Abstract: Nuclear Material Storage Site Selection Using Geo-Cyber Analysis combines facility location and site selection methods to analyze appropriate locations for the storage of nuclear materials, including spent fuel and radioactive waste, under multiple criteria. This project combines traditional techniques in facility location (mathematical modeling, network optimization) with spatial analysis tools and Geographic Information Science (GIS). New nuclear material storage facility locations must balance the location’s potential for cyber exposure with its physical (geographic, environmental) vulnerabilities. It must also meet production and capacity constraints as defined by the production of high-level radioactive waste as a nation and the capacity of a geological repository. Students on this project will work to develop measures for site suitability and facility vulnerability as well as mathematical models for the location of a single nuclear materials storage facility and a set of nuclear materials storage facilities.
# Table of Contents

Introduction ................................................................................................................................................... 3

Initial Research ............................................................................................................................................. 4
  Terms ........................................................................................................................................................ 4
  Constraints and Assumptions .................................................................................................................... 4

Methods ........................................................................................................................................................ 5
  ArcGIS ...................................................................................................................................................... 5
  ArcGIS Methods ....................................................................................................................................... 7

Model ............................................................................................................................................................ 9
  Proof of Concept ..................................................................................................................................... 10
  Genetic Algorithm .................................................................................................................................. 12

Conclusions ............................................................................................................................................... 14
  Summer/Fall Semester Plans .................................................................................................................. 14

Bibliography ............................................................................................................................................... 15
Introduction

Purpose

The purpose of this report is to provide an update on the “Nuclear Material Storage Site Selection Using Geo-Cyber Analysis” research project. The beginning of this report outlines the current status of high-level radioactive waste storage in the U.S. and why a permanent solution for long-term storage is needed. Then, the methods used to create a proof of concept model are discussed in detail. The report concludes with new semester goals as well as future considerations for the project as it continues into its final semester.

Background

One of the greatest technological developments of the 20th century was the discovery of nuclear fission, a process where a large energy surge is created by splitting an atom’s nucleus into smaller parts. Today, nuclear fission is used in nuclear power plants and in 2011 approximately 19.2% of the energy produced in the U.S was from nuclear power plants. However, during the inception of this technology the U.S. used it to create nuclear material for the production of nuclear weapons. Fueled by World War II conflicts, this technology would start a race between the German and American nations to produce the first nuclear weapons. The development of these weapons began in 1939 at sites in Tennessee and New Mexico with nuclear material produced in sites such as the Hanford Nuclear Reservation in Washington State. These two development sites would later become the Oak Ridge and Los Alamos National Laboratories, and the Hanford site is now known as the most contaminated radioactive site in the United States. As of 2007, this site represented two-thirds of the stored high-level radioactive (HLW) waste by volume in the nation. Today 2,000-2,300 metric tons of HLW is produced per year by over 100 currently operational commercial and research nuclear reactors in the United States. Today there is over 70,000 metric tons of HLW being stored in various sites in the United States, and this level is increasing each year. These materials take at least 500,000 years to decompose into non-radioactive elements, with many of them still decomposing after millions of years. In 1982 the U.S. created the Nuclear Waste Policy Act to begin to tackle the problem of storing this waste, and a site at Yucca Mountain, Nevada began being developed as a permanent geological repository. However, construction at this site was recently terminated by the local population. No new sites have been officially approved for long-term disposal since this development even though many short-term disposal projects currently exist.

Scope

This scope of this report includes prior research, new methods, data analysis, a new proof of concept model, and tools to assist with a final prototype to analyze suitable high level radioactive waste facilities. The important terms, assumptions, and accepted community research in the area of high level nuclear waste storage are listed and then data analysis tools proposed. The project methods and tools for model prototyping are described and justified. Finally, future project goals and an updated project timeline are presented for the fall 2013 semester.
Initial Research

A project of this complexity requires some initial assumptions to be made to simplify its complexity. In addition to this, definitions of important terms were required for project clarification. Since this project will ultimately produce a working approximation model, initial constraints were identified for the project as a whole and also to be included in the model as a mathematical constraint.

