VISUALIZATION TOOLS TO AID IN THE UNDERSTANDING OF GEOSTATISTICS

William V. Harper
Otterbein College, United States
Isobel Clark
Geostokos (Ecosse) Limited, United Kingdom
wharper@otterbein.edu

Geostatistics is sometimes a difficult leap for even those individuals well versed in classical statistics. The impact of data location in spatial statistics may be only vaguely understood initially. Visualization tools that allow the student or practitioner to see the impact of moving data, adding additional data, deleting data, adding fault lines, changing search radiuses, and so forth aid the learning of geostatistical concepts. Due to page limitations only a few items are briefly illustrated. This visualization software called the Kriging Game is available free at http://geoecosse.bizland.com/softwares/ . *This site also has other free geostatistical software and tutorials.*

INVERSE DISTANCE WEIGHTS VERSUS KRIGING WEIGHTS

Inverse distance weighting is a common interpolation technique and often the default in GIS packages. It is simple to program, easy to understand, and useful for obtaining an initial feel for one's data. However, it may provide poor estimates. It ignores the spatial structure of the actual data as well as not using the full information available concerning the relative locations of the actual data used to estimate the true unknown value *T* at a point where no sample exists.

The geostatistical technique known as kriging if used properly results in better estimates (optimal in a minimum variance sense if the modelling is correct) based on the development of spatial weights that reflect the spatial variability seen in the actual data of interest. Modelling the spatial variability is usually done by the development of a semi-variogram model reflecting the distance versus variability found in the actual data and also considering if the structure is isotropic (variability the same in all directions) or anisotropic (variability versus distance relationship is a function of direction). This modelling is outside the context of this paper but Clark and Harper (2000) as well as tutorials on the above mentioned web page will help.

The simple topic examined here is how the placement of data impacts inverse distance estimation versus a kriging estimation in terms of the weights assigned to the values selected to interpolate at a given spatial location. Take the simple example illustrated in Figure 1 below. The small dataset in Figure 1 has six data values though only four of these will be used to estimate an unknown value at the location marked with a *T* based on the search radius selected. *T* is often used in geostatistics to represent the true unknown value at a given spatial location. *T** is the estimated value. Obtaining *T** depends on many modelling choices inter alia the spatial dependence model assumed, the search radius used to decide which actual values to include in the estimation, and whether a global trend is factored in.

Figure 1 shows four data values labelled 1, 2, 3, and 4 that are to be interpolated in this simple two dimensional case in which no global trend is modelled and the spatial dependence is assumed to be isotropic. $T^* = \sum w_i g_i$ where the w_i depend on the method used and the g_i represent the values at the observed spatial locations. For this simple system $\sum w_i = 1.0$. The point at (1, 1) seen with a *T* in Figure 1 is equidistant from all four data values. Each data value has corresponding *X* and *Y* locations and a sample value as given in Table 1 below. Inverse Distance estimation computes the distance d_i from the desired estimation location to each sample point used in the interpolation. These are inverted to give the inverse distance and normalized so that the sum is 1.0 as seen in the column "Inverse Distance weights" in Table 1. Kriging uses the semi-variogram values that are functions of the d_i to establish a kriging system of equations that are solved to find the optimal kriging weights. For all the kriging weights shown in this paper a linear semi-variogram was chosen with a nugget of 0.0 and a slope of 1.0 just for simplicity. While in tables other than Table 1 the kriging weights might vary for other semi-variogram models, the basic concepts to be illustrated would be the same. Note that both the inverse distance and kriging approaches result in the same weights in this initial example and thus each data value receives 0.25 weight. The resulting T^* = 5.725 by both methods.

The Kriging Game software illustrated in Figure 1 allows students to quickly assess the kriging weights of each data value used in the estimation as well as many other features not discussed in this article. For example students may easily move a data point and see how both its weight changes as well as the estimated T^* and its standard error. Points may be removed by the student and the resulting impact on the weights of the other points and the estimation seen. Perhaps one of the most interesting items for geologic data is the impact of faults that show discontinuities in the geologic medium. Faults may be easily drawn in and the impact illustrated directly on the screen.

