○ **Security Informatics**

RESEARCH

Open Access

CrossMark

Applying machine learning to criminology: semi-parametric spatial-demographic Bayesian regression

Roman Marchant[1,3]* ⓘ, Sebastian Haan[2], Garner Clancey[3,4] and Sally Cripps[1,5]

Abstract

Objectives: This paper describes the use of machine learning techniques to implement a Bayesian approach to modelling the dependency between offence data and environmental factors such as demographic characteristics and spatial location. The main goal of this paper is to provide a fully probabilistic approach to modelling crime which reflects all uncertainties in the prediction of offences as well as the uncertainties surrounding model parameters.

Methods: The proposed method is based on a Bayesian framework, with a Gaussian Process prior and MCMC, allowing uncertainties in prediction and inference to be quantified via the posterior distributions of interest. By using Bayesian updating, these predictions and inferences are dynamic in the sense that they change as new information becomes available.

Results: We applied the proposed methodology to particular offence data, such as domestic violence-related assaults, burglary and motor vehicle theft, in the state of New South Wales (NSW), Australia. Our results demonstrate the strength of the technique by validating the factors that are associated with high and low criminal activity, including bounds on the degree of the relation.

Conclusions: We argue that this fully probabilistic approach will improve prediction, in the sense that the uncertainties are more accurately quantified, with attendant benefits to policymakers and policing organisations seeking to deploy limited criminal justice resources to prevent and control crime. While limitations and areas for potential improvement are identified, the success of the Bayesian approach, implemented using machine learning techniques, in a criminological context represents an exciting development.

Keywords: Semi-parametric regression, Crime rates, Machine learning, Bayesian methods, Gaussian process

Introduction

For over 150 years, criminologists have aimed to understand crime; why it occurs, where and when. In most cases, this largely social scientific exercise has centred on the belief that to better understand who commits crime is to maximise the chances that social and criminal justice policy can be optimally designed to improve prevention, mitigate risks and manage the efficient allocation of scarce resources. Understanding crime has often involved focusing on longitudinal population information, behaviours and environments including education, employment, family structures, health, and contacts with the policing and justice system. The latest developments in data science and machine learning offer new ways to predict the incidence of crime and to understand the impacts of societal and individual characteristics on criminal behaviour.

In this work we show how to build fully probabilistic models that are able to answer important questions about crime, such as: What is the probability of the occurrence of a crime at a particular location? What are the characteristics of the population that affect the incidence of crime?

*Correspondence: roman.marchant@sydney.edu.au
[1] Centre for Translational Data Science, The University of Sydney, Sydney, NSW 2006, Australia
Full list of author information is available at the end of the article

Marchant *et al. Secur Inform* (2018) 7:1

Page 2 of 19

There are two challenges that need to be addressed in order to properly answer these questions. The first challenge is to define appropriate probabilistic models; the second is to construct machine learning algorithms to estimate these models and quantify the uncertainty around these estimates.

The contributions of this paper are the following:

1. Provide evidence based quantitative methodology that relates crime to environmental and demographic information by coupling the richness of the demographic and historical crime data with state of the art machine learning algorithms and probabilistic models. For the proposed model, the dependent variable is the crime rate at a particular location, which depends on multiple explanatory variables. Our methodology is general enough to allow a wide variety of location-based explanatory variables to be incorporated into the model, including demographic characteristics of the population, environmental features, and transport density, among many others.

2. Combine parametric and non-parametric techniques to model the dependency between the incidence of crime and location-specific factors, as well as to learn spatial correlations without assuming any functional form, which further improves the accuracy of prediction. As Bowers and Johnson [5] and Steenbeek and Weisburd [30] have pointed out, examining the spatial distribution of crimes at different geographical levels is fundamental for achieving an understanding of crime. The methods presented in this paper generate a continuous estimation of crime intensity over space. The underlying spatial estimation will increase its accuracy automatically when higher spatial resolution data are used as input.

3. Propose a fully probabilistic model, which is able to quantify the uncertainty in the predictions as well as model parameters. Accurate quantification of all sources of uncertainty is necessary to achieve informed and appropriate decision-making arising from the output of the models presented in this paper. Most work in this field only report point estimates of the quantities of interest [2, 4, 11], and either ignore or give rough approximations of uncertainty. We note that Weisburd et al. [33, chapter 22], report confidence intervals but these confidence intervals fall short of estimating the true uncertainty on the counts. First they assume the asymptotic normality of the sampling distribution of the regression coefficients and second they are conditional on estimates of other parameters in the model; unlike the Bayesian approach where inference is made via the marginal posterior distribution, where the marginalisation is w.r.t the posterior distribution of all other parameters.

The Bayesian paradigm is the only logistical consistent framework in which to make probabilistic statements regarding the complex models presented in this paper. For this reason Bayesian methods have become the norm in many other fields of study, such as robotics, where decisions are taken based on models learnt from data and the associated uncertainty. It is our hope that the Bayesian treatment of probability models, such as presented in this paper, will lead to informed policy decisions, the evaluation of the impact of specific factors on crime and the acquisition of new information.

The paper is structured as follows. In "Related work" section we review the existing work on models for crime, especially work focusing on demographic and spatial dependencies. "Methodology" section presents the proposed models and the machine learning algorithms used to learn the model parameters from real data, focussing on Bayesian Linear Regression (BLR), Gaussian Processes (GPs) and Markov Chain Monte Carlo (MCMC). Following this, "Application: regression over crime rates" section shows experimental results and comparisons on real world crime data. "Discussion" section presents a discussion of the results and highlights the links to criminological theory. Finally, "Conclusion and future work" section draws conclusion and presents ideas for future work.

Related work

Over the last few decades there has been considerable work on quantitative criminology. Particular interest has been on the study of the occurrence of crime, focussing on the spatial–temporal patterns of crime, and the factors related to criminal activity, including population characteristics and environmental factors. In this section, we briefly describe the relevant literature associated to the quantitative analysis of crime, with particular focus on regression techniques for criminology.

Relevant and popular models for spatial analysis of crime are presented in Chainey et al. [7], Eck et al. [12], Leitner [19], Perry et al. [27] and Piquero and Weisburd [28], include Kernel Density Estimation (KDE), K-means clustering, covering ellipses and other heuristics that result in hot-spot identification and spatio-temporal analysis of crime. Perry et al. [27] detail many other techniques to identify seasonality and periodicity at different resolutions in time series of crime intensity, however Perry et al. [27] fail to explore multivariate representations of crime that couple demographic and environmental effects. Gorr et al. [16] compare several methods for modelling time-series such as the random walk model, and various versions of exponential smoothing for

Marchant *et al. Secur Inform* (2018) 7:1

Page 3 of 19

different crime types. Nogueira de Melo et al. [22] have found different temporal patterns for different crime types. The effect of time coupled with risk terrain modelling is explored by Kocher and Leitner [18] and also noted in Perry et al. [27]. Although these methods are widely used by crime practitioners, they are ad-hoc techniques, in the sense that there is no consistent theoretical underpinning of how point estimates, and the uncertainties associated with these estimates, are obtained.

There are various approaches where authors have opted to model the occurrence of crime as a *solely* spatial–temporal phenomenon. Mohler et al. [23] propose a self-exciting process model of crime, considering a crime intensity function which varies over space and time according to a Poisson Point Process that presents higher values near areas that have experienced crime in the past. Another spatial–temporal approach by Flaxman [13] uses space-time Gaussian Process (GP) over the intensity of a Poisson distribution of event counts to explain the occurrence of crime. Flaxman [13] combines spatial–temporal covariance functions with periodic components that can capture seasonality in the temporal domain. In recent years, Gaussian Processes [29] have been used extensively in machine learning as priors over unknown functions, for modelling spatially and temporally correlated phenomena. Corcoran et al. [9] cluster the occurrence of crimes and uses a Neural Network model for auto-regressive prediction at each cluster. Grubesic and Mack [17] provide a comprehensive review of existing techniques for spatial–temporal modelling of crime, focusing on the importance of coupled space-time model which have varying temporal patterns depending on the location. These approaches model crime solely as a function of space and time, disregarding other sources of explanation. While these techniques may lead to good predictive performance, they do not help understand the factors which drive crime, necessary for the optimal allocation of scarce resources for the prevention of crime.

Other approaches to model crime have been explored by Osgood [26], who applied Poisson regression to crime rates using demographic quantities as explanatory variables. Boessen and Hipp [4] assume crime counts follow a negative binomial distribution, and use a general linear model to model the dependence of crime on the population characteristics of a specific area as well as surrounding areas. Davies [10] considers street network and near-repeat principles to explain burglaries also including the effect of small communities to understand dynamics in the network using differential equations. Deadman [11] has also built a temporal forecasting tool from demographic characteristics without including any spatial dependencies. Tita and Radil [31] recognise that spatial data and other characteristics need to

be considered simultaneously for correct inference and accurate predictions.

Antolos et al. [2] used *Logistic Regression* (LR) for calculating the probability of occurrence of a crime based on previous criminal events and physical characteristics of the environment that reflect connectivity to crime epicentres. Similarly, Berk et al. [3] applied LR, CART and Random Forest models to forecast subsequent domestic violence calls. They found several limitations with LR and identified overfitting problems with CART.

Liu and Brown [20] and Wang et al. [32] have considered demographic, spatial, temporal and social-media dependent models. Particularly Liu and Brown [20] propose a transition density model that takes into account demographic, economic, social, victim and spatial attributes of criminal activity.

Our approach is mainly inspired by the work of Flaxman [20], Liu and Brown [13] and Liu and Brown [32]. We derive a general probabilistic model that can capture generic features across space and that can consider spatial correlations using a non-parametric component. As noted by Weisburd et al. [33], quantitative studies in criminology focus on 'mechanical' reporting of estimates and predictive power. These studies ignore the uncertainty around these estimates. In contrast, our Bayesian approach is fully probabilistic and quantifies all sources of uncertainty, which is necessary for effective policy and decision making.

