2012

Estimativa de densidade amostral para elaboração de mapas de produtividade

REVISTA BRASILEIRA DE ENGENHARIA AGRICOLA E AMBIENTAL, CAMPINA GRANDE, v. 16, n. 4, pp. 449-457, APR, 2012
http://www.producao.usp.br/handle/BDPI/34065

Revista Brasileira de Engenharia Agrícola e Ambiental
v.16, n.4, p.449–457, 2012
Campina Grande, PB, UAEA/UFCG – http://www.agriambi.com.br
Protocolo 138.11 – 24/06/2011 • Aprovado em 30/01/2012

Model to estimate the sampling density for establishment of yield mapping

Graciele R. Spezia[1], Eduardo G. de Souza[2], Lúcia H. P. Nóbrega[2], Miguel A. Uribe-Opazo[2], Marcos Milan[3] & Claudio L. Bazzi[4]

ABSTRACT

Yield mapping represents the spatial variability concerning the features of a productive area and allows intervening on the next year production, for example, on a site-specific input application. The trial aimed at verifying the influence of a sampling density and the type of interpolator on yield mapping precision to be produced by a manual sampling of grains. This solution is usually adopted when a combine with yield monitor can not be used. An yield map was developed using data obtained from a combine equipped with yield monitor during corn harvesting. From this map, 84 sample grids were established and through three interpolators: inverse of square distance, inverse of distance and ordinary kriging, 252 yield maps were created. Then they were compared with the original one using the coefficient of relative deviation (CRD) and the kappa index. The loss regarding yield mapping information increased as the sampling density decreased. Besides, it was also dependent on the interpolation method used. A multiple regression model was adjusted to the variable CRD, according to the following variables: spatial variability index and sampling density. This model aimed at aiding the farmer to define the sampling density, thus, allowing to obtain the manual yield mapping, during eventual problems in the yield monitor.

Key words: precision agriculture, thematic map, spatial variability

Estimativa de densidade amostral para elaboração de mapas de produtividade

RESUMO

O mapa de produtividade representa a variabilidade espacial das características de uma área cultivada e permite intervir na produção dos anos posteriores, na aplicação diferenciada de insumos. Este trabalho teve por objetivo verificar a influência da densidade amostral e do tipo de interpolação na exatidão dos mapas de produtividade, gerados a partir da amostragem manual de grãos, solução que pode ser adotada quando um monitor não pode ser utilizado. Um mapa de produtividade foi obtido com monitor de colheita comercial em lavoura de milho, a partir do qual foram estabelecidas 84 grades e, por meio de três interpoladores, o inverso da distância ao quadrado, inverso da distância e krigagem geraram-se, assim, 252 mapas de produtividade, que foram então comparados com o original, utilizando-se o coeficiente de desvio relativo (CDR) e o índice kappa. A perda de informação do mapa de produtividade aumentou à medida em que se diminuiu a densidade amostral e foi dependente do método de interpolação utilizado. Um modelo de regressão múltipla foi ajustado à variável CDR, em função das variáveis: índice de variabilidade espacial e densidade amostral. Este modelo teve a finalidade de auxiliar o agricultor na definição da densidade amostral e permitir obter-se, manualmente, o mapa de produtividade, caso ocorram problemas eventuais no monitor de colheita.

Palavras-chave: agricultura de precisão, mapas temáticos, variabilidade espacial

[1] Programa de Pós-Graduação em Eng. Agrícola, CCET - UNIOESTE/Cascavel, PR. CEP 85819-110, Cascavel, PR. Fone: (45) 3220-3175. E-mail: gracispezia@hotmail.com
[2] Programa de Pós-Graduação em Eng. Agrícola, CCET - UNIOESTE, Grupos de Pesquisa GROSAP e GGEA, Pesquisador de Produtividade do CNPq. CEP 85819-110 Cascavel, PR. Fone: (45) 3220-3175. E-mail: eduardo.souza@unioeste.br; lucia.nobrega@unioeste.br; miguel.opazo@unioeste.br
[3] Departamento de Engenharia de Biossistemas, ESALQ/USP, CP 09, CEP 13418-900 Piracicaba, SP. Fone: (19) 3429-4165. E-mail: macmilan@usp.br
[4] UTFPR, CEP 85884-000 Medianeira, PR. Fone: (45) 3240-8000. E-mail: bazzi@utfpr.edu.br

Introduction

The spatial variability of soil properties in an agricultural area is something already known by producers. With the advent of precision agriculture (PA), spatial and temporal variability can be considered at large scale aiming to improve the implementation and use of inputs, increase productivity, reduce production costs and environmental impact (Farias et al., 2003). PA provides necessary accuracy and precision in its tools and technologies, making it possible to increase productivity and profitability of crop production while lowering environmental impacts (Corwin & Plant, 2005; Koch et al., 2004). This variability is observed through yield maps that provide parameters to diagnose and correct the causes of lower yields in some areas and study why yield is high in other areas.

However, when a combine is operating in a given area and monitoring yield, problems may occur with the monitor, which prevent the monitoring of the remaining area. In this case, one solution would be to obtain the yield map from a manual sampling. However, there are doubts concerning the sampling density to be used and what would be the loss of information regarding the map that hypothetically would be produced by the monitor. This loss of information, however, can be measured from the comparison of maps produced by both methodologies. From a georeferenced database of yield and using a Geographic Information System (GIS) it is possible to generate yield maps through interpolation. However, one aspect still to be explored is the influence of different types of interpolators in the preparation of thematic maps. Jones et al. (2003) cites that many articles have been published comparing different interpolation methods in a variety of data types. Most of these studies involved comparisons of two-dimensional interpolation methods. The most studied methods were kriging and inverse distance weighting (IDW). Eight studies showed kriging as the best; even when kriging was better "on average", IDW was higher under certain circumstances. Two of the studies showed IDW superior to kriging, and six studies showed little difference between kriging and IDW.

