**Lab 3 Report: Cost Surface**

Title: Cost Surface Model

Notice: Dr. Bryan Runck

Author: Maochuan Wang

Date: 04/17/2023

**Project Repository:**

**Google Drive Link:**

**Time Spent:** 8 hours

**Abstract**

This lab’s purpose consists of creating an ETL pipeline to retrieve data from Google Places and the Minnesota Geospatial Commons (MGC) to develop 3 cost surface models that allow for finding the optimal path for Dory to get from her house to the North Picnic area. The cost of the surface increases due to steep slopes, muddy land cover types, and crossing water bodies without bridges or waders. Map algebra was performed by rasterizing all the variable datasets and reclassifying them on a scale from 1 to 4 where 4 is the most expensive value. Additionally, the pipeline was developed so the user could input different weight factors to each variable. Finally, the Cost Distance and Cost Path tools in ArcPro were used to generate the optimal paths. This exercise found the results yielded by the script are the same as the routes suggested by advanced mapping software such as Google Maps, which indicates high accuracy.

**Problem Statement**

This lab aims to find an optimal route to go from Dory’s house (outside of Whitewater State Park) to the North Picnic area. As several variables are added, such as watercourses, bridges, slopes, etc., the cost surface model plus the cost distance and cost path will show the optimal route where Dory would prefer to walk to get to the park based on different weights assigned by the user to each variable. Figure 1 illustrates the route-finding model.

Diagram, engineering drawing

Description automatically generated

Figure 1. Conceptual route-finding model based on B. Runck’s personal communication (2022)

Table 1. Data required for cost surface analysis

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **#** | **Requirement** | **Defined As** | **(Spatial) Data** | **Attribute Data** | **Dataset** | **Preparation** |
| 1 | Start and end points | Pair of coordinates | Point geometry |  | Google Places | Convert coordinates from strings to points, project to NAD 1983 UTM Zone 15N, create a bounding box, and generate a buffer around the bounding box (AOI). |
| 2 | Elevation dataset | Minnesota DEM | Raster | Elevation | Mn GeoSpatial Commons | Clip to AOI, and generate slope dataset |
| 3 | Farm fields | Cropland data | Polygon | Land type | Mn GeoSpatial Commons | Clip to AOI |
| 4 | Water bodies | Streams or watercourses | Line |  | Mn GeoSpatial Commons | Clip to AOI |
| 5 | Bridges | Bridges | Point |  | Mn GeoSpatial Commons | Clip to AOI, and snap points to streams (lines) |

**Input Data**

First, the start point coordinates were given in the lab instructions, while the end point was obtained from Google Places through a personal API key. From the Minnesota Geospatial Commons portal, 4 datasets were downloaded: i) a state 30-m DEM, ii) a geodatabase containing the cropland layer 2021 raster with information about the different land cover types (including farm fields) in the state, iii) a line shapefile encompassing the state streams, and iv) a point shapefile representing the state bridges.

Table 2. Input data

|  |  |  |  |
| --- | --- | --- | --- |
| **#** | **Title** | **Purpose in Analysis** | **Link to Source** |
| 1 | End Point | Raw input strings containing the coordinates of the North Picnic area | https://maps.googleapis.com/maps/api/place/findplacefromtext/json?input=North%20Picnic%20area%20St%20Charles%20Minnesota&inputtype=textquery&fields=formatted\_address%2Cname%2Crating%2Copening\_hours%2Cgeometry&key=YOUR\_API\_KEY |
| 2 | Minnesota Digital Elevation Model - 30 Meter Resolution | Raw input dataset for slope generation | https://gisdata.mn.gov/dataset/elev-30m-digital-elevation-model |
| 3 | Cropland Data Layer 2021, Minnesota | Extract landcover information, especially farm fields | https://gisdata.mn.gov/dataset/agri-cropland-data-layer-2021 |
| 4 | Public Waters (PW) and Watercourse Delineations | Dataset with streams | https://gisdata.mn.gov/dataset/water-mn-public-waters |
| 5 | Bridge Locations in Minnesota | Locate bridges over water bodies | https://gisdata.mn.gov/dataset/trans-bridges |

**Methods**

To begin with, the North Picnic area (end point) location was obtained through a personal API key from Google Places and stored in a JSON dictionary. These coordinates together with those of the start point (given in the lab instructions) were then stored in lists from which a point shapefile was created. The shapefile was projected to the coordinate system NAD1983 UTM Zone 15 N and an envelope bounding box was used to generate a polygon shapefile. An 8-km buffer was applied to this polygon to create the area of interest (AOI) since the optimal route may not necessarily be within the bounding area. Figure 2 shows the flow diagram of this process.

