

A Localized Method for Predicting Crime Levels

Part 1. Background

“All crimes are local.”

Whether you agree with this statement or not, there are many different methods to model historical crime and to try and predict future crime. These methods range from statistical methods to system dynamics methods to geographical methods. Some of these methods look at certain geographical areas and treat them independent from other areas and others consider each geographical area connected to some or all other geographical areas.

Most of these models match fairly well-known characteristics of crime (Ref. 1). 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.

In Reference 2 I put together a cellular automata model for predicting crime levels. Cellular automata calculations assume that the crime in one geographical area (i.e., a cell in cellular automata terminology) is related to and affected by the crime in neighboring areas. The localization effect may only regard neighboring areas, as in Reference 2, or it may include the effects from further removed geographical areas (with lowered influence factors).

The effort described here relies on a different methodology called the Independent Crime District (ICD) methodology. That is, the crime in a geographical district or area only depends on the features and inherent characteristics of that specific district. There is no direct connection to neighboring geographical areas. The methodology relies on matching a number of damped oscillators to past crime levels for a given geographic area and then projecting forward.

This ICD methodology was developed to overcome the perceived drawback of cellular automata models which do not rely on past levels of crime. The cellular automata model simply starts at a level of crime for each of the geographical districts (or cells) and updates the crime in these geographical areas based upon the crime levels in neighboring areas. There is no accounting for historical crime levels.

The ICD methodology also has the advantage that it allows crime to go up or down in both time and space as opposed to a linear regression model, for example, which has a constant upward or downward slope.

In this work, I utilize a ICD methodology 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.

Part 2. Methodology

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. 3), these districts can be used as the basis for any number of approaches to crime modeling and prediction.

Figure 1 shows the city of Long Beach and its 9 city council districts. These districts would be too large to use in deriving a crime model. The city council districts cover a large area and have diverse land use and populations.

