Aidan Brown

GEOG 523 Fall 2023

Question 1

- The dataset contains data of stream sediment and water geochemical data on Vancouver Island. The data was collected starting in 1976 and continues to 2010.

Question 2

- The spatial features are depicted as river polylines on the map.

Question 3

- The coordinate system is using NAD 1983

Question 4

- 688 total observations in the attribute table.

Most of the points are found on or around any rivers. Looking at the element table of the gathered data, consistencies occur with re-occurring elements, especially those with large found amounts of mercury. This seems like one factor these data points would be chosen.

Question 5

- Field workers go to sites to assess and collect sample materials, the samples are evaluated and prepared for analysis. The samples go through a control process where the parent sample is duplicated for accurate results and to compare if there are any inconsistencies. The data is archived and is updated on a regular basis.

Question 6

- Lat/Long, ppm, Date, PH, List of elements measured in the RGS_92F table:

ZINC
Copper
Lead
Nickel

Cobalt
Silver
Manganese
Iron
Molybdenum
Uranium
Tungsten
Tin
Mercury
Arsenic
Antimony
Barium
Cadmium
Vanadium
Bismuth
Chromium
Selenium

Question 7

- Unit of measurement of pH is ppm, the measurement precision is based on the type. It is to the first decimal place in this case.

Question 8

- In this case the pH variable is used a a *response* because the pH data presented in this lab is based on other data factors, which would be the elemental data associated with each collection point.

Question 9

- All points of the study area from the dataset are covered, except one large area, which doesn't contain any data because Strathcona park is a provincial park.
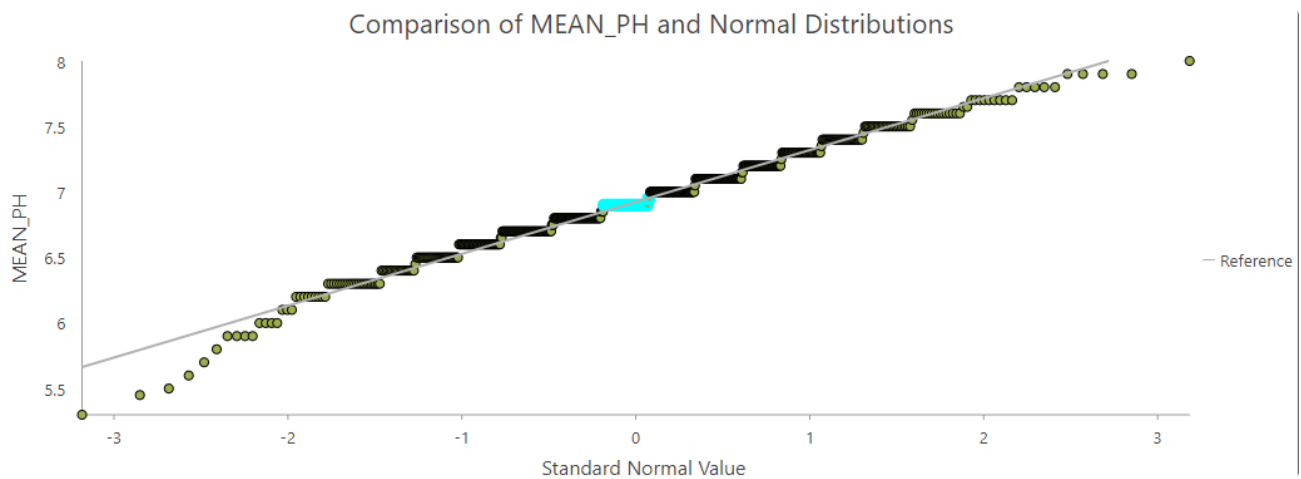
Question 10

- Going to layer properties and into Spatial Reference the coordinate system it lists WGS 1984 as the Geographic coordinate system.

Question 11

- Based on the histogram the model most fits a *bimodel* curve, as there are two distinct peaks of values, being 6.7 on the left side of the curve, and 7.2 on the right side of the curve.

