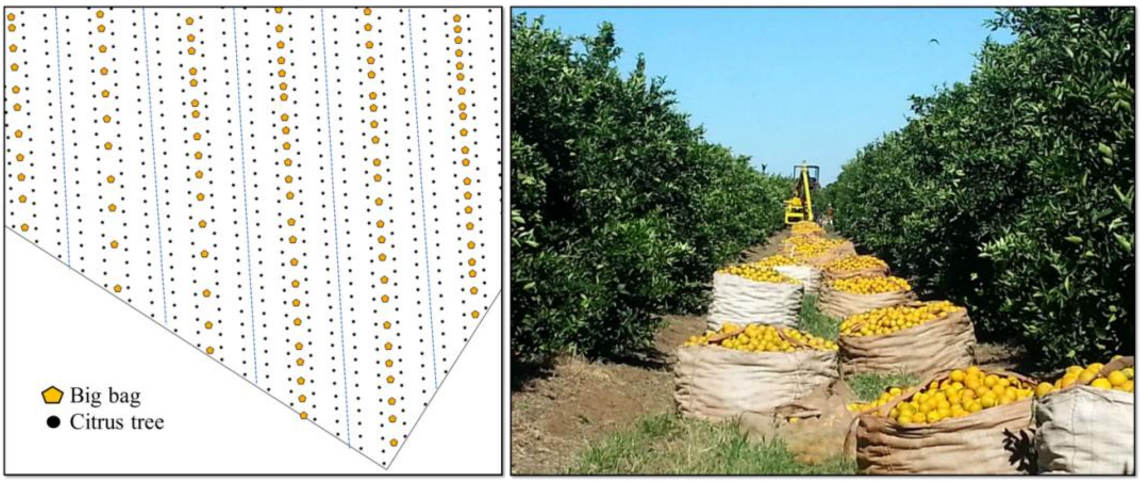

Figure 1: Harvest strip of four tree rows and the placement of bags at the center row

Yield mapping methods

The proposed methods start from the same data set, which is the geographic coordinates of the bags used during the manual harvest. They can be divided into two types (1.i and 2.i). The first type calculates yield at each bag location, based on the mass and coverage area of the bag. The second calculates yield in a given area based on the number of bags found within a searching field. Variation within each method was tested giving a total of six different methods.

Method 1.1 – rectangular bag coverage area

The first method calculates yield at each point based on the mass of each bag and its coverage area in the field – the total mass of the bag is divided by its corresponding area and given in Mg ha^{-1}. The coverage area of each bag is considered as a rectangle. The larger side of the rectangle is the harvest strip width (W). The smaller side is given by the sum of the half distances between one point and its neighbors (d_1 and d_2) (Figure 2). In this method, it is considered that bags are perfectly alligned, although in reality they might not be in a perfect straight line. To generate the final yield map (10 x 10 m pixel), the yield points are interpolated using the inverse distance method.

