LAB - 5

By

Nishant Rodrigues

2348045

5 MSc Data Science

(Geospatial Analysis)

Introduction to Interpolation

Interpolation is a technique used to estimate values at locations where data is not available, based on known values at other locations. It is widely used in creating spatial datasets, such as precipitation maps, elevation models, or population density maps. Spatial interpolation generates a statistical surface by leveraging the spatial relationship between known data points to predict values at unknown locations. For instance, in the absence of evenly distributed weather stations, spatial interpolation can estimate rainfall at unmonitored sites using data from nearby weather stations. Other common applications include estimating temperature, snow accumulation, groundwater levels, and other environmental or demographic attributes.

Inverse Distance Weighted (IDW) Interpolation

IDW is a deterministic interpolation method that estimates values at unknown points based on the proximity of known points, assuming that points closer to the unknown location have a greater influence on its value. The method uses a **weighting coefficient** to control how the influence of known points diminishes with increasing distance. A higher weighting coefficient reduces the impact of distant points, making the interpolation more dependent on nearby values.

The formula for IDW interpolation is:

Z(s0)=∑i=1NZ(si)·di−p∑i=1Ndi−p*Z*(*s*0)=∑*i*=1*Ndi*−*p*∑*i*=1*NZ*(*si*)·*di*−*p*

Where:

- Z(s0)*Z*(*s*0): The interpolated value at the prediction location.
- Z(si)*Z*(*si*): The measured value at location i*i*.
- di*di*: The distance between the prediction location and location i*i*.
- p*p*: The weighting power (coefficient).
- N*N*: The number of known points.

IDW assumes a straightforward distance-based decay, with nearby points having more influence than distant ones. While effective, IDW does not account for spatial correlations or variations beyond simple distance metrics.

Nearest-Neighbor Interpolation

The **nearest-neighbor interpolation** method assigns the value of the closest known data point to an unknown location. It is a simple and fast technique where the interpolation process involves identifying the nearest measured point to the target location and directly assigning its value. This method produces a blocky surface as there is no smoothing or consideration of surrounding points. It is ideal for categorical or discrete data but may not be suitable for continuous variables like elevation or precipitation.

Kriging

Kriging is a geostatistical interpolation method that not only considers the distances between known and unknown points but also incorporates the spatial autocorrelation among data points. Spatial autocorrelation measures how the values at one location are related to values at neighboring locations. Kriging builds a statistical model based on this relationship to provide a prediction surface along with an estimate of prediction accuracy.

Key Steps in Kriging:

1. **Exploratory Statistical Analysis**: Understand the spatial distribution and trends in the data.
2. **Variogram Modeling**: Develop a variogram to quantify spatial autocorrelation, capturing how the data values vary with distance and direction.
3. **Surface Creation**: Use the variogram to fit a mathematical function to the spatial data, estimating values for unknown locations.
4. **Variance Surface Analysis (Optional)**: Create a map showing the uncertainty or reliability of the predictions.

The general formula for Kriging is:

Z(s0)=∑i=1NλiZ(si)*Z*(*s*0)=*i*=1∑*NλiZ*(*si*)

Where:

- Z(s0)*Z*(*s*0): The interpolated value at the prediction location.
- Z(si)*Z*(*si*): The measured value at location i*i*.
- λi*λi*: The weight assigned to each measured value, which depends on distance, the variogram, and the spatial arrangement of the points.
- N*N*: The number of measured points.

Unlike IDW, Kriging incorporates both distance and the spatial relationship among the points through a fitted variogram model, making it more robust and reliable for datasets where spatial correlations significantly impact the results. Kriging is commonly applied in fields like geology, soil science, and environmental modeling, where accurate spatial predictions are essential.

DATASET DESCRIPTION

The dataset in question comes from **Statistics Canada** and consists of **monthly climate summaries** for various provinces and territories of Canada. It is a comprehensive and publicly accessible dataset that provides valuable insights into regional weather patterns. Here's a detailed description of the data:

Dataset Overview:

1. **Source**:
    - Statistics Canada, which maintains and provides open access to a wide array of data, including climate statistics.
2. **Region**:
    - The tutorial focuses on **Alberta**, one of Canada's provinces, with its geographical and climatic variations. However, similar data is available for other regions of Canada.
3. **Time Period**:
    - The dataset includes monthly climate summaries, which capture weather conditions over time. For this tutorial, data for **December 2019** is being used.
4. **Key Features**:
    - **Latitude (Lat)**: Represents the geographical latitude of weather stations.
    - **Longitude (Long)**: Represents the geographical longitude of weather stations.
    - **Mean Temperature (Tm)**: The average mean temperature recorded at each weather station during the specified time frame.
5. **Purpose**:
    - The data is well-suited for creating spatial interpolation models like **Inverse Distance Weighting (IDW)** or **Nearest-Neighbor Interpolation**, which estimate temperature values in areas without direct weather station measurements.
    - By understanding regional temperature patterns, this dataset supports applications in **climate analysis**, **agriculture**, and **urban planning**.
6. **Accessibility**:
    - The dataset is publicly available and accessible through the Statistics Canada website. For further details and similar datasets, you can visit their "About the Data" page.

