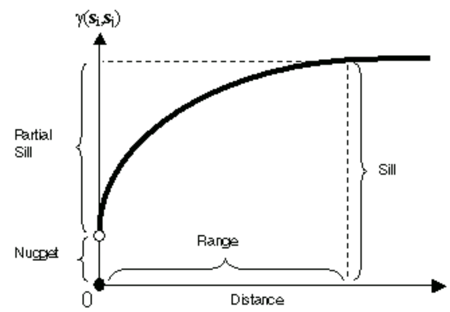
Kriging Exercise

Introduction

In this exercise we will do an exercise on the kriging interpolation technique for spatial data. Kriging is the geostatistical method to estimate the values of unknown points from surrounding sample points. Similar to other weighted average approaches, kriging estimate unknown values by using $Z(x) = \sum_{j=0}^{n} \lambda_j Z(x_j)$. The weights λ_j are assigned in a way that minimizes the variance of the estimation residual: $= K^{-1}k$, where **K** is the **covariance matrix** between the sample points, **k** is the **covariance** vector between the point to be estimated and the sample points. This solution is under the assumption that the values have a constant mean and the mean is known. Under this assumption, the estimation is unbiased (the residual has a zero mean). This is called simple kriging. However, in many situations the simple kriging assumption cannot hold. So an extension to the simple kriging called ordinary kriging assumes that the mean is unknown. This simply adds a constraint $\sum \lambda_j = 1$. The solution is by adding a Lagrange Multiplier.

The covariance matrix K and vector k are estimated from the variance equation fitted to the semi-variogram. The semi-variogram is a plot of the covariances between observation points by the semi-variance equation: $\gamma(s_i, s_j) = \frac{1}{2} var[Z(s_i) - Z(s_j)]$. It is called semi-variance because it takes one half of the variance value. If the two values $Z(s_i)$ and $Z(s_j)$ are similar, they have high spatial autocorrelation. It is assumed that when the distance lag (h) goes high, the spatial autocorrelation goes down, and the semi-variogram will approach its maximum value (sill). An ideal semi-variogram looks like the figure below.



A variance function can be defined by fitting the semi-variogram point cloud. Typically used functions are exponential, spherical, Gaussian, etc. Once the function is defined, the covariance matrix K and covariance vector k can be obtained by measuring the distances between the points. The following exercise will guide you through a manual calculation of the kriging equations and an interpolation exercise using ArcGIS 10.

Tutorial and exercise questions

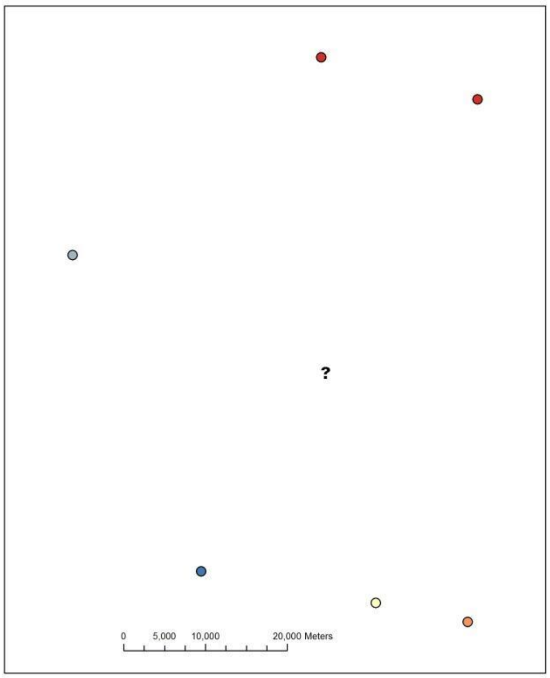
Locating the data

Unzip the data downloaded from Moodle. The data are provided in a file geodatabase called ca_ozone.gdb, the data with the name of O3_Sep06_3pm is the ozone measurement. I have also made a subset of the ozone data with only six points for you to do a manual calculation of kriging interpolation.

