

Predicting Crime Levels with Cellular Automata

Part 1. Background

There have been many attempts to model the geographic distribution of crime and which environmental and social factors affect crime levels. These models range from statistical models to system dynamics models to geographic models. Space-time clusters were used by Bowers and Johnson (Ref. 1) to draw some conclusions regarding the likelihood of additional crime in a region near a recently committed crime. Some, such as the Risk Terrain Modeling of Caplan and Kennedy (Ref. 2), combined the environmental factors of a specific location with the social information to look at the spatio-temporal correlations between successive robbery attempts.

Many of these studies draw a number of similar conclusions. Namely 1) crimes clustered within 1-2 months of a previous crime, 2) crimes tended to cluster up to 300-400 meters of a prior crime, 3) repeat victimization tended to occur in more deprived areas, 4) space-time clustering was more evident in affluent areas.

Thus any attempt at modeling crime has to allow crime to go up and go down in both time and space.

One area that seemed to have been explored but not matured was the use of cellular automata to model and predict crime. Liang and Liu in ref. 3, for example, describe the utilization of cellular automata in crime prediction but don't go on to show simulation or prediction data.

In this work, I utilize a cellular automata based upon police reporting districts and their associated historical crime levels to model and predict crime levels for a medium sized city of approximately 500,000 people. In addition, the effects of policing and public "targeting" of crime levels is explored for their effects on predicted crime.

Part 2. Methodology

Cellular automata are grids of cells that evolve over time based upon a set of rules applied to neighboring cells. Though conceptually simple, cellular automata can model extremely diverse and dynamically rich phenomenon. The values of cells, which would represent a physical phenomenon, vary in both time and across the geometry of the grid.

The city of Long Beach, California is a city in southern California with a population of 450,000 people. Though considered a suburb of Los Angeles, it is a city within itself with its own city government, police department, and fire department. Besides containing a number of zip codes and census tracts, the city has 316 Police Reporting Districts (PoRD) for which crime is reported. Since crime is reported at the Police Reporting District level (Ref. 4), these districts can be used as "the grid" for a cellular automata model.

Figure 1 shows the city of Long Beach and its 9 city council districts. These districts would be too large to use in a cellular automata model. In addition, you wouldn't want to just lay in a grid on top of the city council districts since you wouldn't have a good past history of crime levels for that grid.