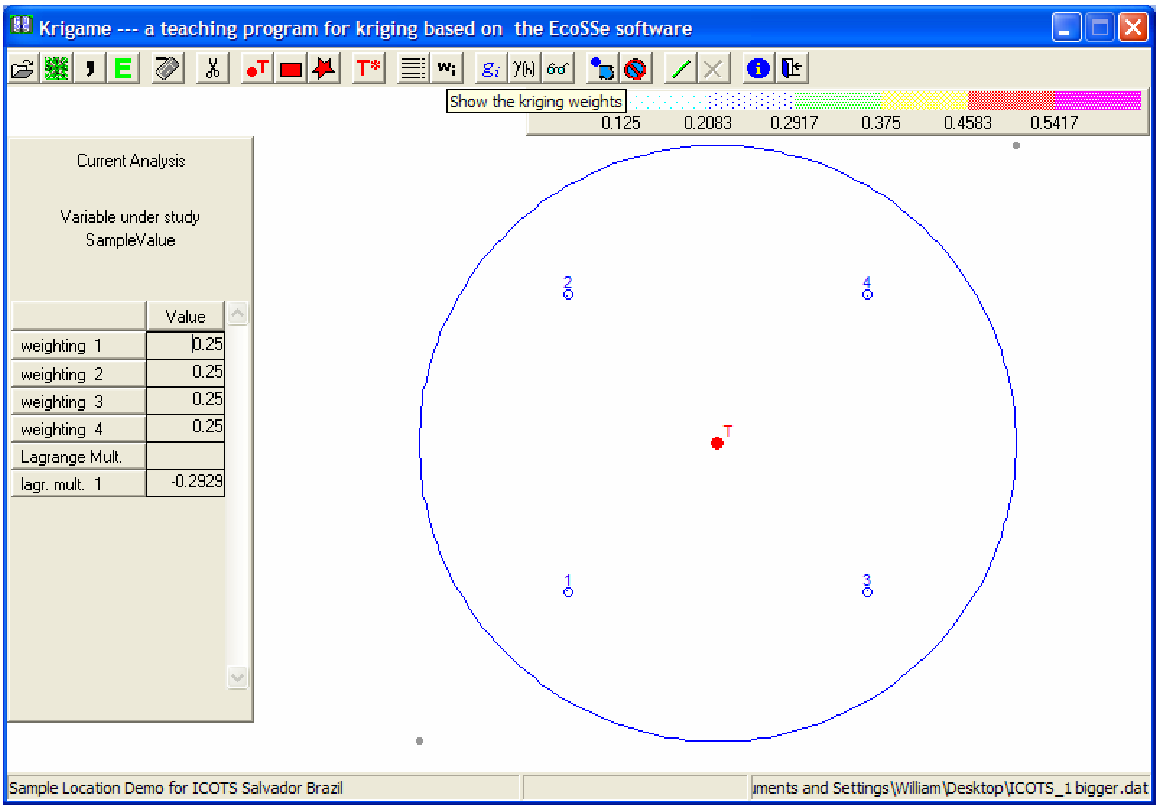

Figure 1: First data set with four equidistant points used to estimate T

Table 1: Weights for data in Figure 1

Point #	X	Y	Sample Value, g_i	d_i	$1/d_i$	Inverse Distance weights	Kriging weights	Inverse distance estimate of T^*	Kriging estimate of T^*
1	0	0	3.2	1.4142	0.7071	0.25	0.25	0.8000	0.8000
2	0	2	5.7	1.4142	0.7071	0.25	0.25	1.4250	1.4250
3	2	0	6.2	1.4142	0.7071	0.25	0.25	1.5500	1.5500
4	2	2	7.8	1.4142	0.7071	0.25	0.25	1.9500	1.9500
					2.8284	1.00	1.00	5.7250	5.7250

