SMOOTHNESS INDEX OF THEMATIC MAPS

Claudio L. Bazzi, Juliano. Carnieletto, and Davi M. Rocha

Universidade Tecnológica Federal do Paraná (UTFPR)
Medianeira/PR, Brazil

Eduardo G. Souza

Universidade Estadual do Oeste do Paraná (Unioeste)
Cascavel/ PR, Brazil
** Corresponding author (godoy@unioeste.br)*

ABSTRACT

The thematic maps are generated with the intention of representing the variables in study, being used interpolators to determine their values in places not sampled. The kriging is usually considered the best interpolation method, but several works found good results for the inverse distance and inverse-square-distance methods. The evaluation of the best method usually consists in comparing the estimated values for interpolation with the sampled values. However, the visual aspect of the map created is also relevant, and the smoothness of the contour curves, an item to be taken into account, because it facilitates the visual interpretation and the site-specific management of agricultural inputs in the agriculture. The objective of this work was to evaluate the smoothness of three thematic maps, created with different interpolators, built through corn yield data in an area of 13.2 ha. The selected interpolators were kriging, inverse of the distance, and inverse-square-distance. The evaluation was accomplished through the smoothness index, proposed in this article that calculates the frequency of classes change of the thematic map in the horizontal, vertical and diagonal directions through software built for this purpose. The index proved efficient in the determination of the smoothness of thematic maps.

Keywords: management zones, interpolators, yield map

INTRODUCTION

The thematic map is the main instrument in the process of making decisions in the precision agriculture (PA). The most common type of thematic map is the yield map, and it is a georeferenced graphich display of the yield of a culture per area unit. According to Molin (2001), the PA has a closed cycle of tasks which must be preferentially initiated through the yield map, which integrates the effects of several spatial variables such as soil properties, fertilizers rate, topographic attributes, atmosphere conditions and illnesses infestation. According to Michelan et al. (2007), the crop is formed by various processes in which errors can occur, for which reason a filtering of data methodology becomes necessary for the attainment of reliable maps. Among the main types of error, we can mention: time of delay, filling and emptying time, positioning of the GPS and effective shorter length of the crop than that showed in the yield monitor. Based on the yield map it is possible to define regions that may have been influenced by some variable on which the yield depends, enabling the application of such attribute in a differentiated manner in accordance with the estimated necessity (Cordeiro et al., 2001).

The thematic maps, such as the yield map and those of related attributes are elaborated taking into consideration samples collected in the evaluated area. In order to estimate the values in sites not sampled, interpolators are utilized, being the main ones the kriging, the inverse distance, and the inverse square distance. Jones et al. (2003) mention that many articles were published comparing different methods of interpolation in a large variety of types of data. These articles utilized various types of data: atmosphere, contents of clay and pH of the soil; rain precipitation; and elevation. Most of these studies involved two-dimensional methods of interporlation. The methods studied in bigger depth were kriging and the inverse distance raised to a power (IDP). Eight studies showed that the kriging was the best method, and that even when the kriging proved more effective “in average”, the IDP was more effective under determined circumstances. Two of the studies concluded that the IDP was more effective than the kriging and six of the studies showed a very minor difference between kriging and IDP.

Bazzi et al. (2008) evaluated the decrease of the precision in yield maps obtained in areas where combine harvesters not equipped with yield monitors associated with combine harvesters equipped with yield monitor were utilized. The decrease in the number of measurements and therefore of the number of points monitored decreased the precision of the maps in the inverse square distance and kriging interpolation methods, being the influence on the kriging more expressive. According to Balastreire et al. (1998), small errors end up having little practical relevance, since what effectively matters is to have regions which can be managed in a differentiated manner and the knowledge of which attribute must be provided for the increase of the yield in that specific region. In the elaboration of thematic maps the appearance of small regions inside other much larger regions and with very different values which render the visual interpretation and its utilization in the site-specific management of the agriculture inputs difficult is frequent.

In practice, the site-specific management of these sub-areas is usually unfeasible and it would be desirable that the interpolation process would eliminate them. However, the visual analysis is subjective and therefore an index which can characterize this situation is desirable. The purpose of this work was to evaluate the smooth index of thematic maps proposed by Souza et al.(2009) and utilize it to evaluate the smoothness in twelve thematic maps, created by the kriging interpolators (KRI), inverse distance (ID), inverse square distance (IQD), referring to 4 distinct areas.