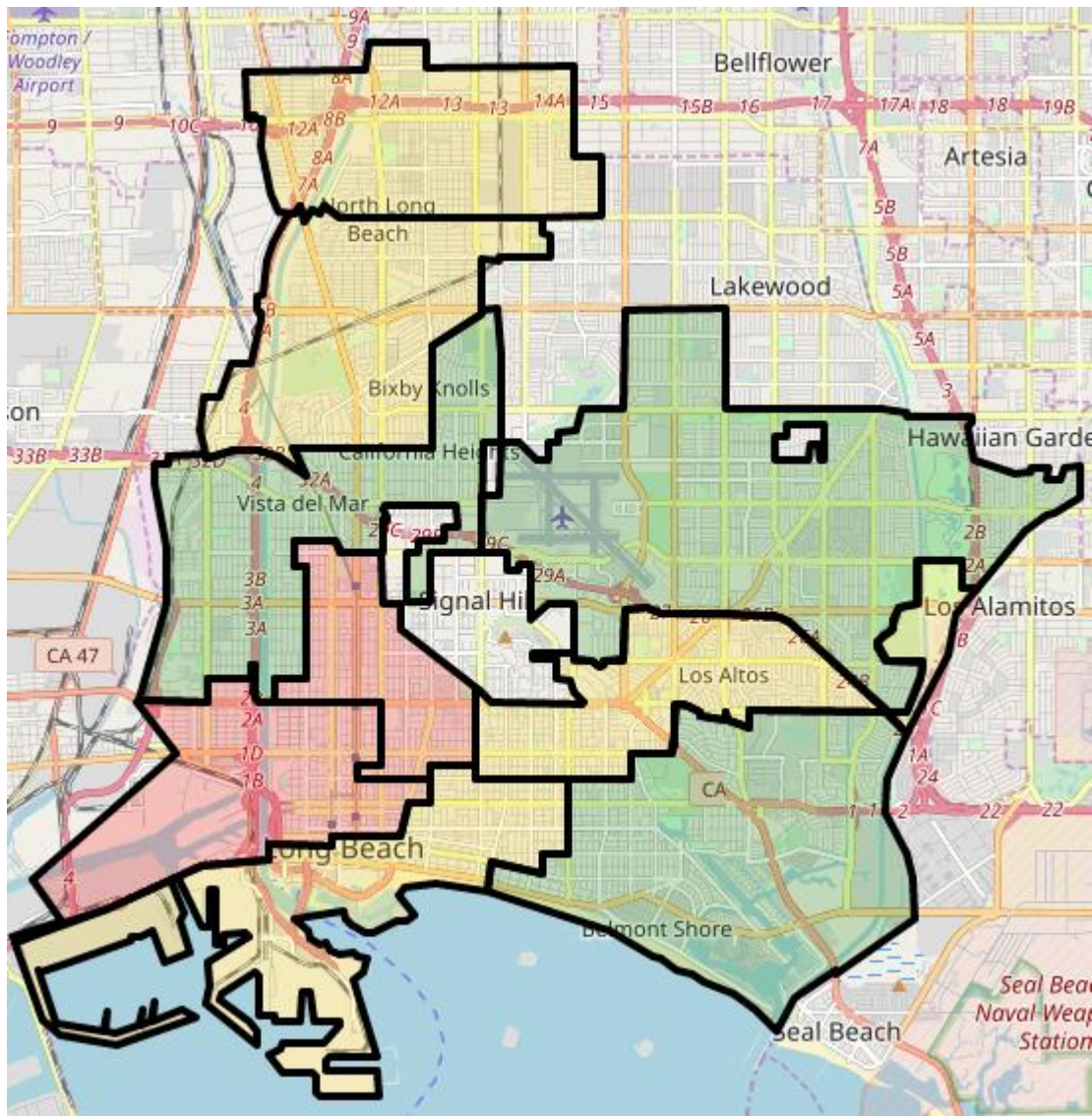


Figure 1. Long Beach City Council Districts

Figure 2 shows the 316 Police Reporting Districts in the city of Long Beach. Now you have both enough interaction between districts as well as past data to utilize for a cellular automata model. Figure 2 is color coded for the crime level in the last reported month of October 2019. The red areas indicate high crime levels and green indicating low crime levels.

From Figure 2 you can see a number of pockets of high crime areas. The first being the downtown area of Long Beach, the second being the North Long Beach area, and two smaller areas in the Anaheim corridor and the Belmont Shore area.

