

Topic 11

Surface Modeling

11.1 The Average Is Hardly Anywhere

Remember your first encounter with statistics? In the days of old before politically correct examples, you might have calculated the average weight of the students in the class by adding up all of the students' weights and then dividing by the number of students. The *average* weight gives you an estimate of how heavy a "typical student" is and the *standard deviation* tells you how typical that typical is.

Now imagine your old classroom with the bulky jocks in the back, the diminutive bookworms in front, and the rest caught in between. Two things should come to mind—1) not all of the students had the same typical weight (some were heavier and some were lighter) and 2) the differences from the typical weight might have followed a geographic pattern (heavy in the back, light in the front). The second point forms the foundation of surface modeling that "maps the variation in geographic data sets."

Figure 11-1 illustrates the spatial and non-spatial character of a set of animal activity data. The right side of the figure lists the number of sightings at sixteen locations for two 24-hour periods (P1 in June; P2 in August). Note the varying levels of activity— 0 to 42 for P1 and 0 to 87 for P2 that can be reduced to their average values of 19 and 23, respectively. A wildlife manager might ponder these findings, and then determine whether the implied typical activity is too little, too much or just right for a particular management action.

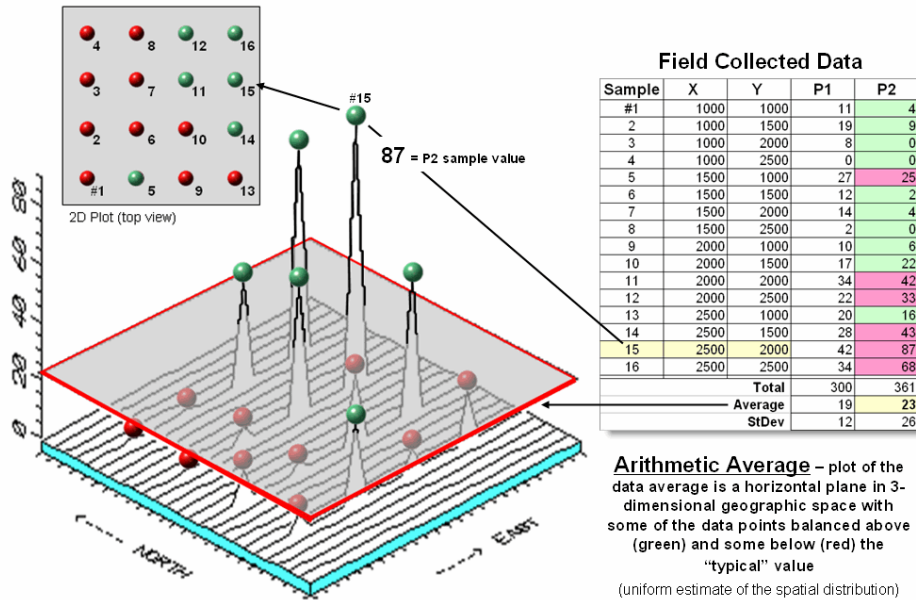


Figure 11-1. Comparison of the spatial distributions of field samples (floating balls) and their arithmetic average (horizontal plane).

But the average does not address the variation in the data set— that's the role of the standard deviation. As a general rule (termed the Coefficient of Variation) "...if the standard deviation is relatively large compared to the arithmetic average, the average cannot be used to make decisions" as there is too much unexplained variation in the data (i.e., the computed average isn't very typical).

Possibly some of the variation in animal activity forms a pattern in geographic space that might help extend the explanation. That is where the left side of figure 1 comes into play with a three-dimensional plot used to show the geographic location (X, Y) and the measured activity levels (Z). I'll bet your eye is “mapping the variation in the data” as high activity in the Northeast, low in the Northwest and moderate elsewhere.

The thick line in the plot outlines a horizontal plane at 23 (arithmetic average) that spatially characterizes the typical animal activity as uniformly distributed in geographic space (horizontal plane). But your eye tells you that guessing 23 around Sample #15 with a measured value of 87, is likely an understatement. Similarly, a guess of 23 around Sample #4 with a value of 0 is likely an overstatement. That is what the relatively large standard deviation indicated— guess 23 but expect to be way off (± 26) a lot of the time.

The non-spatial procedure, however, doesn't provide a hint as to where the average might be guessing too low and where it might be guessing high. That's the main difference between traditional statistics and spatial statistics— traditional statistics characterizes the central tendency (average) of data in numeric space; spatial statistics seeks to map the variation (standard deviation) of data in geographic space.

Figure 11-2 illustrates *Map Generalization* as an approach to mapping the spatial trend in the data using polynomial surface fitting. The average of 23 is depicted as a horizontal plane that does its best to balance half of the balls above it and half below it by minimizing the set of squared deviations from the plane to each floating ball (similar to curve-fitting regression techniques in traditional statistics).

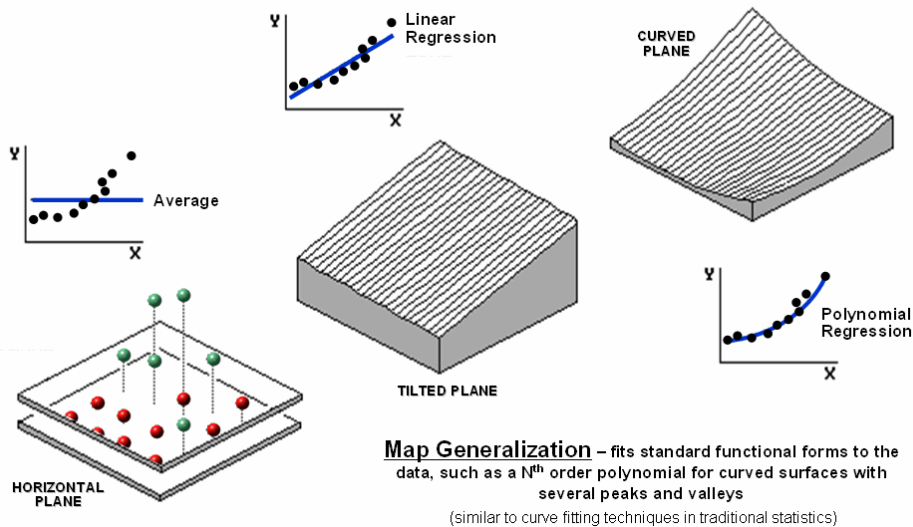


Figure 11-2. Map Generalization can be used to approximate geographic trends.

Now relax the assumption that the plane has to remain horizontal. Tilt it every-which-way until it better fits the floating balls (Tilted Plane). Or even better, the assumption of a flat plane can be relaxed and the surface can curve to better fit the ups and downs of the data points. The right portion of the figure fits a 2nd degree polynomial (Curved Plane).

Figure 11-3 shows an entirely different approach to fitting a surface to the data by using the *Spatial Interpolation* technique of iterative smoothing. Imagine replacing the floating balls with columns of modeler's clay rising to the same height as each ball. In effect this is a blocky first-order estimate of the animal activity throughout the project area derived by simply assigning the closest field sample to each map location.

