Burglaries in Tennessee

Kristen Bales, William Ty Frazier, and Dr. Ariel Cintron-Arias

Department of Mathematics and Statistics
East Tennessee State University

Introduction

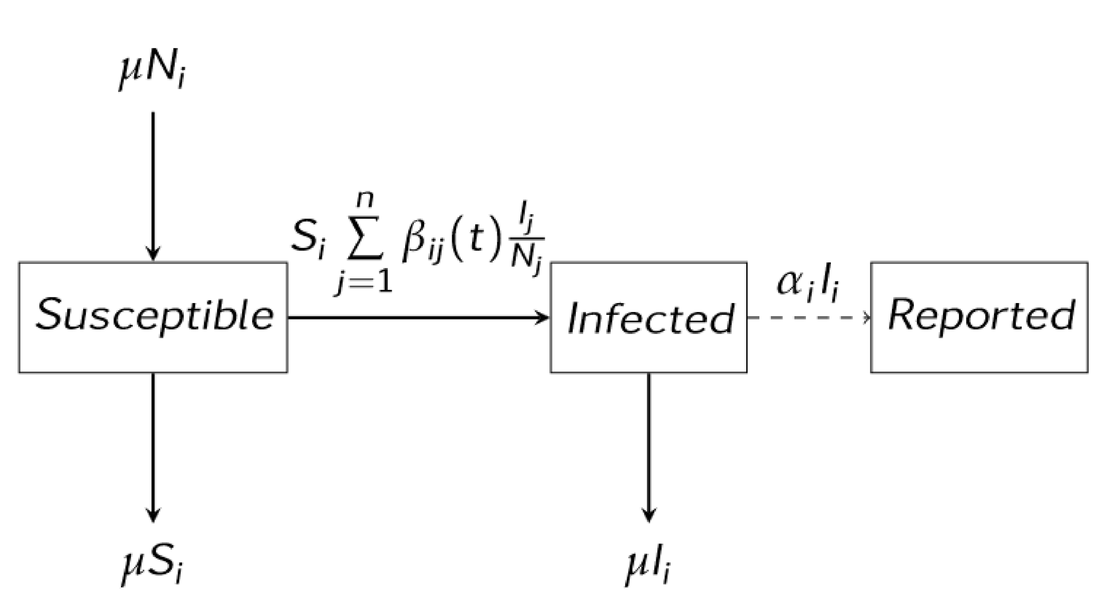
After numerous simulation attempts, Kate Bowers at Jill Dando Institute of Crime Science at the University College London revealed burglaries spread like an epidemic. Her model treated each burglary as an infection and possible source for future infections. She estimated the probability of a given location to be burglarized by combining the risk of all previous burglaries[2].

Also at University College London, Bill Hillier [2] focused on other influences of burglaries. These factors include the geometry of transportation networks and architecture of neighborhoods. He assigned weights to sections of a road based on the traffic it supported. While doing so he revealed clusters of crime. Hillier claims physical and spatial features interact in complex ways [2].

Model Parameters

- A_i = baseline spread
- d_{ij} = distance from patch i to patch j
- ϵ_i = intensity of spread
- μ = rate of change of households in the state
- α_i = reporting rate of patch i
- N_i = number of households of patch i
- $\beta_{ij}(t)$ = spatio-temporal spread
- $\beta_{ij}(t) = A_i e^{-d_{ij}}(1 + \epsilon_i cos(2\pi t))$

