Geographic Profiling Serial Killers with Rossmo's Formula

Clayton Olsen

December 2020

1 Introduction

If someone is committing crimes around a major city on a regular basis, what is an effective way of finding this person? One way to go about this is to try to find patterns among crime scene locations and make a prediction about where the next one will be. This can be tricky though, because there is likely a large amount of randomness in where a serial offender chooses to commit a crime. In 1996, a criminologist named Kim Rossmo suggested that, instead of trying to predict the location of the next crime, we can try to find out where the culprit lives. Kim Rossmo put forth a formula suggesting the pattern of crime locations is likely correlated with the location of a criminal's home, thus using geographic information about the crime scenes, we can make a prediction on the location of the criminal's home.

2 Rossmo's Formula

Rossmo built a formula on two main principles derived from historical criminal data:

1. A serial criminal will not commit crimes too close to their home.
2. A serial criminal will follow the "least effort" principle, which assumes criminals will not travel farther than necessary to find victims.

Rossmo's formula is a loss function that takes a grid over a map with committed crime locations and assigns a value to each grid space referring to the likelihood the culprit lives within that grid space. Given the first principle, the formula uses the idea of a buffer zone, where grid spaces within a buffer around a crime location will be penalized. Given the second principle, the likelihood granted to a given cell will decay with distance away from a crime site. The formula is as follows:

$$p_{i,j} = k \sum_{n=1}^{\text{total crimes}} [\frac{\phi_{i,j}}{(|X_i - x_n| + |Y_j - y_n|)^f} + \frac{(1-\phi_{i,j})(B^{g-f})}{(2B - |X_i - x_n| - |Y_i - y_n|)^g}]$$

$$\text{Where: } \phi = \begin{cases} 1, \text{if} |X_i - x_n| + |Y_j - y_n| > B \\ 0, \text{if} |X_i - x_n| + |Y_j - y_n| \leq B \end{cases}$$

and:

1. $P_{i,j}$ = The score value representing probability at point i,j
2. ϕ = The weighting factor
3. k = Empirically determined scaling constant
4. B = Buffer zone, determined by 1/2 * average nearest neighbor
5. n = Number of crime sites
6. f = Empirically determined exponent (steepness of decay after buffer), recommended 1.2 from Rossmo
7. g = Empirically determined exponent (steepness of decay before buffer), recommended 1.2 from Rossmo
8. x_i, y_j = Coordinates of grid space i,j
9. x_n, y_n= Coordinates of the n^{th} incident location

There are two main parts of the lost function. The first half of the equation is used when $\phi = 1$ or when $|X_i - x_n| + |Y_j - y_n| > B$. What this means is that, if the selected grid space's summed distance, of x and y, away from nth incident location, is greater than the buffer, then the first part of the function is used. Since $|X_i - x_n| + |Y_j - y_n|$ is in the denominator, the farther the distance from an incident location, the lower the value it will produce. The second part of the equation accounts for the principle that criminals do not commit crimes close to their home either. That same equation $|X_i - x_n| + |Y_j - y_n|$ is subtracted from $2B$ which is two times the buffer. So when the grid space is within the buffer, the closer the grid space is to the incident spot, the lower the value will be. So, with these two truncated halves of the equation, we can see that the loss function penalizes grid spaces that are very close to incident locations, and spots that are very far from any incident locations. Running the decay function in R results in a score value where a higher score indicates a higher probability that the point contains the offender's anchor point. For map 1, I used f of 0.2 and g of 10 due to the long distance between the more northern crime locations with the clustered locations. This put a larger penalty for area within the buffer area for individual cases.

3 Albert DeSalvo, The Boston Strangler

In order to utilize this lost function, we need data with a relatively high number of cases committed in a limited amount of time. Additionally, the model is much more effective when these crimes are committed in a condensed area. For this reason, the methodology is primarily used for major city or urban data.

To test the loss function on real data, I utilized the data for Albert DeSalvo, The Boston Strangler. Over a two year span, 1962 - 1964, DeSalvo committed 13 murders within the greater Boston area. This data set is stored in the rgeoprofile package of the CRAN library and consists of a data frame with the name of murder victims, their ages, and the longitude and latitude of their incident locations. I implemented the Rossmo formula on the data to create raster data, with Rossmo's scores distributed across a 3000x3000 pixel area over Boston which encapsulates an area around all of the crime sites related to DeSalvo. In addition, I have the location of DeSalvo's house and the location of the minimum of the euclidean distance from the crime sites for comparison. The map of Boston comes from the Esri Leaflet database which has a world map on the GCS WGS 1984 coordinate system.