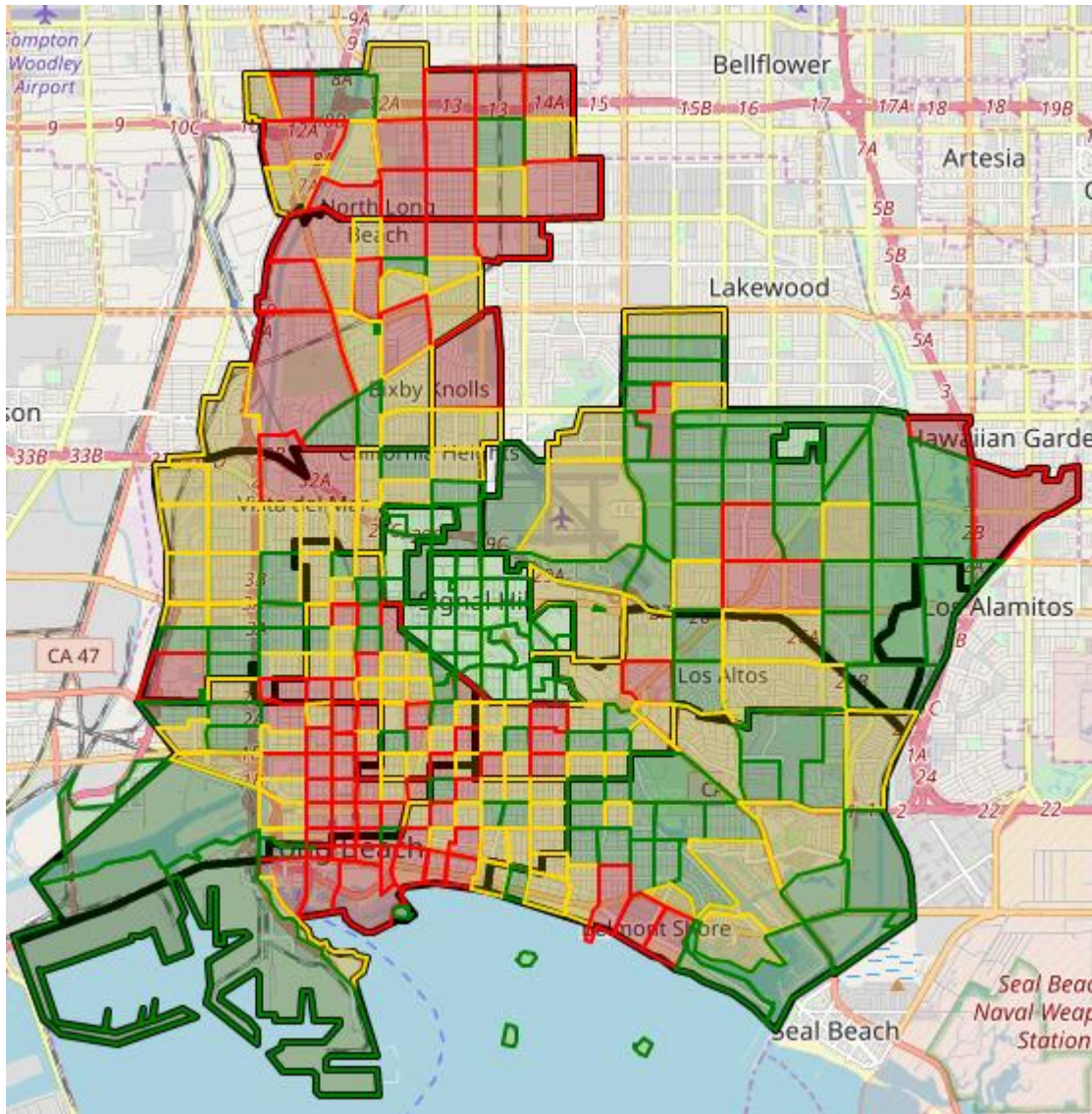


Figure 2. Long Beach Police Reporting Districts

The equation that determines a Police Reporting District crime level for future months is

$$C_{i,M+1} = C_{i,M} + \sum_{j=0}^N k(C_{i,M} - C_{j,M})/N$$

Where $C_{i,M+1}$ is the crime level in PoRD i for month $M+1$, $C_{i,M}$ is the crime level for PoRD i for month M , $C_{j,M}$ is the crime levels for the N PoRDs that neighbor or touch C_i , and k is an interaction factor.

This equation fits in with the understanding that crime is localized. That is, high crime areas breed high crime in nearby areas. This equation also fits in with the understanding that high crime areas in previous months lead to high crime in future months (unless something is done to reduce the crime-inducing factors). The equation has the parameter k that can be used to best match the crime characteristics of a given city or municipality.

Since the cellular automata was seeded with past crime levels, the environmental factors that contribute to crime such as the number of vacant housing units, the presence of gas stations and liquor stores, and grocery stores or bodegas are embedded in the cellular automata.

With this framework, a Python script was written to generate the anticipated crime levels for each of the 316 PoRDs in Long Beach for six months in the future. In addition, the monthly history for each PoRD was generated to clarify the results of the cellular automata. These results inherently assume that there is no specific police or neighborhood action taken to reduce the crime levels.

Furthermore, a number of cellular automata runs were made assuming some police and neighborhood action is taken. For example, let's say that the police and neighborhood residents decide that the crime level in their neighborhood is too high. Thus both groups blanket a single PoRD with extra police patrols and community members standing on every street corner to try and scare off criminals. (This falls in with the Risk Terrain Modeling approach where the focus is on places rather people.) Assuming the action works, a single PoRD crime level is reduced from red (> 12 crimes per month) down to green (< 6 crimes per month) for a single month. What affect does that have on future months? Which PoRD district should they target to have the maximum impact on crime levels?

Part 3. Results

A number of runs were made for the cellular automata model of crime.

A "baseline run" was made using the last measured crime levels (October 2019, Ref. 4) as the starting point for each PoRD and using an interaction factor, k , of unity. Crime levels were predicted for the next six months using the cellular automata level. Overall city crime went from 2727 monthly crime incidents (in October 2019) to 2721 (for April 2020). Thus the overall crime level changed very little. However, the distribution of crime changed significantly. Figure 3 shows the final monthly crime levels (April 2020) for the baseline run. Note that the Anaheim corridor and Belmont Shore areas of crime went from red to yellow. However, the downtown and North Long Beach areas solidified and expanded slightly to the red high crime areas. (Animated gifs for each predicted month of crime are available.)