Model Flow Chart



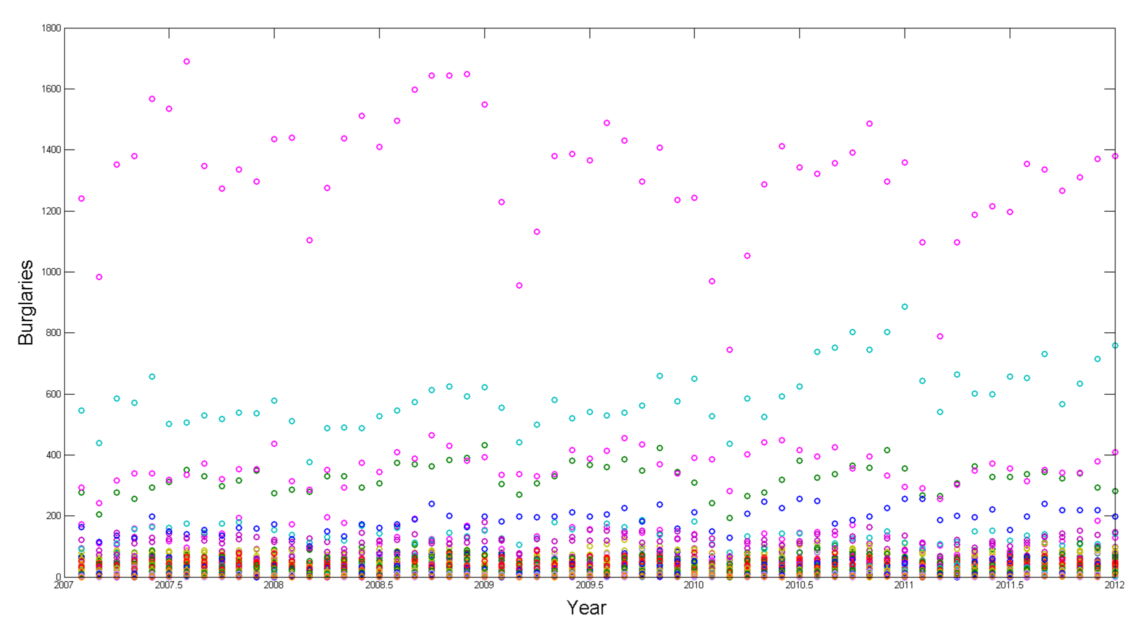
Model Equations

$$\frac{dS_i}{dt} = \mu N_i - S_i \sum_{j=1}^{n} \beta_{ij}(t)\frac{I_j}{N_j} - \mu S_i$$

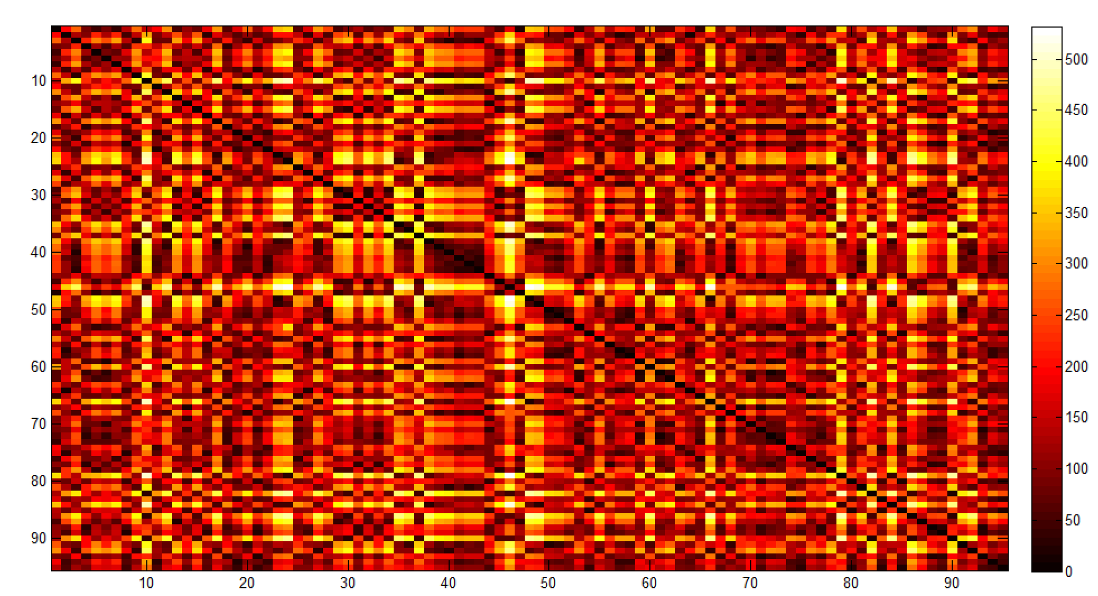
$$\frac{dI_i}{dt} = S_i \sum_{j=1}^{n} \beta_{ij}(t)\frac{I_j}{N_j} - \mu I_i$$

$$\frac{dR_i}{dt} = \alpha_i I_i$$

Longitudinal Burglary Reports



Distances Between Counties



Vector Ordinary Least Squares

Ordinary Least Squares is a method to find a vector of parameters, $\vec{q}$, for one time series of data. However we have one time series for each county. To account for the multiple time series, we use an augmented version called Vector Ordinary Least Squares[1].

