Spectral reflectance of satellite images using geostatistics methods to estimate growth and cotton yield

Ziany N. Brandão*[a], Célia R. Grego[b], Ricardo Y. Inamasu[c], Lúcio A. Jorge[c]

[a]Brazilian Agricultural Research Corporation – Embrapa Cotton, 1143 Oswaldo Cruz, Campina Grande, PB, Brazil 58428-095; [b]Embrapa Satellite Monitoring. 303 Soldado Passarinho Ave., Campinas, SP, Brazil 13070-115; [c]Embrapa Instrumentation, 1452 XV de Novembro, São Carlos, SP, Brazil 13560-970

ABSTRACT

The objective of this study was the spatial identification of the NDVI index and cotton yield distributions through different crop phenological stages using geostatistical methods in Goiás state, Brazil. The experiment was carried out in a commercial field with 47.4 ha, in 80x80m georeferenced grid with 74 plots. Yield monitor data and multispectral satellite images at 56 m spatial resolution were collected in a rainfed cotton field in two dates to monitor the plant vigor. Satellite images of AWiFS sensor were acquired on 08/02/2011 and 01/04/2011, during the first flowering and fruiting cotton stages, respectively, corresponding to 70 and 120DAE (days after emergence). Measures of canopy reflectance, plant height and leaf nitrogen content were determined and cotton yield was obtained by mechanical harvest in August, 2011. Data were analyzed using descriptive statistics, correlation and geostatistical analyses by building and setting semivariograms and kriging interpolation. Best correlation was found between NDVI and cotton yield at 120DAE. At first flowering, the NDVI and cotton yield showed strong spatial dependence, while for 120DAE there was no dependence, probably due to the enlargement of vegetated coverage. There were similarities in the bottom left of the study area with high values of NDVI, as well as the highest values of cotton yield due to excellent plant vigor in the cotton flowering stage. Identifications of spatial differences were possible using geostatistical methods with remote sensing data obtained from medium resolution satellite images, allowing to identify distinct stages of plant growth and also to predict the cotton yield.

Keywords: Precision agriculture, normalized difference vegetation index, geostatistical analyses

1. INTRODUCTION

Conventional measurements of plant nutrients and crop growth are expensive, time consuming and often are performed only on a small portion of the production areas. Furthermore, high variability of these parameters within a single field make the optimization of the plant treatment a difficult task, especially for fertilization, which often lead to an oversupply of nutrients [1].

On the other hand, measurement of spectral indices is fast, nondestructive and could be conducted on a large scale [2]. Remotely sensed data have been widely used to develop vegetation indices as indicators of crop growth, nutrient status assessment and yield prediction [3].

Among the nutrients, nitrogen plays an important role in cotton plant growth and development, leaf photosynthesis, boll retention, and yield development. The growth and yield of cotton depend strongly of the availability of N during the season, and the management of this input is a critical issue in cotton [4]. Moreover, the spatial variability of soil fertility and the different N requirements at various stages of cotton development are additional reasons highlighting the need for practical methods of evaluating N status in commercial fields of cotton.

Cotton visual deficiencies of N have been associated with changes in leaf color [5]. This characteristic combined with plant structure and many other physiological changes can be measured through crop reflectance using multi and hyperspectral instruments and have proven to be a strong estimator of cotton leaf N status [6], [7].

However, many factors influence the N availability for cotton. The task to supply N with the high variability in readings, even with high resolution instruments, is not easy in large areas. So, geostatistic has served to advance analytical methodology for spatial interpolation and to facilitate quantification of spatial features of canopy and soil properties [8].

Geostatistical estimation makes possible to predict values at unsampled locations by taking spatial correlation between estimated and sampled points into account (i.e., spatial variability). In addition, geostatistical estimation minimizes the variance of estimation error, improving the accuracy of spatial prediction. The kriging interpolation method has been tested and it is considered a good method and substantially superior to inverse distance weighting methods [8]. This interpolation method allows the data visualization into maps and became a useful tool to evaluate the variability of many properties. Thus, crop spectral reflectance can be used to evaluate plant nutrient status and estimate plant growth and the measures readings and spectral data can be converted into maps showing the variability for direct application in precision farming.