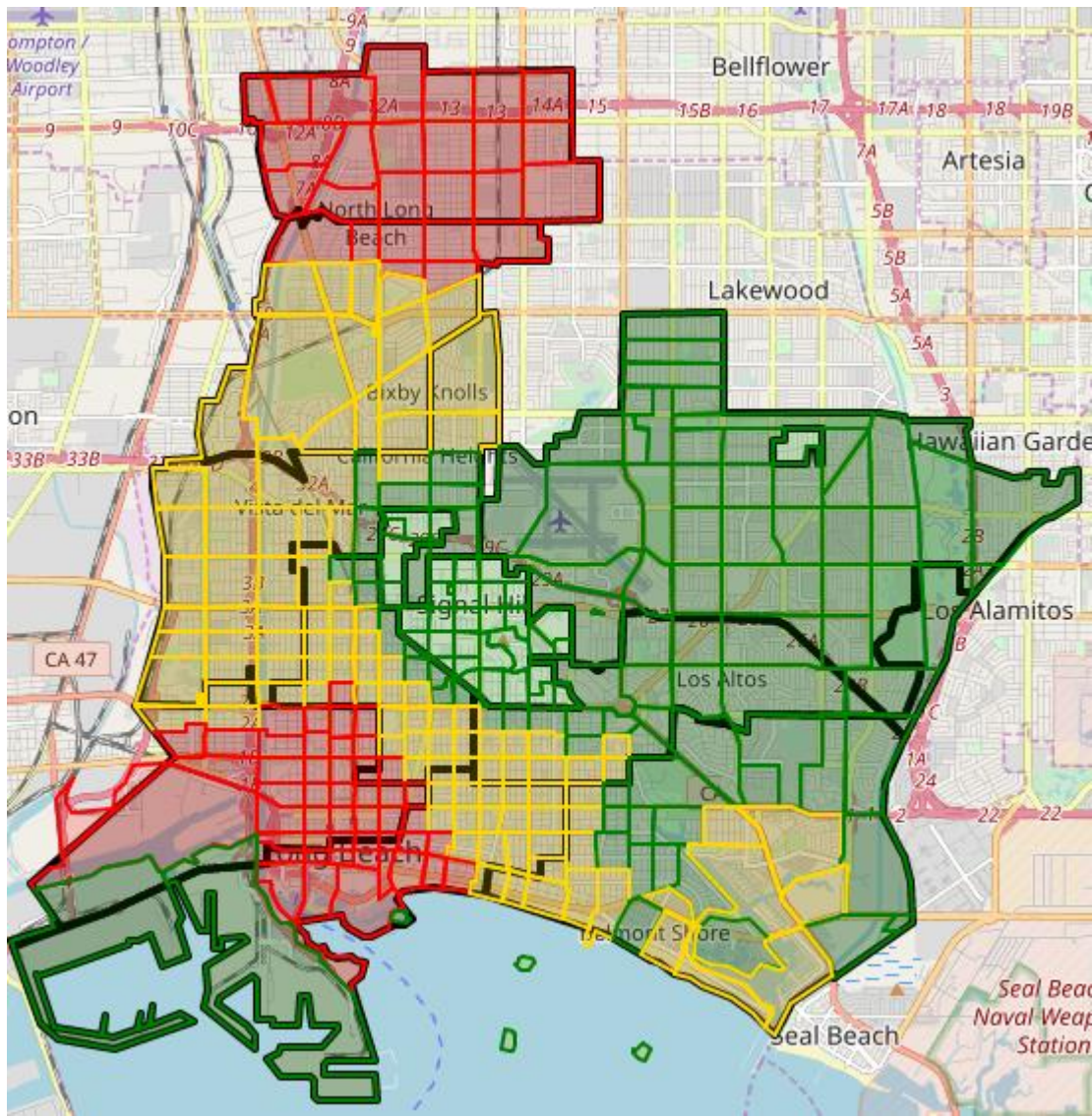


Figure 3. Baseline Run, Predicted Crime Levels

There were a number of simulations/predictions that I ran in order to see the effects of intense police and neighborhood activity on crime levels. Figure 4 shows a zoomed in view of the north Long Beach crime activity. Annotated to the figure are the PoRD identification numbers. All of these districts are red, high crime areas.