4 Other formulations of Rossmo's formula

Since Rossmo publicized his lost function, other variations of decay have been proposed, which vary the relationship between distance from the buffer and decay. Rossmo's formula treats decay exponentially, with the parameters f and g signifying the exponential rate at which probability decay's with distance from the buffer zone. For comparison, I implemented the negative exponential decay function and the linear decay function as well.

4.1 Negative Exponential Decay

This model assumes that likelihood decreases exponentially from the incidents of crime using euclidean distance for the likelihood calculation beyond the two buffer system.

$$P_{i,j} = 0 \text{ if, } d_{i,j} < A \tag{1}$$
$$= B \text{ if } A \leq d_{i,j} < B \tag{2}$$
$$= C \exp\left(-b d_{i,j}\right) \text{ if, } d_{i,j} \geq B \tag{3}$$

1. $\text{P}_{i,j}$ = The score value representing probability at point i,j
2. A = Buffer zone = $\frac{1}{2}$ nearest neighbors
3. B = Plateau transition = 2$*$ buffer zone
4. b = Exponential constant
5. C = Empirically determined constant
6. $d_{i,j}$ = Euclidean distance from $cell_{i,j}$ to $incident_{i,j}$

With this model, A and B are both buffers where the B is just twice the distance of A from the incident sites. This model treats the likelihood within the first buffer A from a site at point i,j as $P_{i,j} = 0$ and the distance between the first and second buffer as the value of B. So, the area between buffer A and buffer B should have the largest contribution to the summed likelihood from incident site i,j. The cells beyond buffer B decrease exponentially with distance from buffer B. For this model, I used a b of -0.2 and an a value of 0.5 to account for the crime location clustering.

4.2 Linear

The linear decay model is a simpler model that does not incorporate a buffer and just decreases the likelihood with an increase in euclidean distance from a crime incident. The formula is as follows:

$$P_{i,j} = A + bd_{i,j} \tag{4}$$

1. $P_{i,j}$ = The score value representing probability at point i,j
2. A = Intercept, default = 1.9
3. b = Slope of linear decay
4. $d_{i,j}$ = Euclidean distance from $cell_{i,j}$ to $incident_{i,j}$

This lost function will act as a base function to see if Rossmo's buffer theory holds up versus a simple linear loss function that only incorporates the distance from a crime site. For this model, I used an a-value of 1.9 and a b-value of -0.05.

5 Comparison of Lost Functions and Takeaways

The raster data produced from the lost functions is displayed on the maps 1, 2, 3. Looking at the maps we can see the weights the different formulas put on distance decay and the effect of the buffers around crime sites. We can compare the raster values for the cell containing DeSalvo's home with the mean of the likelihood distribution to get an idea of the accuracy. In the table below, are the values for the cells containing DeSalvo's home for each given model compared with the mean score of the entire distribution, the standard deviation, and a z-score normalization to compare the results of the three methods on the same scale:

Methods	Means	Score	SD	z.score
Rossmo's Exponential	7.67	8.52	0.62	1.37
Negative Exponential	1.15	2.06	0.95	0.96
Linear	15.69	19.30	2.85	1.26

We can see that for the DeSalvo murders, the exponential decay function was able to produce the highest relative probability for the correct grid location. Despite not including a buffer, the linear function outperformed the negative exponential function. This is likely due to the clustering of the crime sites on the southern side of the map which influenced the likelihood greater than the other crime sites. Since DeSalvo lived closer to these sites, the lack of a buffer did not seem to be as influential as the pull from the further sites were in the negative exponential model. The negative exponential model has constant probability for values within the two buffers which added higher values for cells within the two buffer radius for crime sites, even the ones that were far away from the main cluster.

Looking at the maps we can see that the euclidean mean of the crime sites is located very close to the actual location. In this kind of situation, exponential decay function had the best predictions. Understandably, situations where the serial killer acts most consistently with the two principles that motivated the loss function, the exponential decay function will be a better predictor.

6 Limitations

While this theory presents an interesting argument for geographic profiling, there are a lot of limitations for practical use. To have relevant likelihood estimations, the decay function requires a relatively high number of cases within a short amount of time. This data set had 13 cases within two years, which may be an unreasonable amount of data in a normal circumstances, so the benefit in the case of a low amount of crimes is limited. The formula also depends on killers following the two major principles, which makes practical assumptions for the killers strategy that may not be truthful. So, the model will not be useful against impulsive crimes/criminals. If there are multiple killers in the same area that are not differentiated in the data, the results will be thrown off course. There can also be other factors such as the crime sites being in a lucrative neighborhood that may diverge the killers actions from the assumed principles. Ultimately, if the serial criminals stray away from the assumed principles, the utility of this geographic profiling method will be minimal.