In this context, the objective of this study was to analyze the spatial identification of the NDVI (*Normalized Difference Vegetation Index*) and cotton yield distributions, through different crop phenological dates, using geostatistical methods in Goiás state, Brazil.

2. MATERIAL AND METHODS

2.1 Study area

This study was conducted in an experimental area of 47.4 ha, located at Pamplona Farm (16º10'16'' S, 47º37'47'' W), Cristalina, GO state, Brazil, in a Typic Hapludox soil, with clay, 59.5%; sand, 16.1%; silt, 24.3%; pH in water, 5.67; Ca, 3.11 cmolc dm³; organic matter, 2.93%; available phosphorus, 3.52 mg dm³; and extractable potassium, 0.25 cmol$_c$ dm^{-3}. The climate is characterized as tropical, predominantly Aw at Köppen climate classification, with a dry season from April to September and rainy season from October to March, with an annual average rainfall of 1446.5 mm.
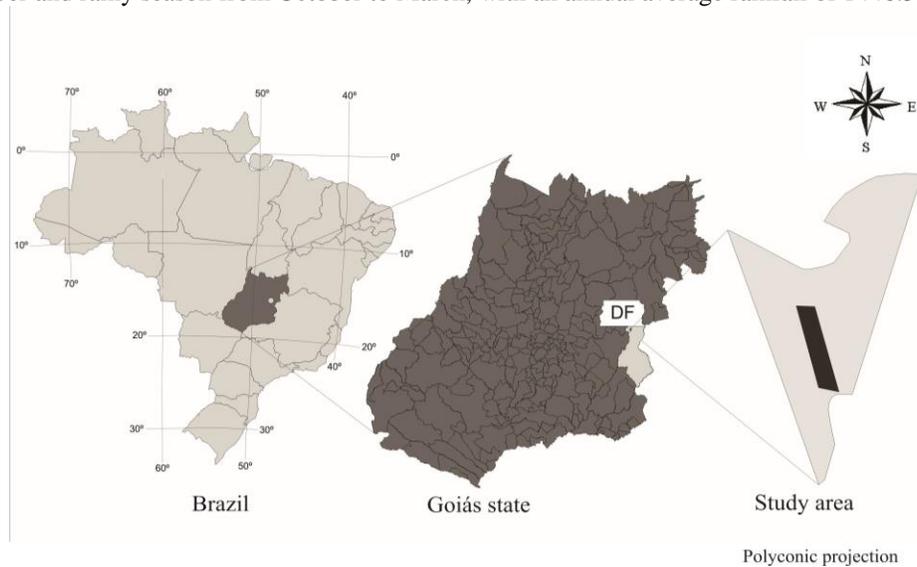
Figure 1. Location of the study area. (Source: IBGE, Brazil).

2.2 Data

Plants were cultivated on the rainy season, and a sampling grid was made before sowing containing 74 points of 80x80m. Data were collected into a 10m radius of the each central point, previously georeferenced with a 76CSx GPS by GARMIN®. The cotton emergence occurred on December 2nd, 2010 and the crop was harvested on August 2011. Nitrogen fertilization was carried out based on soil analysis, in rows spaced by 0.76 m, applied in four stages (pre-planting, sowing, emergence and 45 DAE) adding up to a total of 160 kg ha^{-1}. Diseases and weeds were controlled based on their occurrence, according to the farm's technical recommendations.

Sampling dates are made in clear days took into account the same satellite dates and the phenological stages of plants. For determination of leaf nitrogen content, leaf samples were collected at first flowering (70 DAE - days after

emergence) and fruiting stages (120 DAE), with the leaf of the 5th position of the main stem and collected on 30 plants for each point. Samples were oven dried for 48h at 65°C and ground through a 20 mesh sieve in a Wiley mill. Leaf N concentrations were determined on duplicate samples of 6 mg of ground leaf material.