What were to happen if the police and neighborhood members coordinated an intense policing effort in one (or more) of these districts for one month and brought the crime down? Would the single district crime reduction “spill over” into subsequent PoRDs or would the effort be dissipated once the police left? And which PoRD would produce the greatest and longest lasting crime reduction?

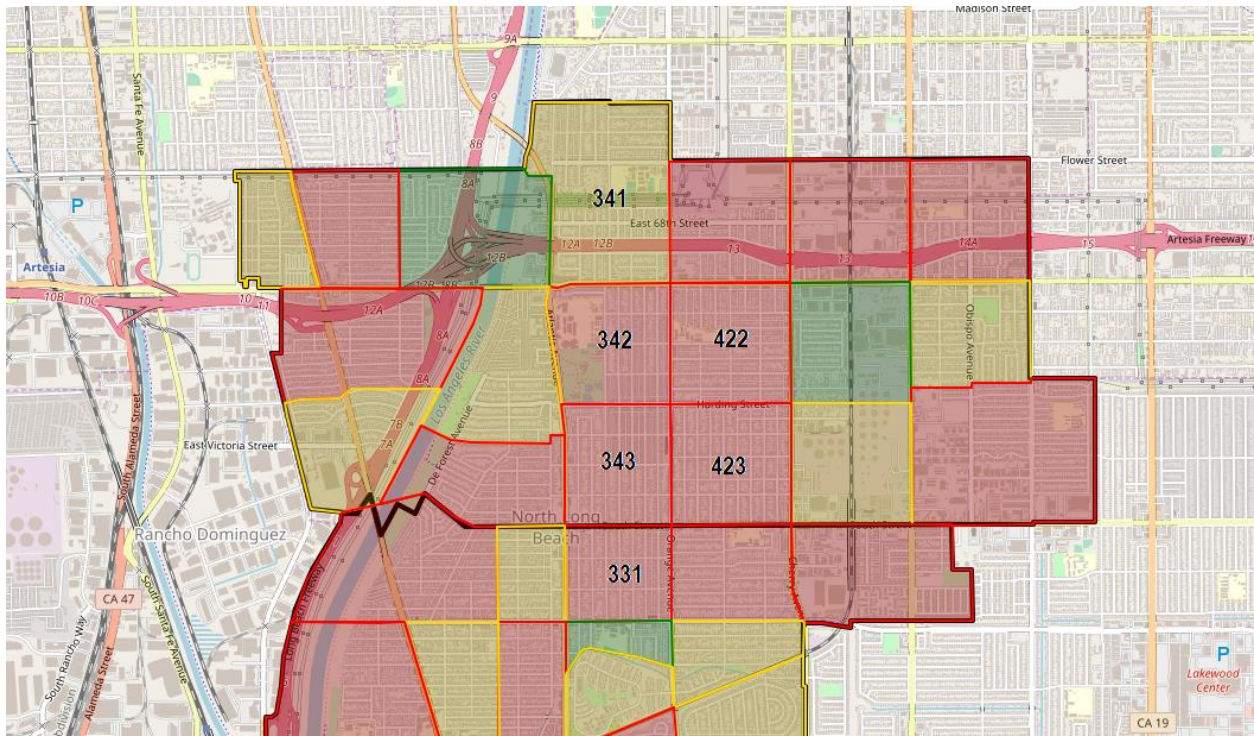


Figure 4. North Long Beach Crime Activity (October 2019)

Single PoRD intense policing for each of the districts annotated in Figure 4 were conducted. In these cases, the October 2019 crime level for the selected PoRD was set at six crime incidents (into the green region) and the next six months of citywide crime computed.

As expected, crime levels in PoRDs removed geographically from the selected PoRD were unchanged. PoRDs near the selected crime district were affected, but not to a significant degree. Figure 5 shows a summary of these results at the end of month six.

PoRD Policing	CityWide Crime	341 Crime	342 Crime	343 Crime	331 Crime	422 Crime	423 Crime
-							
341	2715	14	15	14	14	14	14
342	2696	13	14	14	13	13	14
343	2709	14	15	14	13	14	14
331	2708	14	15	14	13	14	14
422	2704	14	15	14	13	13	14
423	2700	14	14	14	13	13	14

Figure 5. Summary of Single District PoRD Extra Policing

Figure 6 shows the resulting crime distribution from the run where PoRD 342 was set at six crimes per month to start the cellular automata simulation (which resulted in the lowest citywide crime levels). Some north Long Beach PoRDs have turned yellow with lower crime levels than the baseline run, but not significantly lower.

