

SPATIAL MODELING OF TRENDS IN CRIME OVER TIME IN PHILADELPHIA

BY CECILIA BALOCCHI AND SHANE T. JENSEN

University of Pennsylvania

Understanding the relationship between change in crime over time and the geography of urban areas is an important problem for urban planning. Accurate estimation of changing crime rates throughout a city would aid law enforcement as well as enable studies of the association between crime and the built environment. Bayesian modeling is a promising direction since areal data require principled sharing of information to address spatial autocorrelation between proximal neighborhoods. We develop several Bayesian approaches to spatial sharing of information between neighborhoods while modeling trends in crime counts over time. We apply our methodology to estimate changes in crime throughout Philadelphia over the 2006–15 period while also incorporating spatially-varying economic and demographic predictors. We find that the local shrinkage imposed by a conditional autoregressive model has substantial benefits in terms of out-of-sample predictive accuracy of crime. We also explore the possibility of spatial discontinuities between neighborhoods that could represent natural barriers or aspects of the built environment.

1. Introduction. Modeling and prediction of crime has always been of interest to local authorities, police departments and governments to ensure the safety of the population and more efficient law enforcement. Recent availability of detailed crime data has made this effort even more accessible to statistical practitioners and the general public.

As an example, the Philadelphia police department has released detailed information about reported crimes committed from 2006 to the present day.¹ The information about each reported crime includes the type of crime (which we will describe in Section 2), the date and time of the crime and the GPS location of the crime.

Using their reported crime data, many police departments have used statistical modeling procedures and algorithms to help predict locations of crimes for better prevention and faster intervention (Hvistendahl (2016)). The modeling of crime locations is not only useful for law enforcement but also for marketing strategies related both to real estate and commercial activities; for example, Trulia² uses

Received January 2019; revised June 2019.

Key words and phrases. Urbanism, crime, spatial, time trends.

¹<http://www.phlcrimemapper.com/>

²<https://www.trulia.com/>

crime data as part of their evaluation of the relative safety and attractiveness of different neighborhoods. In this paper we will focus on estimating changes in violent crimes over the past decade at a local neighborhood resolution which will involve both temporal and spatial modeling of crime.

Many different approaches have been taken to the modeling of the spatial distribution of crime. These approaches can be subdivided into two general categories, either modeling crime as a spatial point process using the specific locations of each reported crime (Flaxman (2014), Mohler et al. (2011), Taddy (2010)) or modeling crime as *areal* data, that is, totals aggregated within larger regions, as in Aldor-Noiman et al. (2016), Law, Quick and Chan (2014) and Li et al. (2014).

A common method for modeling spatial point processes is kriging or Gaussian process interpolation (Cressie (1990), Stein (1999)). This can be studied either with a classical approach, or with a Bayesian approach (Banerjee, Carlin and Gelfand (2014)). Alternative popular models consider other frameworks such as Gibbs point processes, Poisson processes and Cox processes; see Møller and Waagepetersen (2007).

Common classical methods for modeling areal data are spatial autoregressive models that include the Simultaneous Autoregressive Model (Whittle (1954)), the spatial Durbin model (Anselin (1998)) and the Conditionally Autoregressive Model (Besag (1974)); for a review of these and other methods see Pace and LeSage (2010). Many of these models have also been considered and used in a Bayesian framework (Banerjee, Carlin and Gelfand (2014)).

Our goal in this paper is the estimation of trends in violent crime over the past decade at a high resolution local neighborhood level throughout the city of Philadelphia. As it is well established that crime frequencies are spatially correlated (Brantingham and Brantingham (1984), Herbert (1982)), we need to create a model that allows the change in crime over time to be correlated by locally proximal neighborhoods. Our model will also account for characteristics of each local neighborhood, including the population count of the area and economic health of residents, as measured by median income and poverty level of households.

In addition to aiding law enforcement, accurate estimation of changes in crime at the local neighborhood level would also enable the study of the association between crime trends and changes in the built environment. We are particularly interested in how aspects of the built environment encourage *vibrancy*, a measure of positive human activity, and how vibrancy is associated with safety in local neighborhoods (Humphrey et al. (2017)).

The city of Philadelphia is a particularly interesting case study for estimating trends in crime as it is a large urban area that is currently undergoing substantial development and experiencing population growth for the first time in decades. In addition to our primary goal of estimation of changes in crime in Philadelphia neighborhoods, this application also provides an interesting spatiotemporal data context for comparing different Bayesian shrinkage approaches to spatial areal modeling.

We will take an areal approach to modeling crime since our primary goal is greater understanding of evolving crime dynamics at the local neighborhood level within the city of Philadelphia. Our areal units will be U.S. Census block groups which consist of 10–20 city blocks and which are naturally interpretable as “neighborhoods.” U.S. Census block groups are also the highest resolution for which economic data is available as covariate information.

Compared to previous areal approaches (e.g., Aldor-Noiman et al. (2016), Law, Quick and Chan (2014) and Li et al. (2014)), we are using smaller areal units, and we will focus not only on total crime but also on the trend in crime over time within each local neighborhood. We have a longer time period (10 years) of recorded crimes for estimating time trends than did Law, Quick and Chan (2014) that worked with property crimes over a two year period.

Our methodological contribution is the development of a Bayesian spatial modeling framework to explore global vs. local smoothing for our parameter estimates while also allowing for data-driven discontinuities in our model between proximal areal units. Using a Bayesian approach allows us to induce this smoothing through shrinkage priors for our parameters and also enables us to estimate borders between neighborhoods that have a high probability of being barriers.

In Section 2 we provide details for the neighborhood structure of Philadelphia and describe the detailed crime data that we will use to estimate changes in crime over the past decade. We also outline the demographic, economic and land use measures we will use as neighborhood-level predictors of violent crime in our spatial models. The code for acquiring and cleaning the data that were used in this analysis is available as a GitHub repository at <https://github.com/cecilia-balocchi/Urban-project>. In Section 3 we develop several Bayesian modeling approaches for global or local sharing of information between Philadelphia neighborhoods, as well as a model extension that allows for spatial discontinuities in our parameter estimates between proximal neighborhoods. We then compare these modeling options in terms of both in-sample and out-of-sample predictive accuracy in Section 4. We visualize and discuss the results of our spatial modeling of crime trends for Philadelphia in Section 5 and then conclude with a brief discussion in Section 6.

2. Population, economic and crime data in Philadelphia. The population and economic data are provided by the US Census Bureau, whereas crime data is provided by the Philadelphia Police Department. Our definition of local neighborhoods in Philadelphia will be based upon the “block group” geographical units defined by the U.S. Census Bureau. The city of Philadelphia is divided into 384 census tracts which are divided into 1336 block groups. Shapefiles from the U.S. Census Bureau give the boundaries and area of each census block group. Figure S1 in our Supplementary Material (Balocchi and Jensen (2019)) gives a map outlining the 1336 block groups in Philadelphia.

Our motivation for analyzing trends in crime at this resolution is two-fold: (a) US census block groups consist of 10–20 city blocks which generally matches our concept of a “neighborhood,” and (b) the block group level is the highest resolution of the economic data that we will use as predictors of crime. The average size of block groups in Philadelphia is 0.26 km^2 with an average population of 1142 residents.

Our population data was pulled from the census website³ by setting the geography as all blocks in Philadelphia and setting the data source as “Hispanic or Latino Origin By Race” (which is SF1 P5 in their database). The raw demographic data gives the population count in each block group from the 2010 census. Figure S1 in our Supplementary Material (Balocchi and Jensen (2019)) gives the population count for each block group in Philadelphia.

The same data also has the population count in each block group divided by ethnic categories.⁴ From these ethnicity counts we calculate a measure of the *segregation* in each block group as

$$\text{segregation}_i = \frac{1}{2} \sum_r |p_{i,r} - \bar{p}_r|,$$

where $p_{i,r}$ is the proportion of ethnicity r in block group i and \bar{p}_r is the proportion of ethnicity r across the entire city of Philadelphia. The fraction $\frac{1}{2}$ scales this segregation measure to be between 0 and 1.

In addition to population count and our segregation measure, we will also consider several measures of the economic health of each neighborhood. Our economic data comes from the American Community Survey from the same U.S. Census website as our population data, specifically tables B19301 for income and C17002 for poverty, both from 2013. This data is only available at the resolution of census block groups. For each block group (neighborhood) in Philadelphia, we have income per capita as one predictor of crime.

We also have information about the proportion of households in various states of poverty. Specifically, we have the fraction of the population in seven different brackets of income-to-poverty-line ratios: $[0, 0.5)$, $[0.5, 1)$, $[1, 1.25)$, $[1.25, 1.5)$, $[1.5, 1.85)$, $[1.85, 2)$, $[2, \infty)$. For example, the $[0.5, 1)$ bracket represents families that have income between 50% of the poverty line and the poverty line itself. The poverty line is defined by the Census Bureau according to the size and composition of a household (e.g., a family with two children has a poverty line threshold of \$23,999).