Terms

To clarify the model and projects assumptions, useful terms were defined. They are listed as follows:

- **Terrorist** – person with intentions to harm the biosphere or spread panic in order to reach a religious or political goal.
- **Storage** – Isolation from the biosphere (humans, animals, plants, aquifers, resources, anything that comes into contact with living organisms).
- **Biologically Safe** – less than 50 rem/per year/per organ
- **Minimized terrorism risks** – terrorists ability to access radioactive waste is mitigated
- **Minimize risks from natural disasters** – facility’s location minimizes the risk of geological and atmospheric hazards
  - Research Community consensus on what is considered a “natural disaster”
- **Security** – the facility and transportation methods are designed so that the likelihood of radioactive waste entering the biosphere is decreased
- **High level radioactive waste** – spent radioactive material that is harmful to the biosphere in terms of radioactivity.
- **Heuristically optimal** – as good as is possible in relation to our model.
- **Human error** – actions made that can violate our model

Constraints and Assumptions

Our first major accomplishment was determining assumptions for our model. These were necessary in order to simplify our project and are separated into modeling and non-modeling constraints. The modeling constraints are ranked with decreasing importance from 1 to 3 and will quantifiable in our model. These constraints are listed as follows:

1. Location of facility and proximity to human populations are the most important aspects – Geographically safe with safe transportation routes and a safe distance from human populations
2. Facility is semi-permanent:
   - 50 year life minimum for fiscal reasons and also to look into research about maintaining facilities/waste for extended periods of time
3. High Level Radioactive Waste production for nuclear reactors in the U.S. are assumed to be the same and constant
The non-modeling constraints are organized into “vital” and “hard to get rid of” categories. The vital assumptions are as follows:

- Assume all laws and regulations are followed – there are too many possibilities otherwise
- We are a group of researchers making a proposal on site location.
- Nuclear waste will be stored in its current form (dry cask) and not recycled. This is the worst case scenario and the current situation in the US

The “hard to get rid of” constraints are ranked with decreasing importance so they can be eliminated or changed in that order. These constraints are listed as follows:

1. Only consider storage of high-level nuclear waste
2. Assume no human error, or that the human error is accounted for at an ignorable level
3. Deal only with storage of HLW in the U.S. because of familiarity with culture and government regulations
4. Terrorists know everything about the facility location because we live in a Democratic society and a facility location would need to be approved by citizens
5. Our model does not take cyber security into consideration

Methods

The project team has been conducting research using basic project management tools. A project plan has been created that defines major project goals and guides the team through the research, implementation, testing, and finalization phases of research. Communication is a key factor to project success, so meetings are held twice a week to identify weekly goals, complete research, and discuss findings. At the beginning of the second semester, the team was split into two small groups, one to work with ArcGIS and the other to work with the model, to maximize the team’s productivity.

Despite the split team objectives, both groups had access to and used similar resources to collect quantitative data. The tools used to acquire this research are listed as follows:

- ArcGIS
- Textbooks
- ISEN Dept. Programs
- Engineering Faculty
- Scholarly Papers
- Census and Transportation Networks
- Financial Analysis Methods
- Public domain data
- Nuclear Regulatory Commission (NRC)

ArcGIS

For the ArcGIS team, two of the most useful quantitative tools were public domain data and the actual ArcGIS program. To assist in explaining how these tools were utilized, NRC criteria for
geological repositories and a glossary of ArcGIS terms are provided below. Then, ArcGIS and the functions that were utilized within it are discussed.

