Geospatial Multicriteria Decision Analysis – Evaluating Potential LLM Data Center Sites for Climate Friendliness

Dominik Neumann, Celina Thomé

March 31, 2024

1 Introduction

Data storage and computing has emerged as one of the most rapidly growing industries globally. Especially hosting Large Language Models (LLMs) requires a significant amount of energy, cooling and maintenance. An average request to a LLM consumes up to 60 times more energy than a request to a common search engine (de Vries, 2023). Hence, especially when taking the climate crisis into consideration, it is necessary to evaluate the climate friendliness of potential locations for data centers which are hosting LLMs.

Apart from other ethical concerns like overall necessity and usage, the site selection plays an important role in the climate friendliness potential. In this analysis we provide a proof of concept for our Multicriteria Decision Analysis (MCDA) evaluating the German state of Baden-Württemberg for potential site locations. MCDA is a commonly used method to evaluate site locations of any enterprise, as it provides the possibiliy to weight important location factors against each other to determine the best location. However, using this approach in real life scenarios does mostly not encompass a comprehensive geospatial analysis, instead using an economical approach for already preselected regions or cities.

The goal is to provide an automated workflow of our analysis which makes it possible to up- or downscale the analysis fitting to the purpose. As the factor weights can be amended to the user's needs, it is also possible to adapt the analysis for other purposes than data center site selection.

2 Method

2.1 Multicriteria Decision Analysis

MCDA has been used in a variety of cases, especially for site (pre-) selection. In our case we will use a mixed-method approach to have a large scaled analysis of a certain region, in our case the German state of Baden-Württemberg. We have created a grid of 175 hexagons with the radius of 10 km (centroid to corner) to split the region into equally sized units. Each of the hexagon has a centroid for which the majority of calculations have been done.

2.1.1 Factors

Each MCDA requires a set of criteria, as the name suggests. We call them *factors* in our analysis. Each factor has the potential to influence the outcome of the analysis. In a real world scenario, factors are often selected and evaluated by a CEO, board of a company or a consulting business that seeks to evaluate relocation or expansion of their business or business partners. As this is a theoretical analysis we have taken inspiration by scientific literature and common sense to select our factors. We will briefly describe them now.

- Climate: Especially for data centers, environmental surroundings play an important role. Temperature and humidity heavily influence the performance of servers. A higher temperature might require more cooling and therefore more energy consumption.

- Disaster Risk: Data centers are large structures that require accessibility via roads, but also need electricity and water for e.g. cooling. In disaster prone areas, constant access to those might not be given at all times.
- Infrastructure availability: as mentioned above, electricity and water need to be easily available at all times. Additionally, the power source plays an important role for our analysis, as climate friendliness is key for us. Telecom availability is equally as important.
- Conservation efforts: When there are natural conservation efforts taking place in a location, laws and ethical concerns might prohibit a location from being climate or environmentally friendly.
- Crime: A so-called soft location factor that relates to the general safety of the facility and its employees.
- Employee availability: For construction and maintenance alike, workers and employees are necessary. Hence, it is also important to consider the availability of all talent.

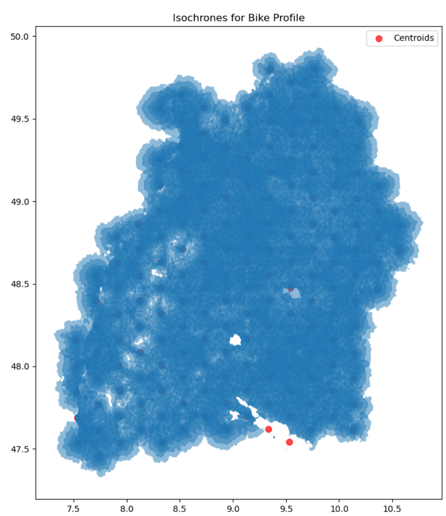

Figure 1: Reachability isochrones via bike for multiple time frames.

