2004 BELTWIDE COTTON CONFERENCES, SAN ANTONIO, TX – JANUARY 5-9

SPATIAL ANALYSIS OF PRECISION AGRICULTURE DATA: AN APPROACH TO IMPROVE MANAGEMENT ZONE DELINEATION PROCEDURES FOR TEXAS COTTON

Margarita Velandia, Roderick M. Rejesus, and Eduardo Segarra
Department of Agricultural and Applied Economics
Texas Tech University, Lubbock, TX
Kevin Bronson
Texas Agricultural Experimental Station
Texas A&M University
Lubbock, TX

Abstract

Cotton production is one of the most important agricultural enterprises in Texas. However, cotton production is subject to many constraints, such as nitrogen (N) fertilizer availability. Precision farming is an alternative technology that farmers can use to more efficiently apply N fertilizer in cotton. One of the most important issues considered in the implementation of precision agriculture, especially in variable rate N application, is the delineation of optimal management zones. The objective of this paper is to develop a method that could potentially give guidelines on how to use site-specific information to establish management zones. In particular, we build upon Exploratory Spatial Data Analysis (ESDA) techniques to develop management zone delineation procedures for precision farming. The data in this study is from an experiment developed in the Southern High Plains of Texas in 2002. The experiment was design primarily to study N use in cotton production. Based on this data, the preliminary methods developed in this study did not show any significant spatial patterns that can be used for delineating management zones. However, it is important to recognize that these results are still preliminary. More work in terms of data manipulation and analysis may be required to streamline the procedure for effective management zone delineation. Nevertheless, we believe that the method is a workable methodology to detect spatial patterns that can potentially help farmers and crops consultants in the definition of management zones.

Introduction

A large proportion of cotton in the U.S. is produced in Texas. Texas cotton production represents about 30% of the total cotton production in the U.S (National Agriculture Statistics Service [NASS], 2003). Also, cotton is one of the commodities that lead the agricultural industry in Texas in terms of production and generation of revenue. Cotton is also the fifth most important export in Texas (TDA, press release, 2001). Cotton production is a profitable enterprise but is subject to many constraints such as availability of water, nitrogen, and phosphorus (Wilson, 1998). Although, cotton growers can not control for the availability of water, they can control other factors by using appropriate production technologies. *Precision farming* is one alternative technology that farmers can use to control inputs better.

Precision agriculture combines different site–specific technologies such as Global Position System (GPS), computer-controlled variable rate technologies (CVRT), geo-referenced yield maps. A simpler technique such as soil sampling is also considered a precision technology (Khanna, et al, 1999). These techniques are used to get information about yield/soil characteristics at different points in a field. This information is typically used to establish potentially more efficient management strategies which consider the differences between the different locations within a field.

Therefore, one of the main issues in implementing variable rate precision agriculture is the delineation of optimal management zones. *Management zones* are geographical areas of a field that can be treated as homogeneous for certain characteristics that are common to them. The input application and general treatments can be managed as different for each zone using variable rate precision technology (Dillon, 2002).

The general objective of this article is to develop a method that could potentially give guidelines on how to use site-specific information (soil characteristics, pH, slope, etc.) to establish management zones; specifically for the application of variable rate precision technology. This method is developed based on *Exploratory Spatial Data Analysis* (ESDA) using information from fields under cotton production in the Southern High Plains of Texas.

Materials and Methods

The data used to develop the method to establish management zones will be based on precision yield data from an agronomic cotton experiment in the Southern High Plains of Texas. The data from the experiment is available for 2002 and the experiment was designed primarily to study nitrogen (N) use in cotton production. The different GPS points for which data was collected can be seen in Figure 1. The experiment consisted of a randomized complete block design with three replicates each replicate was within a center pivot irrigation span. There were three N treatments: variable-rate N, blanket-rate N and zero N. The design of the experiment can be seen in Figure 2.

In order to develop the method to establish management zones, we use the *Exploratory Spatial Data Analysis* (ESDA) approach. ESDA can be defined as a method that combines different techniques to visualize spatial distributions, identify patterns of different locations and association between these locations. This method is based on the concept of spatial autocorrelation that is the relationship between spatial units, and makes use of the concept of distance between locations. Spatial autocorrelation is the idea that points with similar values of a specific characteristic are near in space; therefore, positive autocorrelation means that certain points that are located close to each other share similar characteristics (Mensser & Anselin, 2002, p. 10).

ESDA has measures for spatial autocorrelation at the local and global level. In this work we are going to use measures at the global level using the *Moran's I* statistic. These statistics could potentially guide the farmers on the use of information to establish management zones.