2.3 Hyperspectral data

In order to determine relationships between reflectance and plant growth, the spectral measurements were made at the same two phenological periods. At sampling dates, leaf tissue were collected simultaneously with the canopy reflectance measurements obtained on sunny days above all plots using a portable PAR-NIR Apogee® spectroradiometer, which operates in the range from 350 nm to 1100 nm, with a 0.5 nm sampling interval. Measurements were made after a previous calibration with a white plate, which provides maximum reflectance. The spectral reflectance measurements in each plot at each sampling date were averaged and the mean values were used in statistical analysis. Hyperspectral reflectance index NDVI at 670 and 800nm was calculated by equation (1):

$$NDVI = (\rho_{800} - \rho_{670})/(\rho_{800} + \rho_{670}) \tag{1}$$

2.4 Satellite image

Two images generated by the sensor AWiFS (Advanced Wide Field Sensor) were acquired to identify correlation with cotton yield. The first one on 08/02/2011 (70 DAE), with orbit 331 and point 089, and the second image on 01/04/2011 (120 DAE), with orbit 327 and point 087.

The AWiFS sensor has a spatial resolution of 56 m with four spectral bands and revisit time of 5 days, covering approximately 740 kilometers [9]. The spectral bands used in this work correspond to channel 2 (red), from 620 to 680 nm and channel 3 from 770 to 860 nm (near infrared). The imagery was radiometrically and geometrically corrected, and rectified to the Universal Transverse Mercator (UTM) coordinate system. After corrections and radiometric calibration of the images, reflectance and vegetation index for two sampling dates were determined using the visible and near infrared bands. All procedures for image rectification and calibration were performed using ERDAS IMAGINE® 8.7.

2.5 Geostatistc analysis

Data were submitted to statistical analysis aimed at the determination of average, maximum, minimum, kurtosis coefficient, frequency distribution and coefficient of variation (CV). In order to verify the normality of the data frequency distribution, was used the Kolmogorov-Smirnov test, on which skewness and kurtosis values should be near zero for normal distributions. The coefficents of variation were evaluated based on Warrick and Nielsen criteria [8], which classifies as a low CV < 12%, regular 12% to 60% and high to CV > 60%.

Geostatistical analysis was performed by constructing and adjusting semivariogram and ordinary kriging interpolation using geostatistical package Vesper [10].

Experimental variograms were computed from data to evaluate the spatial dependence analyses, which were conducted by fitting a model to the experimental semivariograms, following the regionalized variables theory [8]. Semivariograms were constructed assuming the intrinsic stationarity hypothesis and the semi-variance, which was estimated by equation (2):

$$\gamma^*(h) = \frac{1}{2\,N(h)} \sum_{i=1}^{N(h)} [Z(x_i) - Z(x_i + h)]^2 \tag{2}$$

where N(h) is the number of pairs of measured values Z(x_i), Z($x_i + h$) separated by a vector h. According to Vieira et al. [11], it is expected that measurements within some neighborhood are more similar than those separated by large distances. This means that γ (h) increases with the distance h up to a sill, which is the maximum value in which the variogram stabilizes, corresponding to the limit distance of spatial dependence. Semivariograms with spatial dependence were fitted with the best mathematical model.

If spatial dependence is showed by the semivariogram, non-sampled data can be estimated by kriging, with minimum variance and without trend [12]. After that, isolines and contour maps were made with the estimated data as a function of geographic coordinates.

2.6 Cotton yield data

Cotton yield data was obtained by mechanical harvesting using a cotton picker, model 9996 by John Deere®, integrated with a GreenStar® Harvest Doc® system, consisting of ultrasonic flow sensor, data monitor (Monitor GreenStar ™ 2), data storage card (Keycard), and Global Positioning System (GPS) with differential signal, StarFire ™ receiver in order to monitor and mapping the cotton yield. Seed cotton yield and GPS data were recorded at 3 s intervals. Data were examined using Apex® software and exported to further treatment, cleaning, filtering and exported to analysis. All the results were presented as two-dimensional maps representing the spatial distribution of the values of leaf N, NDVI and cotton yield.

3. RESULTS AND DISCUSSION

NDVI index were obtained by spectroradiometer (spec) and satellite data (sat). Images were constructed from all AWiFS scenes. Table 1 summarizes the descriptive statistical analysis for cotton yield and all data type (NDVIspec or NDVIsat) in the two sampling date (70 and 120DAE).

