Working Paper Series: Study on Artificial Societies, No.48

The Insurgent Disease? Simulating the Geography of Insurgent Violence*

Gaku Ito†

May 12, 2015

Abstract

There is now a near consensus among students of civil war that violent incidents in civil war clusters in space. However, the underlying generating mechanisms remain disputed. Two primary explanations have been proposed to explain spatial patterns of insurgent violence: first, clusters of insurgent violence stem from contagious nature of insurgent violence; and second, clusters of violence simply mirror a similar distribution of structural factors that attract insurgents. This paper aims to address this question using an agent-based model incorporated with geo-referenced data of Afghanistan. This data-driven computational approach allows for nuanced specification of the micro-level mechanisms of insurgent behavior underlying the observed macro-level spatial patterns of violence. The computational model demonstrates that the observed patterns of insurgent violence in Afghanistan are consistent with a simple mechanism of relocation diffusion constrained by the exogenous conditions. The in-sample predictive performance of the computational model underscores its internal validity. Moreover, the findings are found to be robust to changes in model's modifications.

*All reported analyses were conducted using artisoc 3.0.1 for Mac OS X and R for Mac OS X 3.1.1. Replication codes and data are available upon request. I am grateful to Zorzeta Bakaki, Karsten Donnay, Kaisa Hinkkainen, Atsushi Ishida, Katsuma Mitsutsuji, Heinrich Nax, Susumu Yamakage, and seminar participants at the University of Tokyo and Kobe University for careful reading and valuable suggestions for earlier versions of this paper. Financial support from the Japan Society for the Promotion of Science (Grant-in-Aid for Scientific Research, KAKENHI, Grant # 24243023 and Grant-in-Aid for JSPS Fellows, Grant # 245785) is gratefully acknowledged.

†Ph.D Candidate, Graduate School of Arts and Sciences, University of Tokyo, 3–8–1 Komaba, Meguro-ku, Tokyo 153–8902; Project Academic Support Staff, Graduate School of Medicine, University of Tokyo, 7–3–1 Hongo, Bunkyo-ku, Tokyo 113–8654. Email: gakuatgu@gmail.com.

[O]nce conflict begins, there is some tendency for it to spread out from the "infected" spot.

Norman Z. Alcock (1972), 64.

1 Introduction

Why does insurgent violence in civil war diffuse and cluster in specific locations? Scholars have longly acknowledged that inter- and intra-state conflicts spread and cluster in space (Alcock, 1972; Buhaug and Gleditsch, 2008; Danneman and Ritter, 2013; Iqbal and Starr, 2008; Most and Starr, 1980; Weidmann, 2015). As the research agenda of disaggregating civil war progresses in recent years, scholars have increasingly explored the micro and subnational variations of violence taking place in the context of civil war (Buhaug and Rød, 2006; Kalyvas, 2006, 2008). Among substantial insights, there is now a near consensus within civil war research that disaggregated violence in civil war as well as aggregated conflicts cluster in a manner similar to the spread of a disease (e.g., Braithwaite and Johnson, 2012, 2015; Linke et al., 2012, Forthcoming; O'Loughlin et al., 2010a,b; O'Loughlin and Witmer, 2012; Schutte and Weidmann, 2011; Weidmann and Ward, 2010; Zammit-Mangion et al., 2012).

At the first glance, spatial clustering of violence seems to indicate the existence of diffusion or contagion processes. The subnational risks of violence depend not only on structural attributes of localities that are largely exogenous to conflict processes but also on endogenous, contagion-like processes of insurgent activities. Intuitively, one answer emerges: insurgent violence clusters because thier activities are spatially and/or temporally contagious. Violence itself can alter the prospects for further violence.

However, there are no largely agreed explanations for the micro-mechanisms underlying the observed patterns of violence. Most fundamentally, the answer is not apparent because clusters of violence are also consistent with another, noncontagious mechanism: clusters of insurgent violence need not stem from contagion at all but can result from heterogeneity in the intrinsic tendency of subnational localities to host violence. Put otherwise, clusters of insurgent violence emerge from a similar distribution of violence-facilitating attributes. Although these two mechanisms are conceptually and theoretically distinct, their expressions in empirical data are often observationally indistinguishable, making it extremely difficult to identify the micro-level mechanisms underlying the observed patterns.

Indeed, existing studies tend to merely *control* for temporal and/or spatial dependences in their econometric models rather than *revealing* the underlying generating mechanisms (Schutte and Weidmann, 2011; Zhukov, 2012). Perhaps the most common empirical approach is employing spatial econometric models which incorporates spatially-lagged dependent or independent variables to control and measure the potential diffusion effects. However, although the spatially explicit econometric models allow for controlling and characterizing the spatial nature of the observed data, the oft-employed spatial lags alone do not suffice for examining the generating mechanisms. Consequently, it remains unclear what diffusion process, if any, is at work within insurgent activities (Braithwaite and Johnson, 2015; Linke et al., Forthcoming; Schutte and Weidmann, 2011; Zhukov, 2012).

This paper aims to address this gap drawing on the fine-grained geo-referenced data of violence in the ongoing war in Afghanistan. We first briefly describe the spatial footprints of insurgent violence. Second, we consider two primary explanations for spatial patterns of insurgent violence. Third, we develop a simple agent-based model incorporated with geo-referenced data of Afghanistan and examine the likely determinants and micro-level mechanisms underlying the observed patterns of insurgent violence. This data-driven computational approach enables not only clear specification of competing micro-mechanisms but also making a tight link between hypothesized model and a specific empirical case (Bhavnani et al., 2014; Ito and Yamakage, 2014; Lim et al., 2007; Weidmann and Salehyan, 2013).

In the broader context of civil war literature, this paper joins a growing interest in endogenous processes of civil war. Much of the conventional civil war literature focuses on exogenous determinants of civil war onset, duration, and termination such as geography, economy, ethnic groups, and natural resources. Consequently, the endogenous processes and mechanism through which such factors shape behavior of warring actors, and/or violence feeds itself relatively under-explored (Balcells and Kalyvas, 2014; Kalyvas, 2006; Schutte and Weidmann, 2011; Zhukov, 2012). This gap within the literature has recently facilitated scholars to investigate endogenous conflict processes. Scholars have increasingly acknowledged that civil war research remains incomplete without examination of endogenous dynamics of violence which in turn shape severity, duration, and outcome of conflicts. Taking an approach of empirically explicit computational modeling (Bhavnani et al., 2014; Lim et al., 2007; Weidmann and Salehyan, 2013), we explore the micro-level mechanisms of insurgent violence with a tight link to the specific case of Afghanistan.

The reminder of this paper proceeds as follows. In the next section, we provide a brief overview of the war in Afghanistan and examine the spatial patterns of insurgent violence. Section 3 reviews possible propositions on the clusters of violence in civil war. We then present evidence-driven computational approach and develop a simple agent-based model incorporated with empirical data in Section 4. After seeding the model with fine-grained spatial data from Afghanistan, we validate the computational model using the empirical records of insurgent violence, thereby examining the veracity of the propositions and explanatory power of the model in Sections 5 and 6. The model demonstrates that the observed patterns of insurgent violence in Afghanistan are consistent with a simple mechanisms of insurgent attack and diffusion constrained by the exogenous conditions. The model also yields a fair in-sample predictive performance. Robustness checks in Section 7 ensure that the simulations results are unlikely to be products of arbitrary assumptions. While exogenous factors substantially shape the insurgents' behavior, endogenous diffusion dynamics are likely to be at work, indicating the contagious nature of insurgent violence.

2 Insurgent Violence in Afghanistan

We use the ongoing irregular war in Afghanistan as our case to explore the micro-level mechanisms of insurgent violence in civil war. Following the overthrown of the Taliban regime in 2001 and the external occupation, the Taliban leader Mullah Muhammad Omar vowed to "retake control of Afghanistan" in 2004 (Gall, 2004). The Taliban remnants gradually launched the insurgency against the U.S.-led coalition and the Afghan government forces (Farrell and Giustozzi, 2013). Despite the heavy losses and attrition that the Taliban have suffered and the U.S.-led troop "surge," a massive increase of coalition troops, the counterinsurgency (COIN) campaign is not yet completed as the insurgents pose ever-present threat to the order and security in Afghanistan (Farrell and Giustozzi, 2013; Johnson and DuPee, 2012; Johnson and Mason, 2008).

2.1 Dataset

The following empirical analysis relies on the dataset commonly used in the study of violence in Afghanistan: The U.S. military internal database called "Significant Activities" (SIGACTs). The SIGACTs are the collection of short summary on events that include actors involved, caused casualties, event type, locations, timing, and other related information that have been recorded by individual troops in the course of operations. Most notably, records on locations are accurate at the 1km-level, offering researchers an accurate picture of events on the ground. A subset of the event data that has originally been released as the "Afghan War Diary" (AWD) by WikiLeaks.org in 2010 and is now also available to the public from several news outlets and peer-reviewed journals as a part of replication materials.[1] Although we deeply recognize the potential risks of empirical analysis drawing on the

[1] Although the SIGACTs database offers a rare and noteworthy opportunity for researchers to uncover the trends and patterns of insurgent violence in Afghanistan, it might suffer from potential bias. Two of them are worth noting here (Donnay and Filimonov, 2014; Weidmann, 2013, Forthcomingb). First, there may be a tendency that the military troops under-report collateral damage and civilian casualties caused by their operations. However, this bias is not likely to cause serious problem in the analysis here, since the main focus

data, the already widespread use of the dataset in the academic community leads us to the recognition that empirical analyses relying on the dataset are not likely to harm or endanger government officials, individuals, or institutions involved (e.g., Donnay and Filimonov, 2014; O'Loughlin et al., 2010a; Weidmann, 2013, Forthcomingb; Zammit-Mangion et al., 2012).

2.2 Spatial Patterns of Insurgent Violence

The SIGACTs database covers both violent (e.g., IED explosions) and nonviolent (e.g., information provision from civilians) incidents across the country during the period from January 2004 through December 2009, with $76,910$ entries. Rather than including both violent and nonviolent incidents into the analysis, we opt to focus on the violent incidents from the dataset.[2] This filtering provides us $52,196$ incidents: $45,628$ attacks are coded as insurgent violence and the remaining $6,568$ attacks are coded as ISAF violence, leaving the remaining $24,007$ nonviolent incidents excluded from the analysis.

