

Topic 12

Spatial Data Mining

12.1 Comparing Yield Maps (Discrete)

One of the most fundamental operations in map analysis is the comparison of two maps. Questions like “how different are the maps?”, “how are they different?” and “where are they different?” immediately spring to mind. Quantitative answers are needed because visual comparison cannot fully consider all of the detail in an objective manner.

Discrete maps are comprised of distinct point, line and polygon features, such as a soil type map. Continuous map surfaces, on the other hand, characterize a spatial gradient, such as elevation.

Let’s consider discrete map comparison first. The two maps shown in figure 12-1 identify corn yield for successive seasons (1997 and 1998) on the central-pivot cornfield used in this case study. Note that the maps have a common legend from 0 to 300bu/ac and displayed using the same color pallet. *How different are they? How are they different? And where are they different?*

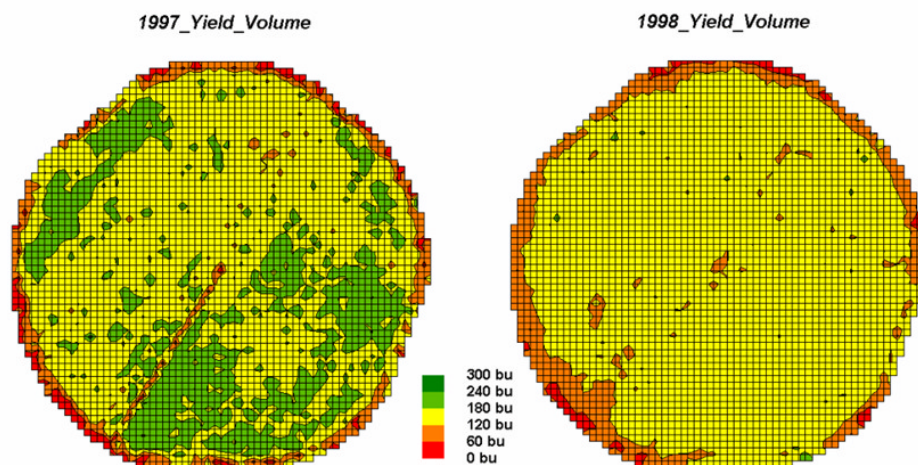


Figure 12-1. Discrete yield maps for consecutive years.

While your eyes flit back and forth in an attempt to compare the maps, the computer approaches the problem more methodically (note the common legend).

The first step converts the point values collected while harvesting the field to a grid value for each cell. An analysis grid resolution is chosen (50ft cells are used in this example) and geometrically aligned with the maps. The dominant yield class within each cell is assigned its interval value (values 1= low yield of 0-60bu/ac through 5= high yield of 240-300 corresponding to the color ranges in the map display).

The next step, as shown in figure 12-2, combines the two maps into a single map that indicates the “joint condition” for both years. Since the two maps have an identical grid configuration, the computer simply retrieves the two class assignments for a grid location, then converts them to a single number.

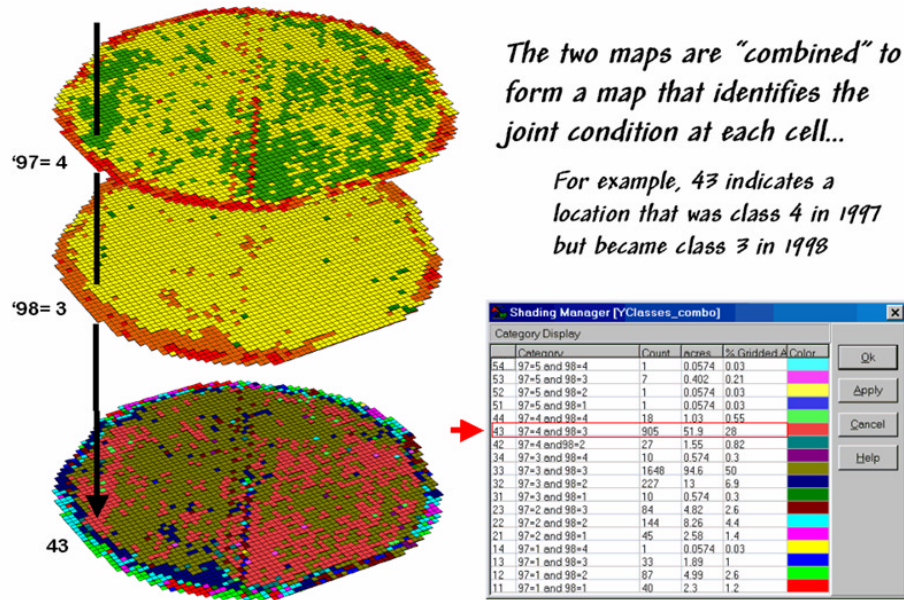


Figure 12-2. Coincidence map identifying the joint conditions for both years.

The map-ematical procedure computes the “first value times ten plus the second value” to form a two-digit code. In the example shown in the figure, the value “forty-three” is interpreted as class 4 in the first year but decreasing to class 3 in the next year.

The final step sums up the changes to generate a coincidence table (see figure 12-3). The columns and rows in the table represent the class assignments on the 1997 and 1998 yield maps, respectively. The body of the table reports the number of cells for each joint condition. For example, column 4 and row 3 notes that there are 905 occurrences where the yield class slipped from level four (180-240bu/ac) to level three (120-180bu/ac).

	A	B	C	D	E	F	G	H	I
1	1997 Yield								
2			1 (0-60)	2 (60-120)	3 (120-180)	4 (180-240)	5 (240-300)	Subtotal	Percent
3	1998 Yield	1 (0-60)	40	45	10	0	1	96	41.67
4		2 (60-120)	87	144	227	27	1	486	29.63
5		3 (120-180)	33	84	1648	905	7	2677	61.56
6		4 (180-240)	1	0	10	18	1	30	60.00
7		5 (240-300)	0	0	0	0	0	0	0.00
8		Subtotal	161	273	1895	950	10	3289	
9		Percent	24.84	52.75	86.97	1.89	0.00		56.25

Figure 12-3. Coincidence summary matrix.

The off-diagonal entries indicate changes between the two maps—the values indicate the relative importance of the change. For example, the 905 statistic for the “four-three” change is the largest and therefore identifies the most frequently occurring change in the field. The 0 statistic for the “four-one” combination indicates that level four never slipped all the way to level 1.

The diagonal entries summarize the agreement between the two maps. The greatest portion of the field that didn’t change occurs for yield class 3 (“three-three” with 1640 cells). The statistic in the extreme lower-right (56.25%) reports that only a little more than half the field didn’t change its yield class.

Generally speaking, the maps are very different (only 56.25% unchanged). The greatest difference occurred for class 4 (only 1.89% didn't change). And a detailed picture of the spatial patterns of change is depicted in the coincidence map shown in figure 12-2.

That's a lot more meat in the answers to the basic map comparison questions (how much, how and where) than visceral viewing can do. The next section looks at even more precise procedures for reporting differences between mapped data.

12.2 Comparing Yield Surfaces (Continuous)

Contour maps are the most frequently used and familiar form of presenting precision agriculture data. The two 3D perspective-plots in the top of figure 12-4 show the color-coded ranges of yield (0-60, 60-120, etc. bushels per acre) and are identical to the discrete maps discussed in the previous section. The color-coding of the contours is draped for cross-reference onto the continuous 3D surface of the actual yield data.