Figure 2 adds a fifth data point such that it is behind sample point 4 from the perspective of the unknown sample location at T. Examine the weights in Table 2 now assigned by both inverse distance and kriging to the 5 values now used in the estimation of T^*. Since the weights are different the corresponding estimates will vary. While the kriging game will show the kriging T^* in one of its screens, the tables included here show how T^* varies between the inverse distance and kriging approaches. The kriging weights do provide the optimal weights in terms of minimizing the standard error of T^* for the assumed spatial variability model contained in the semi-variogram.

The intriguing part of this simple analysis to learning geostatisticians is often twofold. First and perhaps most important concerns the weights assigned to the five points now being used to estimate T by the two methods. Inverse distance weights are solely a function of the distance between the observed point and the location of the point to be estimated. The fifth point added in the second sample seen in Figure 2 has a weight of 0.1818 that is almost equal to the weight of 0.2045 for the nearby fourth point. Thus the northeast corner of the data is accounting for approximately 39% of the weight of the full data set for the inverse distance method. Conversely all 4 corners for the kriging approach allocate approximately 25% to each corner. Why is this? First let's briefly mention the 2nd item newcomers find surprising. Unlike inverse distance weighting, kriging weights may be negative. The concept of a negative weight is hard to swallow for some. But the visualization tools provide an entry door to begin such discussion.

Addressing the unanswered question in the prior paragraph, spatial estimation of any kind has the underlying assumption that points close in space are likely to have similar values. Thus the two points close by (sample points 4 and 5 in Figure 2) are assumed to have similar values and thus sample point 5 is expected to provide little new information over sample point 4. Thus each of the 4 corners in this particular application are fairly equally weighted by kriging. The inverse distance method does not examine how close the observed values are in space to each other and thus loses the opportunity to compensate in the estimation. The combination of the visualization and ability to add, move, and delete data allows quicker understanding and a good stepping off point for subsequent learning that could not be so easily addressed otherwise.

