GEO-107

Database Mgmt and Spatial Stats (Alireza Ghaffari)

Environmental Geomatics

Fall 2024

Submitted to:

Professor Alireza Ghaffari

Submitted by:

Enoch Kayode-Thomas-A00180246

Chetan Bhimsingh Chavan-A00148718

Kirandeep Kaur: - A00177972

Ashish Thankappan : A00175971

(Date: 12-12-2024)

Content

Report on Predictive Modeling for Ontario Wells Depth (2024)

Introduction

The objective of this project was to develop predictive models for estimating the depths of wells in Ontario, Canada, using three different techniques: Inverse Distance Weighting (IDW), Kriging, and Kriging Model 2. This report outlines the methodology, data preprocessing, model implementation, evaluation metrics, and visualizations of the results, along with insights gained during the project.

Data Collection and Preprocessing

The dataset utilized for this study comprises well data from Ontario, including features such as geographic coordinates and recorded depths. The following steps were performed to prepare the data for analysis:

1. **Source of Data:**

The dataset comprised well logs from Ontario, including spatial coordinates (latitude and longitude) and well depths. These records provided the foundation for spatial interpolation techniques

2. Data Cleaning:

- Removal of missing or inconsistent entries.
- Verification of coordinate accuracy to ensure correct spatial referencing.

3. Standardization:

- The well depth values were standardized to ensure comparability across different models and to reduce the impact of outliers.

4. Exploratory Data Analysis (EDA):

- Summary statistics and visualizations were generated to understand the distribution of well depths and to identify patterns in the spatial data.
- Moran's I statistic was calculated to assess spatial autocorrelation, confirming the appropriateness of geostatistical methods such as Kriging.

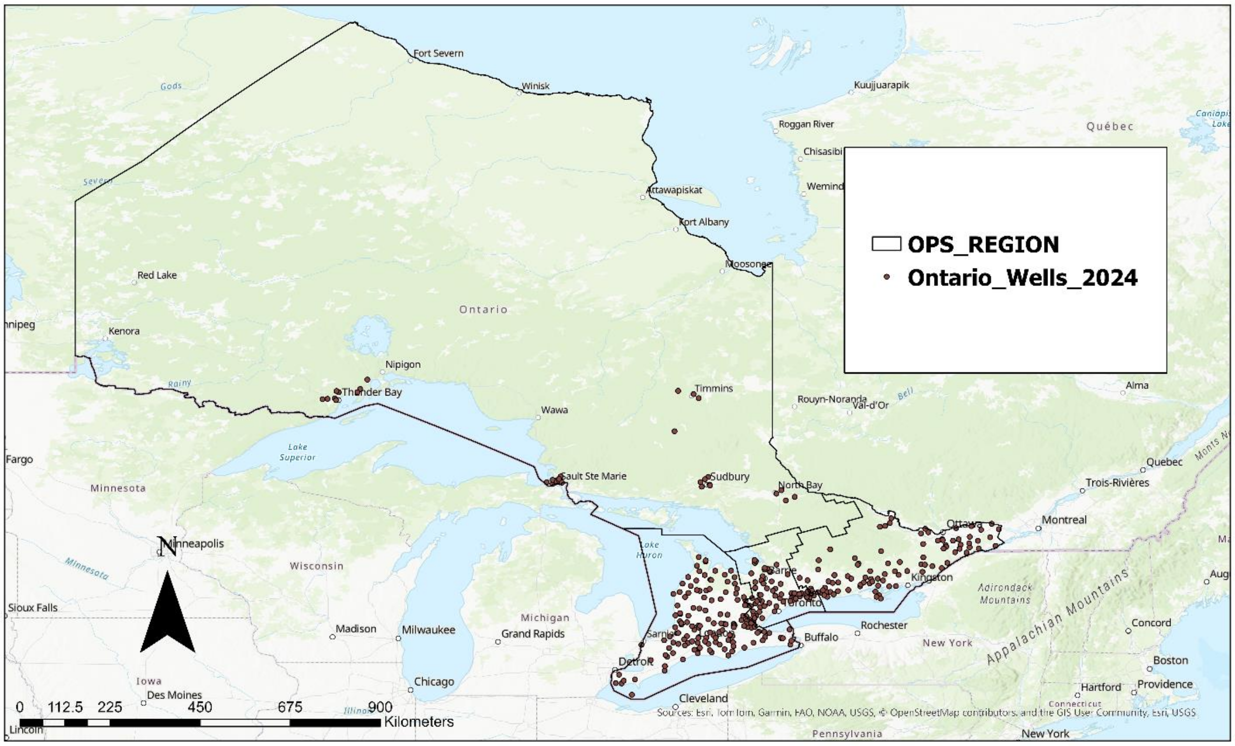

Figure 1. Map showing the Ontario Wells Distribution for the year 2024

Performing a spatial correlation analysis on the Ontario well distribution shows that;

Given the z-score of 4.092423, there is a less than 1% likelihood that this clustered pattern could be the result of random chance.