There are algorithms available to construct maps but not for comparison among maps (Lourenço & Landim, 2004). One way is to compare maps by multiple linear regression analysis as Brower & Merriam (2001) when comparing structural contour maps in order to understand the geological history of a region. According to EMBRAPA (2007), another way of comparing maps is by using the kappa index of agreement, which tests the association among maps. The analysis of accuracy is achieved by confusion matrix or error matrix, and subsequently calculated the kappa index of agreement (Congalton & Green, 1993). According to Coelho et al. (2009), another parameter for comparing two thematic maps is the coefficient of relative deviation (CRD), which expresses the average difference, in module, of interpolated values in each map considering one of them as standard map. The lower the percentage found, the higher the similarity between them.

In this context, the objective was to determine the influence of sample density and the type of interpolator on the accuracy of yield maps to be produced from manual sampling of grains.

Material and methods

Yield data used in this study were collected from a farm located in Cascavel, Paraná, with geographic center under the coordinates 24º 58' 44.4'' S and 53º 31' 26.4'' W, 650 m average altitude. The maize crop with physiological cycle of about 120 days was planted between 25 to 30 January 2004 under no-tillage system and 0.20 m plant spacing and 0.70 m row spacing. The crop was harvested from June 30 to July 2, 2004, separating from a 13.2 ha sub-area to be analyzed in this study. The whole area was harvested with combine and harvest monitor and data used to simulate a manual harvest. The yield monitor used was AgLeader® PF 3000®, installed in a combine harvester New Holland TC 57®, corn platform with six row. 14,760 points were collected, which passed through filtration and removal of inconsistent data following the methodology used by Bazzi et al. (2008). In this analysis, spots with times of filling and emptying the harvester less than six seconds and delay time of less than twelve seconds were eliminated. It was also eliminated points with incorrect platform width, failure to GPS differential signal, outliers or zero water content of grains (below 12% and above 40%), positioning error, and discrepant yield values (through the *box-plot* graphs). When concluded the elimination of data considered inconsistent, remained 13,473 points, which originated the map (Figure 1A) of observed yield. From this, maps were constructed performing the points interpolation, resulting in contour maps which allowed a better analysis (Figure 1B).

For convenience, the map was rotated 53° (Figure 2A) and divided into three smaller and regular areas (Figure 2B), called Area 1 (4.38 ha, 2,728 points), Area 2 (4.86 ha, 2,860 points) and Area 3 (4.99 ha, 2,921 points) in order to obtain better local control.

Maize yield maps using manual sampling are performed by harvesting 1 to 2 linear meters of two adjacent rows. However, in this work, no manual samplings were performed, which were simulated using six grids (Figure 3) from the elimination of points of maps corresponding to areas 1, 2 and 3. This estimate considers that the manual sample (usually between 0.7 and 1.4 m^2) has average productivity similar to the sample taken with the yield monitor (15 to 20 m^2). For more information sources, four replications were performed in each area, taking care not to repeat the same points in each repetition and that they stay as far away as possible. Once four replications were performed in each one of the seven grids and at three areas, it reached a total of 84 data sets. The objective of this simulation was to study the sampling density needed to produce maps of productivity from manual sampling, when sampling for some reason cannot be carried out with a harvest monitor.

Data normality was verified by tests of Anderson-Darling and Kolmogorov-Smirnovs at 5% significance level. Data showing normality in at least one of the tests were considered as having normal probability. The outliers were checked through the box-plot graphs. The coefficient of variation (CV) was considered low when ≤ 10% (homoscedasticity), medium when 10% < CV ≤ 20%, high when 20% < CV ≤ 30%, and very high when > 30% (heteroscedasticity) according to Gomes & Garcia (2002).

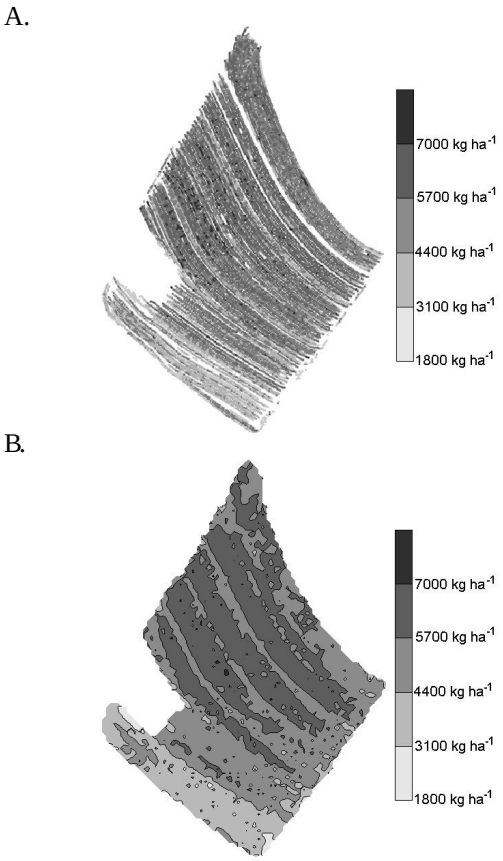

Figure 1. Yield Points Map (kg ha^{-1}) with non-interpolated values (A). Yield contour map (kg ha^{-1}) with values interpolated by the kriging method (B)