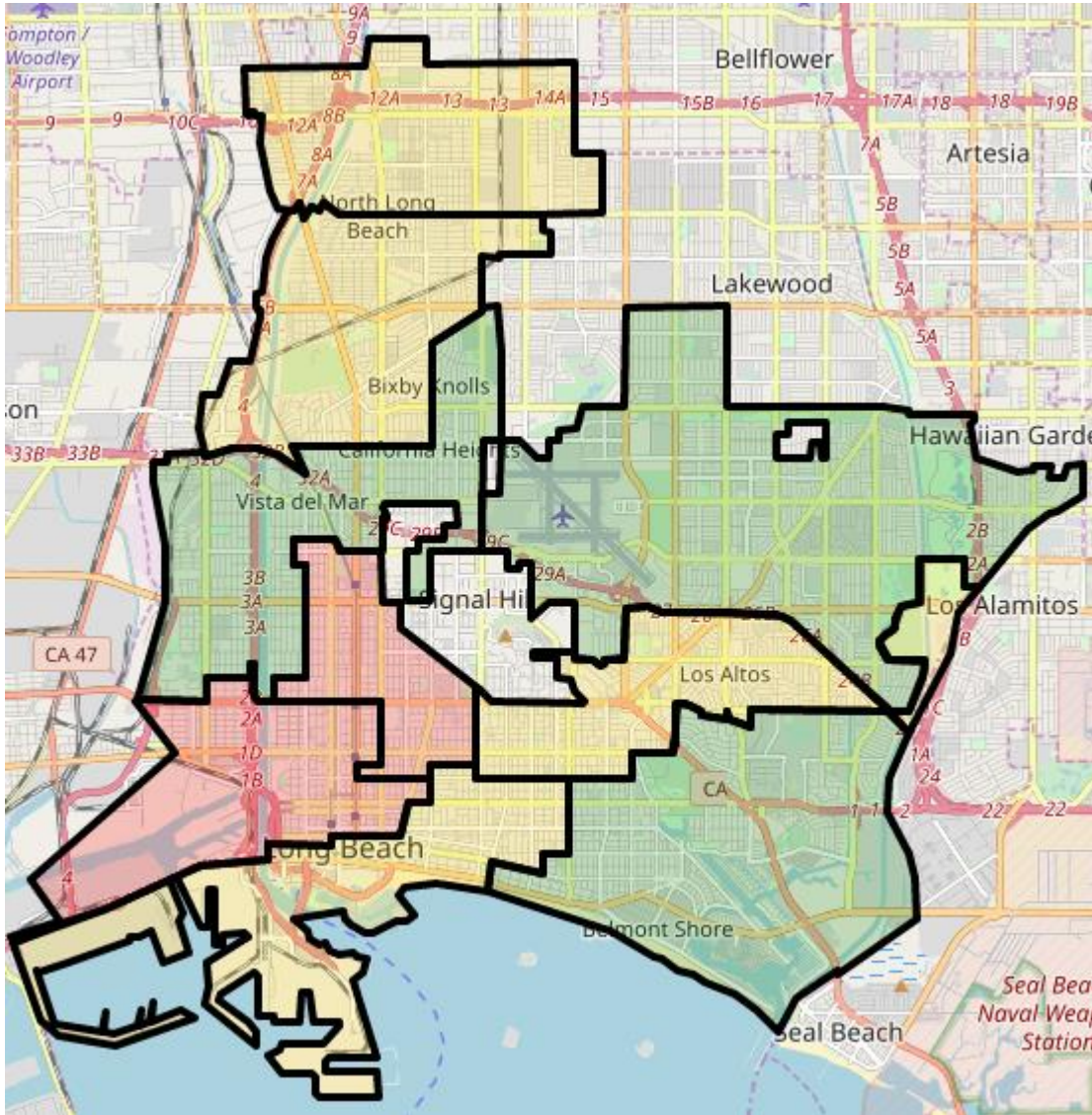


Figure 1. Long Beach City Council Districts

On the other hand, Figure 2 shows the 316 Police Reporting Districts (PoRDs) in the city of Long Beach. Now you have enough granularity in the crime statistics to apply a “bottoms up” model to matching historical crime levels and deriving models that may predict future crime levels. 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 indicate 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.