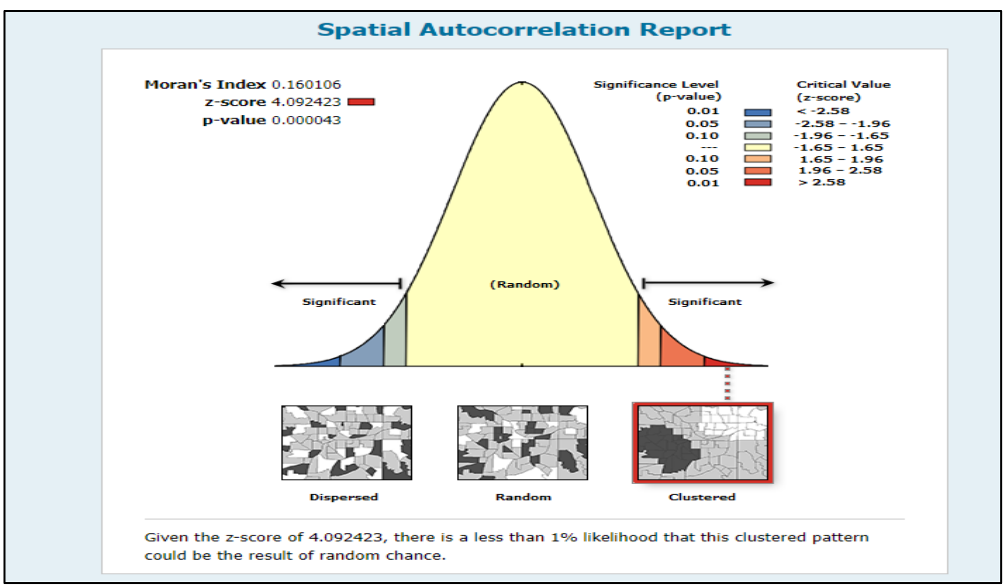

Figure 2 z-score

Predictive Techniques

1. **Inverse Distance Weighting (IDW):**

IDW is a deterministic interpolation method that estimates unknown values based on the weighted average of nearby known values, with weights inversely proportional to distance IDW assumes that points closer to the target location have more influence on the prediction than those farther away.

A power parameter was optimized to determine the best weight distribution. The parameters for the IDW technique are show below:

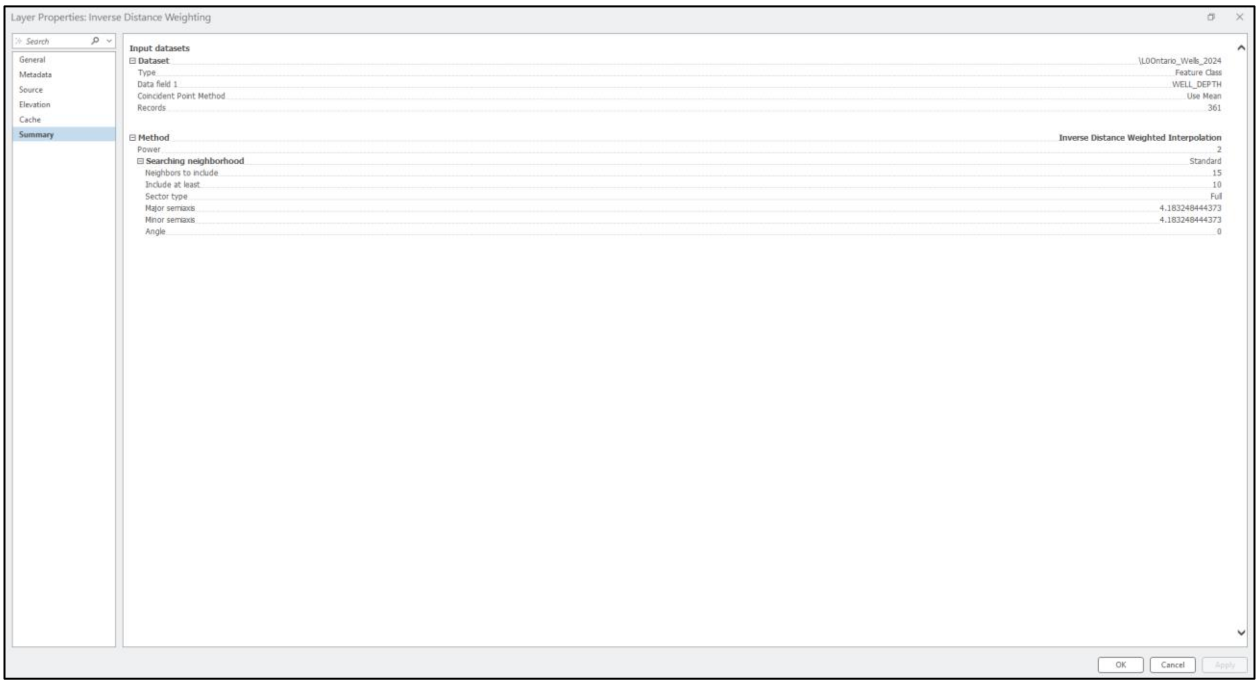

Figure 3: Layer Properties Inverse Distance Weighing

The plot and method report suggest that the inverse distance weighted interpolation model performs well, but there is room for improvement in terms of bias and accuracy for higher measured values.

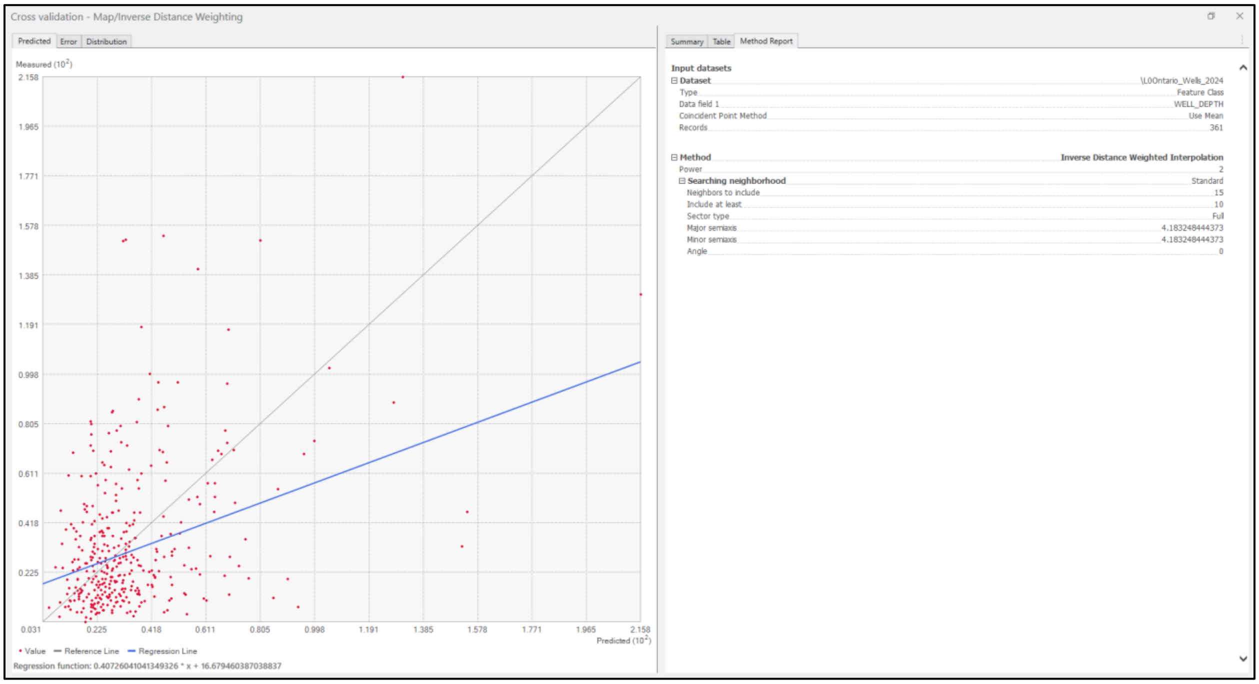

Figure 4: Cross validation: Map Inverse Distance Weighting

The result of the IDW analysis gives the result:

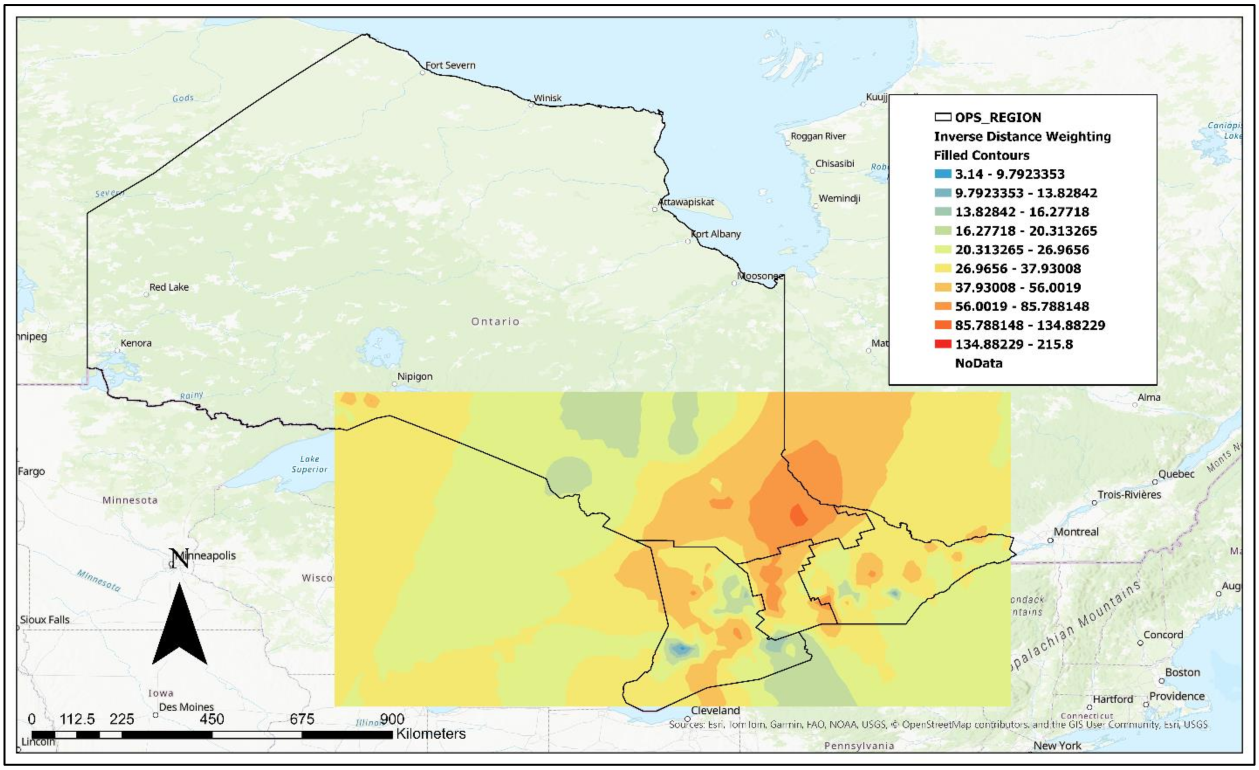

Figure 5: Idw analysis

The principles of IDW and interpreting the resulting map, helps to gain valuable insights into the spatial distribution of the variable of interest.

Inverse Distance Weighting of an Image

This image shows the inverse distance weighting of a geographic region.

- **The colors represent different levels of weighting.** The darker shades indicate higher levels of weighting, while lighter shades indicate lower levels of weighting.
- **The image covers a wide area of Canada,** encompassing regions like northern Ontario, Quebec, and the southern part of Manitoba.
- **The weighting pattern highlights the location of specific features or data points.** This could be related to population density, elevation, or other relevant geographic factors.

2. **Kriging (Ordinary Kriging):**

Kriging is a geostatistical method that considers both the distance and the degree of variation between known data points.

This geostatistical method models the spatial autocorrelation in the data. A semivariogram was constructed to describe the spatial relationship between points, and parameters (range, sill, nugget) were estimated.