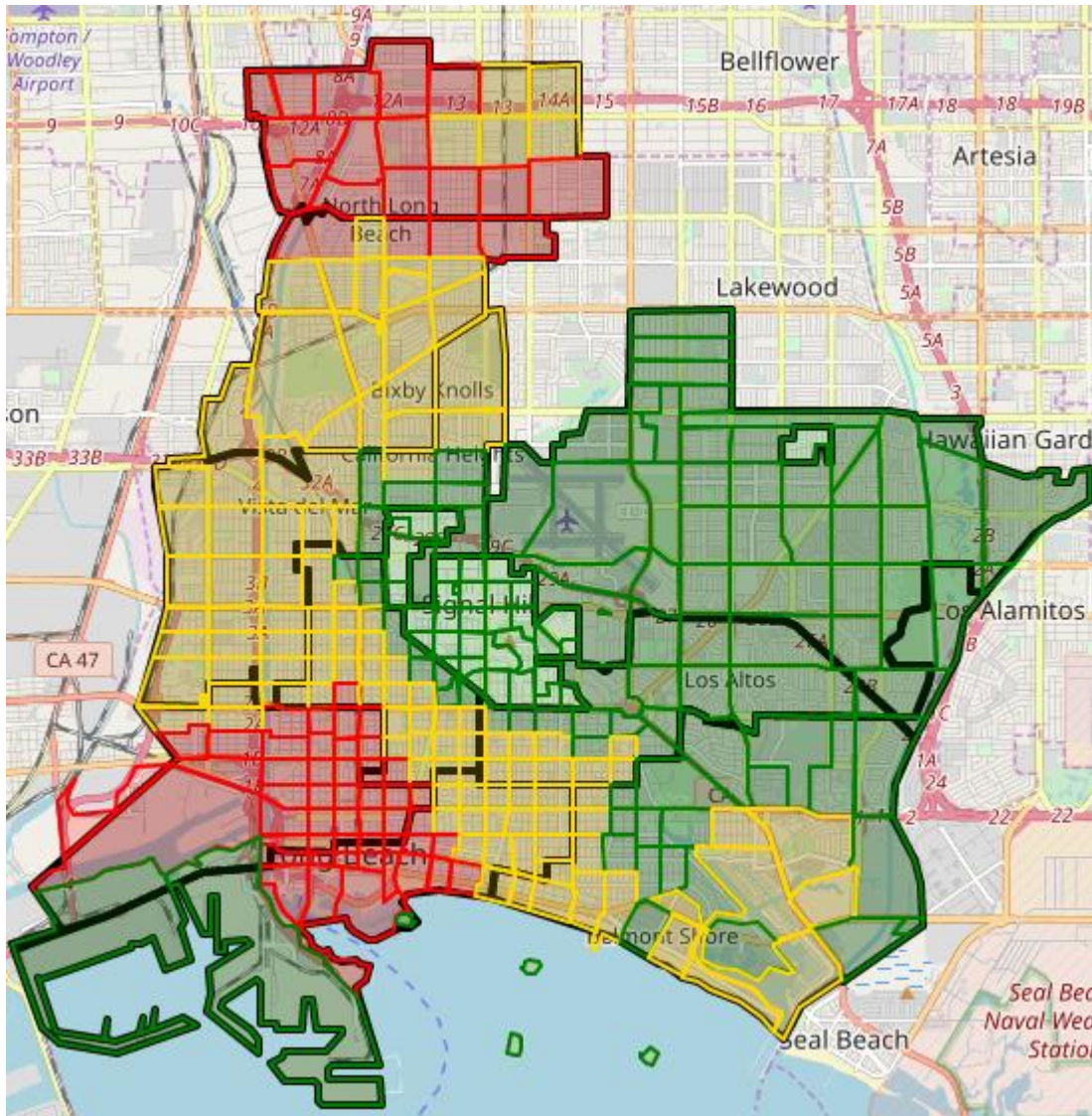


Figure 6. PoRD 342 Intense Policing, Predicted Crime Levels

It was disappointing to see that an intense effort in a single PoRD did not significantly affect nearby PoRD crime levels. In fact, the nearby PoRDs with higher crime levels brought the specified PoRD back up to a higher level.

