KDE in the Congo: Using Kernel Density Estimation to Predict Conflict in the Democratic Republic of Congo

Jessica Moore

June 3, 2016

1 INTRODUCTION

The Democratic Republic of the Congo (DRC) has a history of violent conflict dating back at least to its colonization by Belgium in the late 1800's (Rosen 2013). The colonial rule imposed by the Belgians was notoriously exploitative and left the country largely in ruins (Hochschild 1998). When the Congolese finally gained independence in the 1960's, the country had no government to speak of and, in spite of the DRC's plentiful natural resources, economic underdevelopment reigned (Rosen 2013). While progress has been made in the past half-century, the rampant poverty and lack of government oversight has resulted in high levels of violence. Conflict in the country has killed more than 3.5 million individuals since 1996 (Rosen 2013). In 1998, a rebellion against the already weak Congolese government began in the Kiva region of the DRC, which lies on the country's east, near its borders with Rwanda and Burundi ("MONUSCO Background"). The United Nations (UN) brokered a ceasefire between the government and the rebels in 1999 and has maintained a substantive presence in the country ever since ("MONUSCO Background"). Currently, the UN oversees a force of nearly 20,000 military personnel in the DRC; the force attempts to quell violent uprisings and mitigate potential sources of conflict ("MONUSCO Facts and Figures"). However, in 2014 the UN Security Council identified the need for an exit strategy, the goal of which would be to remove UN forces from the country and transition military and police duties to the Congolese government ("MONUSCO Mandate"). In order to implement such a plan, the UN will need to gradually remove its troops, concentrating its remaining personnel in the most conflict-prone regions. This, however, is no small task, since the UN must first identify (with relative specificity) the areas of the country in which additional conflict is particularly likely.

We suggest that kernel density estimation (KDE) might be an effective tool for doing exactly that. In order to test our theory, we built a variety of density estimates using a number of automatic bandwidth selection methods. We created density estimates both for conflict in general, as well as for specific types of conflict (e.g., riots/protests). We built these densities using the locations of conflicts that occurred in the DRC during 2014. We then tested our models by assessing how well they predicted the locations of 2015 conflicts in the country. With a few exceptions, our models performed quite well and so we feel comfortable concluding that KDE is an effective method for identifying the areas of the country in which violent conflict is likely to occur. As a result, we recommend that the UN use a similar method in order to identify the regions of the DRC that would see the most benefit from military presence. Based on our results, we also suspect that KDE would be an effective tool for predicting conflict in many developing countries (not just the DRC). Thus, it is our hope that this method will also be considered by those attempting to stop or mitigate violent outbreaks in a wide variety of locations.

This paper begins with an overview of kernel density estimation (Section 2). We discuss both the basic mathematics and the intuition behind the procedure. We also review (at a high level) a number of bandwidth selection methods, each of which we employ later in the paper. Next, we discuss previous literature regarding both spatial density estimation as well as the estimation of densities related to conflict (Section 3). Then, we motivate the use of kernel density estimation in this context (Section 4). In particular, we contrast KDE with the parametric methods that are conventionally applied in similar situations. After this, we provide a description of the data employed to create our models (Section 5). We discuss the data's contents and our adjustments to the original data set that we obtained. We also provide a high-level comparison between 2014 and 2015 records, because these years are our primary focus. Next, we review the models that we created based on that data (Section 6). We present visual depictions of the kernel density estimates (in the form of heat-maps), as well as numeric and graphical comparisons showing which models were most effective in predicting the locations of 2015 conflicts. Finally, we provide concluding remarks and suggestions, both for uses of our results and areas for further research (Section 7).

2 Kernel Density Estimation

Kernel density estimation (KDE) is a common nonparametric method for estimating the probability density function of a random variable, based on observed realizations of that variable (Cai 2013). The technique is particularly useful when the data-generating distribution takes a complex form, because KDE imposes relatively few assumptions about the shape of the distribution that underlies the observed data (Silverman 1986). Instead of forcing our probability models to conform to one of the named distributions, the data are allowed to "speak for themselves" as we estimate the density function that generated them (Silverman 1986). In very simple terms, kernel density estimation attempts to represent the probability distribution from which a data set was generated by placing a symmetric "bump" in probability density at every data point (Cai 2013). The bumps are scaled down, such that the sum of their densities will be 1, and are subsequently added together. The height of the resultant density at a given location represents the predicted likelihood that future events will occur there (Cai 2013).

2.1 Kernels

A more mathematically rigorous explanation of KDE requires a definition of the term "kernel." A kernel is a nonnegative, real-valued, even function whose definite integral (over the region on which it is defined) is 1 (Cai 2013). Because kernels are non-negative and integrate to 1, they are simply a subset of probability density functions. Usually, the kernels employed for KDE are unimodal and symmetric about zero (Sheather 2004). One of the most common kernels is the Gaussian kernel, defined as

$$K(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}$$

or, in the case that the data are d-dimensional, as

$$K(\mathbf{y}) = \left(\frac{1}{2\pi}\right)^{\frac{d}{2}} e^{-\frac{1}{2}\mathbf{y}^{\mathbf{T}}\mathbf{y}}$$

with $\mathbf{y}$ a d-dimensional vector (Gutierrez-Osuna N.D.). In KDE, the kernel function determines the general form of the "bumps" that are placed at each data point. (The final shape is also determined by the bandwidth parameter, which controls how wide each "bump" is.) Gaussian kernels result in "bumps" that resemble normal distributions. We use Gaussian kernels in this paper due to their well-understood mathematical properties and good empirical performance (Sheather 2004; Karplus 2015).

To construct a kernel density estimate, a scaled kernel function is applied to each of n data points x_i ($0 \leq i \leq n$) (Cai 2013). Those functions are then summed; their sum is divided by the total number of data points in the sample (Cai 2013). The second step has the effect of placing a weighting of $\frac{1}{n}$ on each of the

summands, so as to guarantee that their combined density is one.[1] In the univariate setting, the resulting density estimate $\hat{f}$, at a point x, takes the form

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left[\frac{(x-x_i)}{h}\right]$$

where K is the kernel function (Silverman 1986). Thus, if one is estimating a univariate density using Guassian kernels and a bandwidth h, the density estimate is $\frac{1}{n}$ times the sum of n normal densities with means x_i $(1 \leq i \leq n)$ each of which has standard deviation h. When the data points are d-dimensional, the density estimate $\hat{f}$, at a point $\mathbf{x}$, is

$$\hat{f}(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n} K_{\mathbf{H}}\left[\mathbf{x}-\mathbf{X_i}\right]$$

where $\mathbf{x}$ and $\mathbf{X_i}$ are d-dimensional vectors, $K_{\mathbf{H}}(\mathbf{z}) = |\mathbf{H}|^{-1/2}K(\mathbf{H}^{-1/2}\mathbf{z})$, and $\mathbf{H}$ is a symmetric d-by-d bandwidth matrix (Duong 2007). Thus, if one is estimating a multivariate density using Gaussian kernels and a bandwidth matrix $\mathbf{H}$, the density estimate is $\frac{1}{n}$ times the sum of n multivariate normal densities with means $\mathbf{X_i}$ $(0 \leq i \leq n)$ each of which has covariance matrix $\mathbf{H}$.

2.2 Bandwidths

It is commonly accepted that bandwidth selection is one of the most important and difficult components of creating a useful density estimate (Cai 2013). The bandwidth parameter, h or $\mathbf{H}$ in the formulas above, controls the smoothness of the estimated density (Chiu 1991). In the univariate case, a relatively small bandwidth results in a taller, more concentrated "bump" around a given data point and a large bandwidth results in a shorter, wider "bump" (Silverman 1986). The multivariate case follows the same rule, with the diagonal elements of the bandwidth matrix controlling smoothing in the directions of the coordinate axes and the off-diagonal elements controlling smoothing in directions oblique to the coordinate axes (Duong 2016).

While it is sometimes the case that the optimal bandwidth can be chosen simply by looking at the density estimates that result from a range of potential bandwidths, this method is impractical in situations where several estimates are needed (Zambom & Dias 2012). In those cases, an automatic selection procedure is usually preferred (Zambom & Dias 2012). We suspect that NGOs may both want to regularly update their predictions of where conflict will occur and want to employ this technique for predicting conflict in many countries (each of which may necessitate a different bandwidth). Thus, we believe an automatic selection method would be necessary if our method is to be widely adopted. Below, we review the bandwidth selection techniques that we tested. For simplicity, the following discussion occurs in a univariate context. The same principles hold in the multivariate case and there exist multivariate analogues of the relevant formulas (Duong 2007; Chacon, et. al. 2009).

As with parameter selection for all statistical methods, the goal in selecting a bandwidth for a kernel density estimator is to minimize the estimator's Mean-Squared Error (MSE). The MSE of an estimator, $\hat{f}$, is often decomposed into the variance of the estimator and its squared-bias, expressed as

$$MSE(\hat{f}(x)) = E\left[(\hat{f}(x) - f(x))^2\right] = Var(\hat{f}(x)) + Bias^2(\hat{f}(x))$$

(Hastie, et. al. 2013). In the case of KDE, taking f to be the true density and $\hat{f}$ to be the estimated density, it can be shown, via Taylor series expansion, that the bias term can be written as

$$Bias(\hat{f}(x)) = \frac{h^2}{2} f''(x)\mu_2(K) + o(h^2)$$

[1] A statistician could also weight certain data points more heavily than others by multiplying the appropriate summands by constants other than 1, as long as all the constants sum to n (the number of data points).

where $\mu_2(K) = \int u^2 K(u) du$ (Zambom & Dias 2012). Similarly, one can show that the variance is

$$Var(\hat{f}(x)) = \frac{1}{nh} R(K) f(x) + o(\frac{1}{nh})$$

where $R(K) = \int K^2(y) dy$ (Zambom & Dias 2012). Combining the above, we obtain the MSE as

$$MSE(\hat{f(x)}) = \frac{h^4}{4} f''^2(x) \mu_2(K)^2 + \frac{1}{nh} R(K) f(x) + o(h^4) + o\left(\frac{1}{nh}\right)$$

which allows us to see that the bandwidth, h, plays a critical role in the "Bias-Variance Tradeoff" (Zambom & Dias 2012). A larger bandwidth results in a smoother (more biased, less variable) density estimate, while a smaller bandwidth yields a more focused (less biased, more variable) function.