$$\vec{Y_j} = \vec{f}(t_j, \vec{q_0}) + \vec{\varepsilon_j}$$

In the equation above, $\vec{Y_j}$, is the vector of time points for county j, which is set equal to the sum of the estimated points given by function $\vec{f}$ and an error for each point $\vec{\varepsilon_j}$. The method to solve the equation for the values of $\vec{q}$ is to minimizes the error $\vec{\varepsilon_j}$. We do this by finding the minimum in $\vec{q_0}$, our initial guesses, yielding the function:

$$q_{OLS} = \underset{\vec{q}\epsilon Q}{arg\ min} \sum_{j=1}^{n} [\vec{Y_j} - \vec{f}(t_j, \vec{q})]^T V_0^{-1} [\vec{Y_j} - \vec{f}(t_j, \vec{q})]$$

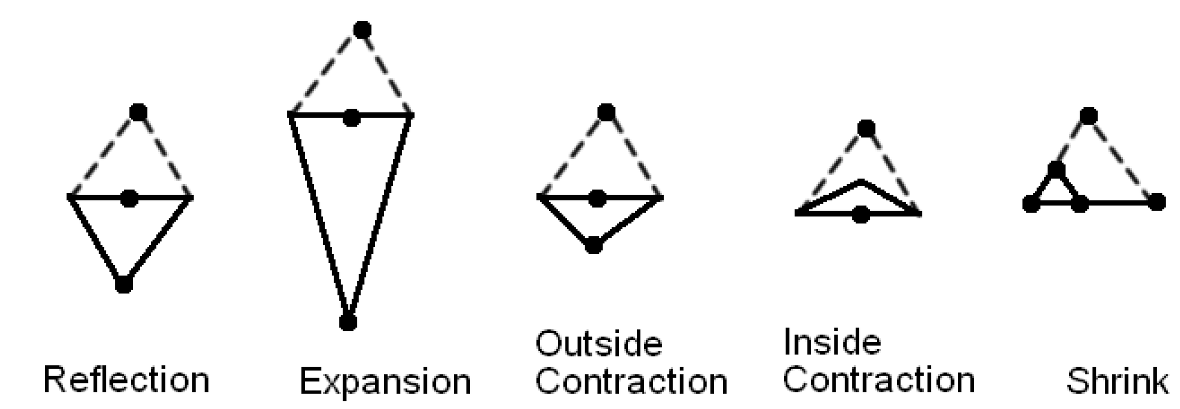
In these equations V represents the weights of each time point. We set V_0 to the identity matrix to find $\hat{q}_0$ from equation the equation above. Then we use the following equations for optimization:

$$\hat{q} = \underset{\vec{q}\epsilon Q}{arg\ min} \sum_{j=1}^{n} [\vec{Y_j} - \vec{f}(t_j, \vec{q})]^T V^{-1} [\vec{Y_j} - \vec{f}(t_j, \vec{q})]$$

$$\hat{V} = diag\left(\frac{1}{n-p}\sum_{j=1}^{n} [\vec{Y_j} - \vec{f}(t_j, \hat{q})][\vec{Y_j} - \vec{f}(t_j, \hat{q})]^T\right)$$

To optimize $\hat{q}_0$ we find use the equation above to find the next $\hat{V}$. We then use the new $\hat{V}$ to compute the next $\hat{q}$. We continue this cycle until the new and old $\hat{q}$ are close enough to each other.