Question 12
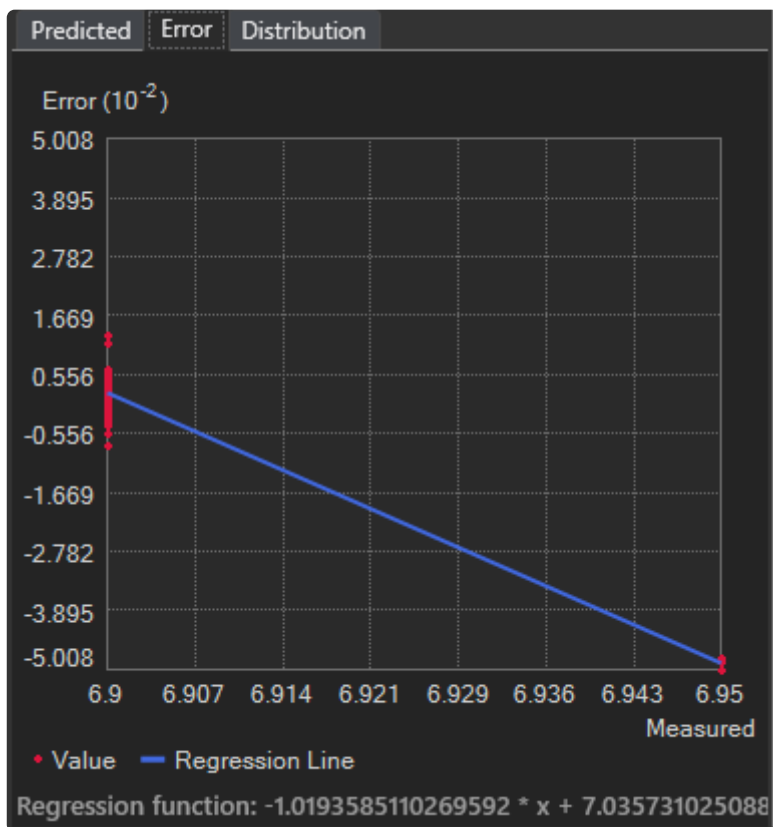

Comparison of MEAN_PH and Normal Distributions

- QQ Plot of Mean pH values
- Based on the QQ Plot above, we see that the mean values roughly follow the same line and are distributed evenly. From this assessment we can say that this data is normally distrubuted.
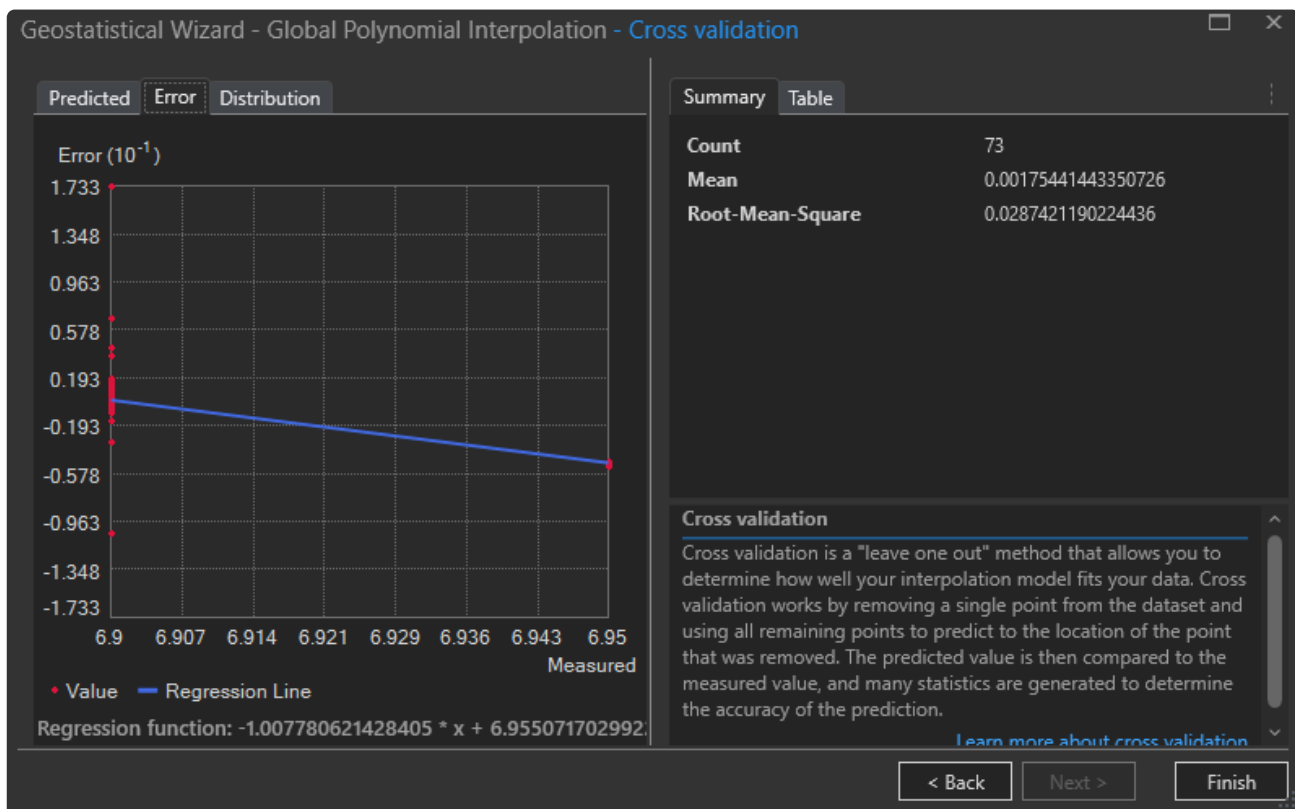
Question 13