Preliminary Results

In order to operationalize the ESDA approach for establishing management zones, we first choose a variable that would let us construct more compact zones. This would allow us to analyze the relationship between locations and values of different variables within these locations. According to Bronson et al (2003) studies have shown that there are different responses to N with regards to different landscape positions for different crops. In cotton, for example, Li et al (2001, 2002) showed that there is a negative correlation between the altitude from the ground and the accumulation of nitrogen on the soil (Bronson, 2003, p. 1). Based on this result we verify this information by running a simple OLS regression between the quantity of N (lb/acre in 0-6 inches) on the ground and the variables that define the landscape position. We found that there is a relation between these variables, and that they are all significant at the 5% level (Table 1). Therefore, we define three main sub-areas for spatial analysis based on landscape position: the bottom area, the south facing slope area, and the north facing slope area. The lower elevation in the bottom area should have systematically higher levels of N accumulation compared to the south and north facing areas that have higher elevation.

Next, we graphically examine whether there are visual evidence of clustering for several key variables that could help establish management zones. The key variables examined are: yield (in lbs/acre), soil N, pH, percentage of sand in the soil, percentage of clay in the soil, soil biomass, and soil electrical conductivity. We first used a *choropleth* map for the initial graphical analysis. A *choropleth* map is a map drawn by Arcview that shows the spatial distribution of points for each variable. The *choropleth* map of yield (Figure 3), defined as handpicked cotton lint after ginning, in lbs/acre, shows that points within the same level curve described by the center-pivot irrigation span have dissimilar yield patterns (i.e. these points are within 2 sq. meters of each other). This result suggests that there is no clustering in terms of cotton yield. This may be due to the design of the experiment where the points within the same level curve are treated under different technologies (variable rate and homogeneous rate of N applications).

Other soil characteristics investigated also visually showed no evidence of clustering. With the information shown by the *choropleth* map for the bottom area, we decided to drop pH as a variable of analysis because there is no significant variation in the data (i.e. all the points has a pH value of eight except for two outliers). There is no visual clustering pattern for sand percentage in the soil the points in the bottom area based on the *choropleth* map. The percentage of sand is between 82% and 86%, and all the points seem to be evenly distributed within the whole area and the points within the same level curves also do not show evident clustering patterns. Even points within a distance of 4 square meters do not show clustering patterns in terms of the variable percentage of sand in the soil. For the variable percentage of clay, the points in the bottom area also do not show significant clustering. The variables N in the soil (Figure 4), biomass, and electrical conductivity also show the similar results as the clay and sand variables. As in the bottom area, all the variables of interest do not seem to have significant clustering patterns in both the north and south facing areas. However, for the case of soil N, points within a distance of 3 square meters seems to show some clustering. In the interest of space, only the *choropleth* map for yield and soil N are reported here (see Figures 3 and 4) but all *choropleth* maps are available from the authors upon request.

The preliminary visual results seem to show that there is no spatial structure in the yield and soil characteristic variables within each landscape area. This result might be due to the size of each landscape area. The landscape areas are too big such that the sampling points are far enough apart and there is no specific spatial structure among them.

There are other maps suggested by Mensser and Anselin (2002) to visually inspect spatial data. One such map is a box map, which shows the position of a location within a distribution of the variable of interest. This graphic is helpful because it can show distributional characteristics of the variable of interest for the whole field. But according to Mensser and Anselin, "they are limited in their ability to identify any significant *spatial* clustering." Using box maps we verify the information given by the *choropleth* maps. The box maps are defined for the whole field and not for each landscape position area. The box map for the pH shows that all points within the field are approximately the same. For the percentage of sand, the low and high outliers are not close to each other (Figure 5). The box map for percentage of sand also shows that there are no similarities for the points close to each other. With respect to the variable percentage of clay, there are only two low outliers. These outliers are not close to each other (Figure 6). Also, the box map for percentage of clay does not show that the points close to each other are similar. On the other hand, the biomass variable seem to have a clustering pattern such that points in the bottom side are in the forth and fifth quartile, and the only upper outlier of the sample is in this part of the field (Figure 7). For the south facing side of the field the measure of biomass is relatively low with respect to the bottom side, where almost all the points are in the fist quartile. For the north facing side of the field, the measure of biomass is lower than the measure in bottom side, but higher than the south facing side. Overall, the results obtained from the box maps do not show any general spatial patterns (i.e. there is no strong visual evidence of clustering) for all the other variables of interest. Note that the box maps for yield and soil N are not reported here in the interest of space, but are available from the authors upon request.

All the maps above are simply visual characterization of the data. More accurate analysis of spatial structure can be made by using spatial statistics. But before we can calculate spatial statistics, we have to define the "neighbors" for each GPS point based on specific criteria. This will allow us to assess if there are any spatial relationships between these points that can serve as a basis for management zones. According to Bivand (1998), we can set the neighbors for each point by different criteria. We can set the membership neighbors by defining the locations that share boundaries with each point. We could also draw bands at different distances of the *centroids*. From the experimental data, we could see that the points of analysis are delineated by different level curves described by a center pivot irrigation span. Therefore we could take advantage of the areas defined by the pivot. Taking the center pivot as a reference we delineate circles with different radius. The points between the level curves are set as the neighbors of the points within the curve.

