Centroids, Clusters, and Crime: Anchoring the Geographic Profiles of Serial Criminals
(Catchy, but informative title)

Team # 12345
(Dated: February 5, 2024)

A particularly challenging problem in crime prediction is modeling the behavior of a serial killer. Since finding associations between the victims is difficult, we predict where the criminal will strike next, instead of whom. Such predicting of a criminal's spatial patterns is called geographic profiling.

Research shows that most violent serial criminals tend to commit crimes in a radial band around a central point: home, workplace, or other area of significance to the criminal's activities (for example, a part of town where prostitutes abound). These "anchor points" provide the basis for our model.

We assume that the entire domain of analysis is a potential crime spot, movement of the criminal is uninhibited, and the area in question is large enough to contain all possible strike points. We consider the domain a metric space on which predictive algorithms create spatial likelihoods. Additionally, we assume that the offender is a violent serial criminal, since research suggests that serial burglars and arsonists are less likely to follow spatial patterns.

There are substantial differences between one anchor point and several. We treat the single-anchor-point case first, taking the spatial coordinates of the criminal's last strikes and the sequence of the crimes as inputs. Estimating the point to be the centroid of the previous crimes, we generate a "likelihood crater," where height corresponds to the likelihood of a future crime at that location. For the multiple-anchor-point case, we use a cluster-finding and sorting method: We identify groupings in the data and build a likelihood crater around the centroid of each. Each cluster is given weight according to recency and number of points. We test single point vs. multiple points by using the previous crimes to predict the most recent one and comparing with its actual location. We extract seven datasets from published research. We use four of the datasets in developing our model and examining its response to changes in sequence, geographic concentration, and total number of points. Then we evaluate our models by running blind on the remaining three datasets. The results show a clear superiority for multiple anchor points.

This is the most important section of the paper. Make it beautiful. The above was the summary page of a winning entry by Damle, West, and Benzel at CU-Boulder in 2010.

I.   MODEL 2: INTERESTS ANALYSIS FOR STAKEHOLDERS

A.   Technique

We consider five kinds of stakeholders that are susceptible to the change in the water levels in the Great Lakes area with special attention to Lake Ontario. They are property owners, recreational boaters, hydro-power stations, shipping companies, and environmentalists. In order to predict the optimal water levels across the year, we convert their demands into numerous linear relations and use simplex method to solve the following linear program:

$$max \quad z = \sum_i f_i x \tag{1}$$

$$s.t. \quad Ax \leq b \tag{2}$$

$$x \geq 73.5 \tag{3}$$

$$x \leq 77 \tag{4}$$

Here z represents the objective function of this model, which is the weighted sum of normalized interests of all stakeholders, x is a vector which represents water level, A is a matrix that contains all the linear constraints that we take into consideration, and b is a vector that encodes thresholds for all constraints.

During the research, we realize that the magnitude of data varies from stakeholder to stakeholder. Therefore, we bring all the numbers to the same level by dividing both sides of linear equations with appropriate constants. This step is called normalization. We assign different weights to different stakeholders to indicate their relative importance in this model. We decide to let environmental and recreational elements subordinate to daily activities and economic development. The table below shows the relationship between stakeholders and their assigned weights.

Property owners	Hydro-power stations	Shipping companies	Boaters	Environmentalists
0.3	0.3	0.2	0.1	0.1

B.   Property Owners

Flooding has always been a major concern for residents and property holders near the lake areas. From "*Great Lakes Flood Thresholds and Impacts*" (Anthony et al., 1997), we can find the flood damage thresholds for all five Great Lakes (Figure 2).

Then we consider how much money flooding can cost. According to media, severe flooding triggered by high water level usually makes "soaked" neighborhoods spend several million dollars. Inspired by the paper, we divide the potential flooding of a lake into three levels based on the water surface height of the lake, and consider the frequencies of each level (Figure 2). We then associate these different levels of catastrophes with the economic damage caused by them. For the linearity of the model, we keep the cost increasing in a linear way when the water level exceeds a certain threshold. Eventually, we take the negative value of the combined result of these functions and fit it into the objective function we are trying to maximize.

| | ENTIRE RECORD | | POST-1955 | | |
	MinMax²	SD Thresh.³	MinMax²	SD Thresh.³	LTM¹
Lake Ontario					
Toronto	74.17	74.72	74.74	74.72	74.6
Kingston	75.47	75.49			74.6
Lake Erie					
Erieau			174.14	174.48	174.11
Kingsville			174.42	174.61	174.14
Port Colborne	174.41	175.06	175.48	175.35	173.93
Port Dover			174.70	174.57	174.09
Port Stanley	173.71	174.24	174.38	174.57	173.96
Lake Huron					
Collingwood	176.97	177.07	177.01	177.18	176.29
Goderich	176.37	176.80	176.88	177.03	176.28
Point Edward	176.44	176.71	176.86	176.90	176.03
Lake St. Clair					
Belle River	174.66	175.12	175.39	175.52	175.01
Tecumseh			175.47	175.52	174.46

