

A New Approach to Statistical Process Control: Identification of Outliers in Yield Maps

Danilo Tedesco de Oliveira¹, Leonardo Bernache¹, Luan Pereira de Oliveira¹, Murilo Aparecido Voltarelli²
& Rouverson Pereira da Silva¹

¹ Faculty of Agricultural and Veterinary Sciences, São Paulo State University of Julio Mesquita Filho, Jaboticabal, Brazil

² Federal University of São Carlos, Lagoa do Sino, Buri, Brazil

Correspondence: Danilo Tedesco de Oliveira, Rural Engineering Department, Faculty of Agricultural and Veterinary Sciences, São Paulo State University of Julio de Mesquita Filho, Paulo Donato Castellane, s/n, Jaboticabal, São Paulo, Brazil. Tel: 55-(14)-997-733-748. E-mail: danilotedesco@outlook.com

Received: March 14, 2018

Accepted: May 24, 2018

Online Published: July 15, 2018

doi:10.5539/jas.v10n8p334

URL: <https://doi.org/10.5539/jas.v10n8p334>

Abstract

The tools of precision agriculture are of utmost importance in the Brazilian agribusiness, enabling increases in yields and reducing production costs. The use of harvest monitoring systems makes it possible due the possibility to identify pontual problems in an area, however, it becomes necessary to be working properly so it does not acquire incorrect information. Therefore, the purpose with this study was to propose a new approach to identify discrepant points in harvesting maps using statistical process control, as well as to define the best multiple of the standard deviation to identificate these points. The work was conducted during the soybean harvesting at São Geronimo farm in an area of 38 hectares in the municipality of Candido Mota, located in the the state of São Paulo. For gathering information, it was used a Stara crop monitoring system (model Topper Maps) set to record information during harvest in each three second. The productivity data were used to generate an individual control chart to identify points that were out of control so they could be removed. Two standard deviation multiples, that presented an average productivity closer to the average real productivity of the area, were selected. The multiples of the deviations that came closest were the 2σ and 3σ . Two multiples of standard deviation presented an average yield closer to the average real yield of the area. Individual control charts can be used to set control limits and identify possible discrepancies. The multiple of standard deviation 3σ presented information with greater reliability.

Keywords: precision agriculture, mechanized harvesting, control charts, yield

1. Introduction

There are technologies that monitor point-to-point productivity at harvest, as the harvest monitors allied to sensors coupled to machines that collect information in large quantities at short intervals of time. The use of high technology in the field directed to the mechanization of process, use of agricultural inputs, systems of direct sowing, biotechnology and the PA (precision agriculture), made that the commercial agriculture of today undergoes numerous modifications. There are many PA tools that have contributed to and still contribute to the development of agriculture; among them we can highlight the use of productivity maps (Santi et al., 2013).

There was an increase in the use of the PA, mainly with the use of harvest maps to contribute to the monitoring of productivity and yield of the crop. After making a map from an attribute, it is possible to identify where problems are located, and thus being able to enter with the relevant corrections for the next harvest.

Harvest monitors used with sensors coupled to machines collect productivity information in large quantities in short time intervals. However, not all the information collected demonstrates the real productivity of the field, and errors in the recording of information are common (Molin, Cremonini, Menegatti, & Gimenez, 2000). Some of these errors are eliminated by computer-generated mapping software. Still, part of the errors is identification and relatively complex characterization (Gimenez & Molin, 2004).

Errors such as smoothing, volume calibration, incorrect platform width, harvester filling time, grain retraction and crop losses were found on harvest maps (Moore, 1998; Blackmore & Marshall, 1996; Larscheid, Blackmore, & More, 1997; Molin, Cremonini, Menegatti, & Gimenez, 2000).

Menegatti and Molin (2003) developed a methodology to identify and characterize errors in productivity maps, where frequency distribution histograms and upper and lower statistical limits were used with the frequency distribution in the characterization of discrepant data in the data set.