Once, we defined the neighbors, we arranged the contiguity relations on a weight matrix based on the arc distance or bands centered on each point. This matrix is necessary to calculate the *Morans I* statistics. According to Bivand :

> It is usual in the literature to define the contiguity relation in terms of sets of N(i) neighbors of zone or site i. These are coded in the form of a weights matrix W, with a zero diagonal, and the off-diagonal non-zero elements often scaled to sum to unity in each row (a.k.a. standardized weights matrices), with typical elements:
>
> $$w_{ij} = \frac{c_{ij}}{\sum_{j=1}^{N} c_{ij}}$$
>
> Where C_{ij}=1 if i is linked to j and C_{ij}=0 otherwise. This implies no use of other information than that of neighbourhood set membership.

The characteristics of the weights matrix based on distance can be seen in Table 2. With the information from the weights matrix, we can then calculate the Moran's I statistic to analyze the spatial autocorrelation between the neighbors. The *Moran's I* statistic tests the null hypothesis, that there is no association between the value observed at a location and the values observed at the neighboring sites, against the alternative hypothesis that values of nearby locations are similar. This spatial autocorrelation indicator was calculated for the different variables at each GPS point. We calculated this statistic for percentage of sand, percentage of clay, biomass, electrical conductivity and quantity of nitrogen in the soil. Based on this test we found that there is no pattern of clustering between the points (See Table 3). The only significant variables were soil N, biomass, and electrical conductivity; but the z-value was negative showing that there is no clustering pattern (i.e. dissimilar points are close to each other). With Moran scatterplots (Figure 8 to 12), we verify the results that most of the clusters of points are negative correlated, so there is no evidence of positive spatial autocorrelation (i.e. clustering of similarly-valued points). This result is consistent with the graphical preliminary analysis done with the *choropleth* and box maps.

Conclusions

The results of the spatial analysis above did not show evident spatial clustering patterns for all the variables examined. This suggests that management zones may not be necessary in this case and blanket application of N may be preferable. This result could be due to the control environment under which the experiment was constructed. For example, the sampling points collected could have been far enough apart such that these locations indeed do not have strong spatial relationships. One possible remedy for this is to create grids (instead of points) using a geostatistics/site-pecific management software like SSToolbox™, which was also done in Anselin, Bongiovanni, and Lowenberg-Deboer (2001). Management zones may then be more easily delineated using this data type.

Although the results did not show any significant spatial patterns that can be used for delineating management zones, it is important to recognize that these results are still preliminary. As mentioned above several other steps may be pursued to further streamline and improve the analysis. Other data sets for different locations or different years will also be examined to evaluate the usefulness of the ESDA method for management zone delineation. Nevertheless, we believe that the method is a workable methodology to detect spatial patterns, which may help guide farmers and crop consultants in the definition of management zones. It is important to note that the results for all the different ESDA measures used (from the *choropleth* maps to the Moran scatterplot) were very consistent, which bodes well for the usefulness and robustness of the approach.

References

Anselin L., Bongiovanni R. and Lowenberg-DeBoer J. (2001). A spatial econometric approach to the economics of site-specific nitrogen management in corn production. Working paper. 2001.

Bivand, R.,(1998, April). A review of spatial statistical techniques for location studies. Retrieved October 20, 2003, from http://www.nhh.no/geo/gib/gib1998/gib98-3/lund.html

Bronson K., Keeling W., Booker J.D.,Chua T., Wheeler T., Boman R., and Robert J. Lascano (2003). Influence of Landscape Position, Soil Series, and Phosphorus Fertilizer on Cotton Lint Yield. Agron. J. 95: 949–957.

Dillon, C. 2002. A Mathematical Programming Model for Optimal Management Zone Delineation in Precision Agriculture. Selected Paper presented at the AAEA Annual Meetings, Long Beach, CA (July 28-31, 2002)

Khanna M., Epouhe, O.F. and Hornbaker, R. 1999. Site-Specific Crop Management: Adoption Patterns and Incentives. Rev. of Agr. Eco. 21:455-472.

Messner, S. and Anselin, L. Spatial analyses of homicide with areal data. Working Paper, 2002.

Texas Department of Agriculture. Online. Available: http://www.agr.state.tx.us/about/link4.htm 04-24-01
.
United States Department of Agriculture (USDA). 2003. Crop Production 2002 Summary. http://usda.mannlib.cornell.edu/reports/nassr/field/pcp-bban/cropan03.pdf

Wilson, L.T. 1998. Precision Agriculture for Cotton. Agroeconomic Research Group, Texas A&M University http://aesrg.tamu.edu/Cotton/PrecAgr.htm

Table 1. Relation of nitrogen (N) and landscape position.

```
reg  n6 nf sf

      Source |       SS       df       MS              Number of obs =     135
-------------+------------------------------           F(  2,   132) =    2.71
       Model |  171.776661     2  85.8883307           Prob > F      =  0.0705
    Residual |  4190.13521   132  31.7434486           R-squared     =  0.0394
-------------+------------------------------           Adj R-squared =  0.0248
       Total |  4361.91188   134  32.5515812           Root MSE      =  5.6341

------------------------------------------------------------------------------
          n6 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          nf |  -1.312089   .6857648    -1.91   0.058    -2.668599    .0444216
          sf |   1.441822   .6857648     2.10   0.037     .0853117    2.798333
       _cons |   7.458622   .4849089    15.38   0.000     6.499424     8.41782
------------------------------------------------------------------------------
```

Table 2. Characteristics of the distance matrix.