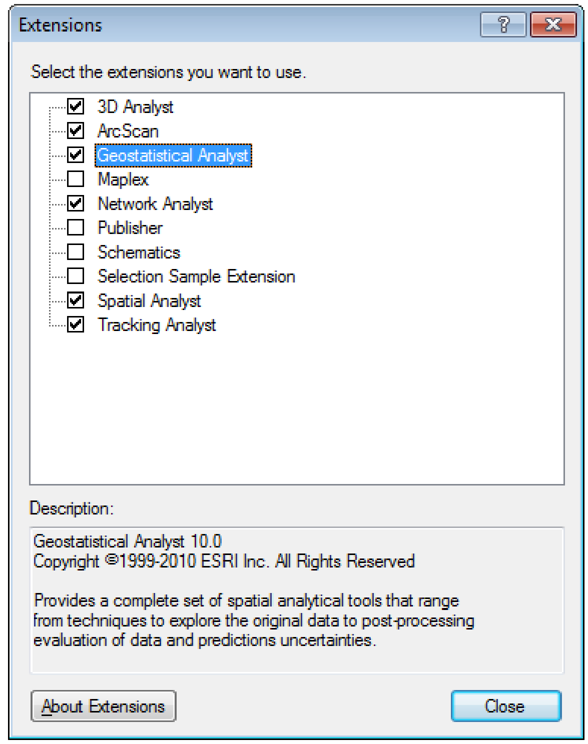
Simple kriging

The scenario of the simple kriging calculation is as the following figure:

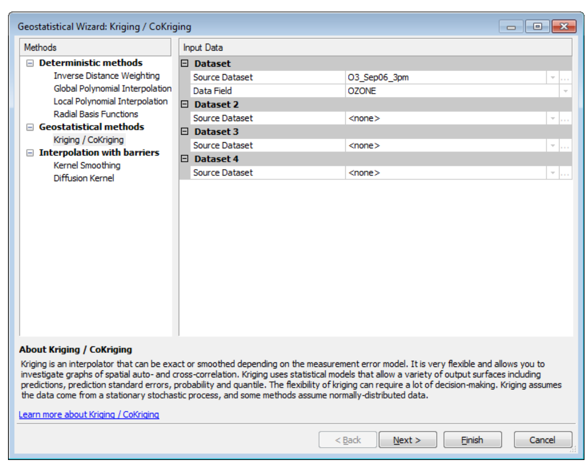
We will use the example data to do a manual calculation.



- → Add the point data "selected" and "interpolation" from the file geodatabase to ArcMap
- → The Excel file contains a distance matrix between six sample points and the distance vector between the point to be estimated and the six points
- → In ArcMap, activate the Geostatistical Analyst toolbar and its license in the Customize->Extension

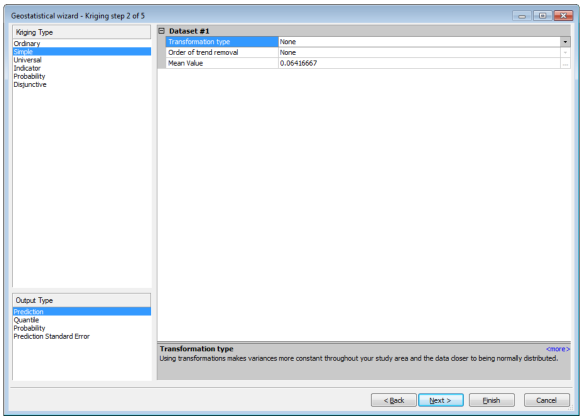


- → Run the Geostatistical Wizard from the toolbar
- → Put the layer "selected" as input data, ozone as the field. Click Next.