To clarify the spatial patterns of insurgent violence, we clipped individual insurgent attacks to the closest population settlements based on their geo-coordinates, aggregating the attacks into the settlement-level ($N_{settlement} = 37,484$).[3] During the period covered by the

of this paper is the numbers of insurgent violence rather than casualties. Second, the reporting standards for SIGACTs may have changed over time and the reporting procedure may vary across units, possibly resulting in a significant measurement error. This concern is partly alleviated by focusing on the temporally aggregated spatial distribution of insurgent violence, rather than the temporal patterns of insurgent violence.

[2]The SIGACTs database contains discrete types of incidents that are relevant for the purpose of this paper. Drawing the "Category" column, we selected the following for insurgent violence, $Violence_{INS}$: "Other (Hostile Action)," "Assassination," "Attack," "Direct Fire," "IED Explosion," "IED False," "IED Found/-Cleared," "IED Hoax," "Indirect Fire," "Mine Found/Cleared," "Mine Strike," "SAFIRE" (Surface-to-Air Fire), "Security Breach," "Unexploded Ordnance," "Sniper Ops." For ISAF attacks, $Violence_{ISAF}$, we selected the following: "Cache Found/Cleared," "Close Air Support," "Counter Insurgency," "Counter Terrorism," "Direct Fire," "Escalation of Force," "Indirect Fire," "Search and Attack," "Show of Force", "Small Unit Actions," "Sniper Ops," "Other Offensive," "Raid." For "Direct Fire" and "Sniper Ops" categories, we match the subset of the data against "Affiliation" variable that contains information of the perpetrator ("FRIEND," "ENEMY," "NEUTRAL," "UNKNOWN"), and code those with "Affiliation"="FRIEND" as ISAF attacks and those with "Affiliation"="ENEMY" as insurgent attacks.

[3]This minimal spatial aggregation is both necessary and widely used in micro-studies of civil war violence (e.g., Hirose et al., 2014; Lyall, 2009, 2014; Zhukov, 2012). Location and population data of individual settlements are taken from USAID MISTI/Humanitarian Response. "Afghanistan: Settlements (villages, towns, cities)," March 2012 – June 2013 (https://www.humanitarianresponse.info/operations/afghanistan/dataset/afghanistan-settlements-villages-towns-cities-0, accessed July 25, 2014). The maximum (minimum)

dataset, $7,644$ settlements out of $37,484$ (20.4%) have experienced one or more insurgent violence. Figure 1 maps the resultant spatial distribution of population settlements with (red dots) and without (gray dots) insurgent attacks.

The "disease map" provides us visual evidence of conflict clustering, and a formal spatial statistical test confirms it. Because individual incidents are aggregated to the closest settlements (discrete locations), the (global) Moran's I static, which measures the spatial correlations between fixed spatial units and each of their neighbors, is suitable here.[4] Figure 2 shows the Moran's I static estimates for (logged) number of insurgent violence across a sequence of distance band neighbor pairs (panel a) and temporal variation of the static during the period between 2004 and 2009 using a half-year temporal window (panel b). For all spatial and temporal windows, the null hypotheses of the nonexistence of spatial autocorrelation were rejected at 5% level. While the levels of spatial autocorrelation decay as the distance band increases, the consistently positive estimates suggest that population settlements with (without) insurgent violence are tend to be located close to violent (peaceful) settlements, indicating the spatial concentration of insurgent violence. Figure 2 (b) indicates an upward trend in the Moran's I estimates, suggesting that insurgent activities show a stronger clustering pattern as the war unfolds.

While this preliminary analysis confirms that insurgent violence significantly clusters in space, it raises the question of why such spatial patterns ever emerge. We will examine the micro-mechanisms underlying the observed spatial patterns in the following sections.

---

distance between coordinates of individual violent incidents and the those of the corresponding nearest settlements is 84.297km (0.002km), and the mean (median) distance is 1.364km (0.767km). Introducing the cut-off distance does not substantially alter the spatial distribution.

[4]The Moran's I static, which is analogous to the standard Pearson's correlation measure, is commonly used to characterize spatial autocorrelations and formally defined as:

$$I = \frac{n}{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \left(x_i - \bar{x}\right) \left(x_j - \bar{x}\right)}{\sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2},$$

where n is the number of spatial units (observations), x_i is the variable of interest, and w_{ij} is a spatial weight which defines the spatial influence or relationship between units i and j.

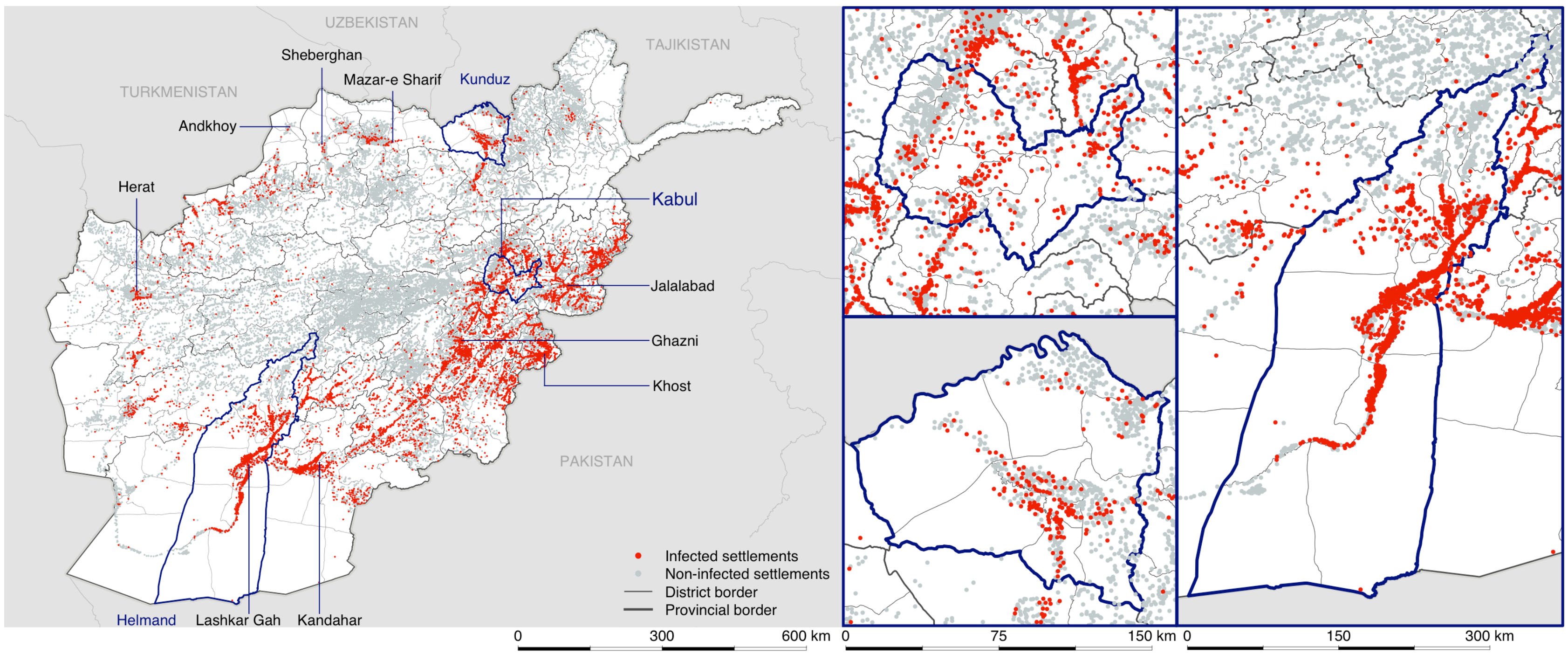
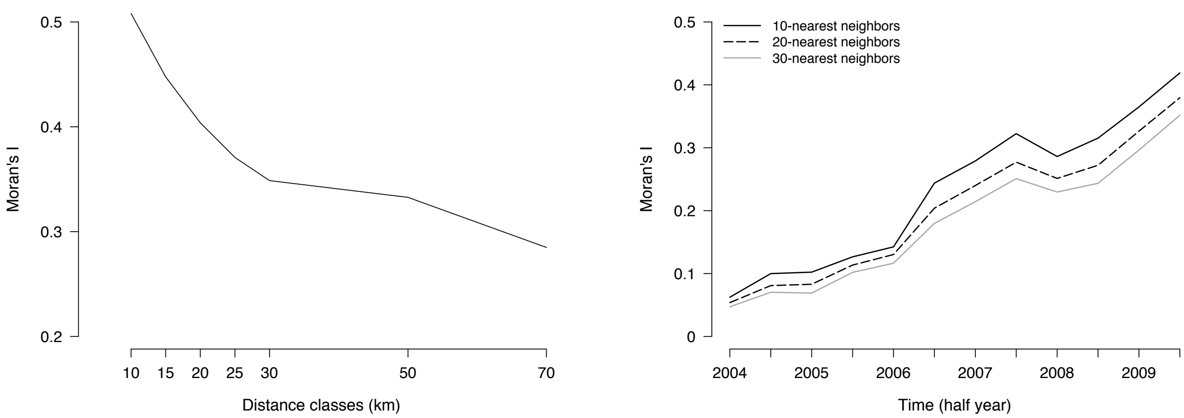


Figure 1: Spatial distribution of insurgent violence in Afghanistan, 2004–2009 ($N^{INS}_{violence} = 45,628$, $N_{settlement} = 37,484$). We clipped individual insurgent attacks to the nearest settlement. Gray dots (●) indicate those settlements that have experienced no insurgent violence, whereas red dots (●) indicate those that have experienced more than one insurgent violence. The solid lines represent district, province, and international borders.



(a) Spatial Window (b) Temporal Window

Figure 2: Spatial autocorrelations in insurgent violence. Panel (a) shows values of Moran's I for a sequence of distance band neighbor pairs (temporally aggregated data, 2004–2009). Panel (b) shows values of Moran's I for insurgent violence within half-year temporal windows varying the neighborhood definition: 10- (black), 20- (dashed), 30-nearest settlements (gray). All estimates are significant at 5% level.

3 Propositions on Violence Clustering

Why does insurgent violence diffuse and cluster? Scholars have proposed two primary explanations how clusters of violence may emerge (Braithwaite and Johnson, 2015; Buhaug and Gleditsch, 2008; Schutte and Weidmann, 2011). The first proposition suggests that clusters of insurgent violence stem from the violence-facilitating, static factors which they themselves cluster such as population size and geographic conditions. The second proposition, on the other hand, argues that endogenous conflict dynamics as well as exogenous factors shape the specific course of insurgent violence; occurrence of violence alters prospects of future violence in the same location and nearby locations.