Dimension:	133
Average distance between points:	35.319
Minimum distance between points:	1
Min. allowable distance cutt off:	1.7388

Table 3. Moran's I test.

	Z-VALUE	PROB
SAND	1.019	0.3079
CLAY	0.1816	0.855
NE	-3.7604	0.00017
BM	-10.229	0.00000
EC	-18.639	0.00000

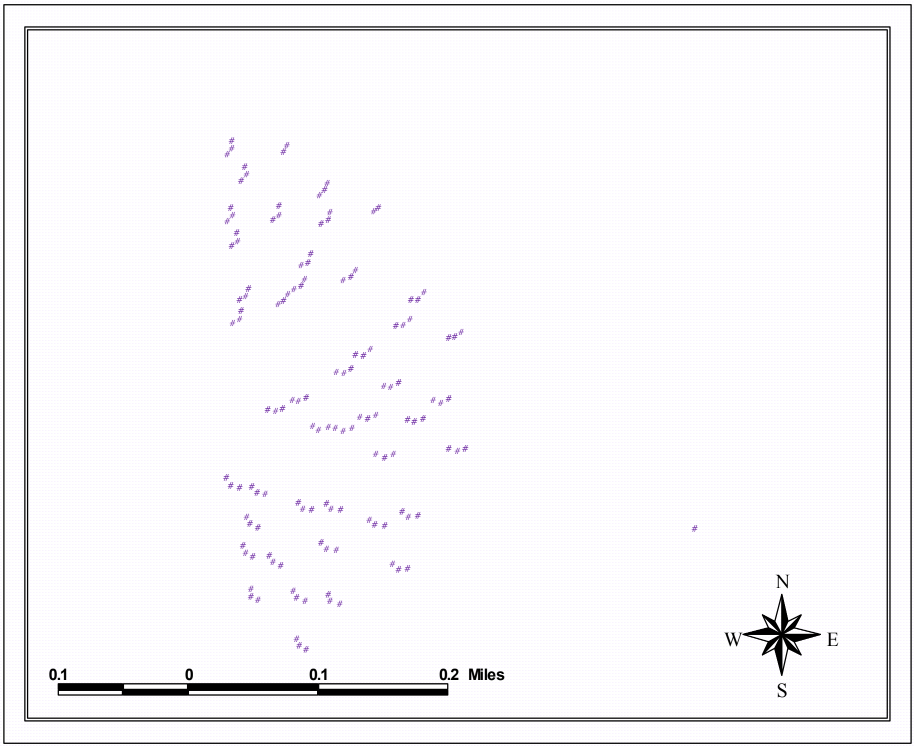

Figure 1. GPS points, Southern High Plains of Texas.

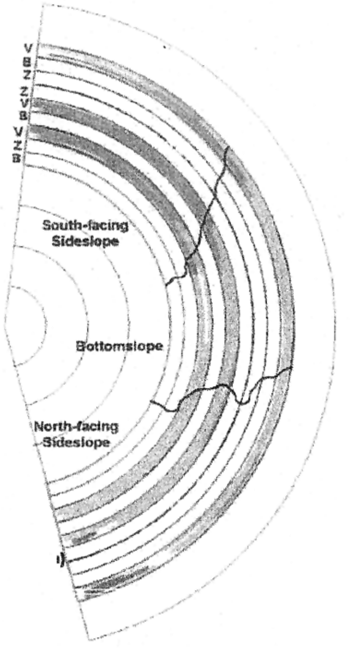

Figure 2. Design of the Nitrogen experiment.