Asymptotically, the "o(-)" terms go to zero and we obtain the Asymptotic Mean Squared Error (AMSE) (Sheather 2004). Integrating AMSE yields the Asymptotic Mean Integrated Squared Error (AMISE),

$$AMISE(\hat{f}(x)) = E \int (\hat{f}(x) - f(x))^2 dx = \frac{R(K)}{nh} + \frac{h^4 \mu_2^2(K) R(f'')}{4}$$

(Sheather 2004). From the above formula, one can obtain a closed-form expression for AMISE-minimizing bandwidth,

$$h_{AMISE} = \left(\frac{R(K)}{\mu_2^2(K) R(f'')} \right)^{\frac{1}{5}} n^{-\frac{1}{5}}$$

(Sheather 2004). This equation is the starting point for many bandwidth selection procedures.

However, the above formula for the AMISE-minimizing bandwidth depends on the unknown function f. In light of this, a number of methods have been proposed for selecting a bandwidth that yields viable density estimates, but which does not require knowing the second derivative of the function one intends to estimate. There are a few primary types of such methods: "rule-of-thumb" methods, "plug-in" methods, and cross-validation methods. Rule-of-thumb methods choose the optimal bandwidth under the assumption that the estimated density takes some specific shape (Sheather 2004). Plug-in methods find an initial estimate of f (often via an initial estimate of the bandwidth) and "plug-it-in" to the formula for AMISE (Sheather 2004; Zambom & Dias 2012). Cross-validation-based methods attempt to approximate the closeness of f and $\hat{f}$ by gauging how well $\hat{f}$, trained on a subset of the data, predicts the outcomes in another subset (Duong 2007). We tested at least one version of all of these methods.

Rule-of-thumb procedures minimize AMISE under the assumption that the density being estimated takes a specific form (Sheather 2004). The most common method is to assume that the estimated density is normal with standard deviation equal to the sample standard deviation (Wang, et. al. 2008). This "normal scale rule" then uses the known properties of the normal distribution in place of the unknown ones of f (Henderson & Parmeter 2015). It produces bandwidth of the form

$$h_{AMISE} = \left(\frac{8\pi^{\frac{1}{2}} C}{3n} \right)^{\frac{1}{5}} \hat{\sigma}$$

where $C = \left(\frac{R(K)}{\mu_2(K)^2} \right)^{\frac{1}{5}}$ and $\hat{\sigma}$ is the sample standard deviation (Wang, et. al. 2008). The potential problem with this method is obvious: the density one is trying to estimate is rarely normal, so the bandwidth one selects with this procedure is unlikely to be truly optimal. However, these methods are quite computationally efficient and have shown good results empirically (Sheather 2004).

Plug-in methods find an initial estimate of f and "plug" that estimate into the equation for the AMISE-minimizing bandwidth. One of the most common plug-in methods employs a relationship, based on the normal distribution, between a pilot bandwidth (an initial guess at the bandwidth) and the true band-

width (Loader 1999; Duong & Hazelton 2003). We refer the reader to the referenced papers for additional details the topic. Because the plug-in method relies on an externally-specified pilot bandwidth, it can be quite susceptible to variations in that value (Loader 1999). The method based on the normal distribution is the least variable of the plug-in methods but, as a consequence, tends to have difficulty selecting small bandwidths, even when those would be optimal (Loader 1999). The result is that the bandwidths selected by this method sometimes oversmooth the density they are being used to estimate, which increases the estimator's bias (Loader 1999).

Cross-validation-based methods for choosing a bandwidth constitute the primary alternative to rule-of-thumb and plug-in selectors. One of the earliest of these was the least squares cross validation (LSCV) method (also referred to as "Unbiased Cross Validation"). LSCV is based on the Integrated Squared Error (ISE), which is expressed mathematically as

$$ISE(f\hat{(}x)) = \int(\hat{f}_h(x) - f(x))^2 dx = \int(\hat{f}_h(x))^2 dx - 2\int \hat{f}_h(x) f_h(x) dx + \int(f(x))^2 dx$$

(Sheather 2004). Since the last term in the expansion does not rely on h, one can choose h so as to minimize an estimate of the other two terms (Sheather 2004). LSCV chooses h to minimize

$$\frac{1}{n}\sum_{i=1}^{n}\int(\hat{f}_{-i}(x))^2 dx - \frac{2}{n}\sum_{i=1}^{n}\hat{f}_{-i}(X_i)$$

the "LSCV score function" (Sheather 2004). In this formula, $\hat{f}_{-i}(x)$ denotes the density estimate for a point x where the estimate was computed on a data set without the observation x_i (Sheather 2004). It has been suggested that bandwidths based on LSCV have a tendency to undersmooth the densities they are being used to estimate, which results in undesirably "bumpy" density functions (Loader 1999; Zambom & Dias 2012). Moreover, LSCV often has difficulty handling data sets in which multiple events occur at the same values of the input variables. (In our context, this would be multiple conflicts happening at the same geographic location.) In the event that there are duplicates in the data, the LSCV score function is minimized at the degenerate bandwidth $h = 0$ (Silverman 1986). One possible solution to this problem is to remove the "duplicates" of a given observation (Bauder, et. al. 2015). This seems problematic, as it will also remove relevant information regarding the conflict density. Another option is to add a small amount of random noise to the data points (Beyer 2014). We tested LSCV by building densities using both unaltered data and data where a small bit of random noise has been added to each data point (thus slightly shifting its geographic coordinates).

Unlike LSCV, biased cross validation (BCV) minimizes an estimate of AMISE, rather than ISE.[2] Specifically, BCV replaces $R(f'')$ with $R(\hat{f}_h'') - \frac{1}{nh^5}R(K'')$ where $\hat{f}_h''$ is the second-derivative of the kernel density estimate, using the bandwidth h (Sheather 2004). Thus, the BCV score function is given by

$$AMISE(f\hat{(}x)) = E\int(\hat{f}(x) - f(x))^2 dx = \frac{R(K)}{nh} + \frac{h^4}{4}\mu_2^2(K)\left[R(\hat{f}_h'') - \frac{1}{nh^5}R(K'')\right]$$

(Zambom & Dias 2012). Since BCV replaces an unknown value in the formula for AMISE with a known estimate computed using the data, this method is often viewed as a hybrid of plug-in and cross-validation methods (Zambom & Dias 2012). It has been shown that the asymptotic variance associated with LSCV is much higher than that associated with BCV (Sheather 2004). We would expect, therefore, that the LSCV-based bandwidths will fluctuate much more than the BCV-based bandwidths in response to small changes in the training data.

[2] Later in the paper, we implement BCV1; however, both BCV1 and BCV2 were tested and they yield nearly identical results. We refer the reader to Duong 2007 for additional information regarding the distinction between the two procedures.

3 Previous Work

Kernel density estimation has been employed in a wide variety of contexts, from locating car-crash hotspots to pinpointing heavy metal hazard zones (Fu, et. al. 2015; Lin, et. al. 2010). Smooth density estimates are useful in such situations, because the individual data points (e.g., car accidents or soil samples) usually reflect information regarding locations proximate to the point where the datum was observed (e.g., that a certain stretch of road needs repair or that there is a large metal deposit in a certain area). For similar reasons, one of the most common applications of KDE is to crime hotspot identification, since the prevalence of crime provides information regarding the area in which it took place, i.e., that something about the area makes it crime-prone (Chainey 2008). In evaluations using historical data, kernel density estimation has outperformed other methods of crime hotspot identification (Chainey 2008). In fact, density estimates based on crime incidence are often used to allocate police patrol units (Chainey 2008; Gerber 2014).

KDE has even been employed in contexts similar to ours, i.e., in order to assess where violence was likely to occur. Downs, Mesev & Shirlow, for example, used KDE to obtain a density of deaths in Belfast related to "The Troubles" – the ethno-religious conflict in Northern Ireland (Downs, et. al. 2009). They then assessed how the probability density depends on other variables, such as a region's primary political affiliation or its level of religious segregation (Downs, et. al. 2009). Another paper created a density estimate for the placement of improvised explosive devices in Afghanistan and then used it to analyze the relationship between public health outcomes and that variety of violence (Curtis, et. al. 2015). The paper most directly related to ours is by Powers, Reeder & Townsen; they created probability densities for conflict in Angola, the DRC, the Ivory Coast, and Sierra Leone (Powers, et. al. 2015). They then used these densities as one input in a regression assessing where peacekeeping forces tend to move, once assigned to a country. Their regression showed that there is a learning curve for these forces; troops do gravitate towards areas in which conflicts have occurred, but only after they have been deployed in the country for several years. (We hope our work might serve to both decrease this learning curve and more accurately pinpoint where troops ought to be moved.) Powers, et. al. noted that they found kernel density estimation to be a powerful tool, because the technique allowed them to incorporate information regarding both the number of conflicts in geographic grid squares (their unit of analysis in the regression) and also the potential for spillover effects from high levels of conflict in proximate areas (Powers, et. al. 2015). (Unfortunately, we do not have access to these authors' data or models in order to compare them to ours.)

In contrast to this paper, these prior works have tended to use kernel density estimation as one input in an otherwise parametric (usually regression) model, rather than as an end in itself. Part of this difference lies in the fact that these papers usually do not strive to predict future events; instead, they attempt to infer relationships between the presence of conflict and other variables of interest. Moreover, the papers we have mentioned tend to employ subject-matter knowledge to choose the kernel's bandwidth, rather than an automated selection technique, or only employ a single bandwidth selection method, rather than trying to optimize across them. We hope to demonstrate the potential for KDE to stand on its own as a predictive model of conflict, even in the absence of a subject-matter experts' contributions to the bandwidth selection process.