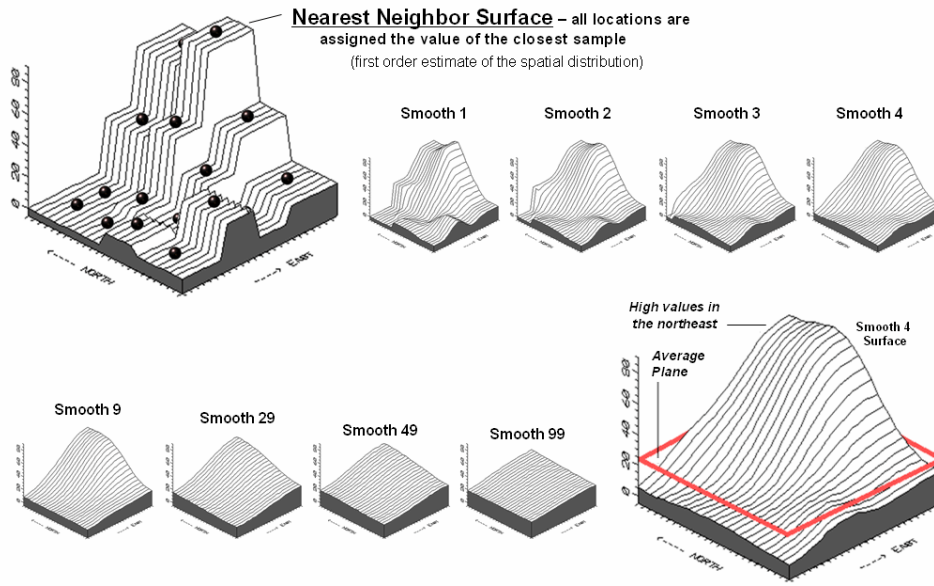


Figure 11-3. Iteratively Smoothed approximations of localized geographic distribution (see Author’s Note).

Now imagine whacking away some of the clay at the top of the columns and filling-in at the bottom. In the example, a 3x3 averaging window was moved around the matrix of values to systematically smooth the blocky surface. When the window is centered over one of the sharp boundaries, it has a mixture of small and larger map values, resulting in an average somewhere in between... a localized whack off the top and a fill-in at the bottom.

The series of plots in the figure show repeated passes of the smoothing window from once through ninety-nine times. Like erosion, the mountains (high animal activity) are pulled down and the valleys (low animal activity) are pulled up. According to theory, the process eventually approximates a horizontal plane floating at the arithmetic average.

The bottom line is that field collected data with geographic coordinates holds a lot more information than simply a reduction to a typical value. Next month’s column will investigate other more powerful methods for estimating the spatial, as well as numerical distribution inherent in mapped data.

Author’s Note: See the online book [Map Analysis](http://www.innovativegis.com/basis/) posted at www.innovativegis.com/basis/, Topic 2 “Spatial Interpolation Procedures and Assessment,” section 1 for a link to an animated slide set of the iterative smoothing depicted in figure 11-3.

11.2 Under the Hood of Spatial Interpolation

Last month’s column described how field collected data (discrete points) can be used to generate a map (continuous surface) of the data’s spatial patterns. The derived surface extends the familiar concept of central tendency to a map of the geographic distribution of the data. Whereas traditional statistics identifies the typical value in a data set, surface modeling identifies "where" you might expect to find the typical and not so typical responses.

The Iterative Smoothing approach described last time is a simple data-driven procedure. However, all of the interpolation techniques share a similar approach that generates estimates of a mapped variable based on the data values within the vicinity of each map location. In effect, this establishes a "roving window" that moves throughout an area summarizing the field samples it encounters within its reach. The summary estimate is assigned to the center of the window, and then it moves on. The extents of the window (both size and shape) influence the result, regardless of the summary technique. In general, a large window

capturing numerous values tends to smooth the data. A smaller window tends to result in a rougher surface.

Three factors affect the window's configuration— its reach, number of samples, and balancing. The *reach*, or search radius, sets a limit on how far the computer will go in collecting data values. The *number of samples* establishes how many data values will be used. The *balancing* of the data attempts to eliminate directional bias by insuring that values are selected from all directions around the window's center.

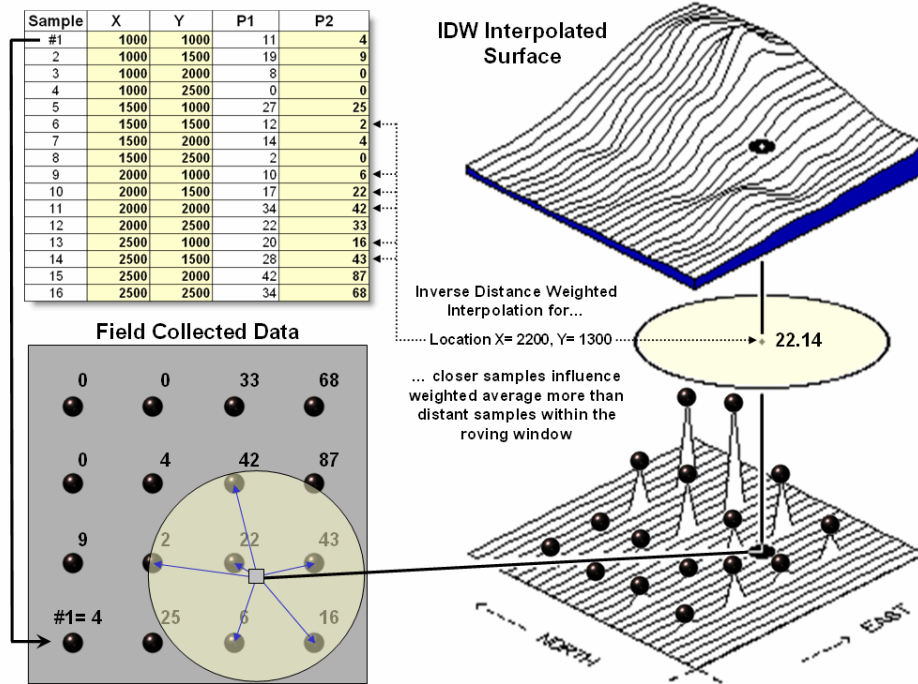


Figure 11-4. A roving window is used to identify and summarize nearby sample values.

Once a window is established, the summary technique comes into play. The *Inverse Distance Weighted (IDW)* technique is easy to conceptualize (figure 11-4). It estimates a value for a location as the weighted-average of the nearby data values within the roving window. The average is weighted so the influence of the surrounding values decrease with increasing distance from the location being estimated.

Figure 11-5 shows the IDW calculations for the location identified in figure 11-4. The Pythagorean Theorem is used to calculate the geographic distance between the location and the six sample values in the window. A weight for each sample is determined as $1/D^2$, then the weighted average of the samples is calculated and assigned to the location. The process is repeated for each location in the project area.

Because IDW is a static averaging method, the estimated values can never exceed the range of values in the original sample data. Also, it tends to "pull-down peaks and pull-up valleys" in the data. Inverse distance is best suited for data sets where the samples are fairly independent of their surrounding locations (i.e., no strong regional trend).