3.1 Structural Conditions

A large and well-established body of aggregated and disaggregated civil war literature has demonstrated that country- and subnational-level structural factors such as population size, economic development, and rough terrain associate to risks of civil war and insurgent violence (e.g., Buhaug, 2010; Buhaug and Rød, 2006; Collier and Hoeffler, 2004; Fearon and Laitin, 2003; Raleigh and Hegre, 2009; Toft, 2003; Weidmann, 2009). As insurgent activities are not self-sufficient but highly constrained by pre-existing structural conditions, a series of exogenous, structural factors are likely to shape the spatial distributions of violence by influencing insurgents' willingness and opportunity to engage violence (Braithwaite and Johnson, 2015, 115–117; Zhukov, 2012, 145–146).

This first approach, which is sometimes referred to as "confusion" or "risk heterogeneity" hypothesis (Braithwaite and Johnson, 2015; Buhaug and Gleditsch, 2008; Johnson, 2008), states that clusters of insurgent violence may simply mirror a similar distribution of time-invariant local conditions (Buhaug and Gates, 2002; Buhaug and Gleditsch, 2008; Buhaug and Rød, 2006). This view can be stated as following:

Hypothesis 1 *The risk of insurgent violence is associated with a set of local structural factors that are independent of one another.*

There are two classes of prime structural covariates of insurgent violence: socioeconomic and geographical factors. We will introduce each class in the following.

Socioeconomic Factors. Major population centers are considered to be a leading predictor of violence since they provide insurgents not merely a large pool of recruitments and physical targets but also "sea" which shelters insurgent "fishes" from incumbent forces (Buhaug and Rød, 2006; Collier and Hoeffler, 2004; Fearon and Laitin, 2003; Mao, 1961; Raleigh and Hegre, 2009; Tollefsen and Buhaug, 2015). The promise of media visibility might also attract insurgents who seek to signal their resolve and capability to their opponents and the public.

Another but related correlate of civil war violence which spatially clusters is local ethnic configuration. Geographically concentrated ethnic groups are hypothesized to consider their territory as their "homeland" and might motivate the population to fight for it; alternatively, geographically concentrated groups are considered to have better social networks that insurgents can use to mobilize the population and fewer difficulties in overcoming the collective action problem (e.g., Toft, 2003; Weidmann, 2009). As the Taliban is commonly characterized as a Pashtun-based insurgent movement (Farrell and Giustozzi, 2013; Johnson and Mason, 2008), the local configuration of Pashtun is another primary predictor of violence in the particular context of Afghanistan.

Spatial variation in levels of economic development is also deemed as a robust predictor of insurgent violence (Berman et al., 2011; Buhaug et al., 2011; Hegre et al., 2009; Østby et al., 2009). Low level of income has been hypothesized to facilitate insurgent activities by providing motivation for rebellion and/or lowering opportunity costs of participation to insurgent movements (Collier and Hoeffler, 2004; Grossman, 1991). On the other hand, strategic considerations might lead insurgents to target wealthier areas to attract public attentions or financial gains. Although there is no widely shared consensus on the causal effects and mechanisms through which it shapes the geography of insurgency, the civil war literature considers local economy to be a primary factors of civil war violence.

Geographic Factors. Geographic conditions also shape how insurgent activities unfold. Inaccessible terrain tends to inhibit the states' reach and thereby create favorable conditions for insurgents to survive (Buhaug, 2010; Buhaug and Rød, 2006; Fearon and Laitin, 2003; Herbst, 2000; McColl, 1969; O'Loughlin and Witmer, 2012; Schutte, 2014; Tollefsen and Buhaug, 2015). As Fearon and Laitin (2003) argue, the essential condition for nascent insurgents' survive is the state's reach into the local areas: as insurgents are often militarily weaker than the incumbents, they are simply "better able to survive and prosper if the government and military they oppose are relatively weak" (80).

One of the leading geographic factors that shape the local balance of power is road networks. Because roads are essential to the projection of power, poorly served road networks have been considered to inhibit capabilities and the reach of incumbents into rural areas, thereby providing opportunities for insurgents' survive and activities (Buhaug, 2010; Buhaug and Rød, 2006; Fearon and Laitin, 2003; Herbst, 2000; Tollefsen and Buhaug, 2015). The former commander of NATO troops in Afghanistan, Lieutenant General Karl Eikenberry, eloquently summarized and contextualized this notion in the current policy debate on the war in Afghanistan: "[w]here the road ends, the Taliban begins" (quoted in Gruber, 2007). Indeed, the United States and the international community have invested heavily in rehabilitating and building Afghan infrastructure over the past decade. Underlying this policy is the notion that underdeveloped roads isolate villages from "basic government services, even police or military protection," which in turn generate favorable conditions for insurgency (USAID, 2014); rehabilitating and paving roads helps extend the central government's reach, thereby bringing peace and economic prosperity across the country (Amiri, 2013).

On the other hand, scholars have also hypothesized that densely served road networks facilitate insurgent activities because they ease the logistical constraints that insurgents face and provide a pool of potential targets for attack while reducing the costs of incumbent operations (Amiri, 2013; O'Loughlin et al., 2010b; Raleigh and Hegre, 2009; Zhukov, 2012). Consequently, easily accessible areas with well served road networks are expected to facilitate insurgent activities rather than containing them. In a sharp contrast with Et. Gen. Eikenberry, an anonymous senior Afghan commander recently expressed this view: "the road is a disaster. It causes obstacles and delays and countless casualties" (quoted in Sieff, 2014). While these two arguments disagree on the direction of the causal effects, they generally agree on the correlation between accessibility to roads and risks of insurgent violence.

Related arguments focus on proximities to the capital and borders. Because a distance from the center of state power as well as poorly served road networks hinges the state's reach, insurgents survive more easily in remote or peripheral areas.[5] In the particular context of Afghanistan, proximities to the Afghanistan-Pakistan border is often deemed to be a prime predictor of insurgent violence due to the presence of the Taliban's "safe heavens" in

[5]Raleigh and Hegre (2009) examines the impact of the distance from the capital in a nuanced way and argues that conflict events cluster in population centers that are distant from the capital.

the Federally Administered Tribal Areas (FATA) of Pakistan (Farrell and Giustozzi, 2013; Johnson and Mason, 2008; O'Loughlin et al., 2010a,b).

3.2 Contagious Nature of Insurgent Violence

There is also compelling evidence that insurgent violence is in fact contagious, suggesting that violence begets violence in the same and/or nearby localities (e.g., Braithwaite and Johnson, 2012; Johnson et al., 2011; Linke et al., 2012; O'Loughlin et al., 2010a,b; O'Loughlin and Witmer, 2012; Schutte and Weidmann, 2011; Zhukov, 2012). Reflecting on such insights, the second category of hypotheses states that violence begets violence in time and/or space.

Hypothesis 2 *The risk of insurgent violence is not simply the product of the local structural factors but is temporally and spatially dependent on one another.*

This hypothesis argues that exogenous structural factors alone are unlikely to explain the spatial patterns of insurgent violence and that the endogenous processes of diffusion processes should also shape how insurgent violence unfolds in the course of civil war; episodes of civil war violence result from previous fighting in the same or nearby locations.

Researchers distinguish four general classes of spatial patterns of violence (Baudains et al., 2013, 214–215; Schutte and Weidmann, 2011, 144–146; Zhukov, 2012, 146–147). First, the dynamics where occurrence of violence spreads to previously peaceful locations while originating place continues to experience violence is characterized as an *expansion* diffusion; violence simply begets violence in both time and space. Second, an *escalation* diffusion or *hot spots* is likely to be produced when violence continues to occur in the originating place while not spreading to other locations. Third, a *relocation* diffusion is a process where occurrence of violence does not facilitate further violence in the originating location while facilitating violence in nearby locations, such that violence "travels" from the originating location to previously peaceful locations. Finally, isolated *flashpoints* of violence appear when short-lived sequence of violence characterizes the conflict episodes.

While existing studies have found empirical evidence of subnational diffusion of civil war violence, the exact nature of diffusion patterns and the underlying generating mechanisms remain disputed. For example, Schutte and Weidmann (2011) explored the diffusion patterns of civil war violence in Bosnia-Heregovina, Burundi, Kosovo, and Rwanda and found that diffusion patterns in irregular civil wars tend to be primarily the escalation diffusion whereas relocation diffusion is predominant in conventional civil wars with clear front line. In contrast, Zhukov (2012) examined the role of road networks in the diffusion of conflict events in the irregular civil war the North Caucasus and found that insurgent violence tends to relocate along the road networks, reflecting on logistical constraints insurgents face on a daily basis. O'Loughlin and Witmer (2012) also scrutinized the violence diffusion in the North Caucasian conflicts and found properties of expansion diffusion, rather than relocation or escalation diffusion alone, although the latter type of diffusion in the temporal dimension is more common. Based on a preliminary analysis drawing on the WikiLeaks AWD, O'Loughlin et al. (2010a) argue that the insurgency in Afghanistan exhibits a expansion diffusion where violence spreads to previously unaffected regions while violence continues to occur in the Eastern and Southern border regions (cf. O'Loughlin et al., 2010b).

We aim to inform the debates between the two distinct approaches and over nature of violence diffusion in civil war by examining the sufficient micro-level mechanisms that generate the patterns of insurgent violence similar to the observed one, using the approach of agent-based modeling incorporated with fine-grained empirical data.

4 Model

We develop an agent-based model to evaluate the plausibility of the mechanisms discussed above. Agent-based modeling specifies hypothesized mechanisms that govern behavior and interactions of constituent elements of a system called agents as a computer program. This technique enables to systematically explore whether and why specific micro-level mechanisms generate a class of macro-level outcomes via computational experiments (Axelrod, 1997).

Another strength of agent-based modeling lies in its flexibility to incorporate with empirical data, which allows for seeding and optimizing hypothesized models using empirical data (Ito and Yamakage, 2014). The empirically-explicit computational modeling approach helps not only contextualize abstract models into the specific case of interest but also validate hypothesized micro-level mechanisms by comparing the model-generated outputs with the empirical records. For example, Lim et al. (2007) and Weidmann and Salehyan (2013) developed agent-based models incorporated with spatial data of ethnic geography and analyzed the patterns of violence and ethnic segregation in the former Yugoslavia, India, and Iraq, and successfully demonstrated that a simple mechanism of ethnically and/or security motivated migration and subsequent violence accounts for the spatial distributions of violence in actual conflicts. Bhavnani et al. (2014) used an agent-based model to not only explain the observed patterns of communal violence but also assess how different levels and forms of Israeli-Palestinian segregation would shape future violence in Jerusalem.