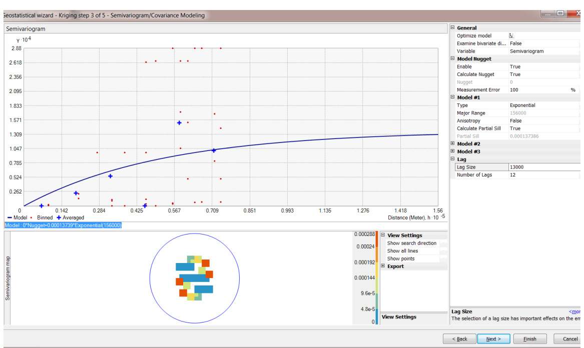


- → Select Transformation type as None

- ➔ Make a note of the reported Mean Value: **0.06417**
- ➔ Click Next



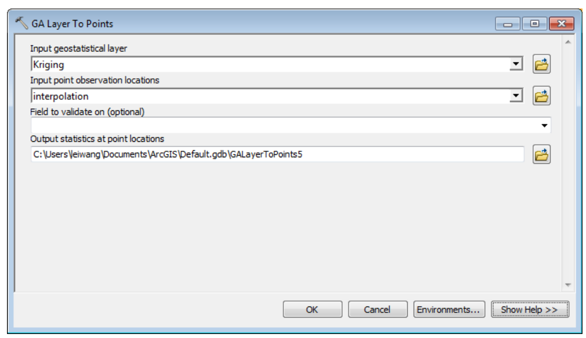
- ➔ Choose Variable as Semivariogram
- ➔ In Model Nugget section, choose "False" for "Calculate Nugget" and put 0 in "Nugget"
- ➔ Select Model #1 Type as Exponential
- ➔ In the lag size, put 13000



- ➔ Make a note of the Model equation:
- ➔ Model : 0*Nugget+0.000013739*(1-Exponential(156000))
- ➔ Write down the values of the three parameters:
- ➔ Nugget (C_0) = ________
- ➔ Partial Sill (C_h) = ________
- ➔ Range (a) = ________

- ➔ The actual semivariogram function is $r(h) = C_0 + C_h(1\text{-Exp}(-3h/a))$ when $h < a$
- ➔ $r(h) = C_h + C_0$ when $h > a$

- ➔ Click Next and Finish to complete the wizard
- ➔ Right click the Kriging layer and Validation/Prediction
- ➔ Use the layer interpolation as the Input point data
- ➔ The output location and name does not matter here. Use the default one
- ➔ Check the predicted value by opening the attribute table of the GALayerToPoint and put the number down
- ➔ The predicted value from the ArcGIS software is: _________
- ➔



- ➔ Go to the Excel file of Kriging.xlsx, use a section of 6 by 6 cells and put the equation "=0.000013739*3*EXP(-3*B2/156000)" in the first cell
- ➔ Question: why do we use such an equation to calculate the covariance values? (hint: Cov = $C_0 + C_h - r(h)$)
- ➔ Select six cells including the current cell to the left and press CTRL-R to autofill the values

→ Select six lines including the current line at the top and press CTRL-D to autofill the matrix. Now you have got the covariance matrix K.

0.000013739	3.97E-06	3.48E-06	3.78E-06	6.51E-06	9.38E-06
3.96747E-06	1.37E-05	7.26E-06	9.05E-06	6.15E-06	3.79E-06
3.48254E-06	7.26E-06	1.37E-05	1.1E-05	3.86E-06	4.02E-06
3.78373E-06	9.05E-06	1.1E-05	1.37E-05	4.64E-06	4.11E-06
6.50539E-06	6.15E-06	3.86E-06	4.64E-06	1.37E-05	4.95E-06
9.38466E-06	3.79E-06	4.02E-06	4.11E-06	4.95E-06	1.37E-05

Now let's calculate k

→ Select the cell as displayed below and put the equation: "=0.000013739*EXP(-3*B9/156000)"

→ Select the six cells including the current one at the top and CTRL-D to autofill

38609	3.21803E-05
28683	3.42945E-05
35070	3.29187E-05
28771	3.42751E-05
34175	3.31081E-05
38248	3.2255E-05

Now you have calculated k. Next, calculate $\lambda = K^{-1} k$

→ Select a matrix of 6 by 6 cells as displayed below

0.000013739	3.97E-06	3.48E-06	3.78E-06	6.51E-06	9.38E-06
3.96747E-06	1.37E-05	7.26E-06	9.05E-06	6.15E-06	3.79E-06
3.48254E-06	7.26E-06	1.37E-05	1.1E-05	3.86E-06	4.02E-06
3.78373E-06	9.05E-06	1.1E-05	1.37E-05	4.64E-06	4.11E-06
6.50539E-06	6.15E-06	3.86E-06	4.64E-06	1.37E-05	4.95E-06
9.38466E-06	3.79E-06	4.02E-06	4.11E-06	4.95E-06	1.37E-05

=MINVERSE(I2:N7)					

→ In the equation box, put '=MINVERSE(' and use the cursor to select the K matrix above it as input. Add the closing parenthesis ')', and press Shift-Ctrl-Enter to complete the equation. You should be able to get the inverse matrix K^{-1}