³<https://factfinder.census.gov/>

⁴The ethnic categories are: White, Black, Asian, Native Americans, Native Pacific Islanders (including Hawaii), Other, Two or more races (non-Hispanic) and Hispanic/Latino. We combined Native Americans, Native Pacific Islanders, and Two or more races into the Other category which leads to five ethnicities in our analysis: 1. White, 2. Black, 3. Hispanic, 4. Asian, and 5. Other.

We use this poverty data to create a single measure of poverty for each block group (neighborhood) by calculating a weighted sum of the proportion of households in each of the seven poverty brackets:

$$\text{poverty}_i = \sum_{j=1}^7 w_j q_{i,j},$$

where $q_{i,1}$ is the proportion of households in block group i that are in the lowest bracket $[0, 0.5)$ and $q_{i,7}$ is the proportion of households in block group i in the highest bracket $[2, \infty)$. We use linearly decreasing weights, $\mathbf{w} = [1, 5/6, 4/6, 3/6, 2/6, 1/6, 0]$, to give higher weight to the brackets with higher poverty. Our poverty measure varies from 0 to 1 with larger values implying higher poverty.

In addition to the demographic and economic predictors described above, we also derive measures of the *built environment* that may also be predictive of crime. Our data on the built environment comes from the zoning designation of each lot in Philadelphia. Zoning data from the city of Philadelphia provides the area and registered land use designation (e.g., commercial, residential, industrial, vacant, transportation, park, civic) of all 560,000 lots in Philadelphia.

We create several land use metrics from these zoning designations that could be predictive of crime. First, we calculate the fraction of area in each block group i that is designated as “Vacant,”

$$\text{vacancy}_i = \frac{\text{Area}_i(\text{Vacant})}{\text{Area}_i}.$$

Second, we calculate the ratio of the area in each block group i that is commercial versus residential,

$$\text{comresprop}_i = \frac{\text{Area}_i(\text{Commercial})}{\text{Area}_i(\text{Commercial}) + \text{Area}_i(\text{Residential})}.$$

To summarize, we have created six neighborhood characteristics that we will use as predictors of crime: population count, segregation, median household income, poverty, vacant proportion and commercial vs. residential proportion. Some block groups in Philadelphia have missing values for the economic predictors due to a very small or zero population count. We exclude these block groups (a total of eight) from our analysis. We additionally exclude one block group containing the detention centers in Philadelphia.

Our crime data comes from the Philadelphia Police Department and includes all crimes reported by the police in the city of Philadelphia from January 1, 2006 to December 31, 2015. For each reported crime, we have the type of crime, the date and time of the crime and the location of the crime in terms of the GPS latitude and longitude (WGS84 decimal degrees). Each crime in our dataset is categorized into one of several types: homicide, sex crime, armed robbery, assault, burglary, theft, motor vehicle theft, etc.

We make a distinction between *violent* and *nonviolent* (property) crimes in our analysis. As defined by the Uniform Crime Reporting program of the [FBI](#), *violent* crimes include homicides, rapes, robberies and aggravated assaults whereas *nonviolent* crimes include burglaries, thefts and motor vehicle thefts.

Our own crime categorization differs from the FBI in two ways. We combine “rapes” and “sex assaults” (which changed in definition in 2013) into a broader “sex crimes” category and consider all “sex crimes” as violent crimes. The FBI also makes a distinction between “aggravated assaults” and “other assaults,” with the latter being where an injury does not occur but the threat of injury is present. In contrast, we combine both “aggravated assaults” and “other assaults” into a broader “assaults” category and consider all “assaults” as violent crimes.

For this paper, we focus entirely on the modeling of violent crimes as they have the most direct impact on human safety and the perception of safety. However, nonviolent crimes are also important to track for law enforcement and are a focus of ongoing research. In the subsequent analyses in this paper, we will use “crime” to mean only violent crimes.

In Figure 1 we give the counts of each type of violent crime within each year in 2006–2015, aggregated over the entire city. We see generally decreasing trends within the assault and robbery categories which are the most numerous types of crimes. Sex crimes and homicides are also somewhat decreasing over this time span though it is harder to see this trend given the low counts for either type of crime.

Clearly, the impression given from Figure 1 is that violent crimes are generally decreasing in the city of Philadelphia over the time period from 2006 to 2015. However, are there specific neighborhoods that show substantially larger decreases or even some neighborhoods that show increases in violent crimes in this period?

As discussed in Section 1, we will model the spatial distribution of crime with an areal approach where our areal units are U.S. Census block groups, which we define as the local neighborhoods of Philadelphia. Violent crimes are aggregated within each U.S. Census block group based on the GPS coordinates of each reported crime.

One issue with this approach is that some crimes occurring near to a boundary between U.S. Census block groups could be aggregated into the incorrect areal unit due to measurement error or ambiguity in their recorded point locations. This possibility is one of several motivations for our hierarchical Bayesian modeling approach that shares information between adjacent block groups when estimating crime totals and trends in crime over time across the city of Philadelphia.

In Figure S2 of our Supplementary Material ([Balocchi and Jensen \(2019\)](#)), we give the count of violent crimes per year in each block group averaged over the years 2006–2015. One can see substantial heterogeneity across block groups in the average counts of violent crimes per year. There are several outlying values—particular block groups that have much higher average violent crime counts.

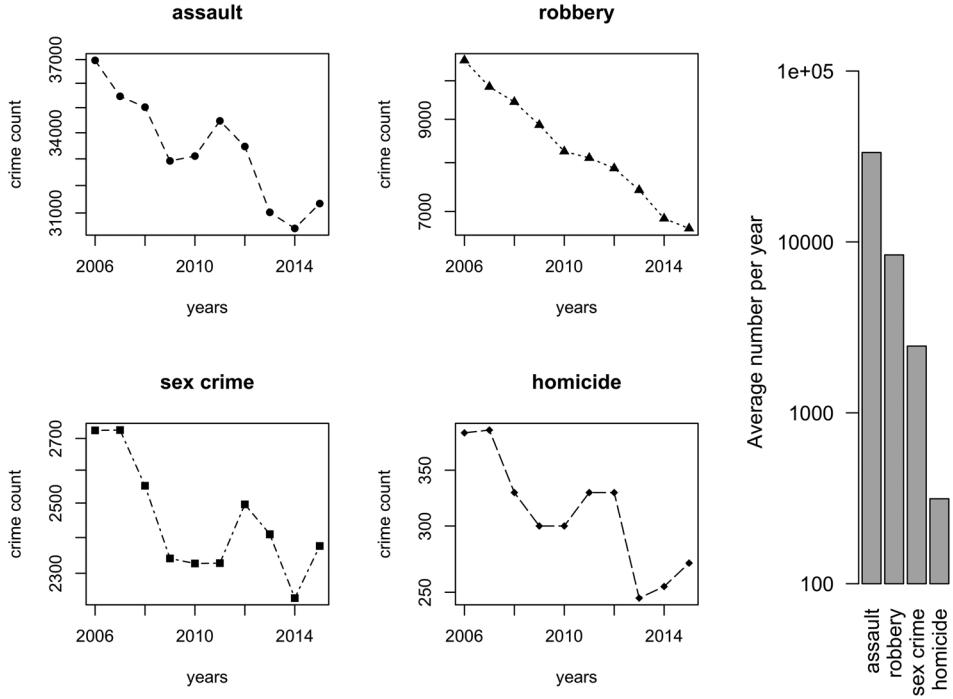


FIG. 1. *Counts of the different types of violent crimes in each year aggregated over the entire city of Philadelphia.*

These outlying neighborhoods motivate us to examine violent crime totals on the log scale. In Figure S2 of our Supplementary Material (Balocchi and Jensen (2019)), we also give the average of the logarithm of the count of violent crimes per year in each block group, averaged over the years 2006–2015. We can see more details of the spatial distribution of violent crime on the log scale. Modeling crime on the log scale has the additional benefit that changes in log crime can be interpreted as percentage changes in crime.

We also see in Figure S2 evidence of spatial correlation in violent crime totals between proximal block groups throughout the city. This is not surprising since the factors that lead to crime likely vary throughout the city in a (mostly) spatially continuous fashion. It is this spatial correlation that will be the focus of our modeling work in Section 3.

To get an idea of the strength of this spatial correlation, one of the standard statistics used for areal data is Moran's I (Banerjee, Carlin and Gelfand (2014), Moran (1950)), which is defined as

$$I = \frac{n}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij}(X_i - \bar{X})(X_j - \bar{X})}{\sum_i (X_i - \bar{X})^2},$$

where $W = (w_{ij})$ is a matrix of weights that capture the spatial proximity of the areal regions. We set w_{ij} to be 1 if block groups i and j share a border and 0 otherwise. We use the *queen* contiguity method so two block groups share a border if they share at least a point on their boundaries.