Nelder-Mead

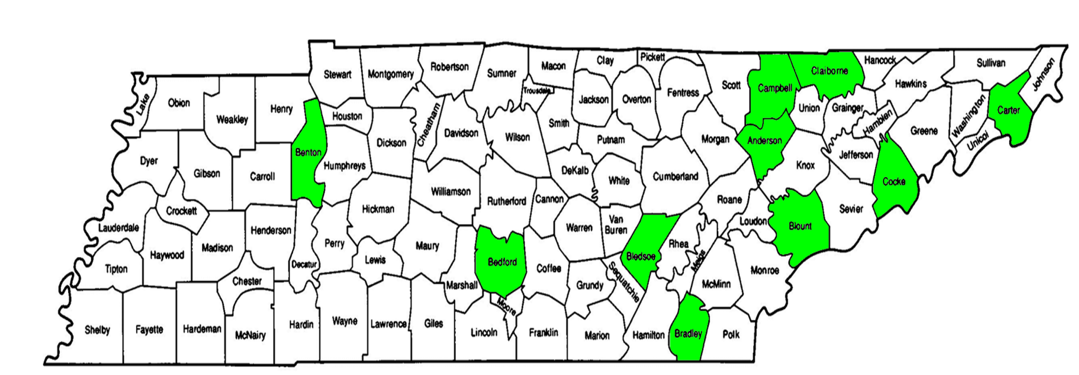


The Nelder-Mead Method is geometric method. It work in the space of $n+1$ dimensions. It has five different strategies each iteration can use. The strategies are: reflection, expansion, outside contraction, inside contraction, and shrink.

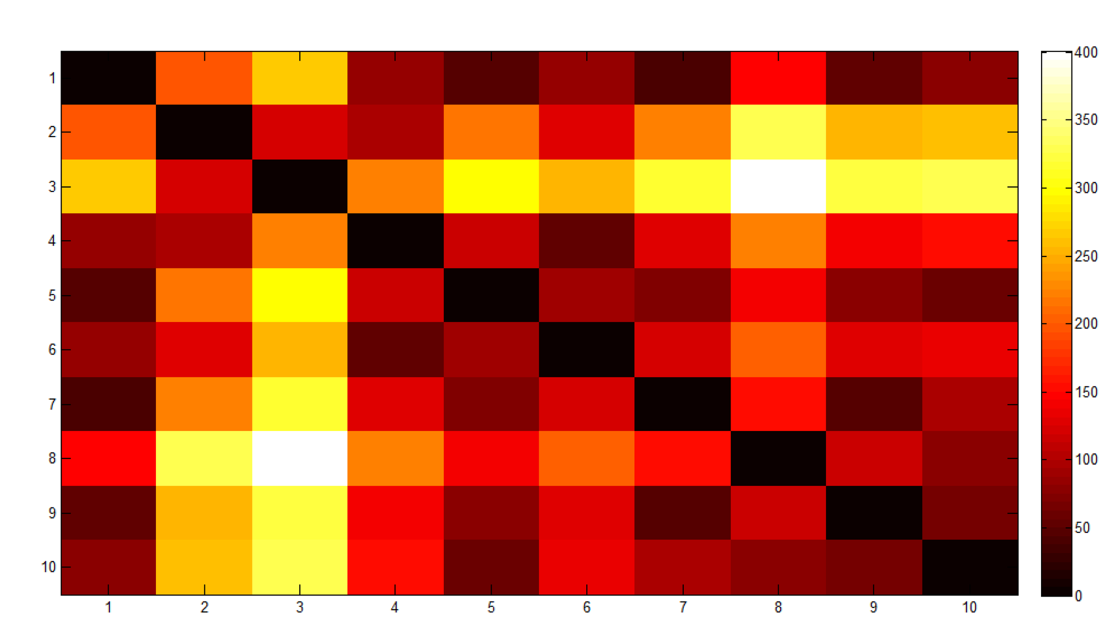
Genetic Algorithm

The Genetic Algorithm creates a population of potential solutions. It then uses a fitness function to determine if the potential solution is a good fit or not. After finding good fits in the population, the next generation is created one potential solution at a time. Each solution is created by a recombination of previous fit solutions. After a new solution is made, there is a possibility for mutation. Mutation takes a fit solution and drastically changes one or more parts to add variety. Mutation is vital to find a global minimum instead of local minimum. This process is continued until the fit solutions converge.