Diagram

Description automatically generated

Figure 2. Flow diagram of AOI generation

Regarding the datasets from Minnesota Geospatial Commons (MGC), first, the DEM was clipped to the AOI. Later, it was used to generate the slope dataset which was eventually reclassified from 1 to 4 by utilizing the geometric interval to keep the optimal path the most gradual in terms of the slope. The cropland layer was also clipped to the AOI and reclassified from 1 to 4 as follows: i) 4 for all types of crops (farm fields), ii) 3 for grasslands as they may be muddy but not as much as the farm fields, iii) 2 for forests and shrublands since the leaves intercept rainwater acting as an umbrella which reduces the amount of mud, and iv) 1 for barren and developed areas which are the least muddy land types.

Likewise, the streams were clipped to the AOI, converted to a raster, and reclassified by assigning 4 to all of them (the least wanted condition). Then, as the bridges were represented with a point geometry and were not exactly over the stream-line layer, the stream vector layer clip was used to create a 30-m buffer to include the nearby bridges. Later, the bridges were snapped to the closest stream edges and converted to raster. Said raster was reclassified by assigning a value of 4, the same as the streams, so they can be multiplied, if desired, by the same weight factor as the streams but negative to counterbalance the cost of crossing water bodies. Figure 3 illustrates the reclassification process. It is worth mentioning that all the raster datasets had the same cell size: 30 m.

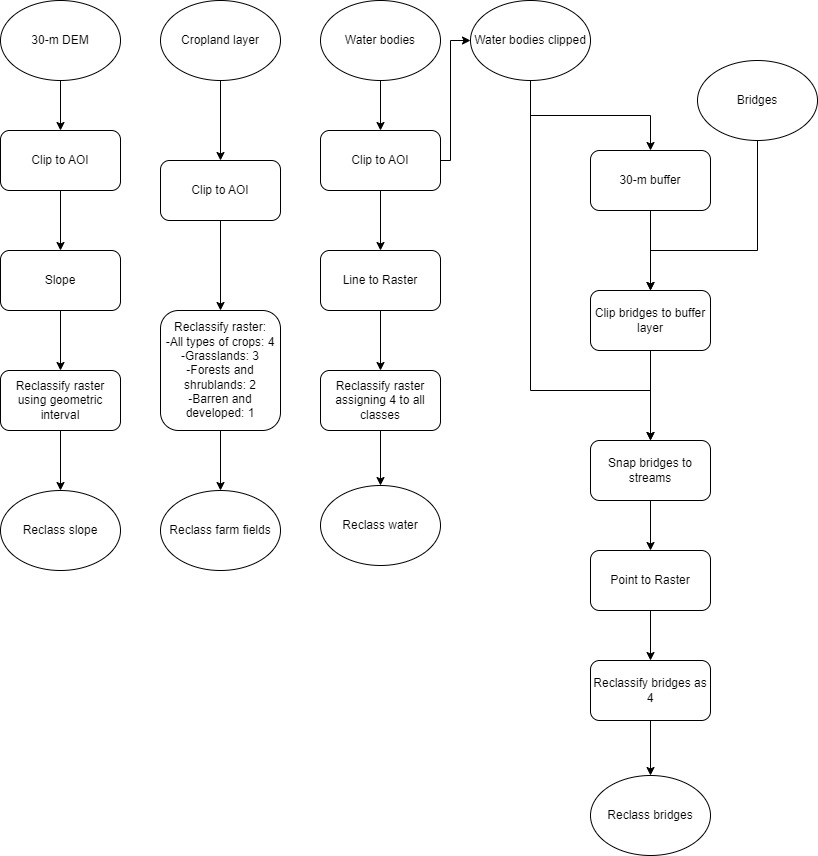
****

Figure 3. Flow diagram of MGC datasets reclassification

A loop was created to generate cost surface models based on different weight factors assigned to each variable (reclassified raster) by the user. As mentioned above, the user is recommended to assign a negative weight factor to the bridges to counterbalance the cost of crossing the streams. All the reclassified rasters are then added up by using map algebra and the output is saved to disk. There is a counter initialized in 0 that keeps track of the number of loops to break the loop after the third cycle. The cost surface created in each loop is saved with the sequence number of the said loop (counter) so as not to overwrite the outputs. This process is summarized in Figure 4.