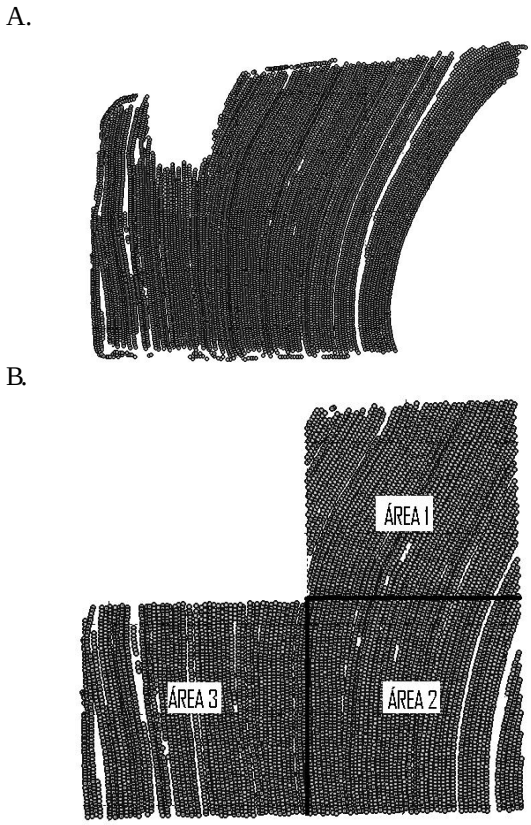

Figura 2. Map rotated 53° with 13,473 points of data collection and non-interpolated values (A); Map clipped from the points map rotated 53°, remaining 8,509 points and non-interpolated values (B)

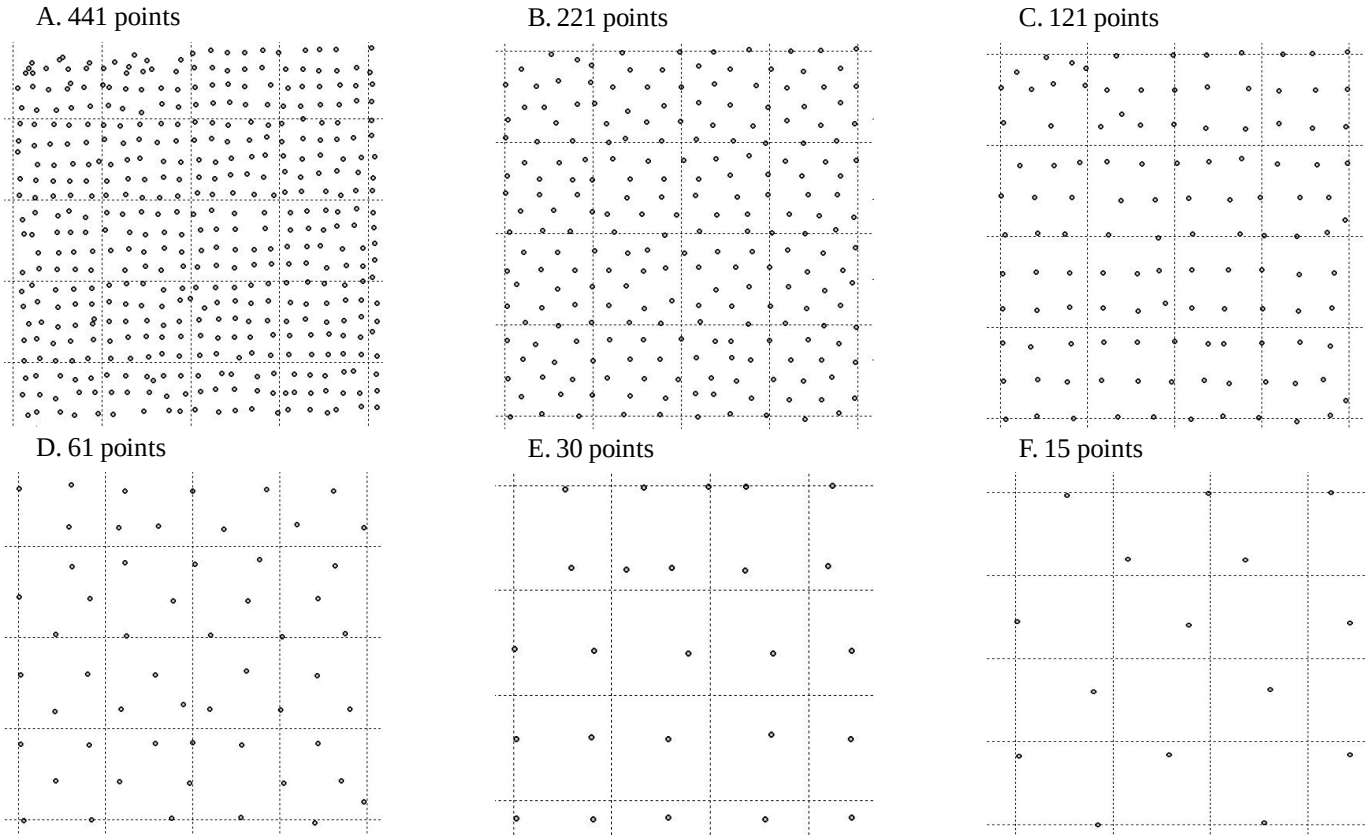

Figure 3. Sampling grids of Area 1 and repetition 1

In geostatistics analysis semivariograms were constructed to verify the spatial dependence on the data. To estimate the experimental semivariance function, the estimator proposed by Cressie & Hawkins (1980) was used. Experimental semivariograms were obtained by applying methods of ordinary least squares adjustment (OLS), adopting the isotropic model (unidirectional semivariogram) with 50% cut-off the maximum distance. In this analysis, the computer program VESPER ® 1.6 Demo was used. Similarly to the spatial dependence index (SDI), proposed by Cambardella et al. (1994), spatial variability index (SVI) Eq. 1, was defined in order to define a proportional index to the spatial variability and not inversely proportional as the SDI. The SVI classification adopted was: very low for SVI < 20%; low for 20 ≤ SVI < 40%; medium for 40 ≤ SVI < 60%; high for 60 ≤ SVI < 80%; and very high for SVI > 80%.

$$SDI = \left(\frac{C_0}{C + C_0} \right).100 \quad (1)$$

$$SVI = 1 - SDI = \left(\frac{C_1}{C_0 + C_1} \right).100 \quad (2)$$

where:

C_o - nugget effect of the adjusted semivariogram

C_1 - scale, $C_0 + C_1$: sill (variance estimate)

