For office use only	Team Control Number	For office use only
T1 ________	**IMMC23618025**	F1 ________
T2 ________		F2 ________
T3 ________	Problem Chosen	F3 ________
T4 ________		F4 ________

2023

(IMMC)

Summary Sheet

Final Plan for Three Square Kilometers of Land

Summary

Determining the use of a piece of land is a challenging task. People need to consider different aspects, such as commercial and environmental aspects. The efficiency of the use of land is influenced by many factors, including geography, climate, etc. A $3km^2$ land in Wayne County is considered to be developed. To determine the best plan of land use, we need to balance negative impacts and positive impacts of the plan, and take into account the unique characteristics of the region.

We first develop a metric IUL to evaluate our plan. We consider three aspects: environmental impacts, business profit, and the community utility, corresponding to the meric of IE, IB, and IC. Each aspect contains multiple indicators. After determining all the indicators for each aspect, we use the **double-level Entropy Weight Method** to obtain the weights of indicators in each aspect, as well as the weights of IE, IB, and IC.

Next, we build prediction models for specific land uses' indicators. The four land uses we consider are: agritourist center, solar array, sports complex, and ranch. We separate them into **Demand-Restricted Land Uses** and **Demand-UnRestricted Land Uses** to build the **prediction models**. For the Demand-restricted Land Uses, we build the **Average Demand Model** by considering a fifty kilometer radius of local area and divide the circular area into grids. We calculate the probability of people in each grid going to different sport complexes or agritourist centers, and obtain the annual total demand of tickets for the specific place. The demand is directly related to many of the indicators including running cost and total revenue. By connecting all the prediction models of indicators with the metric, we construct a rigorous system and its result clearly shows the evaluation of the specific plan.

We obtain the **point cloud data** from USGS, and separate the map of the land into 1,786 grids. We obtain the height variance, signal coverage, and tree area of each grid. Then, we use **K-Means model** to divide the land into two clusters. After modifying the results of clusters, we divide the land into three parts, and get the areas of each part. We use the prediction models and the metric to calculate the results for each of the three parts of land using four different plans. After we obtain the raw data, we correct them according to the suitability of land by considering height variance and signal coverage. At last, we get the final results. The three plans we decide to use are: sports complex, solar array, and agritourist center.

Then, since we found that there will be a semiconductor factory being built nearby, we decided to make some changes to the original data to fit more about the future environment. We changed the demand in electricity, and changed the population density in the nearby area. Therefore, we successfully fit our model to future situation.

Keywords: the Entropy Weight Method, Prediction Models, Average Demand Model, Point Cloud Dataset, K-Means Model

Contents

Contents

1 Introduction

1.1 Background

As the world population continues to grow and urbanize, the demand for land resources keep increasing rapidly. However, the limited availability of land resources and the negative impacts of land use practices, such as deforestation, soil erosion, and loss of biodiversity, have raised concerns about the long-term sustainability of land use.

To address these challenges, there is a need for a metric that can quantify the ”goodness” of land use practices and provide guidance for policymakers, land managers, and other stakeholders to make informed decisions about land use. Such a metric should take into account of various factors, including geography, climate, business options and community needs, and provide a comprehensive and integrated assessment of land use practices.

To develop an effective model for measuring the efficiency of land use, it is important to first identify a specific area of interest. In this case, we are focusing on an example area in Wayne County, New York, which is approximately 3 km^2 in size. By analyzing this area, we can develop a more targeted and specific model that takes into account the unique characteristics of the region.

Figure 1 Different Uses of Land

1.2 Problem Restatement

We take 3 km^2 land located in Wayne County, New York as an example and we are provided with the geometric data of it. From the data, we need to build a model to measure how ”good” is the use of land.

- **Problem 1:** Define a quantitative decision metric to measure how ”good” is the use of land. This metric should consider both short-term and long-term benefits and costs.
- **Problem 2:** Take at least 2 options from the possible use of land given and determine the metric defined in problem 1. We need to determine the factors to focus on for each option.

- **Problem 3:**Given an extra option and some of its data, including the jobs it could provide, both directly and indirectly, and the building cost of it. We need to improve our metric according to this data and re-evaluate the options identified in problem 2 using the new metric.
- **Problem 4:**Briefly discuss how appropriate this model could work in an environment we're familiar with. We need to make some comments about what might have to change if the land belonged to a different country or location. We need to determine how generalizable our model is.

1.3 Our Work

To evaluate the land use plans, we first develop a metric IUL. We consider three aspects: environmental impacts, business profit, and the community utility, corresponding to the meric of IE, IB, and IC. Each aspect contains multiple indicators. We use the **double-level Entropy Weight Method** to obtain the weights of indicators in each aspect, as well as the weights of IE, IB, and IC.

Next, we build prediction models for specific land uses' indicators. The four land uses we consider are: agritourist center, solar array, sports complex, and ranch. We separate them into **Demand-Restricted Land Uses** and **Demand-UnRestricted Land Uses** to build the **prediction models**. For the Demand-restricted Land Uses, we build the **Average Demand Model**. The demand is directly related to many of the indicators including running cost and total revenue. By connecting all the prediction models of indicators with the metric, we are able to have a clear evaluation of different plans.

We obtain the point cloud data from USGS and the height variance, signal coverage, and tree area of each grid. Then, we use **K-Means model** to divide the land into two clusters. After correcting the results of clusters, we divide the land into three parts, and get the areas of each part.

Using the prediction models together with the metric, we obtain the results for each of the three parts of land using four different kinds of plans. After we obtain the raw data, we correct them according to the suitability of land by considering height variance and signal coverage. At last, we get the final results.

We evaluate the impacts of the new factory by taking population change and community utility change into consideration.

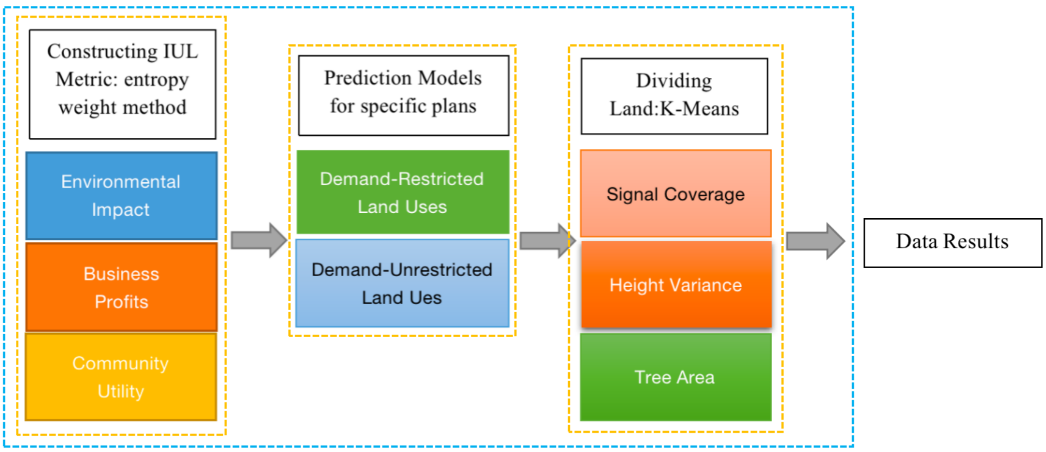


Figure 2 Model Overview

2 Assumptions and Notations

2.1 General Assumptions

- **the average visits to agritourist centers and sports complexes are constants:** In order to calculate the total visits and simplify our model, we assume the average visits are suitable to be used to represent the visit numbers for our plan.
- **the population is evenly distributed in the county:** This can help us calculate the number of people in a specific area easily. Wayne County's population distribution is comparatively even, so we can assume that it is evenly distributed.
- **the three aspects cover all the values we need to evaluate the efficiency of the plan:** We divide the metric into three aspects, and these aspects take into account various kinds of factors. After conducting enough research, we believe all the key factors are covered in the three aspects.
- **the questionnaire is reliable to provide data** When modeling utility, an abstract value, as well as probability, some of the data we collected are through questionnaires that are randomly given to USA residents. Though we couldn't go to a lot of areas, the number of responses, 50, should provide enough data for us to model.

