Image Processing in Precision Agriculture

Dragoljub Pokrajac, Aleksandar Lazarević, Slobodan Vučetić, Timothy Fiez and Zoran Obradović

Abstract---- **A brief review of our signal and image processing application in precision agriculture is presented. A method for determining sampling frequency for agriculture data is proposed, and some initial results based on data simulation and image processing are reported.**

Keywords----**Knowledge discovery in databases, Precision agriculture, Image processing, Signal processing**

I. INTRODUCTION

In the past several years, increased attention has been directed towards the environmental impacts of agriculture [1,2]. A new way of farming called precision farming offers the promise of improving farm profitability and increasing environmental stewardship. Using technical advances such as the Global Positioning System (GPS), precision farmers can collect georeferenced yield, soil, and other important attribute data. The goal is to use these data to make site-specific crop production decisions where production practices are varied throughout a field.

Recent developments in Knowledge Discovery in Databases (KDD) techniques [3] may be extremely applicable to precision farming. An important part of the KDD process is data analysis and preprocessing. Namely, in agriculture we deal with data collected with various levels of resolution and accuracy. Therefore, it is important to optimize the sampling frequency of attributes in order to obtain the optimal cost/benefit balance, and also to minimize the influence of data noise and errors on the learning process. In this paper, we explore the possibility of using signal and image processing in achieving these goals.

II. PRECISION AGRICULTURE

Formally, the task of agriculture management can be stated as follows: Given a two-dimensional field F, and the set of features f_i, $i=1,\dots,m+n$, of which $f_1,\dots, f_m$ are *controllable* (their concentration can be increased by application of treatments $\Delta f_i \geq 0$ $i=1,\dots,m$), determine treatment vector $\Delta f_i(s)=[\Delta f_1 \dots \Delta f_m]^T$ on points $s=(x,y)\in$F, that maximize profit defined as:

$$\mathrm{Profit(F)} = \iint_F \left[c\Delta Yield(\mathbf{s}) - \mathbf{w}^T \Delta \mathbf{f}(\mathbf{s}) - \mathrm{w}_0\right] ds \qquad (1)$$

Dragoljub Pokrajac, Aleksandar Lazarević, Slobodan Vučetić and Zoran Obradović are with Washington State University (WSU), School of Electrical Engineering and Computer Science, Pullman WA 99164, USA. Timothy Fiez is with WSU, Department of Crop and Soil Sciences, Pullman WA 99164, USA. E-mails:{dpokraja, alazarev, svucetic, zoran, tfiez}@eecs.wsu.edu

Dragoljub Pokrajac is on leave from Telecommunications Laboratory, Department of Electronics, University of Nis, Beogradska 14, 18000 Nis,Serbia.

where c is the unit price of the crop, ΔYield(s) is the increment of crop yield due to the treatments, $\mathrm{w}=[\mathrm{w}_1\ \mathrm{w}_2 \dots \mathrm{w}_m]^T$ is a vector of prices per unit of a particular treatment, and w_0 is a fixed unit cost of agriculture management.

In traditional agriculture, treatments are constant for the whole field and are chosen according to the prior knowledge of a practitioner and/or the analysis results of sparse soil samples (typically only one aggregate soil analysis per field). On the other side, the task of precision agriculture is to achieve *site specific* recommendation for treatments, on disjoint regions $B_k \subset F$, such that

$$\mathrm{Profit(F)} = \sum_{B_k \subset F} \mathrm{Profit}(B_k) \qquad (2)$$

and assuming that *Profit*(B_k) in particular regions can be optimized independently and there is no interaction among treatments in distinct regions.

Through the KDD process, the learning algorithm is supplied with data sets F_l from different fields in possibly different geographic regions and years. Each data set consists of features $f_{i,l}(s_l)$, i=1,m and yield data $Yield_l(s_l)$ sampled on points $s_l \in F_l$. The task of the learning algorithm is to estimate from data the function:

$$\mathrm{Yield(s)} = \mathrm{Yield}(f_1,\dots,f_m,f_{m+1},\dots, f_{m+n}) \qquad (3)$$

such that profit optimization can be performed on a given field F. the inductive bias is that there is "enough similarity" between observed fields F_l and a given field F.

