Hierarchical Bayesian Analysis of Property Crime

GONÇALO FARIA and MARGARIDA CAMPOS, Instituto Superior Técnico, Lisboa, Portugal

1 INTRODUCTION

In this paper we present a thorough statistical study of property crime in the United States of America(US), over two years at a state and regional level. The data used was collected from local law enforcement agencies by the Federal Bureau of Investigation(FBI) and published yearly in their Uniform Crime Report(UCR). UCR is the most known and widely cited source of official criminal statistics in the US.

The US has particularly interesting crime dynamics, compared with the other western countries, due to the rampant gun ownership. Almost 50 million households own a gun, more than 90 million, or 49 percent of the adult U.S. population, live in households with guns and about 59 million adults personally own guns [4].

1.1 Problem Statement

Considering the economical and political importance of the United States and the many conjectures that have been made through the years about its elevated crime rates and its relation to gun ownership it seems important to better understand this problem, and furthermore, to model it. This paper's primary goals are to model, in time, the occurences of property crime by region, and to draw conclusions from the regional results to better understand its behavior.

1.2 Objectives

In order to understand and model the problem, we aim at:

- building a model that estimates occurences of each type of property crime by region
- using the pre-established hierarchical model to get the total number of occurences by region
- comparing the results and resulting parameters to draw conclusions about region and population density effects

2 METHODS

2.1 Data Source and Treatment

The dataset consists of:

- two data points for each of the 52 states corresponding to the years 2015 and 2016
- categorical feature - *Region* corresponding to the region each state belongs to
- 5 numerical features:
  - *Population* of each state
  - *Property Crime* - total number of occurences of property crimes
  - *Burglary* - total number of burglary crimes
  - *Larceny Theft* - total number of larceny thefts
  - *Motor Vehicle Theft* - total number of motor vehicle thefts

Figure 1 contains a depiction of the states and regions of the U.S.. The reporting of these crimes follows a hierarchy rule, represented in Figure 2

Authors' address: Gonçalo Faria; Margarida Campos, Instituto Superior Técnico, Lisboa, Portugal.

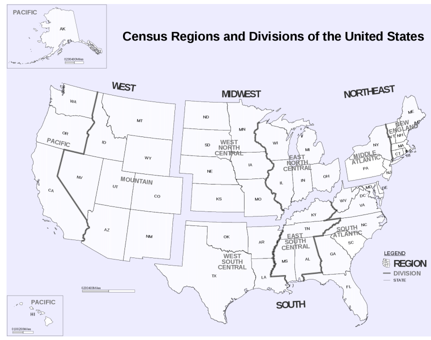

Fig. 1. Depiction of the US regions and states.

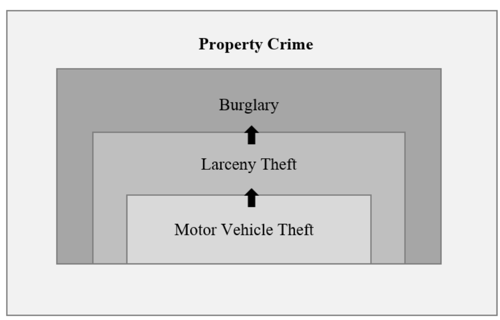

Fig. 2. Property Crime Hierarchy - followed in FBI reports

It is important to mention that despite criminologists' continued reliance on the UCR, its accuracy has been put to question.
Two main concerns in our study are: reporting practices and methodological problems.

2.1.1 Reporting practices. A big problem we can find when dealing with crime data is the possibility of a very incomplete picture. Some criminologists claim that victims of many serious crimes do not report these incidents and therefore, these crimes do not become part of the UCR database.
According to surveys of crime victims, less than 40 percent of all criminal incidents are reported to the police. These findings indicate that the UCR data may significantly underreport the total number of annual criminal events [4].

2.1.2 Methological issues. Another data quality issue arises from the lack of a standard methodology of reporting - not all police departments submit reports. Reports are voluntary and vary in accuracy and completeness. In addition to this, when an offender commits multiple crimes, only the most serious is recorded (following the hierarchy rule) [4], masking some possibly relevant insights in the chain.

2.2 Data Exploration

To better understand if there are differences in property crime rates we can look at Figure 3 to see the property crime numbers per 100 capita per state.

