The Modifiable Areal Unit Problem in Precision Agriculture

by Raymond Struthers and Bruce Erickson

One of the greatest promises of precision agriculture is in its potential to influence decision-making. There are an array of technologies that are associated with precision agriculture, but most aim to utilize some type of computerized mechanism to acquire data, perform some type of analysis, interpret the results, and then implement the decisions based on this information. These steps can be performed in rapid succession in one operation, but are more commonly done one by one over a period of time. But while the equipment and computational sides of precision agriculture are rapidly advancing in their capabilities, our understanding of managing the information collected is still, relatively, in its infancy. The capacity to collect, store, and analyze information in a more spatially and temporally intense way will continue to accelerate these concerns.

One way of managing spatially dense information is in the creation of management zones based on information that varies across a field such as soil test information, yield monitor results, or remote sensing information. Early precision adopters tried to identify management areas within fields most often using equal area grids or management zones based on soil physical and chemical properties or crop yields. Recently, management zones have been proposed as a method of identifying zones that may be consistently managed (Kitchen et al., 2006). It is important to understand how these management zones are created and the implications of creating arbitrary management zones.

A management zone is an “areal” unit, meaning “of an area”, and these can be modified to include different subsets of information, in site-specific agriculture this being different parts of a field—thus the term modifiable areal unit. The process of defining or creating management zones would be acceptable if it were performed using a fixed set of rules or if there were some geographically meaningful basis for creating them. The areal units used in many geographical studies are arbitrary, modifiable and dependent on whoever is doing the aggregating (Openshaw, 1984). Although there is almost an infinite number of different ways by which geographical data can be divided, data are normally presented and analyzed for one particular set of units ignoring other contributions.

Demonstrating the Consequences of Modifying Areal Units

A study was conducted to investigate correlation results by creating arbitrary management zones. These were based on Indiana corn yields and total applied nitrogen, per county, in 2005. The results of this study can be applied to management zones in a field if it is imagined that the state of Indiana is a farm field and each county is a data point that can be aggregated in various ways into management zones, such as Indiana’s nine crop reporting districts.

Corn yield data were collected in 2005 by the United States Department of Agriculture National Agricultural Statistics Service (USDA-NASS) and total nitrogen applied in 2005 by the Office of the Indiana State Chemist. Indiana’s nine crop reporting districts were aggregated two different ways (Figure 1). The first aggregation level was determined by subdividing the existing crop reporting districts into smaller more frequent zones from west to east resulting in twenty zones. The second aggregation level was determined by subdividing the crop reporting districts north to south, again resulting in twenty zones. There was no statistical or scientific motivation for creating the two aggregated zones. An ordinary least squares regression was performed for each of the four levels of aggregation to determine the statistical significance.

