Determining Fixed Sensor Locations For Predicting The Spatial Distribution Of Ndvi Using Esap Software

Article · January 2008

DOI: 10.13031/2013.24735 · Source: OAI

Diaa Eldin M Elshikha
The University of Arizona

Douglas Hunsaker
United States Department of Agriculture

Andrew French
United States Department of Agriculture

2950 Niles Road, St. Joseph, MI 49085-9659, USA
269.429.0300 fax 269.429.3852 hq@asabe.org www.asabe.org

An ASABE Meeting Presentation

Paper Number: 083580

DETERMINING FIXED SENSOR LOCATIONS FOR PREDICTING THE SPATIAL DISTRIBUTION OF NDVI USING ESAP SOFTWARE

D. M. El-Shikha

Agricultural and Biosystems Engineering, University of Arizona, Maricopa Arizona USA

D. J. Hunsaker

USDA-ARS, Arid Land Agricultural Research Center, Maricopa, Arizona USA

S.M. Lesch

US Salinity Laboratory, USDA-ARS, Riverside, California, USA

T. R. Clarke

USDA-ARS, Arid Land Agricultural Research Center, Maricopa, Arizona USA

A.N. French

USDA-ARS, Arid Land Agricultural Research Center, Maricopa, Arizona USA

K. Thorp

USDA-ARS, Arid Land Agricultural Research Center, Maricopa, Arizona USA

**Written for presentation at the
2008 ASABE Annual International Meeting
Sponsored by ASABE
Rhode Island Convention Center
Providence, Rhode Island
June 29 – July 2, 2008**

Abstract. *A software package called ESAP (a salinity modeling software package) was used for designing the optimal sampling i.e., fixed sensors locations to estimate the spatial distribution of*

NDVI in a cotton field. Then it was used to estimate calibration equations and generate maps of the predicted and observed NDVI for the entire field. The ESAP software package consists of three programs: ESAP-RSSD, ESAP-Calibrate and ESAP-Mapping. The ESAP-RSSD was used to generate optimal sampling designs for NDVI spatial distribution estimation. The ESAP-Calibrate was used to estimate calibration equations (i.e. model parameter estimation). The ESAP-Mapper was designed to generate maps of survey and predicted data to compare predicted and observed NDVI data. Model estimation was based on 20 sampling locations. The main objective was to study the ability of the ESAP software to cost-efficiently estimate the spatial distribution of NDVI in a cotton field (i.e., obtaining a small number of sampling locations, consequently a small number of sensors needed to describe the field distribution of NDVI). The field NDVI spatial distribution was predicted for three different dates during the season using DOY163 as a survey to predict DOY176 (called group 1), using DOY226 as a survey to predict DOY246 (group 2) and using DOY246 as a survey to predict DOY261 (group 3). Regression of predicted versus observed NDVI data based on the ESAP sampling sites resulted in R-square values of 0.56, 0.51 and 0.81 for the three profile dates, respectively. These R^2 values and the associated MSE estimates (0.089, 0.079 and 0.045) compared well to the observed R^2 and MSE estimates based on the entire set of NDVI sample data (over 17,000 pixels). The results indicate that the ESAP model can be a reliable tool for locating a relatively small number of fixed sensors, whose combined NDVI data would allow a good prediction of the spatial distribution of NDVI for an entire field.

Keywords. Remote sensing, NDVI, Fixed sensors, Spatial modeling, Cotton

1. Introduction

Accurate assessment of the spatial distribution of crop indices, such as normalized difference vegetation index (NDVI), is an important parameter used in remote sensing energy balance estimation. The required remote sensing imagery can be obtained from aircraft and satellite platforms. However, routine imagery acquisitions from these platforms at the return intervals needed for day-to-day crop management are generally infeasible, as well as operationally expensive for most farming operations. An alternative approach to aerial imagery involves implementing a network of fixed sensors to collect continuous remote sensing data at certain locations within a field.