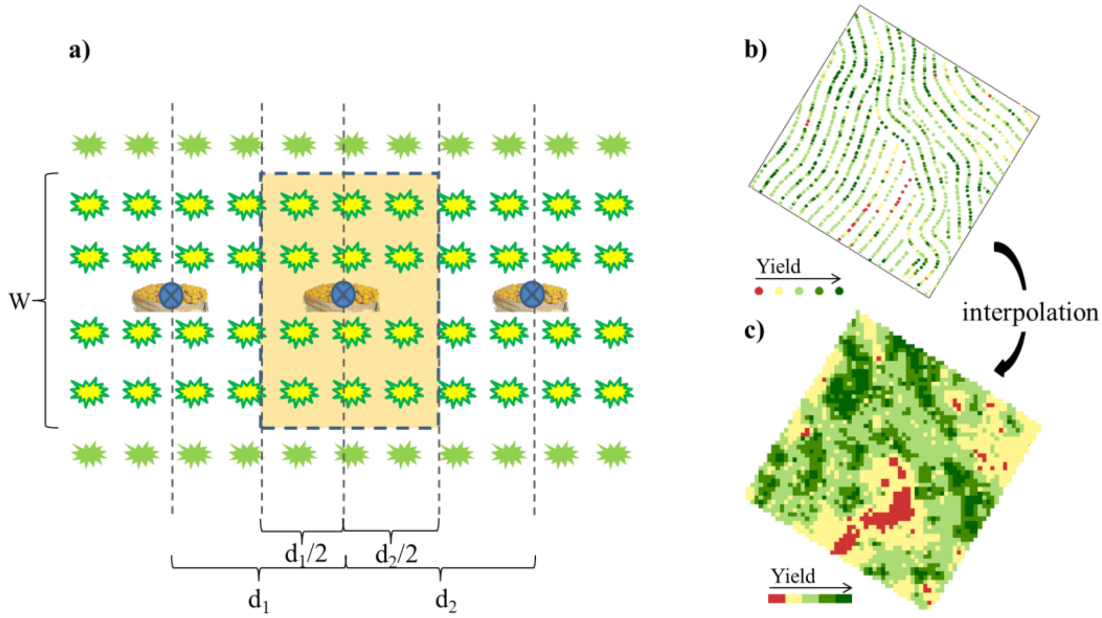

Figure 2: (a)The representative area of bags calculated from method 1.1; (b)yield points; (c)final yield map after interpolation

Method 1.2 – Voronoi *polygon*

The second method is similar to the first one but instead of considering the coverage bag area as a rectangle, it is represented by an unregularly shaped polygon called a Voronoi polygon. To generate such a polygon, a geometry tool available in QGIS 2.4 (Open Source Geospatial Foundation, Beaverton, USA) software was used. The Voronoi polygon tool divides the field into smaller areas, each corresponding to the coverage area of one point (Figure 3).

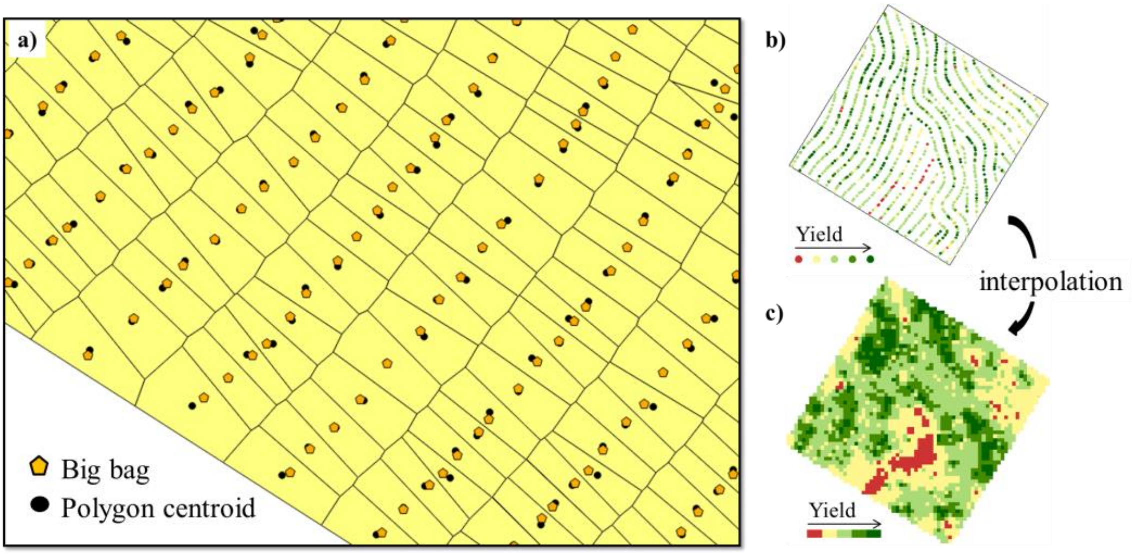

Figure 3: (a)Coverage area of each bag given by the Voronoi polygon; (b)yield points; (c)final yield map after interpolation

Yield is calculated by dividing the mass of each bag by the polygon's area. A centroid point at each polygon is created, which will represent the calculated yield data for that polygon. The final yield map (10 x 10 m pixel) is given by interpolating the yield points using the inverse distance method.