Methodology

In this section we present the methods and probabilistic models used for inference and prediction regarding crime rates. We take a Bayesian approach, and use the posterior mean for prediction and make inference via the marginal posterior distribution of the quantity of interest. To do this, we propose a probabilistic model which is prescribed by a set of parameters which we denote by $\boldsymbol{\theta}$. Inference regarding these parameters is made via the *posterior* probability distribution denoted by $p(\boldsymbol{\theta}|\mathcal{D})$, where $\mathcal{D}$ is a dataset and the notation , means "conditional on". This posterior distribution is given by Bayes theorem to be

$$p(\boldsymbol{\theta}|\mathcal{D}) = \frac{p(\mathcal{D}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathcal{D})}. \tag{1}$$

The term $p(\mathcal{D}|\boldsymbol{\theta})$ represents the *likelihood* of the data being generated, given the parameters $\boldsymbol{\theta}$, and $p(\boldsymbol{\theta})$ is known as the *prior* probability distribution, which encodes prior knowledge about these parameters. The term $p(\mathcal{D})$ is the marginal probability distribution of the data. It is a normalising constant and it is independent of the parameters $\boldsymbol{\theta}$.

Marchant *et al. Secur Inform* (2018) 7:1

Page 4 of 19

In the following sections we:

1. Describe the regression model for crime rate assuming that the noise of this model is spatially independent.
2. Discuss a mixture model that is aware of spatial correlations.
3. Present the algorithms used to learn the model parameters from the data.

Bayesian linear regression with i.i.d. errors

Suppose we wish to predict the crime rate, y, or a particular offence, at location i, conditional on a number location-specific characteristics, contained in $\mathbf{x}$.[1] One approach is to assume that the observed (log) crime rate y_i, is a combination of a signal, f, corrupted by noise, e_i, such that

$$y_i = f(\mathbf{x}_i) + e_i. \tag{2}$$

The noise is assumed to conform to some distribution, which in this case we assume to be Gaussian, so that $e \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_e^2)$, where i.i.d refers to independently and identically distributed samples. In general, the signal f can take any functional form, however in linear regression it is assumed to be linear, so that $f(\mathbf{x}_i) = \mathbf{x}_i \boldsymbol{\beta}$, where $\mathbf{x}_i = (1, x_{i1}, \ldots, x_{iP})$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_P)$, where x_{ik} is the ith observed value of characteristic k, and P is the number characteristics.

The parameters that fully specify this model are given by $\boldsymbol{\theta} = \{\boldsymbol{\beta}, \sigma_e\}$, the data are denoted by $\mathcal{D} = (X, \mathbf{y})$, where $X = (\mathbf{x}_1', \ldots, \mathbf{x}_n')'$, $\mathbf{y} = (y_1, \ldots, y_n)'$ and n is the number of locations with recorded crime rates and their respective location-specific characteristics.

If the e_i's in (2), conform approximately to our assumptions, i.e $e \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_e^2)$, then the likelihood $p(\mathbf{y}|X, \boldsymbol{\theta})$ is the familiar multivariate Gaussian distribution. Similarly, the predictive distribution of the unobserved crime rate at a particular location, $y^\star$, with characteristics $\mathbf{x}^\star$, is given by $p(y^\star|\mathbf{x}^\star, \mathcal{D})$ and is equal to

$$p(y^\star|\mathbf{x}^\star, \mathcal{D}) = \int_{\mathbb{R}^{|\theta|}} p(y^\star|\mathbf{x}^\star, \mathcal{D}, \boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathcal{D}) \mathrm{d}\boldsymbol{\theta}. \tag{3}$$

For computational ease, we follow Zellner (1986) and use a g-prior for $p(\boldsymbol{\theta})$, given in (4). This ensures that the posterior distributions $p(\boldsymbol{\theta}|\mathcal{D})$ and $p(y^\star|\mathbf{x}^\star, \mathcal{D})$ are available analytically [6].

Our model and priors for (2) can be written as,

$$
\begin{aligned}
y_i|\boldsymbol{\beta}, \sigma_e^2 &\sim \mathcal{N}(\mathbf{x}_i \boldsymbol{\beta}, \sigma_e^2) \\
\boldsymbol{\beta}|\sigma_e^2 &\sim \mathcal{N}(0, g\sigma_e(X'X)^{-1}) \\
p(\sigma_e^2) &\propto 1/\sigma_e^2
\end{aligned}
\tag{4}
$$

Bayesian linear regression with spatial dependency

When dealing with spatial data, such as described in this paper, it is unrealistic to assume that the crime rate is only dependent on those location specific characteristics which are measured, because locations that are close to each other in space are likely to be correlated. So we relax the assumption that errors in $\boldsymbol{e} = (e_1, \ldots, e_n)$ are independent and modify (2) to become,

$$y_i = f(\mathbf{x}_i) + h(\mathbf{u}_i) + \epsilon_i, \tag{5}$$

where the error terms are given by[2] $\epsilon_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_\epsilon^2)$ and $\mathbf{u}_i$ is the vector of spatial coordinates of location i and $h(\mathbf{u})$ is a nonparametric function of $\mathbf{u}$. In addition we assume that the relation between crime rate location-specific characteristics in $\mathbf{x}$, is independent of the relationship between crime rate and spatial co-ordinates in $\mathbf{u}$.

We place a Gaussian Process Prior (GPP), over the unknown function $\mathbf{h} = (h_1, \ldots, h_n)'$, which is equivalent to stating that $\mathbf{h} \sim \mathcal{N}(0, K(\boldsymbol{\Phi}))$, and $K(\boldsymbol{\Phi})$ is an $n \times n$ matrix, with hyperparameters contained in $\boldsymbol{\Phi}$, and ith, jth element, $k_{ij}(\cdot, \cdot) = k(\mathbf{u}_i, \mathbf{u}_j)$, equal to cov $(h(\mathbf{u}_i), h(\mathbf{u}_j))$. There are many options for the particular form of k_{ij}, see [29, p. 94]. For example, the isotropic squared exponential covariance function given by

$$k(\mathbf{u}_i, \mathbf{u}_j|\boldsymbol{\Phi}) = \sigma_f^2 \exp\left(-\frac{\|\mathbf{u}_i - \mathbf{u}_j\|^2}{2l^2}\right). \tag{6}$$

In this example $\boldsymbol{\Phi} = (\sigma_f^2, l)$, where σ_f^2 controls the amplitude of the unknown function, and l controls the variability of the function across space. If f is linear in $\mathbf{x}$, the full set of parameters that specify the model in Eq. 5 are $\boldsymbol{\theta} = \{\boldsymbol{\beta}, \sigma_\epsilon, \boldsymbol{\Phi}\}$.

The combination of using a parametric model for the relationship between crime rate and location specific characteristics and an additive nonparametric model for spatial dependencies, serves two purposes. First, the model is interpretable. In particular the regression coefficients, $\boldsymbol{\beta}$, represent the proportional change in crime rate which will result from the same proportional change in a location-specific characteristic, after controlling for other

---

[1] We have chosen to use the log of crime rate as the dependent variable, and the log of the non-zero location-specific characteristics as the independent variables because the relationship between these two sets of variables is approximately linear and the residuals approximately normally distributed.

[2] Note that the spatially independent error terms e_i have been replaced by ϵ_i.

Marchant *et al. Secur Inform* (2018) 7:1

Page 5 of 19

non-observable factors which are a function of space, captured by $h(\boldsymbol{u})$. Second, by placing a flexible, nonparametric prior over the function $h(\boldsymbol{u})$ we are allowing the data to uncover spatial dependencies rather than enforce a parametric form. Thus the model is both parsimonious and flexible and therefore allows for accurate predictions while remaining interpretable.

Our model and priors for (5) is then

$$\mathbf{y}|\boldsymbol{\beta}, \sigma_\epsilon^2, \sigma_f^2, l \sim \mathcal{N}(X\boldsymbol{\beta}, K(\boldsymbol{\Phi}) + I_n \sigma_e^2)$$
$$\boldsymbol{\beta}|\sigma_e^2 \sim \mathcal{N}(0, g\sigma_\epsilon(X'X)^{-1})$$
$$p(\sigma_\epsilon^2) \propto 1/\sigma_\epsilon^2$$
$$p(\sigma_f^2) \propto 1/\sigma_f^2$$
$$p(l) \propto 1/l$$

Inference and prediction are made via the posterior distributions, $p(\boldsymbol{\theta}|\mathcal{D})$ and $p(y^\star|\mathbf{x}^\star, \mathcal{D})$, as in "Bayesian linear regression with i.i.d. errors" section, except that these quantities are no longer available analytically and we use Markov chain Monte Carlo (MCMC) to perform the required numerical multidimensional integration.

Inference via Markov chain Monte Carlo (MCMC)

To carry out inference via the posterior distribution requires a multidimensional integration. MCMC is a very efficient way of achieving this. There are other methods to perform a multidimensional integration such as importance sampling or particle filters, but these are not usually as efficient as MCMC. There are also other methods which approximate the posterior such as variational inference, which is faster than MCMC and therefore particularly useful with very large datasets, but less accurate.

In this section we describe how MCMC is used to approximate the posterior distributions, $p(\boldsymbol{\theta}|\mathcal{D})$ and $p(y^\star|\mathbf{x}^\star, \mathcal{D})$ when no closed form exists for these distributions. The predictive posterior distribution in Eq. 3 can be approximated by

$$p(y^\star|\mathcal{D}) = \int_{\mathbb{R}^{|\boldsymbol{\theta}|}} p(y^\star|\mathcal{D}, \boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathcal{D}) \mathrm{d}\boldsymbol{\theta} \approx \frac{1}{M} \sum_{k=1}^{M} p\left(y^\star|\mathcal{D}, \boldsymbol{\theta}^{[k]}\right), \tag{7}$$

where $\boldsymbol{\theta}^{[k]}$ are drawn from the joint posterior distribution $p(\boldsymbol{\theta}|\mathcal{D})$. Note that the dependency on $\mathbf{x}^\star$ has been obviated for notation purposes. Construction of an efficient MCMC scheme requires the development of a transition kernel which satisfies certain conditions. One of these conditions is that the chain is reversible, and this condition is ensured by using the Metropolis–Hastings algorithm. Draws from the joint posterior are obtained using

MCMC with a Metropolis–Hasting transition kernel to move the chain around the parameter space. Algorithm 1 shows the pseudo code of the Metropolis–Hastings MCMC algorithm.