Finally, a run was made where PoRDs 331, 341, and 342 were set to six crimes for October 2019 to start the simulation. This represents an intense police and neighborhood action across the three vertically-aligned PoRDs shown in Figure 4. In this case, citywide crime dropped from 2692 at the start of the simulation to 2672 at the end of month six. The resulting crime distribution is shown in Figure 7. Note that the large areas of red high crime levels in north Long Beach have been reduced to a majority of yellow medium crime levels and two pockets of red high crime areas. This run shows promise at being able to reduce crime through targeted police and neighborhood action, though the cost of intensely covering three PoRDs with many police and neighborhood personnel is likely to preclude this approach.

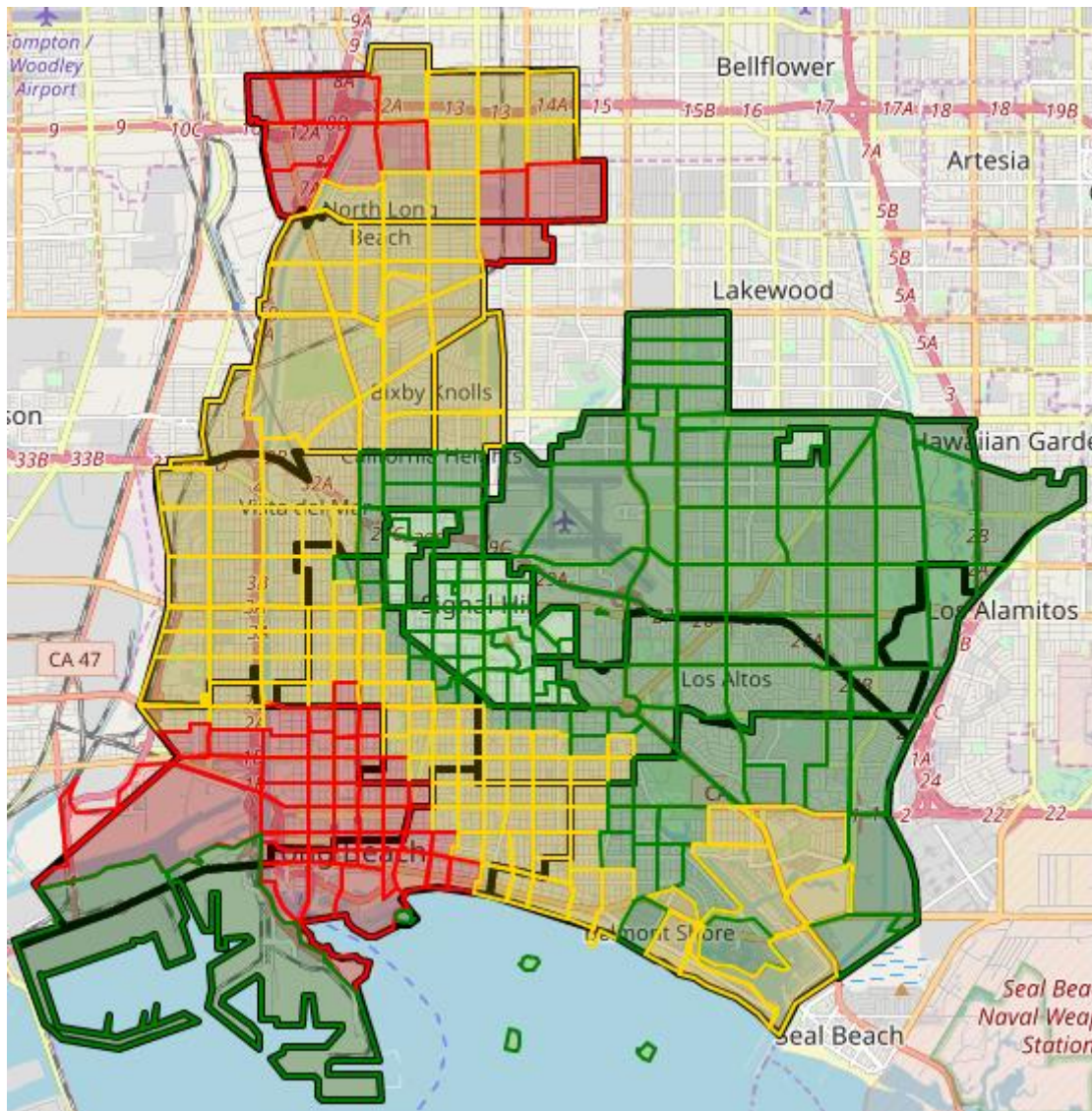


Figure 7. Multiple PoRD Intense Policing, Predicted Crime Levels

For completeness, Figure 8 shows the crime time history in PoRD 342 for this last run. The October 2019 crime level was “seeded in” at six crimes per month to represent an intense policing and neighborhood activity. And the subsequent monthly crime levels go up slightly as a result for neighboring effects. But they never reach the crime levels for that PoRD from the first nine months of 2019.