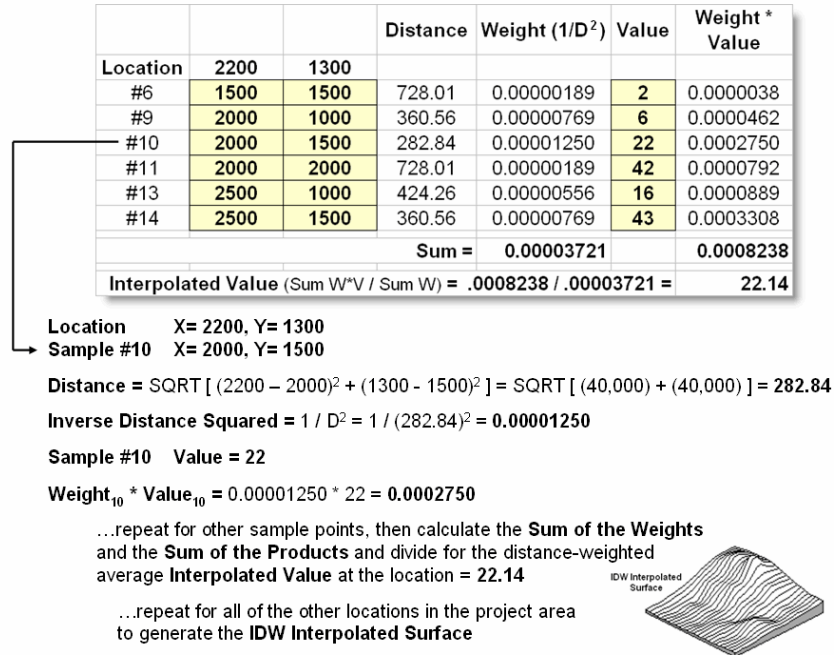


Figure 11-5. IDW calculations use simple geographic distance to weight the average of samples within a roving window.

Another approach, *Kriging (KRIG)* uses regional variable theory based on an underlying variogram. That's techy-speak implying that there is a lot of math behind the approach. In effect, the technique develops custom weights based on trends in the sample data. The moving average weights that define the data trends can result in estimated values that exceed the field data's range of values. Also, there can be unexpected results in large areas without data values. The technique is most appropriate for fairly dense, systematically sampled data exhibiting discernable regional trends.

The center portion figure 11-6 illustrates how the KRIG weights are derived. The distances between each sample point and all of the other sample points are calculated. At the same time, the differences in the sample values are recorded. Common sense suggests that “nearby things are more alike than distant things” and a plot of Distance versus Difference often looks something like the idealized *variogram* in the center of the figure.

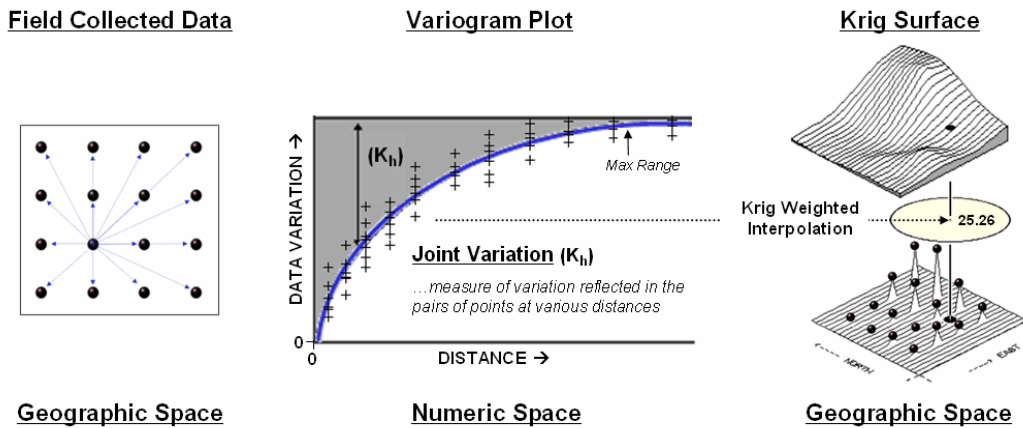


Figure 11-6. A variogram plot depicts the relationship between distance and similarity among sample value pairings.

The origin of the plot at $0,0$ is a unique case. The distance between samples is zero; therefore, there shouldn't be any dissimilarity (data variation = zero) as the location is exactly the same as itself. As the distance between points increase, subsets of the data are scrutinized for their dependency. The shaded portion in the idealized plot shows a rapid deterioration of the spatial dependency among the sample points. The maximum range (*Max Range*) position identifies the distance between points beyond which the data values are considered independent. This suggests that using data values beyond this distance for interpolation isn't useful and sets the reach of the roving window.

An equation is fit to the Distance-Difference data and used to determine the weights used in averaging the sample values within the roving window. Whereas IDW uses a fixed geometric function, KRIG derives the weights by investigating the spatial dependency in the sample data (see Author's Note). Keep in mind that most analytical GIS systems have a dozen or more spatial interpolation techniques— but they all share the common basics of a window configuration and a summary procedure.

Author's Note: Let me apologize in advance for such a terse treatment of a complex subject. Topic 8, *Investigating Spatial Dependency* of the online book *Map Analysis* (www.innovativegis.com/basis/) discusses in greater depth the various measures of spatial autocorrelation, their interpretation and use in interpolation.

11.3 From Point Samples to Map Surfaces

Soil sampling has long been at the core of agricultural research and practice. Traditionally point-sampled data were analyzed by non-spatial statistics to identify the “typical” nutrient level throughout an entire field. Considerable effort was expended to determine the best single estimate and assess just how good the “average” estimate was in typifying a field.

However non-spatial techniques fail to make use of the geographic patterns inherent in the data to refine the estimate—the typical level is assumed everywhere the same within a field. The computed variance (or standard deviation) indicates just how good this assumption is—the larger the standard deviation the less valid is the assumption “everywhere the same.”

Surface modeling, on the other hand, utilizes the spatial patterns in a data set to generate localized estimates throughout a field. Conceptually it “maps the variance” by using geographic position to help explain the differences in the sample values. In practice, it simply fits a continuous surface (kind of like a blanket) to the point data spikes (figure 11-7).

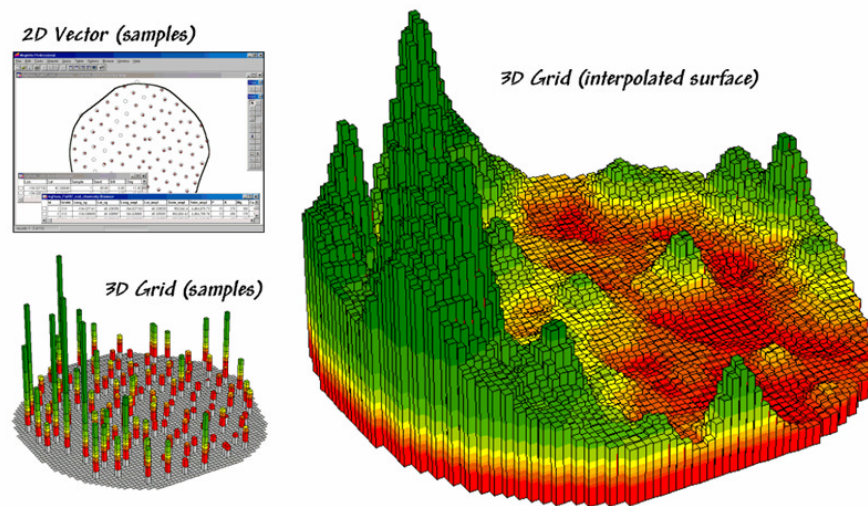


Figure 11-7. Spatial interpolation involves fitting a continuous surface to sample points

While the extension from non-spatial to spatial statistics is quite a theoretical leap, the practical steps are relatively easy. The left side of figure 11-7 shows 2D and 3D *point* maps of phosphorous soil samples collected throughout the field. This highlights the primary difference from traditional soil sampling—each sample must be *geo-referenced* as it is collected. In addition, the *sampling pattern and intensity* are often different than traditional grid sampling to maximize spatial information within the data collected*.