Algorithm 1 Metropolis-Hastings MCMC

1: Initialise $\boldsymbol{\theta}^{[0]}$.
2: **for** $k = 1$ to $M - 1$ **do**
3:     Sample $u \sim U(0, 1)$.
4:     Sample $\boldsymbol{\theta}^p \sim q(\boldsymbol{\theta}^p|\boldsymbol{\theta}^{[k]})$
5:     **if** $u < \alpha = \min\left\{1, \frac{p(\mathcal{D}|\boldsymbol{\theta}^p)p(\boldsymbol{\theta}^{[k]}|\boldsymbol{\theta}^p)}{p(\mathcal{D}|\boldsymbol{\theta}^{[k]})p(\boldsymbol{\theta}^p|\boldsymbol{\theta}^{[k]})}\right\}$
     then
6:         $\boldsymbol{\theta}^{[k+1]} = \boldsymbol{\theta}^p$
7:     **else**
8:         $\boldsymbol{\theta}^{[k+1]} = \boldsymbol{\theta}^{[k]}$
9:     **end if**
10: **end for**

Application: regression over crime rates

In this section, we apply the proposed methodology to model particular types of criminal offences—Domestic Violence (DV) related assaults, Burglaries and Motor Vehicle Theft (MVT), in New South Wales (NSW), Australia. There are two goals in this section, the first is to evaluate the predictive performance of our technique and the second is to evaluate the ability of the model to make meaningful inference regarding the drivers behind specific crime types. The remaining of this section is organised as follows. "The data" section presents a description of the data used for building the models. "MCMC for learning the model" section describes the procedures and specific information for learning the models from the data. Then, "Evaluation of models" section evaluates independent models for each crime type for a specific year and explores in detail the model for DV related assaults across a ten year period. Finally, "Discussion" section presents a discussion of the results and relations with existing research in the area.

The data

Criminal incident data over the time period 1997–2015 was extracted from the Unit-Record Criminal Incident Dataset provided by the NSW Bureau of Crime Statistics and Research (BOCSAR). The spatial information provided on each crime incident is a geographical area identifier, called Statistical Area Level 2 (SA2). SA2s are geographical areas that present a relatively homogeneous population distribution. At this level of granularity, is it possible to visualise interesting patterns while preserving the privacy of the individuals.

Marchant *et al. Secur Inform* (2018) 7:1

Page 6 of 19

Table 1 Demographic features and summary statistics across statistical areas based on the ABS census data 2011

Variable description	Min	Max	Mean	SD
Number of separated males (per 100 total males)	1	4	2	1
Percentage of unemployment	2	14.3	6	2.2
Population density (per km^2)	0.02	14301	1466	2027
Median total household weekly income	618	2610	1264	449
Median mortgage monthly repay	300	3289	1898	535
Median rent weekly	50	690	292	112
Percentage of people with no religion	4	42	18	7
Median age	22	59	39.16	5.31
Percentage of immigrants	2	63	21	15
Percentage of English-only speakers	13	97	78	21
Percentage with vocational education only (Certificate Level 1 or 2 per all levels)	3	13	7	1
Number of families with lonely parent (per 100 total population)	1	10	4	1

Explanatory variables are selected from demographic data at SA2 level, extracted from census data for the years 2001, 2006, and 2011 (the latest census data available at the time of writing). This information is publicly available from the Australian Bureau of Statistics (ABS). The twelve demographic features and their summary statistics are presented on Table 1.

Crime counts for specific crime types are aggregated over space across SA2 and crime rates are calculated using the corresponding population information (per one thousand people). We have excluded regions with a population lower than 1000, such as National Parks and Airports, which results into a total of 512 SA2 regions being subject of the study (from a total of 540). All data are standardised before training the proposed models, which assures that posterior probability distributions for the linear component parameters are comparable.

The method can cope with data of various granularity levels, dealing with the issues described by Andersen and Malleson [1], where they note that the results are different at alternative spatial aggregation scales.

MCMC for learning the model

The implementation of Algorithm 1 and its application to learn the model described in Eq. was conducted by using an existing Python package called *emcee*, which is an affine-invariant ensemble sampler for MCMC that has been well tested for a large range of machine learning applications [14, 15]. The algorithm uses the Metropolis–Hasting acceptance criteria, but rather than having one sampler, the algorithm evolves an ensemble of multiple walkers which explore the parameter space much faster. To propose a new position for one walker, the algorithm selects another walker at random from the rest of the ensemble and chooses a new position that is a random linear combination of the positions of both walkers. We place a uniform distribution for the initial value for each MCMC chain (Line 1 of Algorithm 1) over the relevant range in the parameter space. The overall estimation is conducted with 200 chains, each with 1000 iterations after a burn-in phase of 500 iterations, which removes large initial fluctuations in the parameter space. The convergence of each chain can be inspected on the individual sample plots for each parameter in the "Appendix" (Figs. 8 and 9).

Evaluation of models

This section shows the results of applying the proposed methodology over different scenarios. It presents results on the predictive and generalisation capabilities of the proposed methodology. To evaluate the predictive capabilities of the model and to control for overfitting we split the dataset randomly by geographical regions into train and test data with a ratio of 9:1, respectively. Test data is *hidden* from the model for learning process and the predictive distribution was obtained according to Eq. 7 for these test and train locations. The target variable is the crime rate of each crime type at SA2 areas, while explanatory (or independent) variables are demographic features of the location where the incidents occurred.

Three crime types

We independently modeled three different crime categories: Domestic Violence (DV) assaults, Burglaries (break/enter and stealing) and Motor Vehicle Theft (MVT), for the period 2009–2013.[3] These models were implemented based on spatial dependencies and demographics, as proposed in "Bayesian linear regression with spatial dependency" section.

---

[3] We aggregated crime data over 5 years, around the 2011 census data, to achieve statistical significant inference for long-term decision making.

Marchant *et al. Secur Inform* (2018) 7:1

Page 7 of 19

Figures 1, 2 and 3 plot observed against predicted crime rate and show the diagnostics for each model and the fit for DV-related assaults, Burglaries and MVT respectively. Each point in the plots represents a region, and it can be concluded from visual inspection that there is a correlation between predicted and observed log crime rates for all the selected crime types. These figures also show that the residuals are independent and follow a Gaussian
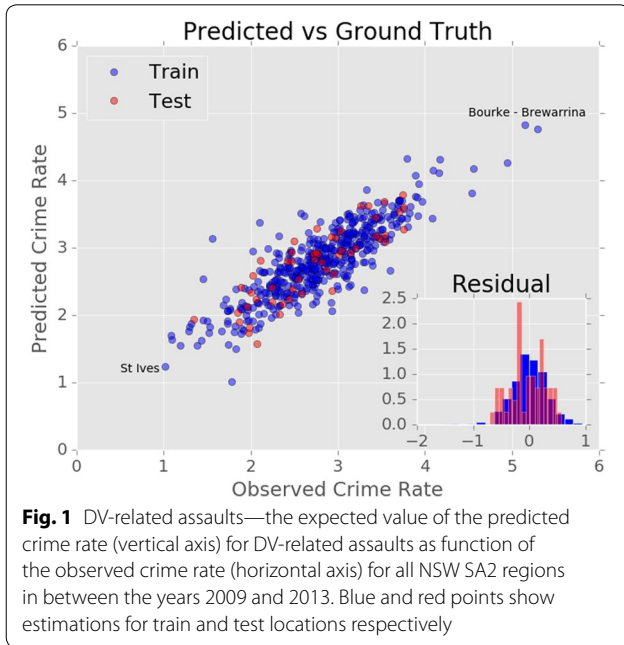
Fig. 1 DV-related assaults—the expected value of the predicted crime rate (vertical axis) for DV-related assaults as function of the observed crime rate (horizontal axis) for all NSW SA2 regions in between the years 2009 and 2013. Blue and red points show estimations for train and test locations respectively
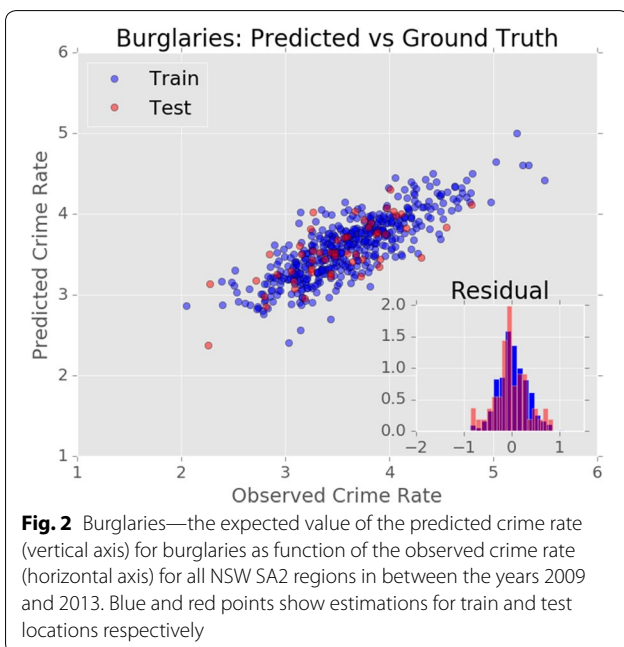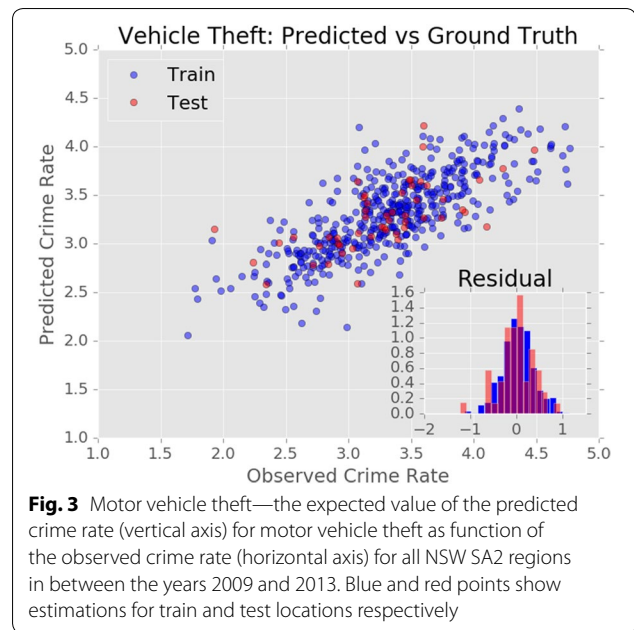

Fig. 2 Burglaries—the expected value of the predicted crime rate (vertical axis) for burglaries as function of the observed crime rate (horizontal axis) for all NSW SA2 regions in between the years 2009 and 2013. Blue and red points show estimations for train and test locations respectively

Fig. 3 Motor vehicle theft—the expected value of the predicted crime rate (vertical axis) for motor vehicle theft as function of the observed crime rate (horizontal axis) for all NSW SA2 regions in between the years 2009 and 2013. Blue and red points show estimations for train and test locations respectively

distribution, as assumed by the model given in "Bayesian linear regression with spatial dependency" section.