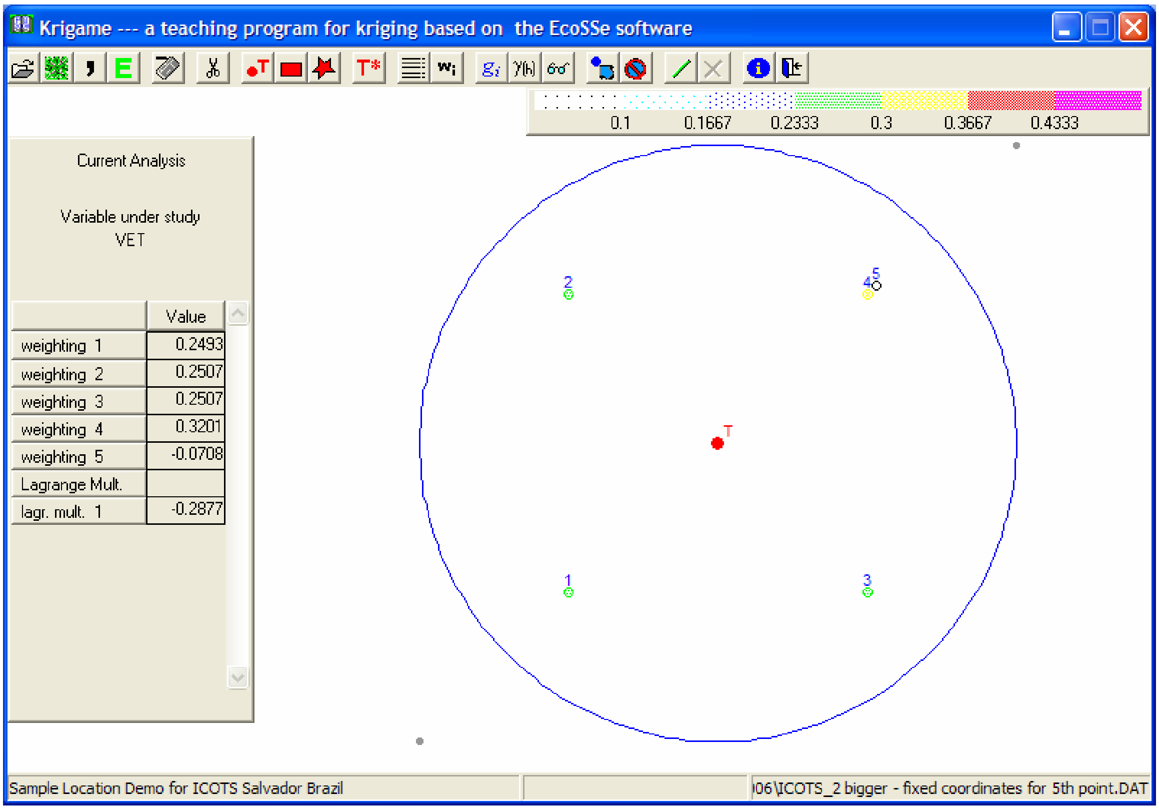

Figure 2: Second data set with fifth point added very close to another point

Table 2: Weights for Figure 2

Point #	X	Y	Sample Value, g_i	d_i	$1/d_i$	Inverse Distance weights	Kriging weights	Inverse distance estimate of T^*	Kriging estimate of T^*
1	0	0	3.2	1.41421	0.70711	0.2023	0.2493	0.6474	0.7978
2	0	2	5.7	1.41421	0.70711	0.2023	0.2507	1.1532	1.4290
3	2	0	6.2	1.41421	0.70711	0.2023	0.2507	1.2543	1.5543
4	2	2	7.8	1.41421	0.70711	0.2023	0.3201	1.5781	2.4968
5	2.06066	2.06066	7	1.5	0.66667	0.1907	-0.0708	1.3352	-0.4956
					3.49509	1.00	1.00	5.9682	5.7823

Figure 3 and the corresponding Table 3 provide another simple but interesting example. The fifth data point has been moved to a location directly to the right of T but at the same distance away that it was in Figure 2. What happens to our weights under the two procedures? Absolutely no changes occur with inverse distance weighting for any of the points; however, a dramatic difference is seen in the kriging weighs – especially for the fifth point that has gone from a negative weight of -0.0708 to a positive weight of 0.0916! Note that these kriging weights are computed automatically and displayed by the free kriging game software.