The basic problems of data acquisition and preprocessing include:

- determination of sampling points, such that sampling is economically acceptable while providing enough information for successful yield prediction
- development of strategies to combat measurement errors and errors due to sampling

For different types of agriculture data, resolution and accuracy are given in Table 1.

TABLE 1
PROPERTIES OF DIFFERENT AGRICULTURE DATA TYPES

type of data	resolution	accuracy
crop yield	high	low
topographic	high	medium /high
remote-sensing	medium	medium
soil sampling	low	high

Data containing crop yield information on field points are collected through the procedure of harvested crop mass measurement within a field [4], while the information about location is determined by GPS [5]. The main causes of error on yield data [6] are: variable width of crop entering combine header, variable time lag, the location error of GPS and yield sensor accuracy. To compensate for the effect of variable crop width, the usage of potential mapping is suggested [6], which is in fact 2D low pass filtering with non-rectangular windows [7]. Time lag can be modeled using low-pass 1-D filter [8,9]. Finally, the influence of sensor accuracy is assessed in [10]. An interesting property of sensor error is that its variance *decreases* with lower resolution, which is exactly opposite to behavior of fractal noise [11]. Although different sources of noise are discussed in literature, the question of an optimal filter to minimize the influence of noise on measured yield is not addressed. Instead, 2-D median filters with different window widths are applied [7].

Topographic data, such as slope, curvature of profile, tangent and planar plane and aspect, are generated using data obtained by GPS through a terrain analysis procedure, [12]. Topographic attributes influence the hydrological characteristics of field [13,14], hence their impact on precision agriculture. The accuracy of topographic data is determined by the accuracy of GPS and by the precision of topographic analysis algorithms. Since high accuracy GPS data are relatively non-expensive [15] and high quality topographic analysis methods are available, the remaining issue of GPS sampling resolution arises, and is discussed in [16].

Remote-sensing data [17] are obtained by observation from airplanes and satellites using photography or radiometric sensors. The resolution of remote-sensing data depends on sensor altitude and the quality of the instrumentation (lenses, etc) and upper limits are imposed by atmospheric effects and the influence of crop shadows [18]. To compensate for dynamic range and calibration problems, as well as to increase particular image properties such as edges and contrast, classical methods of image processing [7] are applied.

Soil sample data describe chemical and physical properties of soil, such as the concentration of nutrients and soil type. Often, a relatively small number of sample points are collected due to the high cost of sampling and laboratory analysis. Hence, the choice of an optimal sampling strategy is crucial to achieve data applicable for KDD purposes [18].

A number of papers regarding sampling frequency of soil tests have been published [19]. They are mainly based on sampling of experimental fields at several different frequencies and comparing the interpolation results obtained by spatial data interpolation techniques [20] with true samples. This approach has two disadvantages: namely, in practice, it is possible to replicate only a few different sampling frequencies, and the results obtained on one experimental field are not generally comparable with results on other fields. Different experiments with soil sampling density varying from 10*10 to 200*200m have been conducted [19], but no conclusive recommendation were made.

Besides classical interpolation techniques, such as kriging [21] inverse distance approximation [20], and spline [22], the application of signal processing theory on interpolation is also addressed in the literature, [23]. There, a 2-D spectral representation of a feature is applied and the spectrum coefficients are determined such that the sum of the interpolation errors at sampling points, weighted by the sizes of corresponding Voronoi polygons [24], is minimal. Although this method is applied to geostatistical data [25], it appears that this has not been applied to agricultural soil sampling data.

III. Signal Processing and the Choice of an Optimal Soil Sampling Strategy

There are several similarities between agriculture data and data originating from image processing:

- both images and agriculture data can be observed as random 2-D signals
- spatial correlation of image pixels is analogous to spatial correlation of agricultural samples, where the correlation is usually expressed by variograms [20]
- picture layers (R,G and B signals) can have some degree of mutual correlation which is often the case with particular agricultural feature layers

On the other side, soil data obtained in a controllable way and with resolution high enough to enable successful experimentation on sampling frequency are sparse due to the high cost of collecting and analyzing soil samples. Therefore, our research is based on a combination of image analysis and data simulation and consists of the following phases:

- Simulate spatial data satisfying different theoretic models a using spatial data simulator [26];
- For each of these models, estimate optimal sampling frequency on simulated data, using standard techniques of image analysis [7,27];

IV. Results

Currently, we are developing methods for the estimation of an optimal sampling frequency, [28]. Here, we present results from the data simulation phase.