2.2 Notations

Symbol	Description	Unit
IE	Index of environmental impacts aspect	N/A
IB	Index of business profit	N/A
IC	Index of community utility	N/A
IE/IB/IC$_i$	Index of the i^{th} indicator of the aspect	N/A
P_c	The influence on the probability one would go to a sports complex	N/A
P_s	The probability of going to an $S\,km^2$ sport complex despite other factors	N/A
P_d	The probability of going to an dkm far sport complex despite other factors	N/A
D	Total demand	N/A
H	Terrain height	m
T	Tree coverage rate	N/A
S	Signal coverage rate	N/A
A	Area in Acres	Acres

3 Land Usage Decision Metric

Please note that our metric for determining the final use of this specific land is related to the plans we want to evaluate in problem 2, so the metric in this section is incomplete without connecting it to the prediction models in the fourth section.

3.1 IUL metric

In problem 1, we are asked to build IUL metric to measure how ”good” the usage of the land is. The symbol of the metric IUL refers to Impacts of the Use of Land.

We evaluate IUL through three aspects: environmental impacts(IE), business profits(IB) and community utilities(IC). Note that “utility” here is an economics term which is “the usefulness or enjoyment a consumer can get from a service or good”(Investopedia).

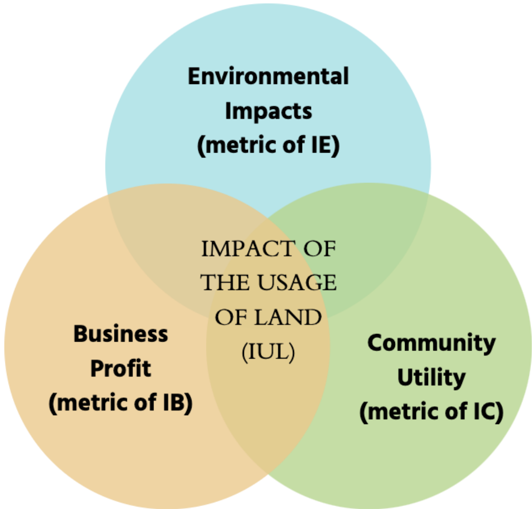


Figure 3 Three Aspects of IUL

These are the most important factors to determine the usage of land, and we will justify each of them in the following parts. Each of the aspects contains several indicators. The indicators include long-term indicators and short-term indicators, so that we will be able to take both the short and long-term impacts of the use of land into consideration.

After we determine the indicators, we would use the Entropy Weight Method to determine the weight of each indicator, as well as the weight of each aspect.

3.2 Indicators for the Metric of IE

To build the metric of IE, Environmental Impacts of the Use of Land, we conducted research on the most important environmental factors when considering the use of land[1]. We found four potential IE indicators for the target $3km^2$ of land: air pollution, wastewater discharge, green area, and saved waste gas.

We replace green area by tree area because for most of the usages of land, grass and other plants can still remain on the ground, while trees are needed to be clear. Therefore, tree area is the determining factor of green area in this case.

The four indicators are listed in the figure below and further explained in table1. Note that for IE_1, since some of the land usages, such as building solar arrays, not only don’t produce air pollution, but also reduce the original air pollution, the index can be a negative number. To simplify the model by avoiding negative numbers, we use another indicator IE_4 to represent the saved waste gas in kg.

The time range indicates whether the index has a continuous influence. Long range indexes have both short-term and long-term effects, while short range indexes have only short-term effects.

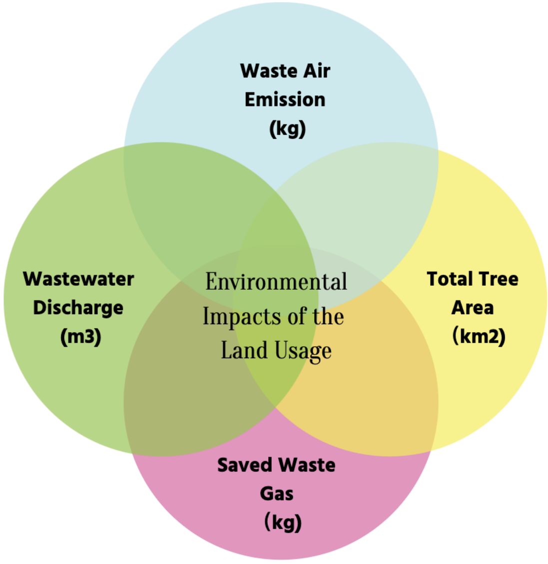


Figure 4 Four Indicators of the Environmental Aspect

The target means whether the increase (↑) or decrease (↓) of the indicator has a positive effect on the final score.

Table 1 Metric of IE

Aspects	Notation	Indicators	Time Range	Notation	Target
Environmental Impacts	IE	Waste Air Emission Index	Long	IE_1	↓
		Wastewater Discharge Index	Long	IE_2	↓
		Saved Waste Gas Index	Long	IE_3	↓
		Reduced Tree Area	Long	IE_4	↓

The function of IE metric is shown below. However, we still don't know the weight of each indicator. The calculation of the weights, using the Entropy Weight Method, will be conducted together when we determine all the functions of the three aspects.

$$IE = f(IE_1, IE_2, IE_3, IE_4) \tag{1}$$

3.3 Indicators for the Metric of IB

Profit is determined by total revenue and total cost. We divide total cost into running cost and building cost.

Table 2 lists the indicators in the IB metric. Building cost is a short-term indicator.

Table 2 Metric of IB

Aspects	Notation	Indicators	Time Range	Notation	Target
Business Profit	IB	Building cost	Short	IB_1	↓
		Total Revenue/year	Long	IB_2	↑
		Running Cost/year	Long	IB_3	↓

Hence, we can write down the function of IB, and decide the weight of each indicator together with other functions using the Entropy Weight Method.

$$IB = f(IB_1, IB_2, IB_3) \tag{2}$$

3.4 Indicator for the Metric of IC

Community utility is used to quantify the usefulness or enjoyment a consumer can get from a service or good. It is a kind of scoring system, and the quantity does not have a unit. In this paper, we get the data by issuing a questionnaire to 50 random American citizens and asking them to rate the community utility in different situations. By using this single quantity, we are able to add all kinds of convenience, happiness and usefulness provided by the new usage of the $3km^2$ land together to value the total community value the final plan has.