A quantitative estimation of the error is the root mean squared error (RMSE) and is calculated over the log crime rate (RMSE$_{rate}$) and over the crime counts (RMSE$_{counts}$) according to the following expressions:

$$\text{RMSE}_{\text{rate}} = \sqrt{\frac{\sum_{j=1}^{N} (y_j - y_j^\star)^2}{N}}, \tag{8}$$

$$\text{RMSE}_{\text{counts}} = \sqrt{\frac{\sum_{j=1}^{N} P_j^2 (e^{y_j} - e^{y_j^\star})^2}{N}}, \tag{9}$$

where y_j is the observed log crime rate at location j, $y_j^\star$ is the posterior mean estimate of the log crime rate at location j, N is the number of locations in the test/train set, and P_j is the total population at location j (in thousands). We calculate the error in the number of crimes to contextualise the magnitudes of crime incidents in the discussion.

We have also calculated the percentage of predictions within Credible Intervals (CI).[4] The CI are calculated based on the posterior predictive density, given by Eq. 7. The % of predictions within CI represents an accuracy measure with respect to uncertainty quantification. If the

---

[4] Credible Intervals differ from Confidence Intervals in that credible intervals are associated with posterior distributions, while confidence intervals often assume that the distribution of the sampling estimates are Gaussian.

Marchant *et al. Secur Inform* (2018) 7:1

Page 8 of 19

assumptions of our model are correct, we would expect that 95% of the actual crime rate at test locations to lie within the 95% predictive posterior distribution.

Table 2 shows several model performance indicators, RMSE, % within CI and correlation between estimated and observed values over the train and test data for every crime type. The main indicators are that there is a high linear correlation between predicted and observed values for every crime type, which indicates that the model is finding associations within the covariates and using that information to explain the crime rate. The first two numeric columns of Table 2 show a similar magnitude of the error for different crime types. We note that the performance indicators suggest that the predictions for DV-related assaults are more accurate than those for Burglaries and MVT, i.e. lower RMSE, higher % within CI and higher correlation between predictions and observations. The reasons behind this will be discussed in "Discussion" section.

Tables 3, 4 and 5 show the summary statistics for the marginal posterior distributions for the model parameters of DV-related assaults, burglaries and MVT

respectively. A graphical representation of these values is shown as box plots in Fig. 4. These values are calculated according to Eq. 7 using Algorithm 1. The dimensionality of $\boldsymbol{\theta}$ is 16 and is composed of: the regression co-efficients, $\boldsymbol{\beta}$, which corresponds to an intercept, β_0, and one parameter for each demographic feature, $\beta_1 \ldots \beta_{12}$; σ_ϵ, that represents the standard deviation of the noise in the process; and $\boldsymbol{\Phi}$, which has a dimensionality of 2 and contains the length- scale and signal variance parameters in the covariance kernel for the Gaussian Process, given by Eq. 6.

DV-related assaults

Further analysis is conducted over DV-related assaults to study the advantages of the proposed methodology and explore variations of the results over a 10 year period.

Advantage of semi-parametric modelling

Firstly, we compare and evaluate the advantages of incorporating spatial dependence in our model. We compare the spatially dependent model, given by Eq. 5, with the

Table 2 Error statistics for the models for different crime types [DV-related assaults, Burglaries, and Motor-Vehicle-Theft (MVT)] for the period 2009–2013

	RMSE log crime rate		RMSE crime count		% within CI		Correlation Pred/Obs	
	Train	**Test**	**Train**	**Test**	**Train**	**Test**	**Train**	**Test**
DV	0.314	0.318	92.1	88.4	94	98	0.86	0.85
Burglaries	0.301	0.350	191.5	208.0	90	90	0.84	0.72
MVT	0.345	0.372	191.7	191.3	89	88	0.78	0.66

Root-mean-squared-error (RMSE) over log crime rate, crime counts, percentage of locations within the 95% Credible Interval (CI) and correlation between predicted and observed values

Table 3 DV-related assaults inference—summary statistics for regression parameters for DV-related assaults between 2009 and 2013

Parameter	Posterior mean	Posterior SD	95% Credible Interval
Percentage of separated males	1.40	0.10	[1.20, 1.59]
Population density	0.96	0.13	[0.68, 1.20]
Percentage of unemployment	0.70	0.14	[0.42, 0.98]
Percentage of English speaking only	0.65	0.14	[0.37, 0.94]
Percentage of people Cert 1 or 2	0.53	0.11	[0.32, 0.75]
Number of families with lone parent	0.23	0.14	[− 0.04, 0.53]
Median total household income	− 0.09	0.18	[− 0.45, 0.28]
Percentage of immigrants	− 0.19	0.15	[− 0.45, 0.14]
Percentage of people with no religion	− 0.49	0.09	[− 0.68, − 0.31]
Median mortgage monthly repay	− 0.66	0.28	[− 1.22, − 0.10]
Median rent	− 0.91	0.24	[− 1.40, − 0.43]
Median age	− 1.05	0.13	[− 1.29, − 0.77]

The table shows the regression coefficient's posterior mean, SD and 95% credible intervals. Rows are sorted based on descending posterior mean

Marchant *et al. Secur Inform* (2018) 7:1

Page 9 of 19

Table 4 Burglaries inference—summary statistics for regression parameters for burglaries between 2009 and 2013

Parameter	Posterior mean	Posterior SD	95% Credible Interval
Percentage of unemployment	1.19	0.15	[0.89, 1.49]
Percentage of people Cert 1 or 2	0.98	0.12	[0.73, 1.21]
Percentage of separated males	0.83	0.11	[0.62, 1.05]
Population density	0.72	0.14	[0.46, 1.01]
Percentage of English speaking only	0.45	0.17	[0.13, 0.80]
Median total household income	0.24	0.20	[− 0.17, 0.62]
Percentage of people with no religion	0.07	0.11	[− 0.14, 0.28]
Median mortgage monthly repay	0.05	0.29	[− 0.54, 0.64]
Number of families with lone parent	− 0.23	0.16	[− 0.55, 0.06]
Median age	− 0.32	0.14	[− 0.63, − 0.05]
Median rent	− 0.63	0.25	[− 1.12, − 0.12]
Percentage of immigrants	− 0.82	0.17	[− 1.15, − 0.47]

The table shows the regression coefficient's posterior mean, SD and 95% credible intervals. Rows are sorted based on descending posterior mean

Table 5 Motor vehicle theft (MVT) inference—summary statistics for regression parameters for MVTs between 2009 and 2013

Parameter	Posterior mean	Posterior SD	95% Credible Interval
Population density	1.81	0.15	[1.52, 2.10]
Percentage of separated males	1.29	0.11	[1.06, 1.51]
Percentage of people Cert 1 or 2	0.78	0.13	[0.52, 1.02]
Median rent	0.70	0.27	[0.14, 1.22]
Percentage of unemployment	0.39	0.16	[0.09, 0.72]
Percentage of English speaking only	0.21	0.17	[− 0.10, 0.56]
Percentage of people with no religion	0.19	0.11	[− 0.01, 0.42]
Median mortgage monthly repay	0.16	0.32	[− 0.50, 0.79]
Number of families with lone parent	− 0.42	0.16	[− 0.73, − 0.10]
Median total household income	− 0.47	0.21	[− 0.87, − 0.05]
Median age	− 0.98	0.15	[− 1.27, − 0.69]
Percentage of immigrants	− 1.20	0.18	[− 1.54, − 0.84]

The table shows the regression coefficient's posterior mean, SD and 95% credible intervals. Rows are sorted based on descending posterior mean

spatially independent one, given by Eq. 2. The results show that the RMSE decreases by ~13% (from 0.35 to 0.31), indicating that the data is indeed spatially dependent and allowing for this results in more accurate predictions. We also compared the performance of a purely spatial model, resulting in the RMSE increasing by 26% with respect to the semi-parametric combination of space and demographics (from 0.31 to 0.39). For reference, the RMSE at test locations for a naive model with no predictors is 0.552 (see Table 6). The naive model represents a basic assumption for the estimation of the crime rate, for which the expected value is independent of space and demographics, by becoming the mean crime rate across the entire state.

Figure 5 shows the posterior distribution of σ_e^2, σ_ϵ^2 and σ^2 (for a purely spatial model). The posterior distributions are shown as histogram across all MCMC chain iterations.

It can be seen that, as expected, by incorporating more information the noise standard deviation is reduced. The worst approach is to use a purely-spatial model. And by merging demographics and space, the explanatory and predictive power of the model is improved.

Even though the distribution of the noise in the semi-parametric model overlaps with the demographic only model, the semi-parametric model with a Gaussian Process over space is consistent with lower noise level.

Another interesting visualisation is shown in Fig. 6, where we plot a heat map of domestic violence across the state of NSW, Australia. The *actual* crime rates, shown in the top section, are compared to the predicted ones. The ability of the model to capture the spatial dependencies and provide accurate estimates of the true crime levels, based only on demographic and spatial information, is striking.