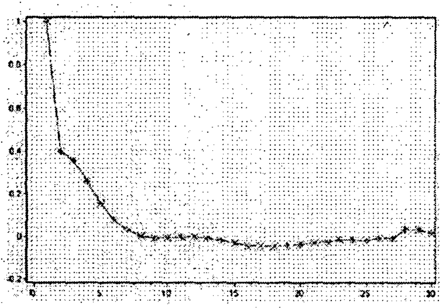
Several simulated data sets were generated and analyzed using methods and software described in [26,29]. Each data set consisted of 1681 simulated data samples taken from a $400*400m^2$ field using a $10*10m^2$ uniform sampling grid. Spatial correlation was determined by spherical semivariograms [20], where range and relative nugget were varied.

For zero-nugget data, the semivariogram range effect (R=50m and R=200m), on spatial correlation is shown in Fig. 1 where 1-D correlograms, assuming spatial isotropy [20] are computed. We can see that the range of the spherical variogram model determines the value of the spatial lag after which points are practically uncorrelated and, as it is known in the theory of random spatial signals, higher variogram range means higher spatial correlation.

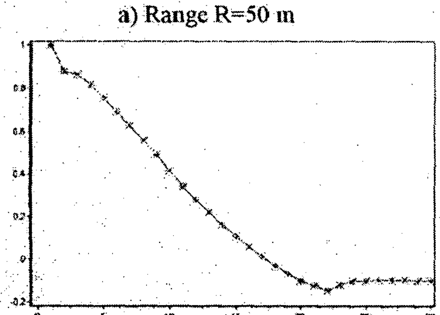
The influence of the relative nugget on spatial data behavior is shown on Figure 2, for range R=200m. The higher nugget means a higher uncorrelated component of a spatial signal, which is reflected as lesser correlogram values for lags smaller than the range.

To examine the relation between the spatial correlation of a feature and the correlation among features, we varied the range of our simulated data, generating 10 datasets with 1681 samples for each range value. Then, means and standard deviations were computed for the correlation coefficients between each of $\binom{10}{2}=45$ pairs of those datasets. Results, shown on Table 2, suggest that the mean value of the correlation coefficients increases with the range of the features. This opens the question of feature independence and suggests application of co-krigging methods for interpolation [20]. In this case it is interesting to examine whether it is possible to use a 3-dimensional Fourier transformation to represent correlated 2-D features [30,31,32,33,34].

Finally, we simulated the influence of sensor error as a multiplicative Gaussian noise with a 15% 3-σ interval and applied a 3*3 median filter to remove noise. Results, shown on Fig. 3, indicate that this method can help to suppress the influence of noise. Research on determining the possible adverse effects of high frequency component suppression on crop yield prediction is in progress.

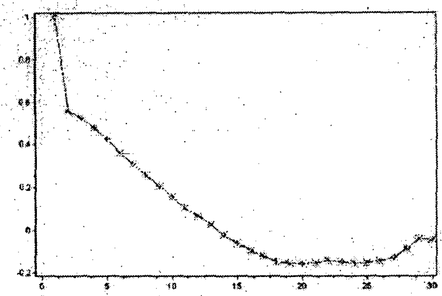


a) Range R=50 m

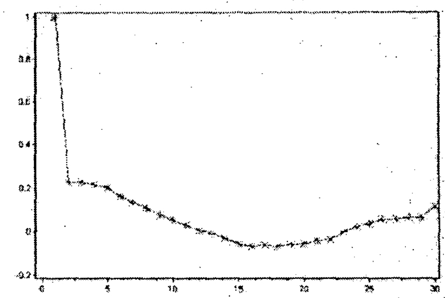


b) Range R=200 m

Fig. 1. Correlograms of simulated features with different variogram ranges (lag is presented in units of 10m)