The metric only contains one indicator, since we put all kinds of community utilities into one variable:

$$IC = IC_1 \tag{3}$$

Table 3 Metric of IC

Aspects	Notation	Indicators	Time Range	Notation	Target
Community Utility	IC	Total Utility	Long	IC_1	↑

3.5 Data Collection

We need a large number of data to conduct the following weight calculation. We gather the data according to how we are going to make final plan in problem 2. To briefly introduce our plan for problem 2, we choose four plans to consider: building solar array, argitourist center, ranch, and sports complex. We will discuss this in details in Section 4. How it is related to this problem is the data we gathered to build the metric for problem 1 are all derived from these four kinds of accommodations. We found samples of these four kinds of accommodations in the vicinity of the land, five samples of each option, to collect enough data.

3.6 Determine the Weights of All Indicators and Aspects: the Entropy Weight Method

We use the entropy weight method to finalize our metrics system.

First, we need to scale the data from 0 to 1 using this formula:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}}$$

Then, we calculate the power of i^{th} sample in j^{th} indicator:

$$p_{ij} = \frac{x_{ij}}{\sum_{i=1}^{n} x_{ij}} \tag{4}$$

Then, we calculate the entropy for the j^{th} indicator:

$$e_j = -k \sum_{i=1}^{n} p_{ij} \ln(p_{ij}) \tag{5}$$

where $k = \frac{1}{\ln(n)}$.

Next, we calculate the entropy redundancy for the j^{th} indicator:

$$d_j = i - e_j \tag{6}$$

And finally, we can calculate the weight of each indicator:

$$w_j = \frac{d_j}{\sum_{j=1} md_j} \tag{7}$$

For this model, we conduct 3 entropy weight calculations which are in two levels. We calculate the second-level-weights of IE_1, IE_2, IE_3 and IE_4 to obtain IE; we calculate the second-level-weights of IB_1, IB_2 and IB_3 to obtain IB; finally, we calculate the first-level-weights of IE, IB and IC to obtain the final formulas of our metric system IUL.

Our results are shown in table4:

Table 4 Metric Aspects and Indicators

Aspects	Aspect Weight	Indicators	Indicator Weight
IE	0.2565	IE_1	0.3034
		IE_2	0.3034
		IE_3	0.3165
		IE_4	0.0767
IB	0.3059	IB_1	0.3621
		IB_2	0.4467
		IB_3	0.1912
IC	0.4376	IC_1	1

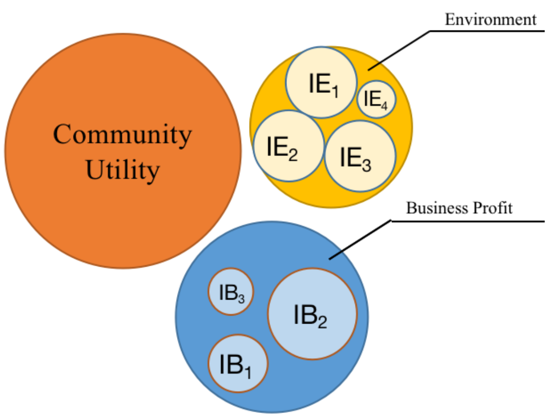


Figure 5 Weights of Variables Indicated by the Sizes of the Circles

And the final formulas are:

$$IDL = 0.3059IB + 0.2565IE + 0.4376IC \tag{8}$$

in which

$$\begin{cases} IB = 0.3621IB_1 + 0.4467IB_2 + 0.1912IB_3 \\ IE = 0.3034IE_1 + 0.3034IE_2 + 0.3165IE_3 + 0.0767IE_4 \\ IC = IC_1 \end{cases}$$

4 Prediction Models of Specific Land Uses' Indicators

In section 3, we have finished our metric to evaluate the plans. However, to evaluate the usage of this specific land, we don't have the data readily available. To obtain the value of the indicators, we build several prediction models for each of the options we choose. By plugging in the results of the prediction models into the metric system, we can have a clear insight of which of the plans is the best one.

Among all the potential options for the use of land, we select four of them to consider: building solar array, argitourist center, ranch, and sports complex. These four options respectively correspond to: environmental protection, educational experience, local people's living, and entertainment. Thus, they together greatly cover the demands of local people as well as those of the government.

Moreover, we divide these four options into two types: demand-restricted and demand-unrestricted. We claim that the sports complex and agritourist center are demand-restricted because both of them sell tickets, and the sales of tickets are controlled by the demand of local people. On the contrary, the implementation of solar array and ranch would not be restricted by the demand since the markets of electricity and food are much larger and less restricted in local area than those of the tickets.

Figure 6 Four Types of Land Uses

4.1 Prediction Models for Demand-Unrestricted Land Uses

We build this prediction based on the past average data and the area planned of the facility.

4.1.1 Ranch

For the revenue, maintenance cost and building cost, we decided to use the original data we found to make the prediction. The data is shown in the table below. (Unit: million USD)

Table 5 Original dataset for the ranch

Area(m^2)	Building cost	Revenue per year	Running cost per year
1214058.102	2.5	1.2	0.6
1011715.085	1.8	0.9	0.45
607029.0508	1.2	0.7	0.3
809372.0677	2	1	0.5
1618744.135	3.5	1.5	0.7

Therefore, we used the following formula to predict the revenue(R), Building Cost(BC) and Running cost(RC):

$$\begin{cases} IB_2 = \frac{S_{predict}}{S_{average}} IB_{2average} \\ IB_1 = \frac{S_{predict}}{S_{average}} IB_{1average} \\ IB_3 = \frac{S_{predict}}{S_{average}} IB_{3average} \end{cases} \tag{9}$$

where the suffix "average" means that the data is the average value of the variable from the original data set.

The utility of the farm is based on it's production. We first modeled the amount of cows. We converted the area into acres, and according to our research, normal ranch has an amount of 2 cows per acre, means that the amount of the cow is $2A$. Therefore we can model the utility, wastewater production and air pollution.

$$IC = 2A \cdot \frac{1}{7} \cdot 0.5$$

where 2A is the amount of the cows, $\frac{1}{7}$ means 1 cow can support 7 people on average for a year, and 0.5 is the base of utility for ranch.

$$IE_1 = 2A \cdot 0.25 \cdot 365$$

where 2A is the amount of cows, 0.25 is the weight of waste air one cow produces one day, and times 365 gets the waste air emission in a year.

$$IE_2 = 2A100 \cdot 365\ 0.0037854118$$

where 2A is the amount of cows, 100 is the weight of waste water produced by a cow one day in gallon, and times 365 gets the waste water discharge in a year, times the 0.0 0378 converts the unit into cubic meter.

All the calculation formulas are shown in the figure below.

Aspects	Indicators	Calculation
IE	IE_1	2A*0.25*365
	IE_2	2A*100*365/0.00378
	IE_3	0
	IE_4	0
IB	IB_1	S*2200000/1052183
	IB_2	S*1060000/1052183
	IB_3	S*510000/1052183
IC	IC_1	2A * 0.5 * 1/7

Figure 7 Formula for ranch indicators

4.1.2 Solar Array

We build this model based on the principle of the solar panel. We take an average of 351 MWh of electrical energy produced by 1 acre of solar panel a year. Therefore, we can calculate the following indicators:

$$\begin{cases} IB_2 = A \cdot 351300 \cdot 0.2 \\ IB_3 = A \cdot 351300 \cdot 0.163 \\ IB_1 = A \cdot 450000 \end{cases} \tag{10}$$

where $A \cdot 351300$ is the energy produced from the solar panel, 0.2(USD) is the average electricity price in New York per kwh, 0.163(USD) is the maintenance cost per kwh of electricity produced, and 450000(USD) is the average price for solar panel per acre.