4 Why use KDE?

Kernel density estimation is a particularly useful modeling technique in this context for a few reasons. Given the current instability in the DRC, the data required for a conventional parametric model are uniquely difficult to obtain. This is especially true given the detailed level at which we would need the data in order to make precise predictions. For example, a typical parametric model might attempt to predict the number of incidents of violence in a region based on the public services, income, or demographics in that region. However, the country's lack of infrastructure would make the collection of such

data difficult, if not impossible, and the lack of residential permanence due to the refugee crisis would make any data regarding a region's demographics suspect shortly after it is collected. Simply put, it is easier to obtain data on previous instances of violence than it is to obtain data on local political, demographic, or economic indicators, because the former are relatively well-publicized and the latter may not even exist.

Moreover, even if one were able to obtain the relevant input variables at the level of detail needed, it is likely that the high degree of interaction between variables would make many parametric models difficult to specify. Spatial data, such as the data set we use, are unlikely to have an easily-defined relationship between predictor variables and the outcome variable. Instead, there exist regions where a series of elements have come together in a specific way to produce the outcome. KDE has the ability to capture these highly nonlinear and interactive relationships, which might go unaccounted for in a parametric model. KDE allows us to effectively construct a model based on "Tobler's First Law of Geography," that "everything is related to everything else, but near things are more related than distant things" (Tobler 1970). Moreover, because data points in particular regions are likely to be similar, spatial data tend to violate one of the fundamental assumptions of regression modeling – that the observations are identically and independently distributed, given the predictor variables. In these cases, conventional parametric models generate inaccurate error estimates (Sun 2014). In contrast, KDE makes no such assumptions, and therefore does not encounter the same issues.

There are, however, parametric methods for analyzing spatial data, which may avoid some of these problems. The basis for many of these techniques is spatial autoregressive modeling, which estimates the output in a given location as partially dependent on the output in other locations (Sun 2014). Most implementations of spatial autoregressive models would involve dividing the country into r individual regions and defining an r-by-r spatial weighting matrix to indicate how the outcome in each region is affected by the outcomes in other regions (Sun 2014; Smith 2015). While this method might be useful if we had only obtained information on the number of conflicts in each region or administrative territory, we in fact have more detailed information (i.e., the exact location of the event) that we could (and should) exploit. Additionally, implementing a spatial autoregressive model would require that we define a weighting matrix (Banerjee 2009). That process begs the same question as bandwidth selection in KDE – how great an effect does one location have on another location X miles away? Creating such a weighting matrix involves either a large number of assumptions or a large number of data points, since one must determine all $\sim r^2$ weights. Moreover, because we believe the presence of conflict should decay smoothly as one moves away from the primary cause, we prefer the smooth prediction provided by KDE to the stepwise one that a region-level spatial autoregressive models would produce. Additionally, spatial autoregressive models are often not useful absent covariates – precisely the data we lack, as discussed previously (WISE 2010).

Kriging is another parametric option; it predicts values at a given point based on a weighted average of values at other points, where the weights are decreasing functions of distance (Bohling 2005). Generally, kriging assumes that responses are distributed according to a Gaussian process (Kie 2010). Kriging then estimates the covariance matrix based on the data at hand (Bachoc 2014). They key to kriging is much the same as it is to nearly all spatial methods – accurately estimating the covariance between values at different points (Atkinson & Lloyd 2010). Therein lies the first problem. Extensions of kriging to Poisson or binomial data (as we have here) do exist, but seem to encounter many problems with covariance matrix estimation, due to the lack of an appropriate joint multivariate distribution on which to base it (Atkinson & Lloyd 2010). Practically, this absence often results in the generation of estimated covariance matrices that are not positive semidefinite or that are numerically unstable, both of which create fundamental problems for inference and prediction (Atkinson & Lloyd 2010). It also appears to be uniquely difficult to constrain kriging estimates to be between 0 and 1, which prevents us from obtaining a true probability distribution, since probabilities can only fall in that range of values (Atkinson & Lloyd 2010). Most significantly, however, kriging performs best with well-spaced data points and tends to encounter difficulties when observations are clumped together (Kie 2010). For this reason, kriging is often used in geostatisti-

cal applications, where the scientists can take well-spaced soil samples (Lin 2010). However, our data are not quite as compatible with this method; the nature of our data is that events tend to occur near other events, because their (similar) environments have made those locations uniquely conflict-prone.

5 DATA

The data used in this paper were obtained from the Armed Conflict Location and Event Data Project (ACLED), an organization associated with the University of Sussex ("About ACLED"). ACLED collects real-time data on political violence in both Africa and South/Southeast Asia ("About ACLED"). This paper employs the subset of ACLED's Africa data set related to the DRC. ACLED records key information (e.g., date, location, actors, number of fatalities, etc.) about instances of conflict or events that may be precursors to it (e.g., the establishment of bases by rebel groups). The organization aggregates data from many sources, including local media outlets and reports by nongovernmental organizations ("About ACLED"). To date, ACLED has produced the most comprehensive public collection of such records ("ACLED"). In the past several years, the organization's data have been used in papers published in a number of major political science journals, such as *World Development*, *Journal of Peace Research*, and *Journal of Conflict Resolution* (Dabalen and Sumik 2014; Landis 2014; Raleigh 2014).

Events recorded in the ACLED data range from nonviolent territory transfers to declared wars (Raleigh & Dowd 2016). We focused on the three most prevalent event categories out of the ones that appear to indicate true instances of conflict (in contrast to events such as rebel base establishment, which are not immediately violent). These categories are "Battle," "Violence against civilians," and "Riots/Protests." ACLED separates battles into three event types ("Battle-Government regains territory," "Battle-No change of territory," and "Battle-Non-state actor overtakes territory"). Since these distinctions are not especially relevant to us, we aggregated them all into one category ("Battle"). The ACLED data also include events categorized as "Headquarters or base established," "Non-violent transfer of territory," "Strategic development," and "Remote violence." We excluded these event categories from our analysis. The first three categories were excluded on the grounds that they do not appear to indicate instances of violent conflict and the last is excluded because it has too few records to produce a reasonable density estimate.

While ACLED maintains data on events from 1997 onward, this paper uses only observations from 2014 and 2015. We anticipate that using more recent data will result in a more accurate assessment of whether our models might be useful in predicting future events. After limiting the data by timeframe and event type, we had 1,737 observations, in total; 960 from 2014 and 777 from 2015. The number of observations in each event category ("Battle", "Riot/Protest", and "Violence Against Civilians") is presented as Table 1. Additionally, the number of monthly events in each category during 2014 and 2015 is presented in graphical form as Figure 1 and tabular form as Table 2. Based on these summaries, we concluded that, while there is relatively high month-to-month variability in the number of events, the difference between 2014 totals and 2015 totals (both overall and within each category) is not particularly large. This suggests that there was not a structural break between the two years, regarding the manner in or frequency with which conflict occurred. Thus, it does not seem unreasonable to use the data from one of these years to build our densities and the other to evaluate them. Additionally, given the high intra-month variation, we concluded that that we should aggregate the data to the annual level for our analysis.

It is also worth noting that the ACLED data sometimes record two different events as having taken place at precisely the same geographic location. This usually occurs because ACLED only has access to information regarding the town or administrative district in which an event occurred, rather than the exact event location (Raleigh & Dowd 2016). It does not appear that ACLED's data contain multiple records that correspond to the same event. We present visualizations of both 2014 and 2015 event locations (aggregated as well as separated by event type) as Figures 3 through 6. The locations of the events in which we are interested do not appear to be radically different between 2014 and 2015. E.g., many of the events

(of all types) seem to be concentrated in the northeastern part of the country. Thus, we feel reasonably confident that 2014 and 2015 event locations are not so different as to clearly invalidate our suggested prediction method.

That being said, it is apparent, even from these initial conflict maps that some event types will be easier to predict than others. For example, both in 2014 and in 2015, riots/protests were scattered throughout the country, with only a slightly greater concentration occurring in the northeast. In contrast, the vast majority of battles and instances of violence against civilians, in both 2014 and 2015, were concentrated primarily in the northeast. However, it is these types of events (battles and violence against civilians) which resulted in the greatest loss of life over the course of 2014 and 2015. (See Table 3 for the number of fatalities associated with each event in each year.) If we take loss of life to be an indicator of the severity of an event, our preliminary maps seem to indicate that we may be able to create predictions regarding the events of greatest importance.

6 Analysis

We created kernel density estimates ("models") based on 2014 conflict locations, using the four different bandwidth selection procedures discussed previously ("rule-of-thumb," "plug-in," LSCV, and BCV). Because our data contain multiple events with the same geographic coordinates, we produced two density estimates using LSCV – one without altering the data and one in which we added a small bit of random noise to each coordinate.[3] (In the interest of concision, we refer to the former as "LSCV" and the latter as "LSCV-R.") Each of the five bandwidth selection techniques was used to create a density estimate for battles, violence against civilians, riots/protests, and "all conflicts." The last category, "all conflicts," is simply the union of the other three categories. In order to test the predictive power of these density estimates, we compared them to the locations of actual 2015 events of the relevant type, (which were not used to create the probability density). E.g., the density estimate for battles (built using 2014 battle locations) was compared to 2015 battle locations.