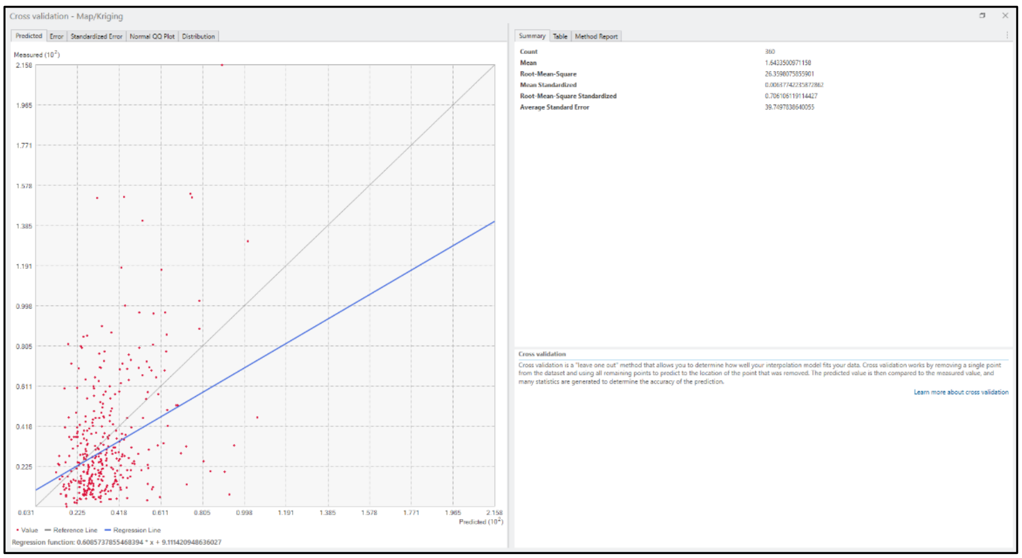

Figure 6: Cross validation: Map/Kriging

The plot displays predicted vs. measured values, with red points representing data, blue line representing regression line, and gray line indicating a 1:1 relationship, indicating model performance.

The results of the Kriging Analysis:

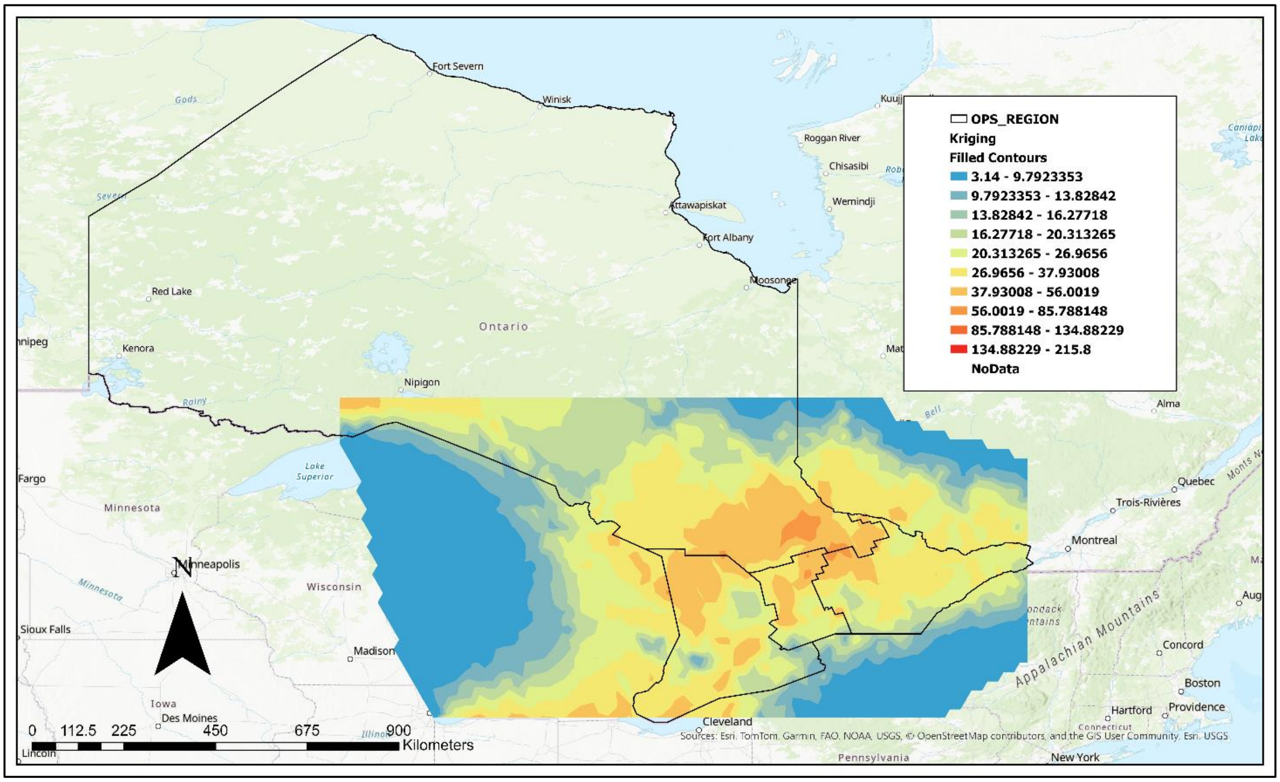

Figure 7

The map's contours reveal a high concentration of kriging values in the midwestern and eastern regions, decreasing towards the north and west, and a data gap in the northeast.

3. **Kriging Model 2(Enhanced Kriging):**

This variant of Kriging incorporated additional refinements in semivariogram fitting, including automated parameter optimization and cross-validation. Enhanced Kriging model with additional refinements in the semivariogram fitting process.

Different variogram models (e.g., spherical, exponential) were tested to select the best fit for the data.

