

A suitability model using GIS to identify optimal housing locations near Bonnyville, Alberta

Zach Ripka

Final project

GEOG 306

Prof. McKenzie

April 2020

Introduction

Buying a home is the biggest investment that most people make in their lifetime. The location is important, but the specific criteria that determine the 'best' location vary person to person. Once a potential homebuyer settles on the factors that are important to their location choice, they face a second dilemma: it is difficult to simply look at a map and visually identify locations that meet or do not meet their criteria. In this study, I develop a decision support system (Albacete et al., 2012: 61) that allows for rigorous consideration of a set of variables relating to housing location in the vicinity of Bonnyville, Alberta.

In my case, to build a house for my family, the important criteria are the following: lots of space (low population), nearness to natural areas, high elevation to avoid flooding, flat and south-facing land for gardening, nearness to lakes and rivers, and clean air and water. Using this decision support system, I identify the 'best' location to build a house based on population density, land use, elevation, slope, aspect, proximity to water bodies, and proximity to industrial facilities that emit pollutants into the environment. I begin by exploring the data sources and discussing suitability analysis, the methodology I use in this study. I then present the model results and finish the report with a discussion of their meaning and applicability.

Data sources

The digital elevation model (DEM) is the GMTED 2010 from the United States Geological Survey (USGS) Earth Explorer (Danielson et al., 2010). Of the USGS's available DEMs, only two cover my study extent, and GMTED 2010 is much more recent and has a finer resolution than the alternative, so I choose it for this study. This dataset is a raster grid storing integer values that represent mean elevation above sea level in meters. The extent of the data is 70.00° N, 120.00° W, 50.00° N, 90.00° W, the cell resolution is 0.0021 degrees, and the coordinate system in which the data is provided is GCS_WGS_1984. See Table 1 for summary statistics.

Per the assignment requirements, I use the GLCC nalucl20.tif land use dataset, also downloaded from the Earth Explorer (USGS, 2018). This raster grid stores nominal data as an integer value associated with each cell. Each value corresponds to a type of land use (USGS, 2018: 16). The extent of the data is 90.00° N, 180° W, 1.67° S, 180.00° E, and the cell

resolution is 0.020 degrees. Note that the dataset is provided in the Lambert Azimuthal Equal Area Sphere_19 projection, but the extent and resolution values I give are measured after re-projection to the GCS_WGS_1984 coordinate system. See Table 2 for the list of land use categories that are represented in the study extent.

I use a population raster from WorldPop (CIESIN, 2018), the floating point value of each cell representing the number of people living in the cell. The extent of the data is 83.14° N, 141.00° W, 41.68° N, 52.62° W, the cell resolution is 0.00083 degrees, and the raster is provided in the GCS_WGS_1984 coordinate system. See Table 1 for summary statistics.

The land use dataset includes a 'water bodies' category, but since that dataset's resolution is low and proximity to water bodies is an important factor for this study, I use an additional water bodies dataset (Statistics Canada, 2011) to increase the resolution and precision of my analysis. The data are vector polygon outlines of water bodies in Canada. The extent is 83.02° N, 141.00° W, 41.96° N, 52.63° W and Statistics Canada provides it in the GCS North American 1983 coordinate system.

I use 2017 data from the National Pollutant Release Inventory (NPRI) to identify sites that emit harmful substances (NPRI, 2017). All records in the dataset represent facility locations that release at least one of more than 350 substances identified by the NPRI as dangerous to the environment or human health. The NPRI makes the dataset available by year and releases it in a table, with latitude and longitude for each site. I include only those points within my study extent, so the extent of this dataset is 55.07° N, 112.11° W, 53.37° N, 109.78° W. I project it to the GCS_WGS_1984 coordinate system, and there are 103 points in the study extent.

Methods

Using spatial models to find optimum locations is referred to as "suitability modeling" and is a type of "process modeling" (McCoy, 2002: 57). In this section I outline the GIS procedure I use to complete this study.

Defining objectives and setting up the data

My goal is to find the "best" location to build a house based on a set of criteria. See Table 4 for my objectives and associated data concerning site choice.