As expected, NDVI average values obtained by satellite image were lower than hyperspectral NDVIspec at two sampling dates (Table 1). The plants were at their early growing stages on February 08, and more soil surface exposure than on April 01. Bare soils usually produce very small and positive NDVI, which is much lower than the NDVI for healthy vegetation. Ground measurements are done above the canopy in the rows with the minimal soil influence [13]. At 120 DAE, the area of soil was reduced due to crop growth, since plant height average was 1.174m, and rows were completely filled. Thus, the difference between NDVI averages was reduced and values were similar, in spite of the low resolution of satellite image.

All variables presented the normal frequency distribution considering that skewness and kurtosis values were close to zero, according Kolmogorov-Smirnov test, as well as low variation was observed for all variable (CV < 12%). The NDVIsat and NDVIspec showed CV around 6% at 70 DAE (6.59 and 6.03%, respect.). Better and uniform canopy distribution with plant growth at 120 DAE provided a CV reduction on this index.

It was observed the highest correlations at 120 DAE. The highest correlation coefficient between NDVIspec and leaf nitrogen content was 0.832, and between NDVIsat and cotton yield was 0.492.

Cotton yield showed a low CV (4.18%), especially close to CV of NDVIspec at 120 DAE (3.95%). Despite of the nitrogen concentration average at 120 DAE was within sufficiency range to promote a good yield, at first flowering it presented value a little below to the low limit for high productivity in cotton [14]. This fact is one of the causes of the experimental area provided small productivity, 2990 kg ha^{-1}, against 3700 kg ha^{-1} for Brazilian national average in cotton yield at the same year.

Table 1. Descriptive statistical analysis for cotton yield, leaf nitrogen content (N) and normalized difference vegetation index (NDVI) obtained through spectroradiometer (spec) and satellite images data (sat) at 70 and 120DAE.

Variable	Average	Variance	CV (%)	Minimum	Maximum	Skeweness	Kurtosis	KS*
NDVIsat 70DAE	0.50	0.0011	6.59	0.43	0.59	0.941	1.609	0.148
NDVIspec 70DAE	0.81	0.0024	6.03	0.71	0.91	-0.007	-0.967	0.107
N (g kg^{-1})70DAE	37.88	5.7650	6.34	33.30	43.30	0.009	-0.504	0.114
NDVIsat 120DAE	0.85	0.0008	3.36	0.80	0.93	0.900	0.311	0.125
NDVIspec 120DAE	0.87	0.0012	3.95	0.80	0.94	-0.245	-0.427	0.142
N (g kg^{-1}) 120DAE	45.54	5.1760	5.00	41.00	51.70	0.274	0.475	0.141
Cotton yield (kg ha^{-1})	2990	15630	4.18	2699	3268	-0.471	-0.176	0.135