In line with these pioneering attempts, we rely on the approach of data-driven computational modeling to explore the micro-level mechanisms underlying the observed patterns of insurgent violence in Afghanistan. In the following subsections, we first describe the structure of the model and then specify the logic of insurgent violence.

4.1 Model Space and Initial Configurations

The model space consists of a set of N population settlements S_i resided by M insurgent agents I_j, with $\boldsymbol{S} = \{S_1, \ldots, S_N\}$ denoting the set of settlements and $\boldsymbol{I} = \{I_1, \ldots, I_M\}$ denoting the set of agents. Model parameter N is set to $N_{settlement} = 37,484$, and settlements S_i are located according to the corresponding geo-coordinates. For a given settlement S_i, we refer to the set of its neighbor settlements as $\mathcal{N}_i$. Insurgent agents I_j are randomly distributed to individual settlements S_i at the beginning of a simulation run. At every discrete time period $t \in [1, t_{\max}]$, insurgent agent I_j makes a binary decision whether to conduct an attack in its current location S_i, or relocate to another settlement $S_j \in \mathcal{N}_i$.[6] Once all agents have made decisions, the model reports the number of attacks conducted in individual settlements and then proceeds to time $t+1$. Each simulation run reports a vector of cumulative numbers of insurgent attacks that have occurred in individual settlements, $\hat{\boldsymbol{Y}} = (\hat{Y}_1, \ldots, \hat{Y}_N)$, which can be directly compared with the empirical records, $\boldsymbol{Y} = (Y_1, \ldots, Y_N)$.

We define the neighborhood structure of the model, an $N \times N$ spatial weight matrix (SWM) $\boldsymbol{W}$, as a distance-weighted k-nearest neighbor (*dwk*NN) sparse matrix in which the diagonal elements $w_{ii} = 0$ and non-diagonal elements $w^S_{ij} \geq 0$ capture the relative degree of (spatial) influence of settlement S_j on settlement S_i.[7] We first construct a 20-nearest neighborhood matrix with $37,484 \times 20 = 749,680$ non-zero entries w^S_{ij}, in which twenty geographically nearest settlements S_j are defined as neighbors of S_i, or $S_j \in \mathcal{N}_i$. We then compute the geodesic distances between neighbor settlements penalized by the additional distances to the nearest roads from individual settlements to specify the weight of settlement S_j for settlement S_i, w^S_{ij}. For simplicity, we opt to rely on the inverse-distance weighting (IDW) scheme: $w^S_{ij} = d^{-1}_{ij}$, where d_{ij} indicates the inter-settlement distance between settlements S_i and S_j penalized by settlement-to-road distances in kilometers.[8] Simply put, w^S_{ij} reflects the accessibility-weighted influence of neighbor S_j on S_i; i.e., a more accessible neighbor is more influential, and *vice versa*. Figure 3 maps the generated neighborhood network.

[6] The attach-or-relocate dichotomy is an arbitrary assumption for simplicity. For example, insurgents may alternatively decide to just stay and hide. We will examine the impacts of this assumption in Section 7.

[7] Zhukov (2012) suggests that insurgent activities are heavily constrained by the logistical factors, and a road distance matrix accurately represents such constraints. However, it is not computationally feasible to construct a computational model on a $37,484 \times 37,484$ origin-destination (OD) matrix with $1,405,050,256$ entries. Although the *dwk*NN scheme is based on an arbitrary selection of neighborhood size k, it allows for nuanced representation of the neighborhood structure at a relatively low computational cost.

[8] For example, if the inter-settlement geodesic distance between S_i and S_j is 30km and settlement-to-road distances are 5km and 2km, respectively, then $d_{ij} = 30 + 5 + 2 = 37$km, and $w^S_{ij} = 37^{-1} \sim 0.027$. The mean (median) value of d_{ij} is 9.431 (7.254) and the maximum (minimum) value is 176.3 (0.167) with the standard error of 7.186. Also note that spatial weight w^S_{ij} is rescaled to the range $[0, 1]$.

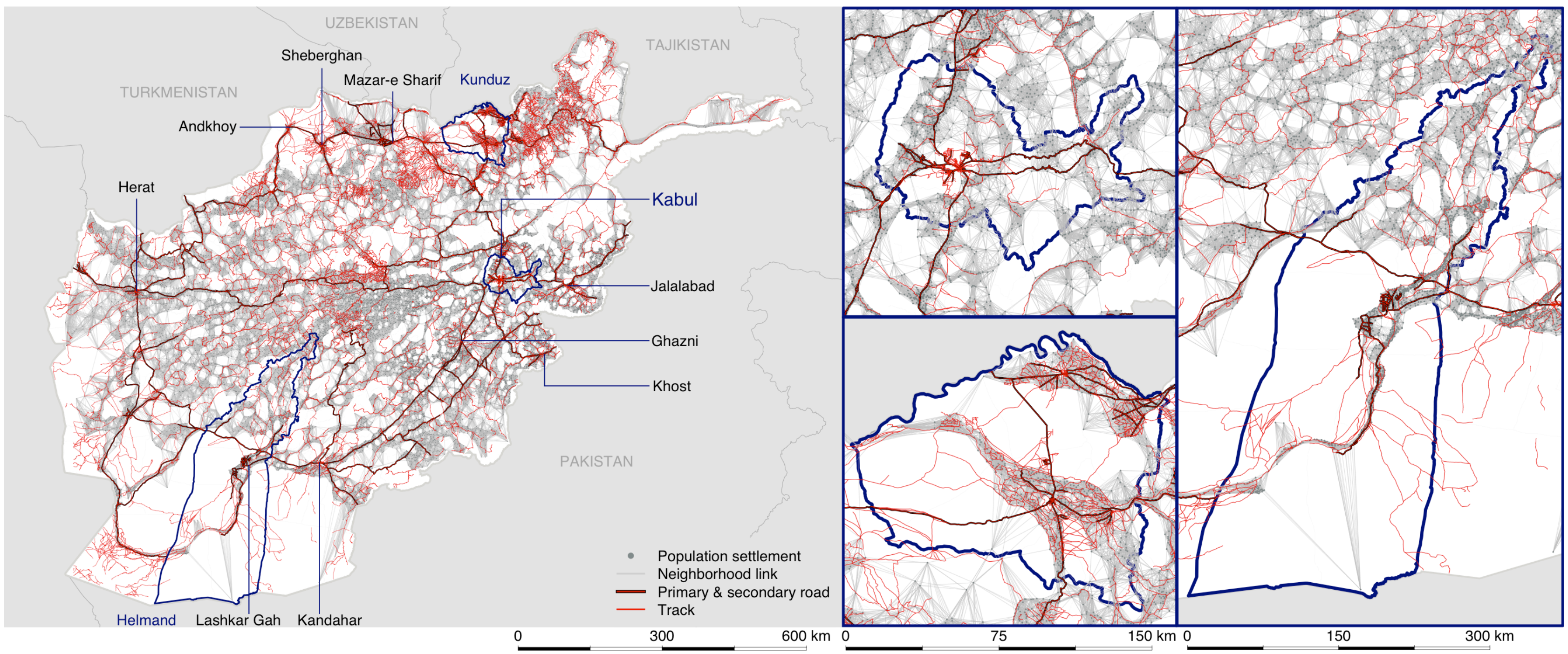


Figure 3: Spatial distribution of population settlements and neighborhood network. Gray dots (•) represent individual population settlements, whereas gray segments linking settlements indicate pairs of neighbor settlements. Red lines represent road networks.

4.2 Insurgent Behavior

The specification of insurgent behavior here is a generalization of the model in Weidmann and Salehyan (2013) who explored the likely determinants of sectarian violence in Baghdad (cf. Bhavnani et al., 2014; Lim et al., 2007). We generalize the model such that it incorporates not only static factors but also diffusion terms as discussed in detail below. Specifically, conditioning on the local environment of the current location S_i, the decision of insurgent agent I_j to carry out an attack at time period $t+1$, or $y_{ijt+1} = 1$, is assumed to be a realization of a Bernoulli process with probability

$$p_{ijt+1} \equiv P\left(y_{ijt+1} = 1 | \boldsymbol{x}_i, \boldsymbol{z}_{it}\right) = h(\alpha + \boldsymbol{x}_i^\top \boldsymbol{\beta} + \boldsymbol{z}_{it}^\top \boldsymbol{\gamma}), \tag{1}$$

where $h(x) = \exp(x)/(1 + \exp(x))$ is the inverse logit, and α denotes a time- and unit-invariant model parameter that determines the baseline probability of insurgent attacks.[9] Recall that decisions of insurgent agents in this model are assumed to be dichotomous: attack or relocate. At every time period t, insurgent I_j decides to carry out an attack in settlement S_i in the subsequent period $t+1$ with probability p_{ijt+1}; otherwise, I_j decides to relocate to a randomly chosen neighbor settlement $S_j \in \mathcal{N}_i$ with probability $1 - p_{ijt+1}$.

The intuition behind is that the probability of insurgent violence depends on two distinct classes of factors: first, the inherent and structural *susceptibility* of the settlements where insurgents are located, and second, temporal and spatial *contexts*. $\boldsymbol{x}_i$ is a vector of settlement-specific, time-invariant covariates (e.g., geographic factors) whereas $\boldsymbol{z}_{it}$ is a vector of time-variant covariates (e.g., local history of insurgent violence). $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are the vectors of the corresponding model parameters that represent the logics behind the propositions presented in the previous section. We call $\boldsymbol{x}$ susceptibility covariates and $\boldsymbol{z}$ diffusion covariates, and the

[9]While the probability of attacks is defined as a function of the proportion of co-ethnics in Weidmann and Salehyan's (2013) model (56), the proposed model incorporates $\boldsymbol{z}_{it}$ covariates which capture the impacts of past insurgents' behavior. Our model collapses to a special case analogous to Weidmann and Salehyan's model without the diffusion terms. We would like to thank Karsten Donnay for clarifying this point.

Table 1: Model Parameters

Parameter	Notation	Description
Model setting (fixed)		
# population settlements	N	
# insurgent agents	M	
Insurgent behavior		
Constant	α	
Susceptibility parameter	$\boldsymbol{\beta}$	Degree and direction to which agents respond to time-*invariant* local conditions $\boldsymbol{x}$
Diffusion parameter	$\boldsymbol{\gamma}$	Degree and direction to which agents respond to time-*variant* local conditions $\boldsymbol{z}$

corresponding coefficient vectors $\boldsymbol{\beta}$ susceptibility and $\boldsymbol{\gamma}$ diffusion parameters. $\boldsymbol{\gamma}$ is further broken into spatial parameter γ_1 and temporal parameter γ_2 that respectively governs the tendency of insurgent violence to diffuse spatially and temporally. While $\boldsymbol{x}$ and $\boldsymbol{z}$ covariates determine the local conditions that insurgents face, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ parameters govern how they respond to their local environment. Table 1 summarizes the model parameters.