[1]Long-term Mean = long-term yearly average for gauge
[2]MinMax = lowest maximum instantaneous water level at which damage occurred
[3]SD Thresh. = threshold level one standard deviation below the mean of maximum instantaneous flood levels

FIG. 1. Summary table of flood damage thresholds by lake and water level gauge. From Ref. [?]. For citations, use Google Scholar's cite function.

| | DAMAGE LEVEL¹ | | | | | |
| | Minor | | Moderate | | Serious | |
	Freq.	(%)	Freq.	(%)	Freq.	(%)
Lake Ontario	13	50	9	34.6	4	15.4
Lake Erie	18	37.5	14	29.2	16	33.3
Lake Huron	13	52	10	40	2	8
Lake St. Clair	3	12.5	12	50	9	37.5
All Lakes	47	38.2	45	36.6	31	25.2

FIG. 2. Summary of flood frequencies and flood thresholds by lake. From Ref. [?]. For citations, use Google Scholar's cite function.

C. Hydro-power Stations

The Moses-Saunders Power Dam is the major source of hydro-power at Lake Ontario. It supports two hydroelectric power generating stations in that region. A hydro-power station transforms the gravitational potential energy into electricity. With the aim of estimating the power generated by the dam, we quote a formula in physics.

$$P = \eta Q g h \tag{5}$$

P is the electric power generated, Q is the discharge of water in kg/s, g is the acceleration due to gravity, h is the height of water level, and η stands for the efficiency of this process. The ability of generating power is considered as the interest of hydro-power stations and is thus included in the objective function. And we later design a mechanism to control the amount of water discharged at the dam.

D. Recreational Boaters

Water levels have a large impact on the economy generated by boating. Based on an analysis of this issue performed by Cornell University (Nancy et al., 2005), we first estimate the loss of total possible boating days of each possible water level, correspondingly. Then we use several sets of linear functions to deduce the consequential economical impacts on boating for each month in the boating season. Figure 3 shows the correlation.
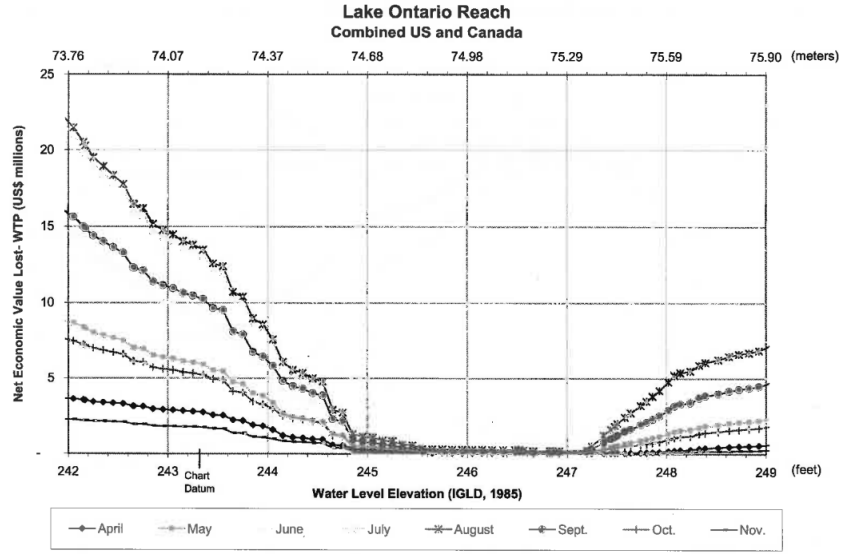

FIG. 3. Relationship between water level and net economic value lost. From Ref. [?]. For citations, use Google Scholar's cite function.

Using the economic data from the same paper, we consider the inflation rate over the past two decades to obtain a reasonable dataset which contains the economic value of boating in the current market. We also applied a more recent boating preference graph of Lake Ontario from Plan 2014 (Figure 4) to our model.