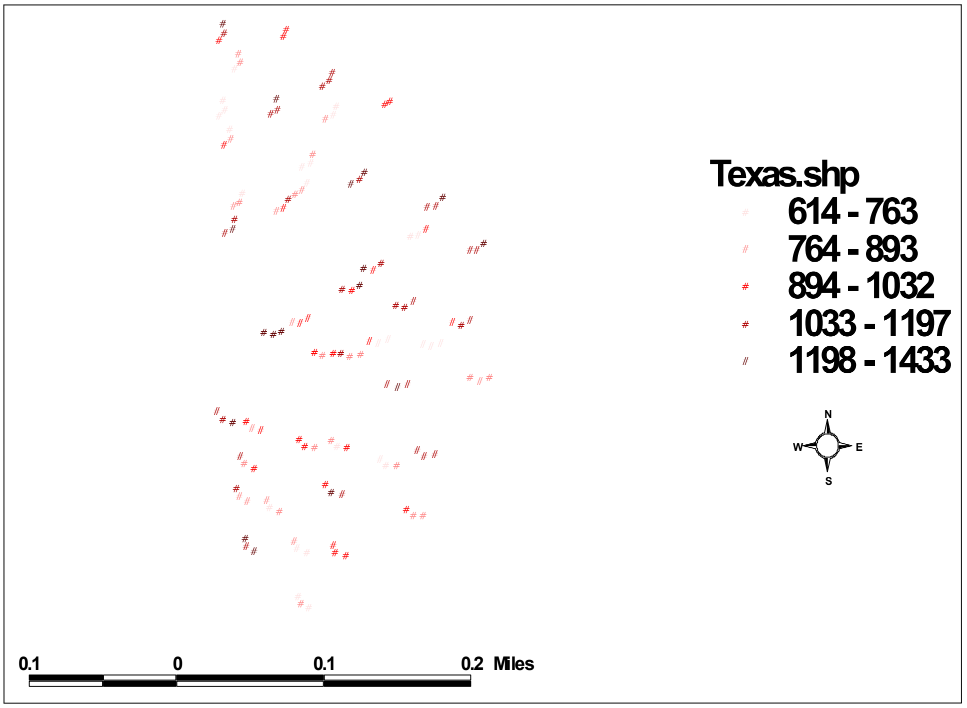

Figure 3. Choropleth map for yield.

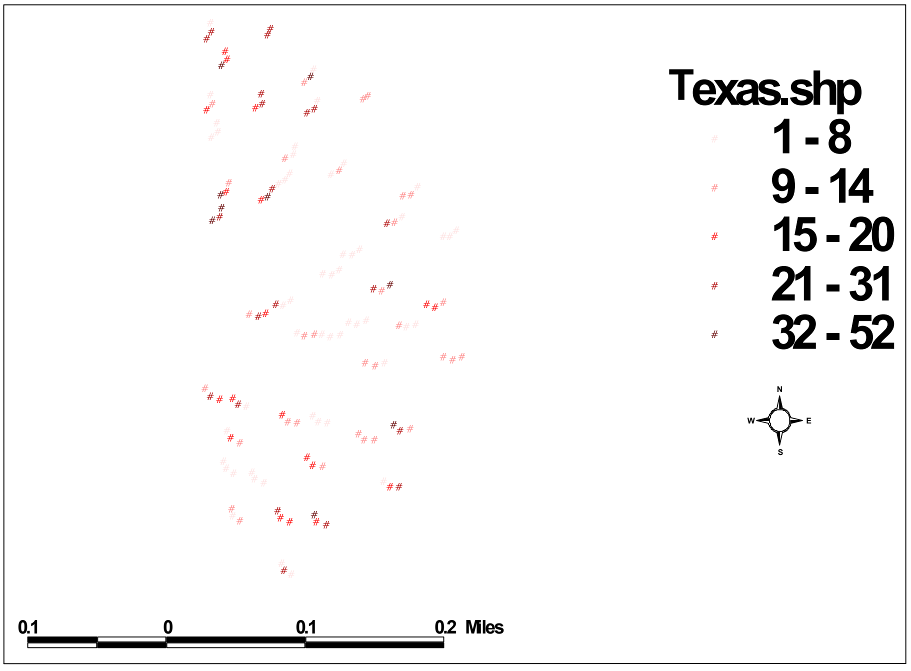

Figure 4. Choropleth map for soil N.