Statistical Process Control (SPC) has, as one of the main objectives, the elimination of variability or part of it (Hessler, Camargo, & Dorion, 2009). This variability can be identified through graphs called Control Charts, which serve to verify when a process is stable or unstable through points inside or outside the control limits. These points when attributed to process errors or special causes, can be considered “outliers”.

Therefore, it is assumed that control charts can be effective on identifying possible discrepant points in harvest maps. Thus, the objective with this study was to propose a new approach to identify discrepant points in harvest maps using statistical process control through individual values control charts, to define the best multiple of the standard deviation to perform the identification of these points.

2. Method

2.1 Description of the Experimental Area

The study was conducted at the São Geronimo farm in a soybean production area of the 2016 crop year with 38 hectares, in the country of Cândido Mota, state of São Paulo, located on coordinates of 22°53'24"S, 50°23'27"W.

2.2 Mechanized Harvest With Harvest Monitor

The harvesting of the area was performed with a John Deere harvester operating at an average speed of 5.0 km h⁻¹, model 1175 equipped with tangential track system, manufactured at 2005 and cutting deck of 19 feet (5.79 meters). The record of productivity information was obtained on a STARA model monitoring system, model Topper 4500, being configured to record points with information every three seconds. In 38 hectares were registered 11,948 points.

The harvest monitor used has a GNSS navigation system where information on latitude, longitude and altimetry positioning is received, which works as reference for the location of the points. It also has a moisture sensor of harvested grains measured by the principle of capacitance; a platform lift sensor which, upon lifting the platform, the monitor stops the information recording, and a volumetric optical type productivity sensor installed in the clean grain elevator of the harvester, performs a reading by interrupting an invisible beam of light in each time interval of 3 seconds.

2.3 Analysis of Discrepant Points (Outliers)

The productivity information recorded by the monitor was downloaded into a spreadsheet and then, inserted into the Minitab software, to create an individual value control chart that is one of the tools of the SPC (Statistical Process Control), testing the multiples of standard deviations (2σ and 3σ) for identification of outliers.

The general average of the individual values is defined, according to Montgomery (2009), according to Equation 1:

$$\mu = \frac{n_1 + n_2 + n_3 + \dots + n_i}{n_i} \quad (1)$$

Where, μ : Average of individual values; n : Values of the observation or sampling point; n_i : Total number of collected points.

The control limits of the individual value charts can be calculated using Equations 2 and 3.

$$UCL = \mu + L \cdot \sigma \quad (2)$$

$$LCL = \mu - L \cdot \sigma \quad (3)$$

Where, UCL : Upper control limit; LCL : Lower limit of control; μ : Mean of individual values; σ : Standard deviation.

All points below the LIC line and above the LSC line were removed using QGIS software (Quantum Gis Version 2.18.3) and then generated descriptive analysis using the measurements of the coefficient of variation, standard deviation and average of the set of points remaining for each multiple of standard deviation.

The data contained in the harvest monitor were tabulated and, with the aid of the QGIS software, a map of the distribution of productivity points was made, classifying low, medium and high productivity points, represented by the red, yellow and green colors, respectively (Figure 1).

The two standard deviation multiples that presented a productivity average closer to the real average obtained in the area (quantified by the producer in the warehouse in $3,471 \text{ kg ha}^{-1}$) would be selected for performing a descriptive analysis at the points that were outside of control

From this selection, the distribution map of these points was generated, identifying possible location of the points, assigning possible errors to explain why these points are outside the control limits.

3. Results and Discussion

The data collected in the harvest monitor were tabulated and, with the aid of the QGIS software, a map of the distribution of the production points was classified, classifying the low, medium and high productivity points represented by the red, yellow and green colors, respectively (Figure 1).