In generating thematic maps three types of interpolators were used: inverse of distance (ID), inverse of square distance (ISD), and ordinary kriging and, using the computer program SURFER® 8.0 were prepared the yield maps. In comparing the maps generated from each grid with the original map were used the coefficient of relative deviation (CRD) (Coelho et al., 2009) Eq. 3, which expresses the average difference in module of interpolated values in each map, considering one of them as a standard map and the kappa index (Cohen, 1960) Eq. 4, which uses a confusion matrix for further index calculation. The aim was to evaluate the quality loss of yield maps when reducing the number of sampling points.

$$CRD = \sum_{i=1}^{n} \left| \frac{P_{ij} - P_{ist}}{P_{ist}} \right| \times \frac{100}{n} \quad (3)$$

where:

n - number of points

P_{ij} - productivity at point i to the map j

P_{ist} - productivity at point i for the standard map (kriging with original data)

$$K = \frac{\left\{ n \sum_{i=1}^{r} x_{ii} - \sum_{i=1}^{r} \left(x_{i+} * x_{+i} \right) \right\}}{\left\{ n^2 - \sum_{i=1}^{r} \left(x_{i+} * x_{+i} \right) \right\}} \quad (4)$$

where:

r - number of rows in a cross-classification table

x_{ii} - number of combinations on the diagonal

x_{i+} - total observations in row i

x_{+i} - total number of observations in column i

n - total number of observations

Landis & Koch (1977) suggested the following interpretation for kappa index values (K): no agreement < 0, poor agreement for 0 ≤ K ≤ 0.19, partial agreement for 0.20 ≤ K ≤ 0.39, moderate agreement for 0.40 ≤ K ≤ 0.59, excellent agreement for 0.60 ≤ K ≤ 0.79, perfect agreement for 0.80 d" K d" 1.00. In practice, the kappa index represents the proportion of pixels that were coincident beyond those that would be by pure chance. To evaluate the behavior of similarity between maps, as measured by the coefficient of deviation (CRD), depending on the density and spatial variability, it was fitted the multiple regression model to the CRD variable, depending on the variables SDI (spatial dependence index) and SD (sampling density). The variable selection method used was the best subset. 0.05 significance for F distribution was used to control entry and exit effects. The adjusted coefficient of multiple determination (adjusted R^2) was used as criterion for selection of the best models.

Results and discussion

For space reasons, only data for Area 1 (chosen by lottery) and those for the three areas together will be shown. The maize's average yield (second season) from Area 1 ranged from 5,795 to 6,082 kg ha^{-1}, with a mean value of 5,919 kg ha^{-1}, 38.7% above the average for Paraná state (4,266 kg ha^{-1}) and 94.6% above the average for Brazil (3,041 kg ha^{-1}) (SEAB, 2007). The coefficients of variation (CV) ranged from 13.8% (medium) to 25.2% (high), with 16.7% average (medium), thus characterizing relative data homogeneity (Pimentel-Gomes & Garcia, 2002). Through the normality tests performed was found that 50% yield data sets showed normal distribution at 5% significance level (Table 1).

Regarding the three areas studied, the average yield ranged from 4,491 to 6,082 kg ha^{-1} (Table 2), averaging 5,440 kg ha^{-1}, 78.9% above the national average and 27.5% above the Paraná average. CVs ranged from 11.8% (average) to 26.1% (high), with 18.5% average (medium), thus characterizing relative data homogeneity. Through the normality tests performed was found that 64% yield data sets showed normal distribution at 5% significance level.

Models and parameters adjusted to the semivariograms for Area 1 are shown in Table 3.

The nugget effect (C_0) for data set of the sampling scheme 61_A1_R1 (61-point grid of Area 1 and repetition 1) was 454,602, which compared to other nugget effects for different purposes sampling schemes showed the lowest value. However, the sampling scheme with data set 15_A1_R3 had showed highest C_0, (1,969,444). The range was the greatest for 121_A1_R3, (361.6 m) while the least range was 25.9 m for 61_A1_R1. The sill ranged between 663,333 for the data set 30_A1_R2 and

Table 1. Descriptive analysis of yield for data sets used in Area 1

Data sets	Number of samples	Mean	SD	CV (%)	Minimum	Median	Maximum	Normality
		kg ha^{-1}				kg ha^{-1}		
	Minimum	5795	815	13.8	1827	5778	7082	
Total Area 1	Maxiumum	6082	1494	25.2	4112	6315	8527	
	Medium	5919	985	16.7	2615	6068	7835	
Original_A1_R1	2728	5974	956	16.0	1827	6116	8527	No
441_A1_R1	441	5926	988	16.7	1827	6100	8129	No
221_A1_R1	221	5902	953	16.1	2630	6069	8055	No
121_A1_R1	121	5890	956	16.2	2630	6021	8055	No
61_A1_R1	61	6014	897	14.9	3851	6049	8055	Yes
30_A1_R1	30	5963	851	14.3	3851	6004	7501	Yes
15_A1_R1	15	5844	984	16.8	3851	5988	7501	Yes
Original_A1_R2	2728	5974	956	16.0	1827	6116	8527	No
441_A1_R2	441	5935	995	16.8	1827	6091	8129	No
221_A1_R2	221	5941	932	15.7	3093	6108	7859	No
121_A1_R2	121	5947	949	16.0	3348	6132	7859	Yes
61_A1_R2	61	5902	951	16.1	3507	6044	7607	Yes
30_A1_R2	30	5906	815	13.8	4112	6039	7501	Yes
15_A1_R2	15	5970	893	15.0	4112	6044	7501	Yes
Original_A1_R3	2728	5974	956	16.0	1827	6116	8527	No
441_A1_R3	441	5892	991	16.8	1827	5999	8129	No
221_A1_R3	221	5887	1009	17.2	1827	5974	8129	No
121_A1_R3	121	5935	1009	17.0	1827	6019	8129	Yes
61_A1_R3	61	5808	1025	17.6	1827	5834	7607	Yes
30_A1_R3	30	5795	1129	19.5	1827	5778	7333	Yes
15_A1_R3	15	5935	1494	25.2	1827	6315	7333	Yes
Original_A1_R4	2728	5974	956	16.0	1827	6116	8527	No
441_A1_R4	441	5903	982	16.6	1827	6107	7894	No
221_A1_R4	221	5890	1017	17.3	1827	6114	7894	No
121_A1_R4	121	5887	1070	18.2	1827	6195	7677	No
61_A1_R4	61	5870	888	15.1	3613	6043	7214	Yes
30_A1_R4	30	5820	969	16.7	3613	6082	7082	Yes
15_A1_R4	15	6082	1024	16.8	3613	6280	7082	Yes