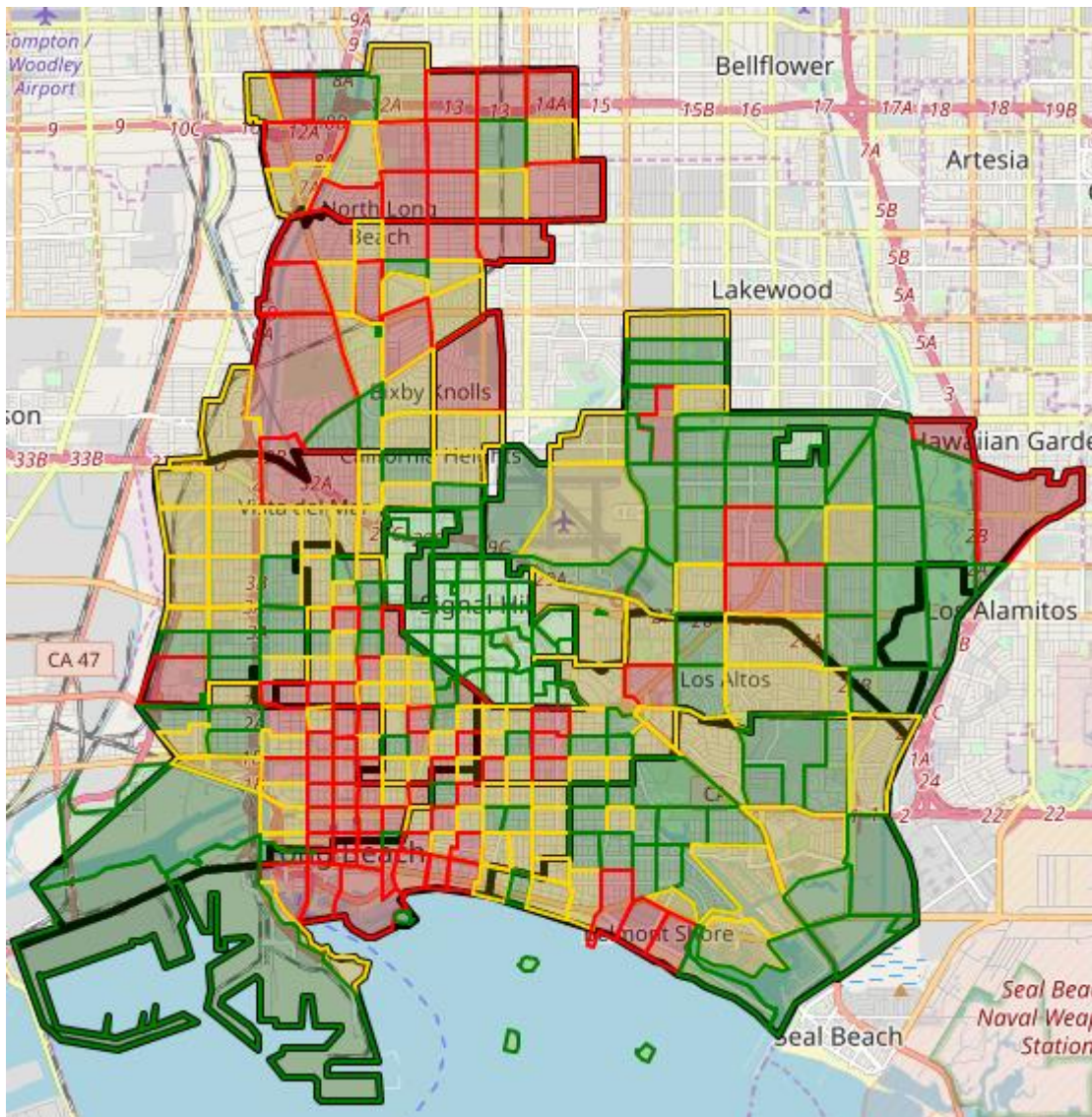


Figure 2. Long Beach Police Reporting Districts

In the ICD methodology, I'm going to take all of the monthly crime statistics starting in May 2018 and try to match a series of second order oscillators to those historical crime statistics. Second order oscillators have the advantage that they can go up or down with time. And neighboring PoRD crime levels can independently go up or down as time marches on. We'll "train" the second order oscillators to match the crime levels in each reporting district for the May 2018 through April 2019 and then we'll use the "trained oscillators" to project the crime levels for May 2019 and beyond (for each crime reporting district and for each month). Note that each crime reporting district is treated

independently and does not care what a neighboring or far-removed crime reporting district is doing.

The equation for the response of a single second order oscillator is shown below

$$y(t) = KM \left\{ 1 - e^{-\zeta t/\tau} \left[\cos\left(\frac{\sqrt{1-\zeta^2}}{\tau} t\right) + \frac{\zeta}{\sqrt{1-\zeta^2}} \sin\left(\frac{\sqrt{1-\zeta^2}}{\tau} t\right) \right] \right\}$$

where y is the monthly crime level for a single oscillator, T is the period of oscillation, and ζ is the damping ratio for the oscillator. The overall crime level for a district is the sum of a number of oscillators.

One can see that the equation has both sines and cosines in it to produce peaks and valleys in the crime response as well as an exponential to produce damping in the response.

Training of the oscillators consists of finding the period of oscillation, damping ratio, and summation coefficient for each oscillator that allows the total model crime level for a PoRD to most closely match the historical crime level.

From the “trained oscillators” for each crime reporting district, future months’ crime levels can be predicted and these crime levels can be aggregated up to city council districts.

With this framework, a Python script was written to learn the best fit series of second order oscillators for the period from May 2018 through April 2019 for each of the 316 Police Reporting Districts. From there, predictions for each of these PoRDs were made for the six months starting in May 2019. Again, each of the 316 Police Reporting Districts is assumed to act independently of any nearby or far-removed Police Reporting Districts. This philosophy goes along with the “focus on places, not on people” approach to crime since each PoRD has its own crime features whether positive or negative embedded in the training data. The inherent drawback to this approach is that it assumes that there has been no significant police and community action to reduce crime levels or no significant actions that increase crime in a Police Reporting District. These actions could possibly invalidate the derived model since the training data is not as applicable.

Part 3. Results

The historical data for each PoRD was matched with six 2nd order oscillators. Figure 3 shows a typical historical crime level and “matched” second order oscillator data for the training data from May 2018 through April 2019. In this case we are looking at PoRD 622 - a moderate crime district. The red bars show the historical crime training data used for May 2018 through April 2019. The blue line shows the match to the red training bars (from the 2nd order oscillators) as well a projection for the next six months. The yellow bars show the actual reported crime from May 2019 through October 2019. Certainly the linear least squares matching to the training data shown in green is not going to capture any peaks or valleys in either the training data or the future (i.e., predicted) data. But the

2nd order oscillator captures peaks and valleys in the crime levels and does a reasonable job at predicting future crime levels (when comparing the blue line with the yellow bars).