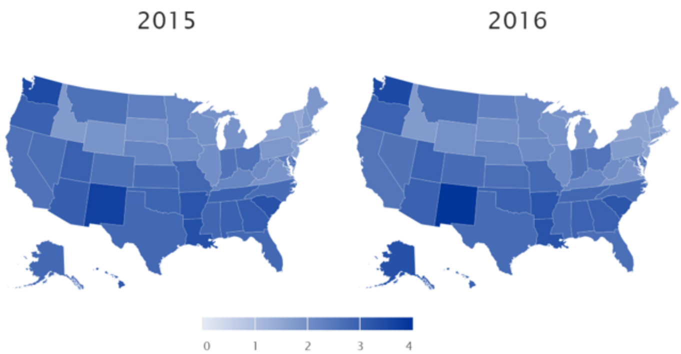

Fig. 3. Property Crime per 100 capita

The first thing that we can take from the map is that there are easy to spot differences between Regions, namely it is clear that crime rates in the Midwest are lower than in the West for example.

Another interesting observation is to see that the distribution seems to remain unchanged between the years. In fact, we can see in Figure 4 that the evolution of property crime rate is very similar across regions, with almost all lines being parallel. Overall there is a slight decrease in crime rate from 2015 to 2016, with the exception of region Mountain.

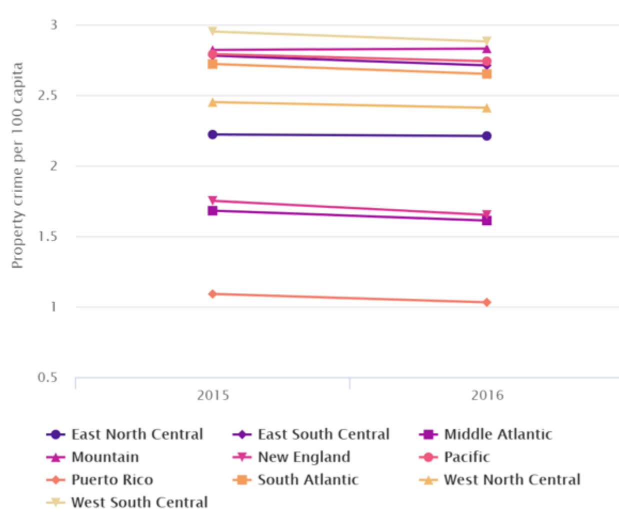

Fig. 4. Property Crime per 100 capita

In Figure 5 we can see how property crime breaks down by types of crime. Larceny theft represents the biggest percentage of crimes in all regions. Note also that the percentages by region are not constant, they vary from 2015 to 2016. This should be taken in consideration when looking at the results of the models by subtype of crime.