MATERIAL AND METHODS

The yield of corn and soybean cultures of four areas (24º 57' S e 53º 27' W, average elevation of 750 m) located in the rural zone of the city of Cascavel, state of Paraná, Brazil, was evaluated in the period from 2003 to 2007. The harvesting was carried out with the utilization of a combine harvester New Holland® TC 57, equipped with yield monitor AgLeader® PF 3000. After the collection of data, the elimination of sampling points presenting very high or very low yield was done, following the procedure adopted by BLACKMORE & MOORE (1999). These points were probably influenced by errors sources such as: delay time, filling and emptying time, positioning of the GPS and effective shorter length of the crop than that showed in the yield monitor. Data with very high or very low water content, originated by humidity sensor reading errors were also eliminated. The maps were elaborated with the utilization of filtered data for each area (Table 1).

Table 1 Metadata collection and filtering

Culture/Havest	Simbology	Area (ha)	Average Speed (km h^{-1})	Time* (s)	Gross Total Points	FinalTotal Points	Point number reduction	Sampling Density (points ha^{-1})
Soybean 2002/2003	Soybean/03	14.8	5.3	1.00	19,351	18,306	5.4%	1,237
Corn 2003/2004	Corn/04	30.3	4.0	3.00	14,693	13,738	6.5%	453
Soybean 2005/2006	Soybean/06	45.3	5.5	3.00	8,089	7,960	1.6%	176
Soybean 2006/2007	Soybean/07	30.0	5.7	3.00	6,037	5,246	13.1%	175

*Time - collection period between the two samples in seconds.

The data were analyzed statistically through descriptive and exploratory data analysis with the calculation of the average, median, minimum, maximum, standard deviation and coefficient of variation. For the verification of the normality of data, the tests of Anderson-Darling and Kolmogorov-Smirnovs at the level of 5% of significance were utilized, being considered normal data which presented normality in at least one of the tests. The coefficient of variation (CV) was considered low when $CV \leq 10\%$ (homocedastici ty), medium when $10\% < CV \leq 20\%$, high when $20\% < CV \leq 30\%$, and very high when $CV > 30\%$ (heterocedasticity) (PIMENTEL-GOMES & GARCIA, 2002).

In the construction of semivariograms the software ArcView 9.2 was utilized in the ordinary kriging module. The theoretical models spherical, exponential and Gaussian were adjusted to the semivariograms through the method OLS (*Ordinary Last Square*), standard of the software utilized, which also provides the cross-validation as a tool of selection of the most adequate model of theoretical semivariogram. Among the estimates provided by the software for the evaluation of models, we have the reduced mean-square error ($\overline{ER}$, equation 1) reduced error standard deviation (S_{ER}, equation 2). In order to avoid the situation in which the estimates point to different models, a new estimate called Index of Comparison of Errors (ICE, equation 3) was proposed, which in the selection of *j* models provides a lower value the lowest closest to zero is the $\overline{ER}$ (Reduced Mean-Square Error) and the closest to one is the S_{ER} (Reduced Error Standard Deviation). Therefore in the selection of various models, the one presenting the lowest ICE is considered the best model.

$$\overline{ER} = \frac{1}{n}\sum_{i=1}^{n}\frac{Z(s_i) - \hat{Z}(s_{(i)})}{\sigma(\hat{Z}(s_{(i)}))} \qquad [1]$$

$$S_{ER} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\frac{|Z(s_i) - \hat{Z}(s_{(i)})|}{\sigma(\hat{Z}(s_{(i)}))}} \qquad [2]$$

Where $\sigma(\hat{Z}(s_{(i)}))$ is the kriging deviation error in the point s_i, without considering the observation $Z(s_{(i)})$.

$$ICE_i = A + B \qquad [3]$$

$$A = \begin{cases} \dfrac{ABS(SME)i}{MAX(ABS(SME))}, when\ MAX(ABS(\overline{ER})) > 0 \\ 1, when\ the\ contrary\ ocours \end{cases} \qquad [4]$$

$$A = \begin{cases} \dfrac{ABS(SDRME)i}{MAX(ABS(SDRME))}, when\ MAX(ABS(S_{ER})) > 0 \\ 1, when\ the\ contrary\ ocours \end{cases} \qquad [5]$$

ICE_i is the index of comparison of errors for the model *i*.