* Normality according to the test of Anderson-Darling and Kolmogorov-Smirnovs at 5% significance level.
Meaning of acronyms: XXX_YY_ZZ. Where: XX - Number of points per grid; YY - area to which the grid belongs to; ZZ - repetition to which the grid belongs to. Ex.441_A1_R1: 441-point grid from Area 1 of repetition 1, SD - standard deviation

Table 2. Global overview of the data descriptive analysis used in the three areas studied

Data sets	Values	Mean	SD	CV	Minimum	Median	Maximum
		kg ha^{-1}		(%)		kg ha^{-1}	
	Minimum	4491	696	11.8	1767	4340	6278
All	Maximum	6082	1494	26.1	4559	6315	8527
	Medium	5440	984	18.5	2641	5584	7499

SD - standard deviation

2,223,757 for 15_A1_R3. The spatial variability was between very low (10.6%) and medium (45.0%) according to the classification adopted for the spatial dependence index.

Regarding the three areas studied, 77% cases used the spherical model while 23% used that exponential. The nugget effect (C_0) ranged between 434,977 and 1,969,444. The major and minor ranges were 361.6 m and 17.0 m respectively. The sill varied between 488,275 and 2,223,757. Spatial dependence showed between very low (5.4%) and high (61.7%).

Based on the fitting parameters and models fitted to individual semivariograms (Figure 4, Area 1 and repetition 1) thematic maps were constructed by applying the interpolation inverse of square distance, inverse of distance and kriging for the study variable (maize yield in kg ha^{-1}). With seven grids made in each area (1, 2 and 3), four replicates per area, and three interpolators, 252 maps were obtained. For similarity analysis of thematic maps was calculated the average difference in module, ie the coefficient of relative deviation (CRD). As examples, Figure 5 shows yield maps generated by kriging method corresponding to the semivariograms in Figure 4, where one can observe the loss of information with decreased number of points used.

Coefficients of relative deviations (CRD) derived from the comparison between simulated and original grids show gradual increase of deviation as the amount of points decreases in the three interpolation methods, ie, CRD decreased with increasing sampling density (division of the number of sampling data and sampling area) (Figure 6). For the same sampling density, CRD, deviation from the original map, was smaller for the interpolation inverse of square distance (ISD), followed by the inverse of distance (ID) and lastly by kriging (KRIG) (except for higher density of points). This result is in agreement with Coelho et al. (2009) and Bazzi et al. (2008), who found better performance for interpolation using inverse distance weighting when compared to kriging. Comparing the explanation of the dependent variable CRD, R^2 assumed values of 0.32 (Kriging), 0.61 (ID), and 0.79 (ISD), ie, explanations ranged from 32 to 79%.

Table 3. Models and parameters adjusted to the semivariogram for yield data sets for Area 1

Data sets	Model	Nugget Effect C_0	Scale C_1	Range a (m)	Sill ($C_0 + C_1$)	Spatial variability index (%) SVI = ($C_1/C_0 + C_1$) . 100
Total Área 1	Maximum	1969444	562092	361.6	2223757	45.0
	Medium	782942	268717	156.0	1051659	25.6
Original_A1_R1	Spherical	606881	328514	36.9	935395	35.1
441_A1_R1	Spherical	790750	214461	105.2	1005211	21.3
221_A1_R1	Spherical	767237	144835	45.6	912072	15.9
121_A1_R1	Spherical	720145	194634	36.9	914779	21.9
61_A1_R1	Exponential	454602	371811	25.9	826413	45.0
30_A1_R1	Spherical	602933	79727	89.2	682660	11.7
15_A1_R1	Spherical	760684	154662	45.3	915346	16.9
Original_A1_R2	Spherical	606881	328514	36.9	935395	35.1
441_A1_R2	Exponential	726741	407337	84.3	1134078	35.9
221_A1_R2	Spherical	700387	285943	253.4	986330	29.0
121_A1_R2	Exponential	756371	290705	109.8	1047076	27.8
61_A1_R2	Spherical	769462	220665	242.5	990127	22.3
30_A1_R2	Spherical	555844	107489	252.5	663333	16.2
15_A1_R2	Spherical	622591	79653	82.4	702244	11.3
Original_A1_R3	Spherical	606881	328514	36.9	935395	35.1
441_A1_R3	Spherical	786200	439273	337.7	1225473	35.8
221_A1_R3	Spherical	852952	279297	260.5	1132249	24.7
121_A1_R3	Spherical	987020	117083	361.6	1104103	10.6
61_A1_R3	Spherical	947470	163414	234.9	1110884	14.7
30_A1_R3	Exponential	1075067	158389	78.6	1233456	12.8
15_A1_R3	Spherical	1969444	254313	100.0	2223757	11.4
Original_A1_R4	Spherical	606881	328514	36.9	935395	35.1
441_A1_R4	Spherical	783758	384415	306.8	1168173	32.9
221_A1_R4	Spherical	780668	511420	287.2	1292088	40.0
121_A1_R4	Spherical	908281	562092	339.7	1470373	38.2
61_A1_R4	Exponential	559228	413513	113.4	972741	42.5
30_A1_R4	Spherical	811941	146416	218.3	958357	15.3
15_A1_R4	Spherical	805082	228469	209.9	1033551	22.1