The utility for the solar panel is determined similar to the ranch, as shown below:

$$IC = \frac{A \cdot 351300 \cdot 0.1}{3500} \tag{11}$$

where 3500 is the average electricity required for 1 person per year, 0.1 is the basic utility for the solar array.

This utility can reduce the air pollution, since the previous energy comes from coal. Producing one kwh will release 0.428 kg of CO_2 using the traditional method. Therefore, saved waste gas can be determined as:

$$IE_3 = 0.428 \cdot A \cdot 351300 \tag{12}$$

The formula for the model are in the figure below.

Aspects	Indicators	Calculation
IE	IE_1	0
	IE_2	0
	IE_3	Area(in Acre)*0.428*351300
	IE_4	0
IB	IB_1	Area(in Acre)*450000
	IB_2	Area(in Acre)*351300*0.2
	IB_3	Area(in Acre)*351300*0.163
IC	IC_1	Area(in Acre)*351300*0.1/3500

Figure 8 Formula for solar array indicatorsk

4.2 Prediction Models for Demand-Restricted Land Uses

We need to obtain the value of indicators when we turn a certain area of land into a sports complex and ranch. Our prediction models need to solve the indicators for all the three aspects: IE, IB, and IC.

For demand-restricted land uses, we need to calculate the average demand of visits per year. It's important to build the average demand models for the two options.

4.2.1 Sports Complex

Based on our research, we assume that the average visits of all the sports areas per person per year is 20. Also, we assume that both the potential customers and other sports complexes that are competing with us are all within a fifty kilometer radius of our sports complex.

For convenience, we assume the population is evenly distributed in the county that the land belongs to. Also, in every spot with 0.01 longitude times 0.01 latitude, the people in it have the same demand for sports complex. As shown in the graph, the circle is with a fifty kilometer radius of our sports complex, and every spot in the circle represents an area of a 0.01 longitude time 0.01 latitude square. (0.01longitude*0.01latitude$\approx 1.11km \cdot 1.11km$)

Thus, the total visits to any sports complex in this area in one year is:

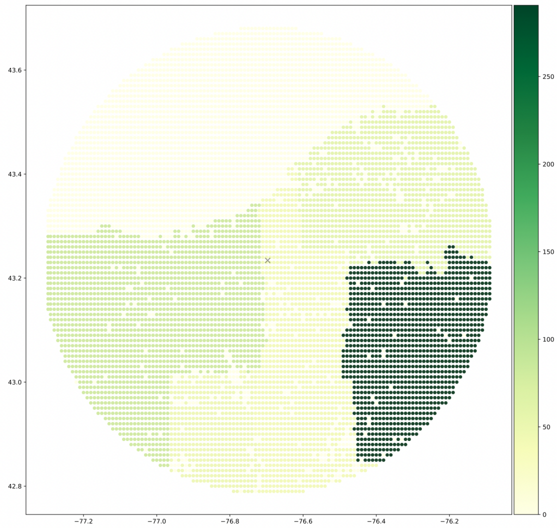


Figure 9 Population Density Map

$$TotalVisits = TotalPopulation \cdot 20 \tag{13}$$

The total visits can be the total demand of all sports complexes' tickets in this area. However, the demand is also affected by the total capacity of all sports complexes in the area, so we have to model the capacity as well. After research, we determine that the average capacity of a sports complex is 50 people per square meter in one year:

$$TotalCapacity = 50000000 \cdot S \tag{14}$$

, where S is the area of the sports complex.

When the number of people reach the capacity, the total demand of tickets would be restricted and equal to the capacity. Thus, the formula of total demand of sports complex tickets in this area of circle with a radius of 50 kilometers in one year is:

$$TD = min[TotalVisits, TotalCapacity] \tag{15}$$

The number of demand in a spot is:

$$D_{spot} = (\frac{Population_{spot}}{TotalPopulation}) \cdot TD \tag{16}$$

The three factors a customer would likely consider when determining which sports complex to go are: percentage of occupied area, size of the sports complex and the distance to the complex.

First of all, the percentage of occupied area (C=number of people in the sports complex/capacity of this sports complex) in the sports complex would affect their decision. If a sports complex is very crowded, some people would not be willing to go there. When it is already full, no one would go there.

We determine the formula of P(C):

$$\begin{cases} P(c) = 1 \quad (0 \leq c \leq 0.75) \\ P(c) = -4(x - 0.25) + 1 \quad (0.75 < c < 1) \\ P(c) = 0 \quad (c \geq 1) \end{cases} \tag{17}$$

P(C) can be regarded as the influence on the probability one would go to a sports complex.

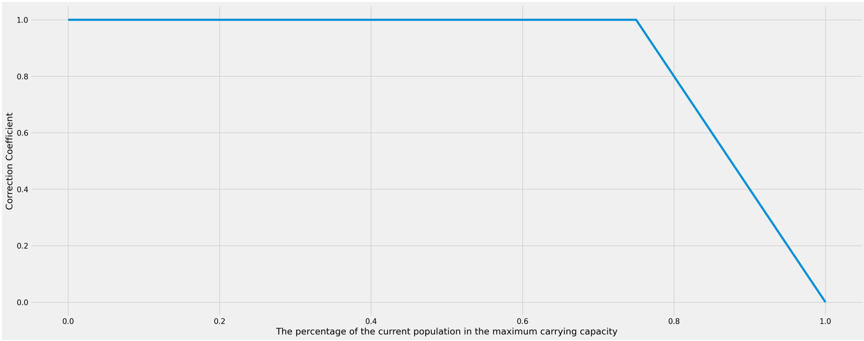


Figure 10 Diagram of $P(C)$

Secondly, the size of the sports complex would affect one's willingness to go because the size indicates the number of facilities, the diversities of sports, and the number of stadiums or courts in the sports complex. These all contribute to the fact that the larger the size, the larger the probability that the customer would go.

Thus, we build the rating model of size to help obtain the final probability.

$$y = \frac{kx}{\sqrt{1 + (kx)^2}} \tag{18}$$

We use the above function to conduct regression.

Thirdly, the distance to the sports complex would also affect the willingness of customers to go. Apparently, people are more likely to go to the sports complex that is closer to them.

We use this function to conduct regression and obtain the rating model of distance.

$$y = -\frac{kx}{\sqrt{1 + kx^2}} + 1 \tag{19}$$

$P(S)$ can be regarded as the relative probability of going to an $S\,km^2$ of sports complex when other factors are the same. $P(d)$ can be regarded as the relative probability of going to a dkm far sports complex when other factors are the same.

These two functions well fit the data. For one thing, as the graph shows, the curve of $P(S)$ shows that the increasing speed of probability decreases when the size keeps increasing. This is because when most of the people are already satisfied by the facilities in a certain size of sports complex, the increasing size would not benefit them much. For another, the decreasing speed of probability decreases when the distance keeps increasing. This is because at this point the distance already exceeds most of the people's tolerance range, the probabilities for these people are already zero.

We gather the data for regression by distributing questionnaires randomly to USA residents.

The final formulas obtained after regressions are:

$$\begin{cases} P(S) = \frac{83.279S}{\sqrt{1+(83.279S)^2}} \\ P(d) = -\frac{0.03d}{\sqrt{1-0.03d^2}} + 1 \end{cases} \tag{20}$$

S in square kilometer is the area of the sports complex, and d in kilometer is a person's distance to the sports complex.