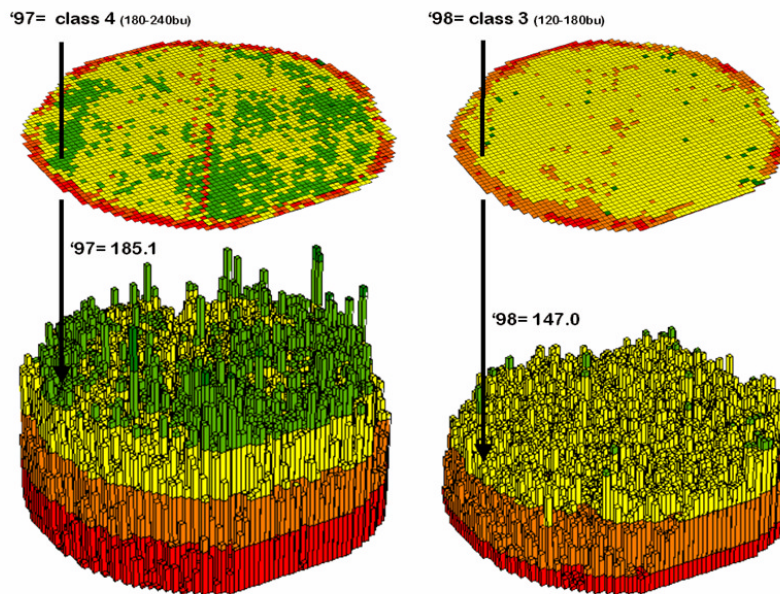


Figure 12-4. 3-D Views of yield surfaces for consecutive years.

Note the “spikes and pits” in the surfaces that graphically portray the variance in yield data for each of the contour intervals. While discrete map comparison identifies shifts in broadly defined yield classes, continuous surface comparison identifies the precise difference at each location.

For example, a yield value of 179 bushels on one map and 121 on the other are both assigned to the third contour interval (120 to 180; yellow). A discrete map comparison would suggest that no change in yield occurred for the location because the contour interval remained constant. A continuous surface comparison, would report a fairly significant 58-bushel decline.

Figure 12-5 shows the calculations using the actual values for the same location highlighted in the previous section's discussion. The discrete map comparison reported a decline from yield level 4 (180 to 240) to level 3 (120 to 180).

The MapCalc command, “*Compute Yield_98 minus Yield_97 for Difference*” generates the difference surface. If the simple “map algebra” equation is expanded to “*Compute (((Yield_98 minus Yield_97) / Yield_97) *100)*” a percent difference surface would be generated. Keep in mind that a map surface is merely a spatially organized set of numbers that awaits detailed analysis then transformation to generalized displays and reports for human consumption.

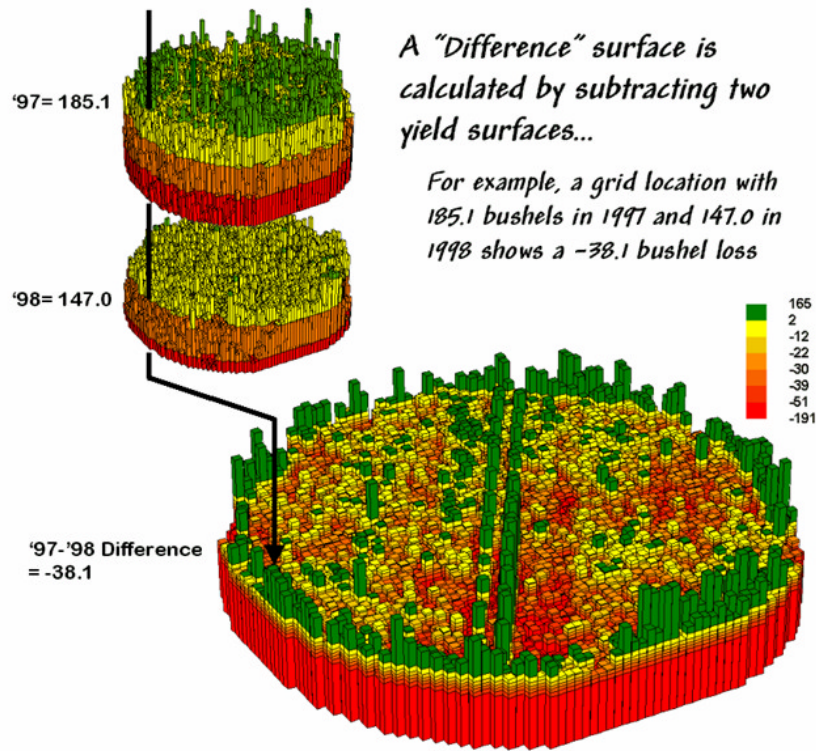


Figure 12-5. A difference surface identifies the actual change in crop yield at each map location.

In figure 12-5, note that the wildest differences (side-by-side green spikes and red pits) occur at the field edges and along the access road—from an increase of 165 bushels to a decrease of 191 bushels between the two harvests. However, notice that most of the change is about a 25 bushel decline (mean= -22.6; median= -26.3) as identified in the summary table shown in figure 12-6.

The continuous surface comparison more precisely reports the change as -38.1 bushels. The differences for other 3,289 grid cells are computed to derive a **Difference Surface** that tracks the subtle variations in the spatial pattern of the changes in yield.

The histogram of the yield differences in the figure shows the numerical distribution of the difference data. Note that it is normally distributed and that the bulk of the data is centered about a 25 bushel decline. The vertical lines in the histogram locate the contour intervals used in the 2D display of the difference map in the left portion of figure 12-6.

The detailed legend links the color-coding of the map intervals to some basic frequency statistics. The example location with the calculated decline of -38.1 is assigned to the -39 to -30 contour range and is displayed as a mid-range red tone. The display uses an *Equal Count* method with *seven intervals*, each representing approximately 15% of the field. Green is locked for the only interval of increased yield. The decreased yield intervals form a color-gradient from yellow to red. All in all, surface map comparison provides more information in a more effective manner discrete map comparison. Both approaches, however, are far superior to simply viewing a couple yield maps side-by-side and guessing at the magnitude and pattern of the changes.

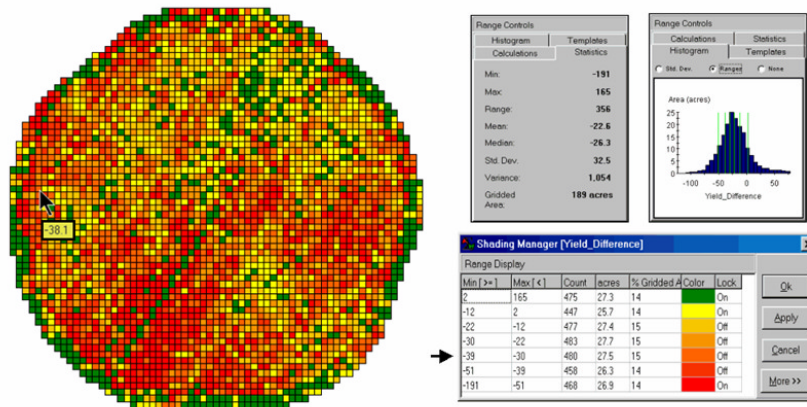


Figure 12-6. A 2-D map and statistics summarize the differences in crop yield between two periods.

The ability to quantitatively evaluate continuous surfaces is fundamental to precision agriculture. A difference surface is one of the simplest and most intuitive forms. While the math and stat of other procedures are fairly basic, the initial thought of “you can’t do that to a map” is usually a reflection of our non-spatial statistics and paper-map legacies. In most instances, precision agriculture is simply an extension of current research and management practices from a few *sample plots* to extensive *mapped data sets*. The remainder of this case study investigates many of these extensions.

12.3. Calculating Field Similarity

How often have you seen a presenter “lasso” a portion of a map with a laser pointer and boldly state “...see how similar this area is to the locations over here and here...” as the pointer rapidly moves about the map. More often than not, there is a series of side-by-side maps serving as the background scenery for the laser show.

Consider the three maps shown in figure 12-7— what areas identify similar data patterns? If you focus your attention on a location in the southeastern portion how similar are all of the other locations?