7 Conclusion

Since his formula's induction in the 1990s, Rossmo has started a company called Environmental Criminology Research Inc.(ECRI) which utilizes geographic profiling techniques for serial crime analysis. Geographic profiling has become more commonplace as a competitor to traditional psychological profiling and has motivated the induction of ECRI's Geographic Profiling Analysis training program which insures that Geographic Profiling remains recognized as a legitimate law enforcement tool, creates a path for meaningful certification for crime analysts, and establishes a code of conduct surrounding this investigative tactic. The utility of geographic profiling may prove more valuable with growing collections of data and more developments in patterned criminal behavior. Beyond crime investigation, the buffer idea has been expanded to other areas of study such as tracking predatory animals like white sharks. It would interesting to see if neural nets could be utilized to develop the parameters for the loss function to apply a more situation specific model. Neural nets may be able to capture patterns that account for situations like clustered points or samples taken over a longer stretch of time.

References

[1] Arcgis.com. 2020. Story Map Journal. [online] Available at: https://www.arcgis.com/apps/MapJournal/index.html?appid=9371ac93690c42cd855836a5567a66d5 [Accessed 14 December 2020].

[2] Artsandscience.usask.ca. 2020. Rossmo'S Formula - Magazine - Arts And Science - University Of Saskatchewan. [online] Available at: https://artsandscience.usask.ca/news/magazine/Spring_2018/rossmos-formula.php [Accessed 14 December 2020].

[3] Rossmo, K. (n.d.). Geographic Profiling and COVID-19 - See News. Retrieved December 14, 2020, from http://ecricanada.com/

[4] Spaulding, Jamie, and Keith Morris. Geographic Profiling Methods for Serial Crime Analysis, 28 May 2020.