153613.0415	-833.2549	420.65413	-1863.22	-39403.1	-90068.9
-833.2548807	143230.98	1375.1105	-82712.01	-35503.8	-1883.84
420.6541309	1375.1105	205394.24	-162156.3	729.9823	-12544.2
-1863.219887	-82712.01	-162156.3	260219.84	-2482.05	-5325.89
-39403.07643	-35503.8	729.9823	-2482.052	108657.5	-1895.79
-90068.89356	-1883.841	-12544.22	-5325.888	-1895.79	140772.3

→ To the right of the k vector, select six empty cells

38609	3.21803E-05	
28683	3.42945E-05	
35070	3.29187E-05	
28771	3.42751E-05	
34175	3.31081E-05	
38248	3.2255E-05	

→ Put the equation 'MMULT(', and use cursor to select the matrix K^{-1} as array1, and the k vector as array2. Press Shift-Ctrl-Enter to finish.

38609	6.53879E-06	=MMULT(I9:N14,C9:C14)
28683	7.91408E-06	
35070	6.99931E-06	
28771	7.90065E-06	
34175	7.12082E-06	
38248	6.58444E-06	

Now, we have the λ vector. Check the sum of the λ vector. Do the λ values sum up to 1?

Next, we need to get the residual vector of the observations by subtracting the mean: 0.06417

- ➔ In the column next to the Point Values, put the equation "=B16-0.06417" and use autofill to calculate all other values

it values		
1	0.07100000000	0.00683000000
2	0.05600000000	-0.00817000000
3	0.06100000000	-0.00317000000
4	0.06000000000	-0.00417000000
5	0.05700000000	-0.00717000000
6	0.08000000000	0.01583000000

Now we are ready to calculate the residual by

using the linear combination of $\sum \lambda(i)R(i)$

- ➔ Find an empty cell and put the multiple of the first two cells in the residual vector and the λ vector

1	38609	6.53879E-06	0.1124372
2	28683	7.91408E-06	0.219019331
3	35070	6.99931E-06	0.092712972
4	28771	7.90065E-06	0.201408316
5	34175	7.12082E-06	0.208118423
6	38248	6.58444E-06	0.179678199
values			
1	0.07100000000	0.00683000000	=C16*D9
2	0.05600000000	-0.00817000000	
3	0.06100000000	-0.00317000000	
4	0.06000000000	-0.00417000000	
5	0.05700000000	-0.00717000000	
6	0.08000000000	0.01583000000	

- ➔ Use autofill (Ctrl-D) to calculate all the cells in the vector

0.07100000000	0.00683000000	0.000767946
0.05600000000	-0.00817000000	-0.001789388
0.06100000000	-0.00317000000	-0.0002939
0.06000000000	-0.00417000000	-0.000839873
0.05700000000	-0.00717000000	-0.001492209
0.08000000000	0.01583000000	0.002844306

- ➔ According to the kriging prediction:

$$\sum_{i=1}^{N} \lambda_i \left[Z(s_i) - mean \right] + mean$$

The sum of the above cells plus the mean 0.06417 will be the kriging interpolation result

- ➔ In the cell below the 6 values above, use the "sum" formula to sum all the 6 values and add the result by 0.06417 to obtain the final interpolation result. Fill the result in the blank below.

0.07100000000	0.00683000000	0.000767946
0.05600000000	-0.00817000000	-0.001789388
0.06100000000	-0.00317000000	-0.0002939
0.06000000000	-0.00417000000	-0.000839873
0.05700000000	-0.00717000000	-0.001492209
0.08000000000	0.01583000000	0.002844306
		=sum(D16:D21)+0.06417

Question 1: The interpolated value from manual calculation is: __________. Compare it to the software-calculated value.

ORDINARY KRIGING

In this section, we will manually calculate an ordinary kriging, which assumes a constant but unknown mean. The solution is to add the Lagrange Multiplier in the calculation to force the sum of weights to equal to one. When this condition is met, the formula of kriging prediction

$$\sum_{i=1}^{N} \lambda_i [Z(s_i) - mean] + mean$$

is rewritten as:

$$= \sum_{i=1}^{N} \lambda_i Z(s_i) + mean \cdot \left(1 - \sum_{i=1}^{N} \lambda_i\right)$$

Under the constraint 1- sum(λ) = 0, the mean value will not affect the interpolation anymore. Therefore, using the ordinary kriging does not require the knowledge of the mean value of the field. This is more convenient for most applications because the actual mean is always not known.