The answers to these questions are far too complex for visual analysis and certainly beyond the geo-query and display procedures of standard desktop mapping packages. The data in the example shows the relative amounts of phosphorous, potassium and nitrogen distributions throughout the case study field.

In visual analysis you move your focus among the maps to summarize the color assignments at different locations. The difficulty in this approach is two-fold— remembering the color patterns and calculating the difference.

Quantitative map analysis does the same thing except it uses the actual map values in place of colors. In addition, the computer doesn’t tire as easily as you and completes the comparison for all of the locations throughout the map window (3,289 grid cells in this example) in a couple seconds.

The upper-left portion of figure 12-8 illustrates capturing the data patterns of two locations for comparison. The “data spear” at map location 45column, 18row identifies the P-level as 11.0, the K-level as 177.0 and N-level as 32.9ppm. This step is analogous to your eye noting a color pattern of burnt-red, dark-orange and light-green. The other location for comparison (32c, 62r) has a data pattern of P= 53.2, K= 412.0 and N= 27.9 ...or as your eye sees it, a color pattern of dark-green, dark-green and yellow.

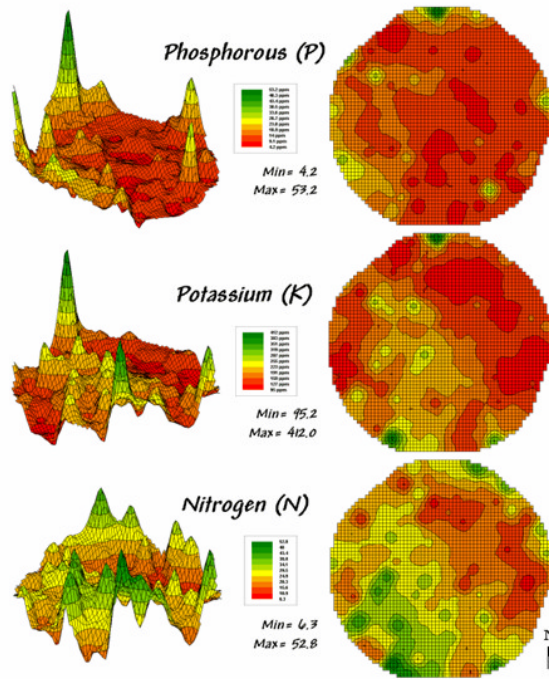


Figure 12-7. Map surfaces identifying the spatial distribution of P,K and N throughout a field.

The right side of figure 12-8 conceptually depicts how the computer calculates a similarity value for the two response patterns. The realization that mapped data can be expressed in both geographic space and data space is key to understanding quantitative analysis similarity.

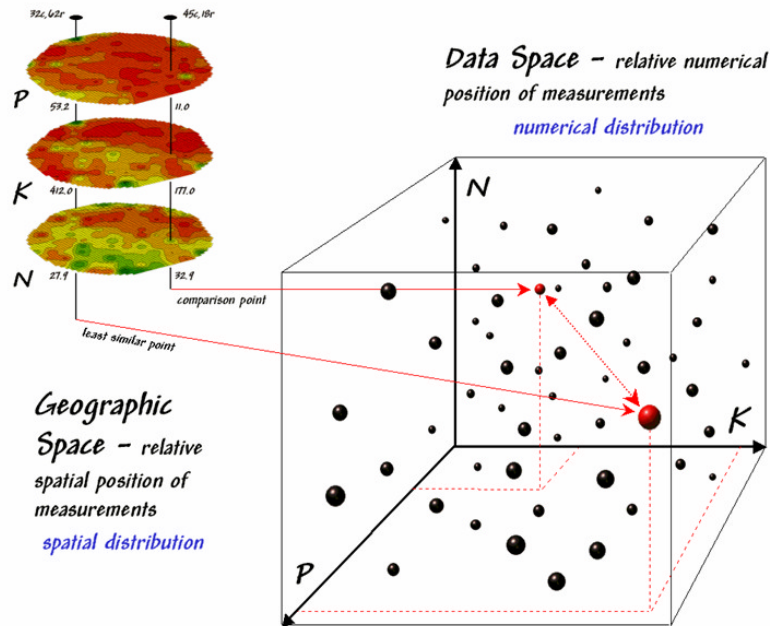


Figure 12-8. Conceptually linking geographic space and data space.

Geographic space uses coordinates, such latitude and longitude, to locate things in the real world—such as the southeast and extreme north points identified in the example field. The geographic expression of the complete set of measurements depicts their spatial distribution in familiar map form.

Data space, on the other hand, is a bit less familiar. While you can't stroll through data space you can conceptualize it as a box with a bunch of balls floating within it. In the example, the three axes defining the extent of the box correspond to the P, K and N levels measured in the field. The floating balls represent grid locations defining the geographic space—one for each grid cell. The coordinates locating the floating balls extend from the data axes—11.0, 177.0 and 32.9 for the comparison point. The other point has considerably higher values in P and K with slightly lower N (values 53.2, 412.0, 27.9 respectively) so it plots at a different location in data space.

The bottom line for data space analysis is that the position of a point identifies its numerical pattern—low, low, low in the back-left corner, while high, high, high is in the upper-right corner of the box. Points that plot in data space close to each other are similar; those that plot farther away are less similar.

In the example, the floating ball closest to you is least similar (greatest distance) from the comparison point. This distance becomes the reference for “most different” and sets the bottom value of the similarity scale (0% similar). A point with an identical data pattern plots at exactly the same position in data space resulting in a data distance of 0 that equates to the highest similarity value (100% similar).

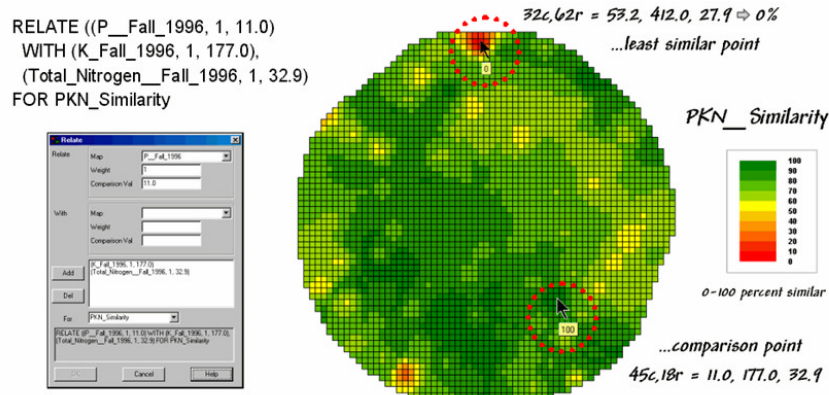


Figure 12-9. A similarity map identifies how related locations are to a given point.

The similarity map shown in figure 12-9 applies the similarity scale to the data distances calculated between the comparison point and all of the other points in data space. The green tones indicate field locations with fairly similar P, K and N levels to the comparison location in the field—with the darkest green identifying locations with identical P, K and N levels (100% similar). It is interesting to note that most of the very similar locations are in the western portion of the field. The light-green to red tones indicate increasing dissimilar areas.

A similarity map can be an invaluable tool for investigating spatial patterns in any complex set of mapped data. The calculation of similarity can handle any number of input maps, yet humans are unable to conceptualize more than three variables (the data space box). Also, the different map layers can be weighted to reflect relative importance in determining overall similarity.

In effect, a similarity map replaces a lot of laser-pointer waving and subjective suggestions of similar/dissimilar areas with a concrete, quantitative measurement at each map location. The technique moves map analysis well beyond the old “...I'd never have seen, it if I hadn't believed it...” mode of visual map interpretation.