In the productivity map (Figure 1), it can be observed that the points of low productivity are in areas of border and contour of obstacles. In these regions, the harvesters carry out the operation at a lower speed to execute maneuvers, which reduces the flow of material that is being harvested, decreasing and being detrimental to the recording of productivity information.

By decreasing the flow of harvested material, there was a reduction in the amount of processed grains in the clean grain elevator system where the optical productivity meter was installed. It explains why the border regions and the contour of obstacles present less productivity, and it is necessary to filter this type of error.

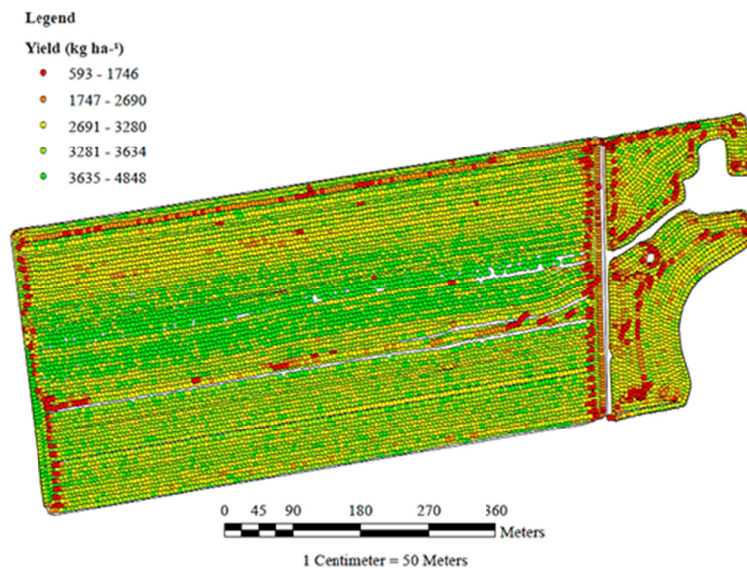


Figure 1. Map of the distribution of soybean yield points. N: North, W: West, E: East, S: South

Based on this context, the original productivity data were used to compose an individual control letter, one of the CEP tools varying from the standard deviation. Through this chart it is possible to identify points that are either inside or out of control (Figure 2). Values that are below or above the LCL (lower control limit) and UCL (upper control limit) lines can be considered outliers.

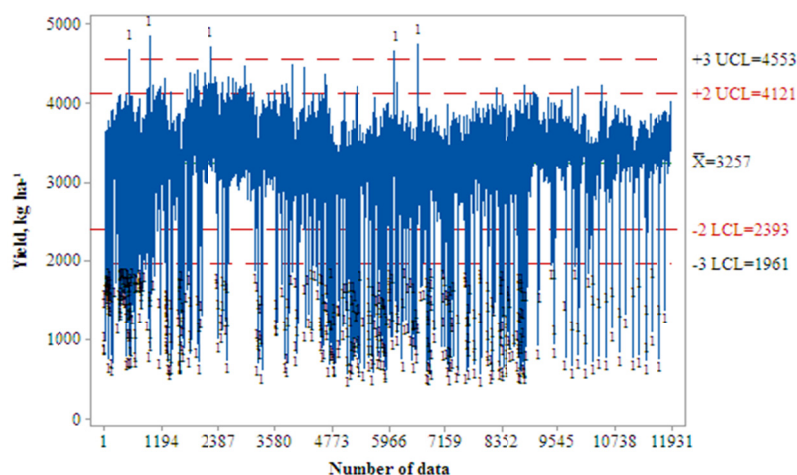


Figure 2. Control chart of individual values for monitoring the yield of the soybean crop as a variation of the standard deviation multiples. UCL: upper control limit, LCL: lower control limit, X: mean

The points out of control that interfere with process quality may be related with six factors: Labor, Machine, Raw Material, Environment, Measurement, Method, as shown by Samohyl (2005). Because it is a volumetric grain flow sensor that records productivity information during harvesting operation, this device is directly related to all the six factors.