- → Make a copy of the current sheet of the kriging worksheet. Name it as ordinary kriging
- → Pad the matrix with 1s to the extra row and column of the K matrix as shown below. The lower right corner is padded with 0.

0.000013739	3.967E-06	3.483E-06	3.784E-06	6.51E-06	9.38E-06	1
3.96747E-06	1.374E-05	7.259E-06	9.055E-06	6.15E-06	3.79E-06	1
3.48254E-06	7.259E-06	1.374E-05	1.101E-05	3.86E-06	4.02E-06	1
3.78373E-06	9.055E-06	1.101E-05	1.374E-05	4.64E-06	4.11E-06	1
6.50539E-06	6.152E-06	3.864E-06	4.642E-06	1.37E-05	4.95E-06	1
9.38466E-06	3.795E-06	4.018E-06	4.106E-06	4.95E-06	1.37E-05	1
1	1	1	1	1	1	0

- → Add one "1" to the k vector too:

1	38609	6.53879E-06
2	28683	7.91408E-06
3	35070	6.99931E-06
4	28771	7.90065E-06
5	34175	7.12082E-06
6	38248	6.58444E-06
ues		1

- → Calculate K^{-1}. It is similar to before but this time the matrix has to be 7 by 7 including the newly added line and column

150283.6075	-4437.982	-4637.694	-2728.172	-43986.8	-94492.9	0.152271
-4437.982443	139328.19	-4101.486	-83648.48	-40466.6	-6673.66	0.164861
-4637.693553	-4101.486	197709.19	-163470.4	-6234.05	-19265.6	0.231342
-2728.17213	-83648.48	-163470.4	259995.14	-3672.87	-6475.2	0.039558
-43986.84201	-40466.58	-6234.048	-3672.867	102346.8	-7986.52	0.209637
-94492.91732	-6673.664	-19265.55	-6475.203	-7986.52	134893.9	0.202331
0.152270561	0.164861	0.2313419	0.0395583	0.209637	0.202331	-7E-06

- → Calculate the λ vector by using a 7-cell vector.

38609	6.53879E-06	0.110400667
28683	7.91408E-06	0.216814407
35070	6.99931E-06	0.089618903
28771	7.90065E-06	0.200879246
34175	7.12082E-06	0.205314645
38248	6.58444E-06	0.176972132
	1	9.31402E-08

Note that the sum of λ(j) should be exactly 1 (not including the last element, which is the lagrange multiplier)

Average: 0.166666667 Count: 6 Sum: 1

→ Calculate $Z = \sum \lambda(j)Z(j)$ with λ(j)

from the last step by calculating the multiplication of the weights and the point values.

nt values		
1	0.07100000000	0.0078384473
2	0.05600000000	0.0121416068
3	0.06100000000	0.0054667531
4	0.06000000000	0.0120527547
5	0.05700000000	0.0117029347
6	0.08000000000	0.0141577705
		=SUM(C16:C21

→

→ Take the sum of the values to get the interpolation result:__________

Question 2. What is the prediction value from the ordinary kriging? Is it significantly different from the simple kriging?

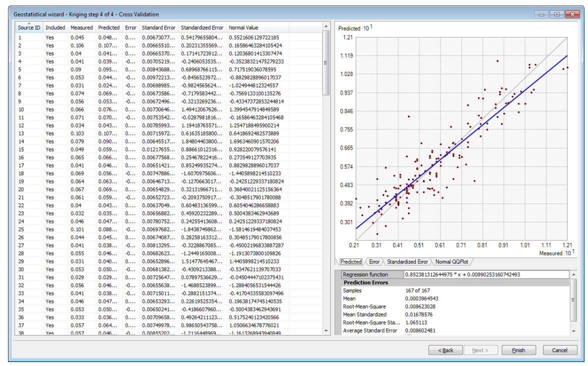
Compare different kriging models

→ Add the ozone data from the geodatabase to the ArcMap layers.

→ Run GeoStatistical Analyst Wizard

→ Select Ordinary Kriging and use the default parameters.

→ Pay attention to the Error analysis page



→ Check the prediction vs. observation scatterplot. The slope of the regression line should be close to 1.