7. **Applications**:
    - **Visualization**: Mapping temperature distributions across regions.
    - **Spatial Interpolation**: Estimating values at unmeasured locations.
    - **Trend Analysis**: Understanding seasonal or geographic variations in weather conditions.
8. **Data Format**:
    - The dataset is provided as a CSV file, which can be easily loaded and manipulated using Python libraries like pandas and geopandas.

Using Python

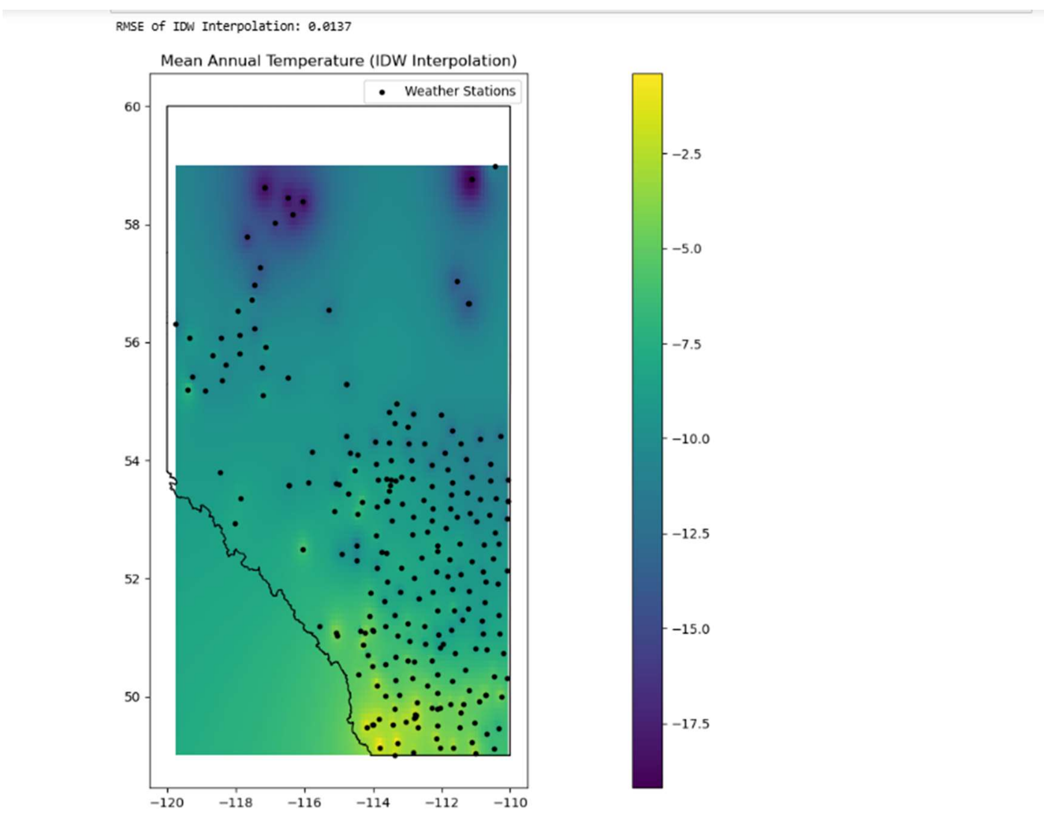
IDW

Methodology:

1. **Data Preprocessing**:
    - **CSV File**: Loaded climate data (Lat, Long, Tm) for Alberta.
    - Converted it to a **GeoDataFrame** for geospatial processing.
    - Defined a regular **grid** for interpolation based on the geographical extent of Alberta.
2. **IDW Interpolation**:
    - Used a **custom Python function** to compute interpolated temperatures across the grid.
    - For each grid point:
        - Calculated the **distance** from all weather station points.
        - Computed **weights** inversely proportional to the distance (powered by a factor, power=2).
        - Calculated the **weighted average** of temperatures based on these weights.
3. **Error Assessment (RMSE)**:

- Predicted temperature values at the weather station locations using IDW.
- Compared predicted values with observed values to compute the **Root Mean Squared Error (RMSE)**, assessing the model's accuracy.