Moran's I can be used for testing for spatial autocorrelation: under the null hypothesis of no spatial association, we can compute exactly the mean (equal to $-\frac{1}{n-1}$) and standard error of Moran's I . Calculating I on the total number of violent crimes from 2006 to 2015 in our data gives an observed value of 0.335, compared to a null mean of 0.0007 and standard error of 0.0127, which suggests a highly significant amount of spatial autocorrelation in violent crime totals.

In the next section we develop several different Bayesian strategies for modeling violent crime over time and spatially between the areal neighborhoods of Philadelphia. We will fit our models on the violent crime data from 2006 to 2014, leaving data from 2015 for model comparison and evaluation.

3. Modeling areal crime data over space and time. As described in Section 2, the areal units of our analysis are the 1336 U.S. census block groups of Philadelphia (shown in Figure S1 in our Supplementary Material (Balocchi and Jensen (2019))).

For the remainder of this paper, we will use the terms “block group” and “neighborhood” interchangeably. The input data for our analysis is the number of violent crimes, c_{it} , reported in year t within neighborhood i . Our temporal range is $t = 1, \dots, T$ with $T = 10$, for the years 2006–2015 and our spatial range is $i = 1, \dots, n$ with $n = 1336$, for all the block groups in Philadelphia.

As seen in the violent crime totals (averaged over time) in Figure S2 of our Supplementary Material (Balocchi and Jensen (2019)), there are some substantial outlying neighborhoods with high violent crime totals relative to most of the city. These outliers (and general skewness in violent crime totals) motivates us to model violent crime totals on the logarithmic scale. This strategy has the additional benefit that linear changes over time in the logarithm of violent crime totals can be interpreted as percentage changes in raw violent crime totals.

However, because there is a small number of neighborhoods with zero crimes in some years, we need to consider a transformation that is defined at zero. Accordingly, we use the inverse hyperbolic sine transformation (Burbidge, Magee and Robb (1988)) that is centered to give values approximately equal to the logarithmic transformation. Specifically, we calculate our transformed violent crime totals as

$$(3.1) \quad y_{it} = \log(c_{it} + \sqrt{c_{it}^2 + 1}) - \log(2),$$

where c_{it} is the total number of violent crimes reported in year t within neighborhood i . A more common solution would be to add a small nonzero value to the

counts, for example, $\log(c_{it} + 1)$. We prefer the inverse hyperbolic sine transformation as it is numerically equivalent to the log transformation for large counts but is a better approximation than the $\log(c_{it} + 1)$ transformation for small counts.

An alternative modeling strategy for count data does not apply a transformation but assumes a Poisson distribution for the counts (Anderson and Ryan (2017), Law, Quick and Chan (2014), Li et al. (2014)). The Poisson model would not work since our data is over-dispersed, and the more flexible negative-binomial distribution does not model mean and variance as intuitively as a normal model. In addition, the normal model is conjugate for the prior distributions we will be considering which eases posterior estimation.

3.1. Accounting for neighborhood level covariates. We use a standard linear regression approach to account for the neighborhood-level economic, demographic and land use predictors of crime. Our transformed violent crime totals y_{it} are modeled as

$$(3.2) \quad y_{it} = \alpha + \mathbf{z}_i^T \boldsymbol{\gamma} + e_{it},$$

where \mathbf{z}_i is the vector of predictor variables for neighborhood i and $\boldsymbol{\gamma}$ is the vector of coefficients for those predictor variables, so $\mathbf{z}_i^T \boldsymbol{\gamma} = \sum_{d=1}^6 \gamma_d z_{id}$.

As outlined in Section 2, we have $d = 6$ predictor variables of crime for each neighborhood: population count, segregation, median household income, poverty, vacant proportion and commercial vs. residential proportion. We used square root transformations of vacant proportion, commercial vs. residential proportion and poverty and a logarithmic transformation of income to give a more linear relationship with the outcome variable.

Although yearly demographic and economic data is available after 2013, we avoid extrapolating values of the predictors to earlier years by modeling each predictor variable as *static* over the 10 year period spanned by our crime data. We examine the estimated partial effects $\boldsymbol{\gamma}$ of these economic, demographic and land use predictors in Section 5.1.

Although there is interest in the partial effects of our crime predictors, our primary interest lies in the temporal trends captured by e_{it} and the spatial correlation in these trends. With these time trends, we will be able to answer questions such as, “What areas of the city are increasing or decreasing most quickly in terms of safety?”

3.2. Time trends with no spatial correlation. We can add a global linear trend over time into our model,

$$(3.3) \quad y_{it} = \alpha + \mathbf{z}_i^T \boldsymbol{\gamma} + \beta \cdot t + \epsilon_{it} \quad \text{where } \epsilon_{it} \sim N(0, \sigma^2),$$

where the scalar coefficient β can be interpreted as the global percentage change in violent crime over time across the entire City of Philadelphia and t takes on integer values from 1 to 10 to represent the years 2006–2015.

However, this model with only a global α and β does not allow for heterogeneity between different neighborhoods in the overall level of violent crime or trend in violent crime over time. We can account for this heterogeneity through neighborhood-specific intercepts α_i and slopes β_i which give us the model

$$(3.4) \quad y_{it} = \alpha_i + \mathbf{z}_i^T \gamma + \beta_i \cdot t + \epsilon_{it} \quad \text{where } \epsilon_{it} \sim N(0, \sigma^2).$$

However, Model (3.4) is overparameterized. In fact, the effect of our static covariates is completely explained by the neighborhood-specific intercepts, α_i , so the same fit can be achieved by removing the covariates,

$$(3.5) \quad y_{it} = \alpha_i + \beta_i \cdot t + \epsilon_{it} \quad \text{where } \epsilon_{it} \sim N(0, \sigma^2).$$

Nonetheless, we can still estimate the partial effects of the covariates with an equivalent two-stage approach where we first fit $y_{it} = \alpha + \mathbf{z}_i^T \gamma + e_{it}$ and then fit the estimated residuals with the neighborhood-specific coefficient model, $\hat{\epsilon}_{it} = \alpha_i + \beta_i t + \epsilon_{it}$.

These neighborhood-specific model coefficients allow us to identify regions of Philadelphia with different levels of crime as well as different trends in crime over the past decade. This richer model is also motivated by fit to the data; a regression model with neighborhood-specific coefficients explains significantly more variation according to an F-test.

That said, we do not expect that every single neighborhood in Philadelphia would have unique coefficients, so we still risk overparametrization with this model. We address this overparameterization by imposing shared prior distributions for the neighborhood-specific coefficients from our time trend model (3.5),

$$(3.6) \quad \boldsymbol{\alpha} \sim N(\boldsymbol{\alpha}_0 \cdot \mathbf{1}, \tau_\alpha^2 \cdot \mathbf{I}),$$

$$(3.7) \quad \boldsymbol{\beta} \sim N(\boldsymbol{\beta}_0 \cdot \mathbf{1}, \tau_\beta^2 \cdot \mathbf{I}),$$

$$(3.8) \quad \boldsymbol{\gamma} \sim N(\mathbf{0}, \tau_\gamma^2 \cdot \mathbf{I}),$$

where we denote our collection of neighborhood-specific coefficients with $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_n)$ and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_n)$. $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_d)$ collects the coefficients (partial effects) of the predictor variables which are shared by all neighborhoods.

We complete this model formulation by placing flat priors on the global means α_0 and β_0 , $p(\alpha_0, \beta_0) \propto 1$, and inverse gamma priors on the variance parameters

$$\sigma^2 \sim \text{Inv-Gamma}(a_\sigma, b_\sigma),$$

$$\tau_\alpha^2 \sim \text{Inv-Gamma}(a_\alpha, b_\alpha),$$

$$\tau_\beta^2 \sim \text{Inv-Gamma}(a_\beta, b_\beta),$$

$$\tau_\gamma^2 \sim \text{Inv-Gamma}(a_\gamma, b_\gamma).$$

The variance hyperparameters are tuned in an empirical Bayes fashion so that the prior mean of the variance parameters is equal to the variance estimated from the

model with no shrinkage, and the prior variance is small. Using noninformative priors for these variance parameters produced nearly identical results. See Section 4 of our Supplementary Material ([Balocchi and Jensen \(2019\)](#)) for details.

This Bayesian hierarchical model shares information between neighborhoods by shrinking the neighborhood specific coefficients α_i and β_i toward global parameters (α_0, β_0) for the entire city. For this reason, we refer to this approach as the *global shrinkage* model.

However, this global shrinkage model does not account for the spatial proximity between neighborhoods when sharing information. We expect close neighborhoods to behave similarly while we want distant neighborhoods to be informative but not as directly influential as adjacent ones. In other words, we may prefer a model that imposes *local shrinkage* rather than global shrinkage.

A model with local sharing of information would also be better able to address the substantial spatial correlation that we see in our application. Testing with Moran's I shows that the residuals from the global shrinkage model are significantly spatially correlated. In the next subsection we will explore conditional autoregressive models for local sharing of information.