Cost-effectiveness being an important goal when designing a sensor network system, the lower the number of sensors the better. However, information from just a few raw sensor readings can be of little value if the sensors are not optimally located throughout the field (i.e., predictions generated from such data may be of low-quality). Therefore, the main challenge is to determine the sensor location that minimizes the cost of sensors and data collection, yet at the same time provides for good sensor coverage across a range of different field conditions. A major restriction to the use of remotely sensed imagery for effective management of agricultural resources is the lack of a reasonable procedure to translate imagery into maps of the crop or soil attributes of interest (Fitzgerald et al., 2006).

Grid-point and management zones are two common techniques for acquiring field spatial variability. In the first sampling approach, sample sites are located at specific grid points within the field (Chang et al., 2003). In contrast, the management zone approach is based on the assumption that a field is a mosaic of different zones, each have exclusive characteristics (Fleming et al., 2000). Then samples are collected from each management zone. Using either of these techniques or a combination of them is limited by cost. Additionally, problems such as spatial dependency should be accounted for, since spatially correlated residual errors can degrade the performance of estimated regression equations (Fitzgerald et al., 2006).

The ESAP-RSSD (Response Surface Sampling Design) software program was developed to generate optimal soil sampling designs from bulk soil electrical conductivity survey information (Lesch et al., 2000). It is a spatial site selection algorithm specifically designed to identify calibration sites that are well suited for multiple linear regression models. The main goal is to choose a subset of n calibration sites at which to sample the crop or soil characteristic of interest. The algorithm selects a limited set of calibration sites (6, 12, or 20) with desirable spatial and statistical characteristics by combining survey site location information with response surface design techniques. The algorithm ensures that the selected set of calibration sites is (1) spatially representative of the entire survey region and (2) suitable for estimating spatially referenced regression models. These regression models can in turn be used to predict the spatial values of the variable of interest across the field.

In this study, the main objective was to test the ESAP sampling algorithm for selecting NDVI sampling locations within an irrigated cotton field. More specifically, NDVI data from 20 ESAP chosen sensor locations were acquired and used to estimate regression models that could in turn predict the NDVI spatial distribution within the entire field. These estimated models and predictions were then compared to the corresponding models and predictions developed from the full set of NDVI readings (17,000+ pixels), in order to assess the reliability of the sampling algorithm.

2. Materials and methods

2.1 Field description

The experimental site was a 1.3-ha cotton field planted on 15 April in 2002 at the University of Arizona, Maricopa Agricultural Center (MAC) located 40 km south of Phoenix (lat 33°04′N, long 111°58′W, 361 m M.S.L.) The soil was a Casa Grande series (Typic Natriargid, fine-loamy, mixed, hyperthermic) with a predominantly sandy loam texture, and the volumetric soil water content at field capacity (θ_{FC}) and wilting point (θ_{WP}) for the upper 1.0 m of the soil profile were determined to be about 0.24 ±0.04 m^3 m^{-3} and 0.12 ±0.01 m^3m^{-3}, respectively (Post et al., 1988). The field was graded level and furrow irrigated from the east.

2.2 Remote sensing

Imagery was acquired using a Duncan MS3100[1] (Redlake Inc., San Diego, CA) with wavebands centered on 670, 720, and 790 nm with 10 nm bandwidths (full width half maximum). The Duncan camera had a 15° ×20° field of view resulting in a pixel resolution of about 0.4 m. Imagery was acquired about 60–90 min after solar noon, between 1315 and 1400 local time. Six imagery (dates) were considered in this study, which were divided into two groups: (1) survey images that were used to select a set of calibration sites, which should allow for accurate regression model parameter estimation (on DOY163, 226 and 246), and (2) Profile imagery (on DOY176, 246 and 261, respectively) which were used to extract observed data at the sample sites for prediction of NDVI spatial values for the whole field.

2.3 Image processing

Image processing, including geo-registration, masking, et cetera, was performed using ERDAS IMAGINE 9.1 software. Raw imagery was first converted to reflectance and then to NDVI using equation (1):

$$NDVI = \frac{\rho_{790} - \rho_{670}}{\rho_{790} + \rho_{670}} \tag{1}$$

Imagine software was used to delete borders before entering data into the ESAP software to avoid selecting sample sites outside of the planted areas. NDVI data files were saved as ASCII text files including three columns (i.e. X-coord., Y-coord., and NDVI values).