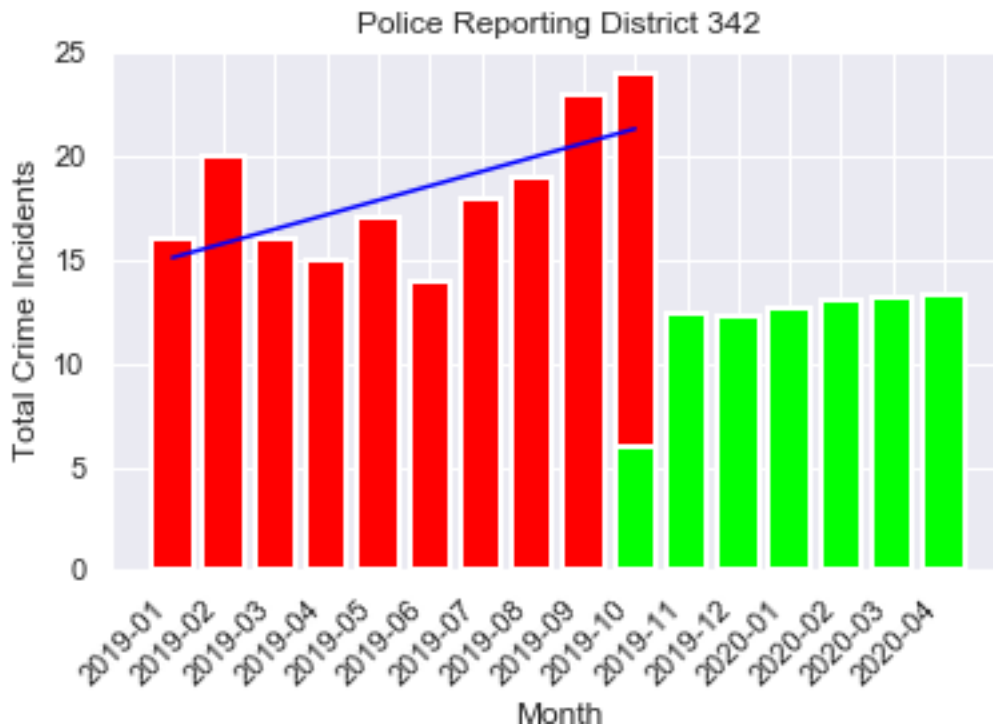


Figure 8. PoRD 342 Monthly Crime Levels

Part 4. Commentary

This effort showed the use of a cellular automata model to match and predict crime levels. The cellular automata has environmental factors built into the model as a result of using historical crime data to start the simulation. Using simple neighboring crime district rules, reasonable results are obtained.

Next I'll be exploring the effects of a non-unity, deterministic interaction factor, k , on crime levels and distributions as well as a non-deterministic k .

References

1. Bower, K. J. and Johnson, S. D., "Domestic Burglary Repeats and Space-Time Clusters: The Dimensions of Risk", *Eur J Criminology*, 2:67-92, 2005, <https://journals.sagepub.com/doi/10.1177/1477370805048631>.
2. Caplan J.M. and Kennedy, L.W, "Risk Terrain Modeling Manual: Theoretical Framework and Technical Steps of Spatial Risk Assessment for Crime Analysis", Rutgers Center on Public Security, 2010.
3. Liang, J. and Liu, L., "Simulating Crimes and Crime Patterns Using Cellular Automata and GIS", 2001, <https://www.semanticscholar.org/paper/Simulating-crimes-and-crime-patterns-using-cellular-Liang-Liu/5acce16ddf27d12689978023af08d29c6dd2a19>.
4. Long Beach Crime Statistics, <http://www.longbeach.gov/police/crime-info/crime-statistics/>