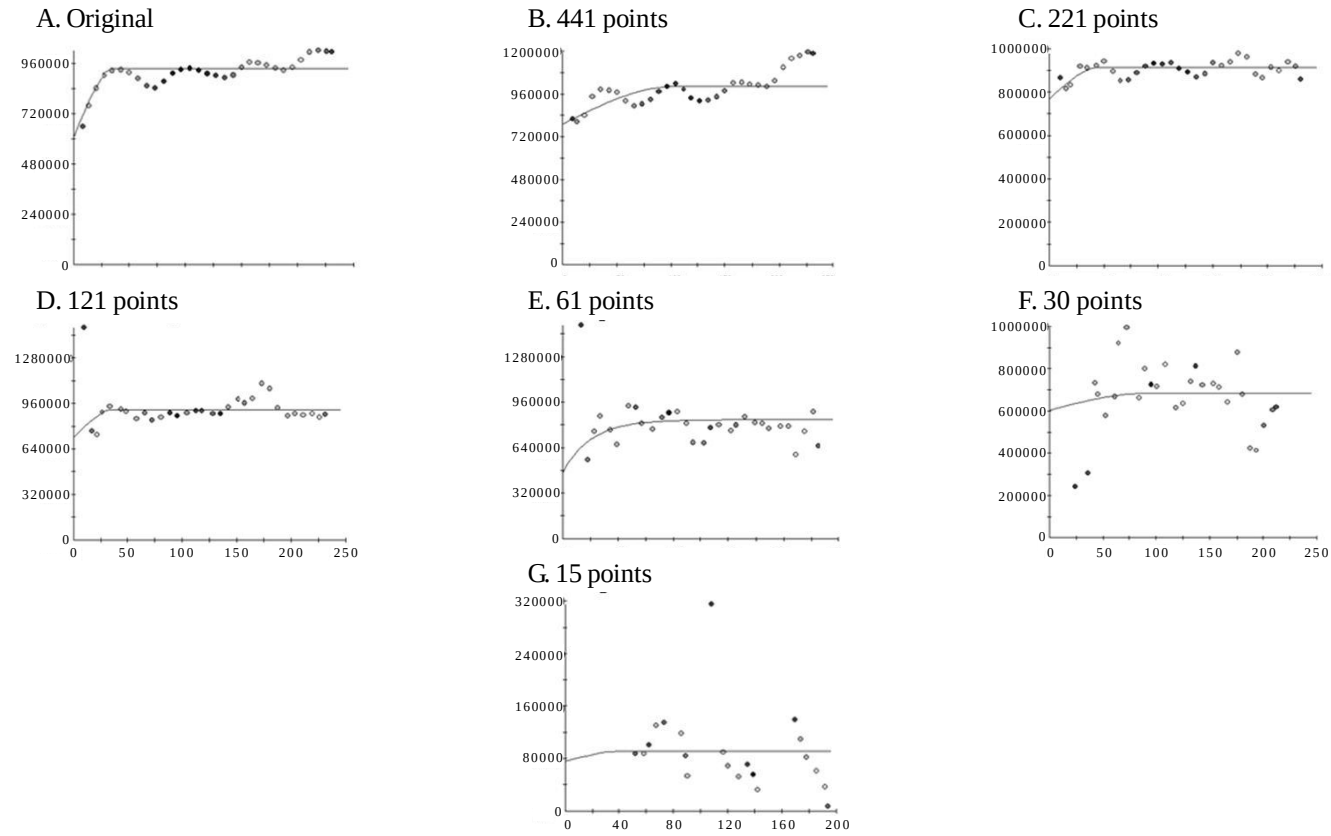

Figura 4. Experimental semivariograms for yield (kg ha^{-1}) of Area 1 and repetition 1

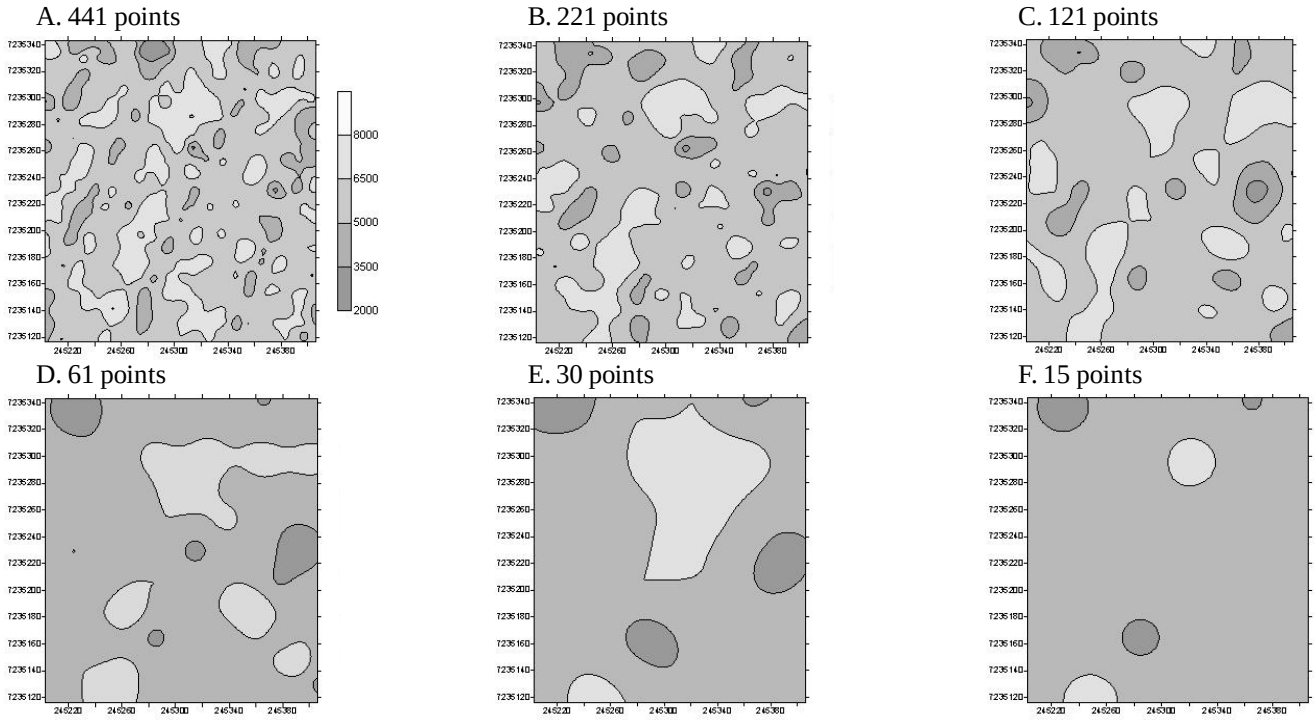

Figure 5. Yield map (kg ha^{-1}) by kriging interpolation method based on the number of points used