4. **Visualization**:
   - Generated a **temperature surface map** using the interpolated data.
   - Added **colorbars** for temperature scales and overlaid weather station points.
   - Superimposed Alberta's boundary for clarity.
5. **Model Validation**:
   - Computed **RMSE** to measure the deviation between actual and predicted values, ensuring reliable interpolation.



The image depicts **Mean Annual Temperature in Alberta** using **Inverse Distance Weighting (IDW) Interpolation**. Here are the key observations:

1. **Temperature Gradient**:
   - Warmer regions (indicated by yellow shades) are observed in the **southern parts** of Alberta.
   - Cooler regions (indicated by dark purple shades) dominate the **northern parts**.
2. **Weather Stations**:
   - Black dots represent the **locations of weather stations** used for the interpolation.
   - A high density of weather stations in certain areas ensures better interpolation accuracy.
3. **Spatial Variability**:
   - The gradual change in colors reflects the spatial variability of temperature across the region, interpolated from weather station data.
4. **Boundary**:
   - The **geographical boundary of Alberta** (shapefile) is overlaid to provide context.

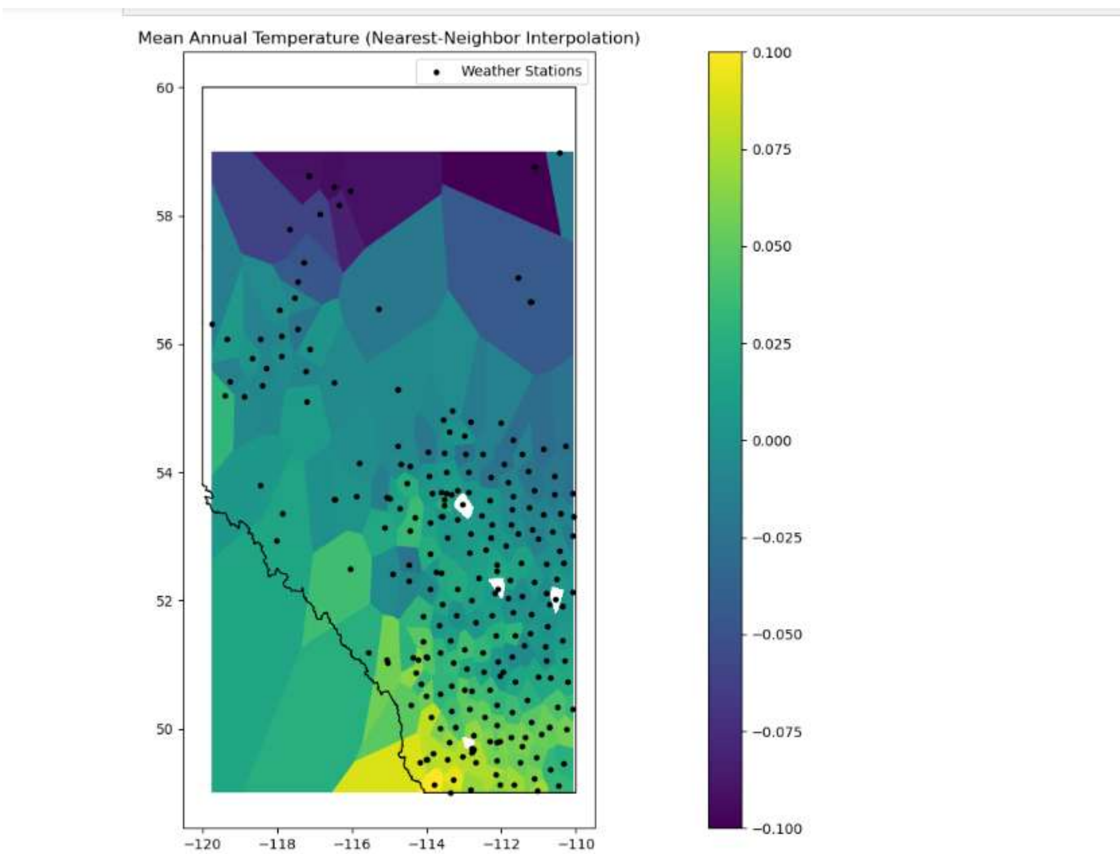
NEAREST NEIGHBOUR

1. **Grid Creation**:
   - Defined a **regular grid** over the data's geographic extent (longitude and latitude) with a resolution of **0.01 degrees** to interpolate temperature values.
2. **Nearest-Neighbor Interpolation**:
   - Used scipy.interpolate.griddata to perform **nearest-neighbor interpolation**, estimating temperature values on the grid from the weather station data.
3. **Boundary Overlay**:
   - Loaded a **shapefile** representing Alberta's boundary to provide a geographical context for the map.
4. **Visualization**:
   - Created a plot:
     - Displayed the interpolated temperature surface with a **color map** (viridis).