We used surveillance plots to evaluate the effectiveness of the different models that we produced. At a high level, the surveillance plot for a model is a curve that shows the percentage of true events (here, 2015 conflicts) that occurred in the x% of the country's area that the model considers to be most-threatened, i.e., the area that the model estimates to have the highest probability of conflict (Gerber 2014).[4] Every point on the curve corresponds to a proportion of the country being surveilled (x-axis) and to the proportion of actual events observed in the surveilled area (y-axis). (See Figures 7, 13, 19, and 25, as examples.) Better models capture a large number of events in relatively little area and therefore produce curves that lie close to the upper left corner of the plot (Gerber 2014). If conflicts occur uniformly at random across the country, our models are unlikely to be particularly effective, since there would be no relationship between an event occurring in a location during one year and an event occurring in a nearby location during the next year. We refer to the model that predicts that events occur uniformly at random as the "null model." It would assign every region of the country the same probability of conflict and therefore would produce (on average) a 45-degree line in the surveillance plot. (This is the case because, if one surveils a random 20% of a country's territory, one would expect to encounter, on average, 20% of the country's total conflicts.) However, if events occur with particular frequency in the same locations

[3] The random noise was drawn from a uniform distribution defined between 0 and 0.1 (degrees of latitude or longitude). This interval was chosen, because it corresponds to the length and width of the grid squares into which we divided the country for surveillance plot analysis. (See footnote 4.) Thus the noise cannot shift an event more than one grid square in any direction.

[4] In order to associate events and estimated densities with areas of the country, we followed the standard procedure for creating surveillance plots. We first divided the country into a grid of squares (with sides of approximately 0.1 degrees – 7 miles – of latitude and longitude) and counted the number of conflicts that occurred within each square. We then found the center of each of the squares and calculated the predicted probability density at that location, using each model. The density predictions for these center points were applied to the full square in which they lie. The squares are small enough that, within each of them, there is not much variation in predicted density.

across years, our density estimates will include some of that information – having obtained it from being trained on 2014 data – and will use that information to more accurately determine the likely locations of future (2015) conflicts. We also produced single-number summaries for each of the models, based on the model's area under the curve (AUC) in the surveillance plot. Higher numbers indicate better predictions (i.e., lines further up and to the left). The "null model" would have an expected AUC of 0.5 (Gerber 2014). The AUCs serve as a convenient method by which to compare models. Table 4 shows the AUCs for all of the models that we produced.

The standard strategy by which one might create confidence intervals for these AUCs – i.e., bootstrap resampling our original data set – does not work here. The bootstrap assumes that the data points in the initial sample are independent from one another (Deutsch N.D.). In our context, however, the data points are spatially correlated (and thus not independent). The traditional bootstrap would be unable to replicate the spatial correlation in the data (Deutsch N.D.). There are a number of adjustments to the bootstrap, most of which rely on the block-bootstrap, that could be employed in order to partially account for the spatial correlation.[5] The block-bootstrap involves dividing one's data into (potentially overlapping) sections, "blocks," and then resampling the blocks ("The Block Bootstrap"). For spatial data, one might resample grid squares defined on a map (Lahiri & Zhu 2006). However, the selection of block size is no small task, as an appropriate choice depends on many factors. More significantly, there are many conditions – e.g., if the data are noisy or irregularly spaced (both of which are true here) – under which the conventional block-bootstrap will fail (Lahiri & Zhu 2006; Mader, et. al. 2013). In those cases, the process requires substantive modification, away from the standard block-bootstrap, in order to produce accurate results (Lahiri & Zhu 2006; Mader, et. al. 2013). In light of time, computational, and page-length constraints, and in order to avoid reporting erroneous confidence intervals, we defer their creation to another paper.

6.1 All Conflicts, Battles, & Violence Against Civilians

The sets of models predicting battles, those predicting violence against civilians, and those predicting "all conflicts" produced thematically similar results, so we explicitly review only the last of these (i.e., the set of models predicting "all conflicts"). Results related to the other two categories of conflict (battles and violence against civilians) are presented in the appendix (and referenced in footnotes throughout this section). First, we display (as Figure 7) the surveillance plot comparing the models for "all conflicts."[6] Next, we present heat maps based on each model's predictions (Figures 8–12).[7] The yellow and orange areas of the heat maps indicate relatively high estimated probability density and the green areas indicate relatively low estimated probability density. We suspect it may be useful for the reader to compare these to Figure 3, which shows the locations of conflict in 2014 and 2015.[8]

Even from the heat maps, it is apparent that if conflict is concentrated in any particular area, the density estimate resulting from LSCV will be a relatively poor predictor. This is confirmed by the surveillance plot and AUC calculations; the LSCV-based model yields an AUC of only about 0.56, close to the AUC that would be expected under the "null model." The problem stems from the fact that the LSCV-based model suggests that nearly all areas are equally likely to see conflict. The results from the LSCV-based model are likely attributable to the presence of "duplicates" in our data. (See Sections 2.1 and 5.) However, adding random noise to the coordinates (in order to "de-duplicate" the data) seems to not only resolve the problems in the LSCV-based density estimate, but actually results in a very strong predictive model. LSCV-R yields an AUC of 0.94; it seems that the random shifts created just enough of a spread for LSCV to work

[5] There are other methods, e.g., the Family-Wise Error Rate-based corrections suggested by Hutson (2004), however, the vast majority of alternatives employ the block-bootstrap in some capacity.

[6] The surveillance plots comparing the models predicting battles is presented as Figure 13 and the surveillance plot comparing the models predicting violence against civilians is presented as Figure 19.

[7] The heat maps for the models predicting battles are presented as Figures 14–18 and the heat maps for the models predicting violence against civilians are presented as Figures 20–24.

[8] Analogous maps for battles and violence against civilians are Figures 4 and 5, respectively.

appropriately (without over-dispersing the data) and that, once working, it is highly effective. While the heat maps for LSCV and LSCV-R do not, at first, appear to be dramatically different, the differences that exist are critical. They both indicate a low probability of conflict for most of the country, which is consistent with reality. Much of the country's territory (because it is of its low population density) does not, in fact, experience widespread conflict. However, the heat map produced by LSCV-R identifies a number of spots in the northeast of the country as having a high probability of conflict and the area surrounding those spots as having a moderate probability of conflict. In contrast, LSCV produces an almost entirely uniform density with only one or two locations of higher predicted probability, which only sometimes seem to correspond to the primary regions in which violence took place during 2014 and 2015.

The other density estimates have AUCs that fall somewhere between the LSCV-based model and the LSCV-R-based model. The plug-in bandwidth produces a density estimate that seems to be relatively concentrated on one region in the northeastern part of the country. However, it is still not nearly as tightly focused as the LSCV-R-based model and, as a result, produces a slightly lower AUC (of around 0.9). The BCV and the "rule-of-thumb" bandwidths result in relatively similar density estimates, in which the primary area of high density is a tilted oval-shaped region in the northeastern part of the country. The region of high density produced by the "rule-of-thumb" bandwidth is just slightly smaller than that produced by the BCV bandwidth. The oval shape of both regions, however, is indicative of a fair bit of over-smoothing. This is borne out in the AUCs for these models (0.85 and 0.86, respectively), both of which are lower than those of the plug-in-based model and the LSCV-R-based model.

6.2 Riots/Protests

The set of models created to predict riots/protests stand in contrast to those created to predict events in the other categories. First, LSCV does not fail in the case of riots/protests, making LSCV-R substantively less useful. Second, there is a different ordering of the bandwidth selection methods, as far as which one generates the highest AUC. We present a surveillance plot comparing the various density estimates for riots/protests as Figure 25 and heat maps corresponding to each of the models as Figures 26 through 30. We suspect it may be useful for the reader to compare these to Figure 6, which shows the locations of riots/protests in 2014 and 2015.

It appears that 2014 riots/protests were sufficiently spread out that LSCV does not result in a degenerate bandwidth. It, in fact, produces a reasonable density estimate with relatively large contiguous areas of high probability, both in the northeast of the country and, to a lesser extent, the southwest. Because LSCV does not result in a problematic bandwidth, the addition of random noise to the coordinates (i.e., LSCV-R) does little to improve the predictive accuracy of the density. Both LSCV and LSCV-R result in similar heat maps and AUCs of around 0.85. The BCV and "rule-of-thumb" bandwidths produce results similar to those of LSCV and LSCV-R, though the areas of high probability density are larger. This tendency towards greater smoothing generates lower AUCs for both the BCV- and "rule-of-thumb"-based models, at 0.83 and 0.84, respectively. The heat map for the plug-in bandwidth produces the model with the highest AUC, at 0.86. The plug-in bandwidth shows regions of high probability density both in the northeast and southwest; these regions are smaller than those produced by the other methods. Its slight improvement in performance over the others seems to be due to its ability to avoid over-smoothing the areas of high estimated probability density.

7 Conclusions

Based on our results, it seems that using kernel density estimation to predict where (within in the DRC) future conflicts will occur is reasonably effective. We were able to achieve surveillance plot AUCs above 0.85 for each event category; KDE generated models that accurately predicted a high number of instances

of future conflict in a relatively small amount of territory. In theory, this means that troops stationed in accordance with a model of the type we suggest would be able to prevent these violent outbreaks. The large differences between our models' AUCs and the AUC expected from the "null model" make it unlikely that our high AUCs are due to chance. Thus, we can say, relatively decisively, that most of our models perform better than the "null model" of random troop allocation. However, because we do not generate confidence intervals, it is difficult to tell whether some of the better-performing models (e.g., LSCV-R and the plug-in method) produce AUCs that are meaningfully different. As mentioned previously, this would be a compelling topic for further investigation. Additionally, due to the lack of data on the subject, it is difficult to know if our models are better than current techniques, which are largely based on subject-matter expertise. Given the seemingly nondecreasing levels of violence (per Figure 2), we have reason to suspect our methods might outperform those currently in use. Assuming one could obtain appropriate data, it would be informative to conduct a comparison of current troop allocation methods with the KDE-based methods in this paper.