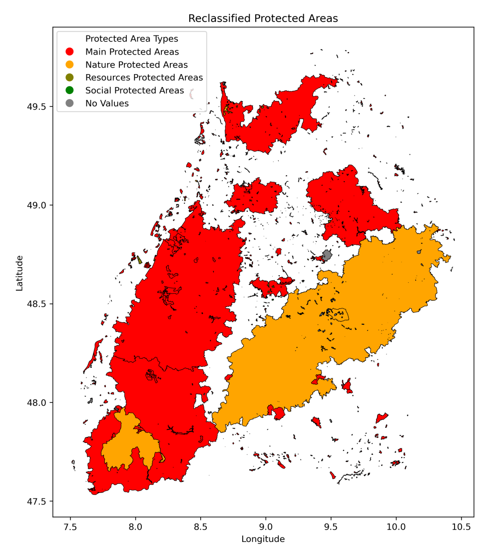

Figure 2: Reclassified protected areas with assigned fitness value.

Of course, all factors can be changed, given there is enough data available for a comprehensive analysis, as is not the case for every type of data. For example, it proved rather difficult to include very important factors like taxes, land and labour cost or availability of public transport options. Those factors often play an important role in real life scenarios, but are not included in our analysis, due to a very high complexity of the available data or little data availability.

2.2 Classification

In MCDA, before weighting the factors, they have to be assigned a class. In our case each hexagon or its centroid is assigned a to a class which is assigned a value (see Fig. 3). The class value between 0 and 10 is typically assigned after comparing all values of the same factor and taking additional scientific evaluation into consideration. For example, for the factor *climate* we have reclassified all sub classes into their main class which was already given by the Köppen-Geiger-Classification. This means instead of the 30 sub classes like *tropical monsoon climate* and *tropical rainforest climate*, which were provided by the data, our new class *tropical climate* is now assigned a value according to its fitness for our purpose. In this case, due to the high humidity and temperature in tropical climates it is not very well suited and receives the value 3.

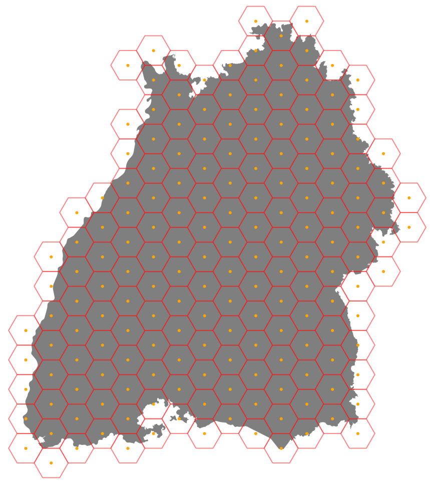

Figure 3: Hexagons and their centroids across Baden-Württemerg

3 Results: Weighting and the final index

Each classified factor is assigned with a weight, which resembles its importance for the overall fitness indicator. The weights add up to 1. They can be seen in the appendix. Some weights are set to 0, either because the data was incomplete (reachability by car) or it was the same for all of the region, e.g. climate class, crime or natural hazards. It must be noted that we could only attain crime and natural hazards data for the whole of Germany, which subsequently means that all of Baden-Württemberg has the same value as well. The result of the analysis is calculated by weighting the value of each classified factor, adding all weighted values to each other and normalizing the result to represent a full number between 0 and 10. This is done for every hexagon in the region. The result is a map of every hexagon that has been processed. The result shows, that even directly adjacent hexagons can be very differently suited for our case. Of course, this is mostly impacted by the heavily weighted factors like energy source or nature conservation measurements as well as reachability.

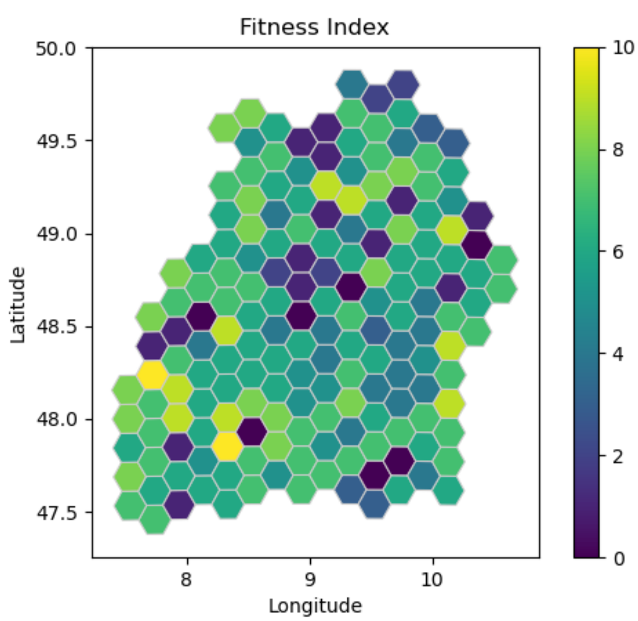

Figure 4: The final Fitness Index for each hexagon.