- Overlayed **weather station locations** (black dots).
- Added Alberta's **boundary** as an outline.

o Included a **color bar** to represent temperature scales and a legend for clarity.

This process enables spatial visualization of mean annual temperature using nearest-neighbor interpolation.



This visualization represents the **mean annual temperature** in Alberta, Canada, as estimated using the **Nearest-Neighbor Interpolation (NN)** method. Here's the significance of the elements:

1. **Interpolated Temperature Surface**:

   o The colored regions indicate the estimated temperature distribution across Alberta.

- The color bar to the right shows the temperature range, where darker shades represent lower temperatures (negative values) and lighter shades indicate higher temperatures.

2. **Weather Stations (Black Dots)**:
   - The black dots mark the locations of weather stations where temperature data was recorded.
   - The interpolation assigns temperature values to the areas between these stations based on the nearest recorded value.

3. **Shapefile Boundary**:
   - The black outline represents Alberta's geographical boundary, contextualizing the spatial data.

4. **Significance of Patterns**:
   - **Northern and higher latitude areas** generally show colder temperatures (darker colors).
   - **Southern and lower latitude areas** tend to have warmer temperatures (lighter colors), as expected due to geographical and climatic variations.

5. **Nearest-Neighbor Interpolation Characteristics**:
   - The **polygonal patterns** are characteristic of the NN method, where each grid cell adopts the value of its nearest weather station.
   - This method is simple but does not smooth gradients, leading to abrupt changes between zones.

6. **Use Case**:
   - This map is valuable for identifying temperature trends in areas lacking direct weather data, aiding climate studies and decision-making in agriculture, forestry, and urban planning.

KRIGING

1. **Set Up Kriging Model**:
   - **Inputs to OrdinaryKriging**:
     - lons: Longitude values of the weather stations (x-coordinates).
     - lats: Latitude values of the weather stations (y-coordinates).

    - temps: Temperature values recorded at those weather stations.
    - variogram_model="spherical": Specifies the variogram model used to represent spatial correlation between points. In this case, a **spherical variogram** is chosen.
    - verbose=False: Disables detailed output.
    - enable_plotting=False: Disables the built-in variogram plotting.

2. **Generate the Interpolation Grid**:
    - The np.meshgrid() function is used to create a grid of longitude (grid_x) and latitude (grid_y) values across the area of interest.
    - These grids define the locations where the Kriging model will predict temperature values.

3. **Interpolate Data Using Kriging**:
    - The ok.execute("grid", grid_x[0], grid_y[:, 0]) method performs interpolation:
        - "grid": Specifies that Kriging should produce a gridded output.
        - grid_x[0] and grid_y[:, 0]: Provide the x and y coordinates of the grid.
        - grid_z: Contains the interpolated temperature values for the grid points.
        - ss: Provides the Kriging variance, which quantifies the uncertainty of predictions at each grid point.

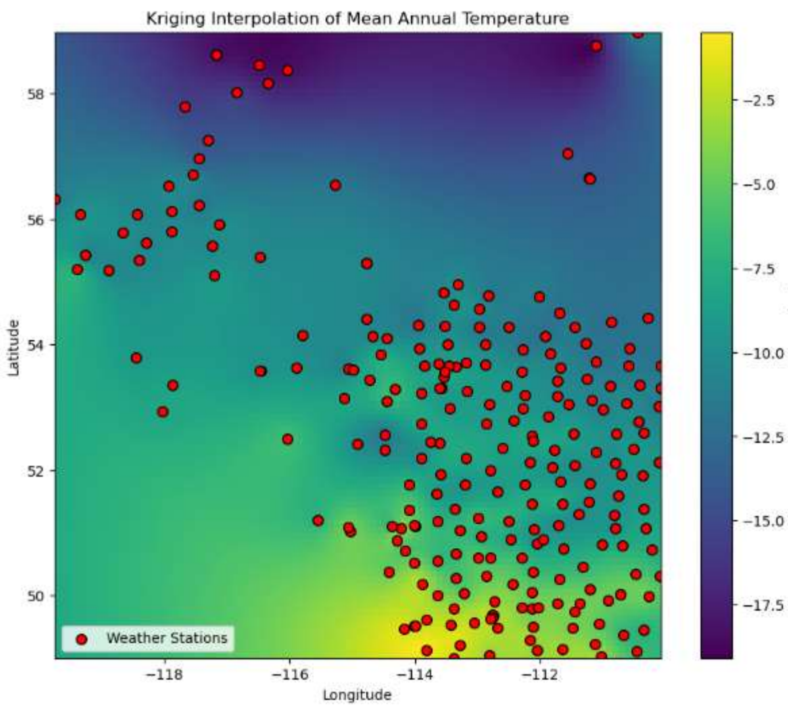
4. **Visualize the Interpolated Surface**:
    - After interpolation, the results (grid_z) are visualized using plotting libraries like Matplotlib.
    - Typically, a heatmap or contour map is created to show the spatial variation in temperatures.