While KDE produces good results in general, some bandwidth selection techniques perform better than others. For our particular problem, LSCV-R and plug-in methods seem to be the most successful of the selection techniques that we have tried. LSCV-R yields the best results in three of four event categories (with the plug-in method coming in second in all three of them) and the plug-in method is best in the fourth event category (with LSCV/LSCV-R coming in second). If an organization were to implement KDE-based troop placement in the DRC, we suspect that both LSCV-R and plug-in methods would be successful. Based on our analysis, it seems that the optimal bandwidth selection method depends, in part, on the event type of interest. As mentioned previously, battles and acts of violence against civilians seem to be responsible for the greatest loss of life. (See Table 3.) It may therefore be advisable to err toward LSCV-R, as this bandwidth selection method tends to perform best in predicting events of those varieties.

Because the optimal bandwidth selection method differs by event category, we suspect that it is also likely to depend on other time- and country-specific factors. Therefore, the precise choice of bandwidth selector ought to depend on the particular scenario at hand and each should be evaluated using historical data. More generally, it appears that the success of both LSCV-R and the plug-in method stems from their ability to avoid over-smoothing the density estimates.[9] Therefore, rather than identifying any particular bandwidth selection method as being the "best," we primarily suggest that anyone implementing KDE for similar purposes ought to actively avoid bandwidths that might cause oversmoothing – erring, when necessary, towards undersmoothing. Along the same lines, a further avenue of research might be to conduct analyses similar ours, using additional bandwidth selectors or data from additional contexts, i.e., other countries, other timeframes, etc.

Ultimately, we hope that our work may be useful to the United Nations' forces in the Democratic Republic of the Congo. We have created models that, with reasonable accuracy, predict the locations of future conflicts within the country. We anticipate that such models could be helpful as the UN proceeds with its removal of troops from the nation, as the models would allow the organization to direct the remaining troops and their local replacements (i.e., DRC police and government forces) to the locations in which these individuals are likely to be most needed. We also hope that the techniques reviewed in this paper will be tried in other contexts, in an effort to improve peacekeeping worldwide. Given the predictive success of these models in the DRC, it seems quite likely that they are more universally applicable with only a mild amount of tuning. More generally, we urge all organizations sponsoring peacekeeping forces to consider a data-based troop allocation strategy, as historical conflict in a given location is clearly a strong indicator of future conflict in the same area.

[9] The results from LSCV-R are consistent with our research regarding the various bandwidth selection methods. (See Section 2.1.) We found that LSCV is known to err towards undersmoothing rather than oversmoothing. True to form, LSCV-R (which is undergirded by LSCV) created very tightly concentrated areas of high probability. In contrast, our research indicated that the plug-in method we were using was likely to oversmooth the density. For our data set, it appears that this is not the case.

8 REFERENCES

Atkinson, Peter M., and C. D. Lloyd. *GeoENV VII – Geostatistics for Environmental Applications.* New York: Springer Science & Business Media, 2010.

"About ACLED." ACLED. Web. 23 Apr. 2016.

Bachoc, François. "Kriging Models with Gaussian Processes - Covariance Function Estimation and Impact of Spatial Sampling." Mar. 2014. Web. 15 May 2016.
<http://www.math.univ-toulouse.fr/ fbachoc/Bachoc_Forge_les_eaux.pdf>.

Banerjee, Sudipto. "Spatial Autoregressive Models." 22 Sept. 2009. Web. 17 May 2016.
<http://www.unc.edu/ rls/s940/Areal.pdf>.

Bauder, Javan M., David R. Breininger, M. Rebecca Bolt, Michael L. Legare, Christopher L. Jenkins, and Kevin Mcgarigal. "The Role of the Bandwidth Matrix in Influencing Kernel Home Range Estimates for Snakes Using VHF Telemetry Data." *Wildlife Research* 42.5 (2015): 437.

Beyer, Hawthorne L. "Kde (Kernel Density Estimation)." Geospatial Modelling Environment. 2014. Web. 17 May 2016. <http://www.spatialecology.com/gme/kde.htm>.

"The Block Bootstrap." (n.d.): n. pag. Web. 25 May 2016.
<http://nccur.lib.nccu.edu.tw/bitstream/140.119/35143/6/51007106.pdf>.

Bohling, Geoff. "Kriging." 19 Oct. 2005. Web. 17 May 2016.
<http://people.ku.edu/ gbohling/cpe940/Kriging.pdf>.

Brunsdon, Chris, and Lex Comber. *An Introduction to R for Spatial Analysis & Mapping.* Los Angeles: Sage, 2015.

Cai, Eric. "Exploratory Data Analysis: Kernel Density Estimation ? Conceptual Foundations." The Chemical Statistican, 9 June 2013. Web. 5 Apr. 2016. <https://chemicalstatistician.wordpress.com/2013/06/09/exploratory-data-analysis-kernel-density-estimation-in-r-on-ozone-pollution-data-in-new-york-and-ozonopolis/>.

Chacón, José E., Tarn Duong, and M. P. Wand. "Asymptotics for General Multivariate Kernel Density Derivative Estimators." *STAT Sinica Statistica Sinica* 21.2 (2011): 807.

Chainey, Spencer, Lisa Thompson, and Sebastian Uhlig. "The Utility of Hotspot Mapping for Predicting Spatial Patterns of Crime." Security Journal 21 (2008): 4-28.

Chiu, Shean-Tsong. "Bandwidth Selection for Kernel Density Estimation." *The Annals of Statistics* 19.4 (1991): 1883-1905.

Curtis, Andrew, Xinyue Ye, Kevin Hachey, Margaret Bourdeaux, and Alison Norris. "A Space-time Analysis of the WikiLeaks Afghan War Diary: A Resource for Analyzing the Conflict-health Nexus." *International Journal of Health Geographics Int J Health Geogr* 14.1 (2015): 14-29.

Dabalen, Andrew L and Sumik, Paul. 2014. "Effect of Conflict on Dietary Diversity: Evidence from Cote D'Ivoire." *World Development* 58: 143-158.

Deutsch, Clayton V. "Statistical Analysis of Longitudinal and Correlated Data." (n.d.): n. pag. Web. 25 May 2016.

<http://www.solvedproblems.com/sitebuildercontent/sitebuilderfiles/spatial_bootstrap.pdf>.

Downs, Joni, Peter Shirlow, and Victor Mesev. "The Geography of Conflict and Death in Belfast, Northern Ireland." *Annals of the Association of American Geographers* 99.5 (2009): 893-903.

Duong, Tarn. "Ks: Kernel Density Estimation and Kernel Discriminant Analysis for Multivariate Data in R." *Journal of Statistical Software* 21.7 (2007): n. pag.

Duong, Tarn. "ks: Kernel density estimation for bivariate data." 4 Apr. 2016. Web. 5 Apr. 2016. <https://cran.r-project.org/web/packages/ks/vignettes/kde.pdf>.

Duong, Tarn, and Martin Hazelton. "Plug-In Bandwidth Matrices for Bivariate Kernel Density Estimation." Nonparametric Statistics 15.1 (2003): 17-30.

Fu, Liping, Tae J. Kwon, and Lalita Thakali. "Identification of Crash Hotspots Using Kernel Density Estimation and Kriging Methods: A Comparison." *Journal of Modern Transportation* 23.2 (2015): 93-106. Web. 16 May 2016.

Gerber, Matthew S. "Predicting Crime Using Twitter and Kernel Density Estimation." *Decision Support Systems* 61 (2014): 115-25. Web.

Gutierrez-Osuna, Ricardo. "L7: Kernel Density Estimation." Web. 16 Apr. 2016. <http://research.cs.tamu.edu/prism/lectures/pr/pr_l7.pdf>.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. The Elements of Statistical Learning. New York: Springer, 2001.

Henderson, Daniel J., and Christopher M. Parmeter. *Applied Nonparametric Econometrics*. 1st ed. New York: Cambridge UP, 2015.

Hochschild, Adam. *King Leopold's Ghost: A Story of Greed, Terror, and Heroism in Colonial Africa*. Boston: Houghton Mifflin, 1998.

Hutson, A. D. "A Semiparametric Bootstrap Approach to Correlated Data Analysis Problems." Computer Methods and Programs in Biomedicine 73.2 (2004): 129-34.

Karplus, Kevin. "Kernel Density Estimates." Gas Station Without Pumps. 22 Jan. 2015. Web. 17 May 2016. <https://gasstationwithoutpumps.wordpress.com/2015/01/22/kernel-density-estimates/>.

Kie, J. G., J. Matthiopoulos, J. Fieberg, R. A. Powell, F. Cagnacci, M. S. Mitchell, J.-M. Gaillard, and P. R. Moorcroft. "The Home-range Concept: Are Traditional Estimators Still Relevant with Modern Telemetry Technology?" Philosophical Transactions of the Royal Society B: Biological Sciences 365.1550 (2010): 2221-231.

Lahiri, S. N., and Jun Zhu. "Resampling Methods for Spatial Regression Models Under a Class of Stochastic Designs." *The Annals of Statistics* 34.4 (2006): 1774-813.

Landis, Steven T. 2014. "Temperature seasonality and violent conflict: The inconsistencies of a warming planet." *Journal of Peace Research* 51: 603-618.

Leverich, Gerald. "Spatial Event Analysis Tool: An Application for Mapping Terrorists' Events." InSPIRE @ Redlands. N.p., Dec. 2008. Web. 25 May 2016.

<http://inspire.redlands.edu/cgi/viewcontent.cgi?article=1116&context=gis_gradproj>.

Lin, Yu-Pin, Hone-Jay Chu, Chen-Fa Wu, Tsun-Kuo Chang, and Chiu-Yang Chen. "Hotspot Analysis of Spatial Environmental Pollutants Using Kernel Density Estimation and Geostatistical Techniques." *International Journal of Environmental Resesearch and Public Health* 8.1 (2011): 75-88.

Loader, Clive. "Bandwidth Selection: Classical or Plug-In?" The Annals of Statistics 27.2 (1999): 415-38.

Mader, Malenka, Wolfgang Mader, Linda Sommerlade, Jens Timmer, and Björn Schelter. "Block-bootstrapping for Noisy Data." Journal of Neuroscience Methods 219.2 (2013): 285-91.