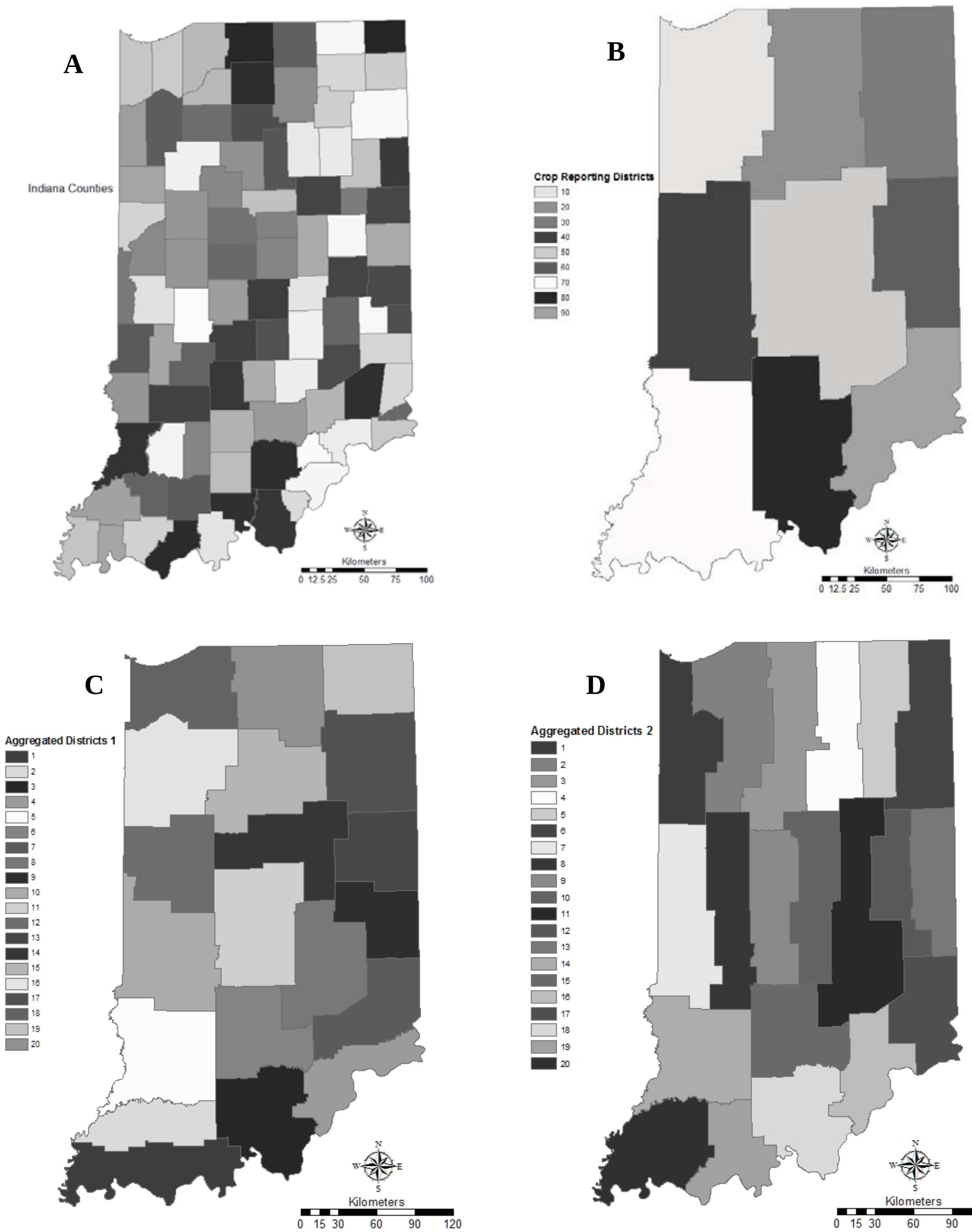

Figure 1. Four levels of aggregation. A) Indiana's ninety-two counties, B) Indiana's nine crop reporting districts, C) East-West aggregation of Indiana's crop reporting districts, and D) North-South aggregation of Indiana's crop reporting districts.

The regression analysis among various aggregation strategies resulted in vast differences in correlation coefficients between corn yield and total applied nitrogen. The most significant correlation exists for the east-west aggregation of crop reporting districts (Table 1). The north-south aggregation of districts shows a weak correlation. The state's 9 crop reporting districts show no significance, and the 92 counties in Indiana show only a weak relationship.

Table 1. OLS regression results for Indiana's 92 counties, 9 crop reporting districts, aggregation level #1 and 2 showing R2 and p-values.

Level of Aggregation	**Map Symbol on Figure 1**	**R^2**	**p-value**
Indiana's 92 Counties	A	0.14	***0.000
Indiana's 9 Crop Reporting Districts	B	0.00	0.975
East-West Aggregation of Crop Reporting Districts	C	0.58	***0.000
North-South Aggregation of Crop Reporting Districts	D	0.15	0.089

*** Significant at the 0.001 probability level.

Possible Solutions to the Problem

Given the influence of the modifiable areal unit problem in spatial analysis, developing a solution is important. A number of ways to understand the impact of the modifiable areal unit problem on spatial analysis have been suggested in the geographical literature (Openshaw and Taylor, 1981; Openshaw, 1984; Fotheringham, 1989). Openshaw (1984) stated that the simplest solution to the modifiable areal unit problem would be to pretend it does not exist and hope that the analysis will be meaningful. This has problems as the areal units being studied should be meaningful in some way relevant to the study. Even if it is possible to agree on a unique set of areal units, statistical variation such as those illustrated by Openshaw and Taylor (1979) will continue to be of interest in order to place a set of results into a meaningful statistical and spatial perspective.

Conclusions

Advances in equipment and computer capacity are allowing agriculturalists working with site-specific agriculture to amass huge amounts of data, hoping to turn that into useful information to aid decision-making. While it might be argued with data that having more is better than less, most would agree that the wrong information is less useful than no information at all.

The modifiable areal unit problem shows how conclusions can be radically different based on different approaches to managing data. In this example, the relationship between corn yields and rate of nitrogen fertilizer could be shown to be highly correlated, or not correlated at all.

References

Indiana State Chemist. http://www.isco.purdue.edu/. Verified 11/2006.

Fotheringham, A.S. 1989. Scale-independent spatial analysis. pp. 221-228. In M.F. Goodchild and S. Gopal (eds.). Accuracy of spatial databases, Taylor and Francis, London.

Kitchen, N.R., S.A. Sudduth and S.T. Drummond. 2006. An evaluation of methods for determining site-specific management zones. http://mpac.missouri.edu/pubs/SoilandCropVar.htm Verified July, 2006.

National Agricultural Statistics Service. http://www.nass.usda.gov/index.asp .Verified 11/2006

Openshaw, S. and P. Taylor. 1979. A million or so correlation coefficients: three experiments on the modifiable areal unit problem. 127-144. In N. Wrigley. (ed) Statistical Applications in the Spatial Sciences. Pion, London.

Openshaw, S. 1984. The modifiable areal unit problem. CATMOG 38. GeoBooks, Norwich, England.