After the Nuclear Waste Policy Act was enacted in 1982, the Department of Energy and related agencies began research on the geological, population, and safety factors that need to be considered when contracting a permanent geological repository. A list of over a dozen factors was considered by the NRC and includes some of the following considerations:

- Population density
- Transportation
- Climate
- Erosion
- Dissolution
- Tectonics
- Human Interface: Natural Resources and Site Ownership and Control
- Surface Characteristics
- Meteorology
- Hydrology
- Geohydrology
- Geochemistry
- Environmental Quality
- Socioeconomics

Terms:

- Geoprocessing - Any operation that creates new data from existing GIS data
- Shapefile - The data file name for geospatial vector data created from GIS software
- Create Fishnet - Creates a fishnet of rectangular cells; At the centroid of each cell a point is created
- Overlay - Information from different layers are combined
- Erase - Type of overlay; Creates a feature class by overlaying the Input Features with the polygons of the Erase Features. Only those portions of the input features falling outside the erase features outside boundaries are copied to the output feature class.
- Intersect - Type of overlay; Computes a geometric intersection of the input features. Features or portions of features which overlap in all layers and/or feature classes will be written to the output feature class.
- Near - Determines the distance from each feature in the input features to the nearest feature in the near features, within the search radius.
- Selection - Identifies/locates a subset of features on your map.

ArcGIS is a geographic information system that is used to analyze the data received from various shapefiles. Included in the program are multiple geoprocessing tools that can be used to restrict the possible locations of nuclear spent fuel repositories. The geoprocessing techniques associated with ArcGIS also allow for the creation of equally spaced data points with multiple attributes relating to seismic activity, soil composition, proximity to water, etc that are considered by the
NRC  The data points created can be exported to an excel file and subsequently imported into the model, which will produce the optimal locations for a repository.

**ArcGIS Methods**

The criteria obtained from the NRC lists the geographical factors that need to be considered when limiting the suitable locations. For example, by using the selection tool, a layer consisting of only the counties with population density greater than 83 people per square mile was created. Then all restricted areas that would bring about potentially adverse conditions, such as high population density counties, underground aquifers, Native American reservations, and natural reserves, are eliminated from the original population density shapefile with the use of the “erase” tool. The layer containing the points from the fishnet is then “[intersected]” with other layers to obtain the associated attributes. Layers to be intersected include: the population density layer with erased areas, seismic activity, and geology. The “near” tool can then be used to find the proximity between the “fishnet” points, and other layers of interest. A “fishnet” creates a web of data points and for the purpose of this project; it is used with the “near” tool to determine the proximity between sources of water and populated areas. After the fishnet layer has undergone its final geoprocess, it can be exported as an excel file for further risk analysis.

In GIS, geospatial data can either be vector or raster. Vector data expresses geographic data in points, lines, and polygons (an enclosed area). A raster data set consists of a matrix of cells organized into rows and columns with each cells having a value that corresponds to a certain attribute, such as population density. It is possible to convert from vector to raster data, however, converting from raster to vector data is much more time consuming.

An advantage of using vector data is that it is represented in its original form; therefore, no data conversion is required. Also, the traditional map representation is easy to look at and understand. Finally, vector data is useful when representing discrete boundaries and can be beneficial due to the accuracy of the data location. Some disadvantages of using vector data include having to store each vertex location explicitly and having to convert the data into a topological structure for effective analysis. There is also a direct proportion between the process time and the amount of analytical data. Furthermore, some data sets are not represented in vector form.

Raster data is beneficial because of its use of cell matrices; each cell’s geographic location is indicated by its point in the matrix. Because no geographic coordinates are stored other than the origin, analysis of the data is easy to program and perform. Raster maps are suitable for mathematical modeling and quantitative analysis. Disadvantages to raster data include the difficulty to represent linear features. Because the cell size controls how the data is represented, it can be difficult to display certain characteristics. Raster maps only display one attribute or characteristic for an area. Also, data must be converted since most input data is in vector form. This leads to longer processing time and concerns of data integrity.