*KS= Kolmogorov-Smirnov test   D (table) = 0.155

Results from the geostatistical analysis are presented in Figure 2. Spatial dependence was observed by experimental semivariograms that were fitted using the best-adjusted model with a smaller root mean square (RMS) and validated by the jack-knifing method [5], [11].
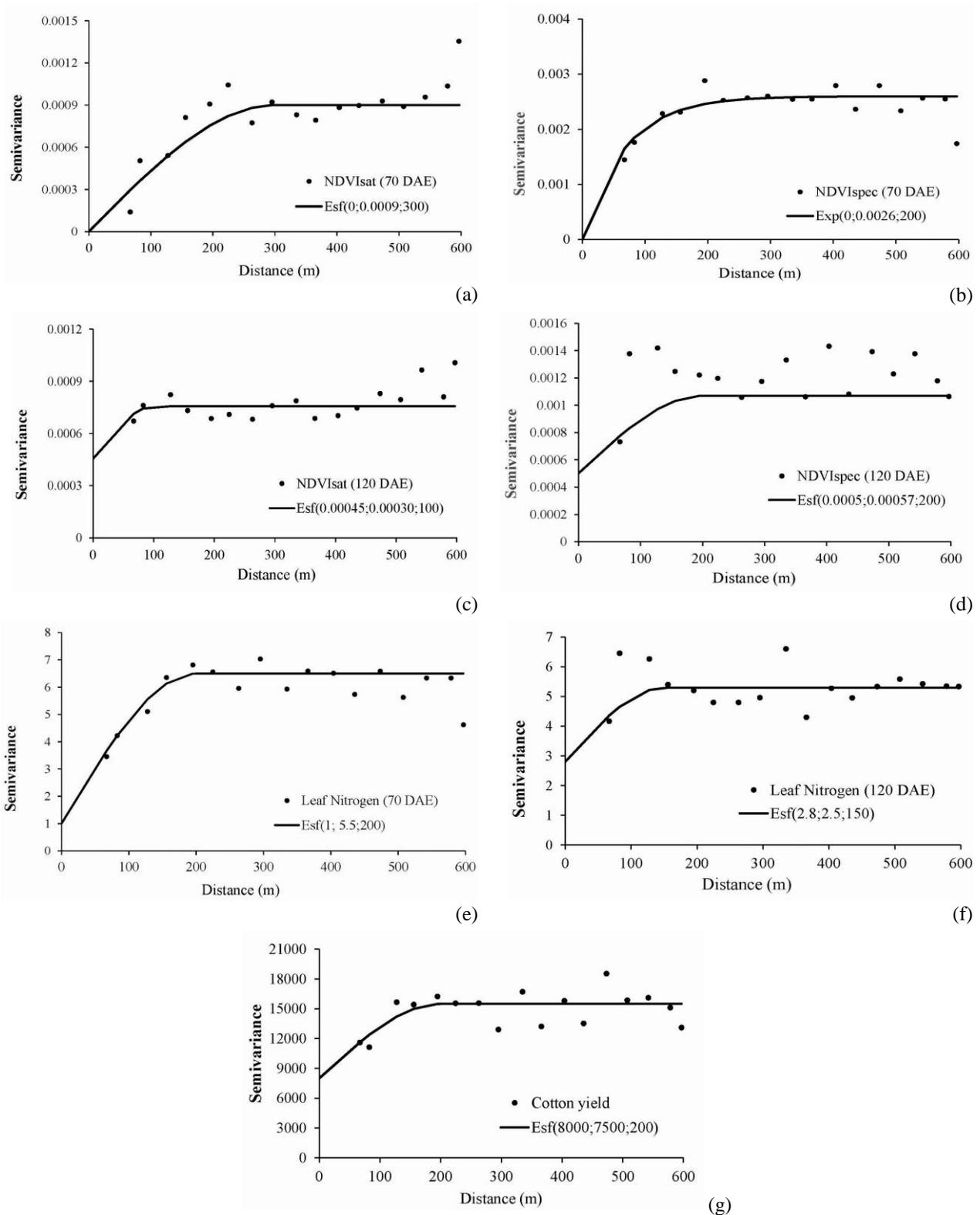
Figure 2. Semivariograms. (a) NDVI by satellite data at 70 DAE; (b) NDVI by spectroradiometer at 70 DAE; (c) NDVI by satellite data at 120 DAE; (d) NDVI by spectroradiometer at 120 DAE; (e) Leaf N content at 70 DAE; (f) Leaf N content at 120 DAE and (g) Cotton yield

With the exception of NDVIspec in first flowering (70 DAE), which was adjusted to exponential model, all data and crop characteristics were fitted to the spherical model. The nugget effect (C0) represents non-explained variance, frequently caused by errors in measurement or by variations of properties not detected in the sampling scale [8]. Cambardella et al. [15] proposed that spatial dependence degree (SDD) be verified by the relationship between the nugget effect (C0) and the sill (C0+C) being classified as weak for values greater than 75%; moderate between 75 and 25%; and strong for values below 25%.

According to this approach, most crop variable presented moderated dependence, with the exception of NDVIsat, NDVIspec and leaf N content at 70 DAE that presented strong spatial dependence (SDD from 0 to 15.3%). The range defines the maximum radius, from which neighboring samples are drawn for interpolation by kriging, since samples are spatially related. Therefore, closer samples are more important in the interpolation because they are more similar than distant samples [8]. Spatial dependence showed a range between 100 and 300 meters, corresponding to the radius of distinct zones on kriging maps.