12.4 Identifying Data Zones

Last section introduced the concept of **Data Distance** as a means to measure similarity within a map. One simply mouse-clicks a location and all of the other locations are assigned a similarity value from 0 (zero percent similar) to 100 (identical) based on a set of specified map layers. The statistic replaces difficult visual interpretation of map displays with an exact quantitative measure at each location.

An extension to the technique allows you to circle an area then compute similarity based on the typical data pattern within the delineated area. In this instance, the computer calculates the average value within the area for each map layer to establish the comparison data pattern, and then determines the normalized data distance for each map location. The result is a map showing how similar things are to the area of interest.

The link between **Geographic Space** and **Data Space** is critical. As shown in figure 12-10, spatial data can be viewed as a map or a histogram. While a map shows us “where is what,” a histogram summarizes “how often” measurements occur (regardless where they occur).

The top-left portion of the figure shows a 2D/3D map display of the relative amount of phosphorous (P) within a farmer’s field. Note the spikes of high measurements along the edge of the field, with a particularly big spike in the north portion.

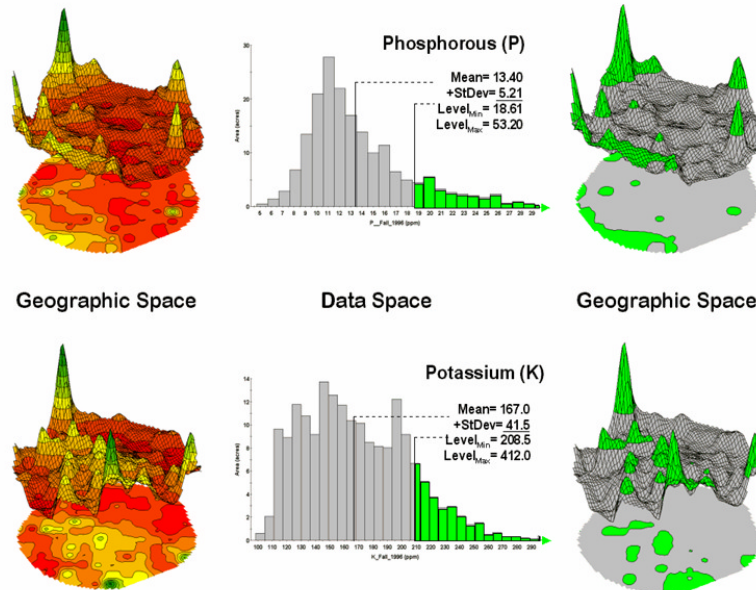


Figure 12-10. Identifying areas of unusually high measurements.

The histogram to the right of the map display depicts a different perspective of the data. Rather than positioning the measurements in geographic space it summarizes their relative frequency of occurrence in data space. The X-axis of the graph corresponds to the Z-axis of the map—amount of phosphorous. In this case, the spikes in the graph indicate measurements that occur more frequently. Note the high occurrence of phosphorous values around 11ppm.

The top-right portion of the figure puts the data and geographic space link to use. The shaded area in the histogram view identifies field phosphorous values that are unusually high—more than one standard deviation above the mean. This statistical cutoff is used to isolate locations of high concentrations as shown in the map on the right. The procedure is repeated for the potassium (K) map surface to identify its locations of unusually high measurements.

Figure 12-11 illustrates combining the P and K data to locate areas in the field that have high measurements in both. The graphic on the left is termed a scatter diagram or plot. It graphically summarizes the joint occurrence of both sets of mapped data.

Each ball in the scatter plot schematically represents a location in the field. Its position in the plot identifies the P and K measurements at that location. The balls plotting in the shaded area of the diagram identify field locations that have both high P and high K. The upper-left partition identifies joint conditions in which neither P nor K are high. The off-diagonal partitions in the scatter plot identify locations that are high in one nutrient but not the other.

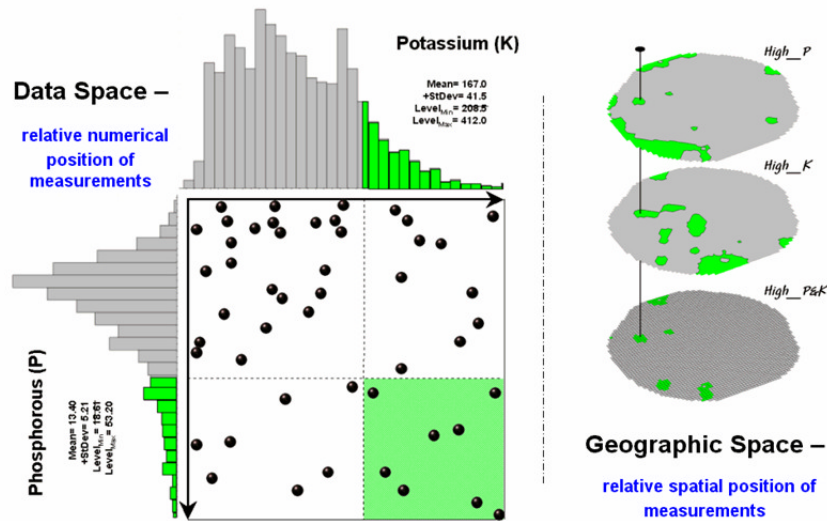


Figure 12-11 Identifying joint coincidence in both data and geographic space.

The aligned maps in the right inset show the geographic solution for areas that are high in both of the soil nutrients. A simple map-ematical way to generate the solution is to assign 1 to all locations of high values in the P and K map layers (bright green). Zero is assigned to locations that aren't high (light gray). When the two binary maps (0/1) are multiplied, a zero on either map computes to zero.

Locations that are high on both maps equate to 1 ($1 \times 1 = 1$). In effect, this “level-slice” technique maps any data pattern you specify—just assign 1 to the data interval of interest for each map variable.

Figure 12-12 depicts level slicing for areas that are unusually high in P, K and N (nitrogen). In this instance the data pattern coincidence is a box in 3-dimensional scatter plot space.

However a map-ematical trick was employed to get the detailed map solution shown in the figure. On the individual maps, high areas were set to $P=1$, $K=2$ and $N=4$, then the maps were added together.

The result is a range of coincidence values from zero ($0+0+0=0$; gray= no high areas) to seven ($1+2+4=7$; red= high P, high K, high N). The map values in between identify the map layers having high values. For example, the yellow areas with the value 3 have high P and K but not N ($1+2+0=3$).

If four or more maps are combined, the areas of interest are assigned increasing binary progression values (...8, 16, 32, etc)—the sum will always uniquely identify the combinations.

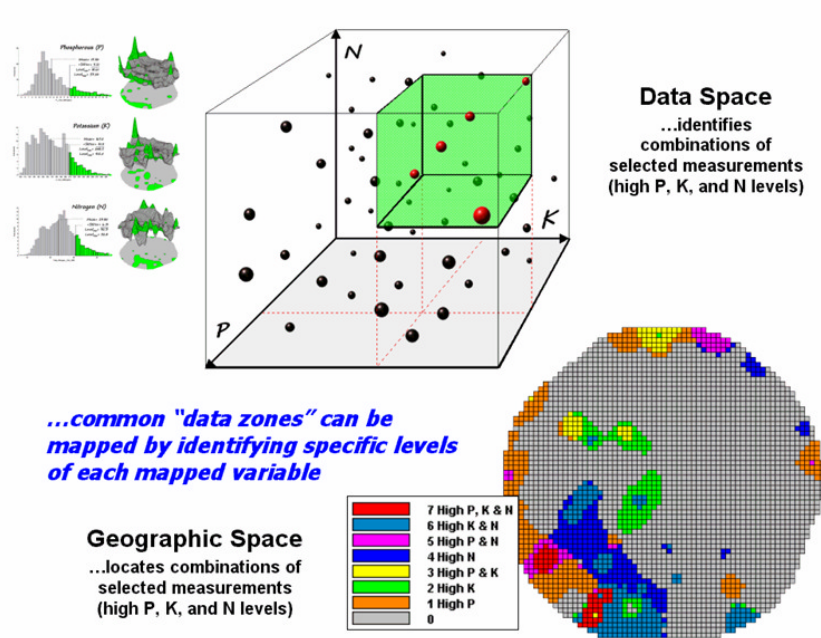


Figure 12-12. Level-slice classification using three map variables.