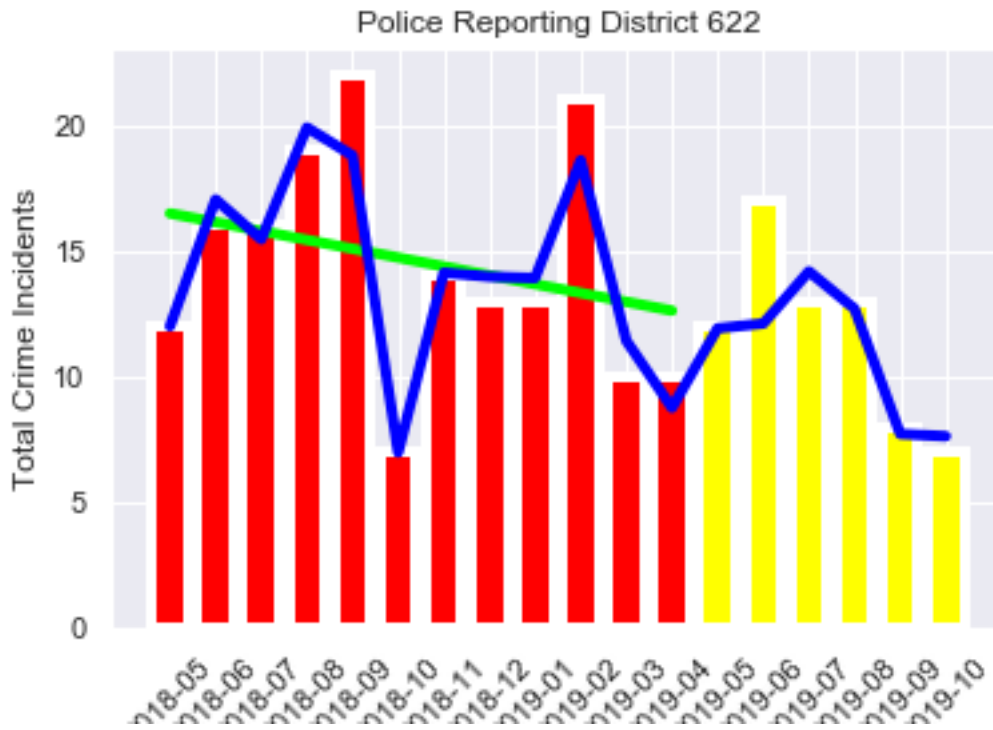


Figure 3. PoRD 622 Crime Levels, May 2018 through October 2019

Figure 4 shows the historical data for a high crime district, PoRD 524. Again, the red bars were used for training and the yellow bars were used as a measure of the prediction capability of the trained ICD model. Note that the ICD oscillators capture a periodic jump in crime every few months and continues that oscillation into the prediction phase. Thus, the ICD methodology is able to predict jumps up or down in crime as opposed to a linear regression method.

For PoRD 524 shown in Figure 4, I do not believe that further training would better capture the high crime spikes for the yellow bars in July 2019 or October 2019. These high crime spikes are somewhat out of family compared to the training data. This is an area that needs some further investigation.