To address these objectives, I begin by loading my data into ArcMap. Some of the datasets are extremely large, so I define my study extent to clip out only the data that is relevant to my area of interest. I aim to identify a housing location within 50 km of Bonnyville, Alberta, and so I create a 100 km buffer around Bonnyville Town Hall. I use the double-size buffer to avoid edge effects in my analysis. After projecting all datasets into GCS_WGS_1984, I clip the DEM, population, and land use datasets to the 100-km buffer right away. The water bodies and pollution datasets are not in raster format, so before I clip them, I need to do some

pre-processing. I first project the water bodies dataset into GCS_WGS_1984, then clip it to the study extent, and finally process it using the Polygon to Raster tool. I then open the .csv file representing pollution points into Arcmap, choose the "show XY" option, and clip the resulting point dataset before using Polygon to Raster to produce a raster version.

At this point I have five raster datasets, with varying cell sizes. I use the Resample tool to standardize all dataset resolutions to match the population grid (the finest resolution). Though this does not refine the richness of the datasets with larger cell sizes, it does allow me to take full advantage of the finer data resolution of the population dataset.

Processing the data: slope, aspect, and distance

To calculate slope, I first project the DEM into NAD 1983 2011 UTM Zone 12N projection so that its linear unit is meters instead of degrees. This step allows me to avoid using the z-factor in the Slope tool as an approximator. The output of the Slope tool is a raster grid storing the angle of each cell's slope, a focal operation. I discard the projected DEM at this point.

To calculate aspect, I use the original DEM as input for the Aspect tool, a focal operation. The output is a raster grid storing the direction that the slope faces at each point in the study extent.

To calculate distance to natural areas, I first identify the land use categories that I consider "natural areas": grassland, shrubland, deciduous broadleaf forest, evergreen needleleaf forest, mixed forest, and wooded tundra. Using the Raster Calculator tool, I produce a grid that stores a value of 1 if and only if the land use value matches one of these types. I then use this grid as input for the Euclidean Distance tool, the output of which is a raster storing each cell's distance to the nearest natural area, a global operation.

I follow the same procedure for water bodies and pollution sources. At this point, I have a set of seven raster grids representing slope, aspect, elevation, population, distance to natural areas, distance to water bodies, and distance to pollution sources.

Reclassification

In order to compare all of these datasets and identify the optimal housing location, I reclassify each dataset into a scale from 1 to 10, with 1 being the least desirable, and 10 being the most desirable. This is a "synthetic" scale that does not have any inherent meaning, but allows me to create "a subjective measure" of the suitability of each location (McCoy, 2002: 64). In other words, it allows me to decide which factors are more important than others.

I achieve this by using the Reclassify tool to bin the values in each dataset. For all the datasets except aspect, I use Jenks natural breaks to define the ten categories, and for the aspect dataset I assign values manually, with north having the lowest value, south having the highest, and the other directions increasing on a spectrum between.

Now all my grids represent the data on the same comparable scale. To decide which datasets are relatively more and less important, I use the Raster Calculator to multiply each grid by a percentage, and add them up. See Table 3 for the breakdown of the weightings. The output of this step is the final suitability raster, where the value of each cell represents how suitable it is as a location for a house based on my subjective criteria, categories, and weighting.

Finally, to identify the 'best' housing location, I only need to find the raster cell with the highest suitability value that is within the 50 km buffer of Bonnyville Town Hall. Because the suitability raster is floating point, no attribute table is available, so I convert it to points using the Raster to Point tool. I then clip the dataset to the 50 km buffer and identify the highest value in the suitability dataset. This is the 'best' location to build a house, based on my subjective criteria.

Results

The 'best' location to build a house within 50 km of Bonnyville is a few kilometres north-east of St. Paul, near a small body of water called Flat Lake. My weighting scheme achieves a compromise between all the input factors: though the identified location is fairly close to a population centre and is not close to any forests, it is fairly flat, fairly high up, south-facing, near several grassland areas, near a lake, and far from most pollution point sources. Figure 1 shows the identified location and the aspect raster; Figure 2 shows the identified location and the land use raster; and Figure 3 shows the identified location, pollution point sources, and water bodies.

Discussion and conclusions

Not everyone would agree that this is the best location to build a house near Bonnyville. However, for a person who values gardening, avoiding flooding, having lots of space, proximity to water and natural areas, and clean air and water, this GIS procedure has identified a reasonably suitable location in a rigorous analytical manner.