Diagram

Description automatically generated

Figure 4. Flow diagram of cost surface model generation.

Finally, for each cost surface, the start point (Dory’s house) was selected through the Select By Attributes tool from the shapefile to use as an input along with the cost surface to generate the cost distance. It is worth mentioning that the Cost Distance tool yields a direction layer too. These two outputs plus the end point (through another selection by attributes) were utilized to run the Cost Path tool which found the final optimal route. The process is illustrated in Figure 5.

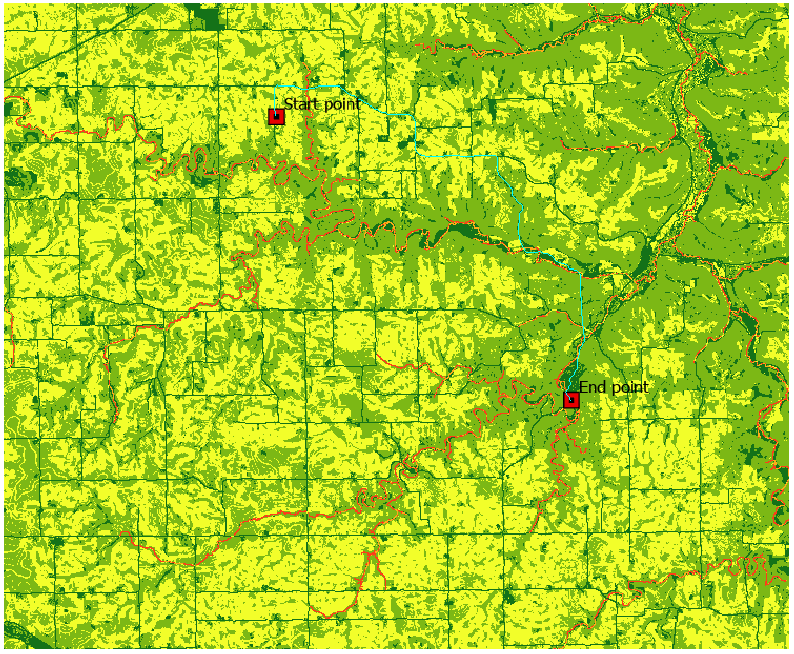
Diagram

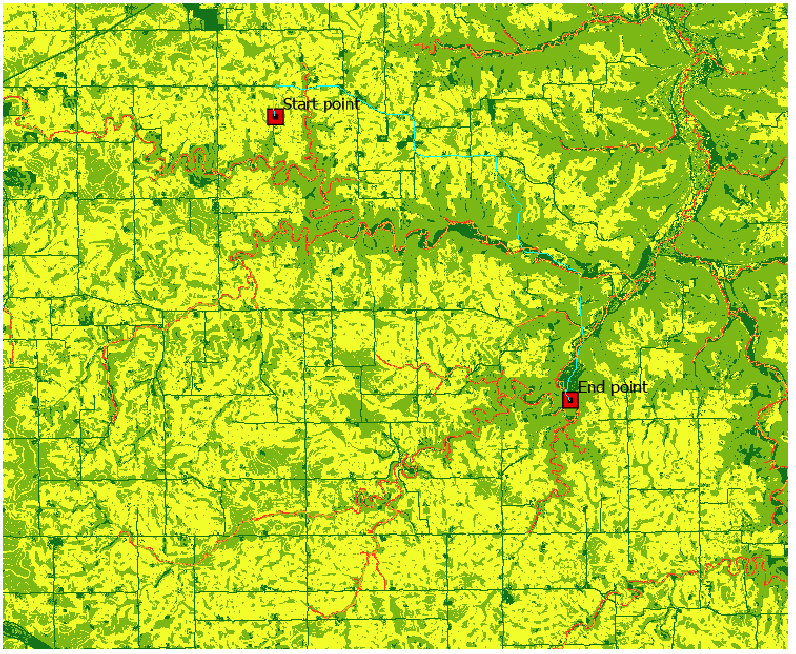
Description automatically generated

Figure 5. Flow diagram of optimal route finding

**Results**

The script finds successfully optimal routes such as those of Figure 6 whose cost surface models received different *w*, *x*, *y*, and *z* weight factors.





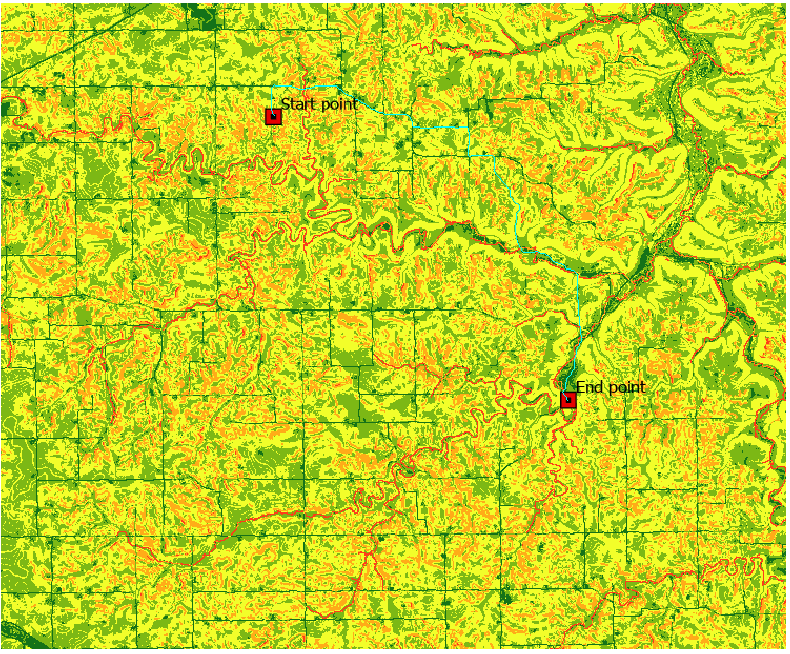


Figure 6. Optimal routes (cyan) for three cost surface models with different weight factors. Top: w=2, x=2, y=-2, z=1, Center: w=3, x=4, y=0, z=2; Bottom: w=1, x=1, y=-1, z=1. Low values are shown in dark, and high values in red.

The results show that the optimal path is lowly sensitive to changes in the weight factors as the 3 optimal routes remained pretty much the same, even though the cost surface models were indeed different in each case.

**Results Verification**

The results obtained here are highly accurate as the features in the cost surface models are displayed as expected. For example, in Figure 6, the streams have the highest cost since Dory is not willing to cross any water body as she is not wearing her waders. The lowest-cost routes are the roads since they are not muddy and are generally built on very gradual slopes. Furthermore, these results match the optimal path suggested by Google Maps as shown in Figure 7.