One of the assets of ArcGIS is the ability to incorporate code to retrieve data. Retrieving data can often require iterative steps, and incorporating a script simplifies this process. Therefore, a Python script is being developed to replicate code in state shapefiles. This will save the team a
great amount of time by efficiently selecting feasible raster data to implement the “Fishnet tool”
on. For example, as stated in the ArcGIS methods, to ensure that the location of a potential
nuclear material storage facility is within a county that has less than 83 people per square mile, a
code can be used to utilize the tools in ArcGIS to constrain the output for feasible county land for
the state of Texas. Furthermore, the Python script will be used to repeat the algorithm for all the
other states. This example hints towards the full vision of (1) implementing Python scripts for
each state shapefile, and (2) combining the output for all the feasible “Fishnet points” into one
Excel file. This data will be employed in the complete model. This will greatly help the
simultaneous run of ArcGIS tools with simple Python scripts.

Figure 1. ArcGIS screen shot of points that represent suitable locations for a nuclear waste
repository. In this figure, no points overlap Native American reservations, aquifers, or counties
consisting of population densities greater than 83 per square mile.
Model

Geoprocessing techniques only eliminate a certain number of potential locations, therefore, a model is needed to determine an “optimal” location for a nuclear waste repository. The model chosen is based off the p-median problem assuming that high level nuclear repositories are un-capacitated and production of high level nuclear waste from nuclear facilities is based off of net megawatts of electrical energy produced by each facility. This type of model is a facility location problem that locates P facilities in relation to customers so that the shortest distance is chosen between facility and repository. This minimizes the product of the weighted Euclidian distance between repositories and facilities and facility production. It is an NP hard problem that requires a heuristic to be solved because as the problem size increases, the time required to solve the problem exponentially increases.

The objective function is as follows:

\[
\text{Minimize Weighted Average (Z)} = \sum_i \sum_j d_{ij} h_i y_{ij}
\]

- \(d_{ij}\) – Euclidian distance between nuclear facility i to high level nuclear waste repository j
- \(h_i\) – Amount of high level nuclear waste produced at facility i
- \(y_{ij}\) – 1 if high level nuclear waste from facility i is satisfied by repository j, 0 otherwise
- \(x_j\) – 1 if repository j is used, 0 otherwise

In this model the following initial constraints and parameters are used:

**CONSTRAINTS**

1. \(\sum_j x_j = P\) Requires that exactly P facilities are located
2. \(\sum_j y_{ij} = 1 \quad \forall \ i, \ i \in I\) Requires that each facility is assigned to only one repository
3. \(y_{ij} \leq x_j \quad \forall \ i,j\) Requires that a repository exist before it is assigned a nuclear facility
4. \(x_i, y_{ij} = 0,1 \quad \forall \ i,j\) Binary requirement
5. \(i \in I, j \in J\) States that the elements i and j are encompassed by sets I and J

**PARAMETERS**

- \(P\) = number of high level nuclear waste repositories to be located
- \(I\) = set of nuclear facilities
- \(J\) = set of candidate locations for high level nuclear waste repositories
Proof of Concept

In order to confidently move forward with this model we created a basic model using excel solver to ensure that the results would emulate our goals. The constraints and objective function formulae were defined and then testing began using randomly generated potential reactor locations. In order to validate the model some simple tests with known results were evaluated. For example, if there is a single reactor the model should result in a waste repository exactly on top of that reactor. When the model was expanded to consider multiple repository locations it became simple to show that if the number of repositories is greater than or equal to the number of reactors then the objective function will have a value of zero associated with transportation risks. This is because a repository can be located at each reactor. This could change once a repository capacity is implemented and should be considered in future development.

One simple implication of this model is that the ideal location would be located inside the smallest convex shape that encompasses all the reactor points. This can be shown mathematically by noting that any repository point located outside of this would benefit the objective function if it were moved to an edge or vertex of that convex shape.

Once the model was validated the ArcGIS team supplied us with the latitude and longitude of each reactor in the United States. By including the energy production we were able to estimate the amount of spent fuel produced and assign heavier weights to the reactors that produced more waste. By transposing the latitude and longitude to ensure only positive number for our calculations we were able to run the proof of concept model using real data. Examples of some of the results are illustrated below.