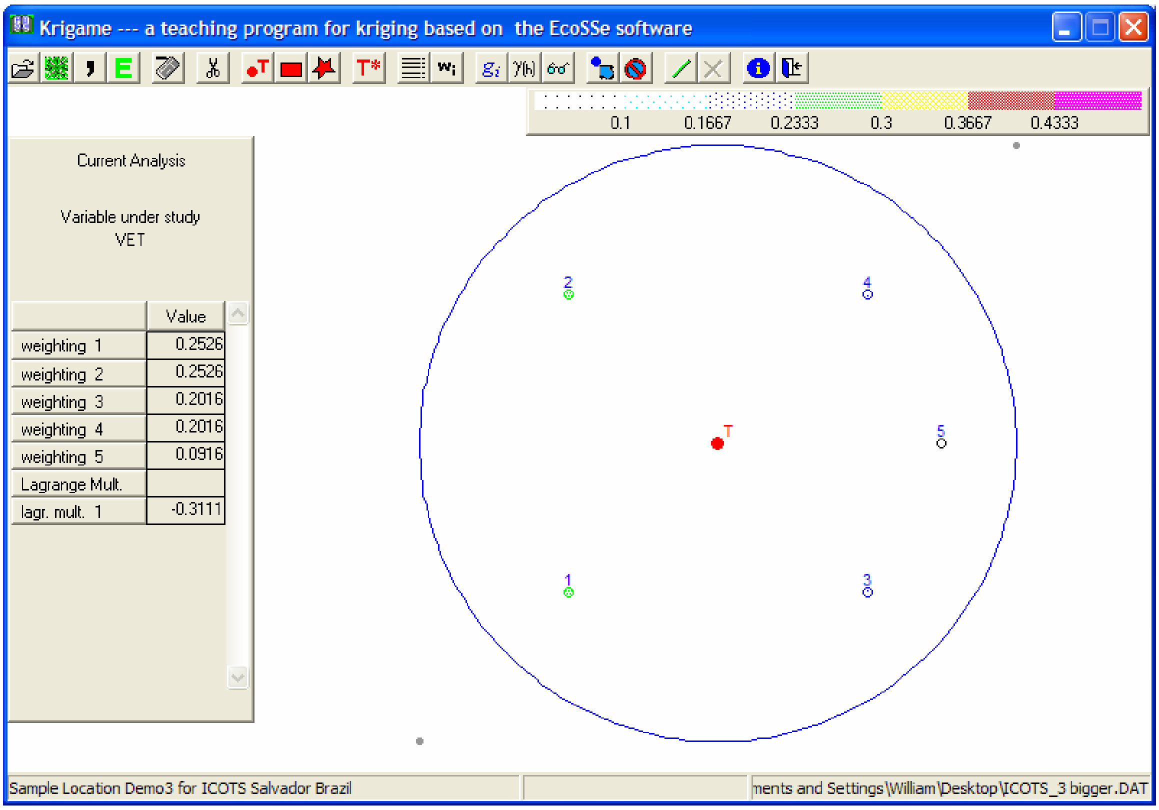

Figure 3: Third data set with fifth point moved to another location exactly as far from T as in Figure 2

Table 3: Weights for Figure 3

Point #	X	Y	Sample Value, g_i	d_i	$1/d_i$	Inverse Distance weights	Kriging weights	Inverse distance estimate of T^*	Kriging estimate of T^*
1	0	0	3.2	1.41421	0.70711	0.2023	0.2526	0.6474	0.8083
2	0	2	5.7	1.41421	0.70711	0.2023	0.2526	1.1532	1.4398
3	2	0	6.2	1.41421	0.70711	0.2023	0.2016	1.2543	1.2499
4	2	2	7.8	1.41421	0.70711	0.2023	0.2016	1.5781	1.5725
5	2.5	1	7	1.5	0.66667	0.1907	0.0916	1.3352	0.6412
					3.49509	1.00	1.00	5.9682	5.7117

The last example for comparing inverse distance weighting to kriging weights moves our beloved mobile fifth point to its final resting place in this too short paper. This fifth point is moved closer to T than any of the five observations to be used in the estimation procedure but directly in a northeastern direction in line with the fourth observation. What happens in this case to our weights? Both give the largest weight now to the newly moved fifth point that is the closest to T. The inverse distance method however equally weights the four original corner observations while the kriging method does not. Note in particular the kriging weight assigned to the fourth observation.