There are some limitations to consider. The data that I chose to include reflected my subjective categories that I consider important in selecting a housing location. However, had I chosen others, the results would certainly have been different. Examples of other data I could have included are:

- current land ownership: the 'best' location may be very difficult to acquire for building a house;
- indigenous land title: it may not even be possible to buy the land;

- Any degree of nuance relating to the pollution dataset. This analysis treats all NPRI points as equally dangerous, and assumes that pollution levels are inversely related to distance. In reality, some of these sites may be quite safe and others may be extremely dangerous, and the distribution of pollution is unlikely to be uniform in all directions, rather being more related to watersheds and wind direction.

A different weighting scheme would equally have produced different results. I weight population as very unimportant, only accounting for 2% of the suitability measure, and as a result the site may be closer to a population centre than I would like. To correct this, I could redo the model and give population a greater weight.

Conclusions

These limitations must be given due consideration, but in general this suitability model is robust and can help any prospective homebuyer identify a location that is suitable for gardening, is unlikely to flood, has lots of space, is close to water and natural areas, and has clean air and water.

Appendix 1. Maps

Figure 1. Aspect map and ideal housing location within a 50 km radius of Bonnyville, Alberta

Aspect map and ideal housing location within a 50 km radius of Bonnyville, Alberta



Geographic coordinate system: Mercator Auxiliary Sphere

Data sources

DEM: Danielson, J.J., and Gesch, D.B., 2011, Global multi-resolution terrain elevation data 2010 (GMTED2010): U.S. Geological Survey Open-File Report 2011-1073, 26

Zach Ripka
Department of Geography
McGill University
April 2020

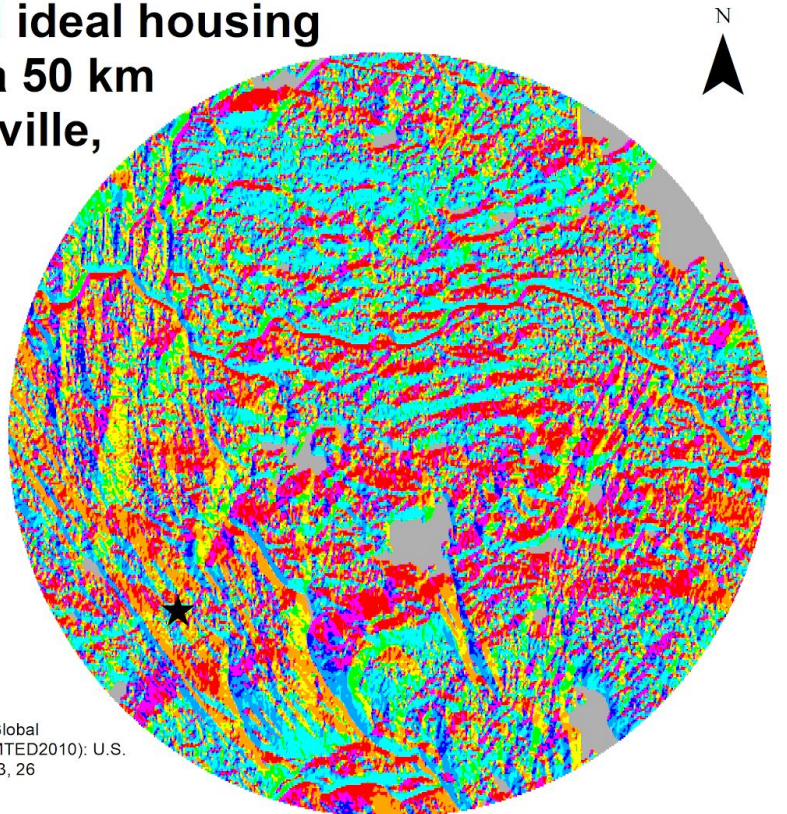


Figure 2. Land use map and ideal housing location within a 50 km radius of Bonnyville, Alberta

Land use map and ideal housing location within a 50 km radius of Bonnyville, Alberta