Menegatti and Molin (2003) points out that this type of sensor is quite sensitive to the variation of grain density. Another factor that also influences the reading is the declivity of the land where the crop is being harvested. Generally, areas with greater declivity may cause the machine to be tilted sideways and an incorrect reading of the productivity values may occur.

The points found above the UCL line and below the LCL were removed using the software QGIS respecting the control limits of each multiple of standard deviation. The remaining points were submitted to a descriptive analysis (Table 1).

Table 1. Descriptive statistics of treatments performed by multiples of standard deviation

Mesures	Original	3 σ	2 σ
Number of points	11934	11396	11010
Mean (kg ha ⁻¹)	3257.03	3346.70	3366.17
Median (kg ha ⁻¹)	3353	3370	3375
Variance (kg ha ⁻¹)	304.046	127.337	90.314
Sd (kg ha ⁻¹)	551.40	356.84	300.52
Curtose	6.31	1.72	0.30
Asymmetry	-2.16	-0.73	-0.29
C.V. (%)	16.93	10.66	8.93
Amplitude	4277	2515	1702

Note. σ : multiple of standard deviation; C.V.: coefficient of variation; Sd: standard deviation.

The averages and medians presented different values among them, with shows to a dispersion of the data outside the central position. According to Silva et al. (2015), high values between the averages and the median can be explained by the high value of amplitude of the data.

It is also observed that the original values presented a negative asymmetry coefficient, representing a behavior of the distribution curve a little more elongated to the left. The standard deviation multiples of 3 σ and 2 σ , presented values close to zero, which represents a symmetrical distribution of the data set, tending to normality. Regarding the kurtosis coefficient, the original values, 3 and 2 multiples of standard deviation presented positive values, representing a sharper curve for the original values and a flatter curve for the remainder of the data.

When testing different multiples of the standard deviation to perform the removal of points out of control, there was a reduction of the standard deviation and variance of the data set and consequently there was an increase in the average of productivity. It can be explained by the elimination of possible discrepant points that when testing a filter to eliminate discrepant points, obtained similar results (Gimenez & Molin, 2004).

From the descriptive statistics analysis, we selected the standard deviation multiples that presented a productivity average closer to the real value of the area indicated by the farmer, which was $3,471 \text{ kg ha}^{-1}$. To analyze the points that were outside the control limits and to make a map of the distribution of these points.

Analyzing the distribution map of points out of control using two multiples of standard deviation (Figure 3), it is possible to notice that the points above the UCL line are located in the center of the area and as observed in the map with original points (Figure 1) it is a region that presented great uniformity of productivity, so it is possible to affirm that these points presented great reliability and to use only two multiples of standard deviation, ended up being discarded in the process of selection and removal.

In the map, it is possible to identify in the selected points below the LCL, a big amount of points of low productivity, located in the areas where maneuvers and obstacle contour were performed.

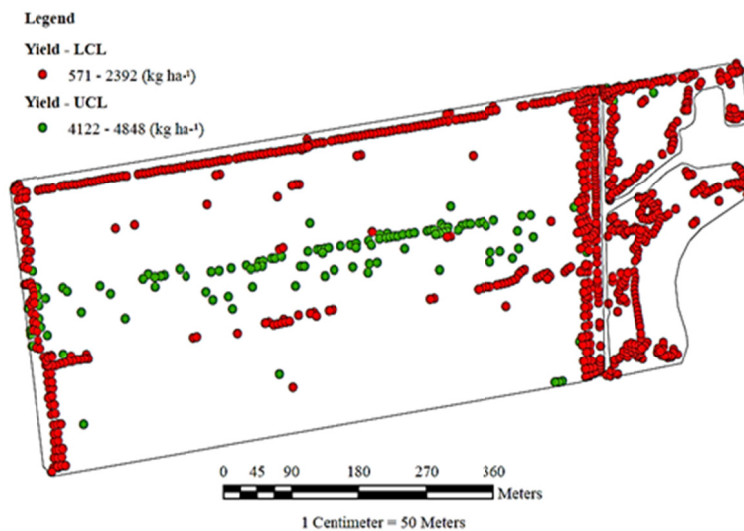


Figure 3. Out of control points selected using two multiples of the standard deviation. UCL: upper control limit; LCL: lower control limit; N: North, W: West, E: East, S: South