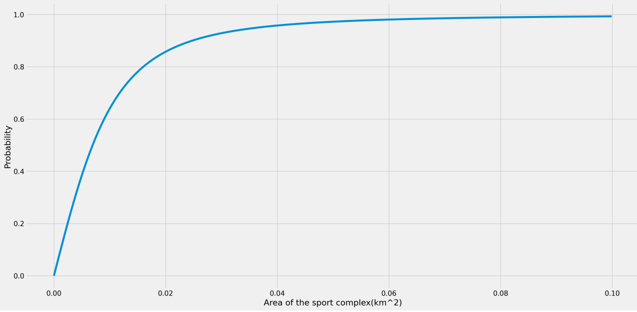


Diagram of $P(S)$

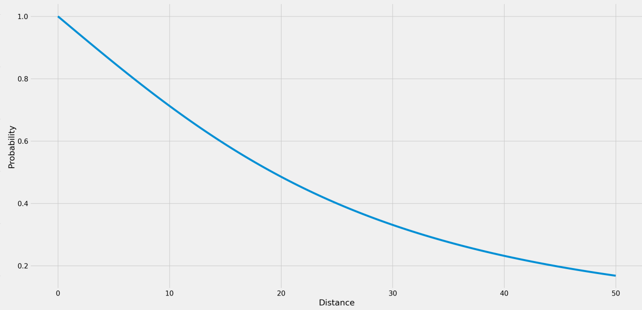


Diagram of $P(d)$

Finally, by putting all three indexed together, we would obtain the true probability:

$$P = \frac{P(S) \cdot P(d) \cdot P(C)}{\sum_{i=1}^{n} P(S_i) \cdot P(d_i) \cdot P(C_i)} \tag{21}$$

P is the probability the people in one spot going to the specific sports complex. n is the total number of local sports complexes. The nominator of function P is the total probability these people in the spot going to any local sports complex, and the numerator is the probability these people in the spot going to this specific sports complex.

Finally, the total demand of tickets to the specific sports complex in one year is:

$$D = \sum_{i=1}^{m} D_{spoti} \cdot P_i, \tag{22}$$

where m is the total number of spots in this circle area.

By adopting this method, we can calculate the sport complex that people are most willing to go to in any spot. Below is the diagram of it. Here, the ”plus” point represents a sport complex. The color of a point, which corresponds the color of a sport complex plus mark, denotes the sport complex people in that spot are most willing to go to. The brown ”plus” point in the middle denotes the new sport complex built in the given land. (The blue points denote the lake)

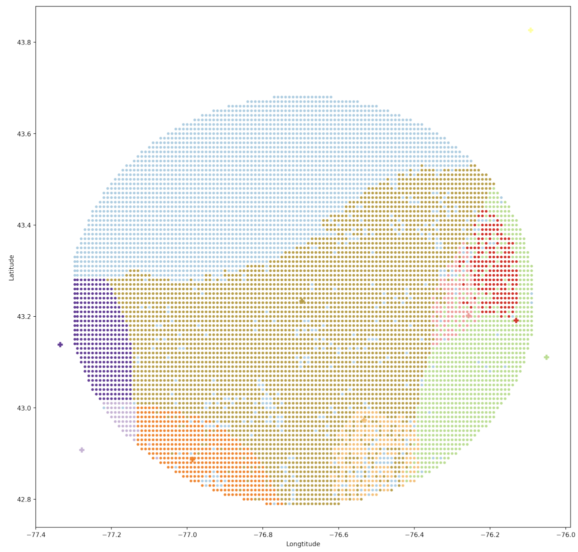


Figure 11 Favourite sport complex distribution

The reason why solving for D is so important is that many indicators are associated with D.

Aspects	Indicators	Calculation
IE	IE_1	0
	IE_2	0
	IE_3	0
	IE_4	0.80*Current Area Percentage of Trees
IB	IB_1	Average Building Cost Per km² ·S = 2000000$ · S
	IB_2	Average Ticket Price · D = 30$ · D
	IB_3	Running Cost Per Tourist · D = 20$ · D
IC	IC_1	D*0.35*20

Figure 12 calculations of sports indicators

The figure below shows the calculation formulas of all indicators for building a sports complex.

The first three indicators equal to 0 because we assume the operation of sports complex would not cause air or water pollution. 0.95 in IE_4 is the percentage of trees needed to be cut.

4.2.2 Agritourist Center

The prediction models of indicators for the ranch are mostly the same as those of sports complex.

We assume every person in the range would visit the agritourist center one time in a year in average.

$$TotalVisits = TotalPopulation \cdot 1 \tag{23}$$

The agritourist centers's average capacity is 18250000 people per km^2 in a year (assuming every day the capacity is $20m^2$ per person).

$$TotalCapacity = 18250000 \cdot S \tag{24}$$

$$TD = min[TotalVisits, TotalCapacity] \tag{25}$$

The number of demand in a spot is:

$$D_{spot} = (Population_{spot}/TotalPopulation) \cdot TD \tag{26}$$

We use the same P(C) function for ranch. And the same function to do regression for $P(S)$ and $P(d)$

$$\begin{cases} P(S) = \frac{0.9427S}{\sqrt{1+(0.9427S)^2}} \\ P(d) = -\frac{0.03d}{\sqrt{1-0.03d^2}} + 1 \end{cases} \quad P = \frac{P(S) \cdot P(d) \cdot P(C)}{\sum_{i=1}^{n} P(S_i) \cdot P(d_i) \cdot P(C_i)} \tag{27}$$

, n is the total number of agritourists center in the area.

The total demand of tickets to the agritourist center in one year is:

$$D = \sum_{i=1}^{m} D_{spoti} \cdot P_i, \tag{28}$$

where m is the total number of spots in this circle area.

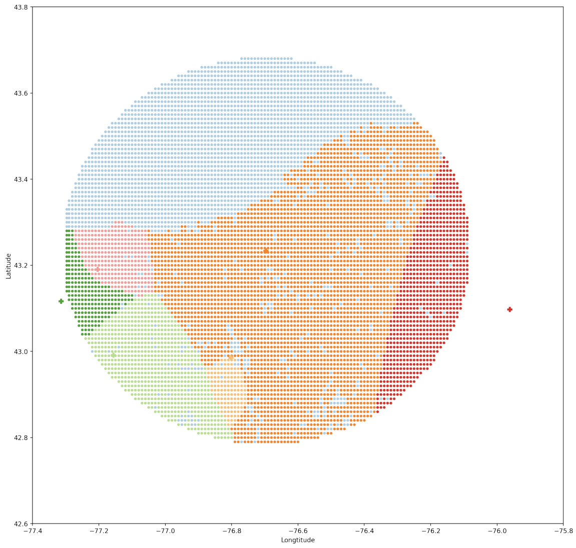


Figure 13 Favourite agritourist farm distribution

Using the same notation as figure.13, below is the distribution of the favourite tourist farm.

