

Lab Report

Title: Alternative Fuel Station Location Optimization in the Twin Cities Metropolitan Area

Notice: Dr. Bryan Runck

Author: Luke Zaruba

Date: September 28, 2022

Time Spent: 3.0

Abstract

With renewable energy and sustainable transportation becoming key focuses for urban planning and development in the coming decades, infrastructure for supplying sustainable alternative fuels for transportation will become a key issue, especially given the United States' reliance upon gas and oil, along with the country's aging energy system. In order to efficiently and effectively provide adequate alternative fueling infrastructure to the nation, GIS and spatial data science techniques can be used to optimize coverage and minimize costs. This project aims to use two prominent methods developed in the field of operations research, for optimizing the rollout of alternative fueling stations in the Twin Cities Metropolitan Area (TCMA) by maximizing coverage and minimizing resources. The project will show how the techniques can be used at a local or regional level, but the analysis is scalable, and can be used at much smaller scales (larger extents), like across the U.S. Interstate System.

Problem Statement

The task at hand is to optimize new alternative fuel infrastructure placement in order to maximize the coverage of potential customers. As new infrastructure is being planned and developed, it should be a priority to create continuous coverage along major road networks, before beginning to implement alternative fuelling into a broader range of gas stations. By optimizing the placement of the infrastructure, fuelling can be implemented as seamlessly as possible into our changing energy system.

Table 1. Requirements for the analysis

#	Requirement	Defined As	(Spatial) Data	Attribute Data	Dataset	Preparation
1	Current Alternative Fuel Stations	Raw input dataset U.S. Department of Energy AFDC	Point geometry	Features, operator, etc.	AFDC	ETL pipeline for big dataset
2	Road Network	TCMA Roads Dataset	Road geometry	Route system	MN Geospatial Commons	Filtering by route system
3	Current Gas Stations	Gas Station results from Google Places API	Point geometry	None	Google Places API	Combining multiple searches and cleaning up duplicates
4	Potential Customers	Randomly generated points along the road network to simulate customers (weighted count by AADT or set at standard distance interval?)	Point geometry	None	Created on own	Point creation and cleaning

Input Data

The data needed in the analysis is comprised of four main components, a network, existing infrastructure, potential locations, and potential customers. The network will be a simple road dataset, likely filtered to only include interstates, and highways, since these are the primary roads that will provide the best coverage. Existing infrastructure will be extracted from the U.S. Department of Energy Alternative Fuels Data Center's API, which contains all alternative fuel stations in the United States and Canada. Potential locations will be determined by using the Google Places API to search for gas stations. Multiple searches will likely need to be performed, and then the data will need to be merged, before finally removing duplicates and creating the final dataset. The easiest way to iteratively perform searches with the Google Places API will be to create a grid pattern that the latitude and longitude parameters can follow and use in the many iterations of the search. Lastly, potential customers will be simulated by creating points along the road network. There are two approaches to accomplish this. The first being to randomly generate points, and the second being to generate points at a set distance interval. The latter is the more simplistic approach and is arguably the best for analyzing network coverage.

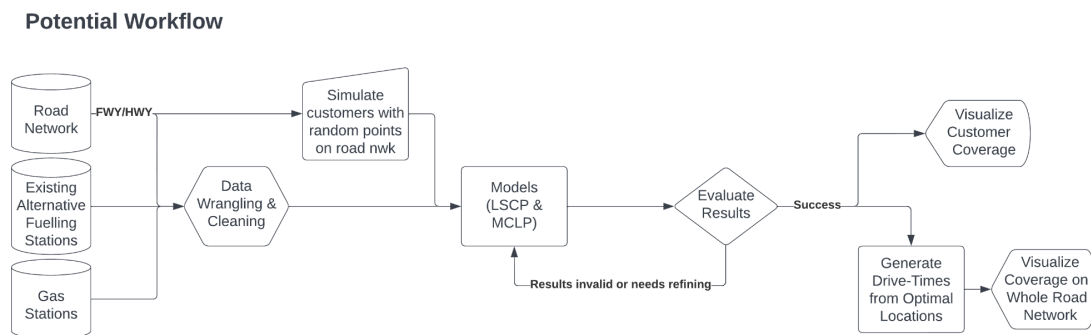
Table 2. Primary datasets that will be used in the analysis

#	Title	Purpose in Analysis	Link to Source
1	TCMA Roads	Dataset for network analysis from Met Council	MN Geospatial Commons
2	Alternative Fuel Stations	Dataset showing existing alternative fuels infrastructure	AFDC
3	Gas Stations	Potential locations for new alternative fuels infrastructure	Google Places API
4	Potential Customers	Simulated customers along road network	Generated in analysis

Methods

There are several potential methods to solve the problem at hand, but two will be focused on, as potential solutions. The first method that can solve the problem is the Location Set Covering Problem (Barcelos et al. 2020a), first proposed by Constantine Toregas (Toregas et al. 1971). The approach seeks to minimize the number of facilities needed to cover all areas within a maximum allowable measure of cost, like time or distance (Barcelos et al. 2020a). With this method, the number of facilities is not fixed, and 100% service coverage is guaranteed. The other approach to solve the problem is the Maximal Coverage Location Problem (MCLP), which was proposed in 1974 by Church and ReVelle (Church and ReVelle 1974). The goal of this method is not to guarantee 100% service coverage, but to maximize coverage with a set number of facilities, based on the fact that resources are often limited and constrained to a budget (Barcelos et al. 2020b).

Figure 1. Data flow diagram.



Results

The expected results of the analysis include two key datasets, which are the two respective outputs from the LSCP and MCLP models. Additionally, several maps displaying customer coverage and drive-time coverage of the entire road network (including local roads) will be produced to visualize the success, or lack thereof, of the models.

Additionally, it is my hope that another result of the analysis is the workflow itself. The workflow could be applied to many other areas across the country, or even other countries, as well. The workflow is also able to be scaled up and down, meaning that it can be performed in neighborhoods or across an entire country.

Discussion and Conclusion

With GIS, it is oftentimes easy to “click-and-go” without understanding the underlying concepts of the models that are used in the analysis. Through this project, it is my hope that I can develop a fundamental understanding of spatial optimization models and apply them to a real-world problem that has great implications for our society’s future. Furthermore, I think that the project will aid me in developing other spatial data science pipelines in other projects that I may work on in the future.

References

- Barcellos, Germano, James D. Gaboardi, Levi J. Wolf, and Qunshan Zhao. “Location Set Covering Problem (LSCP).” PySAL. 2020a. [https://pysal.org/spopt/notebooks/lscp.html#Location-Set-Covering-Problem-\(LSCP\)](https://pysal.org/spopt/notebooks/lscp.html#Location-Set-Covering-Problem-(LSCP)).
- Barcellos, Germano, James D. Gaboardi, Levi J. Wolf, and Qunshan Zhao. “Maximal Coverage Location Problem.” PySAL. 2020b. <https://pysal.org/spopt/notebooks/mclp.html>.
- Church, Richard, and Charles ReVelle. "The maximal covering location problem." In Papers of the regional science association, vol. 32, no. 1, pp. 101-118. Springer-Verlag, 1974.
- Toregas, Constantine, Ralph Swain, Charles ReVelle, and Lawrence Bergman. "The location of emergency service facilities." Operations research 19, no. 6 (1971): 1363-1373.

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? (N/A)
		100	100