4.3 Distinguishing Micro-mechanisms

Reflecting on insights derived from the existing studies, we include six $\boldsymbol{x}$ covariates: population size ($PopSize$), Pashtun population size ($PashtunPop$), accessibility to roads ($RoadAccess$), distance from the capital ($KabulDist$), and distance from the Afghanistan-Pakistan border ($APborder$).[10] We refer to the corresponding coefficient parameters as β_1, β_2, β_3, β_4, β_5, and β_6. Table 2 reports the summary statistics of these covariates considered.[11]

[10] Three distance related covariates, $RoadAccess$, $KabulDist$, and $APborder$, are measured by the geodesic distance from the closest roads, Kabul, and the Afghanistan-Pakistan border in kilometers, respectively. Spatial data on local income level were derived from the "Geographically based Economic data" (G-Econ, Nordhaus, 2006; available at http://gecon.yale.edu, accessed July 25, 2014), and all other settlement-level attributes and data of road networks are obtained and computed using data available from USAID MIST-I/Humanitarian Response.

[11] We log-transformed and rescaled all covariates to the range $[-1, 1]$ to minimize the effect of extreme values and make estimates easily comparable to each other.

Similarly, we include two $\boldsymbol{z}$ covariates: spatial lag $Spread$ and temporal lag $History$. For a given settlement S_i, $Spread_{it}$ is defined as $Spread_{it} = \sum_{j=1}^{k} w_{ij}^{S} \hat{y}_{jt}$, where w_{ij}^{S} is a spatial weight as defined above, and $\hat{y}_{jt}$ denotes the number of insurgent attacks that have occurred at settlement $S_j \in \mathcal{N}_i$ at time period t. Weighting spatially proximate incidents more heavily than those in remote ones, $Spread_{it}$ measures the distance-weighted degree to which insurgent activities spread across immediate neighborhood networks. $History_{it}$ is defined as the temporally weighted number of insurgent attacks that settlement S_i has experienced until time period t: $History_{it} = \sum_{\tau=1}^{t} w_{t\tau}^{T} \hat{y}_{i\tau}$, where $w_{t\tau}^{T} = (1+t-\tau)^{-1}$ is a temporal weight which is analogous to the spatial weight w_{ij}^{S}. Evaluating temporally proximate incidents heavily, $History_{it}$ captures the severity of past insurgent activities it has experienced, which may also shape the context in which future insurgent activities unfold. We refer to the corresponding parameters γ_1 and γ_2 as spatial parameter and temporal parameter, respectively.

$\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ parameters allow for nuanced operationalization of micro-mechanisms hypothesized by the confusion and contagion hypotheses. $\boldsymbol{\beta}$ parameters govern whether and to what extent insurgent agents respond to the structural conditions they face at the local level, while $\boldsymbol{\gamma}$ parameters shape how agents respond to their local context of insurgent activities and thus which diffusion process is at work. A positive estimate of a given parameter indicates that the corresponding covariate positively (negatively) impacts the settlement-level probability of insurgent violence (relocation), whereas a negative estimates indicates otherwise.

5 Parameter Estimates

This section presents the main findings derived from the computational model. The analysis here aims to optimize the model's parameter combinations such that simulation outcomes closely fit the empirical records along the specified dimensions of agreement, thereby identifying the likely determinants and micro-mechanisms of insurgent violence. In the following, we first present the validation strategy employed and then examine the effects of parameters.

Table 2: Summary Statistics

Covariate (logged)	Mean	Std. Dev.	Median	Range
PopSize	5.695	1.103	5.749	$[1.099, 14.750]$
PashtunPop	2.424	2.997	0	$[0, 12.690]$
Development	9.201	0.389	9.265	$[6.483, 9.703]$
RoadAccess (in km)	−0.026	1.461	0.034	$[-5.817, 3.881]$
KabulDist (in km)	5.425	0.742	5.518	$[-6.908, 6.723]$
APborder (in km)	5.017	1.021	5.224	$[-0.962, 6.466]$

Note: All covariates are logged.

5.1 Validation Strategy

The validation strategy here follows Weidmann and Salehyan (2013).[12] We first run N_{run} simulations with randomly drawn parameters from uniform distributions (parameter space $\mathbf{\Theta}_0$) and then select a subset of empirically plausible parameter combinations $\mathbf{\Theta}_1 \subset \mathbf{\Theta}_0$ that generates distributions of insurgent violence similar to the observed one according to the criteria presented below. This procedure allows for identifying the parameters that are necessary to generate realistic patterns of violence and their impacts on simulation outcomes.[13]

For the following analysis, $N_{run} = 20,000$ simulation runs are conducted using randomly drawn parameter combinations and different random seeds. The resultant distributions of insurgent violence are compared and calibrated with the empirical records along two target classes: *location* and *number* of violence (cf. Bhavnani et al., 2014). The agreement between the predicted and observed locations of violence is measured by first collapsing predicted data series $\hat{\boldsymbol{Y}}$ into a vector of binary indicators $\hat{\boldsymbol{V}} = (\hat{V}_1, \ldots, \hat{V}_N)$, with $\hat{V}_i \in \{0, 1\}$ denoting the occurrence of violence at settlement S_i during the simulation period and then computing

[12]See Berk (2008) for a comprehensive discussion on data-based evaluation of computational models.

[13]This parameterization approach allows for examining a large parameter space throughly at a relatively low computational cost compared to sequential parameter sweeping. Recent studies on empirically-explicit agent-based models (e.g., Bhavnani et al., 2014) as well as purely theoretical models (e.g., Siegel, 2011; Weidmann, Forthcominga) rely upon similar approach to specify the parameter space that generates the best fits with empirical records or examine the effects of individual parameters.

true positive rate (TPR), false positive rate (FPR), and accuracy rate.[14] The degree of agreement for the number of insurgent violence is quantified by computing several Root Meas Squared Error (RMSE) measures.

We define empirically plausible simulation runs as those that minimize the deviation of the model outcome from the empirical records, which is operationalized by the following conditions (cf. Bhavnani et al., 2014; Lim et al., 2007; Weidmann and Salehyan, 2013):

(1) accuracy rate > 0.5 (location of violence);

(2) TPR $>$ FPR (location of violence); and

(3) the (weighted) RMSE smaller (better fit) than the 5th percentile value generated by N_{run} random null cases (number of violence).

We first discard the noninformative runs that generated no insurgent attacks, and then select those that meet these three conditions to obtain empirically plausible parameter space $\mathbf{\Theta}_1$. As a general rule, a model with high binary predictive capability has a TPR consistently higher than the corresponding FPR. Similarly, an accuracy rate greater than 0.5 ensures that the model's predictive performance is at least as good as chance. These conditions allow for filtering those runs that correctly classify violent and peaceful settlements.

A "good-fit" run should not only yield a high TPR/FPR and accuracy ratio but also minimize the deviation of predicted numbers of violence from the observed data series (condition 3). Although RMSE $= \sqrt{\sum_{i=1}^{N}(Y_i - \hat{Y}_i)^2/N}$ is often employed as an error metric for a calibration purpose, it might be ill-suited for the validation here given that occurrence of insurgent violence is relatively rare in our dataset; as $29,840$ (79.6%) out of $37,484$ settlements have experienced no insurgent violence during the period of investigation (i.e., $Y_i = 0$), a noninformative prediction simply assigning $\hat{Y}_i = 0$ produces a small RMSE indicating a

[14]These measures are formally defined as TPR (sensitivity) $= \frac{\#\,\text{true positives (TP)}}{\#\,\text{TP}+\#\,\text{false negatives (FN)}}$, FPR $= \frac{\#\,\text{false positives (FP)}}{\#\,\text{FP}+\#\,\text{true negatives (TN)}}$, and accuracy $= \frac{\#\,\text{TP}+\#\,\text{TN}}{\#\,\text{TP}+\#\,\text{FP}+\#\,\text{TN}+\#\,\text{FN}}$, respectively. To compute these measures, empirical records $\boldsymbol{Y}$ are also collapsed into $\boldsymbol{V} = (V_1, \ldots, V_N)$.

"good-fit" (cf. Chadefaux, 2014, 15). Although this problem can be partly alleviated by condition (1), we use the *weighted* RMSE (wRMSE) $= \sqrt{\sum_{i=1}^{N} w_i (Y_i - \hat{Y}_i)^2 / \sum_{i=1}^{N} w_i}$ instead of RMSE as the error metric to address this problem. Introducing an weight $w_i = 1 - p(Y_i)$ which is associated to each observation, penalizes errors between predicted and observed values for rare observations (e.g., $Y_i = Y_{\max} = 323$) more severely than those for abundant ones (e.g., $Y_i = 0$). Although wRMSE tends to tolerate overpredictions compared to RMSE, it helps alleviate the above problem. In order to ensure the robustness of the results to the weighting scheme, we also present results using an alternative (binary) weighting scheme which assigns $w_i = 1 - 0.204 = 0.796$ to those with violence ($\hat{Y}_i \geq 1$) and $w_i = 0.204$ to those without violence ($\hat{Y}_i = 0$; parameter space $\mathbf{\Theta}_2 \subset \mathbf{\Theta}_0$).[15]

$N_{run} = 20,000$ random null cases, which constitute the benchmark of the validation strategy, are generated by randomly assigning the observed number $(45,628)$ of attacks to population settlements. $20,000$ spatial distributions of insurgent violence in hypothetical "random conflicts" are generated by replicating this procedure. If the wRMSE of a given simulation run is smaller than the 5th percentile of the "random conflict" distribution, the corresponding run is considered to outperform random guesses.

5.2 Determinants of Insurgent Violence

Based on the results of the initial parameter sweeping, we restrict the sampling ranges for parameters $\alpha, \boldsymbol{\beta}$, and $\boldsymbol{\gamma}$ to $[-10, 10]$. This restricted parameter region guarantees a vast variation in simulation results. $M = 15,000$ insurgent agents are allocated to randomly selected population settlements at the beginning of a simulation run. Each run continues until either (1) time step count t reaches to $t_{\max} = 300$, or (2) the accumulated number of insurgent attacks N_t^V reaches to the empirical one (i.e., $N_t^V = 45,628$).

Filtering according to the selection criteria above yields a set of $1,034$ out of $20,000$

[15] Chadefaux (2014) developed a revised Brier score to mediate a similar problem by weighting mistakes by the overall probability of a given outcome. The binary weighting scheme follows Chadefaux's scheme.