The degree of spatial dependence was classified in accordance with the Spatial Dependence Index (IDE, Equation 6), and CAMBARDELLA *et al*. (1994) proposed the following intervals to evaluate the percentage of the semivariance of the nugget effect: ≤ 25 % - strong spatial dependence; between 25 % and 75 % - moderate spatial dependence and ≥ 75 % - weak spatial dependence.

$$IDE = \frac{C_0}{C_1 + C_0} x 100 \qquad [6]$$

In the elaboration of thematic maps, the interpolators inverse distance, inverse square distance and kriging were utillized, using the software Surfer 9 (Bazzi et al., 2008). After interpolated, the maps were classified with the utilization of 5 classes (as suggested by Molin, 2001). After the superficial evaluation of each map created, each of them was exported to a *database file* (dbf).

The smoothness index (*SI*, equation 7), proposed by Souza (2009), calculates the frequency of the change of classes in the thematic maps in the horizontal, vertical and diagonal directions, pixel per pixel, through a software created for this purpose. In the hypothesis that the map had a totally homogeneous area, a smooth index of 100% would be obtained due to the absence of change of class. In the same way, if the map was generated with random values, the smooth index would present a value close to zero.

$$SI = 100 - \left(\left(\frac{\sum_{i=1}^{l} NM_{Hi}}{4P_H} + \frac{\sum_{i=1}^{c} NM_{Vi}}{4P_V} + \frac{\sum_{i=1}^{n} NM_{Ddi}}{4P_{Dd}} + \frac{\sum_{i=1}^{n} NM_{Dei}}{4P_{De}} \right) * 100 \right) \qquad [7]$$

where, NM_{Hi} - Number of changes in the horizontal line i; NM_{Vi} - Number of changes in the vertical line j; NM_{Ddi} - Number of changes in the right diagonal in the diagonal k; NM_{Dei} - Number of changes in the left diagonal in the diagonal l; P_H - Number of pixels in the horizontal; P_V - Possibility of change in the vertical; P_{Dd} - Possibility of change in the right diagonal; P_{De} - Possibility of change in the left diagonal.

For evaluation and execution of the process referring to the calculation of the index, the Smoothness Index Calculator 2.0 (SIC2.0, Figure 1), utilizing Java language and the technique of support to the development IDE Eclipse, was utilized.

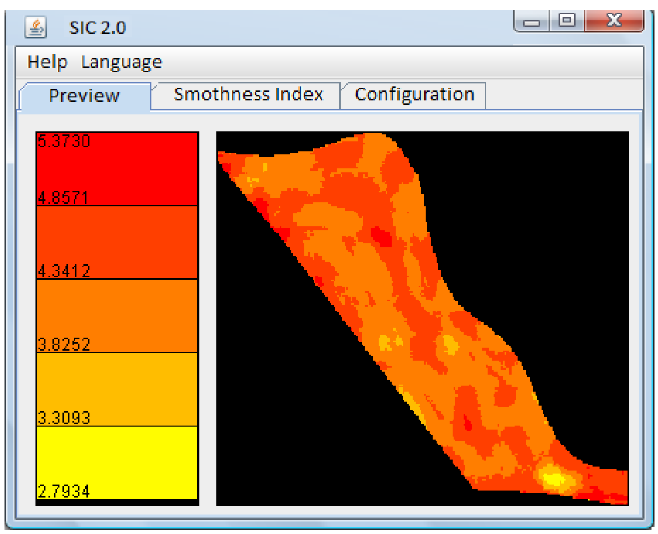


Figure1. Visual environment Smoothness Index Calculator 2.0 (SIC2.0).

The program enables the set up of configurations (Figure 2) so that it is possible to alter the analysis file, to select the quantity of classes or blocks, to alter the change of the colors of the classes and to indicate the maximum and minimum values.