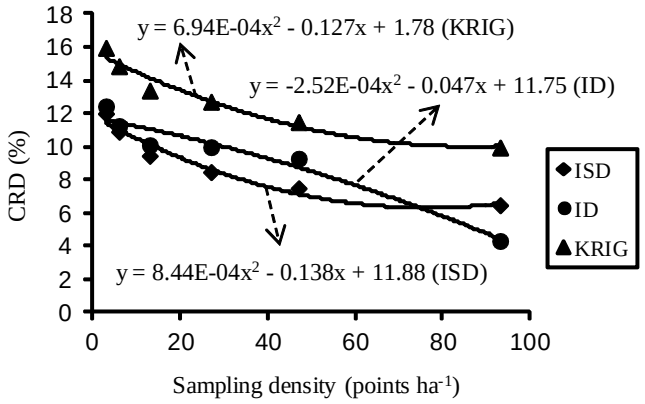

Figure 6. Coefficient of relative deviation (CRD,%) as function of sampling density according to the interpolation method: inverse of square distance (ISD), inverse of distance (ID) and kriging (Krig)

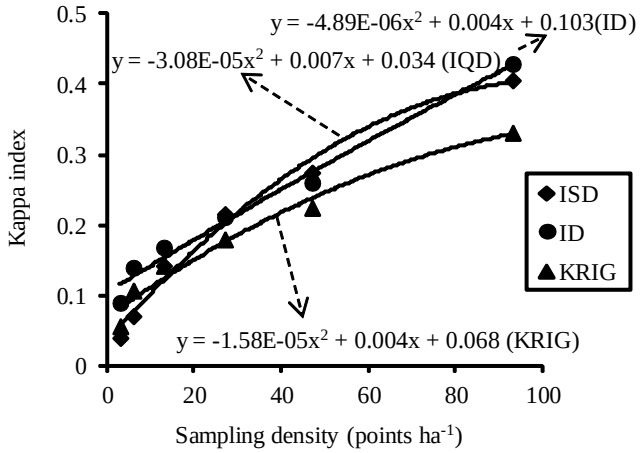

Figure 7. Kappa index as function of sampling density according to the interpolation method: inverse of square distance (ISD), inverse of distance (ID) and kriging (Krig)

Kappa indexes resulting from comparison between simulated and original grids (Figure 7) decreased progressively as sampling density for the three interpolation methods was reduced. The highest index was found in the ID method, i.e. ISD and Krig methods were more influenced by sampling point's elimination. As with the CRD, for the same sample density, kappa index showed higher concordance with the original map for the ISD interpolator; followed by ID and finally by Krig (except for the highest density of points). This result is in agreement with and Bazzi et al. (2008). Comparing the explanation of the dependent variable kappa index, the R^2 assumed values of 0.51 (Krig), 0.54 (ID) and 0.60 (ISD), ie the explanation ranged between 51 and 60%.

The kappa index was associated linearly with CRD (Figure 8), with R^2 between 0.95 and 0.97, ie, explanation of the kappa index above 95%. This means that comparisons of thematic maps performed by both CRD and kappa methods lead to similar results.

The choice for sampling grid depends on the degree of similarity desired between the map that will build and the map that would be produced by the combine equipped with yield monitor. This similarity is estimated by the CRD and the lower the CRD the more similar the maps. To better understand its behavior, a model was developed to describe its dependence as function of the spatial variability index (SVI) and the sampling density (SD, points ha^{-1}). For such an exploratory analysis was carried out initially, which resulted in the following model

$$CDR = a + b.SVI + c.SVI^2 + d.SD + e.SD^2 + \varepsilon \quad (5)$$

where:

CRD - response variable

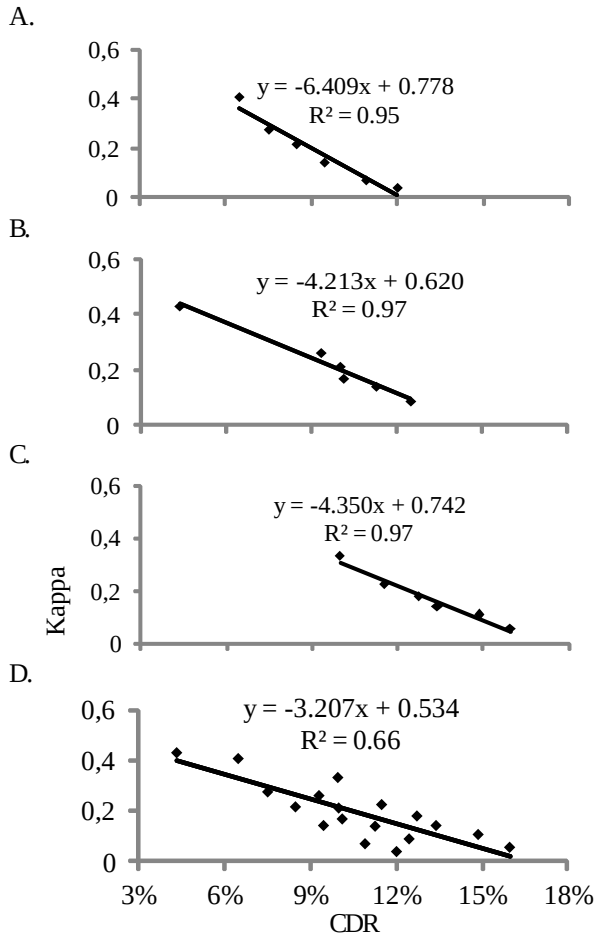

Figure 8. Kappa coefficient as function of the coefficient of relative deviation (CRD) interpolated by (A) inverse of square distance, (B) inverse of distance, (C) kriging and (D) three interpolation methods

SVI and SD - independent variables a, b,...,

a, b, c, d, and e - model parameters to be estimated by the method of least squares

e - random error, which is assumed with normal distribution, zero mean and constant variance σ^2

Applying the aforementioned model to the data collected, an explanation between 87% (ID) and 93% (KRIG) was obtained (Table 4).