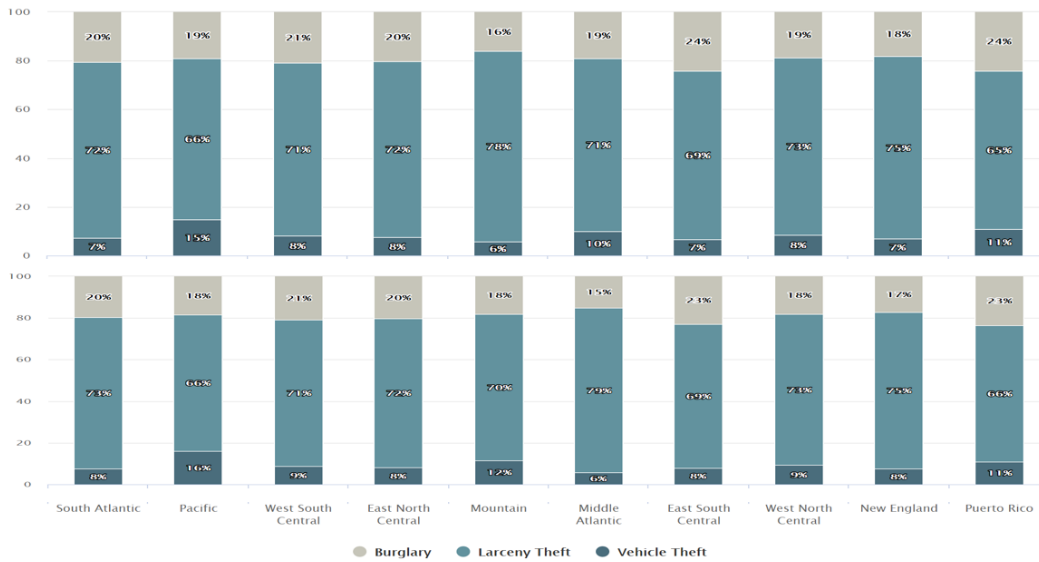

Fig. 5. Percentage of types of Property Crime - 2015 2016

2.3 Model Formulation

2.3.1 Model of Property Crime . To construct a model for property crime we start with a Poisson model for the expected number of crimes E_{it} for region i at timestep t. We assumed this variable to follow a Poisson distribution instead of, perhaps, a Binomial because the crime can not be interpreted as a proportion of the population. Crime is generally committed by a small part of the population, which offends multiple times. Hence the distribution will be as follows:

$$E_{it} \sim \text{Poisson}(n_{it} p_{ij}), \tag{1}$$

where n_{it} represents the population and p_{ij} represents the crime rate, both in region i at time t. With a model for the number of crimes we can use it to model the number of property crime occurences, which is a proportion of the latter with the form:

$$X_{it} | E_{it} \sim \text{Binomial}(E_{it}, r_{it}) \tag{2}$$

The proposed model has a clear interpretation, p_{ij} is the crime rate in region i and timestep t and r_{it} is the proportion of crimes reported which are property crimes. However, when the variables are distributed in this way, we can show that in fact:

$$X_{it} \sim Poisson(r_{it} n_{it} p_{it})$$

Since we don't have the data for E_{it} we will use:

$$X_{it} \sim Poisson(n_{it} \lambda_{it}),$$

where λ_{it} is the property crime rate in region i at time t and amounts to the product between r_{it} and p_{it}.

Therefore, the model will take the form:

$$
\begin{cases}
X_{it} \sim Poisson(n_{it}\lambda_{it}) \\
log(\lambda_{it}) = \theta + \alpha_{r(i)} + \gamma_i + \mu_{it} \\
\mu_{it} = w(\mu_{it-1} - \epsilon_{it}) + \epsilon_{it} \\
\theta \sim N(\beta_\theta, \tau_\theta^{-1}) \\
\epsilon_{it} \sim N(0, \tau_\epsilon^{-1}) \\
w \sim N(0, \tau_w^{-1})
\end{cases} \tag{3}
$$

Where the parameters represent the following:

- θ captures the national effect on the log-relative risk
- $\alpha_{r(i)}$ captures the region-wide homogeneity for the region of state i.
- γ_i captures the state-wide homogeneity for state i.
- μ_{it} captures an heterogeneous effect for state i and timestep t
- w controls the degree of dependence in past observations (w =0 amounts to a stationary model) in the heterogeneous effect.

The parameters β_θ, τ_w and τ_θ have non-informative priors. The remaining parameters, γ_{it} and $\alpha_{r(i)}$ are specified with a conditional autoregressive (CAR) or as independent gaussian variables.

2.3.2 Model of Property Crime Subtypes . Having defined a model for property crime, since Burglary $Y1_{it}$, Larceny theft $Y2_{it}$ and Motor vehicle theft $Y3_{it}$ are subtypes of property crime we use a Multinomial model to estimate them jointly.

$$
\begin{cases}
Y_{it}|X_{it} \sim Multinomial(X_{it}, \lambda_{it}) \\
\lambda_{itk} = \frac{p_{itk}}{\sum_{u=1}^{K} p_{itu}}, k = 1, 2, 3 \\
log(p_{it}) = \theta + \alpha_{r(i)} + \gamma_i + \mu_{it} \\
\mu_{it} = w(\mu_{it-1} - \epsilon_{it}) + \epsilon_{it} \\
\theta_k \sim N(\beta_\theta, \tau_\theta^{-1}), k = 1, 2, 3 \\
\epsilon_{itk} \sim N(0, \tau_\epsilon^{-1}), k = 1, 2, 3 \\
w_{kk} \sim N(0, \tau_w^{-1}), w_{ku} = 0, u \neq k, u, k = 1, 2, 3
\end{cases} \tag{4}
$$

Where here the exponential is taken to be an elementwise operation. Note that apart from the exponential normalization of p_{it} and the *Multinomial* distribution, this model is extremely similar to the one defined for property crime.

The parameter λ_{it} is a discrete probability mass function, which corresponds to the proportions of Burglary, Larceny theft and Motor vehicle theft in that particular region and timestep.