Marchant *et al. Secur Inform* (2018) 7:1

Page 10 of 19

Fig. 4 Box plot of the demographic regression coefficients for three different crime types: DV related assaults, burglaries and MVT
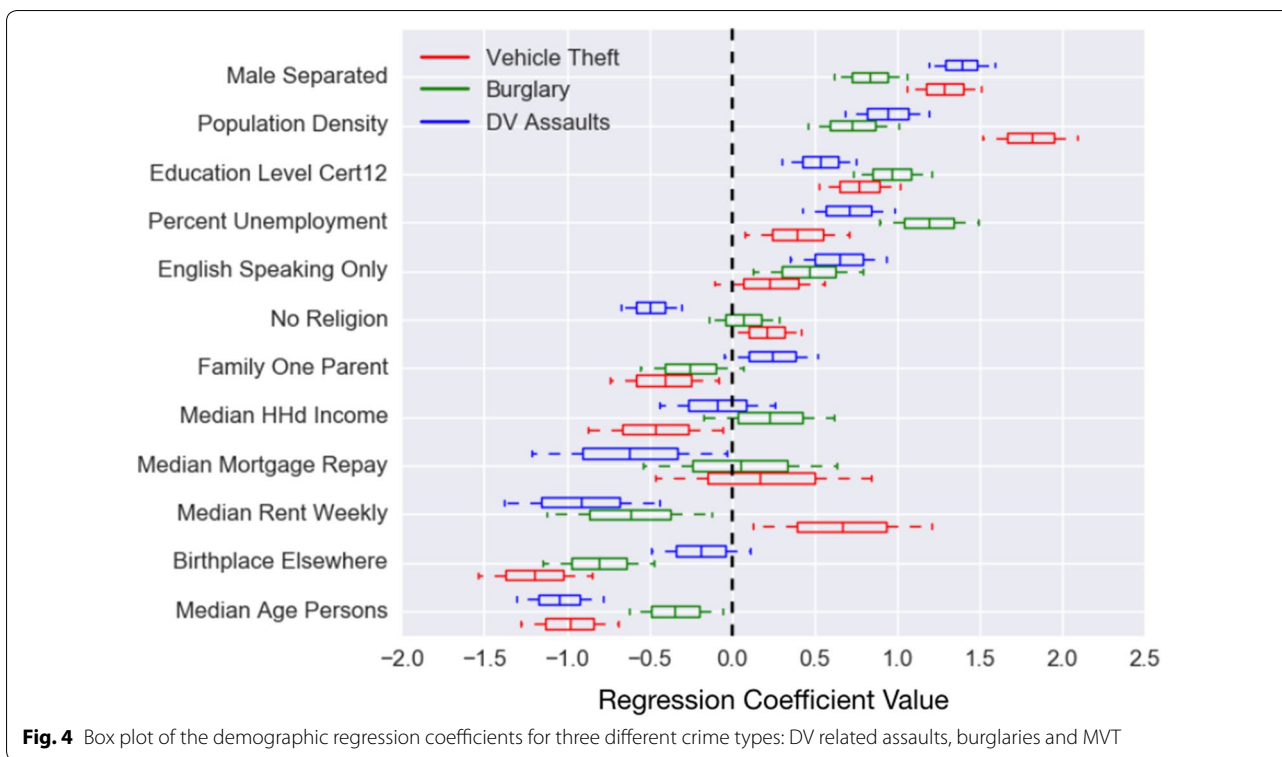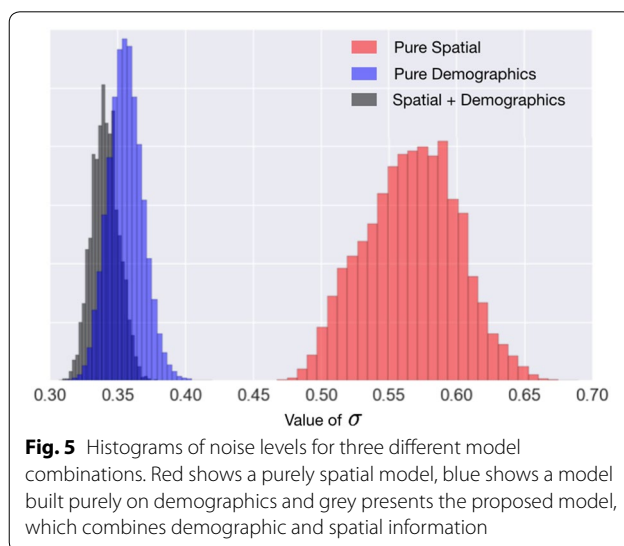
Table 6 Comparison of the RMSE between our spatial-demographic regression model and a naive model (average frequency over crime rates), for DV-related assaults in the period 2009–2013

	RMSE log crime rate		RMSE crime count	
	Train	Test	Train	Test
Spatial-dem. reg.	0.314	0.318	92.1	88.4
Naive	0.544	0.552	162.4	158.1

Robustness over time

We further explore the time-varying nature of the dependency between crime rate and demographic characteristics and spatial location by conducting three cross-sectional studies aggregating crime over three time periods 1999–2003, 2004–2008, and 2009–2013. Each period spans over 5 years and is centred around the Census 2001, 2006, and 2011 data. A boxplot of the draws of the regression coefficients from their posterior distribution for each demographic feature and time period distribution is given in Fig. 7. The box is defined by ± 1 standard deviation for a Gaussian distribution—and the

Fig. 5 Histograms of noise levels for three different model combinations. Red shows a purely spatial model, blue shows a model built purely on demographics and grey presents the proposed model, which combines demographic and spatial information

median as vertical line inside the box. The dashed horizontal line indicates the 96% confidence intervals, i.e. 2 and 98 percentiles. A positive value in a regression coefficient is associated with an increase in crime rate and vice-versa.

Marchant *et al. Secur Inform* (2018) 7:1

Page 11 of 19
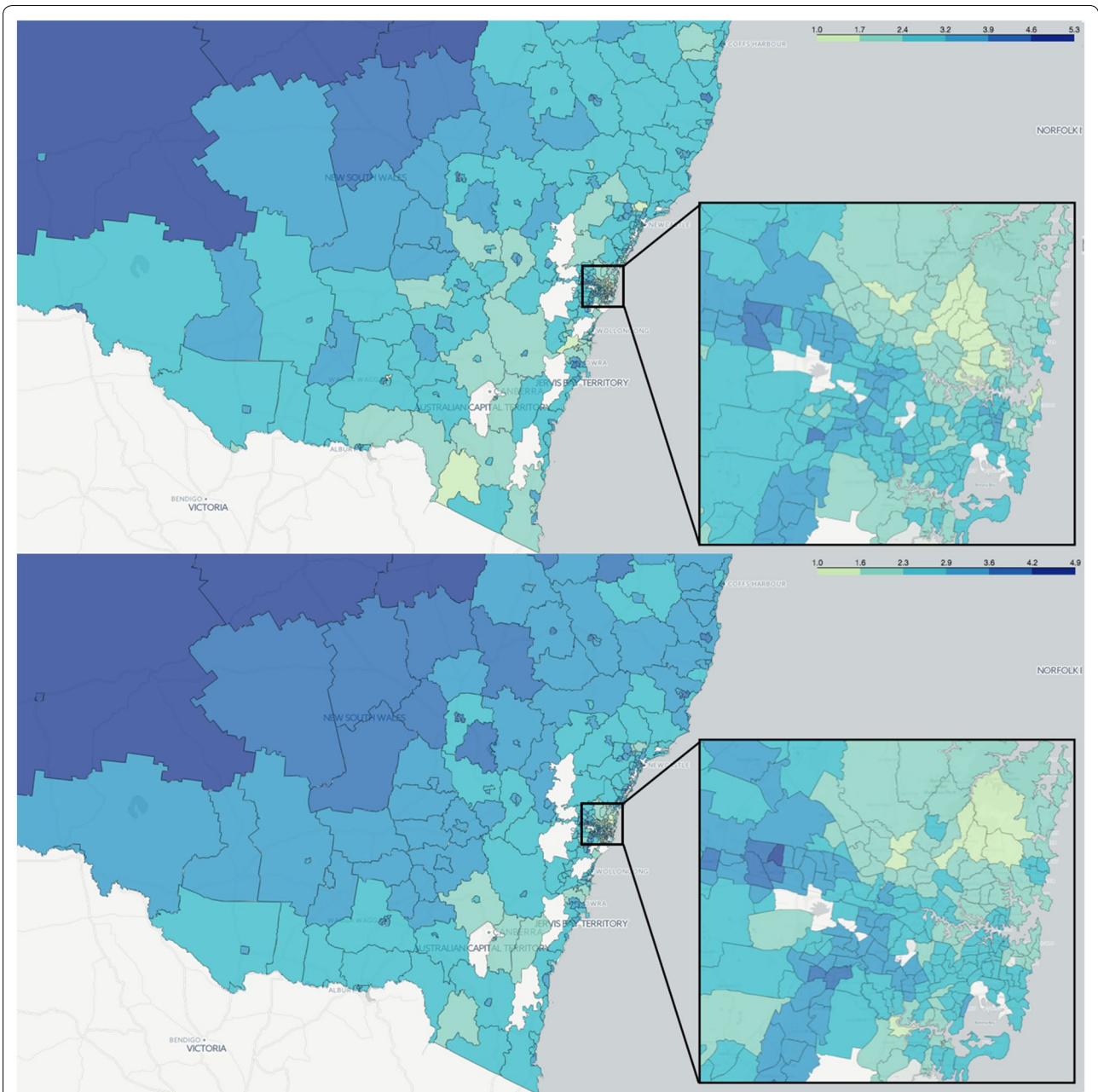
Fig. 6 Crime-rate heat map of DV related assaults (per 1000 people on natural logarithmic scale). Top: Ground Truth crime rate. Bottom: Spatial-demographic semi-parametric model of crime rate. The inset is the Sydney region, while the white areas are those for which there are no data because, for example, the location is in a national park

Marchant *et al. Secur Inform* (2018) 7:1

Page 12 of 19

Generalisation capabilities

In order to validate our model and verify for overfitting in a more principled manner, we have also conducted tenfold cross validation for DV-related assaults. The results show a mean RMSE of 0.30 ± 0.02 over the tenfold evaluations.

Discussion

In this section we analyse the result, link some of the results to existing criminological theory and compare with existing work in the area.

Inference on demographics

To understand how the selected demographic factors contribute to specific crimes type we need to look at the posterior distribution over the regression parameters. As described in "Bayesian linear regression with spatial dependency" section, the values of $\boldsymbol{\beta}$ can be interpreted as percentage increase in crime rate which would result from a percentage increase or each percentage increase

in the independent variable. In this particular case, each β_i represents how a unit increase/decrease in the percentage of demographic variable i is related to the percentage increase/decrease in the log crime rate.