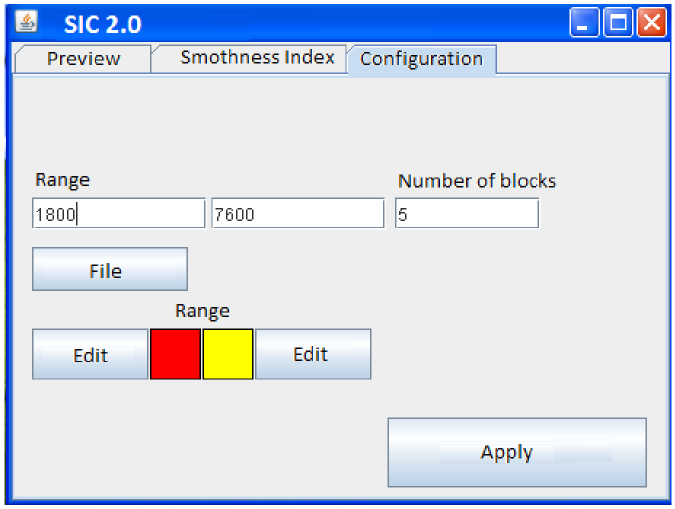


Figure 2. Configuration display of the software Smoothness Index Calculator 2.0 (SIC2.0).

RESULTS AND DISCUSSION

The descriptive analysis of data (Table 5) showed that the four data sets (one for each stand) did not have normal distribution. The productive values presented average (soybean/06, CV = 16.2%; soybean/07, CV = 13.1%) and high (corn/04, CV = 28.3% and soybean/03 CV = 24.3%) variabilities. The maximum value varied from 220% (soybean/07) to 556% (corn/04) from the minimum value, which backs up the premise that even in small areas (from 15 to 45 ha) the variability of the data is very large.

Table 5. Descriptive statistics for the yield data

Culture	Minimum (kg ha^{-1})	Mean (kg ha^{-1})	Median (kg ha^{-1})	Maximum (kg ha^{-1})	StDev (kg ha^{-1})	CV (%)	Amplitude (kg ha^{-1})	Skewness	Kurtosis	N*
Soybean/03	675	1,903	1,836	3,564	540	28.3	2,889	0.56 c	0.01 A	No
Corn/04	1,646	5,549	5,667	9,147	1,350	24.3	7,501	-0.60 c	-0.12 A	No
Soybean/06	2,061	3,741	3,788	5,422	610	16.2	3,361	-0.26 c	-0.09 A	No
Soybean/07	2,414	3,852	3,872	5,314	500	13.1	2,900	1.27 c	11.2 A	No

Skewness: symmetric (a), positive skewness (b), negative skewness (c);
Kurtosis: mesokurtic (A), platykurtic (B), leptokurtic (C);
StDev – Standard deviation; CV – Coefficient of Variation;
* Normality tested with Anderson-Darling and Kolmogorov-Smirnovs tests.

The maps of yield score referring to Soybean/2006 and Soybean/2007 (Figure 3) presented gaps in the data survey, contrary to the Soybean/2003 and Corn/2004. However, this fact could be compensated by the interpolation of data.