After identifying the spatial dependence by adjusting the semivariograms and getting parameters as nugget effect, partial sill and range, data were interpolated by ordinary kriging ensuring not trend and minimum variance.

Kriging maps presented on Figure 3 for NDVIsat and NDVIspec at two sampling date, and Figure 4 for leaf nitrogen content and cotton yield, show dark gray color for the highest values, while light gray color represent the lowest values. Despite the fact of that the fertilization was uniform, both methods reveal distinct plant growth patterns within the fields for each sampling date.
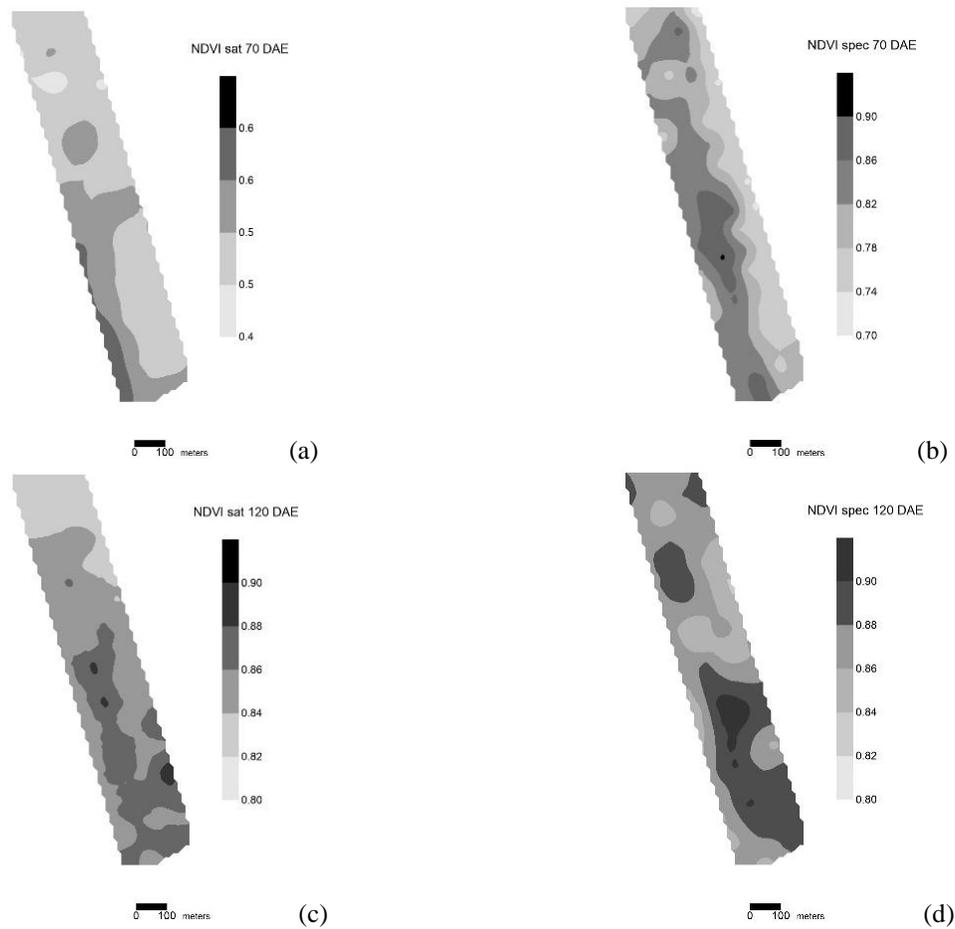

Figure 3. Isoline maps showing spatial distribution of: (a) NDVI by satellite image at 70 DAE; (b) NDVI by spectroradiometer at 70 DAE; (c) NDVI by satellite image at 120 DAE; (d) NDVI by spectroradiometer at 120 DAE.