2.4 ESAP-software

The ESAP software package contains three programs: ESAP-RSSD (used to generate optimal sampling designs); ESAP-Calibrate (used to estimate calibration equations and predict the spatial values of the desired variable; i.e. NDVI in this study) and ESAP-SaltMapper, which can be used to produce spatial maps for both the survey and prediction data. All three programs were used in this study.

[1] Mention of company or trade names does not imply endorsement by the USDA.

2.4.1 ESAP-RSSD (Response Surface Sampling Design)

The ESAP-RSSD program can be used to generate optimal sampling designs for estimating spatially referenced regression models based on survey data. The program was originally designed for generating soil salinity sampling designs using bulk soil conductivity survey data (Lesch, 2005; Lesch et al., 2000; Lesch et al., 1995), but can in principle be used with any type of remote sensor information. Within the ESAP-RSSD program, the survey (NDVI) data can be visualized using different graphical techniques, transformed, decorrelated, and validated. In the decorrelation and validation processes, any outliers are identified and removed before a sample design is created. The software used a principle components analysis to decorrelate the survey data (when more than one signal reading is acquired at each site) and to identify outlier readings. The software used a modified response surface sampling design to select the sampling locations. More detailed information about the statistical methodology used in the ESAP-RSSD program is given in Lesch (2005).

The ESAP-RSSD program is also designed to locate sample sites as far apart as possible to minimize the spatial autocorrelation in the regression model residuals. (In most cases, the software can create an approximately uniform sampling pattern across the field, similar to a space-filling algorithm). Designs can be generated having sample sizes of 6, 12 or 20 sites. In this study, three survey files (groups) containing NDVI data acquired on DOY 163, 226 and 246 were used to generate three 20-location sampling designs.

2.4.2 ESAP-Calibrate

The ESAP-Calibrate program is designed to use remotely sensed survey data and co-located calibration sample data to predict the spatial distribution of a target soil or plant property (in this study, NDVI values). The software does this by estimating a spatially referenced regression model and then using this model to generate target property predictions across the entire field. Additionally, the regression model can be used to generate various field summary statistics, such as the average level of the target property in the field and/or the % area of the field that exhibits property values above a given threshold (referred to in ESAP as a "range interval estimate").

After determining the coordinates of the 20 sample locations based on NDVI data collected on days 163, 226, and 246, the corresponding NDVI values on days 176, 246 and 261, respectively, were extracted at these same locations. These values were then used as input "calibration" data for estimating the regression models for days 176, 246 and 261. This approach generated three groups of "survey" and "prediction" files, defined as follows:

1. Group 1: NDVI survey data from DOY163 and 20 NDVI calibration sites from DOY176 were used to predict the NDVI distribution for DOY176.
2. Group 2: NDVI survey data from DOY226 and 20 NDVI calibration sites from DOY24 were used to predict the NDVI distribution for DOY246.
3. Group 3: NDVI survey data from DOY246 and 20 NDVI calibration sites from DOY261 were used to predict the NDVI distribution for DOY261.

Upon importing the corresponding data files into the ESAP-Calibrate program, the software was used to estimate three simple linear regression models that could predict the NDVI readings for the latter dates using the NDVI readings acquired on the earlier dates. Predicted field averages and range interval estimates were calculated after the calibration equations (regression models) were determined. Three output (NDVI prediction) files were also created by the ESAP-Calibrate software that could in turn be imported into the ESAP-SaltMapper program.

2.4.3 ESAP-Mapper

The ESAP-SaltMapper program is designed to rapidly create 2D spatial raster maps of the target soil or plant property predictions. This program also contains an output ASCII text file feature which allows one to export the prediction data file as a generic ASCII text file (for use in more sophisticated mapping software applications, such as Surfer or ArcGIS). For this study, the ESAP-SaltMapper program was used to generate the final 2D raster maps of the predicted NDVI survey data for days 176, 246 and 261.