The surface map on the right side of figure 11-7 depicts the continuous spatial distribution derived from the point data. Note that the high spikes in the western portion of the field and the relatively low measurements in the center are translated into the peaks and valleys of the surface map.

The traditional, non-spatial approach, if mapped, would be a flat plane (average phosphorous level) aligned within the bright yellow zone. Its “everywhere the same” assumption fails to recognize the patterns of larger levels (greens) and smaller levels (reds). A fertilization plan for phosphorous based on the average level (22ppm) would be ideal for the yellow zone but would likely be inappropriate for a lot of the field as the data vary from 5 to 102ppm phosphorous.

It’s common sense that nutrient levels vary throughout a field. However, there are four important considerations before surface modeling is appropriate—

- the pattern needs to be significant enough to warrant variable-rate application (*cost effective*),
- the prescription methods must be spatially appropriate for the pattern (*decision rules*),
- the derived map surface must fully reflect a real spatial pattern (*map accuracy*), and
- the procedure for generating the pattern must be straightforward and easy to use (*usability*).

Let’s tackle the last point on usability first. The mechanics of generating an interpolated surface involves three steps relating to *geography, data* and *interpolation methods* (figure 11-8).

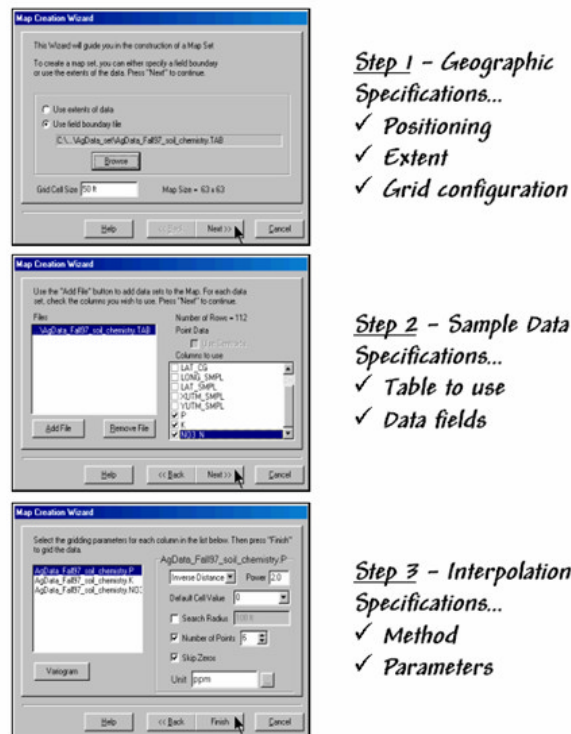


Figure 11-8. A wizard interface guides a user through the necessary steps for interpolating sample data.

The first step establishes the geographic position, field extent and grid configuration to be used. While these specifications can be made directly, it is easiest to simply reference an exiting map of the field boundary (*Positioning* and *Extent*) then enter the grid spacing (*Configuration*—50 feet).

The next step identifies the table and data fields to be used. The user navigates to the file (*Data Table*) then simply checks the maps to be interpolated (*data fields*—P, K, NO3_N).

The final step establishes the interpolation method and necessary factors. In the example, the default *Inverse Distance Squared ($1/d^2$)* method was employed using the six nearest sample points. Other methods, such as Kriging, could be specified and would result in a somewhat different surface.

Figure 11-9 shows the P, K, N map surfaces that were generated by the interpolation specifications shown in figure 11-8. The entire process including entering specifications, calculation of the surfaces and display took less than a minute.

In each instance the yellow zone contains the average for the entire field. Visually comparing the relative amounts and patterns of the green and red areas gives you an idea of the difference between the assumption that the average is everywhere and amount of spatial information inherent in a soil sample data set.

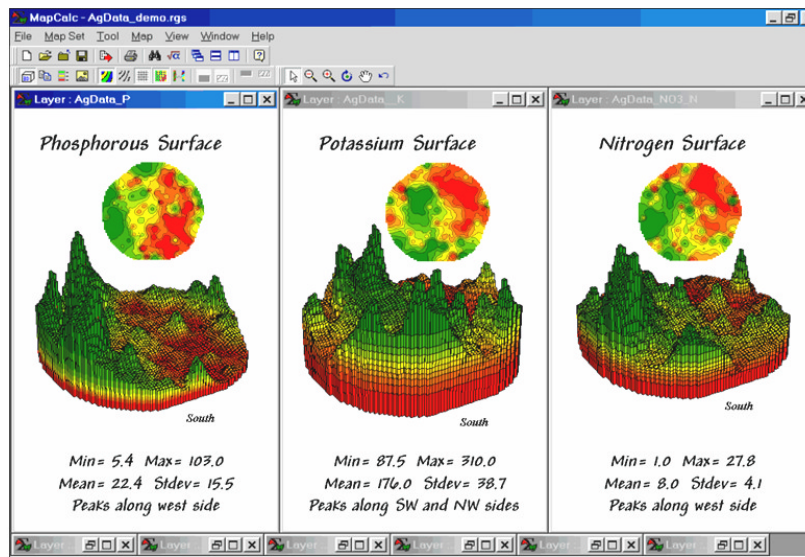


Figure 11-9. Interpolated phosphorous, potassium and nitrogen surfaces.

The next couple of sections investigate how to evaluate the accuracy of these maps—whether they’re good maps or bad maps.

11.4 Benchmarking Interpolation Results

For some, the previous discussion on generating maps from soil samples might have been too simplistic—enter a few things then click on a data file and, *viola*, you have a soil nutrient surface artfully displayed in 3D with a bunch of cool colors draped all over it.

Actually, it is that easy to create one. The harder part is figuring out if the map generated makes sense and whether it is something you ought to use in analysis and important management decisions. This section discusses the relative amounts of spatial information provided in a whole-field average and site-specific map by comparing the average and two different interpolated map surfaces. The next section describes a procedure to quantitatively assess whether a particular map is a keeper.

The top-left item in Figure 11-10 shows the map of a field’s average phosphorous levels. It’s not very exciting and looks like a pancake but that’s because there isn’t any information about spatial variability in an average value—assumes 22ppm is everywhere in the field.

The non-spatial estimate simply adds up all of the sample measurements and divides by the number of samples to get 22ppm. Since the procedure didn't consider where the different samples were taken, it can't map the variations in the measurements. It thinks the average is everywhere, plus or minus the standard deviation. But there is no spatial guidance where phosphorous levels might be more, or where they might be less.