"MONUSCO Mandate - United Nations Organization Stabilization Mission in the Democratic Republic of the Congo." UN News Center. UN. Web. 23 Apr. 2016.

"MONUSCO Facts and Figures - United Nations Organization Stabilization Mission in the Democratic Republic of the Congo." UN News Center. UN. Web. 23 Apr. 2016.

"MONUSCO Background - United Nations Organization Stabilization Mission in the Democratic Republic of the Congo." UN News Center. UN. Web. 23 Apr. 2016.

Powers, Matthew, Bryce Reeder, and Ashley Adam Townsen. "Hot Spot Peacekeeping." International Studies Review 17 (2015): 46-66.

Raleigh, Clionadh. 2014. "Pragmatic and Promiscuous: Explaining the Rise of Competitive Political Militias across Africa." *Journal of Conflict Resolution*: 1-28.

Raleigh, Clionadh, and Caitriona Dowd. "Armed Conflict Location and Event Data Project (ACLED) Codebook." 2016. Web. 9 Apr. 2016.
<http://www.acleddata.com/wp-content/uploads/2016/01/ACLED_Codebook_2016.pdf>.

Rosen, Armin. "The Origins of War in the DRC." The Atlantic, 26 June 2013. Web. 20 Apr. 2016.

Sheather, Simon J. "Density Estimation." *Statistical Science* 19.4 (2004): 588-97.

Silverman, B. W. *Density Estimation for Statistics and Data Analysis.* London: Chapman and Hall, 1986.

Smith, M J De. "SAR Models." Statistical Analysis Handbook. 2015. Web. 17 May 2016.
<http://www.statsref.com/HTML/index.html?sar_models.html>.

Sun, Dennis. "STATS253: Analysis of Spatial and Temporal Data." 2014. Web. 24 May 2016.
<http://web.stanford.edu/class/stats253/lectures_2014.html>.

Wang, Hao, Carlos Scheidegger, and Claudio Silva. "Optimal Bandwidth Selection for MLS Surfaces." IEEE Xplore (2008): n. pag.

WISE. "Lecture 1: Introduction to Spatial Autoregressive (SAR) Models." 2010. Web. 17 May 2016.
<http://www.wise.xmu.edu.cn/UploadFiles/SS/Uploadfiles/2010769818938.pdf>.

Zambom, Adriano Zanin, and Ronaldo Dias. "A Review of Kernel Density Estimation with Applications to Econometrics." ArXiv. Cornell University Library, 12 Dec. 2012. Web. 30 May 2016.
<http://arxiv.org/abs/1212.2812>.

9 APPENDIX

Figure 1: Monthly Event Counts, 2014-2015

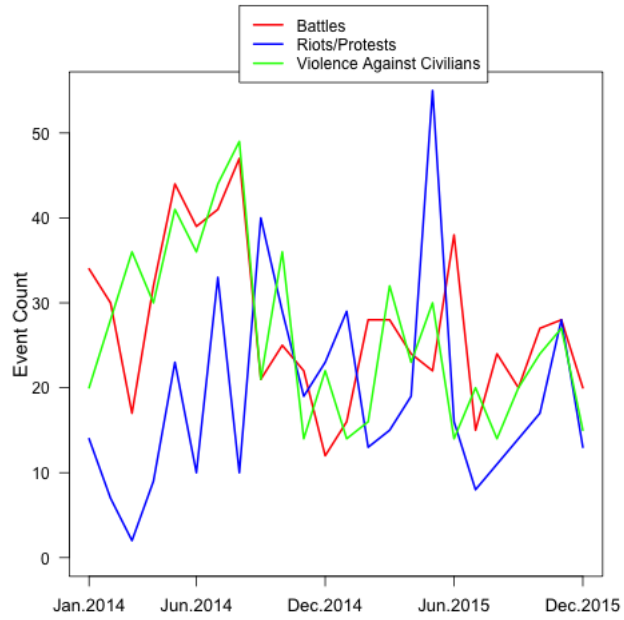

Figure 2: Annual Event Counts, 2010-2015

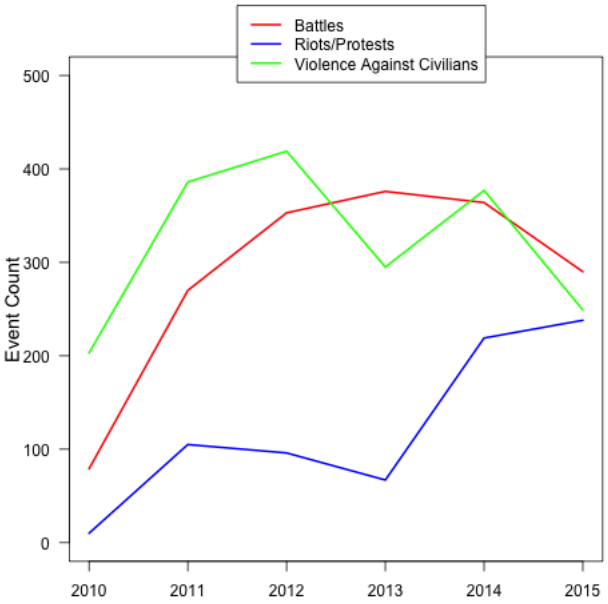

Figure 3: Point Maps of All Conflicts, 2014 (Left) and 2015 (Right)

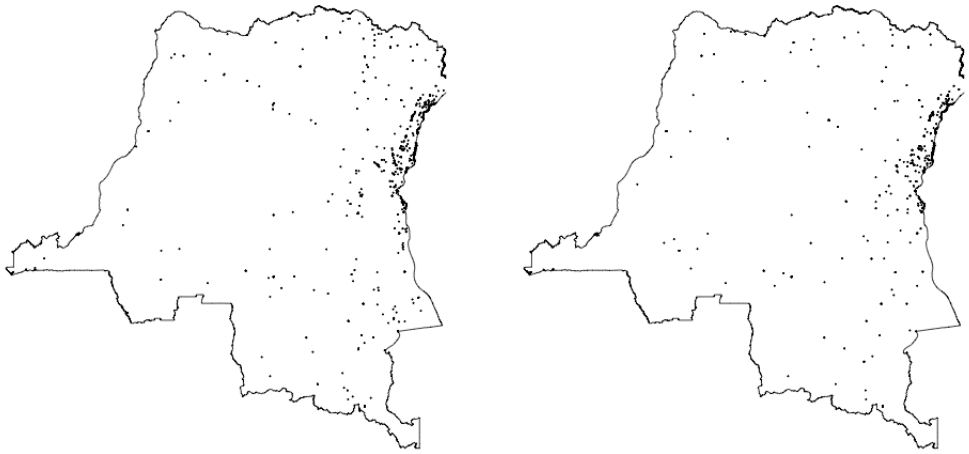

Figure 4: Point Maps of Battles, 2014 (Left) and 2015 (Right)

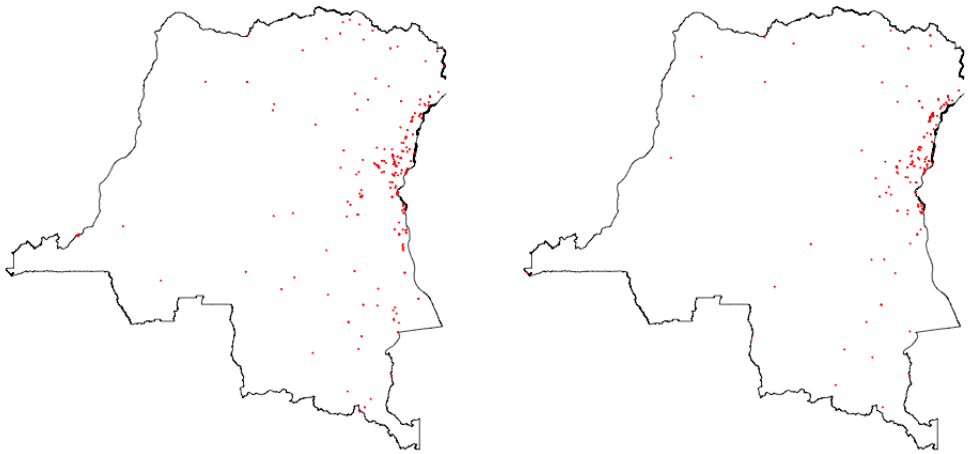

Figure 5: Point Maps of Acts of Violence Against Civilians, 2014 (Left) and 2015 (Right)

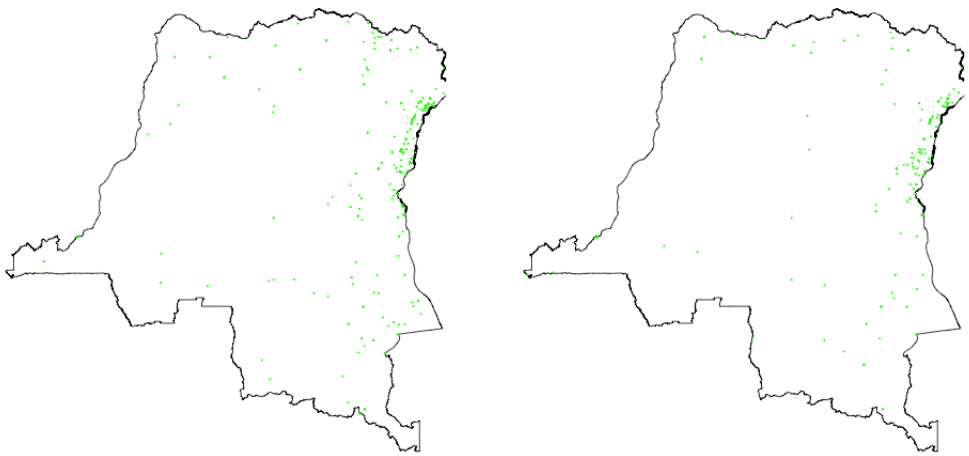

Figure 6: Point Maps of Acts of Riots/Protests, 2014 (Left) and 2015 (Right)