Leaf nitrogen content is a good indicator of healthy plants, which have a dark gray color on Figure 3. Satellite images at 70 DAE presented stressed plants, with NDVI in light gray tone. This low NDVIsat is probably because of large areas with soil exposure, since average plant height was just 0.603m. The low resolution of AWiFS sensor was not good to estimate leaf N content at this sampling date. On the other hand, NDVI obtained by spectroradiometer represented the complete distribution of leaf N content. The best values of NDVIspec were distributed on the west side of study area, which agreed with the nitrogen spatial distribution.

At 120 DAE, both NDVI kriging maps showed similarities in the spatial distribution, where the highest and lowest values practically occur in the same position of the study area. These similarities also occur with NDVI maps to 70 DAE and the cotton yield map. In this case, both NDVI are good indicators of cotton yield especially at 70 DAE.
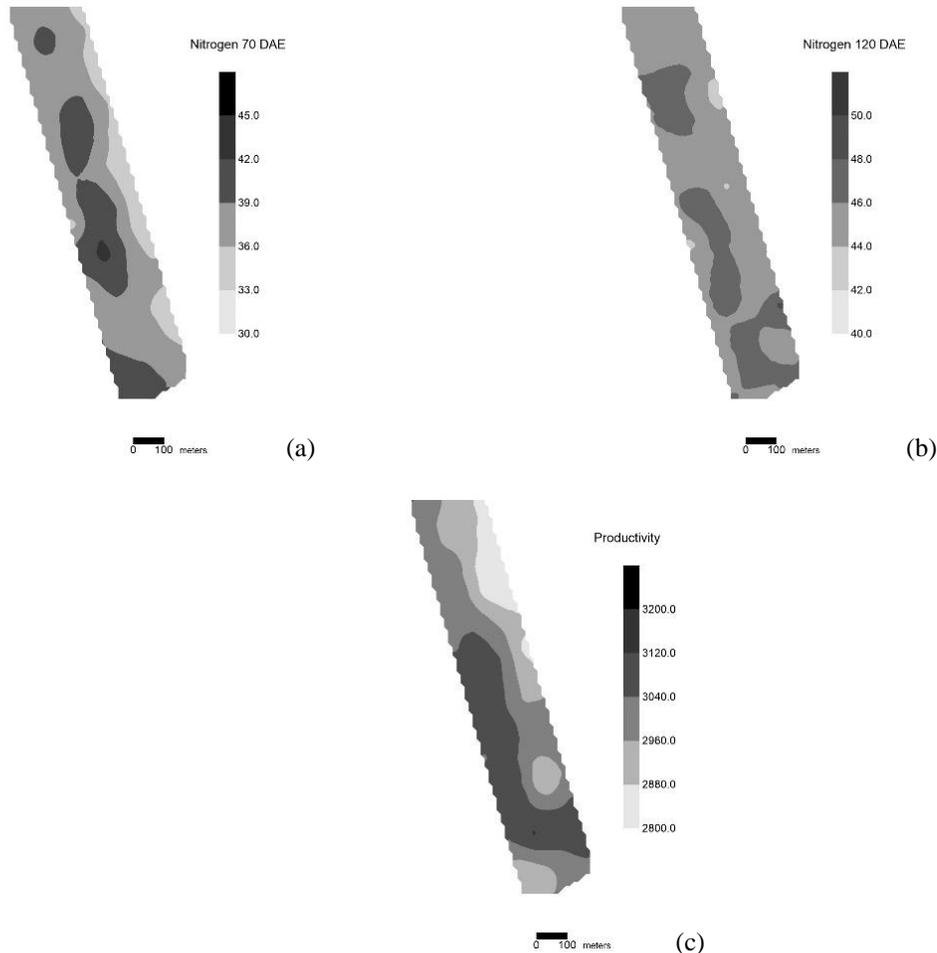

Figure 4. Isoline maps showing spatial distribution of: (a) Leaf N content at 70 DAE; (b) Leaf N content at 120 DAE and (c) Cotton yield.