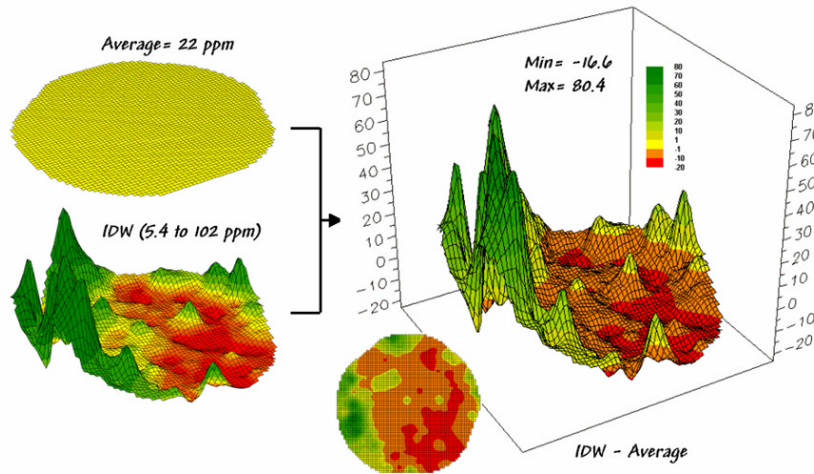


Figure 11-10. Spatial comparison of a whole-field average and an IDW interpolated map.

The spatially based estimates are shown in the map surface just below the pancake. As described in the last section, spatial interpolation looks at the relative positioning of the soil samples as well as their measure phosphorous levels. In this instance the big bumps were influenced by high measurements in that vicinity while the low areas responded to surrounding low values.

The map surface in the right portion of figure 11-10 objectively compares the two maps simply by subtracting them. The colors were chosen to emphasize the differences between the whole-field average estimates and the interpolated ones. The yellow band indicates the average level while the progression of green tones locates areas where the interpolated map thought there was more phosphorous than the average. The progression of red tones identifies the opposite with the average estimate being more than the interpolated ones.

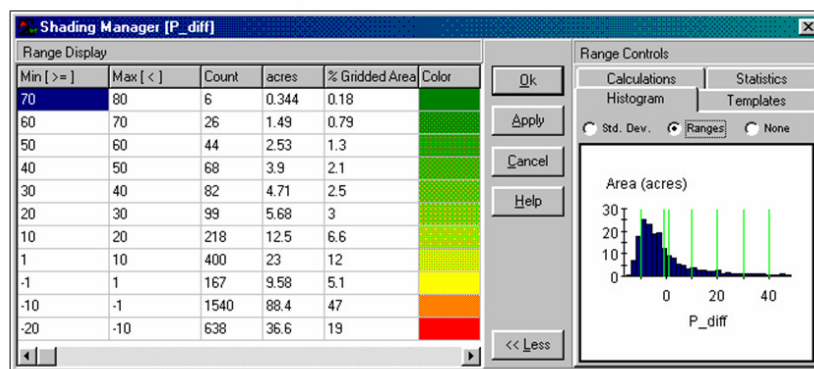


Figure 11-11. Statistics summarizing the difference between the maps in figure 11-5.

The information in figure 11-11 shows that the difference between the two maps ranges from -20 to +80ppm. If one assumes that +/- 10ppm won't significantly alter a fertilization recommendation, then about two-thirds (47 +5.1 +12= 64.1 percent of the field) is adequately covered by the whole-field average.

But that leaves about a third of the field that is receiving too much (6.6 + 3+ 2.5 +2.1 +1.3 +.79 +.18= 16.5 percent) or too little (19 percent) phosphorous compared to a fertilization program based on the whole-field average

Now turn your attention to figure 11-12 that compares maps derived by two different interpolation techniques—IDW (inverse distance-weighted) and Krig. Note the similarity in the peaks and valleys of the two surfaces. While subtle differences are visible the general trends in the spatial distribution of the data are identical.

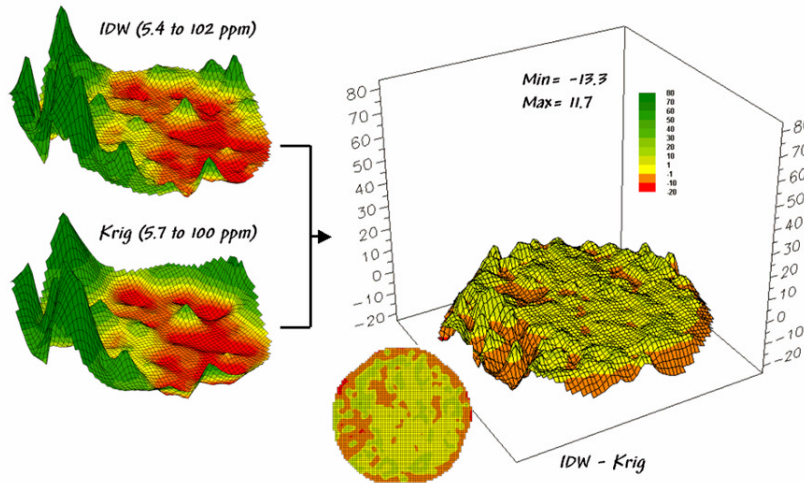


Figure 11-12. Spatial comparison of IDW and KRIG interpolated maps.

The difference map on the right confirms the coincident trends. The broad band of yellow identifies areas that are +/- 1 ppm. The brown color identifies areas that are within 10 ppm with the IDW surface estimates a bit more than the KRIG ones. Applying the same assumption about +/- 10 ppm difference being negligible in a fertilization program the maps are effectively identical.

So what’s the bottom line? That there are substantial differences between a whole field average and interpolated surfaces—at least for this data set. It suggests that quibbling about the best interpolation technique isn’t as important as using an interpolated surface (any surface) over the whole field average. However, what needs to be addressed is whether an interpolated surface (any surface) actually reflects the real spatial distribution. That weighty question is the focus of the next section.

11.5 Assessing Interpolation Results

Last section’s discussion compared the assumption of the field average with map surfaces generated by two different interpolation techniques for phosphorous levels throughout a field. While there was considerable differences between the average and the derived surfaces (from -20 to +80ppm), there was relatively little difference between the two surfaces (+/- 10ppm).

But which surface best characterizes the spatial distribution of the sampled data? The answer to this question lies in **Residual Analysis**—a technique that investigates the differences between *estimated* and *measured* values throughout a field. It’s common sense that one shouldn’t simply accept a soil nutrient map without checking out its accuracy. Cool graphics just aren’t enough.

Ideally, one designs an appropriate sampling pattern and then randomly locates a number of “test points” to assess interpolation performance. Since a lot is riding on the accuracy of the interpolated maps, it makes sense to invest a little bit more just to see how well things are going.

So which surface, IDW or KRIG, did a better job in estimating the measured phosphorous levels in the test set? The table in figure 11-13 reports the results for twelve randomly positioned test samples. The first column identifies the sample ID and the second column reports the actual measured value for that location.