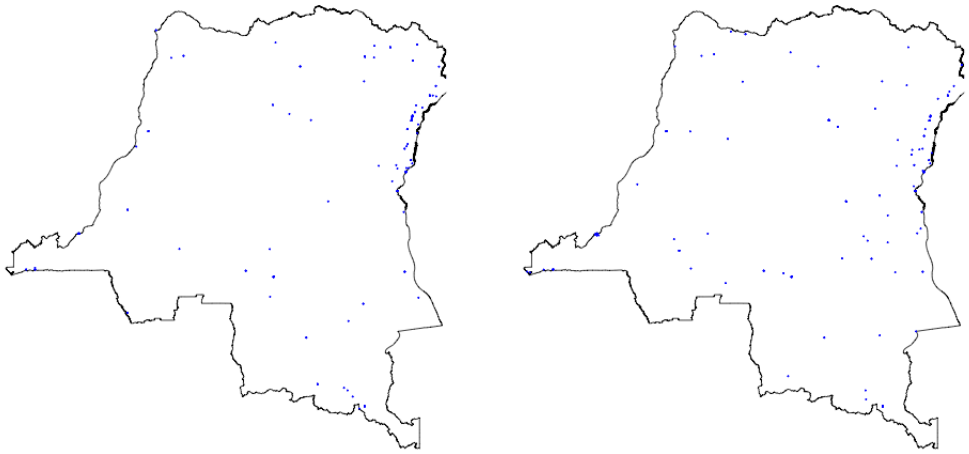

Figure 7: Surveillance Plot for All Conflicts

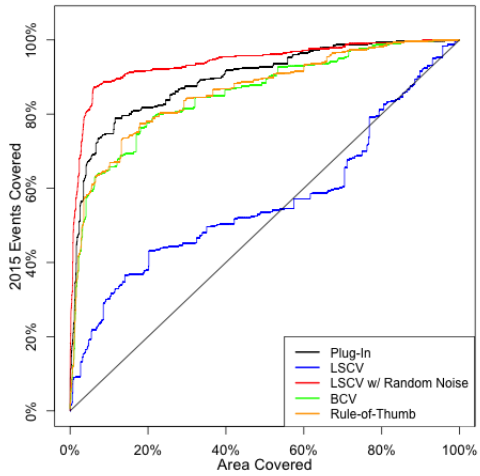

Figure 8: Density Estimate for All Conflicts Using Rule-of-Thumb Bandwidth

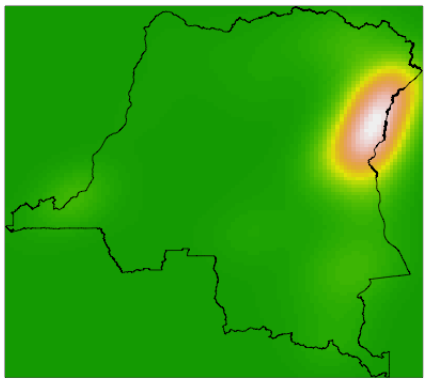

Figure 9: Density Estimate for All Conflicts Using Plug-In Bandwidth

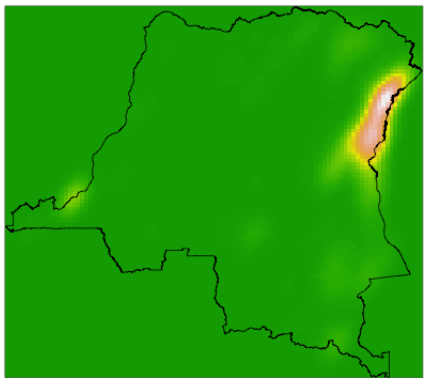

Figure 10: Density Estimate for All Conflicts Using LSCV Bandwidth

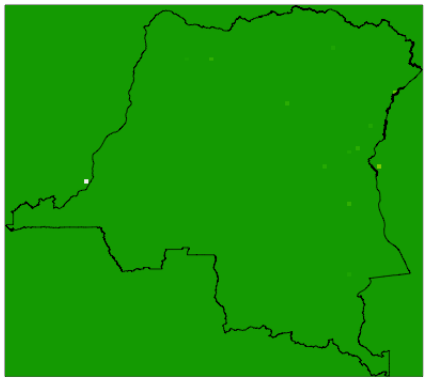

Figure 11: Density Estimate for All Conflicts Using LSCV-R Bandwidth

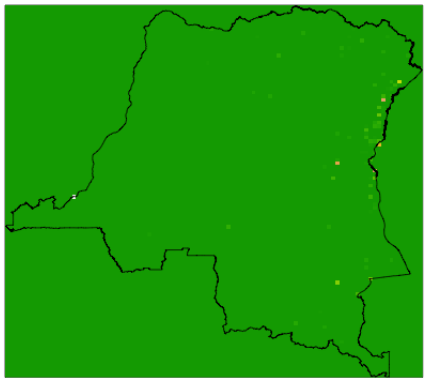

Figure 12: Density Estimate for All Conflicts Using BCV Bandwidth

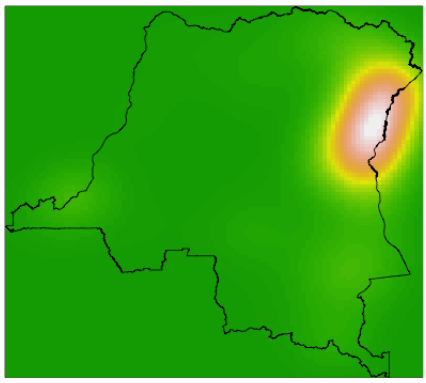

Figure 13: Surveillance Plot for Battles

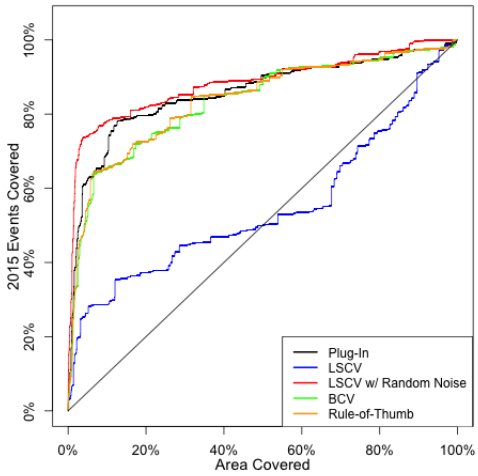

Figure 14: Density Estimate for Battles Using Rule-of-Thumb Bandwidth

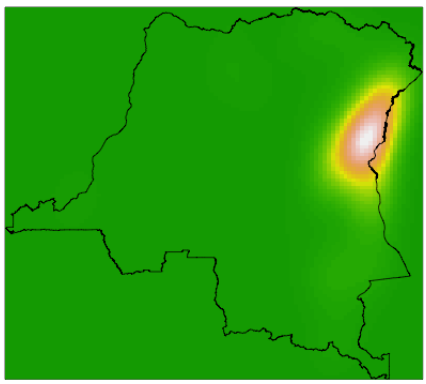

Figure 15: Density Estimate for Battles Using Plug-In Bandwidth

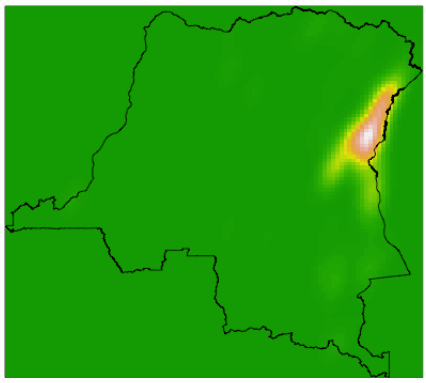

Figure 16: Density Estimate for Battles Using LSCV Bandwidth

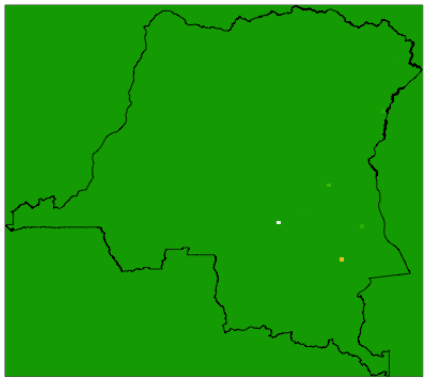

Figure 17: Density Estimate for Battles Using LSCV-R Bandwidth

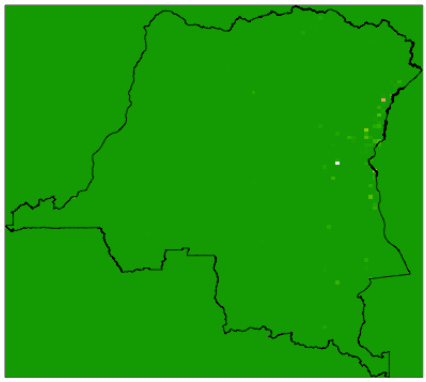

Figure 18: Density Estimate for Battles Using BCV Bandwidth

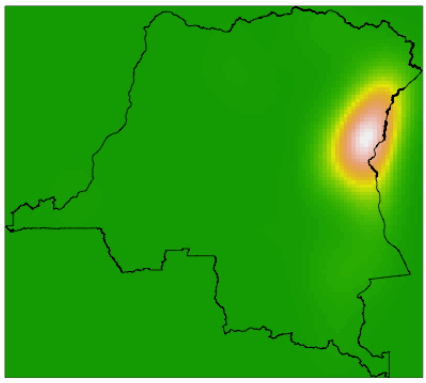

Figure 19: Surveillance Plot for Violence Against Civilians

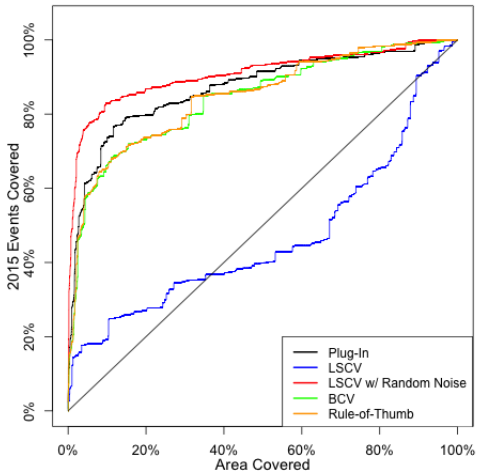

Figure 20: Density Estimate for Violence Against Civilians Using Rule-of-Thumb Bandwidth