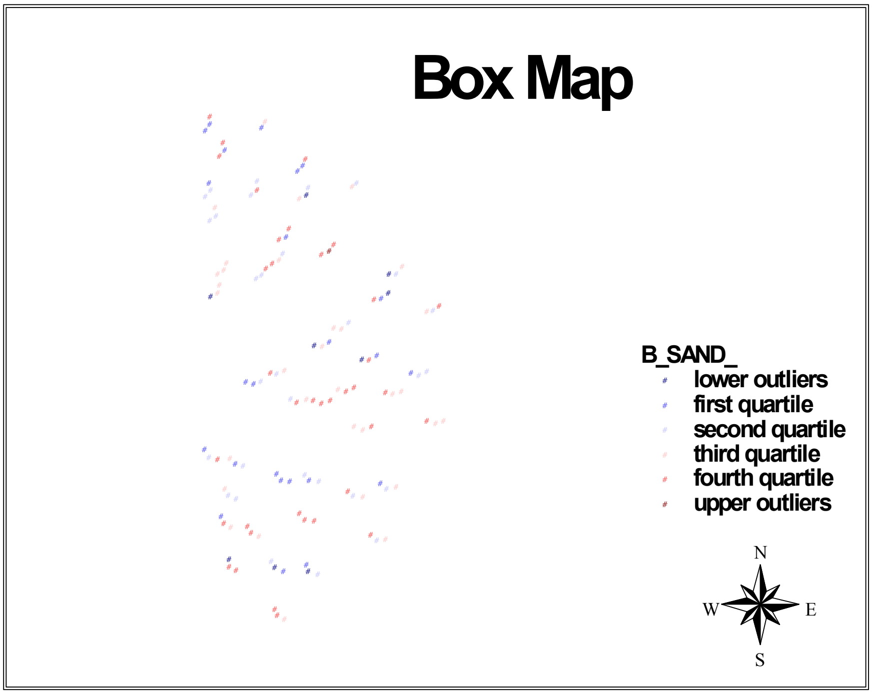

Figure 5. Box map for percentage of sand.

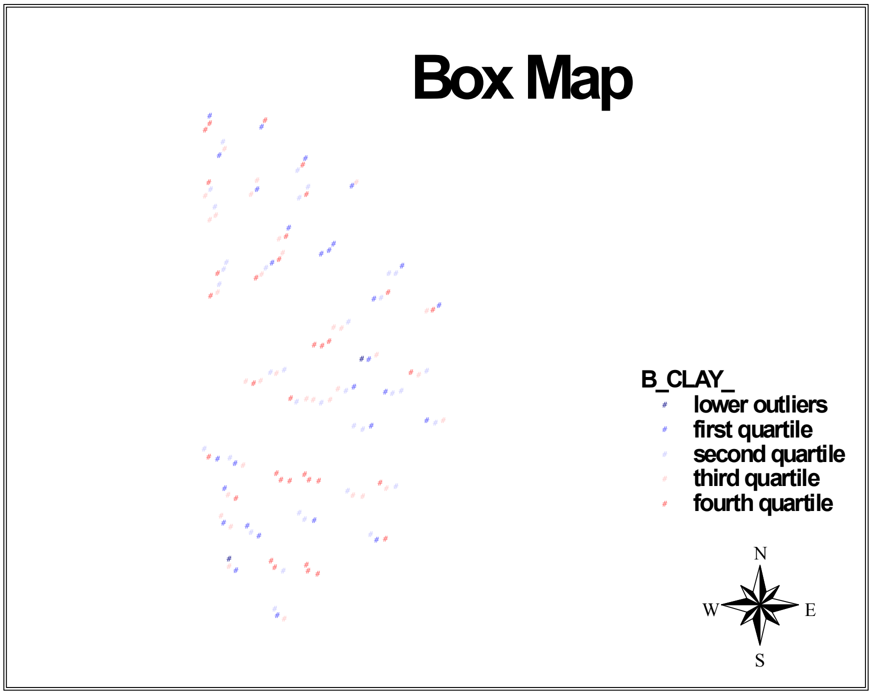

Figure 6. Box map for percentage of clay.

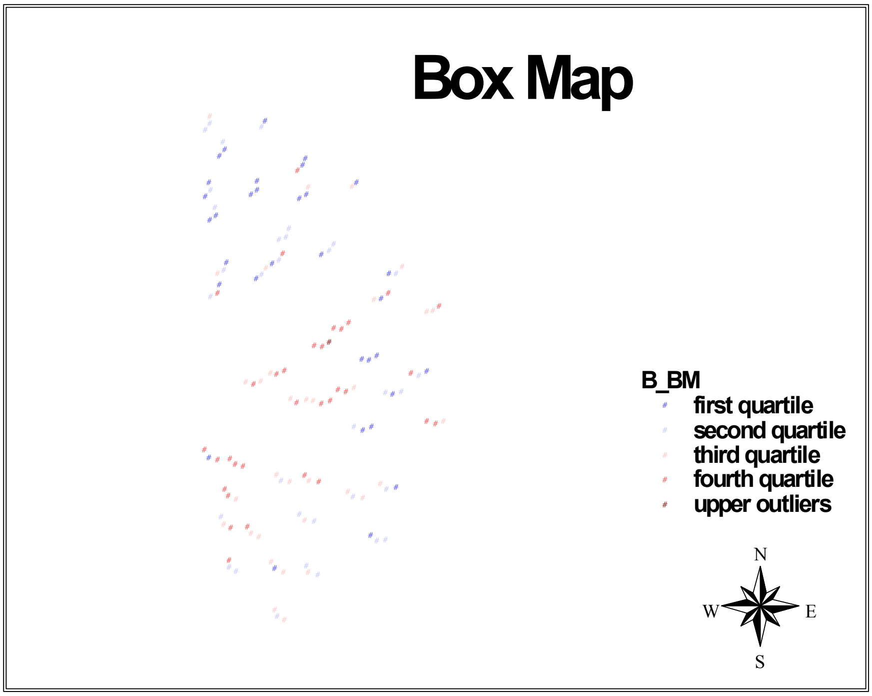

Figure 7. Box map for quantity of biomass.