Why Kriging is Useful:

- **Spatial Correlation**: Kriging accounts for the spatial autocorrelation of the data, meaning it leverages how nearby points influence each other.
- **Prediction with Uncertainty**: It provides not only the interpolated values but also an estimate of uncertainty (via Kriging variance).
- **Flexibility**: The choice of variogram models (e.g., spherical, exponential, Gaussian) allows customization to match the characteristics of the dataset.

Key Characteristics of Kriging:

- Unlike simpler methods (e.g., nearest-neighbor or IDW), Kriging produces smooth interpolations.
- It is particularly suited for applications where understanding spatial variability and prediction uncertainty are critical, such as climate modeling and geostatistics.



1. **Key Observations:**
   1. **Color Gradient:**
      - The color scale indicates variations in mean annual temperature, with purple/blue representing colder temperatures and yellow/green representing warmer temperatures.
      - The gradient suggests that northern areas (higher latitudes) are colder, while southern areas (lower latitudes) are warmer.
   2. **Weather Stations (Red Points):**

- The red dots represent the locations of weather stations where temperature measurements were collected.
- The distribution of stations appears uneven, with a denser concentration in certain areas (e.g., the southern-central region).

3. **Spatial Trends:**
    - There is a spatially correlated pattern in the temperature distribution, suggesting lower temperatures in the northwest and higher temperatures in the southeast.
    - The Kriging method has smoothly interpolated the temperatures in areas without weather stations based on the spatial relationships among the measured points.