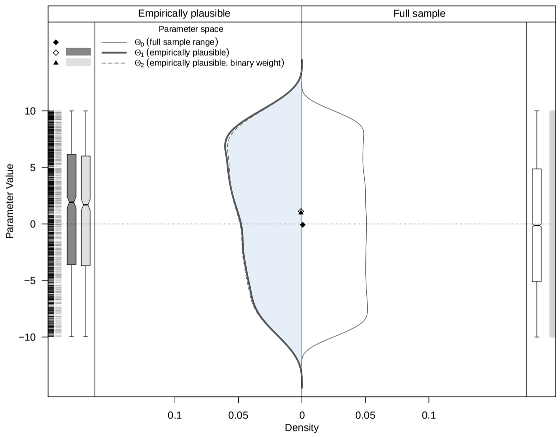
(5.17%) empirically plausible parameter combinations ($\mathbf{\Theta}_1$; 930, 4.65% for $\mathbf{\Theta}_2$). Although the numbers are small, the remaining subsets of parameter combinations are relevant for generating patterns of violence similar to the empirical records, and allow for identifying the likely determinants and micro-mechanisms of insurgent behavior.

Susceptibility Parameter Which parameter shapes the patterns of insurgent violence? The distributions of parameter values in the empirically plausible parameter spaces provide an intuitive indicator: a skewed distribution in the empirically plausible parameter spaces indicates a systematic impact of the corresponding parameter on model fits, whereas a parameter which is not necessary to generate empirically plausible patterns of violence is expected to be indistinguishable from the full sample range (Weidmann and Salehyan, 2013).

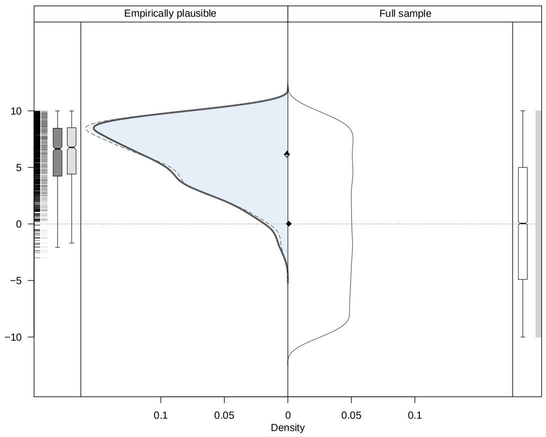
Figure 4 plots the distributions of $\boldsymbol{\beta}$ parameters in $\mathbf{\Theta}_1$ against those in $\mathbf{\Theta}_0$ (uniform distributions), which combines several layers of information. The estimated density curves of parameter values in $\mathbf{\Theta}_0$ are plotted to the right center, whereas those in $\mathbf{\Theta}_1$ are plotted to the left corner. Box plots represent 25th, 50th (median), and 75th percentile values, and notches represent the 95% confidence intervals (CIs) of the medians that provide an eye-guide for a significant difference.[16] The small asterisks and diamonds refer to the means.

The most clear relationship was detected for parameter β_2, which governs the impacts of local ethnic configuration: the density estimate for parameter β_2 (Figure 4b) is consistently positive positive across empirically plausible parameter spaces $\mathbf{\Theta}_1$ and $\mathbf{\Theta}_2$, suggesting that the size of Pashtun population of settlements ($PshtunPop$) would increase the subnational risk of insurgent violence. Significant differences in medians were also found for other $\boldsymbol{\beta}$ parameters in the expected direction, whereas the associations remain weaker than β_2. The density estimates for parameters β_1 (*PopSize*), β_3 (*Development*), and β_5 (*KabulDist*)

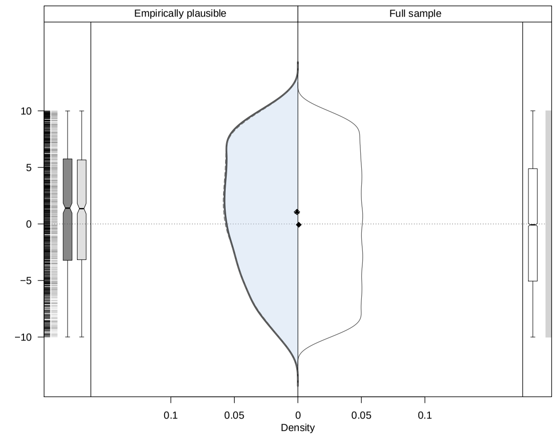
[16]To formally test if distributions of parameters in the full sample range $\mathbf{\Theta}_0$ and those in $\mathbf{\Theta}_1$ ($\mathbf{\Theta}_2$) differ significantly, we ran a non-parametric Mann-Whitney U test (with Bonferroni corrections) comparing the median parameter value in $\mathbf{\Theta}_0$ with those in $\mathbf{\Theta}_1$ ($\mathbf{\Theta}_2$) for each parameter. The corresponding density curve is shaded with blue if the null hypothesis that the samples of parameter values in $\mathbf{\Theta}_0$ and $\mathbf{\Theta}_1$ ($\mathbf{\Theta}_2$) came from the same population was rejected at 5% level. The Welch two sample t-tests yielded similar results.



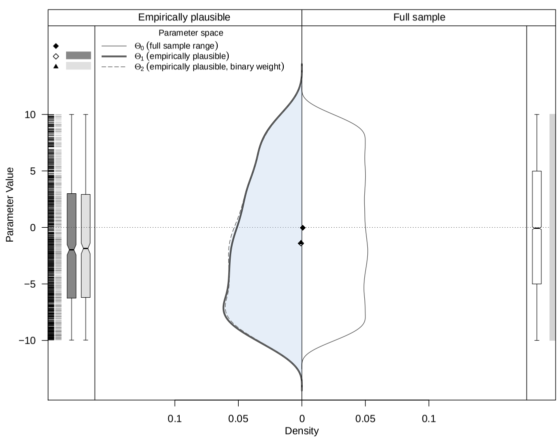
(a) β_1 ($PopSize$)



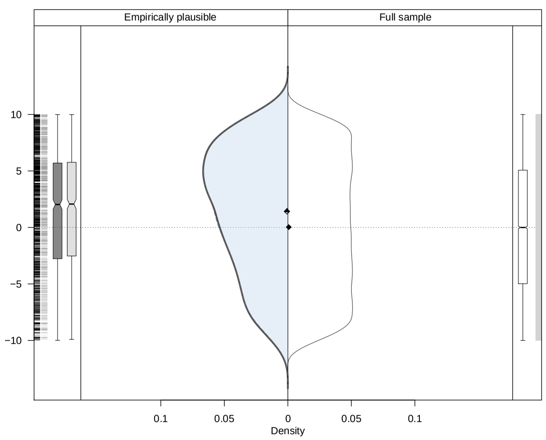
(b) β_2 ($PashtunPop$)



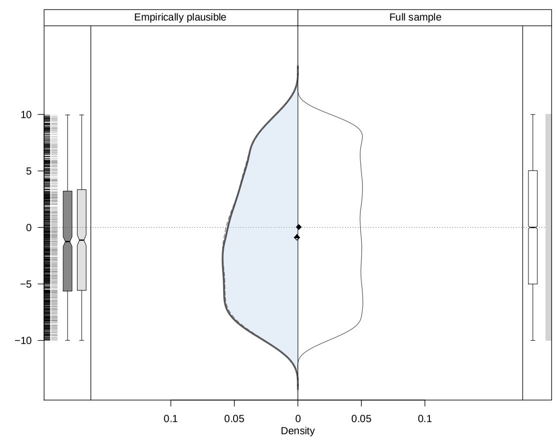
(c) β_3 ($Development$)



(d) β_4 ($RoadAccess$)



(e) β_5 ($KabulDist$)



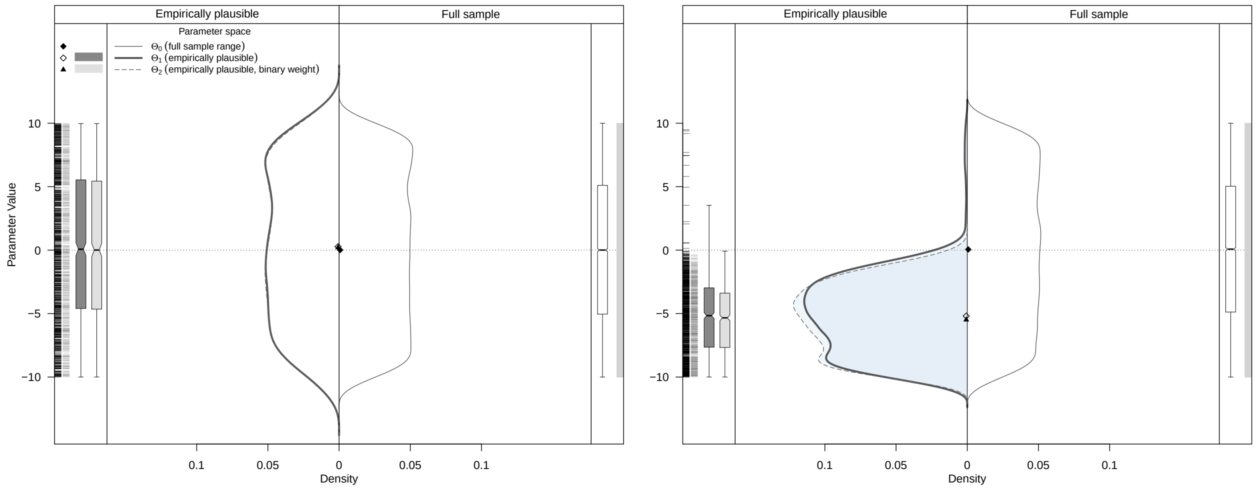
(f) β_6 ($APborder$)

Figure 4: Beanplots of susceptibility parameters. Kernel density estimates of parameter values in $\mathbf{\Theta}_0$ (uniform distributions) are plotted to the right center, whereas those in $\mathbf{\Theta}_1$ and $\mathbf{\Theta}_2$ (empirically plausible parameter spaces) are plotted to the left corner (solid: $\mathbf{\Theta}_1$, dashed: $\mathbf{\Theta}_2$). Diamonds and asterisks refer to the means. The density curve is shaded if a significant difference in median values of the corresponding parameter between $\mathbf{\Theta}_0$ and $\mathbf{\Theta}_1$ ($\mathbf{\Theta}_2$) is found at 5% level (Mann-Whitney's U tests with Bonferroni corrections).

are positively skewed (Figure 4a, c, and e), while the estimates for β_4 ($RoadAccess$) and β_6 ($APborder$) are negatively skewed (Figure 4d and f). Although the levels of economic development and accessibility are postulated to either increase or decrease the local risk of insurgent violence, the model indicates a significant positive relationship for parameter β_3, suggesting that locations with higher levels of development see elevated risk of violence. The negative estimate for parameter β_4 (Figure 4d) demonstrates that accessibility tend to facilitate, rather than containing, insurgent violence. Note that, however, the estimates for $\boldsymbol{\beta}$ parameters excepting for β_2 are not narrow enough to rule out the possibility of opposite effects given that they can take both positive and negative values in $\boldsymbol{\Theta}_1$ and $\boldsymbol{\Theta}_2$.