Land use categories

- Urban
- Dryland cropland and pasture
- Irrigated cropland and pasture
- Cropland/grassland mosaic
- Cropland/woodland mosaic
- Grassland
- Shrubland
- Deciduous broadleaf forest
- Evergreen needleleaf forest
- Mixed forest
- Water bodies
- Wooded tundra
- ★ Ideal housing location

Geographic coordinate system: Mercator Auxiliary Sphere

Data sources

DEM: USGS EROS Archive. (2018). Global Land Cover Characterization (GLCC)

Zach Ripka Department of Geography
April 2020 McGill University

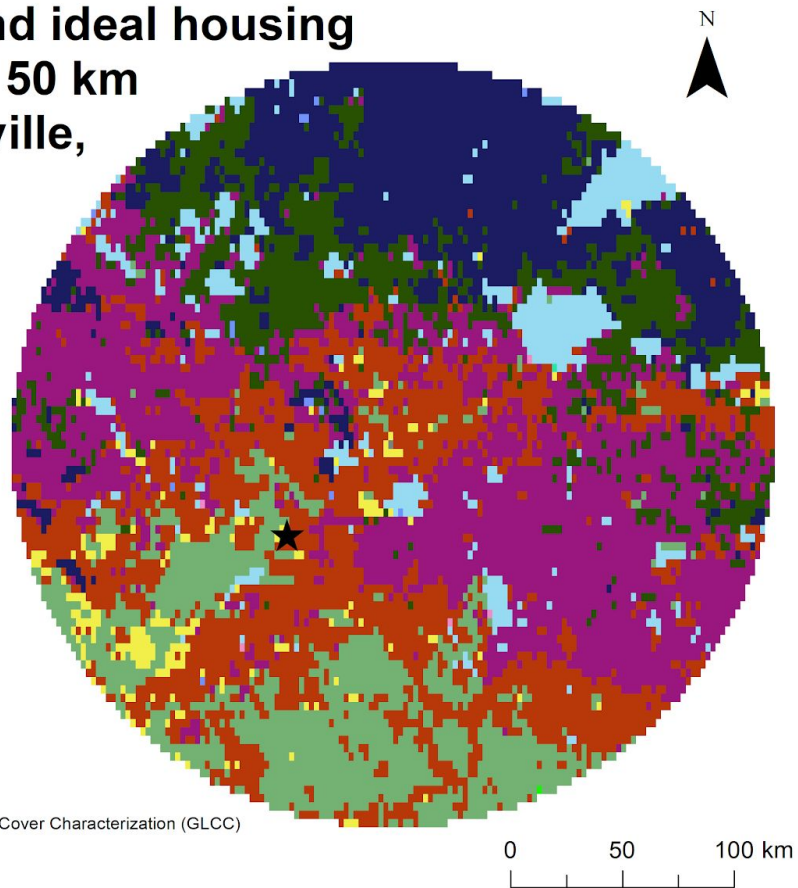





Figure 3. Water bodies, NPRI-reporting pollution sites, and ideal housing location within a 50 km radius of Bonnyville, Alberta

Water bodies, NPRI-reporting pollution sites, and ideal housing location within a 50 km radius of Bonnyville, Alberta

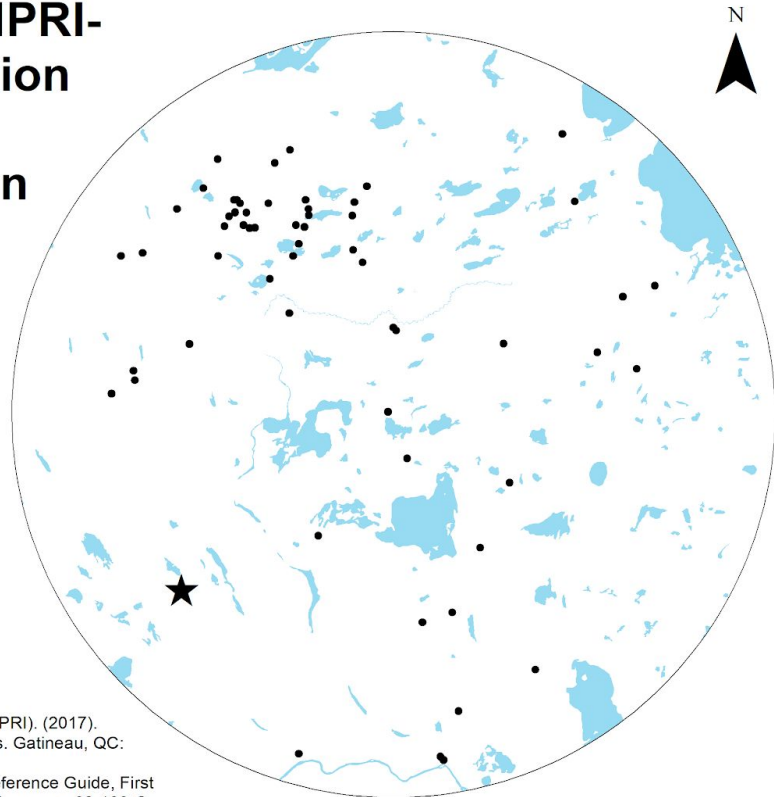
-  Water
-  Pollution sites
-  Ideal housing location