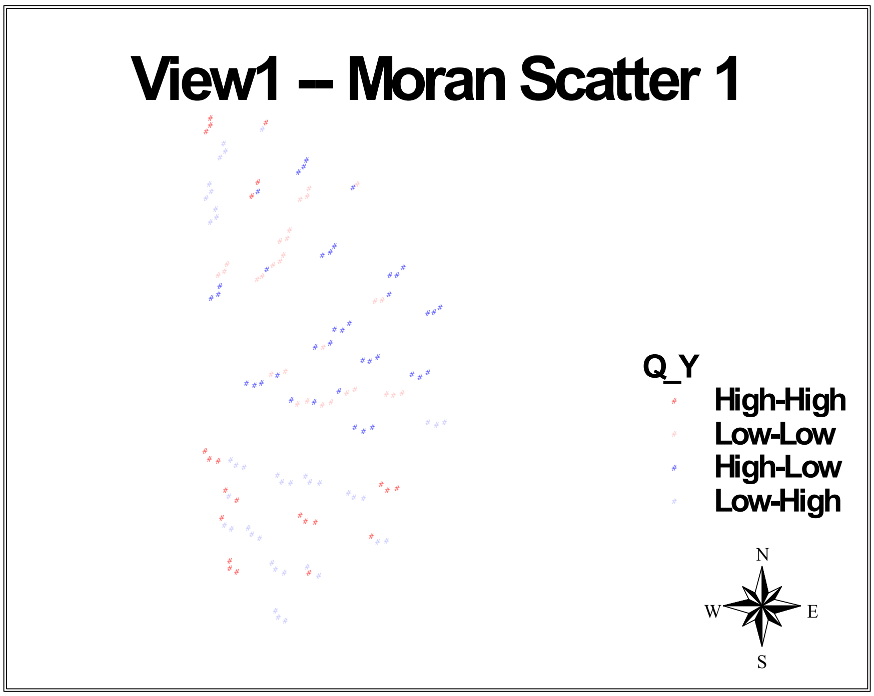

Figure 8. Moran scatterplot for cotton yield.

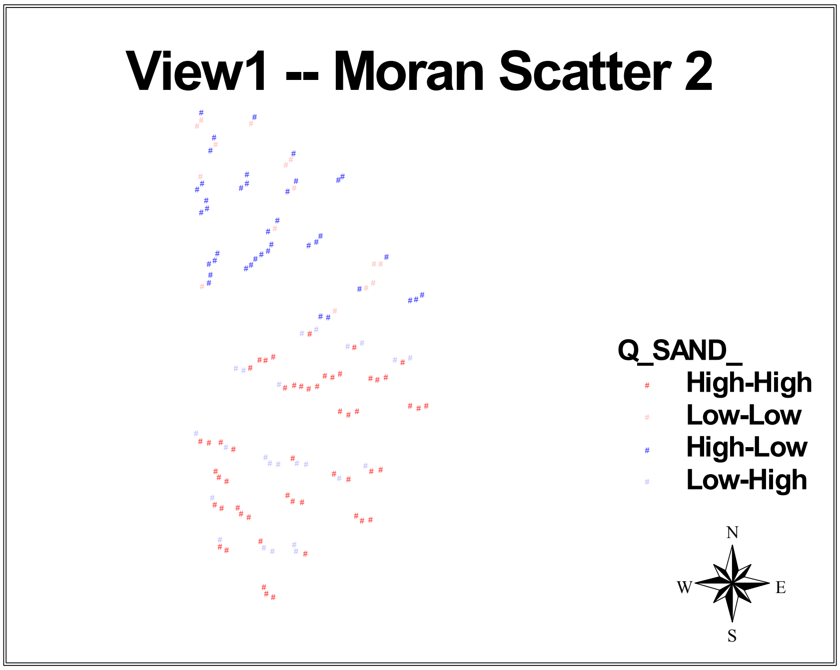

Figure 9. Moran scatterplot for percentage of sand.