Since we are using a fully probabilistic and multivariate Bayesian approach, the MCMC algorithm provides a joint probability density function for the whole parameter space. This joint density can be explored for each independent variable and the marginal distribution for each parameter is plotted in the "Appendix", Figs. 10 and 11 (only for DV-related assaults due to space constraints). It can be seen that all these variables are approximately Gaussian and Tables 3, 4 and 5 show the summary statistics for each variable for each crime type. A positive posterior mean is linked to an increase in the crime rate for the particular crime type. However, attention needs to be drawn to the Credible Interval (CI). If the CI contains zero, then there is a non negligible probability that this parameter is zero, implying that there is no relation between that specific demographic variable and crime rate. A shorter CI also represents lower uncertainty around the value of the specific regression
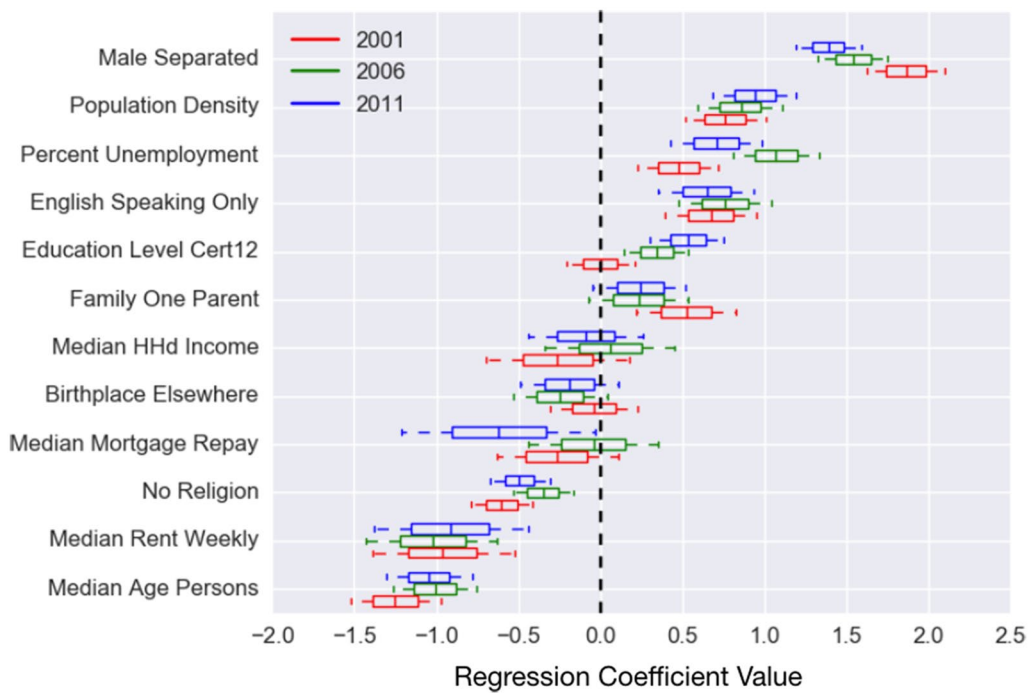
Fig. 7 Box plot of the regression coefficients across multiple time periods for demographic factors and DV related assaults

Marchant *et al. Secur Inform* (2018) 7:1

Page 13 of 19

coefficient, increasing trustworthiness of the relationship between that specific covariate and crime.

Box-plots of the regression coefficient samples drawn from their posterior distributions for the three different types of crime appear in Fig. 4. The box is defined by ± 1 standard deviation for a Gaussian distribution—and the median as vertical line inside the box. The dashed horizontal line indicates the 96% confidence intervals, i.e. 2 and 98 percentiles.

We have grouped variables into three categories. The first category consists of covariates that are unequivocally positively related to an increase in crime: Percentage of Separated Males, Population Density, Education and Unemployment. This is similar to the results reported in Nivette [25], who found that the proportion of males and population density where positively related to crime. The second category is composed by those covariates that have a negative relation with all three crime types, being Age and Immigrants. Lastly, the third category is encloses the covariates for which the impact varies across crime type: Religion, Lone Parent Family, Income, Mortgage and Rent.

The main observation is that some of the demographic factors such as rent, mortgage, and religion have different impacts on certain crime types. Of particular note is the fact that areas which have a high proportion of people claiming to be religious are less likely to experience theft or burglary but more likely to be victims of domestic violence. Similarly, areas with high mortgage/rental payments are less likely to experience domestic violence but more likely to experience theft or burglary. However, living in an area with a high immigrant population is associated with lower crime rates across all three crime types; lower theft, lower burglaries and lower domestic violence. One of the open questions, subject of future research, is whether immigration itself reduces the actual number of crimes committed in these areas, due to the selection process of the immigration office in terms of education and possible pre-offences, or alternatively only reduces the number of recorded crimes, e.g. due to withholding information or less willing to contact police.

It can be seen that the same demographic factors contribute in similar ways to DV related assaults across all years with the largest variation over time in the areas of education and unemployment. These results suggest, if data were available at a finer temporal resolution, that explicitly modelling time may show further variations.

Prediction errors

Regarding prediction errors, we suspect that the larger uncertainties in prediction for non-DV-related crime types is due to the fact that crimes such as MVT and burglaries are not necessarily committed by criminals living in same area, whereas most of DV assaults occur in the residence of the persons of interest (in fact, 81% of DV Assaults occur in a residential area). Since our current demographic model reflects only data of individuals living in that particular area, transient population are not currently taken into account, and thus lead to larger uncertainties in our predictions. For example, motor vehicle theft criminals focus on locations with numerous vehicles and low capable guardianship Cohen and Felson [8]. Thus, inclusion of variables that estimate ambient populations and consider the journey-to-crime literature, will enhance the quality of predictions of offences committed away from an individual's residential address.

Prediction errors and the patterns captured by the model, represented by the parametric regression component, will depend strongly on the selected subset of explanatory variables. We are actively working on including covariates of other domains, such as environmental features, and include these in the system for future research. However, there is no particular changes that need to be done to the proposed methodology, since the strength of our method is that any type of features can be included, i.e. the model does not limit the type of features included in $\mathbf{x}$.

Conclusion and future work

We have presented a fully probabilistic model that is able to accurately predict crime rates and provide uncertainties surrounding those predictions, while simultaneously providing inference over the possible location-specific factors associated with crime. The inference around model parameters is via their posterior distribution which is estimated via MCMC. The main strengths of this approach are that it is fully probabilistic and produces estimates of regression parameters and predictions, all with associated uncertainties and credible intervals. The performance of the proposed methodology has been validated with out of sample data and compared against naive crime models that assume independence with respect to demographics and space. The model also incorporates spatial dependency by placing a non-parametric prior over the evolution of the residuals across

space. The analysis included in this paper is conducted at a SA2 level but is general enough to allow other aggregation at other geographical segmentation units.

The results validate existing theoretical criminological tenets regarding the factors that are associated with high and low criminal activity, including bounds on the degree of the relation. The results also show how this model can be used for understanding different types of crime and what are the limitations depending on the location-specific characteristics used to describe that particular phenomenon.

The study is a cross sectional one, but it compares the results of different cross sections in time. This analysis shows that it would be beneficial to include a temporal component in the model explicitly and this is the subject of future research. The purpose of the model, given its current form, is to capture patterns at the regional and demographic macro levels, which is useful for long term decision making and resource allocation. There are benefits for including this seasonality for shorter term decision making in predictive policing and patrol planning, however, these are not the main applications of the proposed methodology and will be studied in the future.

The are many other areas for future research. For example, an important concern requiring ongoing consideration is the use of biased criminal record data to train the models and how that can affect the interpretation of the inference results. As acknowledged by Lum and Johndrow [21] and Mosher et al. [24], this is a problem widely shared by all quantitative methods that adjust model parameters based on previously collected datasets. In the case of crime, there is unknown over/under policing over certain groups of the population, which can be potentially reinforced when using results from models learnt from these data. This and many other discrimination issues are an active area of research, known as *fairness in machine learning*. In future work, we will include bias quantification and other sources of information that can help uncover the 'dark figure' of crime.

Additionally, future research will include many other factors which may contribute to crime such as green space coverage, street lighting, and transport by placing priors over the inclusion of a factor in a model to gauge the robustness of the finding to prior assumptions. The inclusion of these spatial and the previously mentioned temporal dimensions of crime, consistent with environmental criminological traditions, will further bolster the utility of the approaches adopted here. In so doing, predictions about crime in time and space will be improved, and policymakers will receive the advantage of measurements of uncertainty. This will allow for greater confidence in policy and resource allocation decisions of police, criminal justice and security-related agencies.

Author details
[1] Centre for Translational Data Science, The University of Sydney, Sydney, NSW 2006, Australia. [2] Sydney Informatics Hub, The University of Sydney, Sydney, NSW 2006, Australia. [3] Sydney Institute of Criminology, The University of Sydney, Sydney, NSW 2006, Australia. [4] Law School, The University of Sydney, Sydney, NSW 2006, Australia. [5] School of Mathematics and Statistics, The University of Sydney, Sydney, NSW 2006, Australia.

Appendix: Markov chain Monte Carlo (MCMC) log probabilities

The log probability of the BLR prior is given as

$$\log P_{PriorBLR} = -\left(\frac{N}{2}+1\right)\log\left(\frac{P}{P+1}\sigma_n^2\right) \\ -\frac{1}{2\sigma_n^2}(\beta-\mu)^T\left(\frac{1}{P}\mathbf{X}^T\mathbf{X}\mu\right)(\beta-\mu) \tag{10}$$

with P the number of locations, N the number of demographic features, and

$$\mu = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}. \tag{11}$$

The log likelihood of the Bayesian Linear Regression is

$$\log P_{LhBLR} = -\frac{1}{2}\sum_i^P\left[\log(2\pi\sigma_n^2)+\frac{1}{\sigma_n^2}\left(\mathbf{y}_i-\sum_j^N(\beta_j\mathbf{X}_{ij})\right)\right] \tag{12}$$

where i is the index over the data points and j the number of demographic features. The combined log posterior of the Bayesian Linear Regression is then the sum