2.3.3 Priors that convey spatial association. We can take the regional and state effects to be independent, however, by doing so we would have discarded the knowledge that both states and regions are spatially associated. By including this knowledge in the prior specification of our model we hope to obtain an improved model that can in a better way quantify uncertainty.

By following [1], we were able to identify 1 technique that seems appropriate to our dataset. Particularly, CAR models, originally proposed in [2]. The main idea is that the value for region i should depend on its neighbors($i \sim j$).

$$\alpha_i|\alpha_{-i}, \tau_i^{-1} \sim N(\phi \sum_{i \sim j} \frac{w_{ij}}{d_{ii}} \alpha_j, \frac{\tau_i^{-1}}{d_{ii}}), i = 1, ..., n \tag{5}$$

If we then construct an adjacency Matrix containing the neighbor relations between regions W, where $w_{ij} = 1$ if region i is neighbor of j, we can show that this amounts to the joint probability distribution of the form:

$$\alpha \sim N(\mathbf{0}, [\tau(D - \phi W)]^{-1}) \tag{6}$$

The value D is a diagonal matrix where d_{ii} is equal to the number of neighbors of region i.

The parameter ϕ can be viewed as an additional parameter of the model that has a clear interpretation. When $\phi = 0$, then the full conditionals erode to $\alpha_i \sim N(0, \frac{\tau^{-1}}{d_{ii}})$, which is the spatial independence case. When $\phi = 1$ there is an indication of strong spatial association. Aditionally, the precision τ, regulates the degree to which the random effects shrink to the average of their neighbours.

For this distribution to be proper, we have to make sure that $\tau(D - \phi W_r)$ is positive definite, which is true for undirected graphs composed by regions that always have at least one neighbor and $|\phi| < 1$. Therefore, ϕ is taken to be uniform 0 to 0.99 and τ a Gamma(0.001,0.001).

Even when the distribution of α is not proper, since the full condtionials are proper, we can still use it. When using Gibbs sampler, we need only the full conditionals distributions in order to sample for the joint distribution. However, since we are using JAGS[3] for estimation, the priors must be proper. This presented a challenge given that some regions, such as Porto Rico, Alaska and Hawai have no neighbors, see figure 1. We solved this issue by assigning independent normals for these regions and using the CAR model only for the remaining ones.

The prior we used for α and α_k for $k = 1, 2, 3$ is presented in equation 7. Here Wr, and D_r correspond to the adjacency matrix of the neighbor relations between the regions and the corresponding diagonal matrix with the row sum.

$$\alpha \sim N(\mathbf{0}, [\tau_\alpha(D_r - \phi_r W_r)]^{-1}) \tag{7}$$

Similarly, the prior for γ and γ_k for $k = 1, 2, 3$ is presented in equation 8. Here W_s, and D_s correspond to the adjacency matrix of the neighbor relations between the states and the corresponding diagonal matrix with the row sum.

$$\gamma \sim N(\mathbf{0}, [\tau_s(D_s - \phi_s W_s)]^{-1}) \tag{8}$$

2.3.4 Priors that convey temporal association. Using the tests in the exploratory analysis we were able to conclude that the data was not stationary. Therefore we constructed a heterogeneous effect term which is autoregressive of order one both in 3 and 4.

2.4 Model Selection

Traditionally, in Bayesian statistics, the Bayes factor was used for model selection. However, Bayes factor becomes quite difficult both to compute and interpret for high-dimensional hierarchical models and is not well-defined for models having improper prior distributions. For this reason, several approaches to model choice criteria have been proposed, particularly the *Deviance Information Criterion* [5] and *Widely Applicable Bayesian Information Criterion* [6].

DIC is based on the deviance statistic

$$D(\theta) = -2ln\, f(y|\theta) + 2ln\, g(y) \tag{9}$$

Here $f(y|\theta)$ is the sampling model of the data y given the parameter θ, and f(y) is a standardizing function of the data. Using this methodology the model fit is encapsulated by the posterior expectation of the deviance, $E_{\theta|y}(D)$, while the model complexity is captured by another measure p_D, which is defined as expected deviance minus deviance evaluated at the posterior expectation.

$$p_D = E_{\theta|y}(D) - D(E_{\theta|y}(\theta)) \tag{10}$$

Therefore, DIC is then defined as the sum between model fit and complexity.

$$DIC = E_{\theta|y}(D) + p_D \tag{11}$$

Small values of DIC indicate a better-fitting model and, unlike Bayes factor, we can order a set of models.