After that, to make our simplex solver capable of processing the information, we fit a piece-wise linear function by breaking the dataset into segments. For each segment, we use the following method to calculate its slope m and shift b:

$$m = \frac{y_2 - y_1}{x_2 - x_1} \tag{6}$$

$$b = y_1 - m \cdot x_1 \tag{7}$$

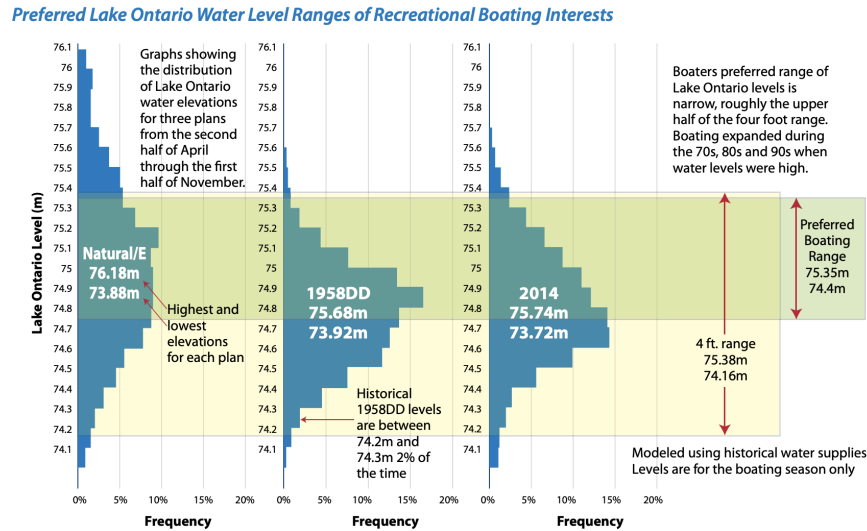The piece-wise linear function is later used a part of the objective function in our linear programming formulation.

Preferred Lake Ontario Water Level Ranges of Recreational Boating Interests

FIG. 4. Preferred boating range: 74.4m to 75.35m. From Ref. [**?**]. For citations, use Google Scholar's cite function.

E.   Shipping Companies

Shipping is a vital industry across the Great Lakes, as it boosts GDP, creates enormous job positions, and supports navigation. Ships all prefer steady and high water, so that they can carry more cargo safely. We examine a paper titled *The impact of water level changes on commercial navigation in the Great Lakes and St. Lawrence River*; in this paper, Millerd mentioned that reducing lake level would increase annual shipping costs by 15 to 33 percent (Millerd, 1996). An earlier study demonstrates the economic effects on commercial navigation caused by several regulatory plans on Lake Erie (International Lake Erie Regulation Study Board 1981), Figure 5 shows the major one of them.

We develop linear equations that calculate the benefits of commercial navigation based on the data provided in study board and by applying the line-fitting technique we use for boaters. Finally, the economic contribution from shipping is added to the objective function.

F.   Environmentalists

The Great Lakes Basin is the home for more than 3,500 species; the diversity of this huge ecosystem makes environmentalists join the team of stakeholders. Since our goal is to predict the optimal water level, we will focus on wildlife that inhabits in the wetlands around the lakes.

Plan		Annual Transportation Cost				Present Worth		Average Annual Amount
		1985		2000		2035		
		$		$		$		$	$
Basis-of-Comparison:								
United States	:	794,596,000	:	1,258,525,000	:	1,610,034,000	:	12,568,500,000	1,086,700,000
Canada	:	553,061,000	:	794,912,000	:	954,839,000	:	8,134,600,000	703,300,000
Total Costs	:	1,347,657,000	:	2,053,437,000	:	2,564,873,000	:	20,703,100,000	1,790,000,000
6L								
United States	:	-494,000	:	-802,000	:	-1,252,000	:	-8,172,000	-707,000
Canada	:	-226,000	:	-392,000	:	-536,000	:	-3,840,000	-332,000
Total Benefits	:	-720,000	:	-1,194,000	:	-1,788,000	:	-12,012,000	-1,039,000
15S								
United States	:	-1,468,000	:	-2,416,000	:	-3,660,000	:	-24,403,000	-2,110,000
Canada	:	-759,000	:	-1,313,000	:	-1,780,000	:	-12,862,000	-1,112,000
Total Benefits	:	-2,227,000	:	-3,729,000	:	-5,440,000	:	-37,265,000	-3,222,000
25N								
United States	:	-4,338,000	:	-7,195,000	:	-10,949,000	:	-72,552,000	-6,273,000
Canada	:	-2,680,000	:	-4,558,000	:	-6,056,000	:	-44,745,000	-3,869,000
Total Benefits	:	-7,018,000	:	-11,753,000	:	-17,005,000	:	-117,297,000	-10,142,000

FIG. 5. Economical impacts on commercial shipping under category 1 of the regulation. From Ref. [?]. For citations, use Google Scholar's cite function.