$$\log P_{BLR} = \log P_{PriorBLR} + \log P_{LhBLR}. \tag{13}$$
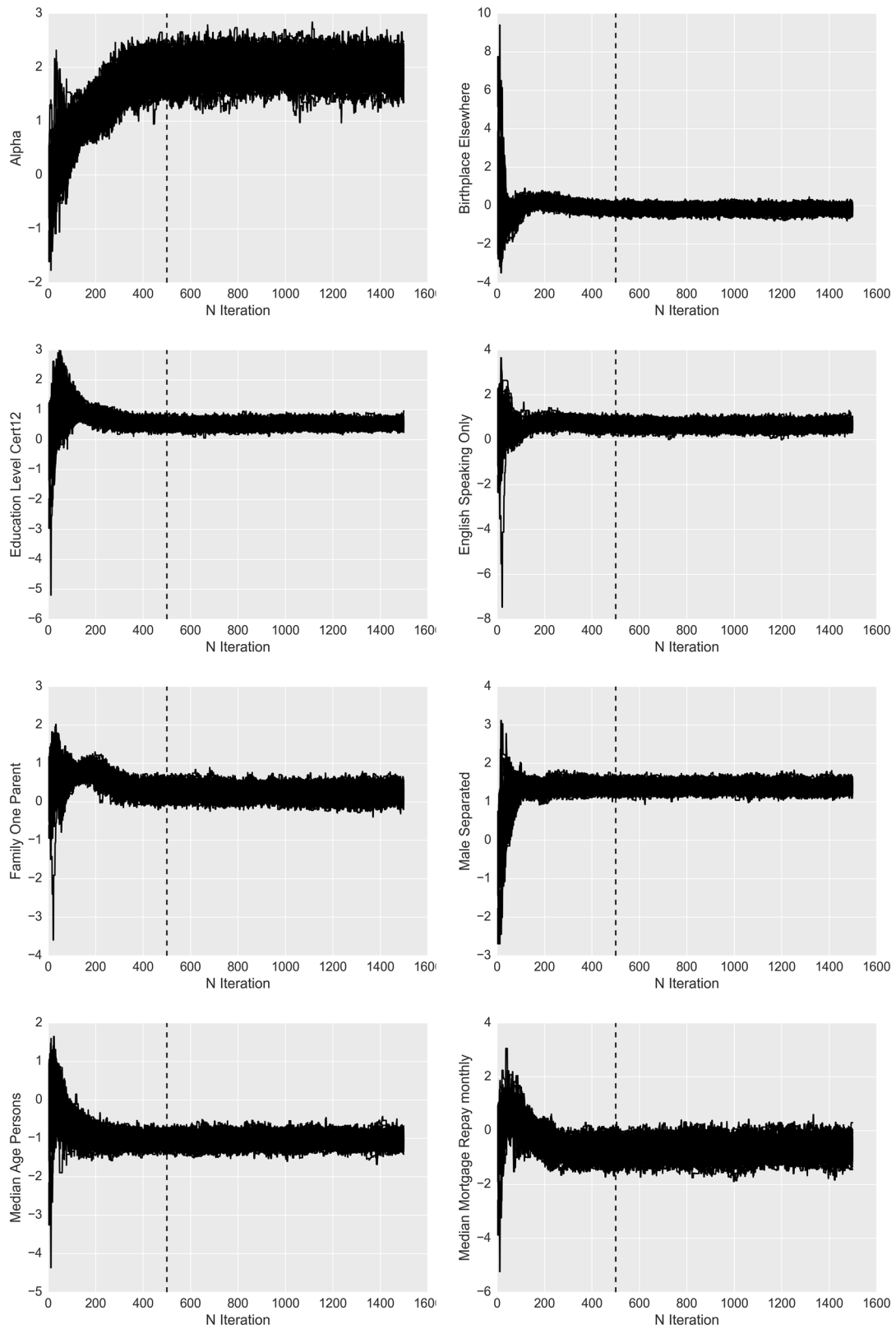
See Figs. 8, 9, 10 and 11.

Marchant *et al. Secur Inform* (2018) 7:1

Page 15 of 19

Fig. 8 Parameter plots for chains for all iterations. The dashed line indicates the MCMC burn-in phase at 500 iterations

Marchant *et al. Secur Inform* (2018) 7:1

Page 16 of 19

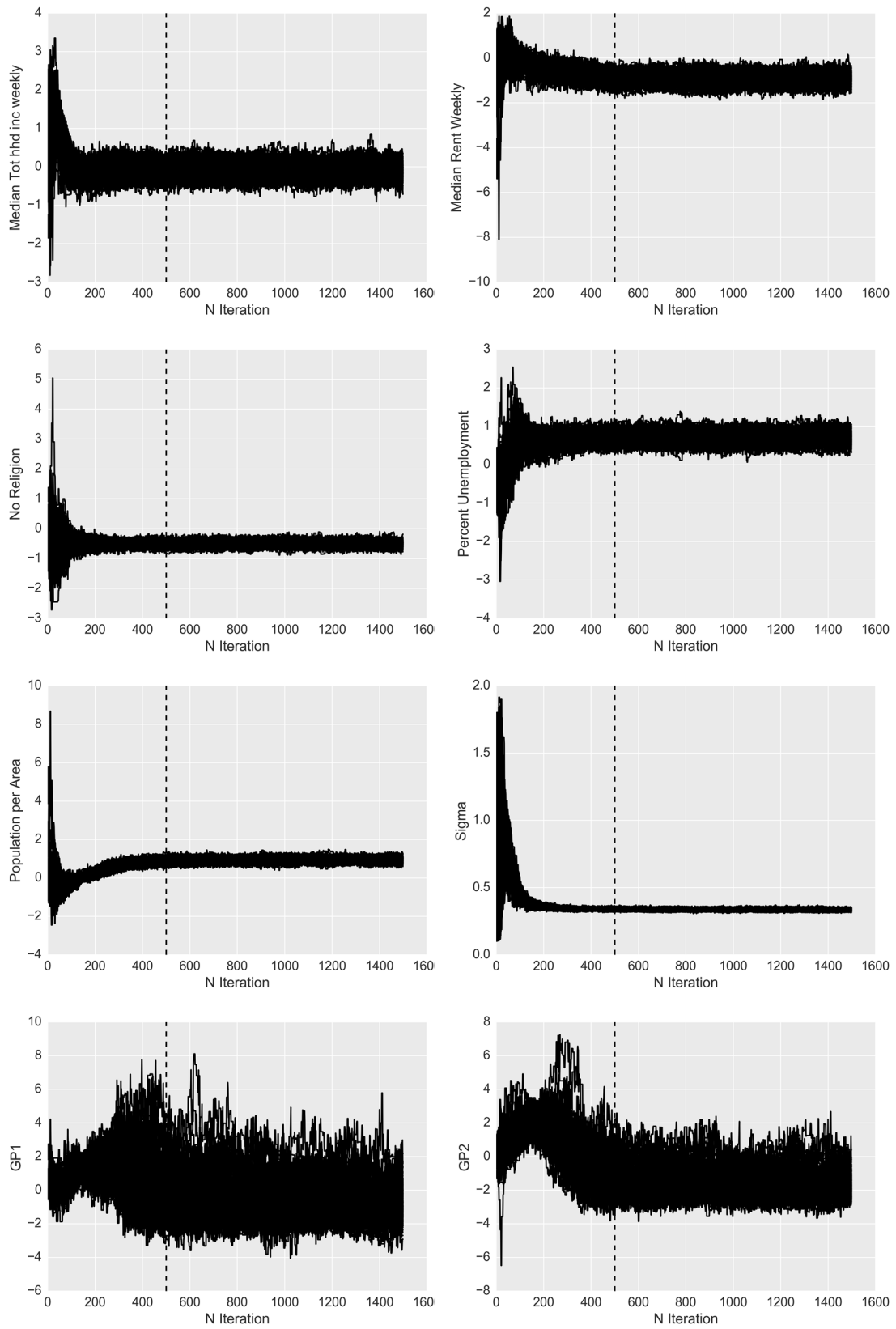**Fig. 9** Parameter plots with all chains for all iterations. The dashed line indicates the MCMC burn-in phase at 500 iterations

Marchant *et al. Secur Inform* (2018) 7:1
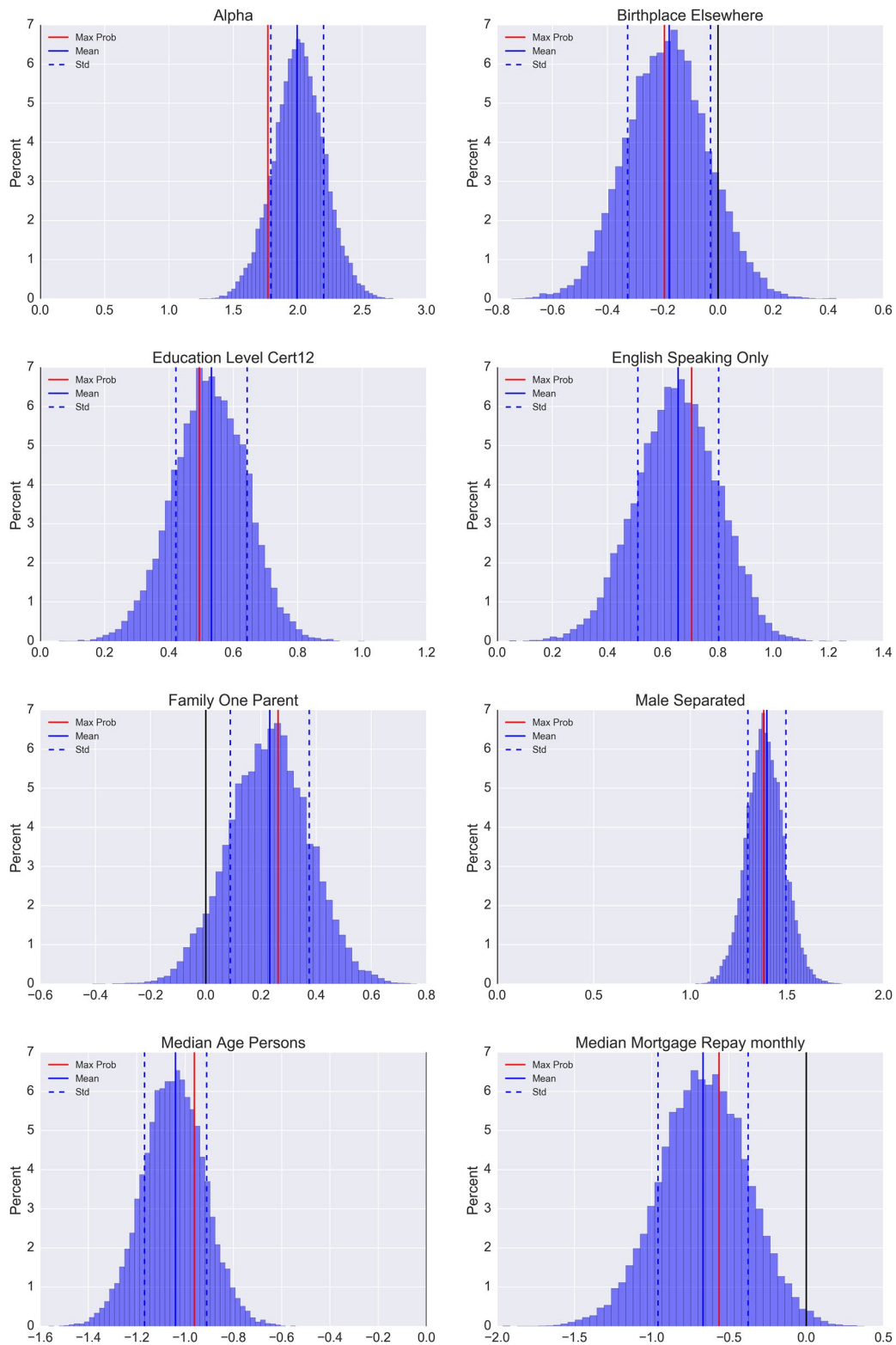
Page 17 of 19