Tennessee Counties



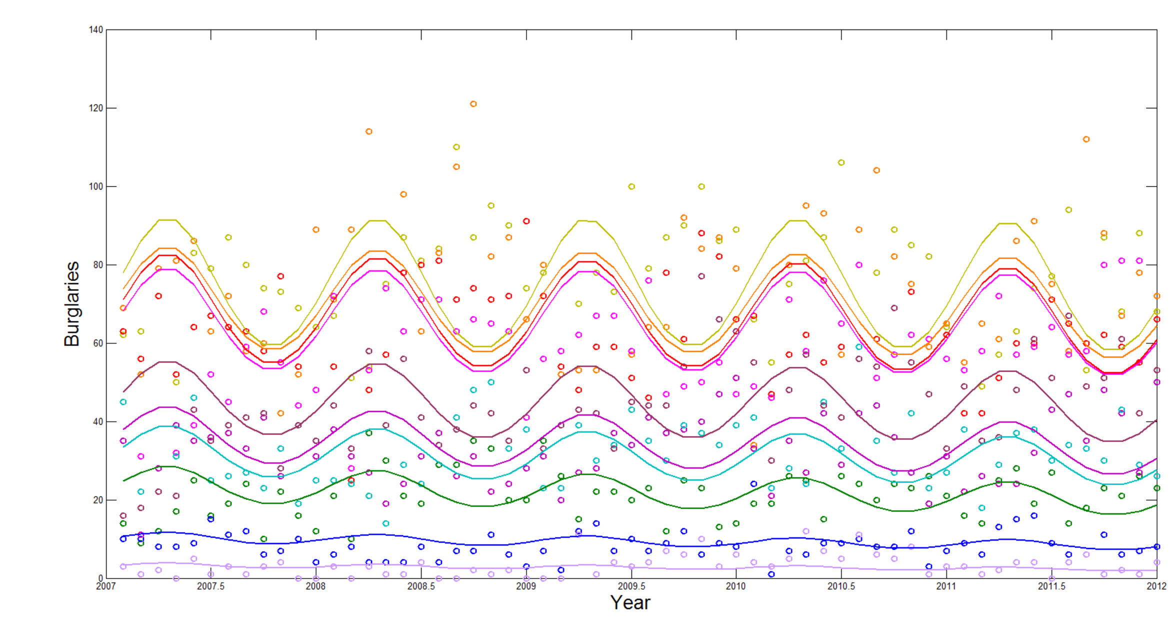
Sample Distance Matrix



OLS Parameter Point Estimates

County	A_i	ϵ_i	α_i	μ
Anderson	1.003	1.581	0.313	0.092
Bedford	0.908	1.466	0.578	0.092
Benton	0.898	1.488	0.588	0.092
Bledsoe	0.855	1.341	0.867	0.092
Blount	0.988	1.481	0.205	0.092
Bradley	0.978	1.441	0.285	0.092
Campbell	0.984	1.452	0.577	0.092
Carter	0.935	1.478	0.392	0.092
Claiborne	0.905	1.488	0.513	0.092
Cocke	0.974	1.508	0.481	0.092

Real Data and Best Fit Solutions



Discussion

We created a spatial epidemiological model from notions of Bower and Hillier[2]. The model has two classes, susceptible and infected, and keeps a record of reported as well. Next, we use Vector Ordinary Least Squares with Nelder Mead and Genetic minimization to estimate the parameter values.

Going forward it will be important to redefine the parameter bounds. I would also like to explore the relation of parameter estimates and standard errors per county as the number of counties in the group expands.

Acknowledgements

This research was supported by the an Access and Diversity Initiative from the Tennessee Board of Regents, grant #210024

References

[1] H.T. Banks and H.T. Tran, eds. *Mathematical and Experimental Modeling of Physical and Biological Processes.* London: Chapman and Hall, 2009.

[2] M. Buchanan. "Sin cities: the geometry of crime". In: *New Scientist* Issue 2810 (2008), pp. 36–39.