3.3. Time trends with a spatial conditional autoregressive model. A popular way of incorporating spatial information is through a prior distribution that is specified according to a Conditional Autoregressive (CAR) model, which was introduced in its most general formulation by [Besag \(1974\)](#). The CAR model is a Gaussian Markov random field which induces spatial dependence through an adjacency matrix for the areal units, which in our case are neighborhoods in Philadelphia.

Several variations of this CAR framework are reviewed and compared in [Lee \(2011\)](#). In this paper, we will use the proper CAR formulation introduced by [Leroux, Lei and Breslow \(2000\)](#).

Let $\boldsymbol{\theta}$ denote a vector of elements that are potentially spatially correlated, such as our neighborhood-specific intercepts $\boldsymbol{\alpha}$ or slopes $\boldsymbol{\beta}$. [Leroux, Lei and Breslow \(2000\)](#) defines the distribution of each θ_i , given the other θ_{-i} , as a normal distribution centered at a weighted average of a global mean and the θ_j 's from bordering neighborhoods,

$$(3.9) \quad \theta_i | \boldsymbol{\theta}_{-i}, \theta_0, \tau^2 \sim N\left(\frac{\rho \sum_j w_{ij} \theta_j + (1 - \rho)\theta_0}{\rho \sum_j w_{ij} + (1 - \rho)}, \frac{\tau^2}{\rho \sum_j w_{ij} + (1 - \rho)}\right),$$

where w_{ij} are adjacency weights that are equal to one if the neighborhoods i and j share a border and equal to 0 otherwise.

We collect these adjacency weights w_{ij} into an adjacency matrix \mathbf{W} that we assume (for now) to be known since we can easily use the shapefiles from the US Census Bureau to determine which of the 1336 neighborhoods (census block groups) share a border.

For now, we consider these adjacency weights w_{ij} to be fixed. However, in Section 3.4 we will extend our model to allow those weights to vary since some borders may represent *barriers* between neighborhoods (e.g., highways or rivers), in which case we would not want to share information across that particular border.

The parameter $\rho \in [0, 1]$ represents the strength of the spatial correlation between the components of θ , where larger values of ρ correspond to a stronger influence of bordering neighborhoods. In the special case of $\rho = 0$, the CAR prior (3.9) reduces to the *global shrinkage* prior (3.6)–(3.7).

It can be proved (Banerjee, Carlin and Gelfand (2014), Chapter 3) using Brook's lemma (Brook (1964)) that the joint distribution of θ is uniquely determined by the set of conditional distributions defined in 3.9:

$$(3.10) \quad \theta | \theta_0, \tau^2 \sim N(\theta_0 \cdot \mathbf{1}, \tau^2 \cdot [\rho(\mathbf{D}_W - \mathbf{W}) + (1 - \rho)\mathbf{I}]^{-1}),$$

where $\mathbf{1}$ is a vector of 1's and $\mathbf{D}_W - \mathbf{W}$ is the Laplacian matrix based on our neighborhood adjacency matrix \mathbf{W} . For values of ρ in $[0, 1)$, the joint distribution is proper, while for $\rho = 1$ the distribution is degenerate (Lee (2011)). By adding the constraint $\sum_i (\theta_i - \theta_0) = 0$, we can get a distribution for a n -dimensional vector, concentrated in a $(n - 1)$ -dimensional subspace; this is known as the intrinsic CAR by Besag, York and Mollié (1991).

We will employ this CAR model as prior distributions for the vectors of time trend coefficients α and β . We assume α and β are a priori independent. In vector form the CAR model (3.9) corresponds to the following prior distributions for α and β :

$$(3.11) \quad \alpha \sim N(\alpha_0 \cdot \mathbf{1}, \tau_\alpha^2 \cdot \Sigma),$$

$$(3.12) \quad \beta \sim N(\beta_0 \cdot \mathbf{1}, \tau_\beta^2 \cdot \Sigma),$$

where $\Sigma^{-1} = \rho(\mathbf{D}_W - \mathbf{W}) + (1 - \rho)\mathbf{I}$.

We use the same prior distributions for α_0 and β_0 and our variance parameters as in the *global shrinkage* model in the previous subsection. For the additional spatial parameter ρ , we choose a Beta(10, 10) prior distribution which has mean equal to 0.5 and a small variance in order to avoid the endpoints of the interval $[0, 1]$.

The posterior distributions for the spatial CAR model and the global shrinkage model (Section 3.2) can be implemented via a Gibbs sampler (Geman and Geman (1984)). Implementation details are given in our Supplementary Material (Balocchi and Jensen (2019)).

3.4. Allowing neighborhood border weights to vary. For most types of areal data, the weights \mathbf{W} that encode the spatial connection between the areal units are considered to be fixed and known. In our data context, the areal units are neighborhoods and the weights \mathbf{W} encode which neighborhoods share a border and hence induce shrinkage on each other in our spatial CAR models outlined in Section 3.3.

However, within any large city, some borders between neighborhoods consist of natural or artificial *barriers* such as rivers, highways or train tracks. These barriers could reduce the similarity in crime trends between neighborhoods, and so we would not want to shrink estimates across those barriers. The implication of these barriers for the spatial CAR models in Section 3.3 are that some weights $w_{ij} = 1$ should really be $w_{ij} = 0$ since those neighborhoods share a border that is actually a barrier.

Attempting to set which borders should actually be barriers manually would be tedious for a large city and also require extensive domain knowledge and subjective decision making. We instead prefer to infer these barriers from the data by allowing a subset of weights w_{ij} to be random variables in our model.

Specifically, we consider the set of indices of pairs of neighborhoods which share a border according to the geography of Philadelphia. The matrix \mathbf{W} is symmetric, so the random variables w_{ij} and w_{ji} are considered to be the same object. We model the w_{ij} for neighborhood pairs that share a border as Bernoulli random variables with an prior probability ϕ of $w_{ij} = 1$. Any weights $w_{ij} = 0$ according to the geography of Philadelphia will remain fixed at $w_{ij} = 0$ since we do not want to form connections between nonproximal neighborhoods.

We expect a priori that the probability ϕ will be close to one, since relatively few borders between neighborhoods actually should be barriers. For this reason, we choose the prior for ϕ to be a Beta(9, 1) distribution which has mean close to one and small variance.

Moreover, we expect that the spatial distribution of the neighborhood-specific crime levels (α_i) may be different from the neighborhood-specific trends in crime over time (β_i), so we allow for different barriers when we model the distribution of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. In particular, we consider two random matrices, \mathbf{W}^α and \mathbf{W}^β , where a subset of the elements of these matrices are random as described above: $w_{ij}^\alpha | \phi^\alpha \sim \text{Bernoulli}(\phi^\alpha)$ and $w_{ij}^\beta | \phi^\beta \sim \text{Bernoulli}(\phi^\beta)$ for neighborhood pairs (i, j) that share a border.

These two weight matrices then determine the local shrinkage of our spatial CAR model from the previous subsection:

$$(3.13) \quad \boldsymbol{\alpha} | \mathbf{W}^\alpha \sim N(\boldsymbol{\alpha}_0 \cdot \mathbf{1}, \tau_\alpha^2 \cdot \boldsymbol{\Sigma}_\alpha),$$

$$(3.14) \quad \boldsymbol{\beta} | \mathbf{W}^\beta \sim N(\boldsymbol{\beta}_0 \cdot \mathbf{1}, \tau_\beta^2 \cdot \boldsymbol{\Sigma}_\beta),$$

where $\boldsymbol{\Sigma}_\alpha^{-1} = \rho \cdot (\mathbf{D}_{W^\alpha} - \mathbf{W}^\alpha) + (1 - \rho)\mathbf{I}$ and $\boldsymbol{\Sigma}_\beta^{-1} = \rho \cdot (\mathbf{D}_{W^\beta} - \mathbf{W}^\beta) + (1 - \rho)\mathbf{I}$.

Allowing variable border weights can lead to overparametrization since we are adding as many parameters as the number of borders, which makes the shrinkage imposed by prior parameters ϕ^α and ϕ^β important. A more sophisticated approach, which is the focus of ongoing work, would be to partition our areal units into clusters with barriers represented as cluster boundaries.

To implement this extended model with some variable border weights, a step is added to our Gibbs sampler that samples each border weight conditional on the

current values of the other model parameters. Details are given in our Supplementary Material (Balocchi and Jensen (2019)).

The idea of detecting discontinuities at boundaries is often referred to as *wombling*, after the seminal work of Womble (1951), and has been very popular in the disease mapping literature. However, most papers have approached detection of boundaries as a selection problem that is performed after inference (see, e.g., Boots (2001), Li, Banerjee and McBean (2011), Li et al. (2012), Lu and Carlin (2005), Lee and Mitchell (2013)).