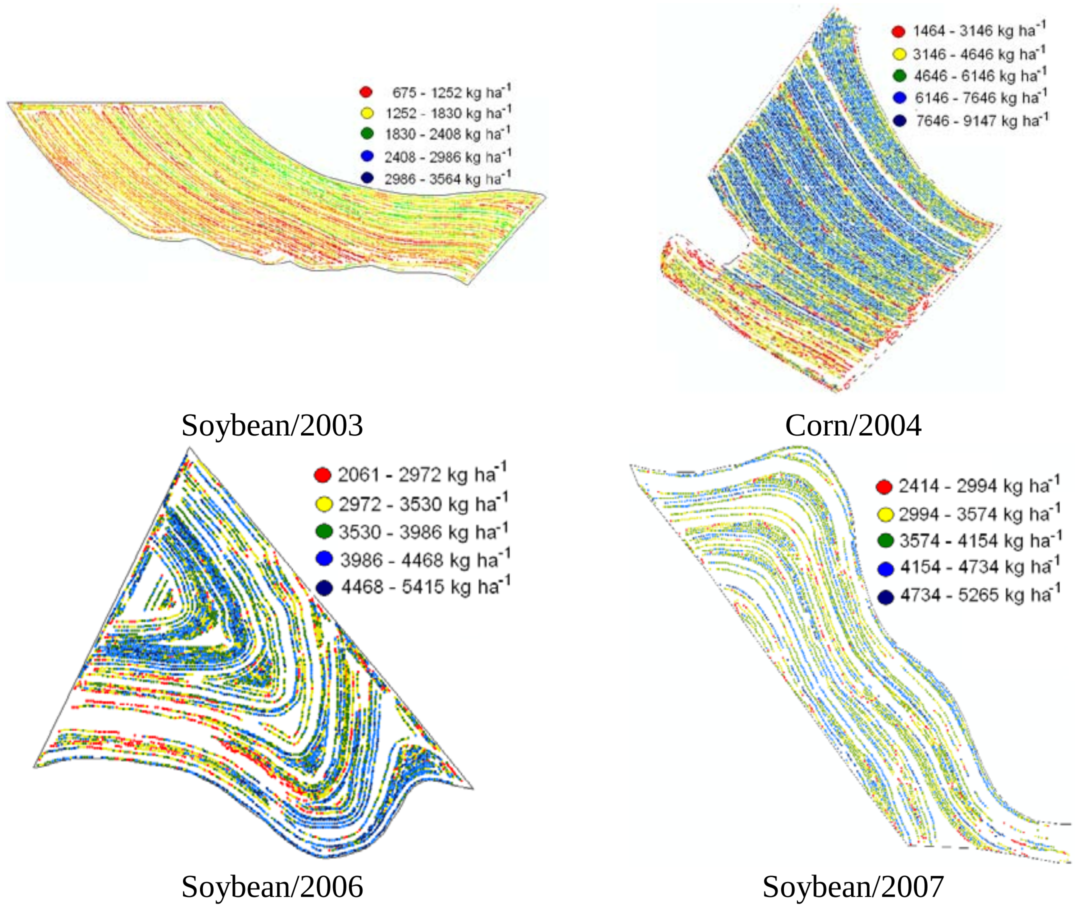


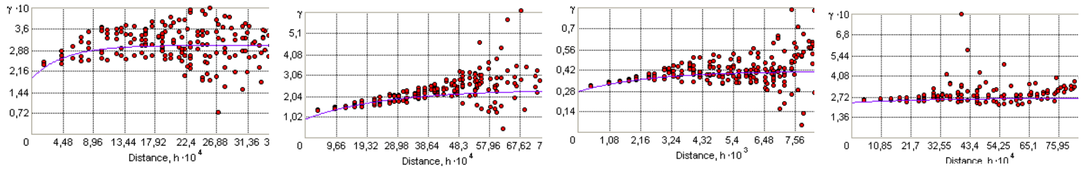
Figure 3 Score Maps

In the geostatistical analysis (Table 3), the method of cross-validation pointed to the exponential model as the best model for adjustment of the semivariograms (Figure 4), as it provided the lowest ICE in all the cases. Date presented mostly medium spatial dependence because for most of cases the spatial dependence index (SDI) varied in the interval from 25 to 75% (CAMBARDELLA et al., 1994).

Table 3. Models and parameters of the semivariograms for each culture

Variable	Model	Co	C1	Sill	a (m)	SDI	$\overline{ER}$	S_{ER}	ICE
	Gaussian	0.2292	0.0731	0.3023	120.4	75.8%	-0.00427	0.8712	2.00
Soybean/03	Exponential	0.1912	0.1128	0.3040	142.6	62.9%	-0.003	0.9098	1.40
	Spherical	0.2148	0.0871	0.3019	138.3	71.1%	-0.00374	0.8839	1.77
	Gaussian	1.2588	1.3694	2.6282	653.9	47.9%	0.00079	0.82	1.75
Corn/04	Exponential	0.9033	1.4391	2.3424	748.2	38.6%	0.00099	0.9405	1.27
	Spherical	1.0609	1.4770	2.5379	748.2	41.9%	0.00105	0.8814	1.66
	Gaussian	0.3010	0.0942	0.3952	410.1	76.2%	-0.00617	0.9069	2.00
Soybean/06	Exponential	0.2728	0.1400	0.4128	727.6	66.1%	-0.00472	0.9359	1.45
	Spherical	0.2826	0.1109	0.3935	461.8	71.8%	-0.00524	0.9257	1.65
	Gaussian	0.2419	0.0248	0.2667	846.5	90.7%	0.00322	0.9034	1.91
Soybean/07	Exponential	0.2327	0.0297	0.2624	846.5	88.9%	0.00337	0.9168	1.86
	Spherical	0.2375	0.0261	0.2636	846.5	90.0%	0.0033	0.9099	1.95

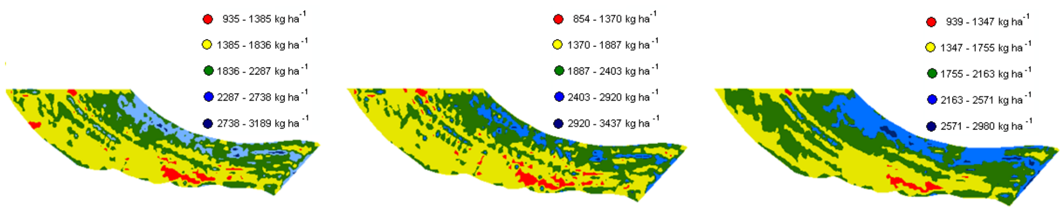
* C0 – Nugget Effect; C1 – Contribution; Sill – C0 +C1; a – Range; SDI – Spatial Dependence Index; $\overline{ER}$ – Average Error; S_{ER} – Average Error Standard Deviation.