When performing the procedure using three multiples of the standard deviation (Figure 4), it is noted that the points below the line of the LCL continued in the regions of the area that occur most maneuvers and obstacle contour, presenting values close to the two multiples of the deviation pattern.

On the points above the UCL line inserted in the map, were identified a reduced number of points in the center of the area, and when compared to the map of original points (Figure 1), it becomes closer to the area map, presenting better information reliability. It agrees with the recommendation of Montgomery (2009b), which states that the value of three multiple of standard deviation provides better results in the performed evaluations.

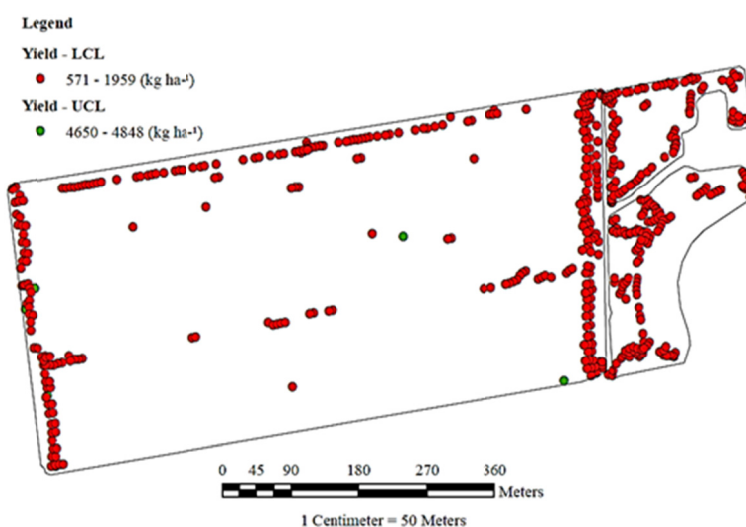


Figure 4. Out of control points with 3 multiples of the standard deviation. UCL: upper control limit; LCL: lower control limit; N: North, W: West, E: East, S: South

Intending to verify the lowest variability among the multiple standard deviations, a descriptive analysis was performed (Table 2). In the table, the number of points removed above the UCL presented smaller amount in the multiple of three standard deviation, considered better compared to the multiple of two standard deviations.

The points removed below the LIC in the multiple of three standard deviation compared to the multiple of 2 standard deviation was better because the volumetric grain flow sensors present more reading errors when there is less grain flow in the elevator of clean grains inside the machine. It was also possible to verify that the coefficient of variation of the multiple of three standard deviations were smaller than the multiples of two standard deviations, both in LSC and SCI. Therefore, both deviations have a CV lower than 10% that according to Pimentel-Gomes and Garcia (2002) presenting more homogeneities in the data.

Table 2. Descriptive analysis of control points presented with two and three multiples of standard deviation

Mesures	+3 LSC	+2 LSC	-3 LIC	-2 LIC
N.P.R.	5	118	533	806
Average (kg ha ⁻¹)	4.725	4.246	1.326	1.621
C.V. (%)	1.62	3.03	30.55	32.87

Note. N.R.P: number of removed points; C.V.: coefficient of variation; UCL: upper control limit; LCL: lower control limit.

4. Conclusions

Therefore, the control charts of individual values of the statistical process control can be used to define control limits for identification of possible discrepancies in harvest maps, and the use of three standard deviation multiples is indicated for data filtering with greater reliability in the mapping of productivity.