In contrast, we incorporate the possibility of discontinuities at boundaries directly into our model through variable \mathbf{W}^α and \mathbf{W}^β which allows us to incorporate potential barriers into our estimation of neighborhood-specific parameters α and β . Lee and Mitchell (2012) and Lu et al. (2007) take a similar approach in the context of disease mapping but with a more elaborate model for $P(w_{ij} = 1)$ that is a function of dissimilarity between covariate values in units i and j .

In Section 4 the different models presented in this section are compared in terms of their accuracy of their in-sample and out-of-sample predictive accuracy. We then visualize the estimated trends in crime over time in Philadelphia and discuss several insights from our results in Section 5.

4. Comparison of predictive accuracy. In the previous section we outlined a no shrinkage model (Section 3.2) and several hierarchical Bayesian models for estimating the neighborhood-level trend in crime over time, including a global shrinkage model (Section 3.2), a spatial CAR models for local shrinkage (Section 3.3) and finally an extension of the spatial CAR model to allow a subset of border weights to vary (Section 3.4).

We now compare each of these model alternatives based on several measures of the accuracy of their predictions on both in-sample and out-of-sample hold-out data. Recall that we have 10 years of crime data for the city of Philadelphia, from the beginning of 2006 to the end of 2015. We estimate each model using the crime data for the first nine years (2006–2014).

We assess the *in-sample* accuracy of each model by computing the mean squared error of the predictions of violent crime totals for 2014, which is a year that was included in model estimation,

$$(4.1) \quad \text{MSE}_{\text{in}} = \frac{1}{1336} \sum_{i=1}^{1336} (y_{i,2014} - \hat{y}_{i,2014})^2.$$

We assess the *out-of-sample* accuracy of each model by computing the mean squared error of the predictions of violent crime totals for 2015, which is a year that was *not* included in model estimation,

$$(4.2) \quad \text{MSE}_{\text{out}} = \frac{1}{1336} \sum_{i=1}^{1336} (y_{i,2015} - \hat{y}_{i,2015})^2.$$

TABLE 1

Comparison of predictive accuracy between the different models outlined in Section 3. The mean squared error for both in-sample and out-of-sample predictions are provided, as well as the percentage change in MSE_{out} relative to model (3.5) without shrinkage. We also provide the Moran's I measure of spatial correlation calculated on the estimated time trends β_i from each model

Model	MSE_{in}	MSE_{out}	% change	MSE_{cv}	Moran's I
Global α, β	0.3558	0.3694	+182.4	0.3043	–
Separate α_i, β_i Models					
No Shrinkage (3.5)	0.0567	0.1308	–	0.1001	0.17
Global Shrinkage (3.6)–(3.7)	0.0698	0.1080	–17.4	0.0928	0.17
Spatial CAR (3.11)–(3.12)	0.0703	0.1052	–19.5	0.0922	0.61
Variable Borders (3.13)–(3.14)	0.0706	0.1069	–18.2	0.0927	0.49

To ensure our evaluation is not overly dependent on any idiosyncratic aspects of the 2015 data, we also calculate the *cross-validated, out-of-sample* accuracy of each model by calculating the mean square error MSE_{out}^t when using year t as the hold out data in the same way that 2015 is used as the hold out data in (4.2), that is,

$$(4.3) \quad MSE_{cv} = \frac{1}{10} \sum_{t=1}^{10} MSE_{out}^t \quad \text{where } MSE_{out}^t = \frac{1}{1336} \sum_{i=1}^{1336} (y_{i,t} - \hat{y}_{i,t})^2.$$

In Table 1 we compare the predictive accuracy of four different models with neighborhood-specific coefficients outlined in Section 3: 1. the time trend model (3.5) without shrinkage between neighborhoods, 2. the global shrinkage model with priors (3.6) and (3.7), 3. the local shrinkage model with spatial CAR priors (3.11) and (3.12) and 4. the local shrinkage spatial CAR model with variable borders (3.13) and (3.14). For additional reference we also provide the mean square error for fitting a single trend (“Global α, β ”) across the entire city.

We see in Table 1 that the model with a global trend over time (“Global α, β ”) for the entire city has very poor predictive accuracy compared to the models that allow neighborhood-specific time trends (“Separate α_i, β_i ”).

Among the neighborhood-specific time trend models, the global shrinkage model has substantially lower out-of-sample mean square errors than the baseline time trend model without any shrinkage between neighborhoods. The best in-sample mean squared error was achieved by the model without shrinkage, as we expect from the least square method, though at a cost of having the worst out-of-sample accuracy.

The model with local shrinkage via the spatial CAR prior further reduces the out-of-sample mean square errors compared to the global shrinkage model. The model that allows variable borders does not further improve the out-of-sample

mean squared errors, though we explore in Section 5.3 that it helps with the interpretation.

Table 1 also provides Moran's I measure of spatial autocorrelation, calculated on the posterior mean of the neighborhood-specific time trends (β_i 's). We see that the spatial CAR model induces a larger spatial correlation in the β_i 's than the models with global shrinkage or without shrinkage. The local shrinkage model has a Moran's I value of 0.61 ($s.e. = 0.016$) which suggests there is substantial spatial autocorrelation in the change in crime within Philadelphia.

In summary, allowing for local shrinkage of the neighborhood-specific crime trend coefficients via the spatial CAR priors (3.11) and (3.12) leads to the best out-of-sample predictive accuracy. In Section 5 we visualize the parameters of this model and discuss the implications of these results for crime in Philadelphia.

Although the variable border model extension does not improve out-of-sample predictive accuracy, we will also see in Section 5 that visualizing the borders that have been turned into barriers by this model provide insight into discontinuities in crime trends in the city of Philadelphia.

5. Interpretation of model parameters. In Section 5.1 we examine the estimated partial effects for the static predictor variables created from the data outlined in Section 2. We then visualize and compare the estimated neighborhood-specific levels (α_i 's) and time trends (β_i 's) on crime from our different models in Section 5.2. In Section 5.3 we examine the results from our model extension outlined in Section 3.4 that allows a subset of neighborhood borders in Philadelphia to be estimated as barriers. Finally, in Section 5.4 we discuss the neighborhoods with the most extreme levels and changes in crime over time over the past 10 years in Philadelphia.

5.1. Partial effects of static predictors. Figure 2 gives the estimated partial effects γ_d for each static predictor variable d from the four models outlined in Section 3. We provide additional numerical details in Table S1 of our Supplementary Material (Balocchi and Jensen (2019)).

We see that among the six predictor variables created in Section 2, only the segregation measure is not a significant predictor of crime. All predictor variables are on the same scale, and so we can directly compare the values of their partial effects.

We see that the strongest predictors of crime are total population and the commercial versus residential proportion, with more populated and more commercial neighborhoods being associated with higher crime. Income and poverty are also significantly predictive of violent crimes, but we must be more cautious about interpreting these partial effects given the high collinearity between income and poverty. Each of these observations on the partial effects γ is relatively consistent across the four models outlined in Section 3.

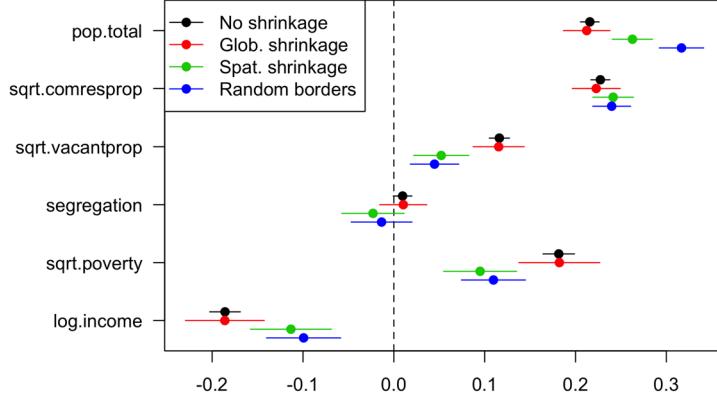


FIG. 2. Estimated partial effects γ_d from four different models indicated in the legend. For the no shrinkage model we plot the maximum likelihood estimate and 95% confidence interval. For three Bayesian shrinkage models we plot the posterior mean and 95% posterior interval.

5.2. Visualizing neighborhood-specific coefficients. Our primary interest in terms of interpretation are the estimated neighborhood-specific coefficients, α_i 's and β_i 's, that represent the level of violent crimes and change in violent crimes over time in Philadelphia, respectively.

In Figures 3 and 4 we give maps where each block group in Philadelphia is colored by the estimated neighborhood-specific levels of crime $\hat{\alpha}_i$ and changes in crime over time $\hat{\beta}_i$, respectively, from the four models outlined in Section 3. We see substantial heterogeneity between neighborhoods in Philadelphia, both in terms of their estimated crime levels ($\hat{\alpha}_i$'s) and changes in crime over time ($\hat{\beta}_i$'s). Regardless of the model, most neighborhoods in the city show decreasing trends in crime over time (negative β 's) with a small subset of neighborhoods showing an increasing trend.