Table 4 Estimated coefficients of multiple regression for the coefficient of relative deviation (CRD) according to the spatial variability index (SVI) and the sampling density (SD, points ha^{-1})

Interpolator	Constant	SVI	SVI^2	SD	SD^2	R^2
ISD	16.09	-0.3885	0.0064	-0.1461	0.0010	0.91
ID	16.45	-0.3970	0.0064	-0.0640	-0.0002	0.87
KRIG	20.59	-0.5027	0.0087	-0.1304	0.0007	0.93

Note: All estimators were significant by t test at 5% probability

In practice, these models serve to enable the farmer to obtain the yield map, even when problems occur on the harvest monitor preventing data storage. For such, the following procedure is suggested: 1) With the data already collected from the monitor is estimated the spatial variability index (SVI) through statistical and geostatistical analysis of these data; if this is not possible, the SVI from previous years can be used or even take the 100% maximum value; 2) Adopting the desired degree of similarity map (CRD); 3) Choosing the method of interpolation using the corresponding model in Table 4 (or other developed for the property) and 4) Finding the sampling density suggested.

Conclusions

1. The loss of information in yield map due to the decreased sampling density depends on the interpolation method, being less significant for the method of inverse of square distance and more significant for ordinary kriging, among the three methods evaluated;

2. The regression model developed to describe the dependence of the coefficient of relative deviation on the spatial variability index and sampling density, allowed explaining between 87% (inverse of distance) and 93% (kriging);

3. This model serves to enable the farmer obtaining a yield map, even when problems occur on the harvest monitor preventing data storage; sampling density to be collected depend on the accuracy desired by the farmer.

Acknowledgements

The authors thank the Coordenação de Aperfeiçoamento de Pessoal de Nível (CAPES) and the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for the support provided.

Literature cited

Bazzi, C. L.; Souza, E. G.; Uribe-Opazo, M. A.; Nóbrega, L. H. P.; Pinheiro Neto, R. Influência da distância entre passadas de colhedora equipada com monitor de colheita na precisão dos mapas de produtividade na cultura do milho. Engenharia Agrícola, v.28, p.355-363, 2008.

Brower, J. C.; Merriam, D. F. Thematic map analysis using multiple regression. Mathematical Geology, v.33, p.353-368, 2001.

Cambardella, C. A.; Moorman, T. B.; Novak, J. M.; Parkin, T. B.; Karlen, D. L.; Turco, R. F.; Konopka, A. E. Field-scale variability or soil properties in Central Iowa Soils. Soil Science Society America Journal, v.58, p.1501-1511, 1994.

Coelho, E. C.; Souza, E. G.; Uribe-Opazo, M. A.; Pinheiro Neto, R. Influência da densidade amostral e do tipo de interpolador na elaboração de mapas temáticos. Acta Scientiarum, v.31, p.165-174, 2009.

Cohen, J. A coefficient of agreement for nominal scales. Educational and Psychological Measurement, v.20, p.37-46, 1960.

Congalton, R.G.; Green, K. A practical look at sources of confusion in error matrix generation. Photogrammetric Engineering and Remote Sensing, v.59, p.641-644, 1993.

Corwin, D. L.; Plant, R. E. Applications of apparent soil electrical conductivity in precision agriculture. Computers and Electronics in Agriculture, v.46, p.1-10, 2005.

Cressie, N.; Hawkins, M. Robust estimation of the variogram: I. Mathematical geology, v.12, p.115-125, 1980.

EMBRAPA – Empresa Brasileira de Pesquisa Agropecuária. Fertilidade de solos. http://www.cnpms.embrapa.br/publicacoes/sorgo/solamostra.htm. 10 de Abr. 2007.

Farias, P. R. S.; Nociti, L. A. S.; Barbosa, J. C.; Perecin, D. Agricultura de precisão: mapeamento da produtividade em pomares cítricos usando geoestatística. Revista Brasileira de Fruticultura, v.25, p.235-241, 2003.

Gomes, F. P.; Garcia, C. H. Estatística aplicada a experimentos agronômicos e florestais. Piracicaba: Biblioteca de Ciências Agrárias Luiz de Queiroz/FEALQ. 2002. 307p.

Jones, N. L.; Davis, R. J.; Sabbah, W. A comparison of three-dimensional interpolation techniques for plume characterization. Ground Water, v.41, p.411-419, 2003.

Koch, B.; Khosla, R.; Westfall, D. G.; Frasier, M.; Inman, D. Economic feasibility of variable rate N application in irrigated corn. Agronomy Journal, v.96, p.1572-1580, 2004.

Landis, J. R.; Koch, G. G. The measurement of observer agreement for categorical data. Biometrics, v.33, p.159-174, 1977.

Lourenço, R. W.; Landim, P. M. B. Análise de regressão múltipla espacial. UNESP/Rio Claro, IGCE, DGA, Lab. Geomatemática. Texto didático 13, 34p. 2004. http://www.rc.unesp.br/igce/aplicada/DIDATICOS/LANDIM/Texto13.pdf. 14 Jan. 2010.

SEAB – Secretaria da Agricultura e do Abastecimento do Paraná 2007. http://www.pr.gov.br/seab. 8 Jan. 2008.