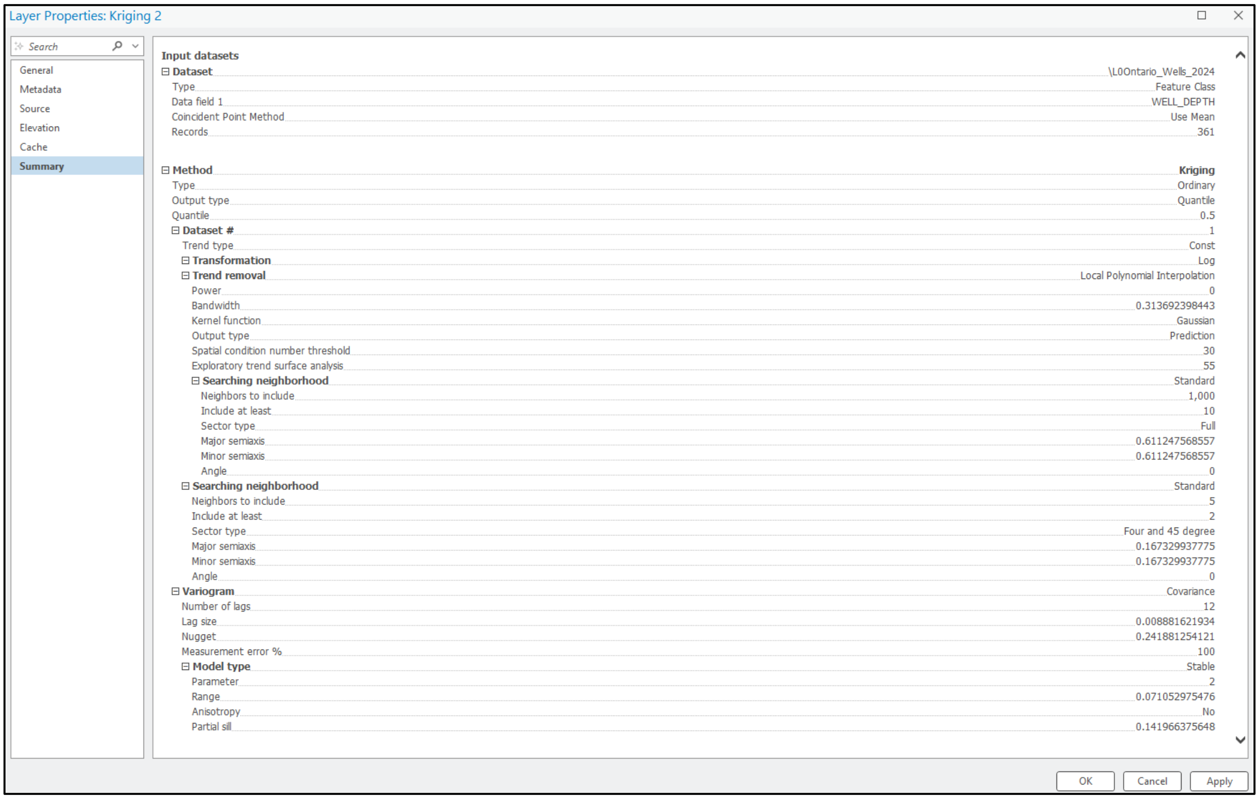

Figure 8 : Layer properties Kriging2

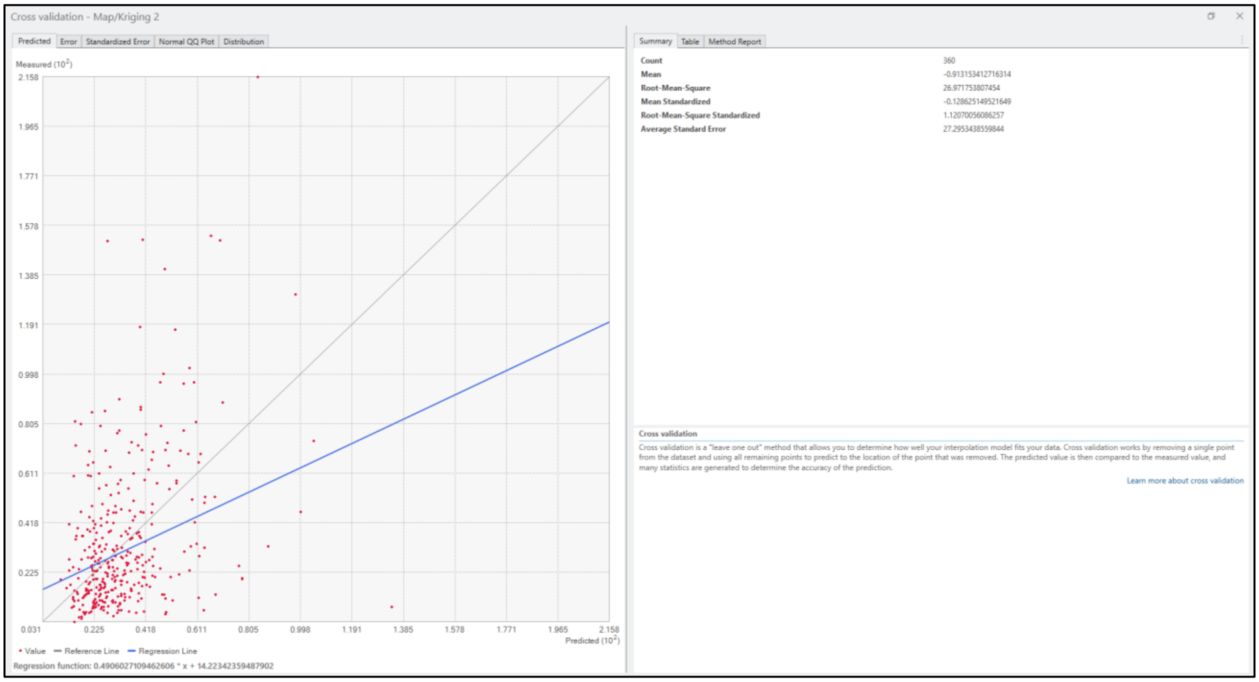

Figure 9 :cross validation

- 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.

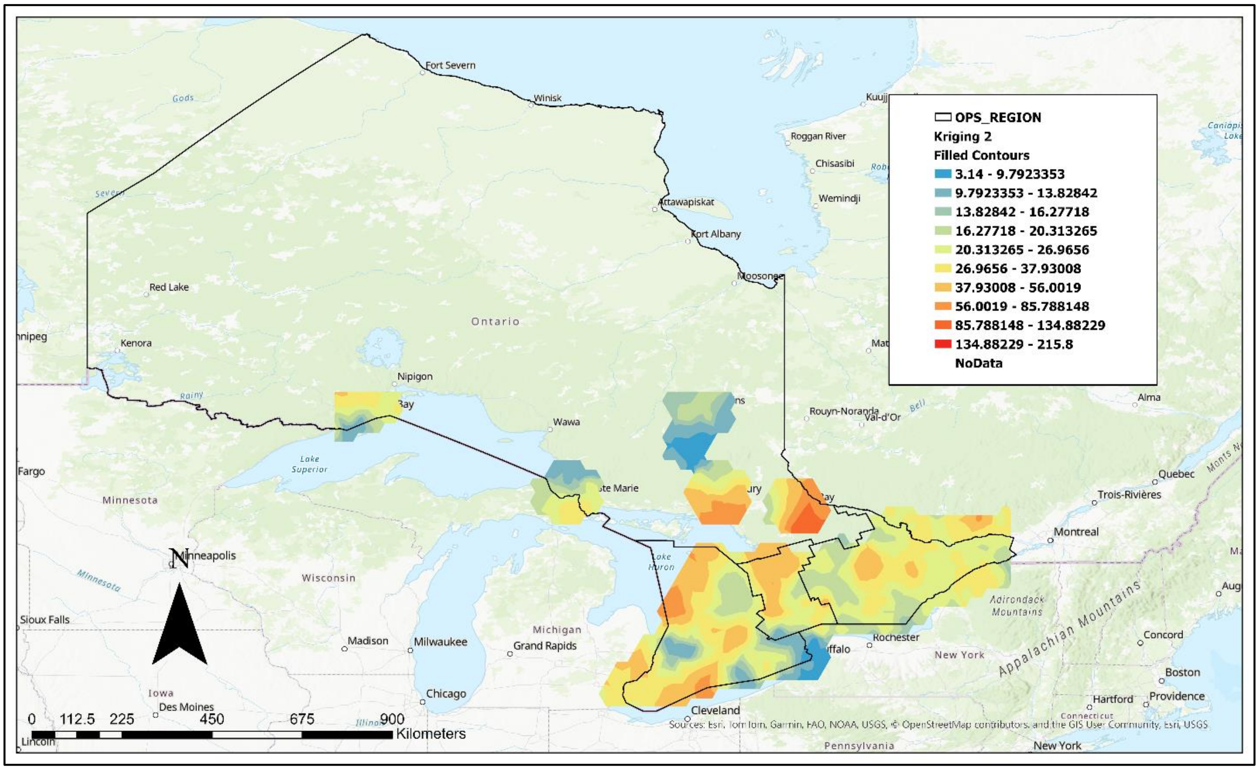

Figure 10

Results and Visualizations

1. **Box Plot Analysis:**

A comparative box plot of standardized well depths and predictions from IDW, Kriging, and Kriging Model 2 revealed the distribution and variance of the predictions relative to the actual data.
From the result below the IDW exhibited the least variance and closest alignment with observed values.