The shrinkage imposed by the global shrinkage model is more visually striking for the change in violent crime over time than the overall level of crime. The maps of the $\hat{\alpha}_i$'s from the no shrinkage and global shrinkage models are almost indistinguishable in Figure 3, whereas the map of the $\hat{\beta}_i$'s from the global shrinkage model has been shifted substantially compared to the no shrinkage map in Figure 4. This observation suggests that there is more substantial heterogeneity between neighborhoods in terms of their overall level of crime compared to their change in crime over time.

This heterogeneity in the mean level of crime is expected as it is influenced by many years of transformation in the city of Philadelphia that led to its current built and social environment. Differences in these overall spatial crime patterns can be addressed by urban planners whose effects are long-lasting (Johnson, Bowers et al. (2008)). In contrast, differences in the trend over time identify shorter-term patterns which can be addressed with interventions by local police departments.

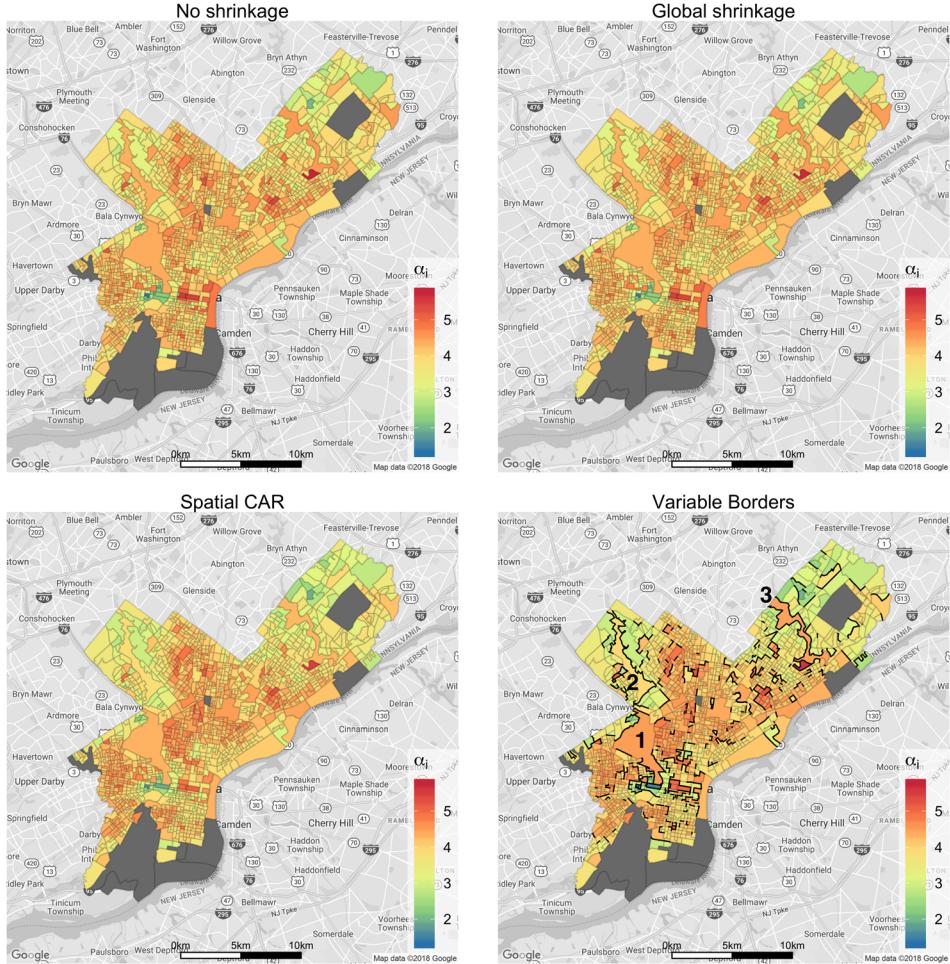


FIG. 3. Maps of Philadelphia colored by the estimated intercept from our four different models. Top-left: Maximum likelihood estimates of α_i from the no shrinkage model (3.5). Top-right: Posterior means of α_i from the global shrinkage model (3.6)–(3.7). Bottom-left: Posterior means of α_i from the spatial CAR model (3.11)–(3.12). Bottom-right: Posterior means of α_i from the spatial CAR model with variable borders (3.13)–(3.14). The black lines represent borders turned into barriers. These maps were created with the R package `ggmap` (Kahle and Wickham (2013)).

The overall level of crime also seems to have a greater inherent spatial correlation between proximal neighborhoods than the change in crime over time. The Moran's I values calculated from the estimated $\hat{\alpha}_i$'s are $I = 0.33$ for both the no shrinkage and global shrinkage models, compared to the value of $I = 0.17$ from the estimated $\hat{\beta}_i$'s for those same models in Table 1. This is clear also from the maps from the no shrinkage model (top left) in Figures 3 and 4; the estimated $\hat{\beta}_i$'s are more “spotty” and less smooth than the corresponding map of the $\hat{\alpha}_i$'s.

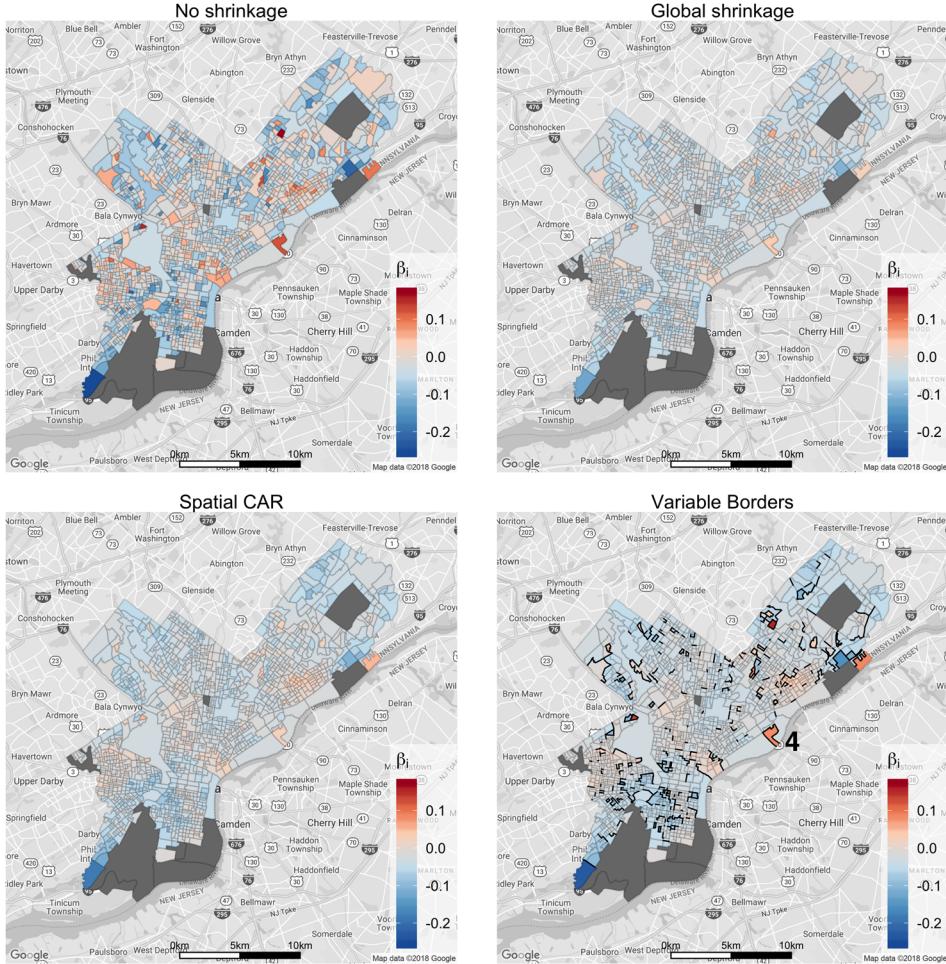


FIG. 4. Maps of Philadelphia colored by the estimated slope on time from our four different models. Top-left: Maximum likelihood estimates of β_i from the no shrinkage model (3.5). Top-right: Posterior means of β_i from the global shrinkage model (3.6)–(3.7). Bottom-left: Posterior means of β_i from the spatial CAR model (3.11)–(3.12). Bottom-right: Posterior means of β_i from the spatial CAR model with variable borders (3.13)–(3.14). The black lines represent borders turned into barriers. These maps were created with the R package `ggmap` (Kahle and Wickham (2013)).

However, once we build spatial correlation into our model via the spatial CAR prior (3.11)–(3.12), the resulting $\hat{\beta}_i$'s are more spatially correlated than the resulting $\hat{\alpha}_i$'s, as can be seen in the lower left of Figures 3 and Figures 4 as well as the corresponding Moran's $I = 0.53$ for the $\hat{\alpha}_i$'s vs. $I = 0.61$ for the $\hat{\beta}_i$'s. Note that all these reported Moran's I values have a standard error approximately equal to 0.016, and so they are all significantly different from the null hypothesis of no spatial autocorrelation.