4. Results and discussion

Figure 1 shows a map of the sampling sites for predicting NDVI for DOY176. Twenty sites were chosen (site IDs are indicated on the map) using ESAP-RSSD software; note that the algorithm achieved a reasonably uniform sampling pattern across the field. Note also that about nine of the twenty locations were close to the edge of the analyzed NDVI plot data. However, this is not problematic, since the true NDVI edge data were removed using Imagine software before data was imported to ESAP-RSSD.

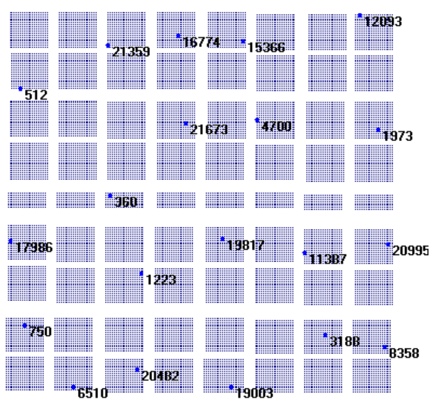

Figure 1. An example of the sampling map as an output from ESAP-RSSD for the prediction of NDVI for DOY176.

The regression model summary statistics are shown in Table 1 for all three sampling dates. A simple linear regression model was used in this study and each regression function was estimated using twenty calibration sites. The models for groups 1 and 2 exhibited lower R^2 values, 0.56 and 0.51, respectively, compared to 0.81 for group 3. Additionally, the MSE estimates were higher for the first two groups (average of 0.084) compared to the third group (0.045). Nevertheless, the MSE for all three dates were considered within an acceptable limit (<0.09) for most NDVI applications. In other words, ESAP can predict NDVI spatial values within 0.09 using the 20 sampling locations. Regression of predicted versus observed NDVI data (over 17,000 pixels) resulted in R-square values of 0.48, 0.63 and 0.73 for the three profile dates (DOY176, 246 and 261, respectively). The corresponding MSE estimates for these three dates were found to be 0.073, 0.057 and 0.046; these correspond reasonably well with the 20-site ESAP regression models discussed above.

Table 1: Model statistics

Group	R-square	Root MSE	Prob > F	Parameter estimates (std.error)	
				b_0	b_1
1	0.56	0.089	0.0001	0.378 (0.020)	0.076 (0.016)
2	0.51	0.079	0.0004	0.784 (0.018)	0.059 (0.014)
3	0.81	0.045	0.0001	0.768 (0.010)	0.066 (0.008)

An analysis of the ESAP regression model residuals suggested that the residual errors were normally distributed and exhibited no influential observations or outliers. The corresponding Moran tests for spatial independence were found to be non-significant for the residuals associated with groups 2 and 3, but significant below the 0.05 level for group 1. This latter test result implies that the intercept and slope parameter estimates for group 1, while still unbiased, are not estimated as efficiently as they could be (if additional data were collected and a geo-statistical modeling approach was adopted).

The ESAP predicted versus observed NDVI field averages are shown in Table 2 for all three sampling dates. All of the ESAP predicted means are statistically equivalent to the observed means; i.e., 95% confidence intervals for the predicted means contain the observed NDVI means (calculated using all 17,000+ observations).

The ESAP generated range interval estimates (RIE's) are shown in Table 3. Note that an RIE represents a prediction of the number (or percentage) of NDVI pixels falling within a specific range, after adjusting for the uncertainty in the regression model. For example, in group 1 (DOY176), we expect that 7.15% of the field should have NDVI values below 0.2, 50.67% of the field should exhibit values between 0.2 and 0.4, 38.08% of the field should exhibit values between 0.4 and 0.6, etc. For groups 1 and 2, these predictions were somewhat different from the actual (observed) values (as shown in table 3). However, for group 3 (DOY261), where the regression model produced much more accurate individual NDVI predictions, these RIE predictions were quite accurate.

Table 2: Predicted vs. observed NDVI field averages (including data at all sites).