Method 2.1.1 and 2.1.2 – counting bags within grid cells

The third and fourth methods divide the field into squared grid cells of 2500 m² (method 2.1.1) and 625 m² (method 2.1.2) (Figure 4). The number of bags inside each cell is counted. The total mass of fruit is divided by the cell's area and yield is given in Mg ha^{-1}. These methods do not use interpolation to generate the final yield map as the entire area of each cell presents its respective average yield value.

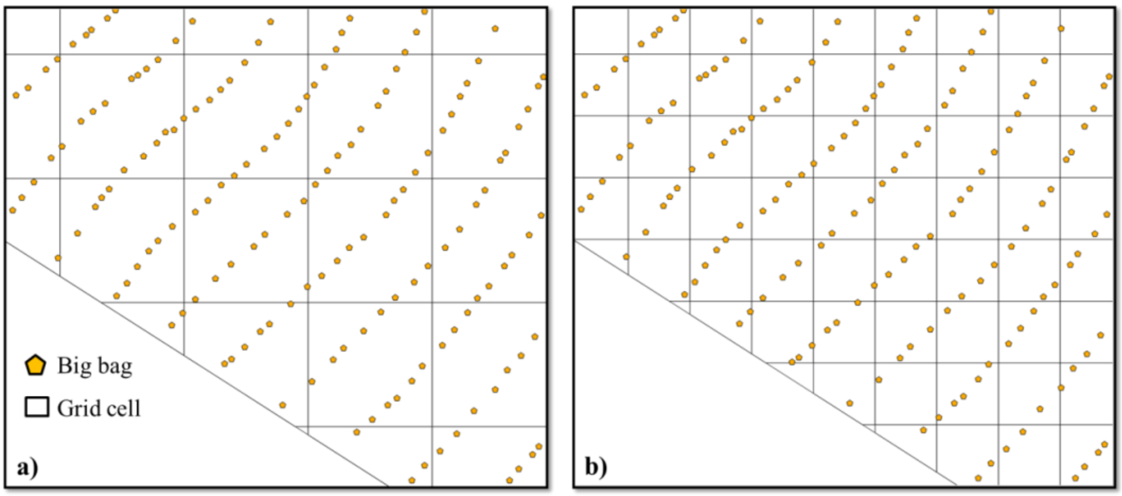

Figure 4: Grid cells used on methods (a) 2.1.1 and (b) 2.1.2

Methods 2.2.1 and 2.2.2 – counting bags within a given radius (heat map)

The fifth and sixth methods are based on a GIS tool called "heat map". It searches the number of bags within a given radius – 15 m (method 2.2.1) and 30 m (method 2.2.2). The searching field is centered at every pixel (10 x 10 m) of the final yield map, so that each bag will be counted more than once (Figure 5). Different weight is given to the bags inside the radius according to its proximity to the center of the searching field (higher weight for closest points). Yield is calculated at each pixel based on the weighed mass of fruit and the size of the searching area. The final yield map presents pixels of 10 x 10 m like those from methods 1.i, but no interpolation is used since the yield value is given at all pixels.