WAIC, while sharing the same idea of designing a criterion with both model complexity and model fit in a bayesian setting, it calculates model fit and model complexity p_W in a slightly different way.

$$p_W = -2\sum_{i=1}^{n}\{E_{\theta|y}[lnf(y_i|\theta)] - lnE_{\theta|y}[f(y_i|\theta)]\} \tag{12}$$

$$WAIC = -2\sum_{i=1}^{n} lnE_{\theta|y}[f(y_i|\theta)] + 2p_W \tag{13}$$

2.5 Markov chain monte carlo for Estimation

In order to estimate the parameters for the models Gibbs Sampling - a Markov chain Monte Carlo algorithm - will be used.
The idea is to create **ergodic** Markov chains whose **equilibrium distribution** is our desired *posterior* distribution, and later apply Monte Carlo estimation on the final samples of the chains to obtain the wanted parameters.

The Gibbs Sampling algorithm, uses the full *posterior* conditional distributions,*i.e.*, the distribution of each variable conditioned by all remaining variables.

Regarding implementation, four parallel, initially overdispersed MCMC chains were run for 40.000 iterations each, for all of the presented models. In every case, we employed graphical monitoring of the chains for a representative subset of the parameters, along wih sample autocorrelations. All of the presented models indicated an acceptable degree of convergence.

3 RESULTS AND INTERPRETATION

3.1 Property crime model

We fitted seven property crime models with different combinations of random effects and also random effects with independent priors(not using CAR), and accessed their performance using Bayesian approaches, Mean Squared Error(MSE) and Log-Likelihood.

Property crime	pD	DIC	Log-likelihood	WAIC	MSE
Baseline	inf	inf	-427061.683	inf	1856642858.871
Model I	102.984	1586.738	-690.386	1516.082	546.972
Model II	103.714	1588.032	-690.302	1516.981	648.406
Model II (i)	102.961	1586.469	-690.273	1515.750	674.307
Model III	103.411	1587.931	-690.555	1517.192	738.764
Model III (i)	102.882	1586.349	-690.293	1515.686	**545.870**
Model IV	103.130	**1585.085**	-690.413	**1514.827**	742.634
Model IV (i)	103.246	1587.028	**-690.268**	1516.178	563.010

Table 1. Results for the property crime models.

The baseline comprises only the national fixed effect. It assumes no heterogeneous effect. The model I has a national fixed effect and the heterogeneous state and time effect. It assumes no homogeneous effects between members of the same region and the same state across time. Model II drops some of the assumptions and introduces a regional homogeneous effect with CAR prior. The Model III, instead of introducing to model I regional homogeneity, introduces state homogeneous effect with CAR prior. The full model, Model IV, introduces both state homogeneity and regional homogeneity. To access whether CAR priors are useful, we also experiment with independent prior versions of II, III and IV, denoting them with (i). The results are presented in Table 1.

From the results shown in the Table1, we are able to conclude that using priors that convey the spatial associations is advantagous. We selected model IV, since it was the best model both in DIC and WAIC. In figure 6 we presente the relative residual plot for this model.

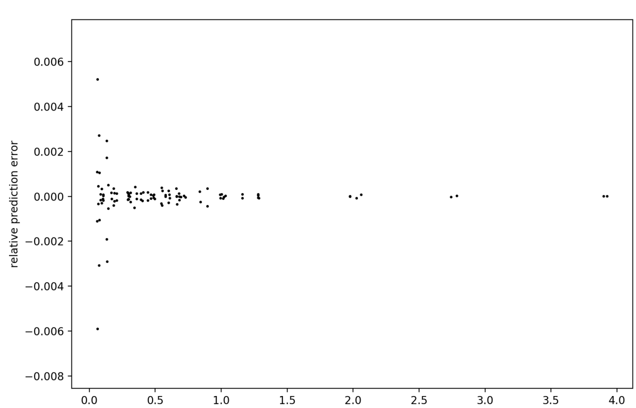

Fig. 6. Relative residual plot for the property crime model.

3.2 Property crime Subtypes model

For the property crime subtypes we fitted four models also with different combinations of random effects. Simillarly we accessed wheir performance using Bayesian approaches, Mean Squared Error(MSE) and Log-Likelihood.

