MAPPING SOYBEAN AREAS WITH REMOTE SENSING DATA IMAGES USING A SIMPLIFIED BAYESIAN NETWORK

Marcio Pupin Mello[1]
Marcos Adami[1]
Anibal Gusso[1]
Gustavo Bayma Siqueira Silva[1]
Rodrigo Rizzi[2]
Bernardo Friedrich Theodor Rudorff[1]
Leila Maria Garcia Fonseca[1]

[1]Instituto Nacional de Pesquisas Espaciais – INPE
Av. dos Astronautas, Jd. Granja, 12227-010 – São José dos Campos-SP, Brazil
emails: {pupin, adami, anibal, bayma, bernardo}@dsr.inpe.br ; leila@dpi.inpe.br

[2]Universidade Federal de Pelotas – UFPel
Campus universitário S/N 96001-970 – Pelotas-RS, Brazil
e-mail: rodrigo.rizzi@ufpel.edu.br

1. INTRODUCTION

The Brazilian Institute of Geography and Statistics (IBGE) and the National Supply Company (CONAB) are the Brazilian official agencies for agricultural statistics. Currently, these agencies carry out Brazilian crop area estimation survey subjectively, based on information of agricultural agents, that do not enable quantitative analysis of the error [1]. An alternative to reduce the current area estimation subjectivity is with remote sensing (RS) images and geographic information systems (GIS) integration, using either sampling ([2]) or mapping ([3]).

With both RS and GIS tools it can be considerate a large number of data inputs (i.e., slope, vegetation indices, etc) that are related to the occurrence of some crop, assembly the probability model as a Bayesian Network (BN) approach [4]. However, few researches have mentioned the use of BN techniques applied to the RS image classification (e.g., [5]). The mathematical modeling of Bayes theorem utilizes the probability calculations based on prior knowledge and probabilistic conditionality, i.e., it connects the rational inference (posterior probability) to the subjective (a prior probability) and empirical knowledge (conditional probability). In another words, the Bayes theorem connects the human reason to the physical universe [6]. BN are directed acyclic graphs, i.e., representations of the causal relationships between random variables on probabilistic models [7].

Thus, this work aimed to propose an initial methodology to identify and map soybean crop using a Bayesian Network technique.

2. MATERIALS AND METHODS

2.1. Study area and materials

The study area (≈ 100,000 km^2) is located in the north portion of Rio Grande do Sul State, southern Brazil, limited by geographic coordinates W 50º40' to W 56º20' and S 27º03' to S 30º13' and at beginning of this decade comprised more than 90% of the State's soybean production. Soybean season goes from early November (sowing) to early April (harvest) with peak canopy development in mid February for most crop fields.

A prior investigative study has indicated that soybean growing season can be monitored by Enhanced Vegetation Index (EVI). In this study, we used the MOD13Q1 from MODIS sensor (tile h13v11) in order to investigate the EVI temporal behavior profile for the soybean growing season. At this phase it was identified two periods of minimum and maximum EVI values, which indicate the sowing and maximum canopy development of soybean fields, respectively. An image named $Minimum_{EVI}$ (M) was generated by the average of EVI values of early November. A second image was generated by using the average EVI values from January to mid-February (this image corresponds to the maximum EVI for soybean). The difference between the maximum and $Minumum_{EVI}$ images corresponds to the $Range_{EVI}$ image (R). The average was preferred instead single minimum or maximum for minimizing the noise interference proposes. Another input we used for mapping soybean areas was the slope (L) derived from Shuttle Radar Topography Mission (SRTM) data.

A reference map was generated by [8] using medium spatial resolution images from Thematic Mapper (TM) and Enhanced TM plus (ETM+) sensors on board of Landsat-5 and -7, respectively. Images were acquired at two key periods to identify soybean crop (early February and early March), during the same analyzed crop year (2000/2001). The reference map was resampled to the MODIS spatial resolution (250 m) for the accuracy assessment.

2.1. The Bayesian Network approach

The applied BN model in this research (Fig. 1) was chosen considering some meaningful simplifications. Each input variable (M, R and L) had its probabilities functions considered as Binomial where the success (*s*) for each variable was defined by satisfying some criterion. These criteria were $0.05 \leq m < 0.45$ for M, $r \geq 0.2$ for R and $s \geq 0.15$ for L. When a criterion did not achieved the evidence it was tagged as "failure" (*f*). The soybean variable (S) was considered only for the *s* case (S = *s*, i.e., soybean occurrence). Both the prior probabilities values and the joint probabilities were obtained by counting the pixels that satisfied the conditions (success or failure) for the input variables. Even that M has an intrinsic relationship with R, for example, all input variables were assumed to be independent for model simplification.

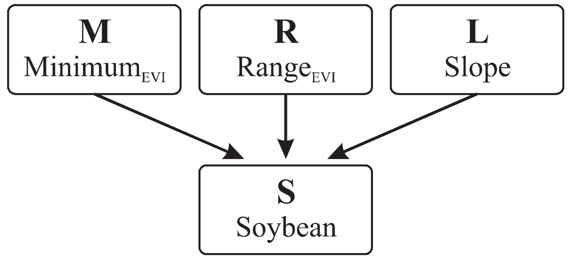

Fig. 1. Graphic representation of the Bayesian Network used.