...Residual Analysis is used to evaluate interpolation performance (IDW at .18 appears best)

	A	B	C	D	E	F	G	H
1	Test Set (randomly selected)							
2	Id	P_actual	Average	Avg-Actual	P_idw	IDW-Actual	P_Krig	Krig-Actual
3	59	20.0	21.6	1.6	18.80	-1.20	17.80	-2.20
4	2	15.0	21.6	6.6	11.70	-3.30	13.50	-1.50
5	25	5.0	21.6	16.6	5.39	0.39	5.69	0.69
6	63	27.0	21.6	-5.4	28.80	1.80	27.40	0.40
7	18	9.0	21.6	12.6	9.68	0.68	7.90	-1.10
8	34	14.0	21.6	7.6	12.80	-1.20	13.30	-0.70
9	22	10.0	21.6	11.6	10.00	0.00	10.20	0.20
10	91	20.0	21.6	1.6	31.60	11.60	42.30	22.30
11	109	103.0	21.6	-81.4	91.10	-11.90	90.70	-12.30
12	46	12.0	21.6	9.6	16.80	4.80	15.80	3.80
13	89	93.0	21.6	-71.4	67.60	-25.40	68.60	-24.40
14	50	24.0	21.6	-2.4	24.60	0.60	27.40	3.40
15								
16	Average values=	29.3	21.6		27.41		28.38	
17	Residual sum=			-92.8		-23.1		-11.4
18	Average error=			19.0		5.24		6.08
19	Normalized error=			0.65		0.18		0.21

Figure 11-13. A residual analysis table identifies the relative performance of average, IDW and Krig estimates.

Column C simply depicts estimating the whole-field average (21.6) at each of the test locations. Column D computes the difference of the “estimate minus actual”—formally termed the *residual*. For example, the first test point (ID#59) estimated the average of 21.6 but was actually measured as 20.0 so the residual is 1.6 (21.6-20.0= 1.6ppm)... not bad. However, test point #109 is way off (21.6-103.0= -81.4ppm)... nearly 400% under-estimate error.

The residuals for the IDW and KRIG maps are similarly calculated to form columns F and H, respectively. First note that the residuals for the whole-field average are generally larger than either those for the IDW or KRIG estimates. Next note that the residual patterns between the IDW and KRIG are very similar—when one is way off, so is the other and usually by about the same amount. A notable exception is for test point #91 where KRIG dramatically over-estimates.

The rows at the bottom of the table summarize the residual analysis results. The *Residual sum* row characterizes any bias in the estimates—a negative value indicates a tendency to underestimate with the magnitude of the value indicating how much. The -92.8 value for the whole-field average indicates a relatively strong bias to underestimate.

The *Average error* row reports how typically far off the estimates were. The 19.0ppm figure for the whole-field average is three times worse than KRIG’s estimated error (6.08) and nearly four times worse than IDW’s (5.24).

Comparing the figures to the assumption that a ±10ppm is negligible in a fertilization program it is readily apparent that the whole-field estimate is inappropriate to use and that the accuracy differences between IDW and KRIG are minor.

The *Normalized error* row simply calculates the average error as a proportion of the average value for the test set of samples (5.24/29.3= .18 for IDW). This index is the most useful as it allows you to compare the relative map accuracies between different maps. Generally speaking, maps with normalized errors of more than .30 are suspect and one might not want to make important decisions using them.

So what’s the bottom line? That Residual Analysis is an important component of precision agriculture. Without an understanding of the relative accuracy and interpolation error of the base maps, one can’t be

sure of the recommendations derived from the data. The investment in a few extra sampling points for testing and residual analysis of these data provides a sound foundation for site-specific management. Without it, the process can become one of blind faith and wishful thinking.

11.6 The Bigger Picture

Precision Agriculture is a new field that uses geotechnology to help understand and utilize the spatial relationships involved in crop production and management. Since its emergence in the early 1990s, yield mapping has gaining commonplace status for many crops and locales. Site-specific management of field fertilization has a growing number of users. Remote sensing applications from satellite imagery to aerial photos and on the ground “proximal” sensing are coming onboard. Irrigation control, field leveling, variable rate seeding, on-farm studies, disease/pest modeling, stress maps and a myriad other computer mapping uses are on the horizon.

Fundamental to all of these applications is the digital nature of a computer map and the procedures required to turn these data into useful information for decision-making. The bottom line for precision agriculture appears to be an understanding that maps are numbers and what one can do with the digital data to enhance their interpretation.

Site-specific farming isn’t just a bunch of pretty maps, but a set of new technologies and procedures linking mapped variables to appropriate management actions. It requires the integration of three spatial technologies: *global positioning system* (GPS), *geographic information systems* (GIS) and *intelligent devices and implements* (IDI) for on-the-fly data collection (monitors) and variable-rate application (controls) as depicted in the left-side of figure 11-14.

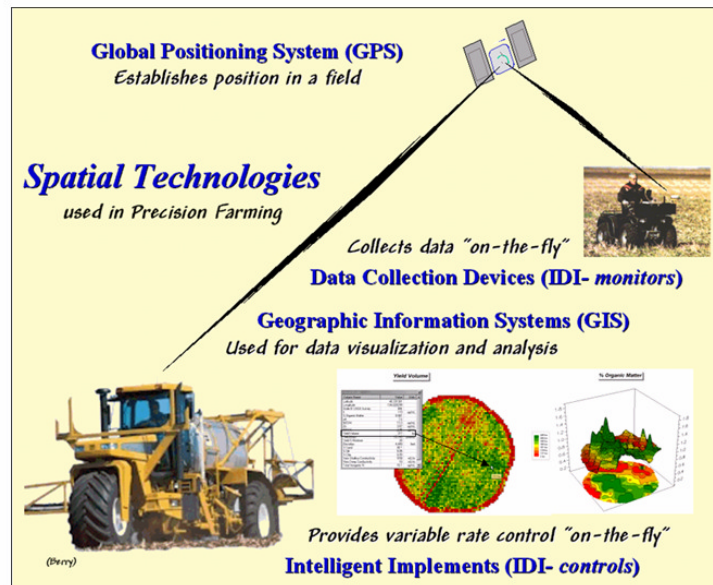


Figure 11-14. Precision Agriculture involves applying emerging spatial technologies of GPS, GIS and IDI.

Modern GPS receivers are able to establish positions within a field to few feet. When connected to a *data collection device*, such as a yield/moisture monitor, these data can be "stamped" with geographic coordinates.


As previously noted, GIS is used to map the field data so a farmer can see the variable conditions throughout a field. The GIS also is used to extend map visualization of yield to analysis of the relationships among yield variability and field conditions.

Once established these relationships are used to derive a "prescription" map of management actions required for each location in a field. The final element, *variable rate implements*, notes a tractor's position through GPS, continuously locates it on the prescription map, and then varies the application rate of field inputs, such as fertilizer blend or seed spacing, in accordance with the instructions on the prescription map.

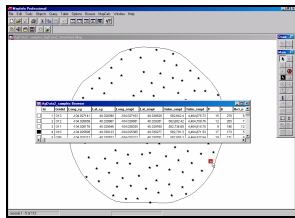
The following two topics in this book investigate how these data are analyzed and the relationships uncovered are utilized for better crop management and stewardship.

11.7 Exercises

Access *MapCalc Learner* without specifying an existing database by selecting **Start** → **Programs** → **MapCalc Learner** → **MapCalc Learner** → **Create a new map set**.

 If MapCalc is actively using another database, select **File** → **Close** then press the **Create a new file** button on the main menu bar to pop-up the first dialog box in the *Map Creation Wizard*.

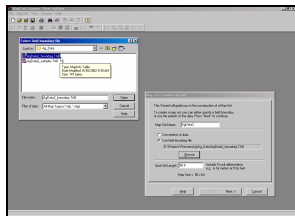
11.6.1 Interpolating Point Data (IDW)



Point Samples (MapInfo)

Note: point data for interpolation must be in either ESRI (.shp) or MapInfo (.tab) format.