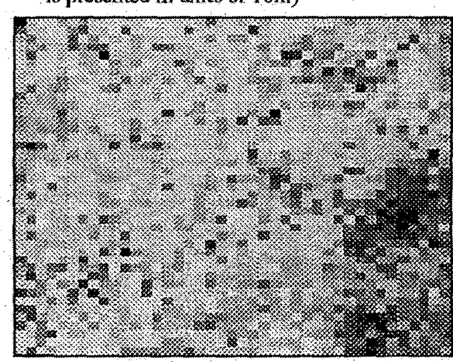


a) Nugget=0.2

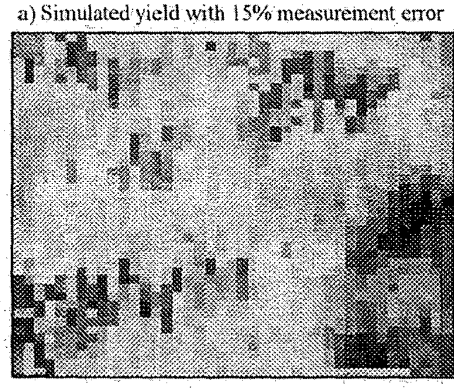


b) Nugget=0.8

Fig. 2. Influence of relative nugget on the correlograms of spatial data with R =200m(lag is presented in units of 10m)



a) Simulated yield with 15% measurement error



b) Filtered yield using 3*3 median filter

Fig. 3. The influence of median filtering on noise reduction of simulated yield

TABLE 2
MEANS AND STANDARD DEVIATIONS OF ABSOLUTE VALUES OF CORRELATION COEFFICIENTS BETWEEN SIMULATED FEATURES WITH SPECIFIED RANGE

range	0m	50m	100m	200m	300m
mean	0.015	0.040	0.094	0.176	0.202
std	0.014	0.030	0.071	0.147	0.142

V. CONCLUSION

Precision agriculture is a multidisciplinary field, which incorporates numerous science areas, such are agronomy, computer science, statistics, economics, environmental science, automatic control, telecommunications and microelectronics. The purpose of this paper is to emphasize the role that signal processing methods and approaches could have in achieving the ultimate goal of precision agriculture, and at the same time to introduce this emerging scientific branch to a signal processing and telecommunication engineers' community.

ACKNOWLEDGEMENT

Partial support by the INEEL University Research Consortium project No.C94-175936 to T. Fiez and Z. Obradovic is gratefully acknowledged.

REFERENCES

[1]G. Hergert, W. Pan , D. Huggins,J. Grove ,T. Peck , "Adequacy of Current Fertilizer Recomndation for Site-Specific Management," In F. Pierce, *The state of Site-Specific Management for Agriculture*, American Society for agronomy, Crop Science Society of America, Soil Science Society of America,pp. 283-300, 1997.

[2]R. Drossu R., T. Fiez, *A Framework for Yield Data Analysis in Precision Farming*, Pullman: WSU, 1997.

[3]P. Adriaans, D. Zantinge, *Data Mining*, Reading, MA: Addison Wesley Longman. 1996.

[4] F.J. Pierce at al, "Yield mapping," in *The site specific management for Agricultural systems*, 1997 ASA-CSSA-SSSA

[5] B. Hofmann-Wellenhof, H. Lichtenegger, J. Collins, *Global Positioning System: Theory and Practice*, New York NY: Springer Verlag, 1997

[6] B.S. Blackmore, C.J.Marshall, Yield mapping, Errors and Algorithms", in *Precision Agriculture, Proceedings of 3rd Int'l Conf.*, Minneapolis MN, June 1996, pp. 403-415

[7] M. Topalovic, *Televizija*, Radio-televizija Srbije 1993.

[8] J.V.Stafford at al, "Mapping and interpreting the yield variation in cereal crops," *Computers and Electronics in Agriculture*, vol 14 pp. 101-119, 1996.

[9] M. V. Popovic, *Digitalna obrada signala*, Beograd: Nauka, 1994.

[10] B. Missotten, G.Stubbe, J. De Baerdemaeker, "Accuracy of grain and straw yield mapping," *Precision Agriculture, Proceedings of 3rd Int'l Conf.*, Minneapolis MN, June 1996, pp.713-722

[11] M.F.Barnsley at al, *The science of fractal images*, New York NY: Springer Verlag, 1988

[12] J.C. Gallant, J.P. Wilson, "TAPES-G: A grid-based terrain analysis program for the environmental sciences," *Computers and Geosciences*, vol. 22, pp 713-722, July 1996.