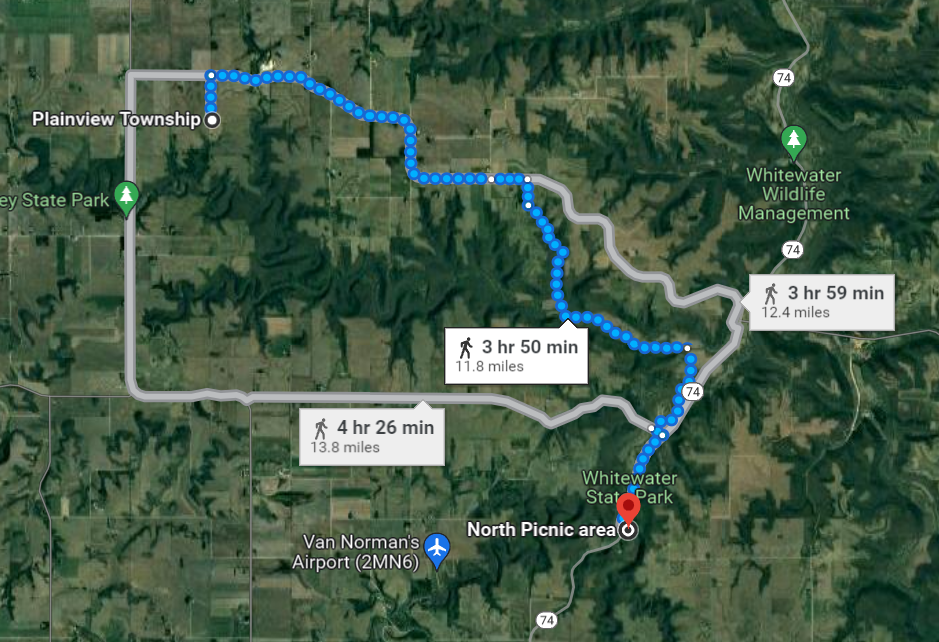


Figure 7. Optimal route suggested by Google Maps

**Discussion and Conclusion**

This exercise builds on Part 2 of Lab 2 and integrates the cost surface models with the start and end points to find the best route to go from Dory’s house to the North Picnic Area by utilizing Cost Distance and Cost Path tools in ArcPro. When comparing the outputs obtained here with those of advanced mapping software, such as Google Maps, the results are the same, which means this script has high accuracy. Furthermore, the workflow followed here can be applied to other scenarios as long as all the inputs overlap.

Overall, by carrying out this lab, I discovered and learned to use Cost Distance and Cost Path tools and interpret their outputs. Likewise, I learned to iterate through different folders and name the outputs following a sequence number.

**References**

Google. (2022). *North Picnic area*. Retrieved from Google Places: https://maps.googleapis.com/maps/api/place/findplacefromtext/json?input=North%20Picnic%20area%20St%20Charles%20Minnesota&inputtype=textquery&fields=formatted\_address%2Cname%2Crating%2Copening\_hours%2Cgeometry&key=YOUR\_API\_KEY

MNDNR. (2008). *Public Waters (PW) Watercourse Delineations*. Retrieved from Minnesota Geospatial Commons: https://gisdata.mn.gov/dataset/water-mn-public-waters

MNDOT. (n.d.). *Bridge Locations in Minnesota*. Retrieved from Minnesota Geospatial Commons: https://gisdata.mn.gov/dataset/trans-bridges

USDA. (2022). *Cropland Data Layer 2021, Minnesota*. Retrieved from Minnesota Geospatial Commons: https://gisdata.mn.gov/dataset/agri-cropland-data-layer-2021

USGS. (2004). *Minnesota Digital Elevation Model - 30 Meter Resolution*. Retrieved from Minnesota Geospatial Commons: https://gisdata.mn.gov/dataset/elev-30m-digital-elevation-model

**Self-score**

|  |  |  |  |
| --- | --- | --- | --- |
| **Category** | **Description** | **Points Possible** | **Score** |
| **Structural Elements** | All elements of a lab report are included **(2 points each)**:  Title, Notice: Dr. Bryan Runck, Author, Project Repository, Date, Abstract, Problem Statement, Input Data w/ tables, Methods w/ Data, Flow Diagrams, Results, Results Verification, Discussion and Conclusion, References in common format, Self-score | 28 | 28 |
| **Clarity of Content** | Each element above is executed at a professional level so that someone can understand the goal, data, methods, results, and their validity and implications in a 5 minute reading at a cursory-level, and in a 30 minute meeting at a deep level **(12 points)**. There is a clear connection from data to results to discussion and conclusion **(12 points)**. | 24 | 24 |
| **Reproducibility** | Results are completely reproducible by someone with basic GIS training. There is no ambiguity in data flow or rationale for data operations. Every step is documented and justified. | 28 | 28 |
| **Verification** | Results are correct in that they have been verified in comparison to some standard. The standard is clearly stated **(10 points)**, the method of comparison is clearly stated **(5 points)**, and the result of verification is clearly stated **(5 points)**. | 20 | 20 |
|  |  | 100 | 100 |