- Adjusting the search neighbourhood, and specifying a minimum and maximum neighbours concentrated the statistical output of the distance weighing. These options along with optimizing the power value improved the overall RMSE value.
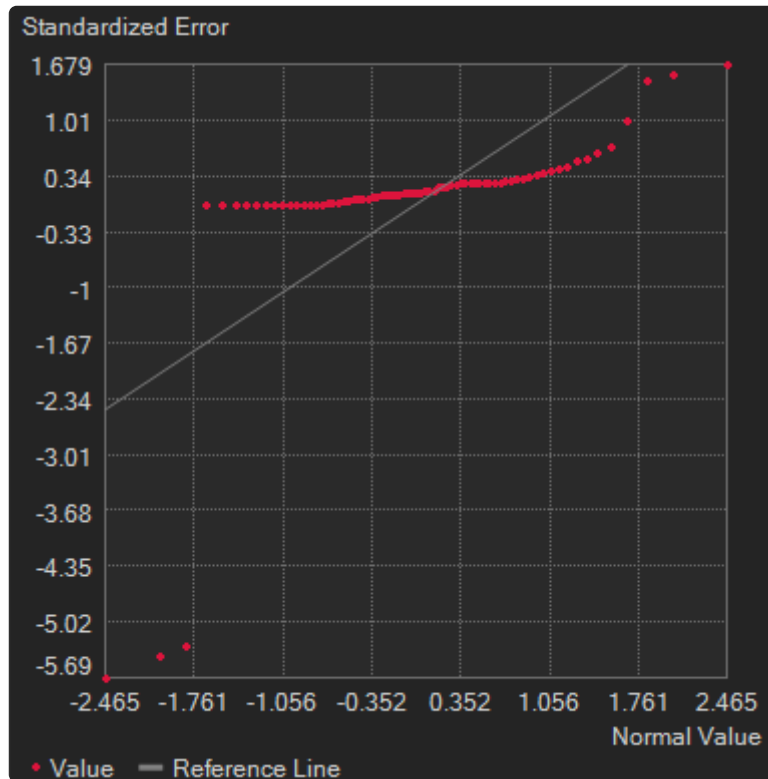
Question 14

- Order of Polynomial 2:

Predicted Error Distribution

Error (10^{-2})

5.008
3.895
2.782
1.669
0.556
-0.556
-1.669
-2.782
-3.895
-5.008

6.9   6.907   6.914   6.921   6.929   6.936   6.943   6.95
Measured

• Value — Regression Line

Regression function: -1.0193585110269592 * x + 7.035731025088

- Order of polynomial 6:

Geostatistical Wizard - Global Polynomial Interpolation - Cross validation

Predicted Error Distribution

Error (10^{-1})

1.733
1.348
0.963
0.578
0.193
-0.193
-0.578
-0.963
-1.348
-1.733

6.9   6.907   6.914   6.921   6.929   6.936   6.943   6.95
Measured

• Value — Regression Line

Regression function: -1.007780621428405 * x + 6.955071702992

Summary Table

Count	73
Mean	0.00175441443350726
Root-Mean-Square	0.0287421190224436

Cross validation

Cross validation is a "leave one out" method that allows you to determine how well your interpolation model fits your data. Cross validation works by removing a single point from the dataset and using all remaining points to predict to the location of the point that was removed. The predicted value is then compared to the measured value, and many statistics are generated to determine the accuracy of the prediction.