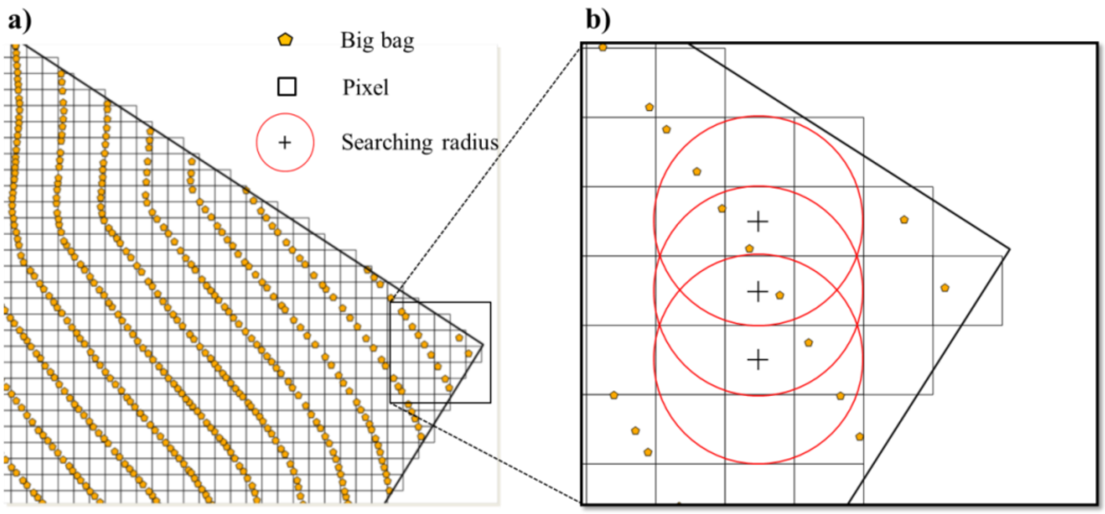

Figure 5: (a)Big bags locations and pixels; (b)searching radius centered at each pixel

Example case study
To demonstrate and test the proposed methods, they were carried out over data from one orange field of approximately 25 ha. The total number of bags was 1439. The descriptive statistics and a visual comparison between maps were carried out over the final yield maps.

Accuracy assessment
To verify the accuracy of each method, a modeled yield map was created and used as a reference. The six methods were carried out over the modeled bag data and compared with actual yield from the reference map.
To create the modeled yield map, a map of the tree locations was generated following a tree spacing of 4 x 7 m. The fruit production of each tree followed a spatial distribution given by a pre-defined semivariogram. This procedure resulted in the reference yield map. To obtain the location of the bags, a mathematical routine was created as follows. The width of the harvest strip was defined as four tree rows. The mass of fruit from each tree is iteratively summed until it reaches the total mass of one bag (540 kg). At this point, the location of the bag is assigned. All the modeling steps were carried out using the R 3.1.2 (R Foundation for Statistical Computing, Vienna, Austria) and QGIS 2.4 software. A correlation analysis and the error calculation were carried out between the final result from each method and the actual modeled yield from each tree.

Results and Discussion

Results from the yield mapping procedures over one 25 ha citrus block showed similar average yield among different methods (Table 1). Higher coefficients of variation (CV) were found for the methods based on the grid cell searching field, mainly because of the high yield range, that went from zero (no bags found at a certain grid cell) up to 255 Mg ha^{-1} – which is unusual for an orange grove.

Table 1: Descriptive statistics from yield mapping methods carried out over one citrus field.

	Rectangular bag coverage	Voronoi	Heat map (15 m)	Heat map (30 m)	Grid cell (25 m)	Grid cell (50 m)
Count	2501	2501	2501	2501	446	122
	------------------------------ Mg ha^{-1} ------------------------------					
Mean	25.40	24.75	24.53	24.53	23.99	25.40
Minimum	14.70	8.97	16.88	18.07	0.00	0.00
Maximum	37.35	39.78	40.99	35.20	255.88	255.88
Range	22.65	30.81	24.11	17.13	255.88	255.88
Standard Deviation	3.31	4.99	3.39	2.88	14.48	22.35
	------------------------------ % ------------------------------					
Coefficient of Variation	13.02	20.17	13.82	11.73	60.35	88.01

All methods showed visually similar variability patterns (Figure 6). The first two methods showed high similarity. Between them, the Voronoi method showed higher variability (higher range and CV, Table 1). In this type of data processing, the calculation of yield is carried out at every bag location so it is more affected by the position of each bag, which is defined by the worker during the harvest. Methods 2.i showed much smoother maps, especially the heat map with 30 m radius of searching field. In these cases, the actual location of each bag is decreases in importance as the searching area for bags is increased. Methods based on the squared grid cells were the easiest and less complex data processing methods. They revealed yield variability but are unable to show smooth variability since the pixels are larger. Oversimplifying the yield map (like the 50 m grid cell), might affect the map accuracy but can help during variable rate applications that follow a prescription map based on this kind of yield map.