While level-slicing isn't a very sophisticated classifier, it does illustrate the useful link between data space and geographic space. This fundamental concept forms the basis for most geo-statistical analysis—including map clustering discussed in the next section.

12.5 Mapping Data Clusters

The last couple of sections have focused on analyzing data similarities within a stack of maps. The first technique, termed **Map Similarity**, generates a map showing how similar all other areas are to a selected location. A user simply clicks on an area and all other map locations are assigned a value from 0 (0% similar—as different as you can get) to 100 (100% similar—exactly the same data pattern).

The other technique, **Level Slicing**, enables a user to specify a data range of interest for each map in the stack then generates a map identifying the locations meeting the criteria. Level Slice output identifies combinations of the criteria met—from only one criterion (and which one it is), to those locations where all of the criteria are met.

While both of these techniques are useful in examining spatial relationships, they require the user to specify data analysis parameters. But what if you don't know what Level Slice intervals to use or which locations in the field warrant Map Similarity investigation? Can the computer on its own identify groups of similar data? How would such a classification work? How well would it work?

Figure 12-13 shows some examples derived from **Map Clustering**. The "floating" map layers on the left show the input map stack used for the cluster analysis. The maps are the same P, K, and N maps identifying phosphorous, potassium and nitrogen levels throughout a cornfield that were used for the examples in the previous sections.

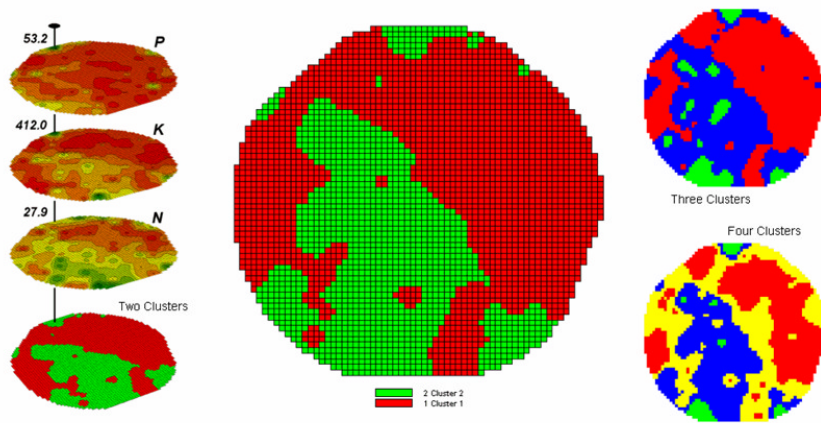


Figure 12-13. Examples of map clustering.

The map in the center of the figure shows the results of classifying the P, K and N map stack into two clusters. The data pattern for each cell location is used to partition the field into two groups that are 1) *as different as possible between groups* and 2) *as similar as possible within a group*. If all went well, any other division of the field into two groups would be worse at balancing the two criteria.

The two smaller maps at the right show the division of the data set into three and four clusters. In all three of the cluster maps, red is assigned to the cluster with relatively low responses and green to the one with relatively high responses. Note the encroachment on these marginal groups by the added clusters that are formed by data patterns at the boundaries.

The mechanics of generating cluster maps are quite simple. Just specify the input maps and the number of clusters you want then miraculously a map appears with discrete data groupings. So how is this miracle performed? What happens inside cluster's black box?

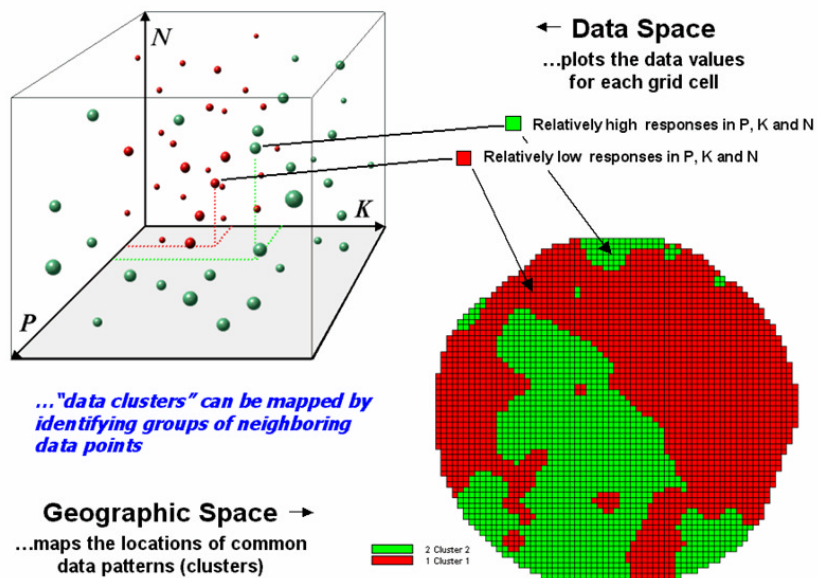


Figure 12-14. Data patterns for map locations are depicted as floating balls in data space.

The schematic in figure 12-14 depicts the process. The floating balls identify the data patterns for each map location (geographic space) plotted against the P, K and N axes (data space). For example, the large green ball appearing closest to you depicts a location with high values on all three input maps. The tiny red ball in the opposite corner (near the plot origin) depicts a map location with small map values. It seems sensible that these two extreme responses would belong to different data groupings.

While the specific algorithm used in clustering is beyond the scope of this discussion*, it suffices to note that data distances between the floating balls are used to identify cluster membership—groups of balls that are relatively far from other groups and relatively close to each other form separate data clusters. In this example, the red balls identify relatively low responses while green ones have relatively high responses. The geographic pattern of the classification is shown in the 2D map in the lower right portion of the figure.

Identifying groups of neighboring data points to form clusters can be tricky business. Ideally, the clusters will form distinct “clouds” in data space. But that rarely happens and the clustering technique has to enforce decision rules that slice a boundary between nearly identical responses. Also, extended techniques can be used to impose weighted boundaries based on data trends or expert knowledge. Treatment of categorical data and leveraging spatial autocorrelation are additional considerations.

So how do you know if the clustering results are acceptable? Most statisticians would respond, “...you can’t tell for sure.” While there are some elaborate procedures focusing on the cluster assignments at the boundaries, the most frequently used benchmarks use standard statistical indices.

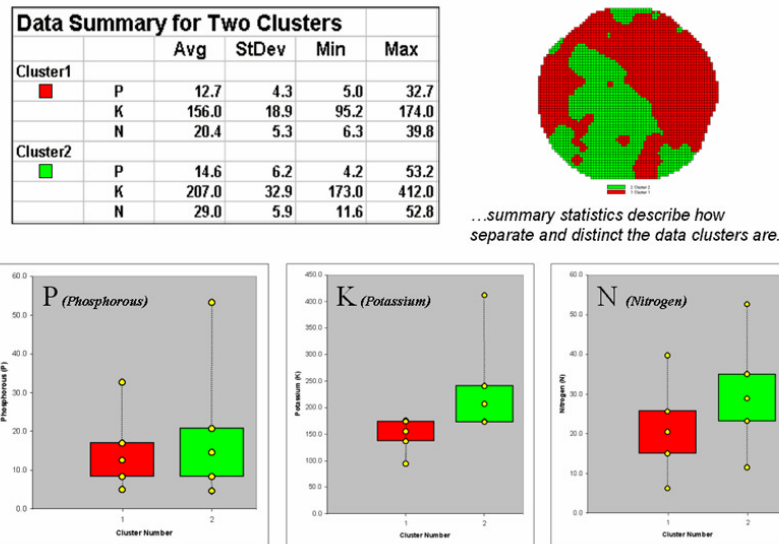


Figure 12-15. Clustering results can be roughly evaluated using basic statistics.