[13] I.D.Moore, R.B. Grayson, A.R. Ladson, "Digital terrain modelling: a review of hydrological, geomorphological and biological applications," *Hydrological processes*, vol. 5, pp.3-30, Jan. 1991.

[14]D.J.Pennock, E.de Jong, "The influence of slope curvature on soil erosion and deposition in hummock terrain," *Soil science*, vol. 144, pp.209-217, March 1987.

[15] S.P.Saunders, G. Larscheid, B.S.Blackmore, J.V.Stafford, "A method for direct comparison of differential global positioning system suitable for precision farming," *Precision Agriculture, Proceedings of 3rd Int'l Conf.*, Minneapolis, MN, June 1996, pp. 663-680

[16] J.Gallant, M.Hutchinson, "Towards an understanding of landscape scale and structure," CD -ROM The Third Int'l Conf./Workshop on Integrating GIS and Environmental Modeling, Santa Fe 1996, NCGIA Publications Office, Santa Barbara, also on http://www.ncgia.ucsb.edu/conf/SANTA_FE_CD-ROM/sf_papers/gallant_john/paper.html

[17] R. A. Schowengerdt, *Remote sensing, models, and methods for image processing*, San Diego,CA : Academic Press, 1997

[18]M.S.Moran, U.Inoue, E.M.Barnes, "Opportunities and limitations for image-based remote sensing in precision crop management," *Remote.sens.environ.*, vol. 61, pp.319-346, 1997

[19]*Precision Agriculture, Proceedings of 3rd Int'l Conf.*, June 23-26, 1996, Minneapolis, Minnesota.

[20] N. Cressie, *Statistics for spatial data*, New York, NY: John Willey and Sons, Inc, 1993.

[21] E.H. Isaaks, R.M. Srivastava, *An Introduction to Applied Geostatistics*, London: Academic press, 1989.

[22] M. Voltz, R.Webster, "A comparison of kriging, cubic splines and classification for predicting soil properties from sample information," *The Journal of soil science*. v. 41 (3) p. 473-490,Sept 1990.

[23] T.Strohmer, "Computationally Attractive Reconstruction of Bandlimited Images from Irregular Samples," *IEEE Trans. image processing*, vol. 6, pp. 540-548. Apr. 1997.

[24] R. Duda, P. Hart, *Pattern Classification and Scene Analysis*, New York, NY: John Willey and Sons, Inc, 1973.

[25] M. Rasuth, "Application of 2-D methods for scattered data approximation to geophysical data sets," in *Proc. SampTA-Sampling Theory and Applications*, Riga/Latvia, Lithuania, 1995, pp.38-43.

[26] D.Pokrajac, T.Fiez, Z.Obradovic, "A spatial data simulator for knowledge discovery in agriculture," in review, *Precision Agriculture*

[27]A.V. Oppenheim, R.W. Schafer, *Discrete-Time Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1989

[28] S. Vucetic,T. Fiez, Z. Obradovic, "Analyzing the influence of data aggregation and sampling density on spatial estimation," in review, *Water Resources Research*

[29] A. Lazarevic, T. Fiez, Z.Obradovic, "A software system for spatial data mining andanalysis," submitted to *IEEE Hawaii Int'l Conf. on System Sciences*

[30] A. Jerri, "The shannon sampling theorem, its various extensions and applications: A tutorial review," *Proc. IEEE*, vol 65, pp. 1565-1596, Nov. 1997.

[31] N.Gaardner, "A note on multidimensional sampling theorem," *Proc. IEEE*, vol. 60, pp.247-248, Feb. 1972.

[32]D.Petersen, D. Middleton, "Sampling and reconstruction of wave-numbered limited function in N-dimensional Euclidean spaces," *Inform. Contr.*, vol. 5, pp.279-323, Dec. 1962.

[33]H. Miyakawa, "Sampling theorem of stationary stochastic variables in multidimensional space," *J.E.C.E. Japan*, vol. 42, pp.421-427, 1959.

[34]D.Petersen, D.Middleton, *Stochastic Fields and Signals*, in preparation, Cambridge, MA: MIT Press.