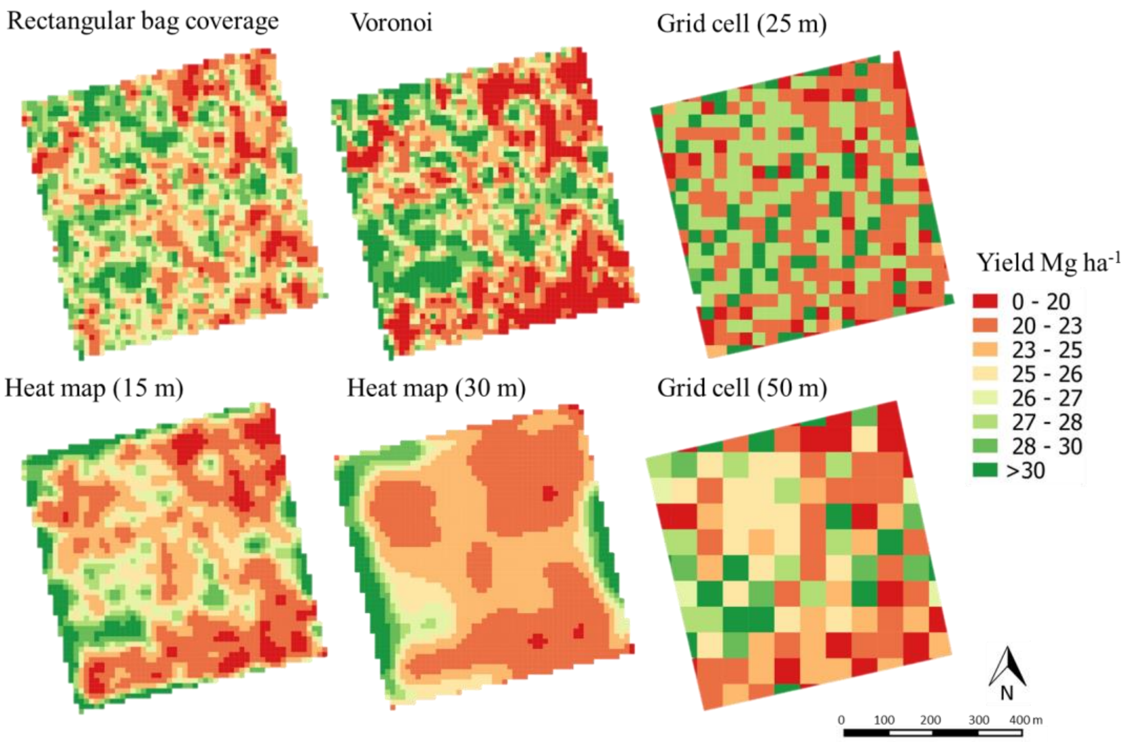

Figure 6: Orange yield maps from different processing methods

The performance and accuracy of the proposed methods was assessed by comparing their results against a reference yield map. The descriptive statistics from the yield mapping methods over the modeled bag location data showed that average yields were close among different methods and also similar to the reference yield map (Table 2). Higher CV was found for the grid cell-based maps. The other methods showed CV values lower than the one from the reference yield map, which means that none of them captured the actual level of variability from the reference yield map.

Table 2: Descriptive statistics from reference yield map and yield mapping methods carried out over modeled bag location

	Reference yield map	Rectangular bag coverage	Voronoi	Heat map (15 m)	Heat map (30 m)	Grid cell (25 m)	Grid cell (50 m)
Count	8450	2360	2360	2360	2360	416	113
	------------------ Mg ha^{-1} ------------------						
Mean	34.85	35.09	35.35	34.94	34.94	33.93	35.85
Minimum	3.28	14.55	16.54	16.26	21.30	0.00	10.80
Maximum	64.21	53.63	54.11	67.75	51.28	108.01	108.01
Range	60.93	39.08	37.57	51.49	29.97	108.01	97.21
Standard Deviation	8.92	6.17	6.49	6.40	5.48	16.12	12.16
	------------------ % ------------------						
Coefficient of Variation	25.59	17.58	18.36	18.31	15.68	47.51	33.91

From the plotted yield maps, it can be seen that all methods (Figure 7) showed spatial variability patterns similar to the reference map (Figure 8). Methods 1.1 and 1.2 presented good performance since they reproduced the existing yield variability with greater details. The two methods also got the highest correlation with actual yield (R^2 = 0.7), and lowest average error (around 15%) (Table 3). The heat map with 15 m of searching radius was the third best data processing method. The methods based on the grid cell and also the heat map with 30 m of searching radius are not recommended to be adopted. Their final result inheres greater error and does not represent detailed spatial variability like the first tree methods.