Our main purpose in this research was to obtain the probability of each pixel in representing the soybean crop, given the conditions of $Minimum_{EVI}$, $Range_{EVI}$ and Slope. In formal notation: P (S / M , R , L) = ? According to Bayes theorem, we may write

$$P(S/M,R,L)=\frac{P(S\cap M\cap R\cap L)}{P(M\cap R\cap L)}. \tag{1}$$

Although the equation is the same for all pixels, the result matches to the evidences. Thus, for a pixel that satisfies the *s* criteria only for, e.g., M and L, (1) becomes

$$P(S/M=s,R=f,L=s)=\frac{P(S\cap M=s\cap R=f\cap L=s)}{P(M=s\cap R=f\cap L=s)}. \tag{2}$$

The calculation of all image will retrieve a probability map where its pixels values represent the probability of soybean occurrence.

3. RESULTS AND DISCUSSION

The probability values were obtained by counting the pixels that satisfied the determined conditions. Table 1 shows the probability values for each possible combination.

Table 1. Probability values for each evidences combination of the input variables.

	Combination number and Evidences							
Variable and Probs.	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
M	*s*	*s*	*s*	*f*	*s*	*f*	*f*	*f*
R	*s*	*s*	*f*	*s*	*f*	*s*	*f*	*f*
L	*s*	*f*	*s*	*s*	*f*	*f*	*s*	*f*
P (S ∩ M ∩ R ∩ L)	0.3657	0.0146	0.2583	0.0045	0.0402	0.0004	0.2379	0.0784
P (M ∩ R ∩ L)	0.2312	0.0057	0.0467	0.0022	0.0035	0.0001	0.0157	0.0022

The probability map (Fig. 2) was generated according to the criteria for each pixel on the input variables, following as (1).

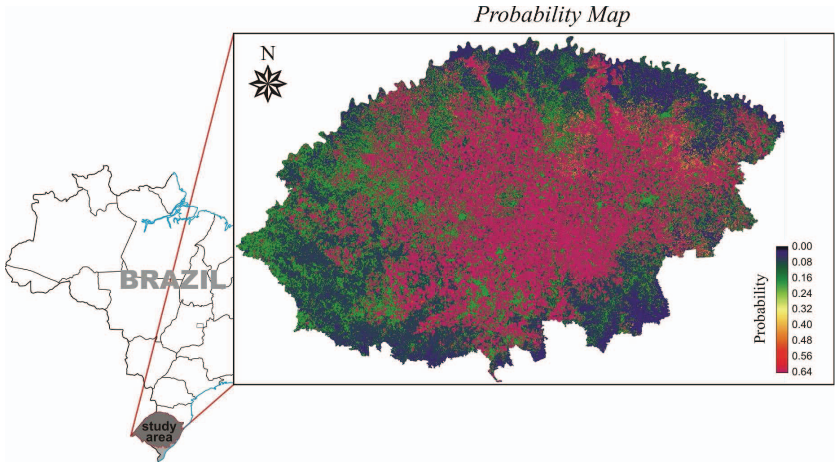

Fig. 2 – Probability map of soybean occurrence at study area.

In general, the probability map showed features visually very similar to the reference map. A slicing process was performed for probability map where pixels values greater than 60% were mapped as soybean crop. The similarity was assessed with a confusion matrix where the overall index was 79%. It was observed that the probability map overestimated in almost 20% the soybean area. Although inaccurate results the methodology indicates a new approach of potential for applications in the mapping of soybean crops.

4. CONCLUSION

This research presented an initial methodology to be applied in soybean identification and mapping using the Bayesian network technique. Despite all the simplifications, our approach has shown promise for mapping soybean areas in southern Brazil. In further researches, we suggest that the BN approach should considerate a more detailed model, with appropriate probabilistic models. It is expected that this issues improves the soybean crop areas identification and mapping.

5. REFERENCES

[1] F. A. Pino, “Estimação subjetiva de safras agrícolas,” *Informações Econômicas*, vol. 31, no. 6, pp. 55-58, 2001.

[2] M. Adami, M. A. Moreira, B. F. T. Rudorff et al., “Painel amostral para estimativa de áreas agrícolas,” *Pesquisa Agropecuária Brasileira*, vol. 42, no. 1, pp. 81-88, 2007.

[3] G. A. Ippoliti-Ramilo, J. C. N. Epiphanio, and Y. E. Shimabukuro, “Landsat-5 Thematic Mapper data for pre-planting crop area evaluation in tropical countries,” *International Journal of Remote Sensing*, vol. 24, no. 7, pp. 1521-1534, 2003.

[4] M. P. Mello, C. A. O. Vieira, L. A. Peternelli et al., "Redes Bayesianas no delineamento de culturas agrícolas usando informações contextuais," *in Proceedings of the 23th Congresso Brasileiro de Cartografia*. Rio de Janeiro, Brazil: SBC, 2007, pp. 1289-1295.

[5] P. V. Gorsevski, P. Jankowski, and P. Gessler, “Spatial prediction of landslide hazard using fuzzy k-means and Dempster-Shafer theory,” *GIS*, vol. 9, no. 4, pp. 455-474, 2005.

[6] S. D. Pena, “Bayes: "o cara"!,” *Ciência Hoje*, vol. 38, no. 228, pp. 22-29, 2006.

[7] E. J. M. Lauría, and P. J. Duchessi, “A methodology for developing Bayesian networks: An application to information technology (IT) implementation,” *European Journal of Operational Research*, vol. 179, no. 1, pp. 234-252, 2007.

[8] R. Rizzi, and B. F. T. Rudorff, “Estimativa da área de soja no Rio Grande do Sul por meio de imagens Landsat,” *Revista Brasileira de Cartografia*, vol. 57, no. 3, pp. 226-234, 2005.