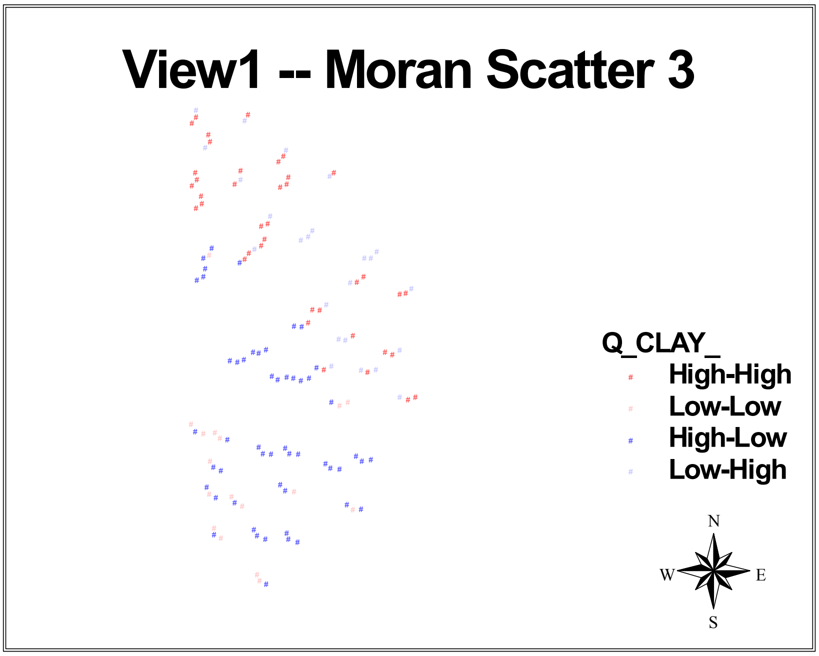

Figure 10. Moran scatterplot for percentage of clay.

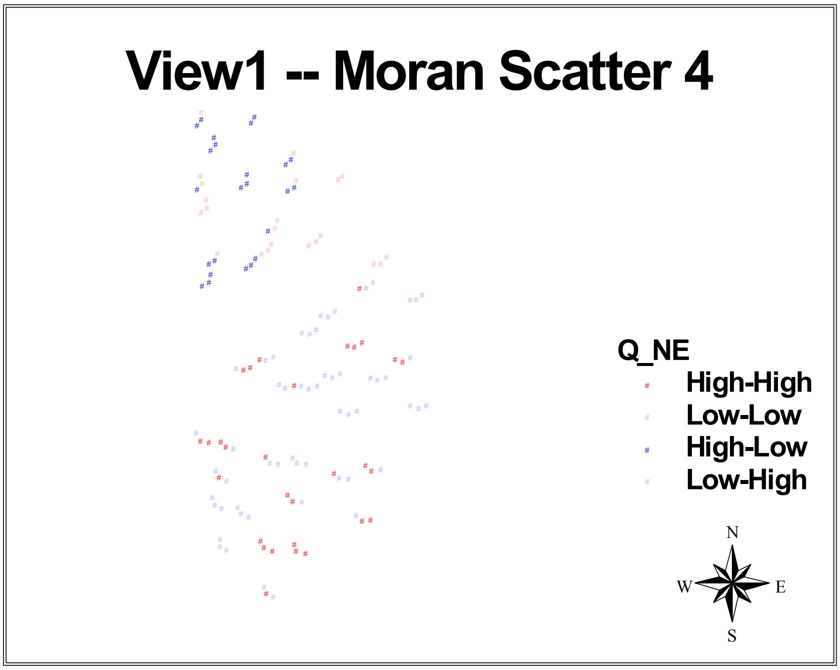

Figure 11. Moran scatterplot for soil N.

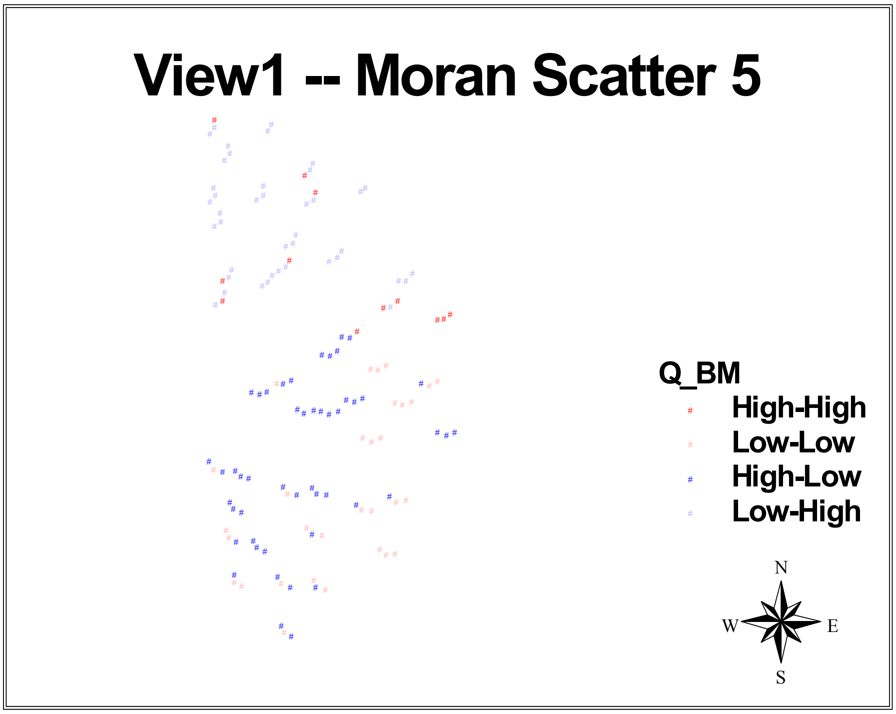

Figure 12. Moran scatterplot for biomass.