Soybean/2003 Corn/2004 Soybean/2006 Soybean/2007

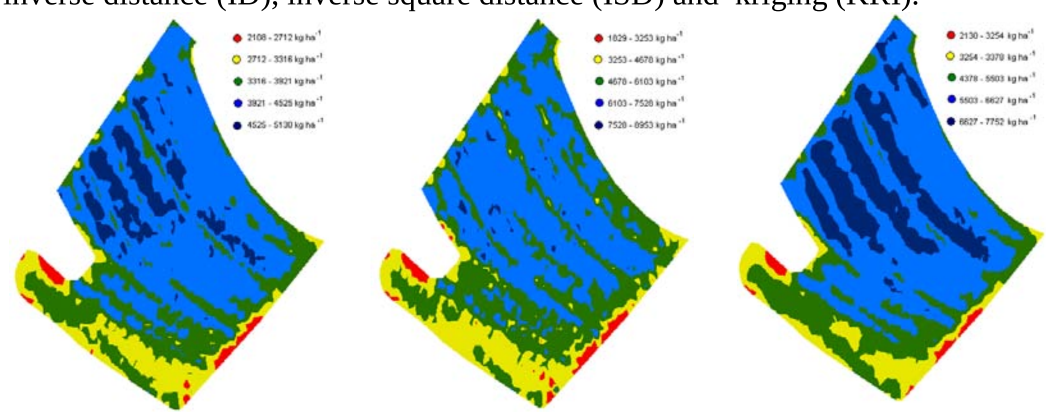
Figure 4 Experimental semivariograms

For each data set (soybean/2003, Figure 5; corn/2004,Figure 6; soybean/2006, Figure 7; soybean/2007, Figure 8), three thematic maps referring to the yield were generated, utilizing the interpolation methods inverse distance (ID), inverse square distance (ISD) and kriging (KRI)



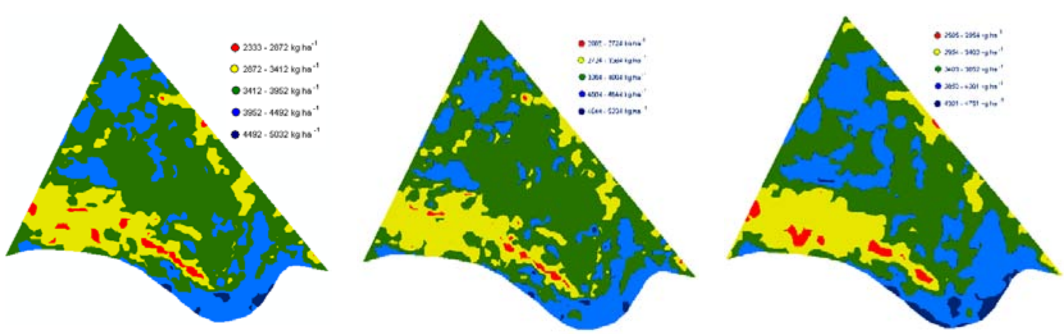
ID ISD KRI

Figure 5 Yield map for the crop Soybean/03, utilizing the interpolation methods inverse distance (ID), inverse square distance (ISD) and kriging (KRI).



ID ISD KRI

Figure 6 Yield map for the crop of Corn/04, utilizing the interpolation methods inverse distance (ID), inverse square distance (ISD) and kriging (KRI).



ID ISD KRI

Figure 7 Yield map for the crop Soybean/06, utilizing the interpolation methods inverse distance (ID), inverse square distance (ISD) and kriging (KRI).