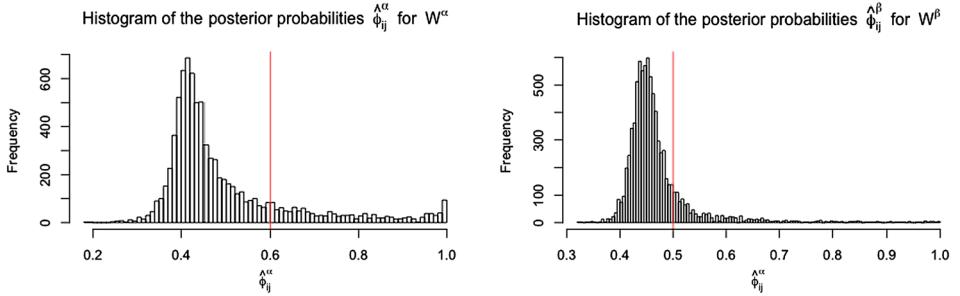


FIG. 5. *Histograms of the posterior probabilities of each border being turned into a barrier. Left: Probabilities for barriers for the α_i 's; the threshold to identify the borders turned into barriers is 0.6 (red line). Right: Probabilities for barriers for the β_i 's; the threshold to identify the borders turned into barriers is 0.5 (red line).*

Although the smoother maps from the spatial CAR model (lower left of Figures 3 and 4) ease interpretation by identifying larger regions of the city with similar crime dynamics, there is the potential to overshrink certain neighborhoods that should actually stand out from their neighbors. In any large city natural or artificial *barriers*, such as rivers, highways or rail lines, create discontinuities between neighborhoods which should not be smoothed over. In Section 5.3 we examine the results from our model extension that allows a subset of borders between neighborhoods to be turned into barriers.

5.3. Borders turned into barriers. In Section 3.4 we extended the spatial CAR model to allow a subset of the weights w_{ij} to vary which allows the *borders* ($w_{ij} = 1$) between some neighborhoods to be changed into *barriers* ($w_{ij} = 0$); the latter prevent shrinkage between two bordering neighborhoods. Our model has separate weight matrices \mathbf{W}^α and \mathbf{W}^β , so a particular border can be turned into a barrier either for the level of crime (α_i 's) or the change in crime over time (β_i 's) or both. Using this model, we estimate the posterior probability that we change a border into a barrier for each border between proximal neighborhoods in Philadelphia.

Figure 5 gives the distribution of the estimated posterior probability of a border being turned into barrier for each border encoded in the weight matrices \mathbf{W}^α and \mathbf{W}^β . These distributions seem to have two components: a main mode representing the behavior of the majority of the borders which has a low probability of being turned into a barrier, and a “tail” component which has a higher probability of being turned into a border.

It is clear that many more borders have a high probability of being a barrier for the level of crime (α_i 's) compared to the change in crime over time (β_i 's). In other words, our variable border model is detecting more discontinuities between bordering neighborhoods in the level of crime compared to the change in crime over time. In Section 5 of our Supplementary Material (Balocchi and Jensen (2019)),

we explore an alternative model that only allows variable borders for the mean level of crime.

In the lower right panels of Figures 3 and 4, we provide maps of Philadelphia where we have highlighted any borders between neighborhoods that have been inferred by our model to have a high probability of being barriers. These particular highlights are based on posterior probabilities larger than 60% for \mathbf{W}^α and larger than 50% for \mathbf{W}^β .

We see in the lower right panel of Figure 3 that barriers have been detected around several parks including Fairmount Park, Wissahickon Valley Park and Pennypack Creek Park (indicated by the black numbers 1, 2 and 3, respectively, in the lower right panel of Figure 3). In these cases our model has automatically detected several natural geographic structures within Philadelphia as locations which have discontinuities in the level of crime.

We also see that some estimated barriers have isolated particular neighborhoods from their proximal neighbors. For example, the neighborhood of Bridesburg (indicated by the black number 4 in the lower right panel of Figure 4) seems to have a much more positive trend on crime over time than its surrounding neighborhoods.

As barriers highlight the boundaries of regions that display differences in either in the level of crime or the trend in crime over time, these barriers can be used by police departments and city planners for delineating the possible limits of effectiveness for interventions or as potential targets for interventions themselves.

5.4. Neighborhoods with most extreme crime trends. To further understand which regions of Philadelphia have the most extreme levels of crime and trends in crime over time, we can examine the most extreme intercepts (α_i 's) and slopes (β_i 's) found by our fitted models. Specifically, we focus on the estimated α_i 's and β_i 's from the local shrinkage spatial CAR model (3.11)–(3.12) that had the best out-of-sample predictive performance in Table 1.

Figure 6 provides maps that highlight the most extreme (largest 50 and smallest 50) neighborhoods in terms of the estimated level of crime ($\hat{\alpha}_i$'s) and in terms of the estimated change in crime over time ($\hat{\beta}_i$'s).

We see that the region of University City in West Philadelphia (black number 1 in the top panel of Figure 6) is an interesting transitional area that contains both neighborhoods with the highest and lowest levels of crime in the city. We also see that the area of Frankford (black number 2 in the top panel of Figure 6) has neighborhoods with high levels of crime. This area is a major transportation hub for the Northeast region of Philadelphia.

The SW region of Philadelphia, specifically the Elmwood and Eastwick neighborhoods (black number 3 in the bottom panel of Figure 6) have seen some of the largest reductions in crime over the past decade in Philadelphia. We also see some regions of the city that are showing increases in crime over that same time period, such as the Wissinoming and Tacony neighborhoods, (black number 4 in

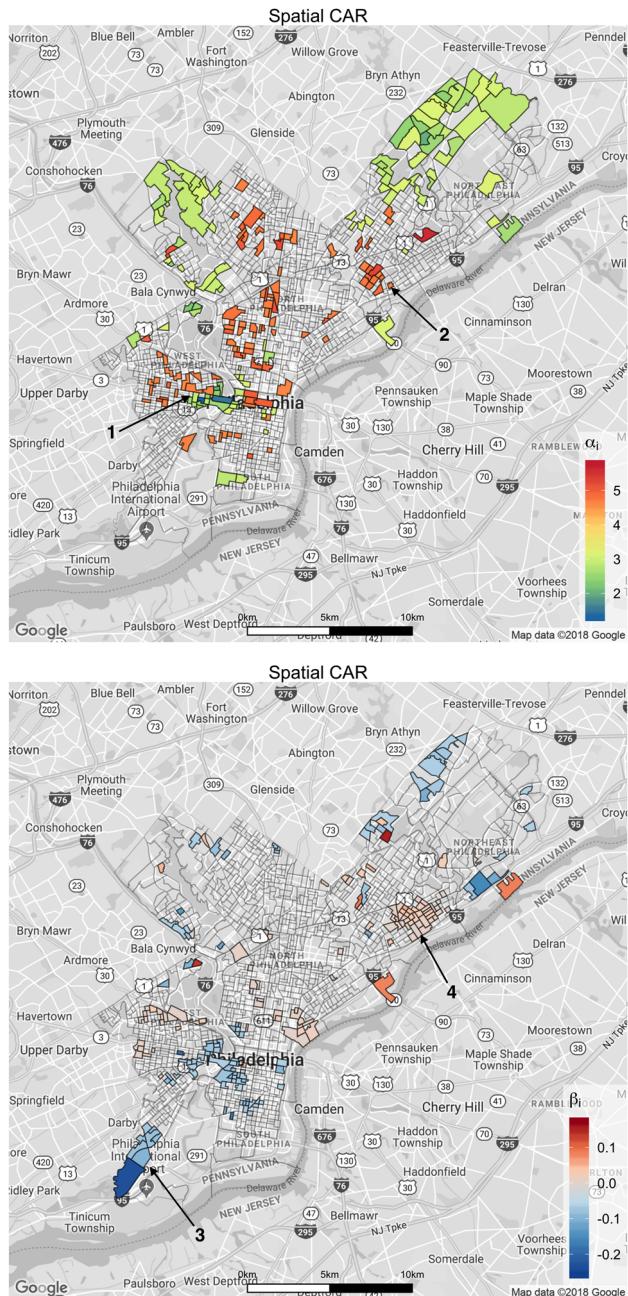


FIG. 6. Top: The 50 neighborhoods with the largest $\hat{\alpha}_i$'s (red) and 50 neighborhoods with the smallest $\hat{\alpha}_i$'s (green). Bottom: The 50 neighborhoods with the largest $\hat{\beta}_i$'s and 50 neighborhoods with the smallest $\hat{\beta}_i$'s. These maps were created with the R package `ggmap` (Kahle and Wickham (2013)).

the bottom panel of Figure 6) that are just to the northeast of the high crime neighborhoods of Frankford (black number 2 in the top panel of Figure 6).