Taken together, these parameter estimates suggest that a simulation run tends to generate patterns of insurgent violence similar to the observed one; and the effect of parameter β_2 is most consistent across the empirically plausible parameter spaces when insurgent agents are likely to conduct attacks in those settlements with large population (β_1) as well as large Pashtun population (β_2) and higher levels of income (β_3) located further from Kabul (β_5) but closer to roads (β_4) and the Afghanistan-Pakistan border (β_6). These detected relationships are fairly robust to the choice of the weighting scheme of wRMSE, indicating that they are unlikely to be a product of the specific selection criteria.

Diffusion Parameter Figure 5 shows the density estimates for diffusion parameters γ_1 and γ_2, which govern whether and how insurgent agents respond to location specific context ($Spread$) and history of violence ($History$). An apparent association is found for temporal parameter γ_2: the distributions of γ_2 in empirically plausible parameter spaces $\boldsymbol{\Theta}_1$ and $\boldsymbol{\Theta}_2$ are consistently negative, indicating that marked history of violence facilitates insurgents' migration rather than further violence in the originating settlements. In contrast, the density estimate for spatial parameter γ_1 is statistically indistinguishable from the uniform distribution, suggesting that γ_1 is not likely to have a systematic impact on model's fit with the empirical records. This result leads us to a conclusion that a negative value assigned to γ_2



(a) γ_1 (*Spread*) (b) γ_2 (*History*)

Figure 5: Beanplots of diffusion parameters. Kernel density estimates of parameter values in $\mathbf{\Theta}_0$ are plotted to the right center, whereas those in $\mathbf{\Theta}_1$ (solid) and $\mathbf{\Theta}_2$ (dashed) are plotted to the left corner. Diamonds and asterisks refer to the means. The density curve is shaded if a significant difference in median values of the corresponding parameter between $\mathbf{\Theta}_0$ and $\mathbf{\Theta}_1$ ($\mathbf{\Theta}_2$) is found at 5% level (Mann-Whitney's U tests with Bonferroni corrections).

is likely to be a necessary condition for the model to generate "good-fit" distributions. As with the estimates for $\boldsymbol{\beta}$ parameters reported above, these results appear to be fairly robust to the selection of error metric employed.

The strongly negative estimate for temporal parameter γ_2 indicates that occurrence of violence tends to facilitate migration of insurgent agents from their current locations to nearby locations in those simulation runs that minimize the deviations of generated distributions from the observed one. This is consistent with the pattern of relocation diffusion, suggesting that occurrence of violence facilitates insurgents' relocation and thereby spreads "seeds" of violence from the originating locations.

6 Explanatory Power

The simulation exercise in the previous section helps us identify the likely determinants of the insurgent violence. However, the analysis alone informs us little about the explanatory power of the model. Whether and to what extent does the model correctly predict the location and number of insurgent violence across the population settlements? Although the main purpose here is not to generate extremely accurate predictions, an assessment of the model's predictive performance, even in-sample, is likely to be a valuable heuristic of its explanatory power (Ward et al., 2010; Weidmann and Ward, 2010).

The model's capability to correctly classify violent and peaceful settlements can be quantified using the Receiver Operating Characteristic (ROC) curve and the area under the ROC curve (AUC) score. A ROC curve plots TPR and FPR as the output of each possible probability threshold for positive prediction, thereby evaluating the model's binary classification performance. The resultant curve displays the balance between TPR and FPR where a highly predictive model (with high TPR and low FPR) produces the curve up in the top left corner. An AUC score ranges between 0 and 1 and provides a single number summary of the model's classification performance. A random coin toss produces an AUC score of 0.5, whereas a model with higher classification performance should yield a greater score.

The predicted probability of violence for each settlement is computed by simply averaging the binary output $\hat{V}_i$. The probability assigned to a given settlement reflects the fraction of simulation runs with optimized parameter combinations where insurgent violence has occurred in the corresponding settlement. The ROC analysis using the computed probability yields an AUC score of 0.77 (95% CI: $0.764, 0.776$), indicating that the model's capability to classify those settlements with and without events is well beyond that of a random coin toss.[17] Using the resultant best threshold value for positive prediction that maximizes the AUC score, Figure 6 maps the predicted spatial distribution of insurgent violence. The model correctly predicted $6,016$ out of $7,644$ observed locations of insurgent violence (true positives), while producing $9,330$ false positives (TPR $= 0.787$ and FPR $= 0.313$; 1.55 false positives per true positive). Overall, the computational model correctly classified $26,526$

[17]The 95% CI was obtained by bootstrap using R's pROC package (Robin et al., 2011). Parameter space $\mathbf{\Theta}_2$ yields an AUC score of 0.769 (95% CI: $0.763, 0.775$), with $6,005$ true positives and $9,234$ false positives.

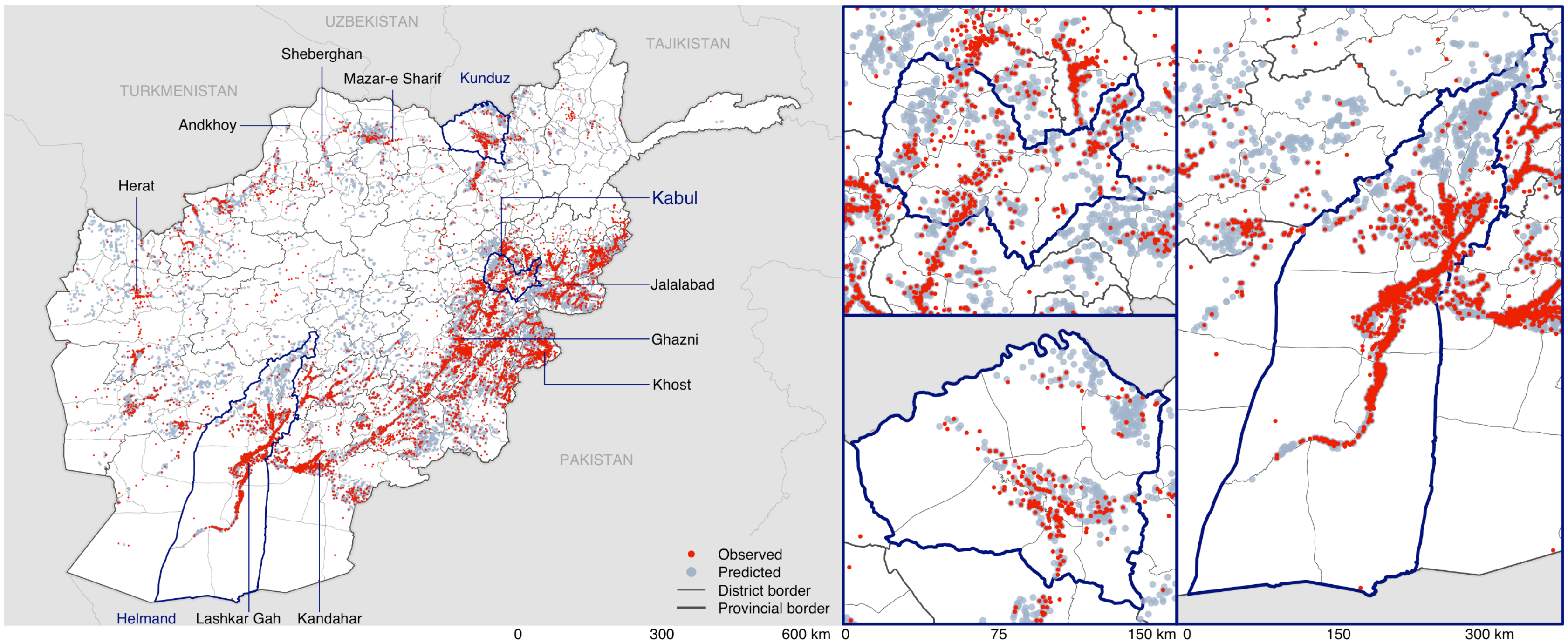
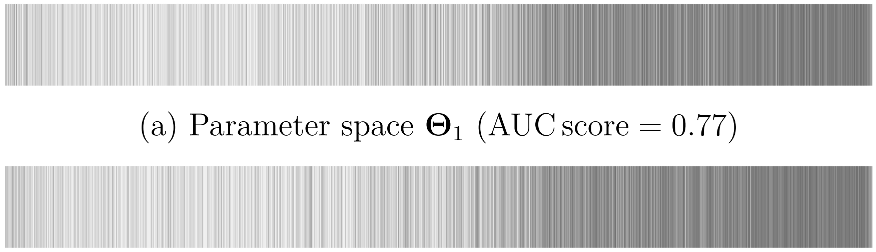


Figure 6: Predicted distribution of insurgent violence. Blue dots (●) represent the predicted locations of population settlements with insurgent violence, whereas red dots (●) represent the observed locations of settlements with one or more insurgent violence. The solid lines represent district and provincial borders. The best threshold value (0.427) obtained by the ROC analysis was used to classify the predicted “violent” and “peaceful” settlements.



(a) Parameter space $\mathbf{\Theta}_1$ (AUC score = 0.77)

(b) Parameter space $\mathbf{\Theta}_2$ (AUC score = 0.769)

Figure 7: Separation plots. Dark gray lines represent observed instances of events ($V_i = 1$), whereas light gray lines represent nonevents ($V_i = 0$). The predicted probability of insurgent violence increases from left to right.

out of 37,484 (70.8%) violent and peaceful population settlements (accuracy = 0.708).[18] Separation plots (Greenhill et al., 2011) in Figure 7 visualize the model's binary classification performance, which again underscore its internal validity.