Property crime Subtypes	pD	DIC	Log-likelihood	WAIC	MSE
Baseline	2.054	349263.785	-174630.594	inf	108132266.612
Model uI	203.638	2702.076	-1147.400	2564.223	**216.839**
Model uII	204.487	2701.667	**-1146.347**	2562.522	248.922
Model uIII	201.677	**2698.485**	-1147.635	**2562.489**	309.370
Model uIV	202.787	2700.270	-1147.348	2563.068	316.266

Table 2. Results for the property crime subtypes models.

The baseline comprises only the national fixed effect. It assumes no heterogeneous effect. The model uI has a national fixed effect and the heterogeneous state and time effect. It assumes no homogeneous effects between members of the same region and the same state across time. Model uII drops some of the assumptions and introduces a regional homogeneous effect with CAR prior. The Model uIII, instead of introducing to model uI regional homogeneity, introduces state homogeneous effect with CAR prior. The full model, Model uIV, introduces both state homogeneity and regional homogeneity. The results are presented in Table 2.

We selected model uIII, since it was the best model both in DIC and WAIC. In figure 7 we presente the relative residual plot for the diferent subtypes of property crime.

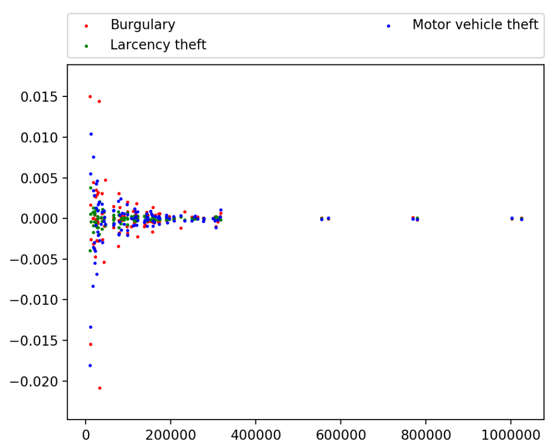

Fig. 7. Relative residual plot for the property crime subtypes model.

We can also look at the values of γ_i per state (see Figure 8), and see

3.3 Interpretation

In Table 3 we present the posterior summaries for a subset of the model IV parameters. At first glance, we are able to ascertain that both parameters ϕ for regional and state homogeneous effects are far from 0, therefore, the data favors spatial association. The precision τ is high in the regional case and low in the state case. Therefore, following the interpretation given above, we can say that the regional random effects will tend to follow their neighbor's average while the state random effects less so. The national effect θ is negative. This is expected given that when in comparison with the population size, the propery crime is small. The parameter w, which captures the temporal dependence of the heterogeneous effect, is negative. This is means that the heterogeneous effect, the

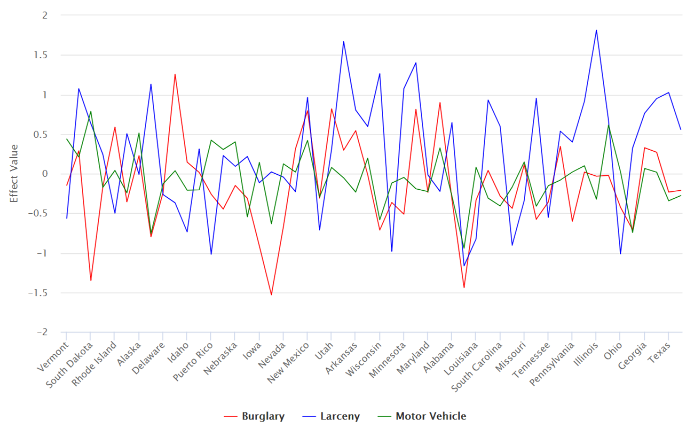

Fig. 8. State-wide homogeneity per state, by crime type

only temporal effect of our model, is decreasing. This makes intuitive sence, since in our exploratory analysis we seen that the the response variable for 2016 was lower that in 2015.

In Table 4 we present the posterior summaries for a subset of the parameters of the model uIII. The values of ϕ indicate the presence of spacial association, as expected. Interestingly, the values of τ follow the reporting hierarchy, Bugulary has the highest, followed by Larceny theft and then Motor vehicle theft. As expected, the values for w are also negative indicating a downward trend.