This result indicates that mapping yield based on bag locations has accuracy limitations. However, they can provide enough information for assessing yield variability enabling further site-specific management. Practitioners can use the yield maps to either delimit management zones or create variable rate prescriptions.

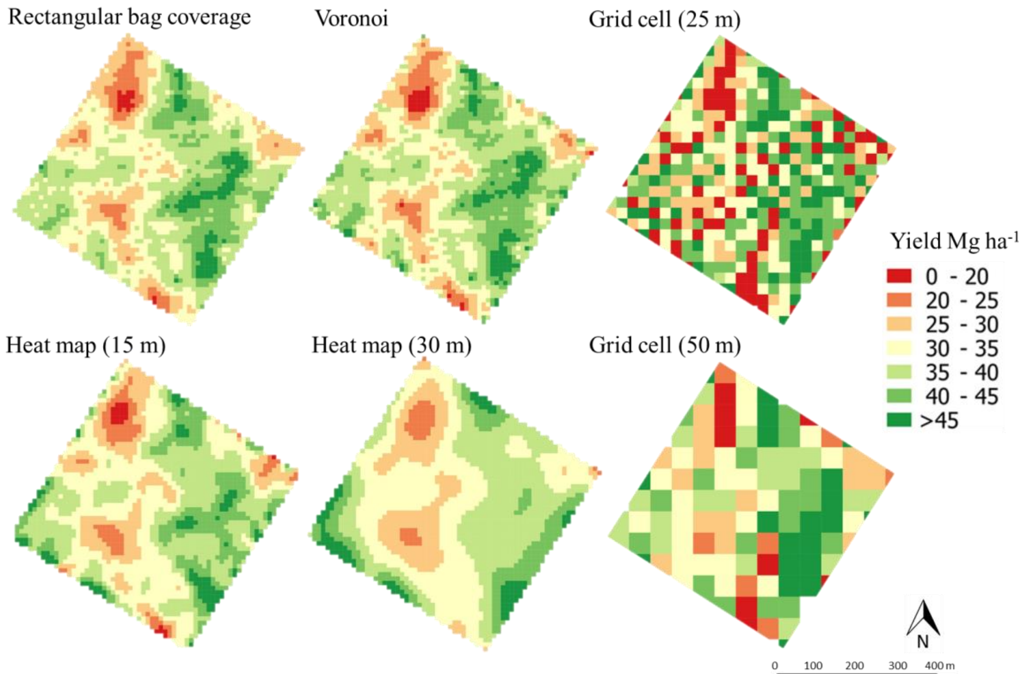

Figure 7: Yield maps from different processing methods carried out over modeled bag location