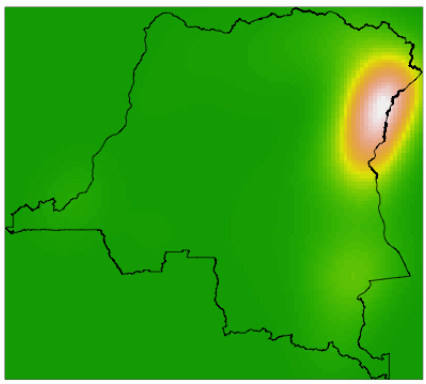

Figure 21: Density Estimate for Violence Against Civilians Using Plug-In Bandwidth

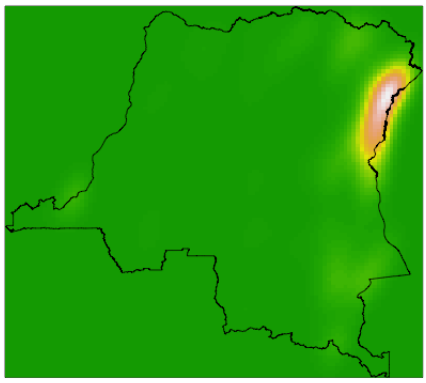

Figure 22: Density Estimate for Violence Against Civilians Using LSCV Bandwidth

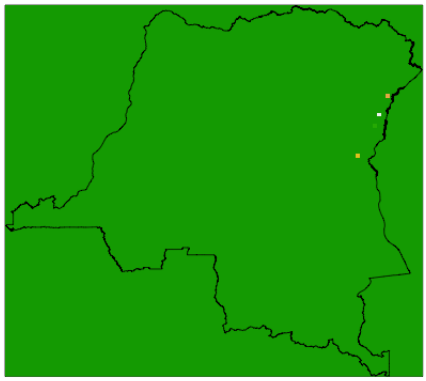

Figure 23: Density Estimate for Violence Against Civilians Using LSCV-R Bandwidth

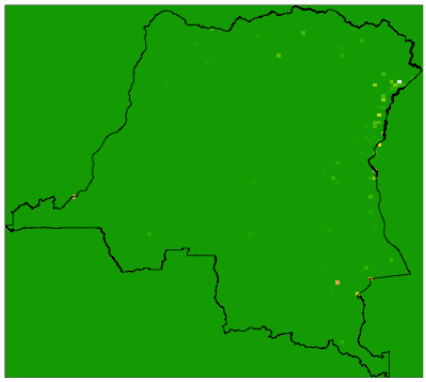

Figure 24: Density Estimate for Violence Against Civilians Using BCV Bandwidth

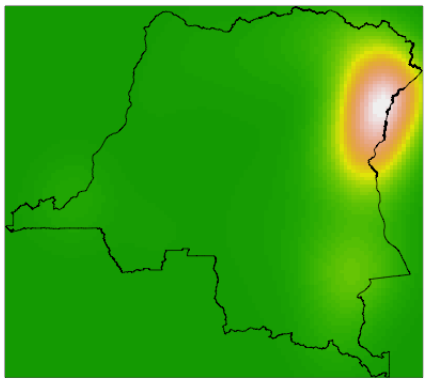

Figure 25: Surveillance Plot for Riots/Protests

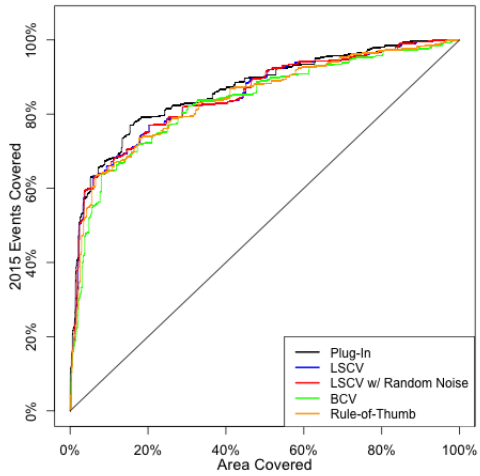

Figure 26: Density Estimate for Riots/Protests Using Rule-of-Thumb Bandwidth

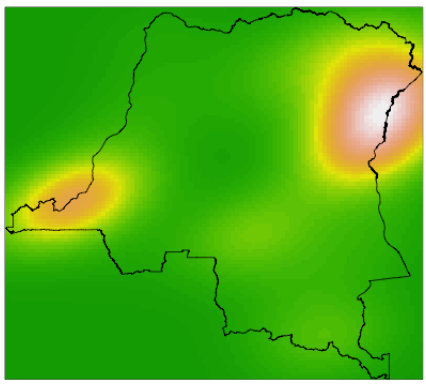

Figure 27: Density Estimate for Riots/Protests Using Plug-In Bandwidth

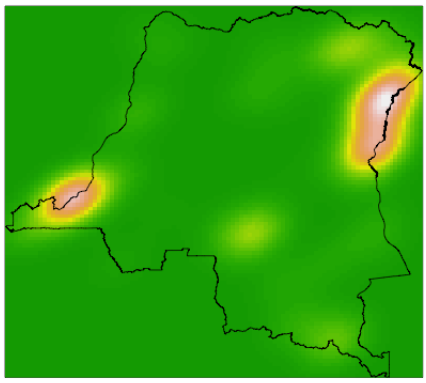

Figure 28: Density Estimate for Riots/Protests Using LSCV Bandwidth

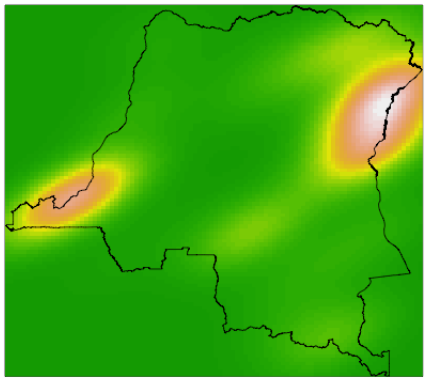

Figure 29: Density Estimate for Riots/Protests Using LSCV-R Bandwidth

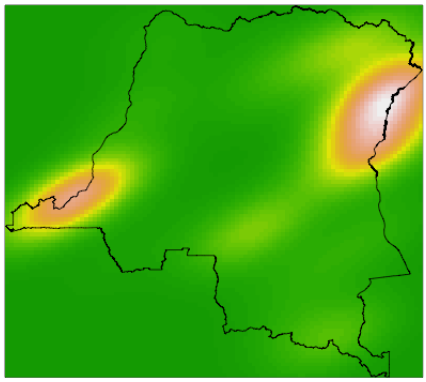

Figure 30: Density Estimate for Riots/Protests Using BCV Bandwidth

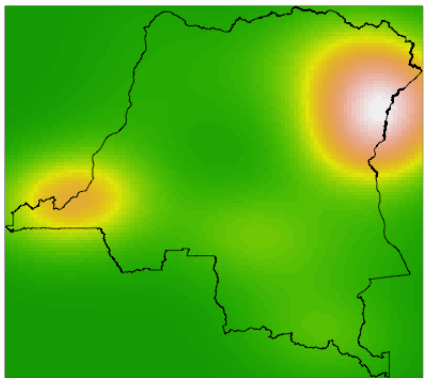

Table 1: Observation Count, By Event

Event Type	2014 Count	2015 Count
Battle	364	290
Riots/Protests	219	238
Violence Against Civilians	377	249
Total	**960**	**777**

Table 2:Monthly Event Counts, 2014-2015

Month	Battles	Riots/Protests	Violence Against Civilians	Total
January 2014	34	14	20	68
February 2014	30	7	28	65
March 2014	17	2	36	55
April 2014	32	9	30	71
May 2014	44	23	41	108
June 2014	39	10	36	85
July 2014	41	33	44	118
August 2014	47	10	49	106
September 2014	21	40	21	82
October 2014	25	29	36	90
November 2014	22	19	14	55
December 2014	12	23	22	57
Total: 2014	**364**	**219**	**377**	**960**
January 2015	16	29	14	59
February 2015	28	13	16	57
March 2015	28	15	32	75
April 2015	24	19	23	66
May 2015	22	55	30	107
June 2015	38	16	14	68
July 2015	15	8	20	43
August 2015	24	11	14	49
September 2015	20	14	20	54
October 2015	27	17	24	68
November 2015	28	28	27	83
December 2015	20	13	15	48
Total: 2015	**290**	**238**	**249**	**777**

Table 3: Fatality Count, By Event[10]

Event Type	2014 Count	2015 Count
Battle	611	736
Riots/Protests	22	52
Violence Against Civilians	597	956
Total	**1,230**	**1,744**

[10] Note that these are exclusively lives lost due to immediate, physical violence. These figures do not account for the indirect effects of the conflicts, such as poverty and lack of basic public services, which result in far more deaths each year.

Table 4: AUC Summary

Event Type	Rule-of-Thumb	Plug-In	LSCV	LSCV-R	BCV
All Conflicts	0.8598	0.8950	0.5644	0.9402	0.8546
Battles	0.8371	0.8570	0.5462	0.8839	0.8336
Violence Against Civilians	0.8461	0.8700	0.4665	0.9081	0.8418
Riots/Protests	0.8408	0.8650	0.8510	0.8510	0.8331