![Optimal Waste Site Locations](image)

**Figure 2.** An example of where the model would locate four repositories given the transposed coordinates. An outline of the USA can be seen in the blue points.
Euclidian distance is a simple, easy to calculate method of measurement that can be determined using only the latitudes and longitudes of the reactor locations. This simplifies the proof of concept model by making the calculations quicker and the data easier to obtain. However, there are downsides to using this method as the high level radioactive waste cannot always travel in a straight line. Road networks, train networks, or any other shipping methods follow different routes and thus the estimates are not entirely accurate. As the model is expanded next semester another method called the Minkowski distance metric will be used to more realistically approximate real road networks. This method will keep calculations and the necessary data simple but will still produce a reliable estimate of the transportation risks. Furthermore, this model does not account for the curvature of the earth. The influence of this is minimal on our results, but we will attempt to account for it next semester by using the Haversine formula, which is common in navigation and gives the great circle distance between two points on a sphere. We will use the World Geodetic System (WGS) 1984 Earth ellipsoid in order to match our ArcGIS projections as all of our data uses this projection.

In addition, we found while testing the model that there are diminishing returns in terms of transportation safety as the number of repositories increase. While each new facility does decrease the objective function, realistically the cost of building a new facility would need to be considered in comparison to how much safer the overall system will be. The efficient frontier can be seen below:

![Efficient Frontier](image)

**Figure 3.** The efficient frontier of the model as more waste sites are considered. The goal is to minimize the objective function, but each new facility does not produce the same benefit as the previous facility. This should be taken into account when considering cost.
Genetic Algorithm

A Genetic algorithm is a heuristic algorithm based on the principle of survival of the fittest. In a natural evolutionary process, how much a creature fits the environment depends on its genetic makeup. In a population, new genes are generated by mutation and crossover. Because both mutations and crossovers are omnidirectional, some new genes would be less fit, while some others would be fit. Those with good genotypes have a larger probability of surviving so that they have better chances to produce offspring. Individuals with bad fitness are eliminated, while those with good fitness survive and reproduce. Gradually, the whole population has better fitness to the environment and the population is optimized by natural selection.

In a genetic algorithm, independent variables represent genotypes and dependent variables represent phenotypes, which are decided by genotypes. Independent variables are transformed into binary numbers, where 0 and 1 stand for two different forms of a gene (allele). The use of a binary system makes evolutionary operations like mutation (a gene changes from 0 to 1 or from 1 to 0) and crossover (two different individuals exchange parts of their genes) easily computated. The target function, which is to be optimized, is made into a fitness function. An individual’s probability of surviving is determined by this fitness function so that all the solutions (individuals) evolve toward a pre-specified.

The main steps in a genetic algorithm are

- **Encoding**: Based on the boundaries and precision requirements, determine the length of genes. Transform the value of independent variables into binary codes (0 and 1).
- **Decoding and Fitness Evaluation**: Transform binary numbers into decimal numbers to evaluate their fitness values.
- **Selection**: Use fitness values and roulette algorithm to decide which individuals reproduce and which individuals are eliminated.
- **Mutation**: Assign mutation probability. Still use roulette algorithm to decide which genes mutate.
- **Crossover**: Assign the crossover probability. Use the probability and roulette algorithm to decide which individuals would exchange their genes with others. Use random number generator to decide which part of their genes would be exchanged.
- **Repeating**: Repeat the steps above until the assigned generation.

The reasons a genetic algorithm was chosen for optimization are as follows:

- **Heuristic algorithms** are good at dealing with large-scale computations and can reduce computation time. When the computational load is small, violent search algorithms can easily find an optimal solution with high precision. But even in the simplest p-median model, as the problem size increases, the time needed by violent search algorithm exponentially increases. In order to reduce the computation time back to normal, we would need a heuristic algorithm.
- **Genetic algorithms** have excellent robustness. Although analytical methods can provide a precise solution, different problems would require different analytical method and not all problems have optimal solutions. A genetic algorithm can be used to deal with many
different types of problems while the users need only to adjust the fitness function and the parameters.