	NDVI	**Mean**	**STD**	**Min**	**Max**
Group1	**Obs**	0.41	0.1007	0.03	0.71
	Prd	0.38	0.0764	0.14	0.69
Group2	**Obs**	0.79	0.0934	0.32	0.97
	Prd	0.78	0.0590	0.52	0.92
Group3	**Obs**	0.76	0.0885	0.36	0.93
	Prd	0.77	0.0665	0.47	0.89

Table 3: NDVI range interval estimates (RIE)

	RIE (DOY176)			RIE (DOY246)			RIE (DOY261)		
Interval	Actual (%)	ESAP (%)	Diff.	Actual (%)	ESAP (%)	Diff.	Actual (%)	ESAP (%)	Diff.
0.0 - 0.2	2.80	7.15	4.35	0.00	0.00	0.00	0.00	0.00	0.00
0.2 - 0.4	42.71	50.67	7.96	0.04	0.15	0.11	0.06	0.04	-0.02
0.4 - 0.6	51.39	38.08	-13.31	4.27	4.76	0.49	6.33	3.41	-2.92
0.6 - 0.8	3.10	4.00	0.90	38.30	49.59	11.29	55.16	58.32	3.16
0.8 - 1.0	0.00	0.11	0.11	57.40	45.50	-11.90	38.45	38.23	-0.22

The ESAP predicted versus observed NDVI maps for each sampling date are shown in Fig. 2. These side-by-side map comparisons indicate that the predicted reflects most of the features of the observed maps. In other words, there is generally a good correspondence between the observed and predicted maps, although the predictions are of course less variable (see Table 2). For example, note that while the maximum values of the predicted and observed NDVI readings were relatively close (i.e. difference of 0.02 to 0.05), the minimum readings was different by a magnitude between 0.11 and 0.20.