Figure 12-15 shows the performance table and box-and-whisker plots for the map containing two data clusters. The average, standard deviation, minimum and maximum values within each cluster are calculated. Ideally the averages between the two clusters would be radically different and the standard deviations small—large difference between groups and small differences within groups.


Box-and-whisker plots enable us to visualize these differences. The box is centered on the average (position) and extends above and below one standard deviation (width) with the whiskers drawn to the minimum and maximum values to provide a visual sense of the data range. When the diagrams for the two clusters overlap, as they do for the phosphorous responses, it tells us that the clusters aren’t very distinct along this axis.

The separation between the boxes for the K and N axes suggests greater distinction between the clusters. Given the results a practical precision farmer would likely accept the classification results... and statisticians hopefully will accept in advance my apologies for such a introductory and terse treatment of a complex topic.

12.6 Exercises

Access MapCalc using the *Agdata.rgs* data set by selecting **Start** → **Programs** → **MapCalc Learner** → **MapCalc Learner** → **Open existing map set** → **AgData.rgs**. The following set of exercises utilizes this database.

12.6.1 Comparing Yield Maps

 Press the **Map Analysis** button to access the analytical operations, select **Reclassify** → **Renumber** and enter the MapCalc command:

```
RENUMBER 1997_Yield_Volume
ASSIGNING 1 TO 0 THRU 60
ASSIGNING 2 TO 60 THRU 120
ASSIGNING 3 TO 120 THRU 180
ASSIGNING 4 TO 180 THRU 240
ASSIGNING 5 TO 240 THRU 300
FOR 97Yield_classes
```

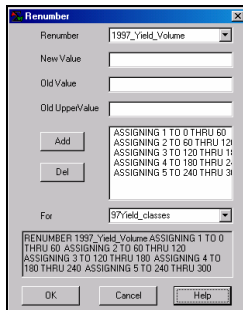
...that identifies five levels of yield from low (class 1= 0-60bu) to high (class 5= 240-300bu).

First specify **1997_Yield_Volume** as the input map. Next, enter the first *Assignment Phrase*:



ASSIGNING 1 TO 0 THRU 60 then press the **Add** button to append the phrase to the assignment list. Repeat for the “*NewValue, OldValue, OldUpperValue* then *Add*” sequence for the other four assignment phrases.

Finally, specify **97Yield_classes** as the output map and press **OK** to derive the map. Close the *Map Analysis* window to view the result.



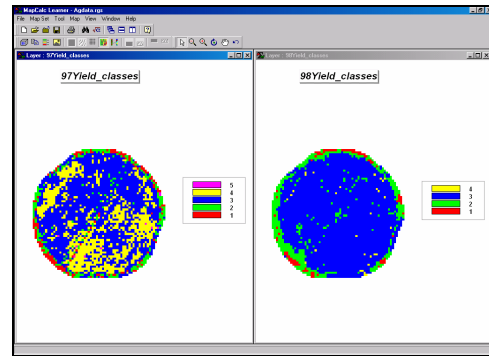
```
RENUMBER
1997_Yield_Volume ASSIGNING 1 TO 0 THRU 60
ASSIGNING 2 TO 60 THRU 120 ASSIGNING 3 TO 120
```


```
THRU 180 ASSIGNING 4 TO 180 THRU 240 ASSIGNING 5 TO
240 THRU 300 FOR 97Yield_classes
```

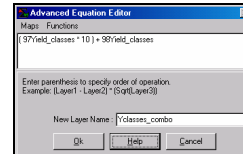
Repeat the same classification procedure using the *1998_Yield_Volume* map to derive the *98Yield_classes* map:

- ü Click on the **Map Analysis** button
- ü Select **Reclassify** → **Renumber**
- ü Specify **1998_Yield_Volume** as the input map
- ü Complete the *Assignment Phrases* using **Add** button
- ü Specify **98Yield_classes** as the output map
- ü Click **OK** to derive the result
- ü Close the *Map Analysis* window

Display as side-by-side maps in 2D Grid format (*Use Cells* button...remember?).





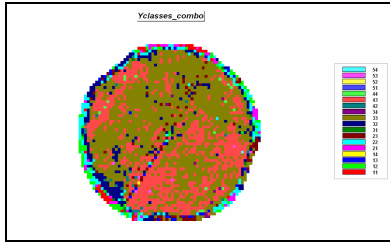
 Press the **Map Analysis** button to access the analytical operations, select **Overlay** → **Calculate** and complete the dialog box shown below.



*Calculate (97Yield_classes * 10) + 98Yield_classes For Yclasses_combo*

For convenience, use the *Maps* menu pull-down to select the input and output maps as you construct the equation. You can use the *Functions* pull-down menu to select mathematical operators or simply enter them via the keyboard.

  Click on the **Use Cells** and **Data Type** buttons to switch the default 2D continuous lattice display to 2D discrete grid format.




The values on the derived map *Yclasses_combo* form a two-digit code with the first value (tens digit) identifying the 1997 yield class and the second value (ones digit) identifying the 1998 yield class. For example, the value “forty-three” is interpreted as yield class 4 (180-240bu) in 1997 but decreasing to yield class 3 (120-1280bu) in 1998.


Clicking on the **Shading Manager** button pops-up a summary of the joint coincidence of the two yield maps.


Category	Count	acres	% Grided Area	Color
54	1	0.0574	0.03	Cyan
53	7	0.402	0.21	Magenta
52	1	0.0574	0.03	Yellow
51	1	0.0574	0.03	Blue
44	18	1.03	0.55	Green
43	905	51.9	28	Red
42	27	1.55	0.82	Orange
34	10	0.574	0.3	Purple
33	1648	94.6	50	Dark Green
32	227	13	6.9	Light Green
31	10	0.574	0.3	Dark Blue
23	94	4.82	2.6	Brown
22	144	8.26	4.4	Light Blue
21	45	2.58	1.4	Pink
14	1	0.0574	0.03	Yellow
13	33	1.89	1	Blue
12	87	4.99	2.6	Red
11	40	2.3	1.2	Green

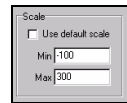
Note that most of the field was classified as class three (120-180bu) in both periods— code 33= 1648 cells representing 94.6 acres representing 50 percent of the field.

12.6.2 Comparing Yield Surfaces

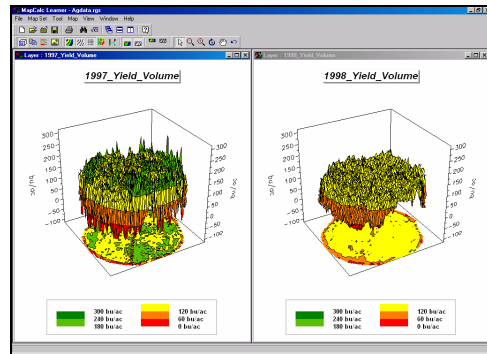
 Create side-by-side displays of the **1997_Yield_Volume** and the **1998_Yield_Volume** maps.

 Use the **3D** buttons to change the display format of both maps. Right-click on the **1997_Yield_Volume** map and select **Properties** → **Legend Tab**.


 Select **Bottom** in the *Position Legend* window.

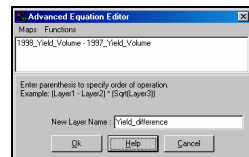
 Select the **Plot Cube** tab and un-check the *Use Default Scale* checkbox then enter **-100** as the *Min* and **300** as the *Max* for the Z-axis.