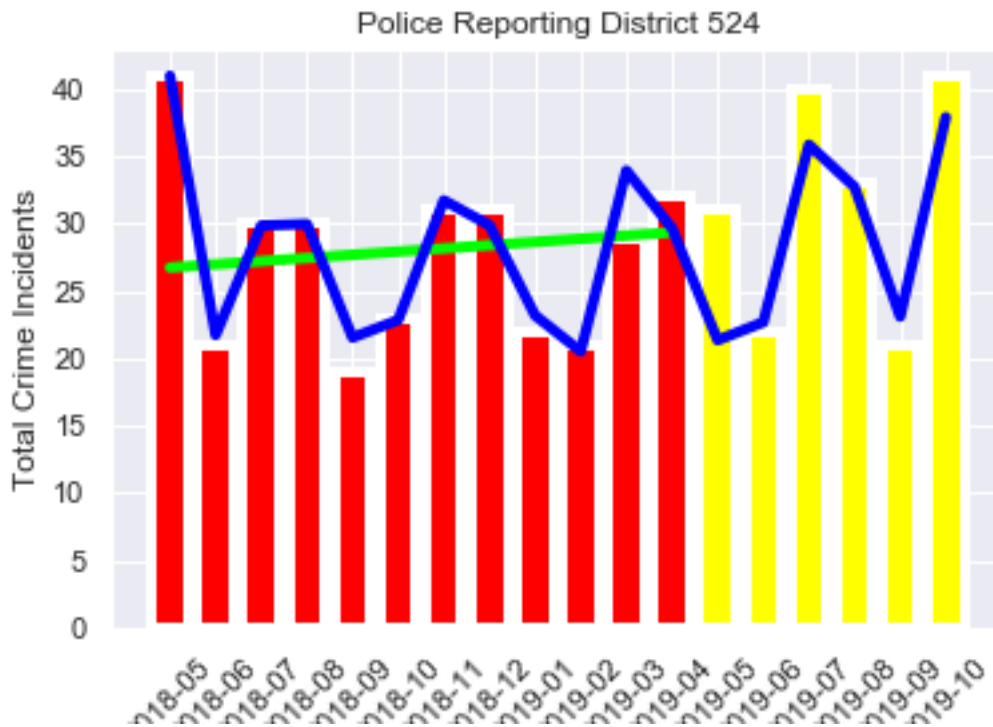


Figure 4. PoRD 524 Crime Levels, May 2018 Through October 2019

Finally, the citywide crime data can be aggregated from the 316 individual Police Reporting Districts. In this case, the historical crime levels as well as the predicted crime levels were simply added month by month to yield the citywide results. Figure 5 shows the training data in red, the actual future crime levels in yellow, and the ICD predictions for crime. The ICD predictions show good promise when accumulated all of the way up to the citywide level. But enough small differences in each and every PoRD, when accumulated up to the city level, contribute to discrepancies larger than desired.

I'll keep working on adjusting the methodology including the number of oscillators used for each PoRD as well as the convergence criteria for training each PoRD and see if that improves the citywide data.

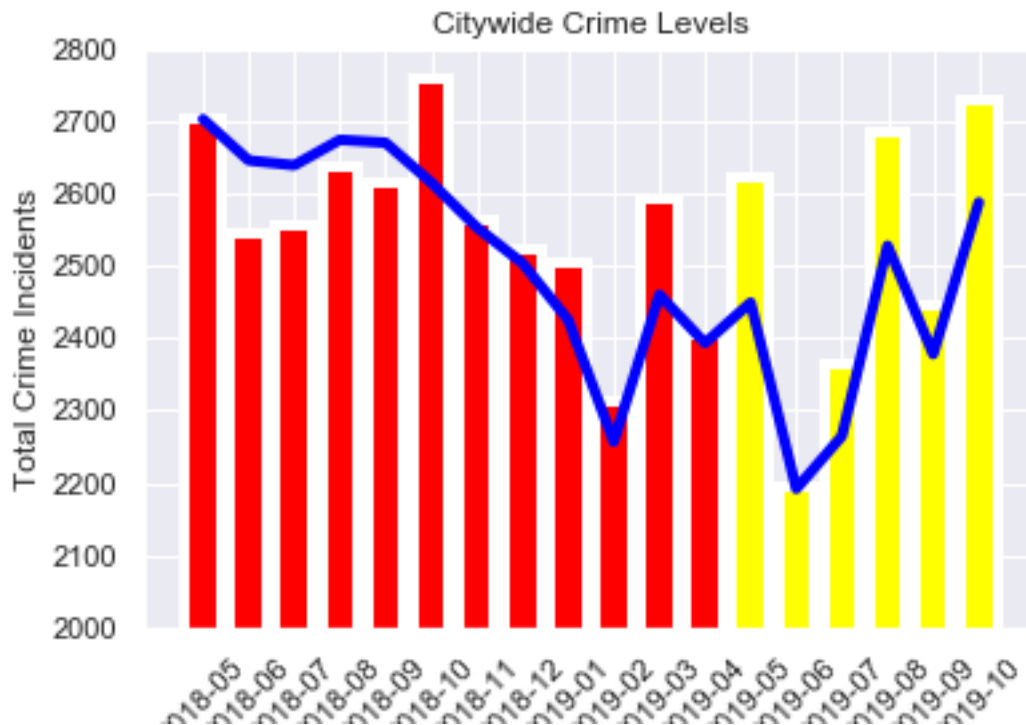


Figure 5. Citywide Crime Levels, May 2018 Through October 2019

Part 4. Commentary

This effort showed the use of matching a series of second order oscillators to historical crime data and use the resulting oscillators to predict future crime levels. The oscillators inherently allow crime to go up or down and replicate past crime levels. The model used second order oscillators for each Police Reporting District independently and did not have any interaction with neighboring Police Reporting Districts.

The effort indicates that the Independent Crime District model can be trained to match historical crime trends and be used to predict future crime levels.

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. Manning, R. A., "Predicting Crime Levels with Cellular Automata", Medium, <https://medium.com/@ray.90807/predicting-crime-levels-with-cellular-automata-ffb3928f1be5>
3. Long Beach Crime Statistics, <http://www.longbeach.gov/police/crime-info/crime-statistics/>