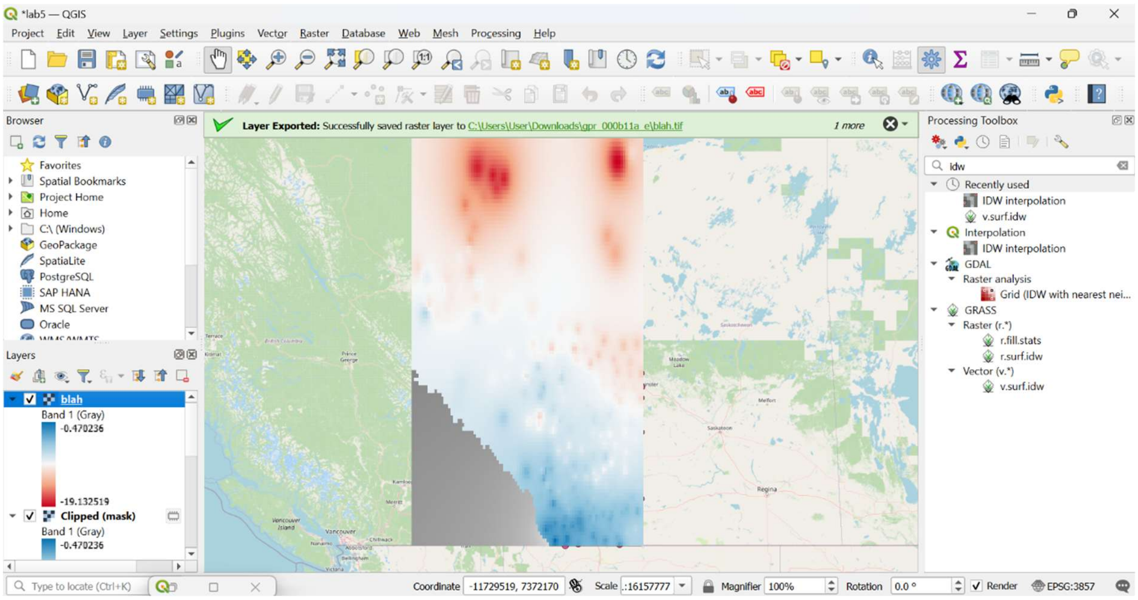
RMSE

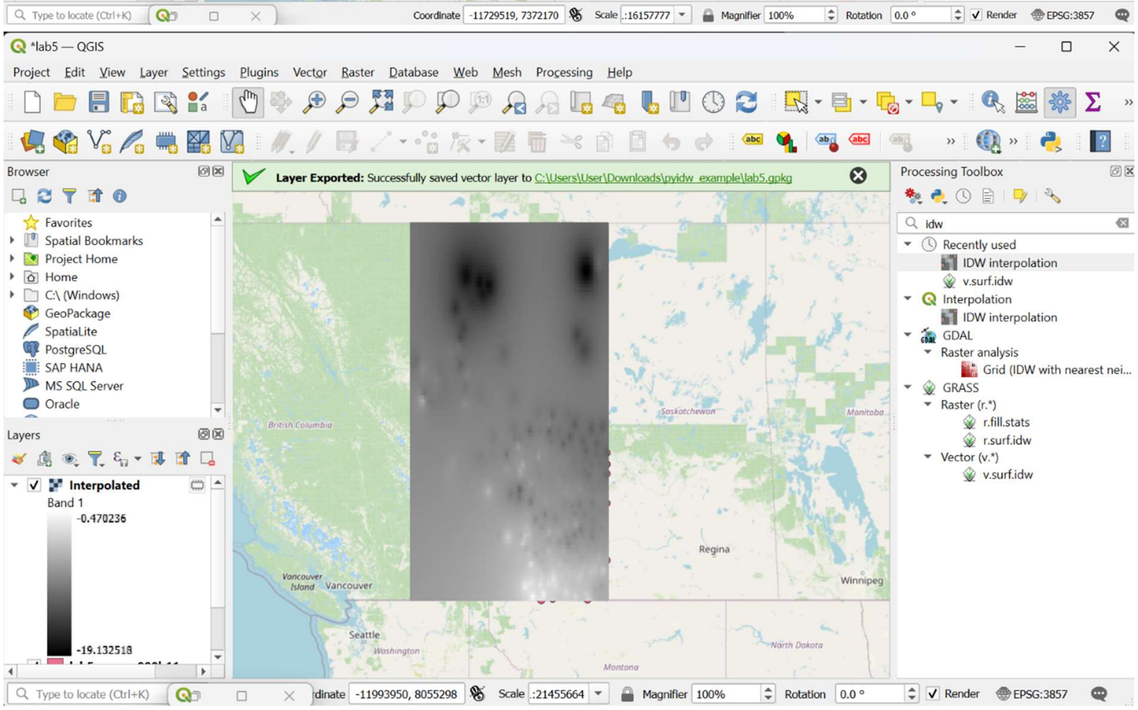
The Root Mean Squared Error (RMSE) values indicate the accuracy of the interpolation methods. Kriging and IDW, both with an RMSE of 0.0137, demonstrate higher precision compared to Nearest Neighbor, which has an RMSE of 0.0194. This suggests that Kriging and IDW effectively capture spatial variability and provide smoother predictions. Nearest Neighbor, while simpler, lacks the nuanced modeling of distance or spatial correlations, resulting in less accurate interpolation. Thus, for this dataset, Kriging and IDW are better choices for generating reliable spatial predictions.

Conclusion

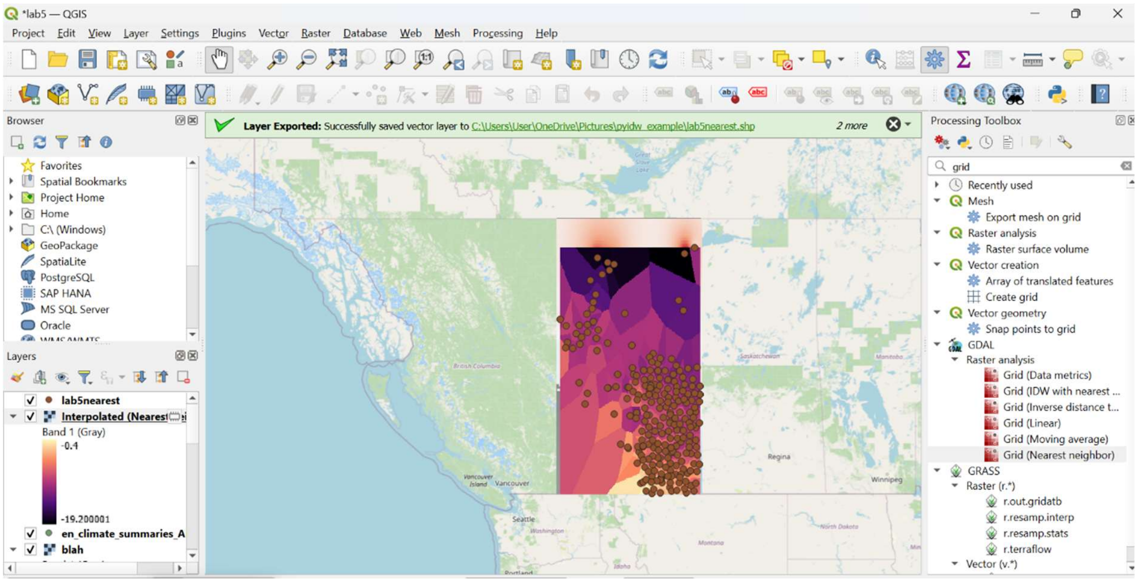
In this lab, we explored three spatial interpolation techniques—Kriging, Inverse Distance Weighting (IDW), and Nearest Neighbor—and learned how they estimate values at unmeasured locations. Nearest Neighbor provides a simple, discrete interpolation by assigning the value of the closest point, while IDW produces smoother results by weighting nearby points more heavily based on distance. Kriging, however, stands out by incorporating spatial autocorrelation through variograms, offering a more sophisticated and accurate interpolation along with uncertainty estimates. This comparative study highlights how the choice of method depends on the nature of the dataset and the level of precision required.