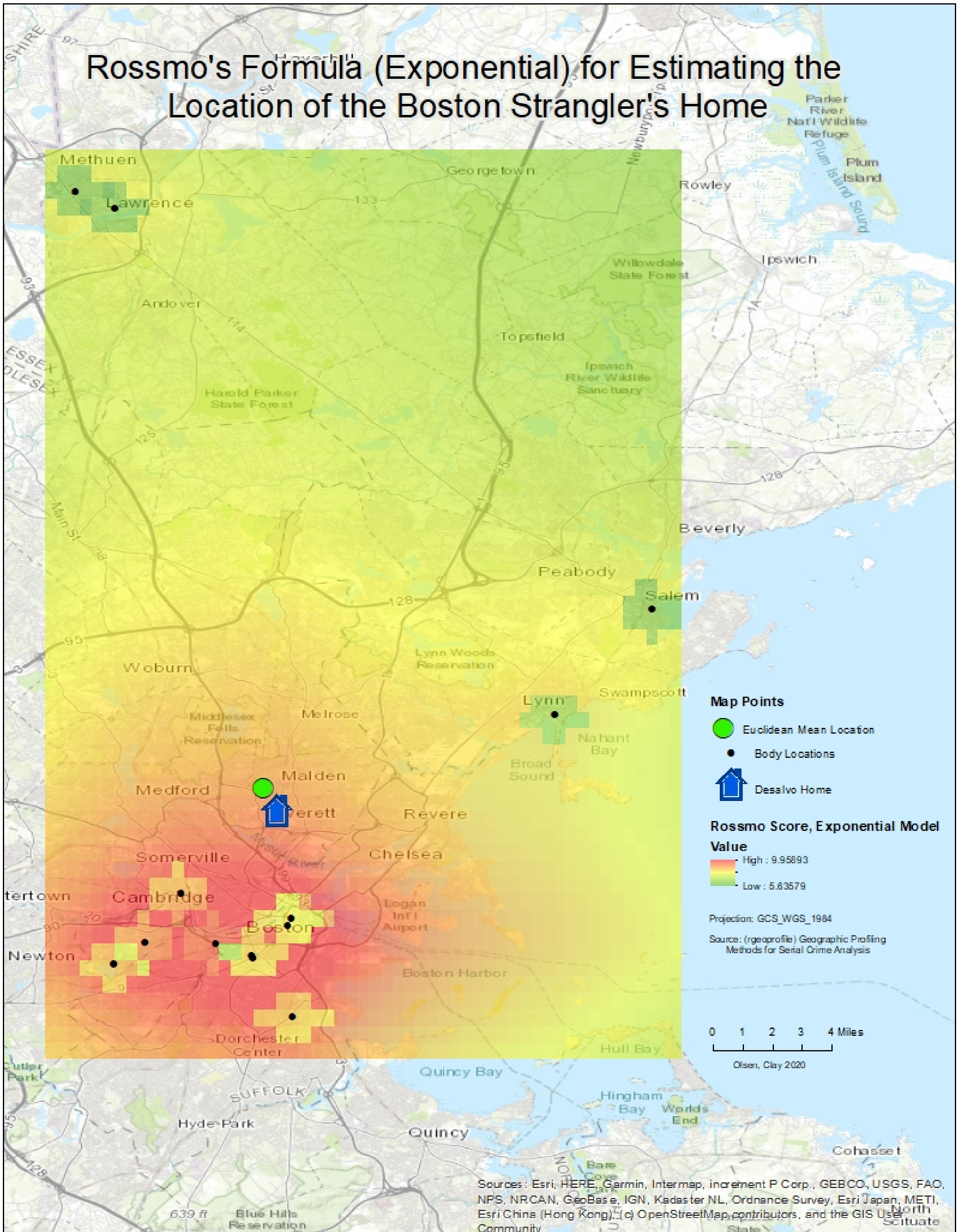

Figure 1: Exponential Formula

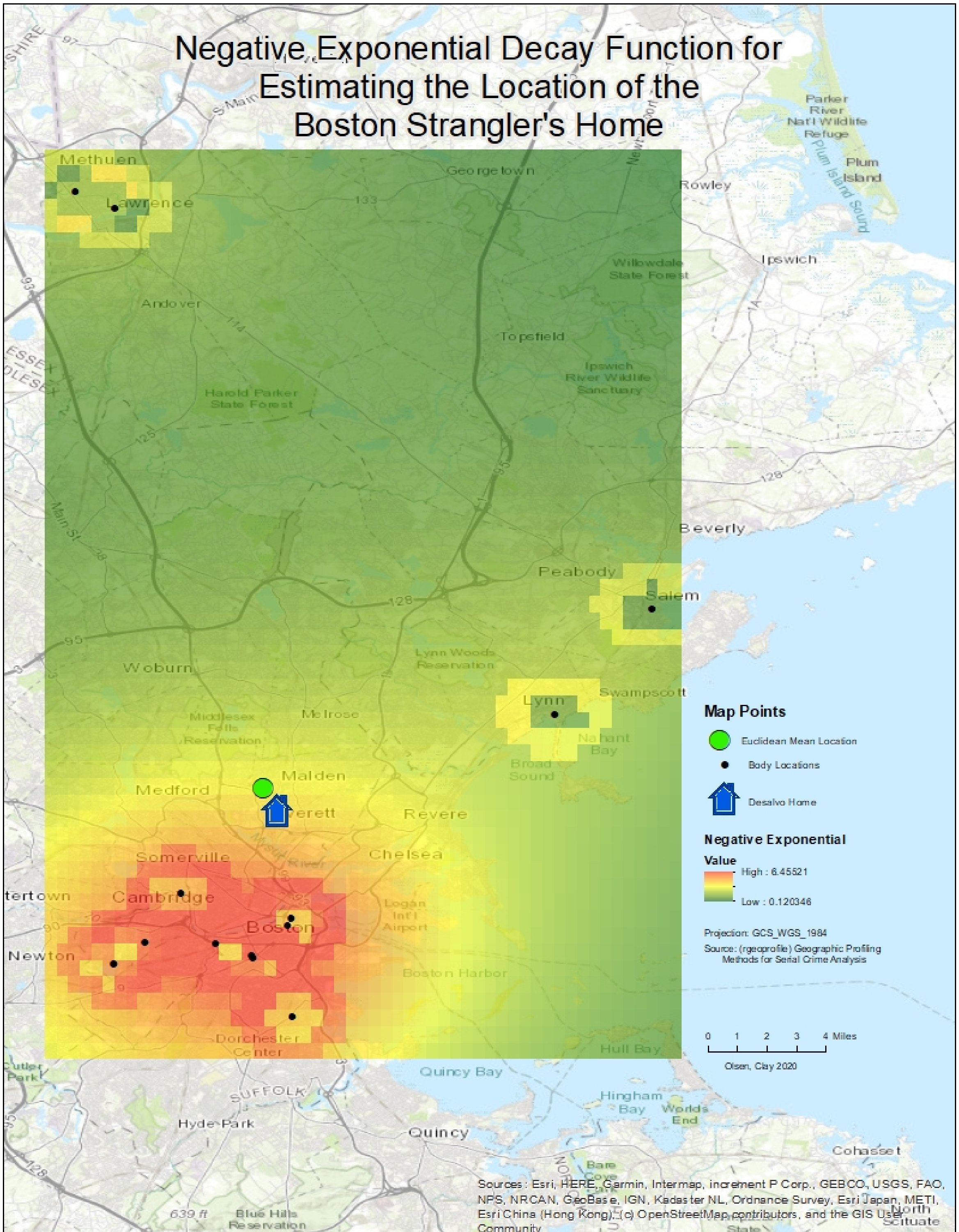