Fig. 10 Histograms of MCMC samples for each parameter. The parameter value with maximum probability is indicated as red solid line; the mean and the standard deviation is indicated with a blue solid and dashed line, respectively
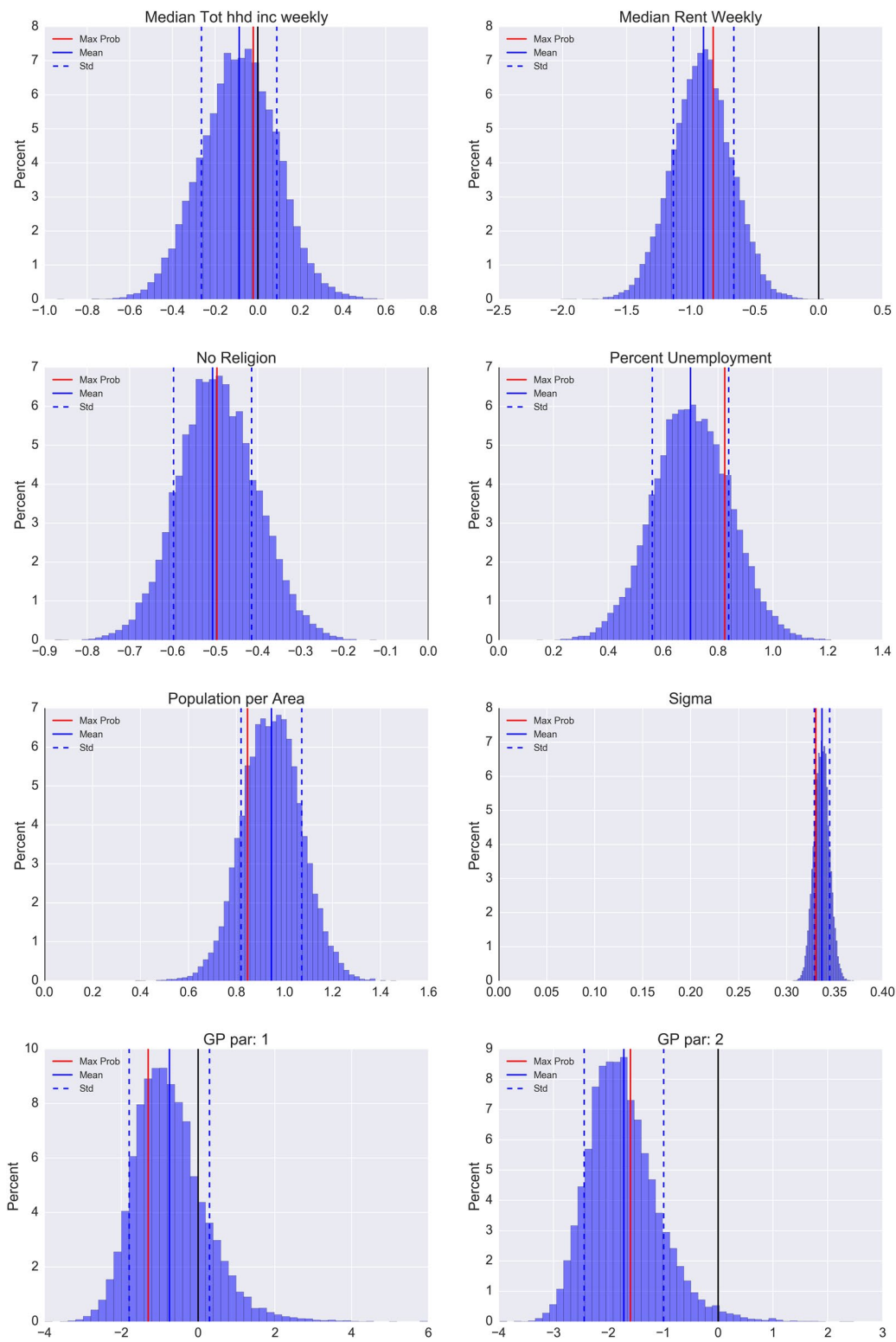
Marchant *et al. Secur Inform* (2018) 7:1

Page 18 of 19

Fig. 11 Histograms of MCMC samples for each parameter. The parameter value with maximum probability is indicated as red solid line; the mean and the standard deviation is indicated with a blue solid and dashed line, respectively

Marchant *et al. Secur Inform* (2018) 7:1

Page 19 of 19

Publisher's Note

References

1. Andersen MA, Malleson N (2013) Spatial heterogeneity in crime analysis. In: Leitner M (ed) Crime modeling and mapping using geospatial technologies. Springer, Berlin, pp 3–23
2. Antolos D, Liu D, Ludu A, Vincenzi D (2013) Burglary crime analysis using logistic regression. In: Human interface and the management of information, pp 549–558
3. Berk R, He Y, Sorenson SB (2005) Developing a practical forecasting screener for domestic violence incidents. Eval Rev 29(4):358–383
4. Boessen A, Hipp JR (2015) Close-ups and the scale of ecology: land uses and the geography of social context and crim. Criminology 53(3):399–426
5. Bowers K, Johnson S (2014) Crime mapping as a tool for security and crime prevention. In: Gill M (ed) The handbook of security. Springer, Berlin, pp 566–587
6. Box G, Tiao G (1973) Bayesian inference in statistical analysis. Addison-Wesley, Boston
7. Chainey S, Tompson L, Uhlig S (2008) The utility of hotspot mapping for predicting spatial patterns of crime. Secur J 21(1–2):4–28
8. Cohen L, Felson M (1979) Social change and crime rate trends: a routine activity approach. Am Sociol Rev 44:588–608
9. Corcoran JJ, Wilson ID, Ware JA (2003) Predicting the geo-temporal variations of crime and disorder. Int J Forecast 19(4):623–634
10. Davies TP (2015) Spatio-temporal modelling for issues in crime and security. Ph.D. thesis, University College London
11. Deadman D (2003) Forecasting residential Burglary. Int J Forecast 19(4):567–578
12. Eck JE, Chainey S, Cameron JG, Leitner M, Wilson RE (2005) Mapping crime: understanding hotspots. Department of Justice, Technical report, U.S
13. Flaxman SR (2014) A general approach to prediction and forecasting crime rates with Gaussian processes. Technical report, Carnegie Mellon University
14. Foreman-Mackey D, Hogg DW, Lang D, Goodman J (2013) emcee: the MCMC hammer. Publ Astron Soc Pac 125(925):306
15. Goodman J, Weare J (2010) Ensemble samplers with affine invariance. Commun Appl Math Comput Sci 5(1):65–80
16. Gorr W, Olligschlaeger A, Thompson Y (2003) Short-term forecasting of crime. Int J Forecast 19(4):579–594
17. Grubesic TH, Mack EA (2008) Spatio-temporal interaction of urban crime. J Quant Criminol 24(3):285–306
18. Kocher M, Leitner M (2015) Forecasting of crime events applying risk terrain modeling. J Geogr Inf Sci 2015:30–40
19. Leitner M (ed) (2013) Crime modeling and mapping using geospatial technologies. Springer, Berlin
20. Liu H, Brown D (2003) Criminal incident prediction using a point-pattern-based density model. Int J Forecast 19(4):603–622
21. Lum K, Johndrow JE (2016) A statistical framework for fair predictive algorithms. In: Workshop on fairness, accountability, and transparency in machine learning
22. Nogueira de Melo S, Pereira DV, Andresen MA, Fonseca Matias L (2017) Spatial/temporal variations of crime: a routine activity theory perspective. Int J Offender Ther Comp Criminol 62(7):1967–1991
23. Mohler G, Short M, Brantingham P, Schoenberg F, Tita G (2011) Self-exciting point process modeling of crime. J Am Stat Assoc 106(493):100–108
24. Mosher CJ, Miethe TD, Hart TC (2011) The mismeasure of crime. Sage Publications Inc, Thousand Oaks
25. Nivette AE (2011) Cross-national predictors of crime: a meta-analysis. Homicide Stud 15(2):103–131. https://doi.org/10.1177/1088767911140639 7
26. Osgood DW (2000) Poisson-based regression analysis of aggregate crime rates. J Quant Criminol 16(1):21–43. https://doi.org/10.1023/A:10075 21427059
27. Perry WL, McInnis B, Price CC, Smith SC, Hollywood JS (2013) Predictive policing, the role of crime forecasting in law enforcement operations. RAND, Santa Monica
28. Piquero AR, Weisburd D (eds) (2010) Handbook of quantitative criminology. Springer, New York
29. Rasmussen CE, Williams C (2006) Gaussian processes for machine learning. The MIT Press, Cambridge
30. Steenbeek W, Weisburd D (2015) Where the action is in crime? An examination of variability of crime across different spatial units in the Hague, 2001–2009. J Quant Criminol 32(3):449–469. https://doi.org/10.1007/s10940-015-9276-3
31. Tita GE, Radil SM (2009) Spatial regression models in criminology: modeling social processes in the spatial weights matrix. In: Piquero AR, Weisburd D (eds) Handbook of quantitative criminology. Springer, Berlin, pp 101–121
32. Wang X, Brown D, Gerber MS (2012) Spatio-temporal modeling of criminal incidents using geographic, demographic, and Twitter-derived information. In: International conference on intelligence and security informatics (ISI)
33. Weisburd D, Cave B, Piquero AR (2016) How do criminologists interpret statistical explanation of crime? A review of quantitative modeling in published studies. In: Piquero AR (ed) The handbook of criminological theory. Springer, Berlin, pp 395–414