- Genetic algorithms have strong global search abilities. There are many other heuristic algorithms we can choose, and some of them have better efficiency (computing speed). But those algorithms’ global search abilities are not as good as genetic algorithm; they have a larger chance to terminate at local optimal solutions. A genetic algorithm can keep its evolutionary ability to the end because there are always new genes generated by mutation and crossover.

However, due to the disadvantages of genetic algorithms, we must also pay attention to the points listed below:

- Genetic algorithms are sensitive to parameters. If the mutation probability or crossover probability are set too large, convergence would be very slow (sometimes the population may even not converge); if they are set too small, the ability to generate new genes is weak and soon all the individuals would become the same so that the population would lose its evolutionary ability (named premature phenomenon).

- Despite the excellent global search ability, genetic algorithms have worse local search ability compared with other heuristic algorithms. Due to Hamming cliff (the number of mutations needed to change one gene to another is named Hamming distance and the phenomenon that Hamming distance between two near numbers, like 1 and 2, is usually not 1 is Hamming cliff), it is hard for an individual to mutate into a near number.

We used the target function to test the global search ability of genetic algorithm.

\[
f(x_1, x_2) = 2\pi e^{-0.2\left(\frac{x_1^2 + x_2^2}{2}\right)} + e^{\cos(2\pi x_1) + \cos(2\pi x_2)} - 2\pi
\]

The reasons this function is used are:

1. We know the theoretical maximum of the function is e, so we can compare the maximum we get with this value;
2. This function has many local maximum points. Without excellent global search ability, the algorithm would return a value far from the theoretical one.

Using the codes in Appendix, we get 2.70466 as the maximum. The relative error is 0.5%, suggesting the algorithm’s ability to avoid local maximum values.
Conclusions

Summer/Fall Semester Plans

The Spring 2013 semester ended on April 23rd with the final report due date being May 6th. The following objectives were met this semester:

- Displayed understanding of nuclear waste management techniques at the Spring 2013 engineering research showcase
- Created a poster that summarized major research and was peer reviewed
- Utilized the ArcGIS database to refine feasible locations
- Developed a working proof of concept model
- Developed a Matlab genetic algorithm code
- Developed a python script to simplify the data retrieval process

The goals for the Fall 2013 semester are listed as follows:

- Create a graphic user interface tool that determines an optimal location for storing high level nuclear waste
- Move from proof of concept model to Matlab Genetic Algorithm
- Expand Model:
  - Include capacity and production considerations
  - Move from regional (Texas) to national scale (United States)
  - Add more constraints
- Sensitivity analysis
  - Determine which variables in the objective function affect the final location the most
- Begin removing assumptions
- Write out scales for seismic activity, distance to water, etc.
- Write out an “elevator speech” to help prepare for the Fall 2013 engineering showcase

Initial tasks for the Fall 2013 semester will begin with the targeting the following objectives:

1. Determine new meeting times
2. Implement a graphic user interface with python script
3. Add scales and constraints to the proof of concept model and convert it to a Matlab code

As always, we will continue to seek peer evaluation of our resource and update our master plan as our objectives and project scope change. Considerable progress was made this semester in ArcGIS and model components. While there was no “optimal” location determined this semester, the work completed lays a foundation to easily compute this location in the Fall. Thus, the main concern for next semester is to begin consolidating the different resources compiled throughout the Spring into a user friendly interface that computes an “optimal” location based on varying (and easily changeable) criteria.
Bibliography


<http://en.wikipedia.org/wiki/Nuclear_weapons_and_the_United_States>.


<https://www.census.gov/geo/maps-data/data/tiger-line.html>