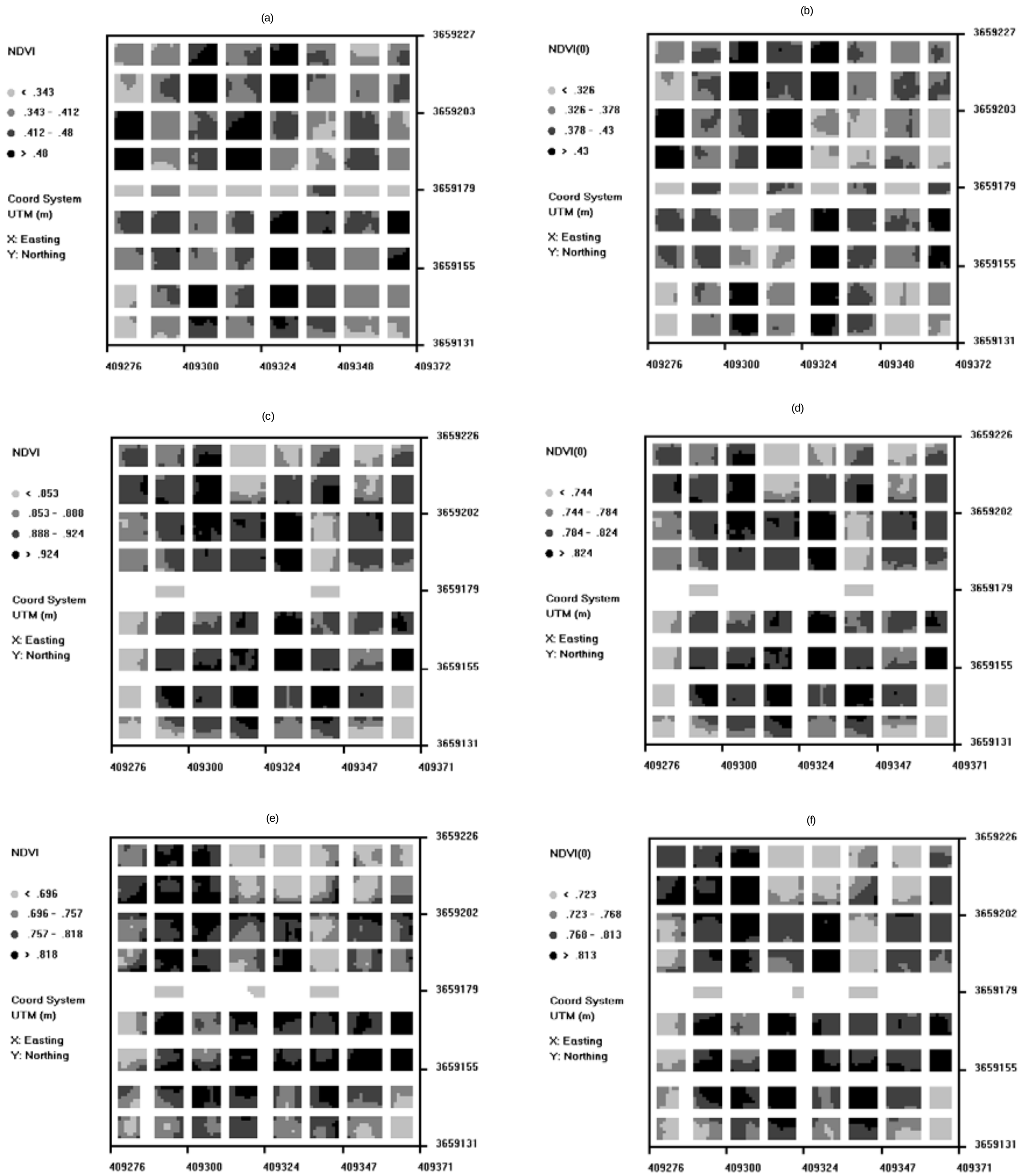

Figure 2. Maps of NDVI for (a) observed-group1, (b) predicted-group1, (c) observed-group2, (d) predicted-group2, (e) observed -group3 and (f) predicted-group3.

5. Conclusions

Using a relatively small sample size (20 locations), the ESAP software package was able to reliability estimate NDVI prediction equations for all three sampling dates. These models were capable of estimating the spatial NDVI distribution across the entire field with reasonable agreement between the observed and predicted values. These results suggest that the algorithm can be reliably used to generate sensor location that minimizes the cost of data collection, yet at the same time provide for good sensor coverage across a range of different field conditions.

References

Chang, J. Y., D. E. Clay, C. G. Carlson, S. A. Clay, D. D. Malo, R. Berg, J. Kleinjan, W. Wiebold. 2003. Different techniques to identify management zones impact nitrogen and phosphorus sampling variability. Agron. J. 95: 1550-1559.

Fitzgerald, G. J., S. M. Lesch, E. M. Barnes, W. E. Luckett. 2006. Directed sampling using remote sensing with a response surface sampling design for site-specific agriculture. Computers and Electronics in Agriculture 53: 98-112.

Fleming, K. L., D. G. Westfall, and W. C. Bausch. 2000. Evaluating management zone technology and grid soil sampling for variable rate nitrogen application [CD-ROM computer file]. In P.C. Robert et al. (ed.) Proc. of the 5th Int. Conf. on Precision Agriculture, Bloomington, MN. 16–19 July 2000. ASA, CSSA, and SSSA, Madison, WI.

Lesch, S. M., D. J. Strauss and J. D. Rhoades. 1995. Spatial prediction of soil salinity using electromagnetic induction techniques 1. statistical prediction models: a comparison of multiple linear regression and cokriging. Water Resources Research 31(2): 373-386.

Lesch, S. M., J. D. Rhoades, D. L. Corwin. 2000. ESAP-95 Version 2.01R: User Manual and Tutorial Guide. Research Rpt. 146, USDA-ARS, George E. Brown, Jr., Salinity Laboratory, Riverside, CA, USA.

Lesch, S. M. 2005. Sensor-directed response surface sampling designs for characterizing spatial variation in soil properties. Computers and Electronics in Agriculture 46: 153-179.

Post, D. F., C. Mack, P. D. Camp, A. S. Sulliman. 1988. Mapping and characterization of the soils on the University of Arizona, Maricopa Agricultural Center. In Proceedings of hydrology and water resources in Arizona and the Southwest (pp. 49–60). The University of Arizona, Tucson, AZ, USA.