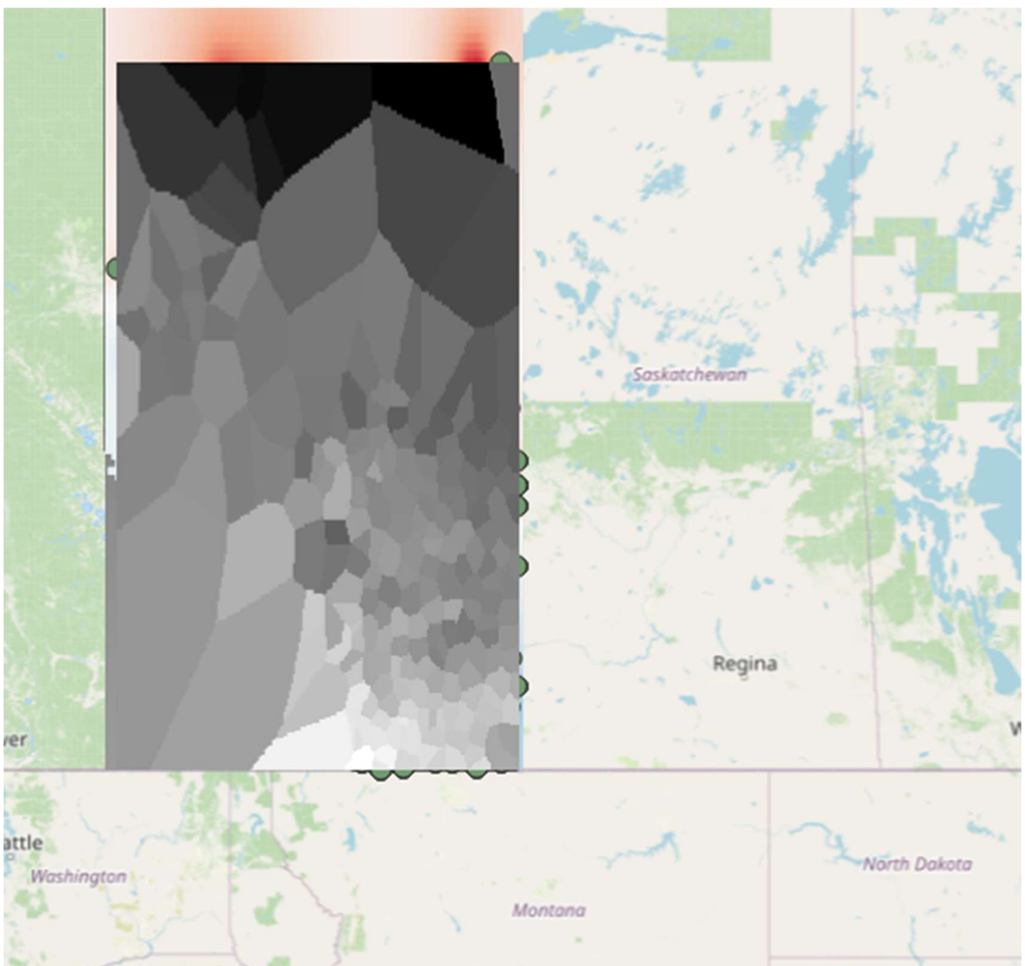
Using QGIS



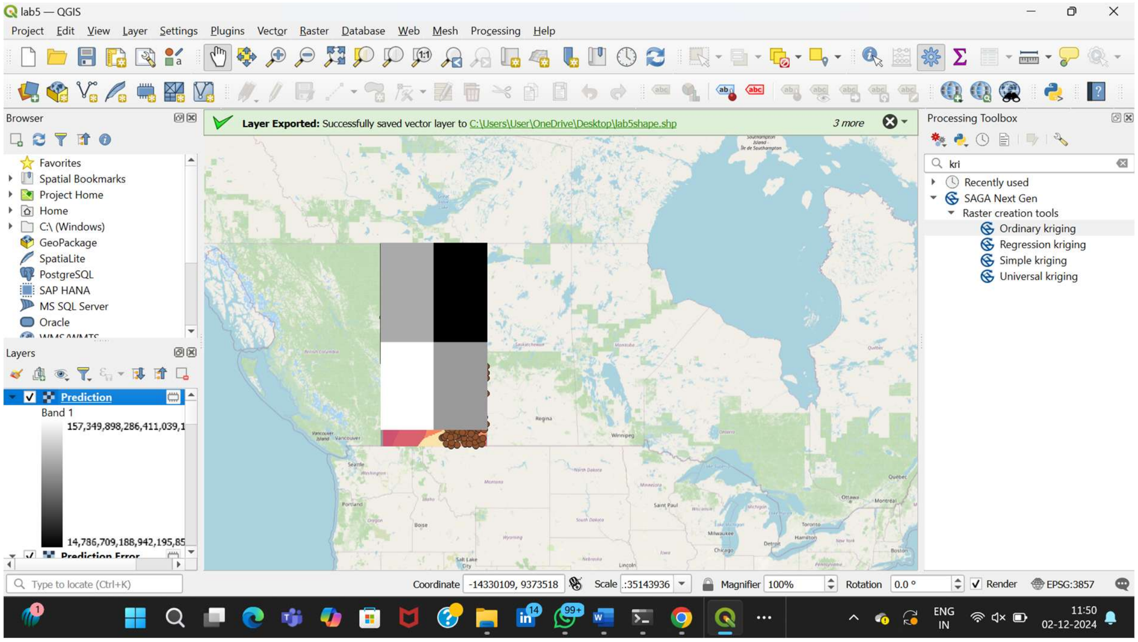


Nearest Neighbour


*lab5 — QGIS
Project
Edit
View
Layer
Settings
Plugins
Vector
Raster
Database
Web
Mesh
Processing
Help
Browser
Layer Exported: Successfully saved vector layer to C:\Users\User\OneDrive\Pictures\pyidw_example\lab5nearest.shp
2 more
Processing Toolbox
Favorites
Spatial Bookmarks
Project Home
Home
C:\ (Windows)
GeoPackage
SpatiaLite
PostgreSQL
SAP HANA
MS SQL Server
Oracle
Layers
lab5nearest
Interpolated (Nearest
Band 1 (Gray)
-0.4
-19.200001
en_climate_summaries_A
blah
grid
Recently used
Mesh
Export mesh on grid
Raster analysis
Raster surface volume
Vector creation
Array of translated features
Create grid
Vector geometry
Snap points to grid
GDAL
Raster analysis
Grid (Data metrics)
Grid (IDW with nearest ...
Grid (Inverse distance t...
Grid (Linear)
Grid (Moving average)
Grid (Nearest neighbor)
GRASS
Raster (r.*)
r.out.gridatb
r.resamp.interp
r.resamp.stats
r.terraflow
Vector (v.*)
Regina
Seattle
Vancouver


Saskatchewan
Regina
Washington
Montana
North Dakota

Kriging


lab5 — QGIS
Project Edit View Layer Settings Plugins Vector Raster Database Web Mesh Processing Help
Browser
Layer Exported: Successfully saved vector layer to C:\Users\User\OneDrive\Desktop\lab5shape.shp
3 more
Processing Toolbox
Favorites
Spatial Bookmarks
Project Home
Home
C:\ (Windows)
GeoPackage
SpatiaLite
PostgreSQL
SAP HANA
MS SQL Server
Oracle
kri
Recently used
SAGA Next Gen
Raster creation tools
Ordinary kriging
Regression kriging
Simple kriging
Universal kriging
Layers
Prediction
Band 1
157,349,898,286,411,039,1
14,786,709,188,942,195,85
Prediction Error
Type to locate (Ctrl+K)
Coordinate -14330109, 9373518
Scale 1:35143936
Magnifier 100%
Rotation 0.0 °
Render
EPSG:3857
ENG
IN
11:50
02-12-2024