Figure 2: Negative Exponential Formula

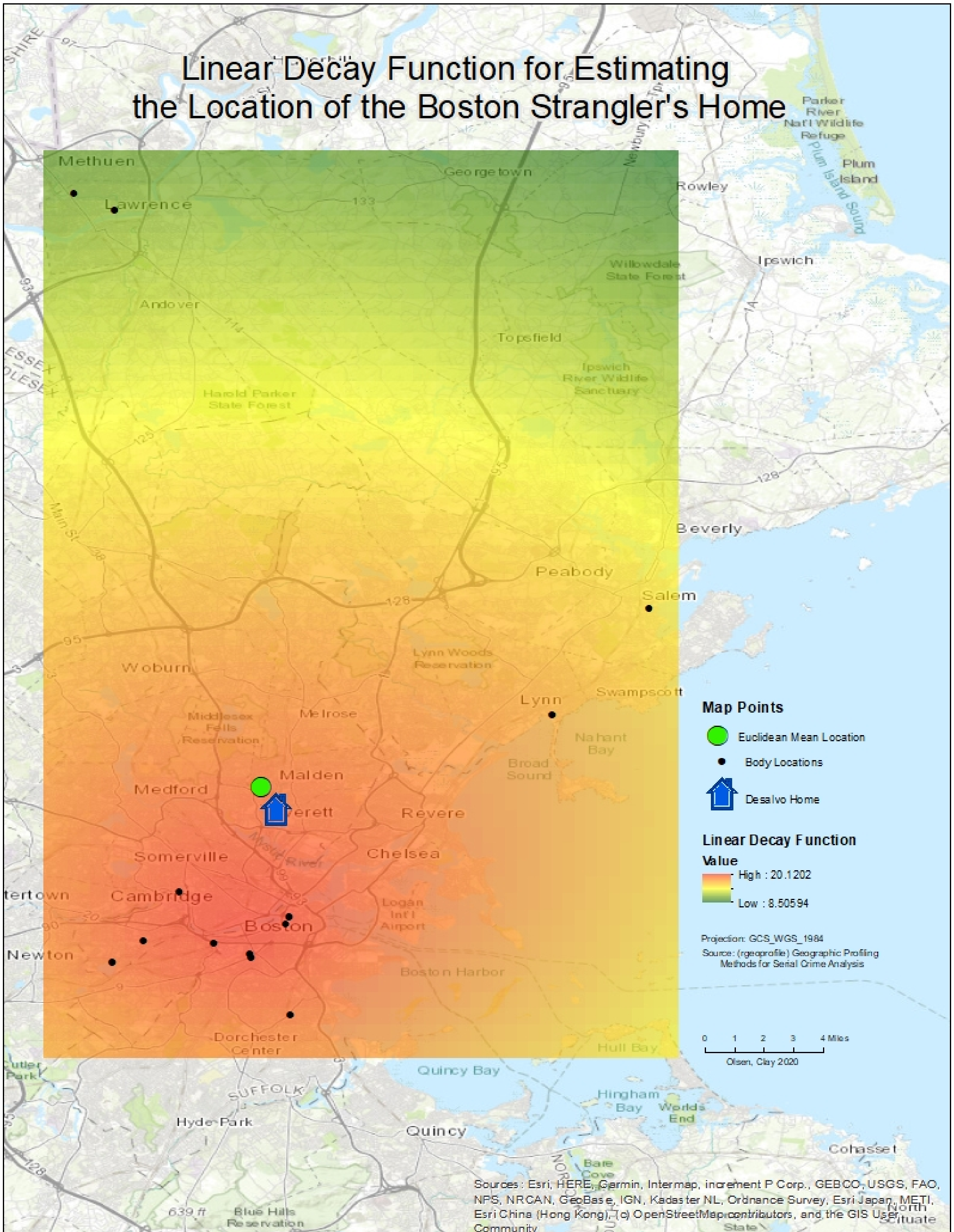

Figure 3: Linear Formula