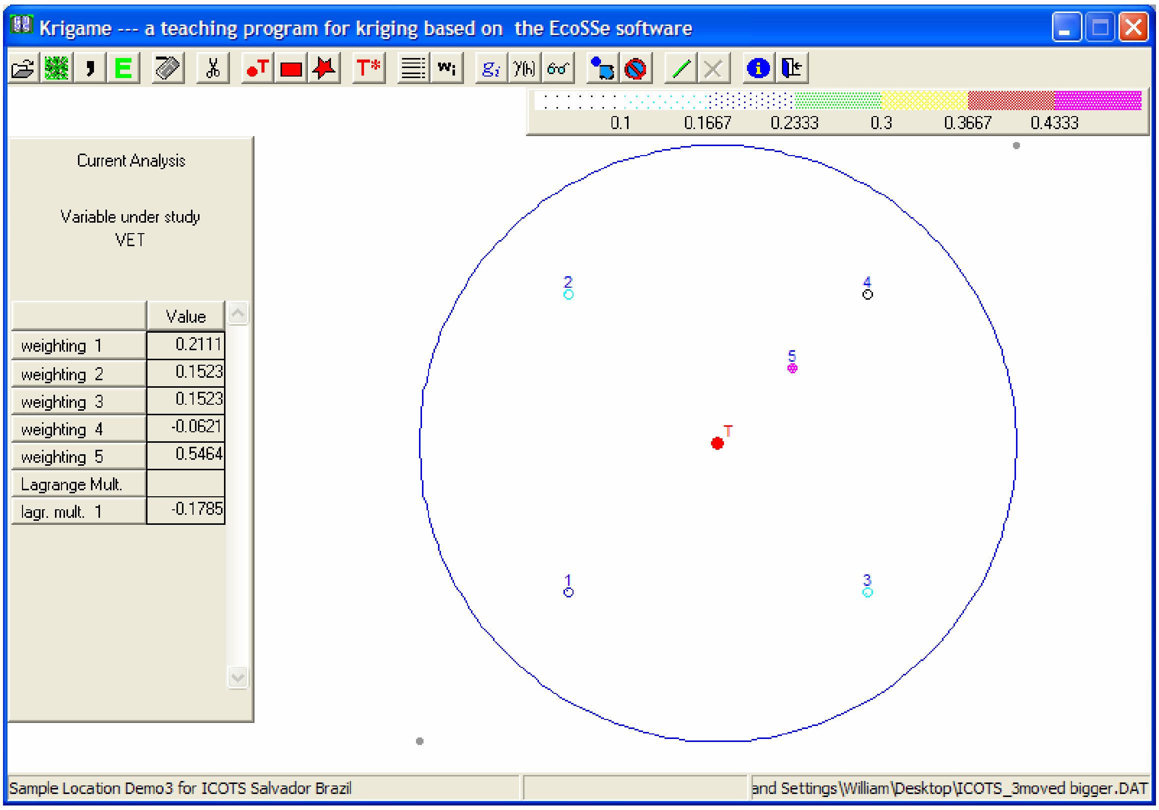

Figure 4: Fifth data point moved closer to T

Table 4: Weights for Figure 4

Point #	X	Y	Sample Value, g_i	d_i	$1/d_i$	Inverse Distance weights	Kriging weights	Inverse distance estimate of T^*	Kriging estimate of T^*
1	0	0	3.2	1.41421	0.70711	0.1667	0.2111	0.5333	0.6755
2	0	2	5.7	1.41421	0.70711	0.1667	0.1523	0.9500	0.8681
3	2	0	6.2	1.41421	0.70711	0.1667	0.1523	1.0333	0.9443
4	2	2	7.8	1.41421	0.70711	0.1667	-0.0621	1.3000	-0.4844
5	1.5	1.5	7	0.70711	1.41421	0.3333	0.5464	2.3333	3.8248
					4.24264	1.0000	1.00	6.1500	5.8283

This relatively straightforward comparison of inverse distance weights versus kriging weights can be done without the visual tools in the kriging game, but students typically learn more by exploring on their own and visually seeing what happens as points are added, deleted, or moved. All of this may be done interactively with this software. It is a shame that space limits the coverage of this game as it has the ability to illustrate many complex operations such as the impact of adding fault lights, the often mysterious Lagrange multipliers and the perhaps scary kriging equations directly. It is recommended that individuals download the kriging game and use it personally or for their classes to take advantage of the visualization options to enhance the understanding of geostatistical techniques.

REFERENCES

Clark, I. and Harper, W. V. (2000). *Practical Geostatistics 2000*, Columbus, OH: Ecosse North America, LLC.