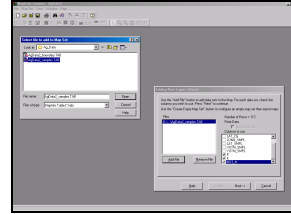
The first step in the *Wizard* identifies the *Name* of the new data set, the *Extent* (boundary) and the *Gridding Resolution* (cell size).



Wizard Step 1

- ü Enter the new map set name as **AgData2**
- ü Select **Use Field Boundary File**, browse to the *MapCalc Data* folder and specify **AgData2_boundary.tab**
- ü Specify the *Grid Cell Size* as **50** feet
- ü Press the **Next** button.

The second step identifies the point sampled data you want to map (**P** <phosphorous>, **K** <potassium>, **NO3_N** <nitrogen>). These specifications identify the “data fields” (columns) in the sample data table that will be used—the X,Y (Longitude, Latitude) positions each data point and the measured value reports the results of chemical analysis for each soil sample.

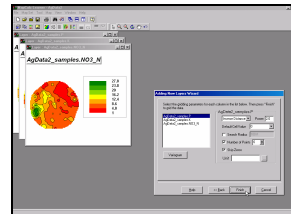


Wizard Step 2

Note: The data must be in lat/lon WGS84 projection for the *MapCalc Learner* and *Academic* versions. *MapCalc Professional* contains on-the-fly projection conversion.


- ü Press the **Add File** button and browse to the **AgData2_samples.tab** table containing the sample data
- ü Check the **P, K** and **NO3_N** as the *Columns to use* at the bottom of the listing
- ü Press the **Next** button.

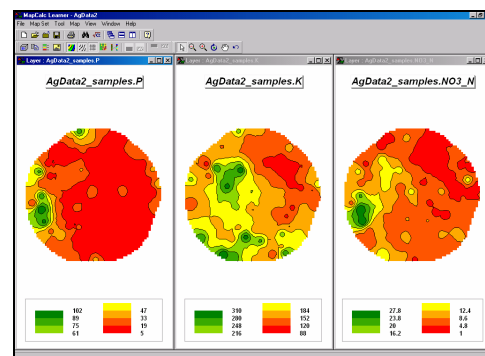
The final step specifies the interpolation technique and appropriate parameters to use.



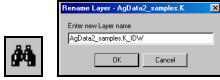
Wizard Step 3

- ü Press the **Finish** button to interpolate the three maps using the default *Inverse Distance Square* technique and parameters.

 Use the **Tile Vertically** button to generate side-by-side displays of all three maps.

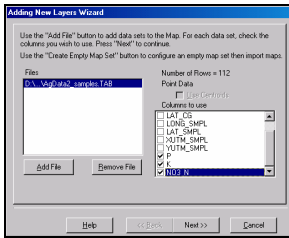


11.6.2 Interpolating Point Data (Kriging)



Click on the **View** button, select the **AgData2_samples.K** map and press the **Rename** button. Rename the map to **AgData2_samples.K_IDW** to identify the interpolation technique that was used. Repeat to add the **_IDW** extension to the **P** and **NO3_N** map names then **Close** the window for managing map layers.

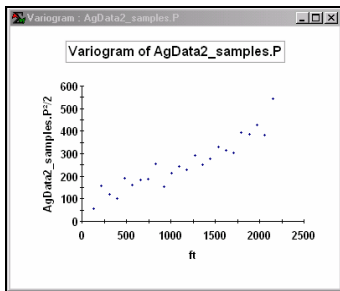
Select **Map Set** → **Add New Layers** to re-access the *Interpolation Wizard*.



Click on the **Add File** button and specify **AgData2_samples.tab** and **P**, **K** and **NO3_N** columns as before. Press **Next**.

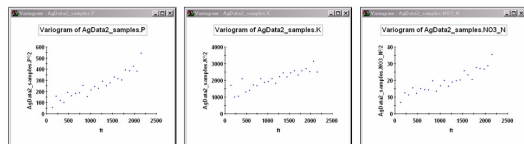


Highlight the **AgData2_sample.P** set of samples and press the **Variogram** button to generate a plot of the data's spatial autocorrelation.

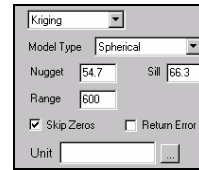


Generally speaking, a horizontal (flat) variogram indicates minimal information for interpolation and the map surface generated may be of negligible value. A variogram exhibiting a tightly-clustered linear pattern suggests better map surfacing results.

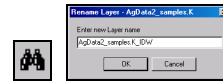
Repeat to generate *Variograms* of the other two variables.



Note that the samples for all three variables exhibit an upward-tilted linear trend that suggests that the data likely contains sufficient spatial autocorrelation for interpolation.



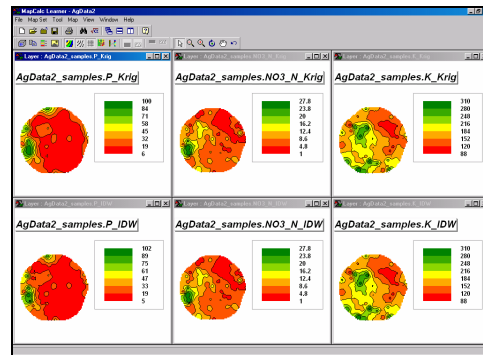
Change the technique from **IDW** to **Kriging** using the default set of parameters. Press the **Finish** button to generate another set of interpolated maps.



Rename the new set of maps with the extension **_KRIG**.



Close the windows containing the *Variogram* plots and press the **Tile Vertically** button to display all six maps at the same time. Move the map windows into a sensible order by clicking along their top edge and dragging them.



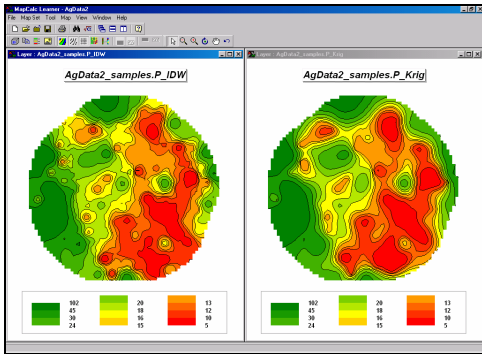
11.6.3 Comparing Interpolation Results



Click on the **Shading Manager** to generate a display of the **AgData2_samples.p_IDW** with the following formatting:

- Ü 2D Continuous lattice
- Ü Equal Count
- Ü 11 Ranges
- Ü Color Ramp with the lowest range set to **red**, highest range set to **green** and a **yellow** color inflection point at the mid range (6th interval).

Use the *Templates* tab to save the color ramp then display the **AgData2_samples.p_Krig** map and apply the template. Generate side-by-side displays of both maps as shown below.



IDW and Krig interpolated maps ...can you see a difference?

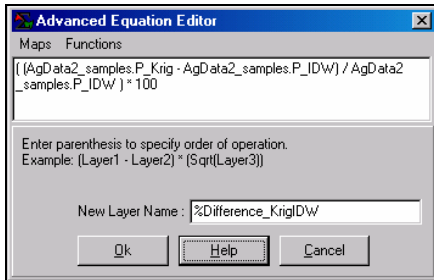
Note that 44% of the field estimates differ by less than plus or minus 1ppm (98.2% differs by less than +/-10ppm)—the predicted phosphorous surfaces for this data using default parameters for IDW and Krig exhibit minimal differences.