Parameter	Mean	sd	2.5%	Median	97.5%
ϕ_r	0.436	0.273	0.020	0.414	0.942
τ_r	1309.368	875.913	288.052	1089.169	3580.998
ϕ_s	0.283	0.200	0.012	0.250	0.725
τ_s	4.018	3.693	2.984	2.984	13.480
w	-0.998	0.001	-0.997	-0.998	-0.999
θ	-2.700	1.956	-6.040	-1.952	-0.870

Table 3. Posterior summaries of the Property crime Model.

Looking at Figure 9 we can compare the numbers of property crime generated by the final model compared to the actual numbers. We can see that all the predicted points overlap the actual values, validating the precision of the selected model.

4 CONCLUSIONS & FUTURE WORK

The results obtained fot the objective model are very good, specially when we take into account the lack of volume in data. This represents a good example of how a Bayesian approach for inference can be vastly superior to many modern Machine Learning techniques, that rely heavily on big amounts of data and provide no means for interpretation.
The big advantage one can find in this method is that now we have a model that not only can make predictions, but actually models the behavior of property crime. From this one can draw conclusions, evaluate the behavior, even study alternative scenarios by changing the parameters.

An important conclusion we can take from this exercise is that there is in fact a benefit in using spatial association, which is interesting in a criminology point of view. It would be interesting to

Parameter	Mean	sd	2.5%	Median	97.5%
Burgulary - ϕ_s	0.372	0.254	0.014	0.338	0.887
Burgulary - τ_s	392.806	692.369	0.0514	6.648	26225.022
Burgulary - w	-0.900	0.040	-0.968	-0.892	-0.829
Burgulary - θ	0.079	1.074	-1.183	-0.084	1.742
Larceny theft - ϕ_s	0.326	0.246	0.011	0.273	0.888
Larceny theft - τ_s	155.360	433.031	0.0203	0.226	1659.095
Larceny theft - w	-0.897	0.073	-0.992	-0.917	-0.742
Larceny theft - θ	1.950	1.692	-0.371	1.906	4.355
Motor vehicle theft - ϕ_s	0.335	0.257	0.012	0.274	0.919
Motor vehicle theft - τ_s	67.576	128.028	0.206	5.729	507.854
Motor vehicle theft - w	-0.879	0.061	-0.948	-0.902	-0.739
Motor vehicle theft - θ	-0.546	0.355	-1.109	-0.466	-0.113

Table 4. Posterior summaries of the Property crime subtypes model.

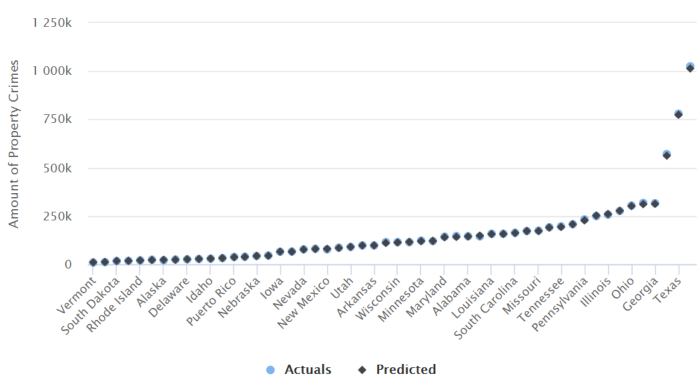

Fig. 9. Actual values of property crime compared with generated ones

see if this relation exists for different types of crimes.

In future work it would be interesting to use the obtained model to make predictions on more recent data and see how well it performed. Another interesting approach would be to try different priors to more extensely assert if the standard approach used is in fact superior.

REFERENCES

[1] S. Banerjee, B. Carlin, and A. E. Gelfand. Hierarchical modeling and analysis for spatial data. 2003.
[2] J. Besag, J. York, and A. Mollié. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43:1–20, 1991.
[3] M. Plummer. Jags: A program for analysis of bayesian graphical models using gibbs sampling. 2003.
[4] L. J. Siegel. Criminology: Theories, patterns, and typologies. 1983.
[5] D. J. Spiegelhalter, N. G. Best, B. P. Carlin, and A. van der Linde. Bayesian measures of model complexity and fit. 2002.
[6] S. Watanabe. A widely applicable bayesian information criterion. *J. Mach. Learn. Res.*, 14:867–897, 2012.