Geographic coordinate system: Mercator Auxiliary Sphere

Data sources

DEM: National Pollutant Release Inventory (NPRI). (2017). Geographic Distribution of Reporting Facilities. Gatineau, QC: Environment and Climate Change Canada.
Statistics Canada. (2011). Boundary Files, Reference Guide, First edition, 2011 Census. Statistics Canada Catalogue no. 92-160-G.

Zach Ripka Department of Geography
April 2020 McGill University



0 25 50 km

Appendix 2. Tables

Table 1 - Summary statistics (limited to study extent)

Dataset	Minimum	Maximum	Mean
DEM	488 m	824 m	597 m
population	0 people	62 people	0.03 people

Table 2 - Land use categories represented in the study extent (USGS, 2018: 16)

Value	Description
1	Urban and built-up land
2	Dryland cropland and pasture
3	Irrigated cropland and pasture
5	Cropland/grassland mosaic
6	Cropland/woodland mosaic
7	Grassland
8	Shrubland
11	Deciduous broadleaf forest
14	Evergreen needleleaf forest
15	Mixed forest
16	Water bodies
21	Wooded tundra

Table 3 -Relative dataset weights

Dataset	Weight
Population density	2%
Elevation	8%
Slope	10%
Aspect	10%
Distance to natural areas	20%
Distance to water	30%
Distance to pollution sources	20%

Table 4. Study objectives: desired location characteristics

Characteristic	Associated data
Flat for gardening	Slope
South-facing for gardening	Aspect
High up to avoid flooding	Elevation
Lots of space	Population density
Near natural areas for walks and hiking	Distance to natural areas
Near water for spending time on the beach, skating in the winter	Distance to water bodies
Far from pollution sources like mines and factories	Distance to pollution sources

Bibliography

Albacete, X., Pasanen, K., Kolehmainen, M. (2012) A GIS-based method for the selection of the location of residence, *Geo-spatial Information Science*, 15:1, 61-66.

Center for International Earth Science Information Network (CIESIN), Columbia University (2018). Global High Resolution Population Denominators Project - Funded by The Bill and Melinda Gates Foundation (OPP1134076). <https://dx.doi.org/10.5258/SOTON/WP00660>

Danielson, J.J., and Gesch, D.B., 2011, Global multi-resolution terrain elevation data 2010 (GMTED2010): U.S. Geological Survey Open-File Report 2011–1073, 26 p
<http://pubs.usgs.gov/of/2011/1073/>

McCoy, J., Johnston, K., Kopp, S., Borup, B., Willison, J., Payne, B. (2002) *ArcGIS 9: Using ArcGIS Spatial Analyst*.

National Pollutant Release Inventory (NPRI). (2017). *Geographic Distribution of Reporting Facilities*. Gatineau, QC: Environment and Climate Change Canada.

Statistics Canada. (2011). *Boundary Files, Reference Guide, First edition, 2011 Census*. Statistics Canada Catalogue no. 92-160-G.

USGS EROS Archive. (2018). *Global Land Cover Characterization (GLCC)* doi:
<https://doi.org/10.5066/F7GB230D>

USGS (2018). *Global Land Cover Characteristics Data Base Version 2.0*.