Below are the calculation formulas of all indicators for building an agritourist center

Aspects	Indicators	Calculation
IE	IE_1	0
	IE_2	0
	IE_3	0
	IE_4	0.80*Current Area Percentage of Trees
IB	IB_1	Average Building Cost Per km² ·S = 2000000$ · S
	IB_2	Average Ticket Price · D = 30$ · D
	IB_3	Running Cost Per Tourist · D = 20$ · D
IC	IC_1	D·0.3

Figure 14 calculations of agritourist center indicators

5 Determine Specific Plan for the Use of Land

5.1 Point cloud data

To find the accurate terrain data for the given array, we went to the official website of the US Geological Survey, "National Map Downloader", and we downloaded the point cloud data, the .las file provided. A LAS file is a standardized format for storing and sharing LiDAR (Light Detection and Ranging) data. LiDAR is a remote sensing technology that uses laser beams to measure the distance between the sensor and the object or surface it hits. This technology is widely used in various applications such as topographic mapping, forestry, civil engineering, and urban planning.

The LAS file format contains both the raw point cloud data ($x, y, zcoordinates$) and additional information such as the intensity of the laser returns, classification information (ground, vegetation, buildings, etc.), and GPS time stamps. It also includes header information that describes the data properties, such as the data format, coordinate reference system, and data acquisition parameters.

Figure3 is the original point cloud data consisting of six square shapes, with the given area outlined in black.

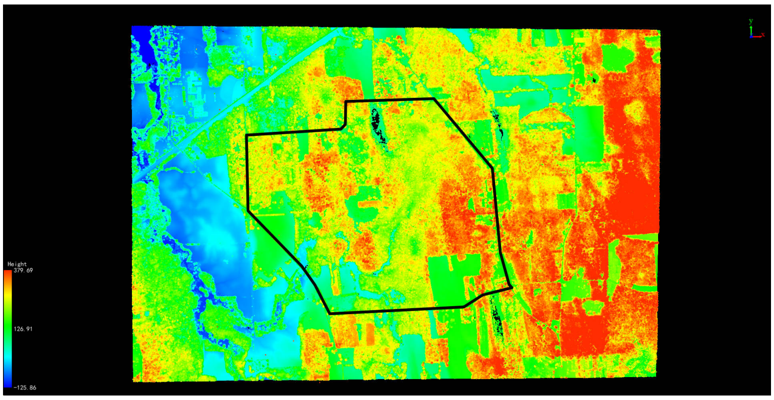


Figure 15 Primary data from USGS

In order to conduct an in-depth analysis of the point cloud data, we utilize the machine learning classification module of LiDAR360. This module employs machine learning classification using a machine learning model with the suffix ”vcm”. The “vcm” machine learning model used in LiDAR360 is designed to handle large and complex point clouds. It is trained on various LiDAR data sets, enabling it to classify points based on their attributes, such as height, intensity, and color. By utilizing this module, we conduct a more detailed and accurate analysis of the point cloud data, providing insights and information that would have been difficult to obtain through manual classification. And we get the classification of the ground glass(orange points) and the plant class(green points).

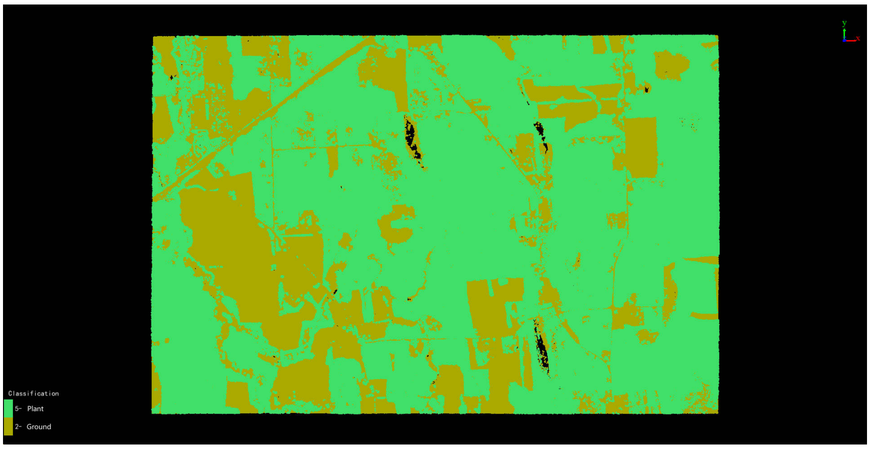


Figure 16 Classification Result

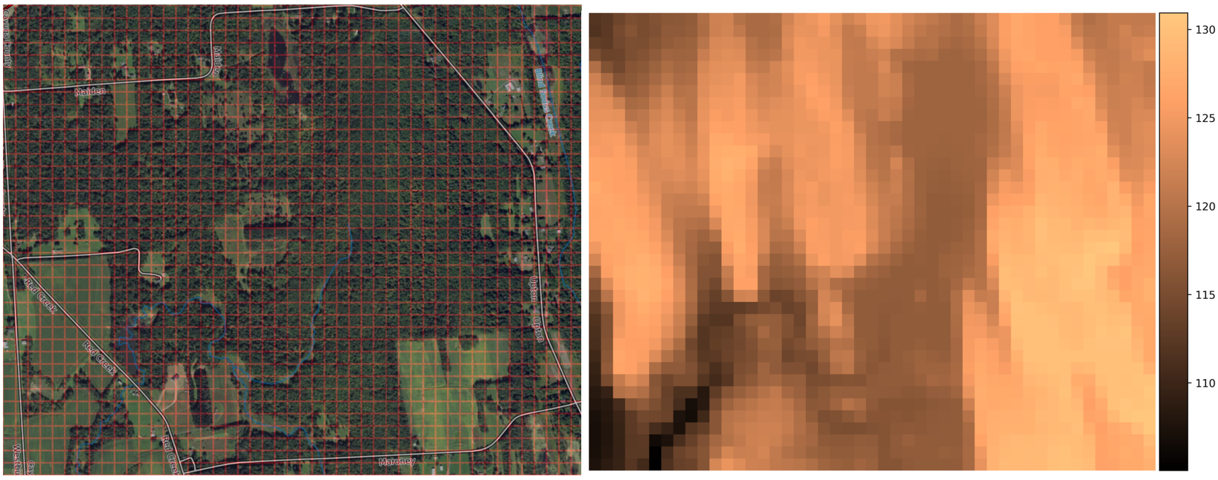
Then we delete most of the points that are not located in the given region, then convert the terrain into a mesh grid, dividing it into 1,786 grids, each with a size of 50 meters by 50 meters.

We conduct an analysis of the terrain height, and tree coverage, and based on the map provided by the organizer, we solve the signal coverage for each grid point.

$$H(Grid_{i,j}) = \frac{\sum_{(x,y)\in Grid_{i,j}} H(x,y)}{\#\,of\,(x,y) \in Grid_{i,j}} \tag{29}$$

H represents height, S represents signal coverage, and T represents tree area.

$$T(Grid_{i,j}) = \begin{cases} 0 & \frac{\#\,of\,Tree}{\#\,of\,Points \in Grid_{i,j}} \leq 0.6 \\ 1 & \frac{\#\,of\,Tree}{\#\,of\,Points \in Grid_{i,j}} > 0.6 \end{cases} \tag{30}$$