Repeat the same custom display settings for the *1998_Yield_Volume* map.



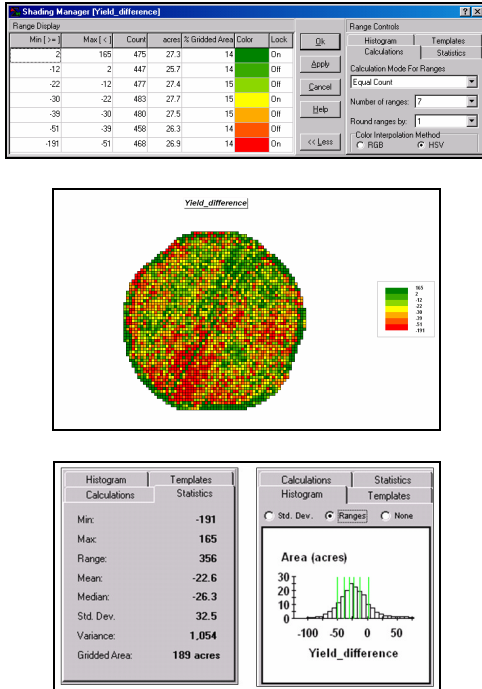
Note the relationship between the 2D projected display and the 3D surface in both displays. Also note the relative height of both plots using a common Z-axis scale. The 1998 plot displays considerably less yield variability (peaks and valleys).

 Press the **Map Analysis** button to access the analytical operations, select **Overlay** → **Calculate** and complete the following dialog box.

 Calculate **1998_Yield_Volume - 1997_Yield_Volume** for *Yield_difference*

 Use the **Use Cells** button to switch to grid display format.

Right-click on the map, select the **Shading Manager** option and change the *Calculation Mode* from *Equal Ranges* to **Equal Counts**. Press **OK** to re-display.



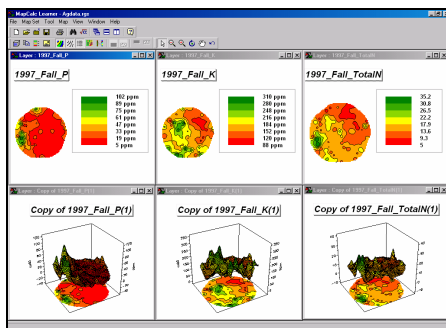
Note that the yield difference ranges from a -191 bu/ac decrease to a 165 bu/ac increase with an average difference of -22.6.

12.6.3 Characterizing Map Similarity

Based on your previous experience in map analysis create a map of total nitrogen for the 1997 period by summing...

1997_Fall_NH4N
 + 1997_Fall_NO3N
 1997_Fall_TotalN *Hint: Calculate command.*


Based on your previous experience in displaying maps generate the following side-by-side 2D and 3D displays of the 1997_Fall_P, 1997_Fall_K and 1997_Fall_TotalN maps...

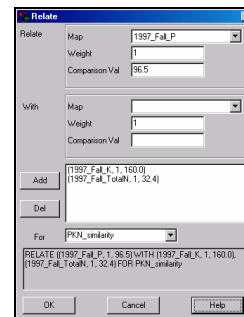


Hint: Use the View button (binocular icon) to clone the P, K and TotalN maps. Use the 3D Toggle button to switch to 3D display for the cloned map displays. Click on the Title Vertically button and click/drag to arrange the six displays.

Make a mental note of any similarities in the P, K, and TotalN levels in the graphical displays. For example you might “see” generally low values for all three nutrients occurring in the northeastern portion of the field.

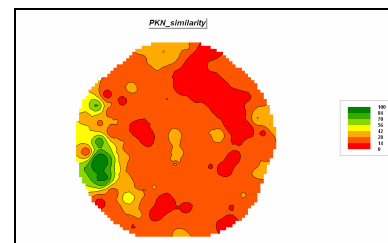
Maximize the 1997_Fall_P window, press the Use Cells button to switch to grid display format and then press the Layer Mesh button to turn on the analysis grid for reference. Move the cursor to the field location Column=8, row= 24 using the coordinate reference in lower left corner of the screen. Double-click at this location and note the 1997_Fall_P, _K and _TotalN values—96.5, 160, and 32.4, respectively—that represent the “data pattern” for the location.

 Press the **Map Analysis** button to access the analytical operations, select **Statistics** > **Relate** and complete the dialog box shown below to generate a similarity map to the data pattern at field location column=8, row=24.

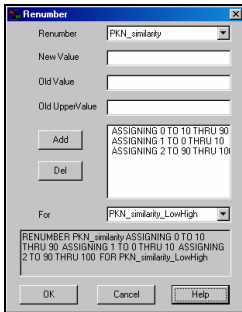


RELATE ((1997_Fall_P, 1, 96.5) WITH (1997_Fall_K, 1, 160.0), (1997_Fall_TotalN, 1, 32.4) FOR PKN_similarity

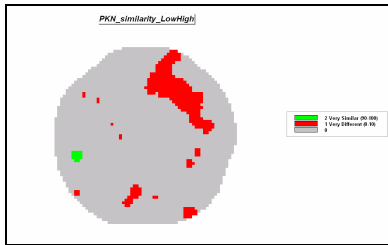
The resulting map shows how similar (red= low to green= highly similar) all map locations are to the data pattern that was entered—the western portion of the field has the most similar data patterns.



Select **Map Analysis** → **Reclassify** → **ReNUMBER** and complete the following dialog box to isolate the locations that are very different (0-10 percent similar) and are very similar (90-100 percent similar).



*RENUMBER PKN_similarity
ASSIGNING 0 TO 10 THRU 90 ASSIGNING 1 TO 0 THRU 10
ASSIGNING 2 TO 90 THRU 100 FOR PKN_similarity_LowHigh*

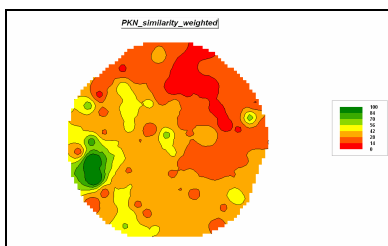


P, K and N conditions are very different in the northeast portion of the field with very similar location localized around the comparison location.

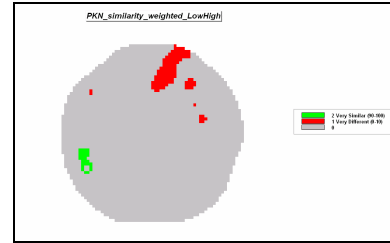
Repeat the similarity analysis but this time weighting the TotalN levels as five times more important in determining similarity than P and K levels...

- 1997_Fall_P weight= 1
- 1997_Fall_K weight= 1
- 1997_Fall_TotalN weight= 5

...to generate a weighted similarity map—



...and a map isolating Low/High similarity—

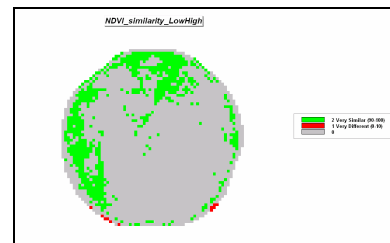
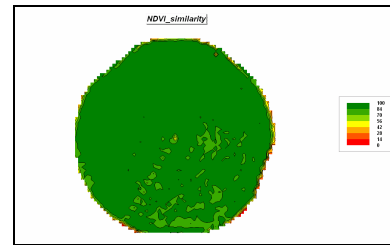


Did the change in relative weighting of the data pattern make a significant difference in the similarity maps produced? How could you quantify the difference (Hint: see Topic 3, Comparing Mapped Data)?


On your own, generate another similarity analysis (map variable weights= 1) using the normalized density vegetation index (NDVI) map series...