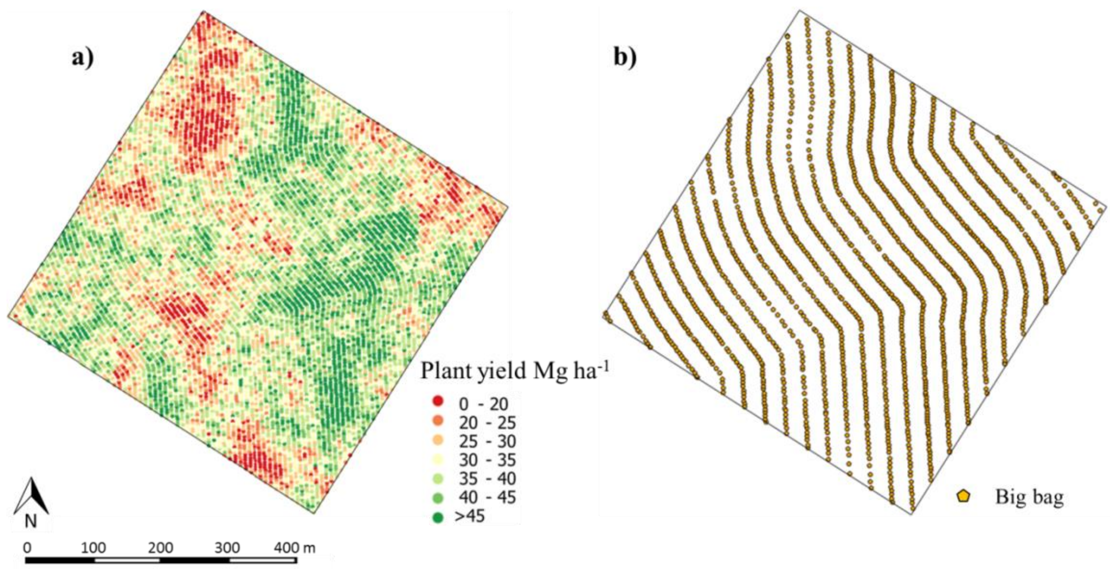

Figure 8: Modeled yield map (a) and bag locations (b)

Table 3: Correlation and average error between yield mapping methods and reference yield

	R^2 *	Average error (%) *
Rectangular bag coverage	0.77	15.19
Voronoi	0.75	15.72
Heat map (15 m)	0.68	17.09
Heat map (30 m)	0.54	20.12
Grid cell (25 m)	0.44	31.14
Grid cell (50 m)	0.55	21.37

* Against reference yield map

Conclusion

Different data processing methods were proposed to generate yield maps from bag location on manually harvested crops. Six methods were tested with real orange bag location data and also with a modeled yield map. All methods presented similar average yield and spatial variability patterns, but different level of detail and accuracy. The best yield mapping performance was found on methods that calculate yield at every bag location. They also presented the highest correlations with actual yield and lowest average error. These maps can show yield variability with good level of detail.

References

Bansal, R., Lee, W. S., Satish, S. 2012. Green citrus detection using fast Fourier transform (FFT) leakage. *Precision Agriculture*, 14(1), 59–70.

Colaço, A. F., Spekken, M. Molin, J. P. 2013. Obtaining yield maps in orchards by tracking machine behavior**. In: Stafford, J. V. (Ed.)** Proceedings of the 9th European Conference on Precision Agriculture, Wageningen Academic Publishers, The Netherlands, pp.683-690.

Colaço, A. F.; Molin, J. P. 2014. A five years study of variable rate fertilization in citrus. In: Proceedings of the 12th International Conference on Precision Agriculture, ISPA, IL, USA. https://www.ispag.org/icpa/Proceedings/

FAO. The Statistics Division. FAOSTAT 2012: core production data. Available in: < http://faostat.fao.org/ >. Access in: December, 10, 2014.

Gong, A., Yu, J., He, Y., Qiu, Z. 2012. Citrus yield estimation based on images processed by an Android mobile phone. *Biosystems Engineering*, 115(2), 162–170.

Okamoto, H., Lee, W. S. 2009. Green citrus detection using hyperspectral imaging. *Computers and Electronics in Agriculture*, 66, 201–208.

Schueller, J. K., Whitney, J. D., Wheaton, T. A., Miller, W. M.,Turner, A. E. 1999. Low-cost automatic yield mapping in hand-harvested citrus. *Computers and Electronics in Agriculture*, 23, 145–153.

Tumbo, S. D., Whitney, J. D., Miller, W. M., Wheaton, T. A. 2002. Development and testing of a citrus yield monitor. *Applied Engineering in Agriculture*, 18(4), 399–403.

Whitney, J.D., Ling, Q., Miller, W.M., Wheaton, T.A. 2001 A DGPS yield monitoring system for Florida citrus. *Applied Engineering in Agriculture*, 17(2), 115-119.

Whitney, J. D., Miller, W. M., Wheaton, T. A., Salyani, M., Schueller, J. K. 1999. Precision farming applications in Florida citrus. *Applied Engineering in Agriculture*, 15(5), 399–403.

Ye, X., Sakai, K., Manago, M., Asada, S., & Sasao, A. 2007. Prediction of citrus yield from airborne hyperspectral imagery. *Precision Agriculture*, 8(3), 111–125.

Zaman, Q. U., Schumann, A. W., Hostler, K. H. 2006. Estimation of citrus fruit yield using ultrasonically-sensed tree size. *Applied Engineering in Agriculture*, 22(1), 39–44.