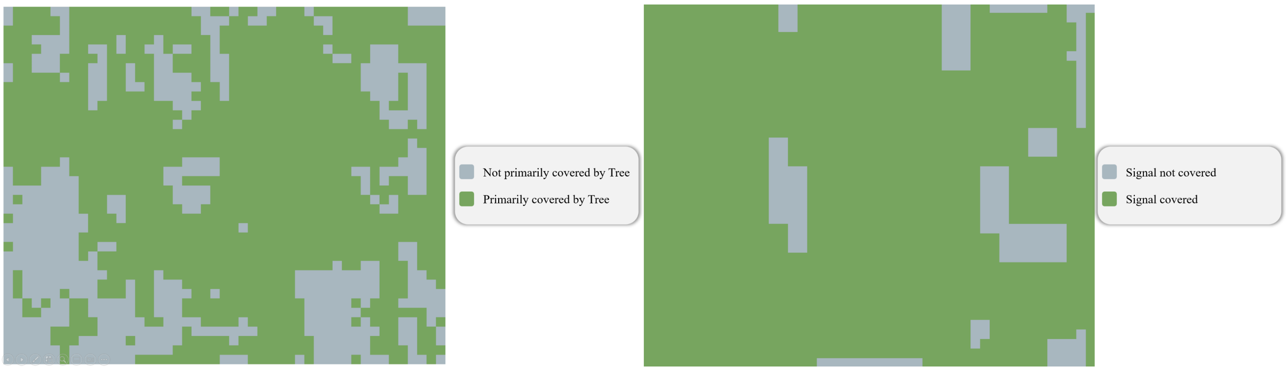


Mesh Grid with Real map

Grid Terrain Result

$$‘S(Grid_{i,j}) = \begin{cases} 0 & No \quad signal \quad covered \\ 1 & Signal \quad covered \end{cases} \tag{31}$$

The resulting visualized graph is showned below:



Grid Tree Result

Grid Signal Result

5.2 Dividing Land: K-Means model

We use the K-Means cluster analysis model to divide the land. The K-Means algorithm is used to divide the original data set into several clusters. Suppose that the clusters are $C_1, \quad C_2, ... \quad C_k$, then our goal is to minimize the sum of squared errors(SSE). It can be explained as the following equation:

$$SSE = \sum_{i=1}^{k} \sum_{x \in C_i} ||x - \mu_i||_2^2 \tag{32}$$

where

$$\mu_i = \frac{1}{|C_i|} \sum_{x \in C_i} x$$

In our method, we consider 3 topographic characteristics to build this cluster analysis model: signal coverage, height variance and original plant/tree coverage. We first need to determine how many clusters we need to sort the data into. We used the "elbow" method, which is to find the turning point ("elbow") of the graph. We plot a graph with the number of clusters as the x-axis and SSE as the y-axis, then we can find the turning point of this graph. In our process, the turning point appears when the number of clusters is 2, meaning that the best cluster number is 2.

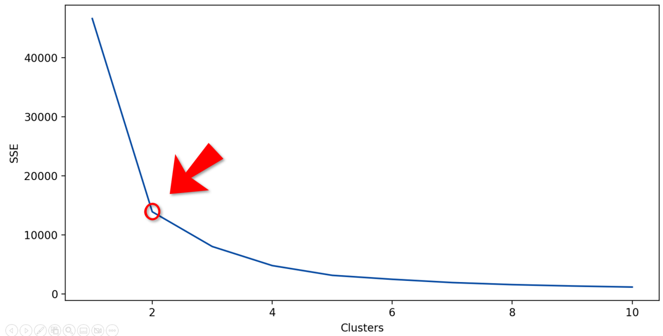


Figure 17 Result of the Elbow Method

Finally, the land is divided into three parts which belong to two different types of land. The figure below on the left is the clustering result:

But because such a classification is not very friendly for large-scale planning, as there are many scattered grids, we made some artificial adjustments. The figure on the rightshows the clusters after artificial adjustment.

Primary k-mean cluster result

Artificial modification result

From left to right, we defined the three large-scale clusters as $(LEFT, MID, RIGHT) \in CLUSTERS$. Hence, we have:

Table 6 Detailed data of the sorted land

	LEFT	MID	RIGHT
σ_H	14.996368020423233	10.379969699851694	7.950415453617519
$T\%$	0.6548223350253807	0.7470881863560732	0.6618705035971223
$S\%$	0.9458544839255499	0.7504159733777038	0.8111510791366906

5.3 Final Plans for Each of The Three Parts of Land

Finally, we use the prediction models and the metric to calculate the results for each of the three parts of land using four different methods:

	Land No.1	Land No.2	Land NO.3
Sports Complex	0.392	0.385	0.398
Agritourist Center	0.090	0.087	0.088
Ranch	-0.163	-0.169	-0.154
Solar Array	0.124	0.122	0.116

Figure 18 Raw Results

5.3.1 Suitability

Since the building has some limit on land condition, we added a variation called the suitability to measure how good does the land fits the usage of land.

We correct our raw data by taking signal coverage and height variance into consideration.

Because a flat land is most suitable for the building of sport complex, tourist farm, and animal ranch, we want to build a function that has the property of $f(0) = 1$ and when approaches infinity, $f(x)$ approaches 0. Therefore we decide to use the same function as we used for P_d. By plugging $f(9) = 0.9$, as a value we estimate based on data, we obtain:

$$f(\sigma_H) = -\frac{0.011\sigma_H}{\sqrt{1 + 0.011\sigma_H^2}} + 1, \tag{33}$$

where σ_H is the height variance. The diagram below shows our suitability correction formula:

	Correction by Multiplying	Justification
Sports Complex	$f(\sigma_H)$*signal coverage	Sport complexes need to be built on flat land. Tourists there need signal.
Agritourist Center	$f(\sigma_H)$*signal coverage	Agritourist farms need to be built on flat land. Tourists there need signal.
Ranch	$f(\sigma_H)$	Ranches need to be built on flat land. Grazing does not require signal.
Solar Array	1	Solar farms can be built on any terrain. Solar power generation does not require signal.

Figure 19 Correction of suitability

The final results of our models are:

	Land No.1	Land No.2	Land NO.3
Sports Complex	0.332	0.270	0.289
Agritourist Center	0.080	0.082	0.079
Ranch	-0.138	-0.118	-0.112
Solar Array	0.124	0.122	0.115

Figure 20 Modified Results

It is reasonable that sports complex obtain the best evaluations for all three parts of lands due to its high demand of local people.

Since we want the uses of land be more diverse, we choose land No.1 to be a sports complex, since it has the highest evaluation. We choose No.2 to a solar array, since solar array has the second highest evaluation, and land No.2 is the most suitable land for solar array besides land No.1, which is already chosen. Finally, we choose No.3 to be the agritourist center, which has the third highest evaluation.

6 Change the model: New Factory

In problem 3, we will have a new semiconductor built Onondaga County, which is near to our land. According to our analysis, we have 2 data changes:

1. The electricity demand. Since it is a semiconductor factory, it has a high requirement on electricity. So, we improve the base utility point of solar array from 0.1 to 0.3.

2. The population of the county. According to the given data from the problem, the new factory requires 49000 people working here. So the population in Onondaga County will increase from 476516 to 525516.

After the change in data, we have the raw results shown below.

	Land No.1	Land No.2	Land NO.3
Sports Complex	0.392	0.385	0.398
Agritourist Center	0.040	0.038	0.039
Ranch	-0.163	-0.169	-0.154
Solar Array	0.127	0.126	0.119

Figure 21 Raw results after change in data

The modified result(raw result times suitability) is shown in graph below.