In Section 5 of our Supplementary Material (Balocchi and Jensen (2019)), we provide additional visualizations of the neighborhood-specific parameters that are significantly different from the overall mean across the city as well as the widths of the credible intervals for these parameters.

6. Discussion. Reliable estimation of the change in crime over time at the local neighborhood level is a crucial step toward a better understanding of the determinants of public safety in large urban areas. With a focus on the city of Philadelphia, we have explored several Bayesian approaches to modeling crime trends within the areal units of neighborhoods while sharing information either globally or locally across the city.

Imposing local shrinkage between proximal neighborhoods via a spatial conditional autoregressive (CAR) prior gives the best out-of-sample predictions of violent crime compared to models that impose global shrinkage or no shrinkage at all between neighborhoods. We also explore allowing the weight matrix of our spatial CAR model to vary in order to detect neighborhood borders that represent spatial discontinuities in the level of crime or change in crime over time. In this way, we automatically detect several natural barriers in the geography of Philadelphia. Our model estimates also identify the regions of Philadelphia with the most extreme levels of violent crime as well as the largest increases and reductions in crime over the period of 2006–2015.

Acknowledgments. We thank Rachel Thurston and Theresa Smith for helpful contributions to our modeling efforts and interpretation of our results.

SUPPLEMENTARY MATERIAL

Supplementary material for “Spatial modeling of trends in crime over in Philadelphia” (DOI: 10.1214/19-AOAS1280SUPP; .pdf). We provide maps outlining the block group structure, population count and distribution of violent crimes in Philadelphia. We give implementation details for our fitted models. We compare results under an alternative choice of prior distributions and under an alternative model that only allows variable borders for the mean level of crime. We provide additional numerical details about our estimated partial effects. We also provide visualizations of the significance and widths of credible intervals for the neighborhood-specific parameters.

REFERENCES

- ALDOR-NOIMAN, S., BROWN, L. D., FOX, E. B. and STINE, R. A. (2016). Spatio-temporal low count processes with application to violent crime events. *Statist. Sinica* **26** 1587–1610.
MR3586230

- ANDERSON, C. and RYAN, L. M. (2017). A comparison of spatio-temporal disease mapping approaches including an application to ischaemic heart disease in New South Wales, Australia. *International Journal of Environmental Research and Public Health* **14** 146.
- ANSELIN, L. (1998). *Spatial Econometrics: Methods and Models*. Springer, Dordrecht.
- BALOCCHI, C. and JENSEN, S. T (2019). Supplement to “Spatial modeling of trends in crime over time in Philadelphia.” DOI:10.1214/19-AOAS1280SUPP.
- BANERJEE, S., CARLIN, B. P. and GELFAND, A. E. (2014). *Hierarchical Modeling and Analysis for Spatial Data*, 2nd ed. *Monographs on Statistics and Applied Probability* **135**. CRC Press, Boca Raton, FL. [MR3362184](#)
- BESAG, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *J. Roy. Statist. Soc. Ser. B* **36** 192–236. [MR0373208](#)
- BESAG, J., YORK, J. and MOLLIÉ, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Ann. Inst. Statist. Math.* **43** 1–59. [MR1105822](#)
- BOOTS, B. (2001). Using local statistics for boundary characterization. *GeoJournal* **53** 339–345.
- BRANTINGHAM, P. J. and BRANTINGHAM, P. L. (1984). *Patterns in Crime*. Macmillan, New York.
- BROOK, D. (1964). On the distinction between the conditional probability and the joint probability approaches in the specification of nearest-neighbour systems. *Biometrika* **51** 481–483. [MR0205315](#)
- BURBIDGE, J. B., MAGEE, L. and ROBB, A. L. (1988). Alternative transformations to handle extreme values of the dependent variable. *J. Amer. Statist. Assoc.* **83** 123–127. [MR0941005](#)
- CRESSIE, N. (1990). The origins of kriging. *Math. Geol.* **22** 239–252. [MR1047810](#)
- FBI. Uniform crime reporting program, definitions. Available at <https://ucr.fbi.gov/crime-in-the-u-s/2011/crime-in-the-u-s-2011/offense-definitions>. Accessed: 2016-09-15.
- FLAXMAN, S. R. (2014). A general approach to prediction and forecasting crime rates with Gaussian processes. Heinz College Technical Report, Carnegie Mellon Univ., Pittsburgh, PA.
- GEMAN, S. and GEMAN, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.* **6** 721–741.
- HERBERT, D. T. (1982). *The Geography of Urban Crime*. Longman, London.
- HUMPHREY, C., JENSEN, S. T., SMALL, D. S. and THURSTON, R. (2017). Analysis of Urban Vibrancy and Safety in Philadelphia. Working paper. Available at [arXiv:1702.07909](#).
- HVISTENDAHL, M. (2016). Crime forecasters. *Science* **353** 1484–1487.
- JOHNSON, S. D., LAB, S. P. and BOWERS, K. J. (2008). Stable and fluid hotspots of crime: Differentiation and identification. *Built Environment* **34** 32–45.
- KAHLE, D. and WICKHAM, H. (2013). ggmap: Spatial visualization with ggplot2. *The R Journal* **5**.
- LAW, J., QUICK, M. and CHAN, P. (2014). Bayesian spatio-temporal modeling for analysing local patterns of crime over time at the small-area level. *J. Quant. Criminol.* **30** 57–78.
- LEE, D. (2011). A comparison of conditional autoregressive models used in Bayesian disease mapping. *Spat Spatiotemporal Epidemiol* **2** 79–89.
- LEE, D. and MITCHELL, R. (2012). Boundary detection in disease mapping studies. *Biostatistics* **13** 415–426.
- LEE, D. and MITCHELL, R. (2013). Locally adaptive spatial smoothing using conditional autoregressive models. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **62** 593–608. [MR3083913](#)
- LEROUX, B. G., LEI, X. and BRESLOW, N. (2000). Estimation of disease rates in small areas: A new mixed model for spatial dependence. In *Statistical Models in Epidemiology, the Environment, and Clinical Trials* (Minneapolis, MN, 1997). *IMA Vol. Math. Appl.* **116** 179–191. Springer, New York. [MR1731684](#)
- LI, P., BANERJEE, S. and MCBEAN, A. M. (2011). Mining boundary effects in areally referenced spatial data using the Bayesian information criterion. *Geoinformatica* **15** 435–454.
- LI, P., BANERJEE, S., MCBEAN, A. M. and CARLIN, B. P. (2012). Bayesian areal wombling using false discovery rates. *Stat. Interface* **5** 149–158. [MR2928066](#)

- LI, G., HAINING, R., RICHARDSON, S. and BEST, N. (2014). Space-time variability in burglary risk: A Bayesian spatio-temporal modelling approach. *Spat. Stat.* **9** 180–191. [MR3326838](#)
- LU, H. and CARLIN, B. P. (2005). Bayesian areal wombling for geographical boundary analysis. *Geographical Analysis* **37** 265–285.
- LU, H., REILLY, C. S., BANERJEE, S. and CARLIN, B. P. (2007). Bayesian areal wombling via adjacency modeling. *Environ. Ecol. Stat.* **14** 433–452. [MR2405556](#)
- MOHLER, G. O., SHORT, M. B., BRANTINGHAM, P. J., SCHOENBERG, F. P. and TITA, G. E. (2011). Self-exciting point process modeling of crime. *J. Amer. Statist. Assoc.* **106** 100–108. [MR2816705](#)
- MØLLER, J. and WAAGEPETERSEN, R. P. (2007). Modern statistics for spatial point processes. *Scand. J. Stat.* **34** 643–684. [MR2392447](#)
- MORAN, P. A. P. (1950). Notes on continuous stochastic phenomena. *Biometrika* **37** 17–23. [MR0035933](#)
- PACE, R. K. and LESAGE, J. (2010). Spatial econometrics. In *Handbook of Spatial Statistics. Chapman & Hall/CRC Handb. Mod. Stat. Methods* 245–260. CRC Press, Boca Raton, FL. [MR2730939](#)
- STEIN, M. L. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*. Springer Series in Statistics. Springer, New York. [MR1697409](#)
- TADDY, M. A. (2010). Autoregressive mixture models for dynamic spatial Poisson processes: Application to tracking intensity of violent crime. *J. Amer. Statist. Assoc.* **105** 1403–1417. [MR2796559](#)
- WHITTLE, P. (1954). On stationary processes in the plane. *Biometrika* **41** 434–449. [MR0067450](#)
- WOMBLE, W. H. (1951). Differential systematics. *Science* **114** 315–322.

DEPARTMENT OF STATISTICS
THE WHARTON SCHOOL
UNIVERSITY OF PENNSYLVANIA
400 JON M. HUNTSMAN HALL
3730 WALNUT STREET
PHILADELPHIA, PENNSYLVANIA 19104
USA
E-MAIL: balocchi@wharton.upenn.edu
stjensen@wharton.upenn.edu