On your own, calculate percent difference maps comparing the average phosphorous level for the field (20.9ppm) with the .P_IDW and .P_Krig interpolated maps. What is the range of percent differences for both calculations? What is the average percent difference? What is the standard deviation of percent differences?

Press the **Map Analysis** button, select **Overlay** → **Calculate** and complete the following dialog box that generates a percent difference map between the *IDW* and *Krig* interpolated surfaces.

Hint: use the percent difference equation—

$$\%Diff = ((estimated - actual) / actual) * 100$$
 ...substituting the average value for the *actual* term and the interpolated map for the *estimated* term.

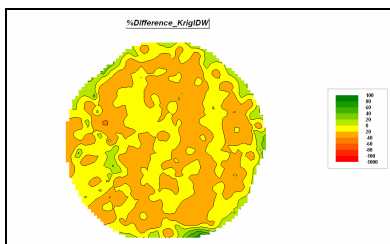
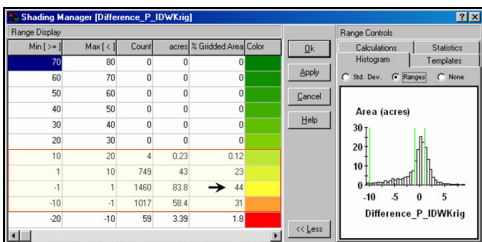


11.6.4 Evaluating Interpolation Performance

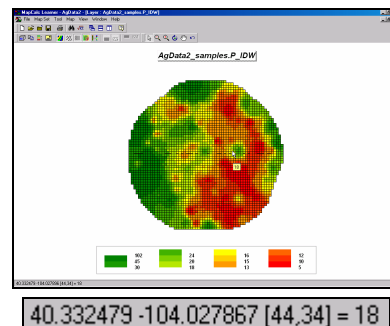
The phosphorous levels for ten randomly selected “test” locations were measured as follows:

Column	Row	Value
44	34	20.1
24	27	13.5
17	47	19.2
28	61	50.8
43	6	9.0
5	27	63.1
56	46	17.8
33	31	18.3
41	58	17.6
5	38	49.9

Click on the **Shading Manager** and use the *Statistics* and *Histogram* tabs to get an idea of the *numeric distribution* of the map data— Min, Max, Mean, Std. Dev., etc.— of the map you just created. Enter the settings for display as shown below to get an idea of the *spatial distribution* of the data.



Display the *AgData2_samples.P_IDW* map in *2D continuous grid* format with the *Layer Mesh* turned on.



As you move the cursor about the map notice that the Lat/Lon and Column/Row position is reported in the lower-left corner of the display. Move the cursor until you are over the first test sample location at *Column=44, Row= 34* and note the *.P_IDW estimated value= 18.0*. On a sheet of paper or an Excel spreadsheet enter these values in the *.P_IDW* column as shown in the table below.

	A	B	C	D	E	F	G	H	I
1	Column	Row	Value	.P_IDW	IDW_diff	.P_Krig	Krig_diff	Average	Avg_diff
2	44	34	20.1	18.0	-2.1	17.7	-2.4	27.9	7.8
3	24	27	13.5					27.9	14.4
4	17	47	19.2					27.9	8.7
5	26	61	50.8					27.9	-22.9
6	43	6	9.0					27.9	18.9
7	5	27	63.0					27.9	-35.1
8	56	46	17.8					27.9	10.1
9	39	31	18.3					27.9	9.6
10	41	68	17.6					27.9	10.3
11	5	38	49.9					27.9	-22.0
12		Average=	27.9						0.0
13		StdDev=							19.0

Once you have entered the other nine estimated values for the “test” location calculate the difference between the estimated value and the actual value (*IDW_diff= 18.0 – 20.1= -2.1*). Complete the calculated differences for column **IDW_diff** then compute the column average and standard deviation. Complete the same processing for the **.P_Krig** and **Krig_diff** columns in the table.

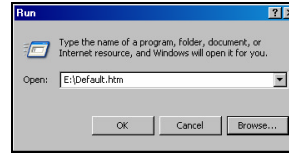
The best technique (whole-field *Average, IDW* or *Krig*) is the one that has the smallest **Average difference** and **Standard Deviation**. So which technique is best? How much better is it than the other two?

You can exit the program by selecting **File à Exit** or by clicking on the “X” in the upper-right corner of the MapCalc program window. If you want to save your work, specify a new file name, such as *Topic11_exercises.rgs*.

11.6.5 Using Surfer for Surface Modeling

Surfer is an advanced system for contouring, gridding, and surface mapping. It contains several additional spatial interpolation procedures beyond IDW and Kriging introduced in these exercises. You are encouraged to complete the Surfer Tutorials.

A printer formatted version is available on the CD accompanying this book. Insert the CD and access the main menu by pressing **Start à Run à E:\Default.htm**.



<enter your CD drive>

From the “Installing Surfer...” section, click on the link for the tutorial exercises. Print a copy of the exercises.

Analyzing Natural Resources Data

A Hands-on Case Study in Spatial Analysis and Data Mining

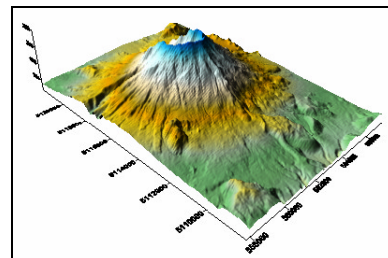
Note: Materials on this CD and companion hardcopy workbook are not for further distribution without prior written consent of Joseph K. Berry (jberry@newsunivcc.edu/Berry)

Item	Description
 	<p>This item contains installation guidelines for installing—</p> <p><i>MapCalc Learner</i> evaluation software for grid-based map analysis used in exercises throughout the book.</p> <p><i>Surfer</i> demonstration software for contouring, gridding, and surface mapping primarily used in the exercises accompanying Topic 11, “Surface Modeling.”</p> <p><i>Snatch</i> evaluation software for screen capture and creation of composite presentation graphics in exercises throughout the book.</p> <p>See Topic 1.5.3, “Installing Companion Software” for detailed instructions.</p>
<p style="font-size: x-small; color: #000080;">Example Applications</p>	<p>This item accesses several annotated descriptions and example applications of grid-based map analysis. See Topic 1.3.1, Example of Analysis Capabilities for more information.</p> <p>Click on any of the links in the listing to access the examples. All of the example applications were developed using MapCalc Learner and can be replicated using the tutorial database identified with each example.</p>
<p style="font-size: x-small; color: #000080;">Text Figures</p>	<p>This item contains a full color PowerPoint slide set of all of the workbook figures. Permission to copy is granted. Reference as “Figure x-xx from Analyzing Precision Ag Data, J.K. Berry, 2002.”</p>

Install the Surfer system if you have not already—see Topic 1.5.3. To access the system click on **Start à Programs à Golden Software Surfer 8 à Surfer 8**.

From Surfer’s main menu, select **Map à Surface** and specify **..\Samples\Helens2.grd** as the grid layer to open. Press the **Open** button to plot the data.

If Surfer has been properly installed a three-dimensional plot of Mt. St. Helens will appear (see below).



You are now ready to complete the Surfer Tutorials.