Acknowledgements

To the owners of the São Geronimo Farm, Agnaldo Favoni and Vinicius Favoni, and the Professor M.sc. Edson Massao Tanaka of the GECOM (Mechanized Harvest Study Group), for their collaboration in obtaining the data.

References

Blackmore, B. S., & Marshall, C. J. (1996). Yield mapping; errors and algorithms. In P. C. Robert, R. S. Rust, & W. E. Larson (Eds.), *Precision Agriculture: Proceedings of the 3rd International Conference* (pp. 403-416). Madison, WI, USA: American Society of Agronomy, Crop Science Society of America, Soil Science Society of America. <https://doi.org/10.2134/1996.precisionagproc3.c44>

- Gimenez, L. M., & Molin, J. P. (2004). Algoritmo para redução de erros em mapas de produtividade para agricultura de precisão. *Revista Brasileira de Agrocomputação*, 2(1), 5-10.
- Hessler, C. V., Camargo, M. E., & Dorion, E. (2009). Application control charts for analysis of quality in industrial painting. *Proceedings ... IEEE Xplore* (pp. 1330-1334). International Conference on Computers & Industrial Engineering 29, Troyes.
- Larscheid, G., Blackmore, B. S., & Moore, M. (1997). *Management decisions based on yield maps* (pp. 895-902). European Conference on Precision Agriculture, Warwick. Precision Agriculture'97 ... Oxford: BIOS Scientific Publishers Ltd.
- Menegatti, L. A. A., & Molin, J. P. (2003). Metodologia para identificação e caracterização de erros em mapas de produtividade. *Revista Brasileira de Engenharia Agrícola e Ambiental*, 7(2), 367-374. <https://doi.org/10.1590/S1415-43662003000200031>
- Molin, J. P., Cremonini, L. C. M., Menegatti, L., & Gimenez, L. (2000). *Acurácia de um monitor de produtividade com sensor de fluxo volumétrico*. Congresso Brasileiro de Engenharia Agrícola, Fortaleza. Anais/CD-Rom Fortaleza: SBEA.
- Montgomery, D. C. (2009a). Control charts for variables. In D. C. Montgomery (Ed.), *Introduction to statistical quality control* (6th ed., pp. 226-268). Arizona: Wiley.
- Montgomery, D. C. (2009b). Process and measurement system capability analysis. In D. C. Montgomery (Ed.), *Introduction to statistical quality control* (6th ed., pp. 344-390). Wiley, Arizona.
- Moore, M. (1998). *An investigation into the accuracy of yield maps and their subsequent use in crop management* (PhD. Diss., Silsoe College, Cranfield University, Silsoe, UK).
- Pimentel-Gomes, F., & Garcia, C. H. (2002). *Estatística Aplicada a experimentos agrônômicos e florestais: Exposição com exemplos e orientações para uso de aplicativos* (p. 309). Piracicaba: FEALQ.
- Samohyl, R. W. (2005). Controle estatístico de processo e ferramentas da qualidade. In M. M. Carvalho, & E. P. Paladini (Eds.), *Gestão da qualidade: Teoria e casos* (Cap. 9, pp. 261-269). Rio de Janeiro: Elsevier.
- Santi, A. L., Amado, T. J. C., Eitelwein, M. T., Cherubin, M. R., Silva, R. F., & Da Ros, C. O. (2013). Definição de zonas de produtividade em áreas manejadas com agricultura de precisão. *Revista Brasileira de Ciências Agrárias*, 8(3), 510-515. <https://doi.org/10.5039/agraria.v8i3a2489>
- Silva, R. P., Compagnon, A. M., Voltarelli, M. A., & Gígiro, L. A. S. (2015). *Controle de Qualidade em Operações Agrícolas Mecanizadas* (1st ed.). Associação Brasileira de Engenharia Agrícola, Brazil.

Copyrights

Copyright for this article is retained by the author (s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).