4 Discussion

The results show a general feasibility of the MCDA for a data center site location. We have successfully evaluated 175 equally sized regions or their centroids in Baden-Württemberg for the climate friendliness when selecting a site. Even differences between adjacent hexagons are clearly shown, depending on the factors. First estimations show, that it is likely that those are influenced by the conservation efforts factor and the power source. Due to the scope of the project we were not yet able to further analyse our results. However, we would like to compare the results with real life data center locations and also cross examine for the impact of the factors per hexagon. There might be common ground across similarly suited areas. Additionally, there are limiting factors to the analysis. Apart from API key restrictions, the computed data is quite large. For more finely grained analyses of the same area an increased computational power is needed, even though the analysis technically allows for more hexagons over the same area as well as same-sized hexagons for a larger area. Additionally, the quality of the data is difficult to evaluate. OpenStreetMap data can be rather complete and up-to-date in some areas of the world, but certainly not everywhere. Data like crime rates or natural disasters vary depending on statistical methods and might have a significant number of unreported cases, be it material damage, injuries or even deaths. The analysis itself proved difficult to execute, as it relied heavily on small steps towards the final result. Hence, there are currently many different scripts, notebooks and data files, which makes the analysis difficult to reproduce at the moment. Overall, we are quite happy with the result of this complicated analysis and look forward to improving it further.

5 Conclusion and Outlook

Generally, the approach proved successful. However, there is still work to be done to streamline the analysis, e.g. refactoring the scripts that were used. The approach bears the potential to be used in a larger scale, be more fine grained and even be adapted to other use cases, like other location analyses. Of course there can be factors added and weights changed. We eagerly await improvements from our side to enhance the repository structure and take our analysis to the next level. The ultimate goal is to embed it into the ohsomeHeX service by HeiGIT which already implemented similar analyses.

6 References

- Daim, T. U., Bhatla, A., & Mansour, M. (2013). Site selection for a data centre–a multi-criteria decision-making model. *International Journal of Sustainable Engineering*, *6*(1), 10-22.

- de Vries, A. (2023). The growing energy footprint of artificial intelligence. *Joule*, *7*(10), 2191-2194.

- Fetanat, A., & Khorasaninejad, E. (2015). A novel hybrid MCDM approach for offshore wind farm site selection: A case study of Iran. *Ocean & Coastal Management*, *109*, 17-28.

- Malczewski, J. (1999) GIS and multicriteria decision analysis.

- Önüt, S., Efendigil, T., & Kara, S. S. (2010). A combined fuzzy MCDM approach for selecting shopping center site: An example from Istanbul, Turkey. *Expert systems with applications*, *37*(3), 1973-1980.

- OpenStreetMap contributors (2024).

- Owen, S. H., & Daskin, M. S. (1998). Strategic facility location: A review. *European journal of operational research*, *111*(3), 423-447.

6.1 Online resources

- Project repository: https://github.com/celthome/MCDA-for-data-center-location

- https://heigit.org/analyzing-bikeability/

- https://giscience.github.io/openrouteservice/

- https://github.com/GIScience/ohsome-api

Factor	weight
Energy source	0.3
Powerplant distance	0.004
Powerline distance	0.001
Telecom line distance	0.05
Water area nearby	0.1
Type of protected area	0.3
Average deaths/y in last 30 years	0.0
Bike 10 min	0.04
Bike 20 min	0.02
Bike 30 min	0.01
Bike 45 min	0.007
Bike 60 min	0.003
Car 10 min	0
Car 20 min	0
Car 30 min	0
Car 45 min	0
Car 60 min	0
Foot 10 min	0.05
Foot 20 min	0.035
Foot 30 min	0.02
Foot 45 min	0.01
Foot 60 min	0.005
Climate Class	0.0

Table 1: Weights for each factor. A weight of 0 indicates incomplete or non-differential data across the region.