	Land No.1	Land No.2	Land NO.3
Sports Complex	0.332	0.270	0.289
Agritourist Center	0.036	0.035	0.035
Ranch	-0.138	-0.119	-0.112
Solar Array	0.127	0.126	0.119

Figure 22 Modifies Results

7 Generalizability Analysis

I think our model can provide a good approach towards planning the usage of the land. Let's take our familiar environment, Shanghai, as an example.

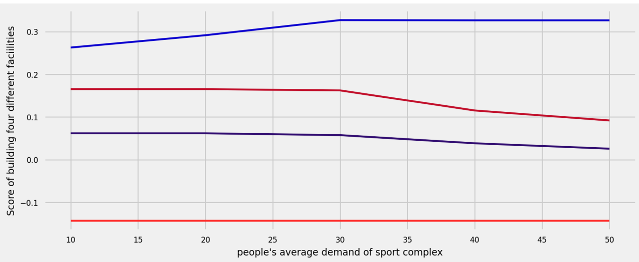
Firstly, the demand in Shanghai is much more than that in other regions. Shanghai has a high density of population, which means that the demand will increase greatly compared to the given location in Wayne County.

Secondly, Shanghai has a very flat land, so height variance doesn't need to be taken into consider. However, we need to consider about the rivers or water in Shanghai, since shanghai has a complicated water network and such network can influence the planning greatly. The solar radiation in Shanghai is also higher, meaning that building a solar array could get more profit.

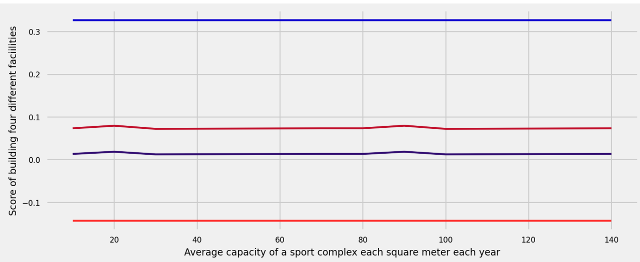
Finally, Shanghai is more developed than the location given. For example, Shanghai has a signal coverage everywhere, which means that we no longer need to consider about the signal coverage rate. The transportation is also more convenient, which means that a good sports complex might get much more people than in the given location. However, the building cost could be higher, since Shanghai is lack of land so that the price of land will be higher.

Overall, the model provides a solid framework for evaluating utility infrastructure installation in various regions, but it may need to be adjusted based on the specific conditions of each location. Factors such as demand for the utility, terrain variance, tree cover rate, and community involvement should be tailored to each region to ensure accurate and effective decision-making.

8 Robustness Analysis



Change the demand



Change the average capacity

We take the change in sport complex's original data as an example. We can see that the line is almost flat, which means the change in sport complex's data does not affect the score of the four different facilities much. Therefore, our model is not sensitive to the given facility's condition, and our model is quite stable.

9 Strengths and Weaknesses

9.1 Strengths

Our model considers three aspects and eight indicators. The metric system is quite complete and rigorous. This enables us to consider all kinds of situations.

Our model can adapt different situations quickly, since our model is built without any context, so we need to think about the aspects and indicators as completely as possible.Our model can fit into different situations, since we have a complete model, we only need to change the basic data of the program or model and we can make decent planning and prediction.

The Average Demand Model makes precise prediction of the IE indicators. It's both creative and rigorous when dividing the circle range into grids and calculating probabilities.

We build prediction models for all indicators being used.

Our data of terrain is very complete. The point cloud dataset enables us to easily model the terrain and divide the land.

Our model is also very stable. Our model doesn't fluctuate very much when some of the given facility's condition changes. This gives our model a higher suitability to different location, countries, etc.

9.2 Weaknesses

The main weakness of the model is that we didn't have any external data to test our model's accuracy. In the future, we need to find some examples that is already planned, and we can test our model based on completed area planning examples. We can also make modifications to our model to make it more precise and adaptive on every kind of model.

Some quantities we model are not objective quantities, such as relative probability and utility. They might include errors.

10 Conclusion

We develop the metric IUL which contain three aspects: IE, IB, and IC. Each aspect contains multiple indicators. We use the **double-level Entropy Weight Method** to obtain the weights of indicators in each aspect, as well as the weights of IE, IB, and IC.

Next, we build prediction models for specific land uses' indicators. The four land uses we consider are: agritourist center, solar array, sports complex, and ranch. We separate them into **Demand-Restricted Land Uses** and **Demand-UnRestricted Land Uses** to build the **prediction models**. We build the **Average Demand Model** for agritourist center and sports complex.Using all the prediction models of indicators together with the metric, we are able evaluate different plans.

We obtain the point cloud data from USGS and the height variance, signal coverage, and tree area of each grid. Then, we use **K-Means model** to divide the land into two clusters by considering three variables: height variance, signal coverage and tree area. After correcting the results of clusters, we divide the land into three parts, and get the areas of each part.

After we obtain the raw data of the three lands with four usages, we correct them according to the suitability of land by considering height variance and signal coverage. At last, we get the final results. Sports complex is the dominating option, but we want the land to be more diverse, so we also take solar array and agritourist center, which are the second and third highest score plans, into account.

We evaluate the impacts of the new factory by taking population change and community utility change into consideration. We increased the population density, since the new factory will support new jobs, resulting in more people living in the county, and therefore the population density increases. We also increased the base point for the solar array, since the new factory will consume a lot of electricity.

Recommending Plan of the Land

Dear employers:

We are here to help you determine the final use, and also the best use of the cute 3 square kilometers of land. We spent a lot of energy on it, and built reliable models that take many factors into accounts.

Here are our recommending plans according to our maths model.

We recommend you to divide the land into three parts, each with a different usage. The lower left figure can give you a general idea how it will be divided.

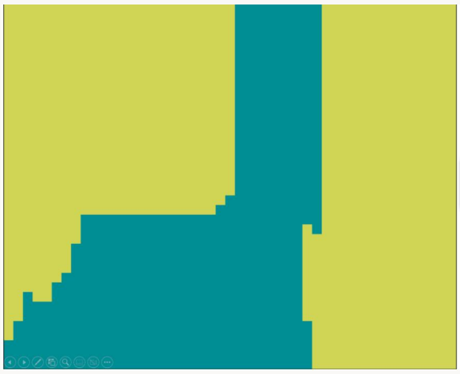
The three usages of land are: sports complex, solar array, and the agritourist center.

Among all the usages of land, sports complex has the highest evaluation for all parts of land. This is because of the high demand of doing sports. However, we want to fully utilize the land by giving it diverse roles. Therefore, the usages that have the second and third highest evaluatios, solar array and agritourist center, are also recommended to you.

We truly hope that after reading our paper, you will have a better insight of determining the land use.

Sincerely,

A group of passionate students



References

- Pollution. Education. (n.d.). Retrieved March 12, 2023, from https://education.nationalgeographic.org/resource/pollution/ [1]
- Top 5 solar farm land requirements. YSG Solar. (2021, June 10). Retrieved March 13, 2023, from https://www.ysgsolar.com/blog/top-5-solar-farm-land-requirements-ysg-solar [2]
- Ost, I. (2023, January 24). How much energy does a solar panel produce? Solar.com. Retrieved March 13, 2023, from https://www.solar.com/learn/how-much-energy-does-a-solar-panel-produce/ [3]
- TNM Download V2. (n.d.). Retrieved March 13, 2023, from https://apps.nationalmap.gov/downloader// [4]