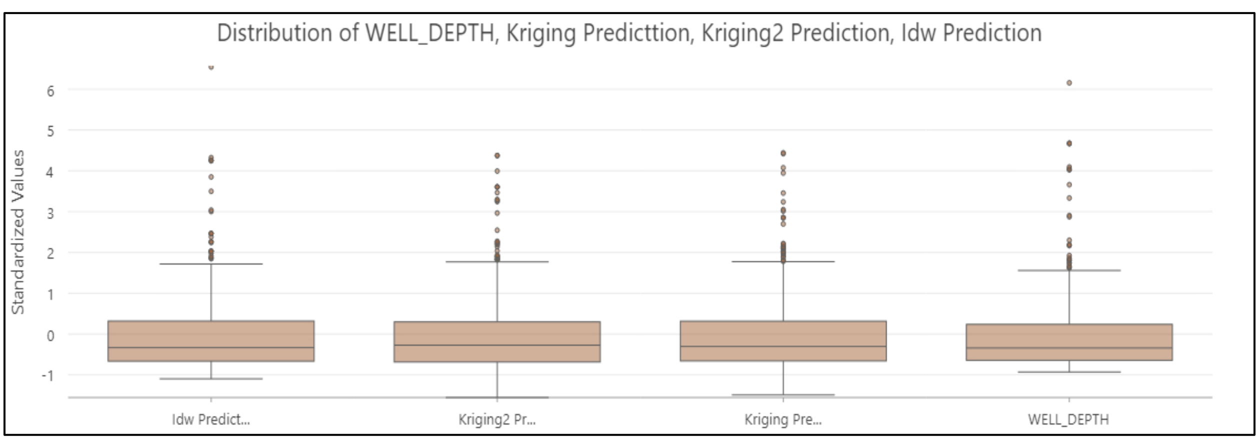

Figure 11. Box plot Anakysis

2. **Scatter Plot Analysis:**

Scatter plots were created to examine the relationships between observed and predicted depths for each model.

Idw prediction displayed tighter clustering around the line of equality compared to Kriging and Kriging 2, indicating better performance.

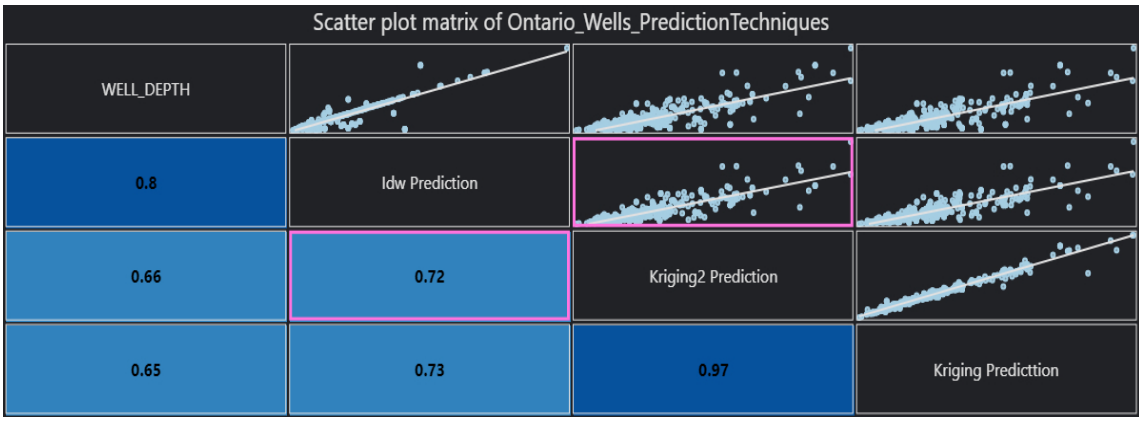

Figure 12. Scatter plot Analysis

Limitations:

- Sparse data points in some regions may reduce the accuracy of predictions.
- Assumptions of stationarity in Kriging may not fully capture spatial heterogeneity.

Conclusion and Recommendations

The project demonstrated that geostatistical models, particularly Kriging Model 2, are well-suited for predicting well depths in Ontario. Future work could explore:

- Incorporating additional covariates (e.g., geological features).
- Using machine learning methods for comparison with geostatistical approaches.
- Increasing data density in underrepresented areas to improve prediction accuracy.

Reference:

1. "C:\Users\User\Downloads\GEO 107 Final Project\GEO 107 Final Project\Ontario_Wells_PredictionTechniques Report.pdf"