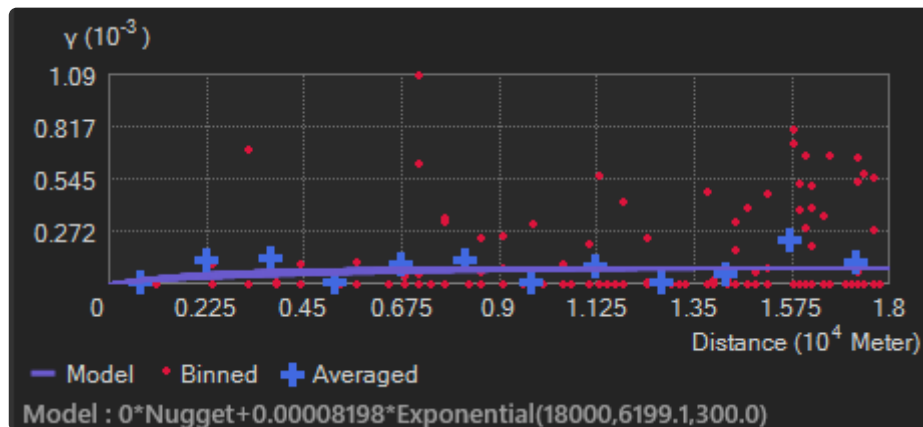Learn more about cross validation

< Back   Next >   Finish

- Based on the two option the error trend using a 2 polynomial value the trend is skewed downwards a lot more than the 6 value.

Question 15

- The neighbouring pH values are more likely to be similar than the distant values are, as the semivariogram.

Model : 0*Nugget+0.00008198*Exponential(18000,6199.1,300.0)

Question 16

Based on the results of the semivariogram pH is anisotropic. This is because The distance between points are greater than they are together. This is the definition of an anisotropic spatial dependency.

Question 17

- Universal Kriging:
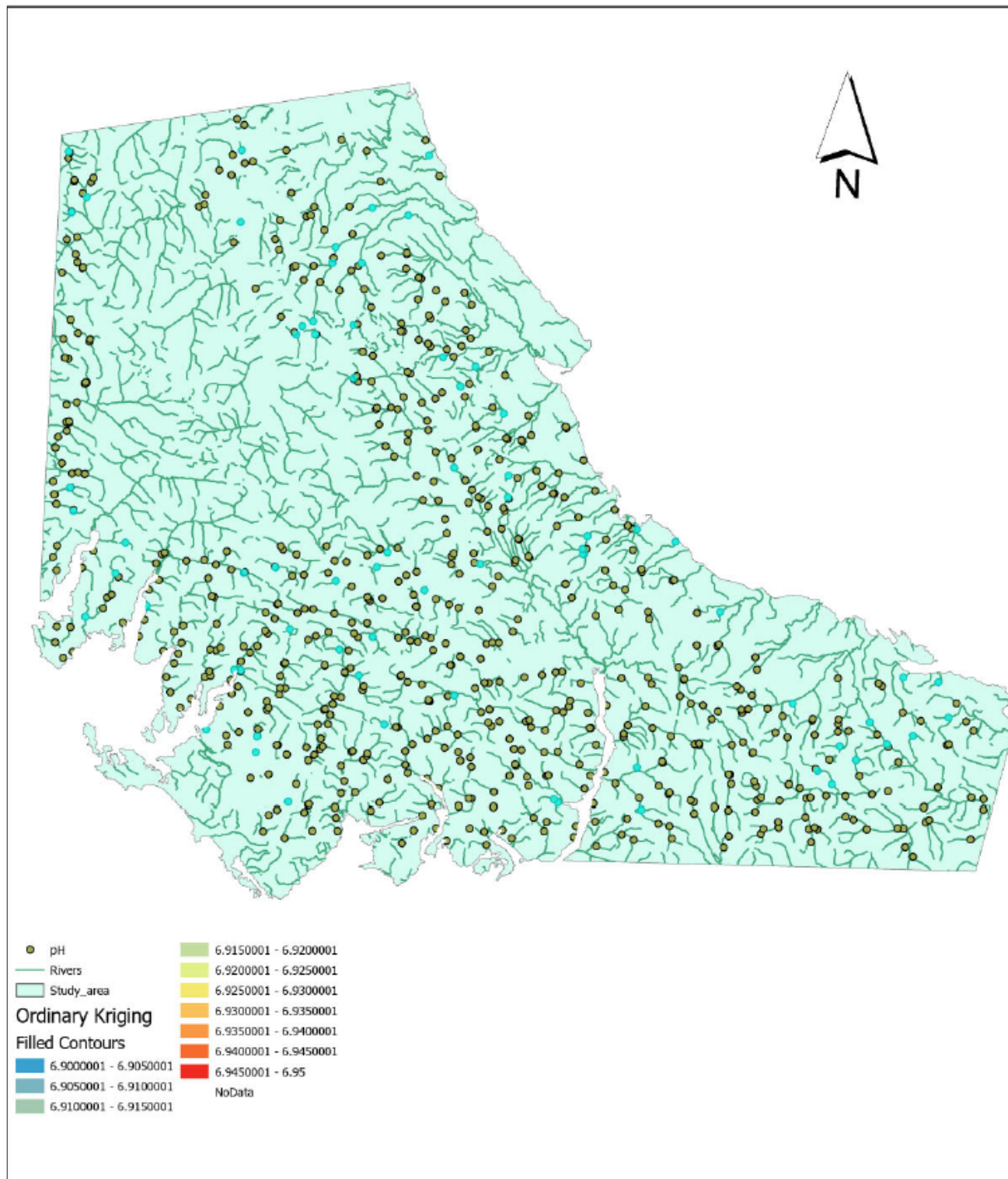
- Universal kriging with anisotropy:

Value changes slightly, from 1.18 to 1.21 using universal kriging.

Question 19

Inverse Distance Weighing ranked the highest as the isotropic model, best approximating the source points. Using Inverse Distance weighing, Global polynomial interpolation, and IDW Ellipse Neighbourhood models.

Question 20

Central Vancouver Island pH Data Analyzed using Ordinary Kriging Geo-statistics

Legend:

- pH
- Rivers
- Study_area

Ordinary Kriging

Filled Contours

- 6.9000001 - 6.9050001
- 6.9050001 - 6.9100001
- 6.9100001 - 6.9150001
- 6.9150001 - 6.9200001
- 6.9200001 - 6.9250001
- 6.9250001 - 6.9300001
- 6.9300001 - 6.9350001
- 6.9350001 - 6.9400001
- 6.9400001 - 6.9450001
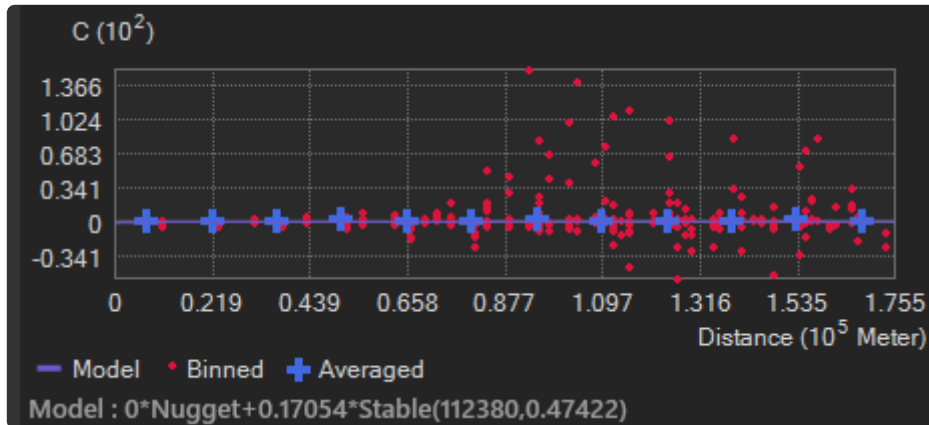- 6.9450001 - 6.95
- NoData

Map Produced By: Aidan Brown
Data source: BC Ministry of energy,
mines and petroleum

Based on running the Geostatistical layer geoprocessing, Ordinary Kriging ranked the first as the best anisotropic method, with the input parameters of the ordinary kriging model, and universal kriging models.

Question 21

- Semivariogram results of comparing Mean pH to Elevation:


```
C (10²)

1.366
1.024
0.683
0.341
    0
-0.341
        0    0.219   0.439   0.658   0.877   1.097   1.316   1.535   1.755
                                                    Distance (10⁵ Meter)

— Model   • Binned   ╋ Averaged
Model : 0*Nugget+0.17054*Stable(112380,0.47422)
```

Question 23

- Farming or mining can affect pH levels by introducing chemicals and pollution into bodies of water.
- Variations temperature can also affect carbon levels in water at higher elevations.

Question 26

Square root of the Error2 column = 19.288987868 / count of records = 0.02553443619

Question 27

- Kriging Standard Error can be used as a stochastic result, which measures the reliability of predicted values and spatial variability between points.