4. CONCLUSIONS

There was spatial dependence for all data analyzed, but the spatial dependency to 70DAE was higher than the 120 DAE;

Similarities observed in the spatial distribution at 70 DAE to both satellite and spectroradiometer NDVI maps, as well as cotton yield map, indicate that NDVI data can predict cotton yield at first flowering;

Identifications of spatial differences were possible using geostatistical methods with remote sensing data obtained from medium resolution satellite images, allowing to identify distinct stages of plant growth and also to predict the cotton yield.

REFERENCES

[1] Aubert, A. B., Schroeder, A. and Grimaudo, J., "IT as enabler of sustainable farming: an empirical analysis of farmers' adoption decision of precision agriculture technology," Decision Support Systems 54(1), 510–520 (2012).

[2] Eitel, J.U.H., Long D. S., Gessler, P. E. and Hunt, E. R., "Combined spectral index to improve ground-based estimates of nitrogen status in dryland wheat," Agronomy Journal 100(6), 1694-1702 (2008).

[3] Zarco-Tejada P. J., Ustin S. L. and Whiting M. L., "Temporal and spatial relationships between within-field yield variability in cotton and high-spatial hyperspectral remote sensing imagery," Agronomy Journal 97(3), 641-653 (2005).

[4] Rosolem, C. A. and Mellis, V. van. "Monitoring nitrogen nutrition in cotton," Revista Brasileira de Ciência do Solo," 34(5), 1601-1607 (2010).

[5] Rochester, I. J., Constable, G. A., Oosterhuis, D. M. and Errington, M., [Flowering and Fruiting in Cotton]. In: Oosterhuis D., M., and Cothren J., T., (Eds.). The Cotton Foundation Publisher, Cordova, Tennessee, USA, 35-50 (2012).

[6] Read, J. J., Tarpley, L., McKinion, J. M. and Reddy, K. R., "Narrow-waveband reflectance ratios for remote estimation of nitrogen status in cotton," Journal of Environmental Quality, 31(5), 1442-1452 (2002).

[7] Zhao, D., Reddy, K. R., Kakani, V. G., Read, J. J. and Koti S., "Selection of optimum reflectance ratios for estimating leaf nitrogen and chlorophyll concentrations of field-grown cotton," Agronomy Journal 97(1), 89–98 (2005).

[8] Oliver, M. A. and Webster, R., "A tutorial guide to geostatistics: Computing and modelling variograms and kriging. Catena," 113(1), 56–69 (2014).

[9] Seshadri, K.; Rao, M.; Jayaraman, V.; Thyagarajan, K. and Murthi, K., "Resourcesat-1: A global multi-observation mission for resources monitoring," Acta Astronautica, 57(8), 534-539 (2005).

[10] Whelan, B. M., McBratney, A. B. and Minasny, B., "Vesper - Spatial Prediction Software for Precision Agriculture," Proc. European Conference on Precision Agriculture - ECPA, 3, 139-144 (2001).

[11] Vieira, S. R.; Millete, J. A.; Topp, G. C. and Reynolds, W. D., [Handbook for Geostatistical analysis of variability in soil and meteorological paramaters]. In: Tópicos em Ciência do Solo. Alvarez, V. H Publisher, Viçosa, MG, Brazil, 1-45 (2002).

[12] Vieira, S. R., Carvalho, J. R. P. and González, A. P., "Jack knifing for semivariogram validation," Bragantia 69(1), 97-105 (2010).

[13] Balzarolo, M., Anderson, K., Nichol, C., Rossini, M., Vescovo, L., Arriga, N., Wohlfahrt, G., et al., "Ground-Based Optical Measurements at European Flux Sites: A Review of Methods, Instruments and Current Controversies". Sensors, 11(1), 7954–7981 (2011).

[14] Yamada, T., Malavolta, E., Martins, O. C., Zancanaro, L., Casale, H. and Baptista, I., [Teores foliares de nutrientes observados em áreas de alta produtividade], Potafos, Piracicaba, SP, Brazil, 1-25 (1999).

[15] Cambardella, C. A., Moorman, T. B., Nowak, J. M., Parkin, T. B., Karlen, D. L., Turco, R. F. and Konopka, A. E., "Field-scale variability of soil properties in central Iowa soils," Soil Science Society of America Journal 58(1), 1501-1511 (1994).