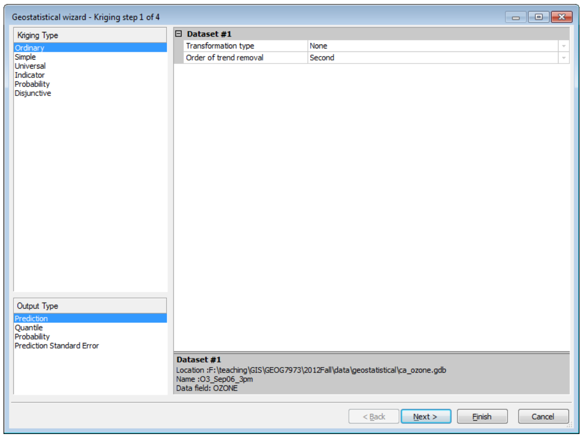
→ Also, check the standard error plot. The mean should be close to 0 and follow a normal distribution

→ Click next and finish the kriging model

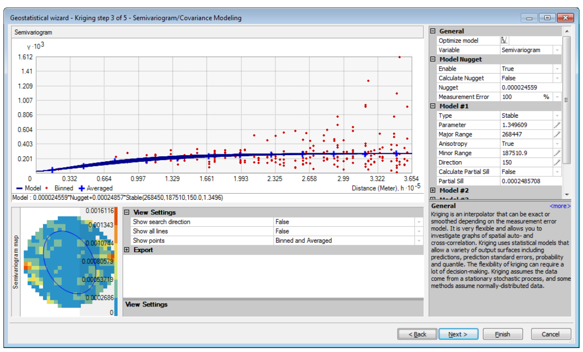
→ Now we will need to consider the trend and anisotropic neighborhood searching options

→ Run the Wizard again

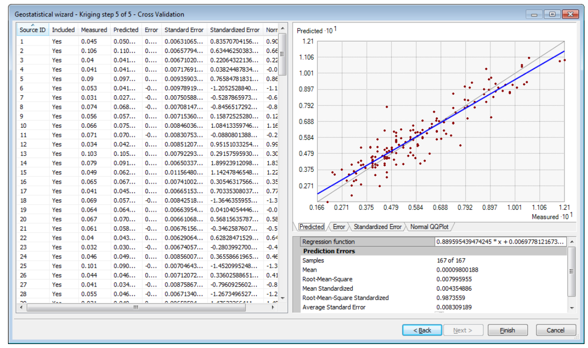
→ Use the ordinary kriging with a second order trend

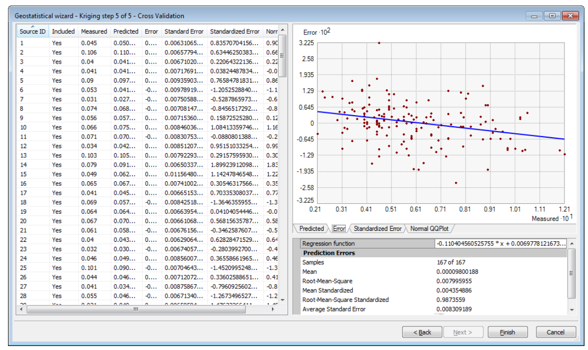


→ Click next and next

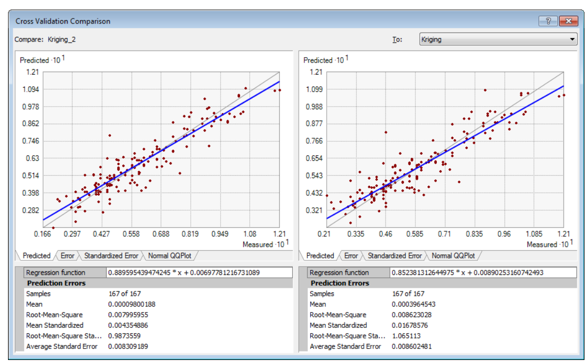


→ In Model # 1, change the Anisotropy parameter to True
→ Click the pencil icon of the Direction parameter in Model #1 and update the value to 150 degree
→ Check the quality of the model and the error distribution





→ Right click the Kriging_2 model and Compare…
→ Compare the two models



→ Put the kriging maps into your report

Question 3. By comparing the two models, which one do you think is better? Why?

Question 4. Why in the second model the directional search is defined as 150 degree from the east direction?

Question 5. Submit the report with two maps from kriging embedded and the two additional excel tables of your manual kriging calculation.