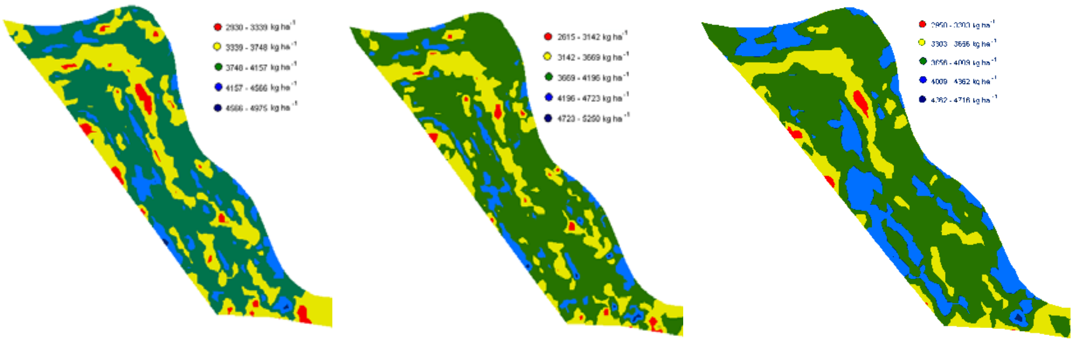


Figure 8 Yield map for the crop Soybean/07, utilizing the interpolation methods inverse distance (ID), inverse square distance (ISD) and kriging (KRI).

For the cultures soybean/03 and soybean/06 and corn/04 (Table 4), the kriging interpolator provided the maps with the highest smoothness (higher SI), and the ISD interpolator presented the lowest smoothness indexes. For the year 2007, in the culture of soybean, the ISD interpolator presented a higher smoothness index, probably influenced by the shorter amplitude and data standard deviation in relation to the other cultures. It is further noticeable that this set of data presented asymmetry and kurtosis indexes more significant than the other sets.

Table 4 Smooth Index in relation to the interpolator (inverse distance (ID), inverse square distance (ISD) and kriging (KRI)) for each set of data

Product/Crop	Soybean/03	Corn/04	Soybean/06	Soybean/07
ID	73%	86%	88%	88%
IQD	61%	71%	76%	94%
KRI	79%	88%	90%	89%

Minimum value Maximum value

CONCLUSIONS

Through the smoothness index, it can be concluded that in three of the four cases analysed the kriging interpolator presented the maps with the highest smoothness and the inverse square distance interpolator, the maps with the lowest.

ACKNOWLEDGEMENTS

The authors thank for the support provided by the State University of Western Paraná (Unioeste), the Araucária Foundation (Fundação Araucária), the General office of State of Science, Technology and Higher Education - SETI/PR, Capes Foundation –Ministry of Eduacation of Brazil, Technological Foundation Park Itaipu (for the doctorate scholarship); and the National Council for Scientific and Technological Development (CNPq)

BIBLIOGRAPHICAL REFERENCES

Balastreire, L.A. Estudo de caso, uma pesquisa brasileira em agricultura de precisão. 1998. In: Silva, F.M., P.H. de M. Borges. Mecanização e agricultura de precisão. UFLA; SBEA. 1(1):302-231.

Bazzi, C.L., E.G. Souza, M.A. Uribe-Opazo, L.H.P. Nóbrega, and R.P. Neto. 2008. Influência do número de passadas da colhedora com monitor de colheita nos mapas de produtividade na cultura do milho. Revista Brasileira de Engenharia Agrícola, 28(2): 355-363.

Cordeiro, R. 2001. Efeito do desenho em amostragem de conglomerados para estimar a distribuição de ocupações entre trabalhadores. Rev. Saúde Pública. 35(1):10-5.

Jones, N.L., R.J. Davis, and W. Sabbah. 2003. A comparison of three-dimensional interpolation techniques for plume characterization. Ground Water. 41(4):411-419.

Matheron, G. Principles of geostatistics. 1963. Economic Geology. Paris, 58(1):1246-1266.

Michelan, R.; E.G. Souza; and M.A. Uribe-Opazo. 2007. Determinação e remoção do tempo de atraso em mapas de colheita de milho. Acta Scientiarum, (UEM). 29(1):147-155,

Molin, J. P. 2001. Agricultura de precisão – o gerenciamento da variabilidade. LTC. Piracicaba – Sp.

Pimentel-Gomes, F.; and C.H. Garcia. 2002. Estatística Aplicada a Experimentos Agronômicos e Florestais. Biblioteca de Ciências Agrárias Luiz de Queiroz-FEALQ. Piracicaba-Sp.