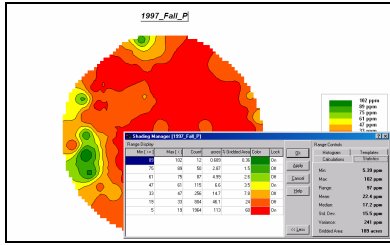
- 2000_Image_7_21_NDVI
- 2000_Image_8_10_NDVI
- 2000_Image_8_30_NDVI

...at field location column=38, row=51. Complete the analysis by identifying the portions of the field that have very different and very similar data patterns.



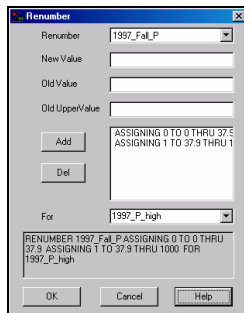
12.6.4 Identifying Data Zones

 Display the **1997_Fall_P** map, press the **Shading Manager** button and select the **Statistics** tab.



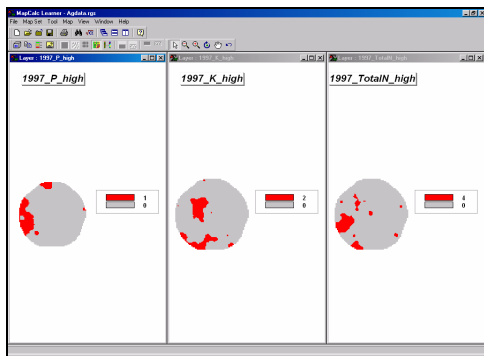
Note that the Mean value of phosphorous is 22.4ppm with a standard deviation of 15.5ppm. The cutoff for unusually high levels of phosphorous is 37.9ppm (22.4 + 15.5).

Select **Map Analysis** → **Reclassify** → **Renumber** to generate a binary map of unusually high 1997_Fall_P levels by completing the following.

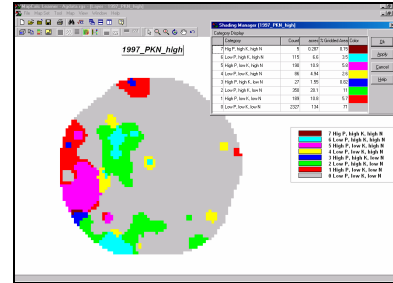


*RENUMBER 1997_Fall_P
ASSIGNING 0 TO 0 THRU 37.9
ASSIGNING 1 TO 37.9 THRU 1000
FOR 1997_P_high*

Repeat the same processing to isolate the areas of *high K* levels in fall 1997 as **2** and *high TotalN* levels as **4**.



Select **Map Analysis** → **Overlay** → **Calculate** and add the three binary maps together.



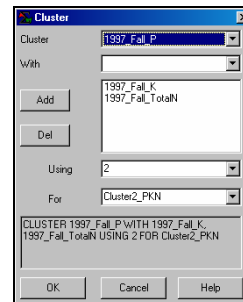
The summed values identify unique combinations of the three maps—

- 7= High P, high K, high N
- 6= Low P, high K, high N
- 5= High P, low K, high N
- 4= Low P, low K, high N
- 3= High P, high K, low N
- 2= Low P, high K, low N
- 1= High P, low K, low N
- 0= Low P, low K, low N

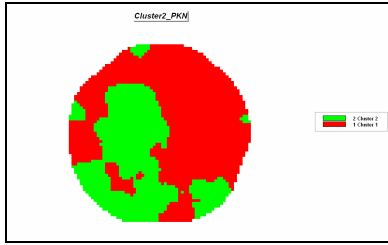
Most of the field (71% grey) does not contain areas with unusually high levels of nutrients. Note that most of the high nutrient coincidence occurs in the western half of the field. How might you use this information in planning a site-specific fertilization program?

12.6.5 Mapping Data Clusters

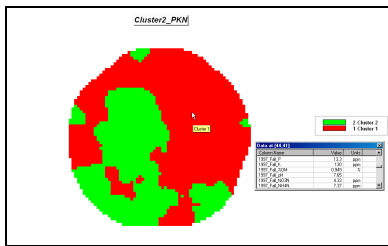
The Cluster command can be used to partition the field into inherent data pattern groups. Relative positioning in data space determines the membership of the clusters. Generate a two-cluster grouping of the field using the 1997_Fall P, K and N maps by selecting **Map Analysis** → **Statistics** → **Cluster** and entering—



*CLUSTER 1997_Fall_P WITH
1997_Fall_K, 1997_Fall_TotalN USING 2 FOR Cluster2_PKN*

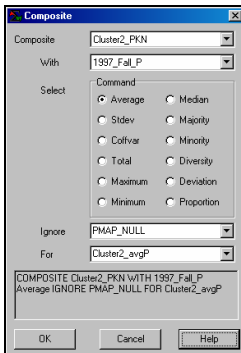


Use the “Data Drill-down” feature (double-click on the Cluster2_PKN map) and note how the P, K and N values change within the two clusters by moving the cursor around the map. What was the typical range of P values for Cluster 1? For cluster 2?



Select **Map Analysis** → **Overlay** → **Composite**

...



COMPOSITE Cluster2_PKN WITH 1997_Fall_P Average IGNORE PMAP_NULL FOR Cluster2_avgP

... to calculate the average P value for both clusters. Repeat the procedure to calculate the standard deviation in phosphorous levels in both clusters. Based on the avgP and stdP results is there a “substantial” difference between the two clusters in their phosphorous levels? Hint: *avg + std* ranges should not overlap very much.

Note the careful use of the word “substantial” difference as the simple range test does not allow the statement that there is or isn’t a statistical “significant” difference. A statistical test, such as an T-test or F-test, between the two data

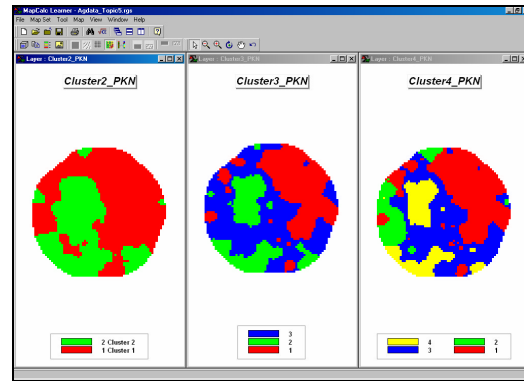
groups would be required (see *Appendix B, Exchanging Data With MapCalc* for procedures to export the data for further statistical testing in Excel or other system).

Repeat the analysis for the K and N levels in the two clusters. Is there a “substantial” difference in the levels within the two clusters for either of these nutrients?

Now create cluster maps containing three and four clusters—

CLUSTER 1997_Fall_P WITH 1997_Fall_K, 1997_Fall_TotalN USING 3 FOR Cluster3_PKN

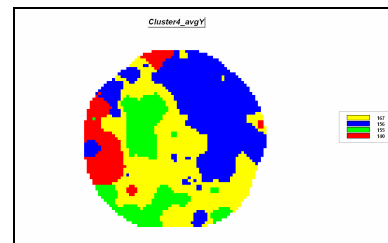
CLUSTER 1997_Fall_P WITH 1997_Fall_K, 1997_Fall_TotalN USING 4 FOR Cluster4_PKN



—see any consistency in the partitioning as the number of clusters gets larger?

Use the Composite command to calculate the average yield within each of the four clusters depicted on the Cluster4_PKN map.

COMPOSITE Cluster4_PKN WITH 1997_Yield_Volume Average IGNORE PMAP_NULL FOR Cluster4_avgY



Does there appear to be a progression in the average yield among the four clusters?

Repeat the processing to determine the average yield within the clusters on the Cluster3_PKN and

Cluster2_PKN maps. Can you discern any consistent patterns in the results?