The agreement between the simulated and the observed numbers of violence can be quantified using standard correlation measure. The Spearman's (Peason's) correlation between the mean simulated and the observed numbers of violence is 0.405 (0.367, logged) for parameter space $\mathbf{\Theta}_1$ and 0.405 (0.367, logged) for $\mathbf{\Theta}_2$. While the correlations remain modest, this is at least the level of agreement that is not reached in any of 20,000 randomized trials.[19]

Although the fair in-sample predictive performance underscores the model's internal validity, over- and under-predictions were generated by the optimized parameter combinations. Why do some predictions deviate from the empirical records? Given that the model is completely governed by the predetermined behavior rules and parameters, possible omitted variables or interactions are likely to account for the deviations. Most notably, the model does not incorporate any counterinsurgent (COIN) efforts, which has been hypothesized to shape insurgents' behavior (e.g., Braithwaite and Johnson, 2012; Linke et al., 2012; O'Loughlin and Witmer, 2012). Indeed, over-predictions are concentrated around the center part of the

[18]This point must be interpreted with caution because a comprehensive test of predictive accuracy requires an *out-of-sample* validation while the analysis here is essentially an *in-sample* validation. Although we deeply recognize the danger of overfitting, we leave testing out-of-sample validation for future research.

[19]The highest correlation reached by the "random conflicts" remains $\rho = -0.044$.

country (Figure 6), which roughly corresponds to the vacuum of COIN activities.

A straightforward preliminary analysis provides tentative support for this speculation: the model systematically over-predicts insurgent violence in settlements with few COIN efforts while under-predicting violence in those with marked COIN activities. The average number of ISAF attacks is considerably higher in the false negative settlements (0.504) than in the false positive ones (0.065).[20] Similarly, under-predictions in numbers of attacks are positively correlated with the observed numbers of ISAF violence, with Spearman's (Peason's) correlation of 0.403 (0.465, logged). Given the positive correlation between the observed numbers of insurgent and COIN violence ($\rho = 0.409, r = 0.582$), this comparison implies that the deviations are likely to be the consequence of omission of COIN efforts.

This is consistent with the findings of existing studies in support of the "tit-for-tat" associations between insurgent and COIN activities (e.g., Braithwaite and Johnson, 2012; Linke et al., 2012; O'Loughlin and Witmer, 2012). This is the interaction that is not modeled within the presented framework that exclusively focuses on insurgents' behavior.

7 Robustness Checks

The analyses in the previous two sections identified the likely determinants of insurgent violence and demonstrated the veracity of the computational model. However, if it were the case that the results depended on some specifications, this would question the theoretical and empirical plausibility of the model. Specifically, potential sensitivities are likely to lie along the two dimensions: the topology of neighborhood networks and the behavior of insurgent agents. We provide an overview of the sensitivity tests regarding these two dimensions in the following subsections. Reassuringly, none of the following sensitivity tests yields results that deviate markedly from those reported in the previous sections.[21]

[20] In contrast, the opposite holds for the correct predictions: the average number of ISAF attacks is higher in the true positive settlements (0.814) than in the true negative ones (0.012).

[21] See Appendix for more detailed information on the following robustness checks <URL>.

7.1 Neighborhood Network

The neighborhood networks are the pathways through which insurgent agents move around and violence diffuses. Naturally, one might wonder whether and to what extent the network topology influences the simulation results. Do alternative definitions of neighborhood network substantially alter the results reported in the previous section? In order to examine potential sensitivities of the results, additional $20,000 \times 2$ simulation runs have been conducted using alternative network sizes $k = 10$ and $k = 30$ instead of the baseline value of $k = 20$ while holding all other parameters at the baseline values.

Figures **??** to **??** in Appendix plot the density estimates for simulation runs using these alternative network sizes. As these density estimates are substantially indistinguishable from those with $k = 20$, it can be concluded that the parameter estimates are fairly robust to the changes of neighborhood definitions. The alternative parameter settings did not alter the model's predictive performance either. Optimized parameter combinations yield the AUC scores of 0.762 (95% CI: $0.756, 0.768$) for $k = 10$ and 0.775 (95% CI: $0.77, 0.781$) for $k = 30$, respectively. The levels of correlations between predicted and observed numbers of violence also remain at the level reported in the previous section, with Spearman's (Pearson's) $\rho =$ 0.394 (0.355) for $k = 10$ and 0.413 (0.374) for $k = 30$. These results guarantee that the results presented in the previous sections are not products of the specific network sizes.

7.2 Binary Decision

Thus far, insurgent agents are assumed to make binary decisions: attack at the current location or relocate to another settlement. Although this dichotomy applies as long as there are several actions of which only one is subject to the analysis (cf. Siegel, 2011, 995), insurgents might alternatively decide to just stay and hide among civilians while not conducting attacks. We examine the potential sensitivity of the simulation results to the dichotomy assumption by allowing for the third option of "stay and hide."

Specifically, we extend the baseline model such that it incorporates an additional model parameter q which determines the probability that insurgent agents decide to stay at their current locations. If insurgent agent I_j decides not to conduct an attack with probability $1-p_{ijt+1}$, I_j decides to stay at its current location S_i with probability q; otherwise, it decides to migrate to another settlement S_j with probability $1-q$. This extended model coincides with the baseline model when $q = 0$.

Another $20,000$ simulation runs were conducted with q set at 0.5 while holding all other parameters as in the baseline setting. The results generally agree with those presented in the previous section, suggesting that our results do not depend crucially on the binary-decision assumption. The same set of structural factors significantly impacts the model's fit with the empirical records, while relocation diffusion process also shapes the insurgent behavior (Figures **??** and **??** in Appendix). The extended model yields an almost identical explanatory power with the baseline model, with the AUC score of 0.769 (95% CI: $0.763, 0.775$) and the correlation estimate of $\rho = 0.409$. These results suggest that the main findings are not likely to be an artifact produced by the dichotomy in the baseline model. These lead us to the conclusion that the simple dichotomy employed in the baseline model is sufficient to generate realistic spatial patterns of insurgent violence.

8 Conclusion

Violence in the context of civil war diffuses and clusters, but the micro-mechanisms underlying the observed macro-outcome have been remained disputed. On the one hand, contagious nature of insurgent activities can alter the prospects for future violence at the same and nearby locations, thereby generating clusters of violence. On the other hand, clusters of violence at the macro level are consistent with the proposition that clusters of violence simply emerge from a similar distribution of violence-attracting attributes. Drawing on fine-grained georeference data and agent-based computational modeling technique, this paper explored the micro-level mechanisms underlying the macro-level patterns of insurgent violence. The computational model demonstrated that while exogenous structural conditions such as local ethnic configuration substantially constrain insurgents' behavior, endogenous diffusion processes are also likely to shape how insurgent activities unfold. Specifically, the model demonstrated that relocation diffusion process is likely to be consistent with insurgent violence in Afghanistan, suggesting that such an endogenous dynamic cannot be simply assumed away in the study of civil war violence. The fairly good agreement between the simulated and observed distributions of insurgent violence suggests that the model captures the plausible micro-level mechanisms of insurgent behavior. Moreover, these results and findings were found to be robust to the changes in parameter settings.

The findings derived from computational model provide support for relocation diffusion and run counter to the earlier insights of Schutte and Weidmann (2011) who argue that the diffusion patterns of civil war violence are primarily escalation rather than relocation diffusion. Even though Schutte and Weidmann (2011) do not include Afghanistan within the scope of analysis and imply that the particular diffusion process observed should vary from a civil war to another, the simulation exercise in this paper suggests that relocation diffusion would also shape how irregular warfares unfold.

The limitations of this paper include the purely in-sample validation strategy and the assumption of the agent-based model. First, the current analysis remains essentially in-sample validation where the pitfall of overfitting cannot be ruled out. An over-fitted model tends to reproduce idiosyncratic patterns of the training data rather than capturing the systematic features of the generating mechanisms. This would produce accurate in-sample predictions while yielding poor out-of-sample predictive performance. An out-of-sample, in addition to the current in-sample, validation scheme is likely to be required to further validate the computational model.

Another issue concerns the assumption of the computational model: the exclusive focus on insurgent behavior. Put another way, it remains a model of insurgency without counterinsurgency. This simplification might be problematic given that existing studies have consistently found the "tit-for-tat" associations between insurgent violence on the one side and counterinsurgent efforts on the other (e.g., Braithwaite and Johnson, 2012; Linke et al., 2012; O'Loughlin and Witmer, 2012; Toft and Zhukov, 2012; but see Braithwaite and Johnson, 2015). The systematic relationship between the observed COIN efforts and over- and under-predictions also suggests that it is a main challenge for future research is to incorporate the model with COIN activities as well as interactions with insurgents. Despite its preliminary character, the data-driven computational model will contribute to the emerging research agenda of disaggregation of civil war and better understanding of conflict process.

References

Alcock, Norman Z. 1972. *The war disease.* New York: CPRI.

Amiri, Mohammad Abid. 2013. "Road reconstructions in post-conflict afghanistan: A cure or a curse?" *International Affairs Review* **21**(2): 2–16.

Axelrod, Robert. 1997. *The complexity of cooperation: Agent-based models of competition and collaboration.* Princeton, NJ: Princeton University Press.

Balcells, Laia, and Stathis N. Kalyvas. 2014. "Does warfare matter? Severity, duration, and outcomes of civil wars." *Journal of Conflict Resolution* **58**(8): 1390–1418.

Baudains, Peter, Shane D. Johnson, and Alex Maves Braithwaite. 2013. "Geographic patterns of diffusion in the 2011 London riots." *Applied Geography* **45**: 211–219.

Berk, Richard. 2008. "How you can tell if the simulations in computational criminology are any good." *Journal of Experimental Criminology* **4**(3): 289–308.

Berman, Eli, Michael Callen, Joseph H. Felter, and Jacob N. Shapiro. 2011. "Do working men rebel? Insurgency and unemployment in Afghanistan, Iraq, and the Philippines." *Journal of Conflict Resolution* **55**(4): 496–528.

Bhavnani, Ravi, Karsten Donnay, Dan Miodownik, Maayan Mor, and Dirk Helbing. 2014. "Group segregation and urban violence." *American Journal of Political Science* **58**(1): 226–245.

Braithwaite, Alex, and Shane D. Johnson. 2012. "Space-Time Modeling of Insurgency and Counterinsurgency in Iraq." *Journal of Quantitative Criminology* **28**(1): 31–48.

———. 2015. "Terrorism and Political Violence The Battle for Baghdad: Testing Hypotheses About Insurgency From Risk Heterogeneity, Repeat Victimization, and Denial Policing Approaches." *Terrorism and Political Violence* **27**(1): 112–132.

Buhaug, Halvard. 2010. "Dude, where's my conflict? LSG, relative strength, and the location of civil war." *Conflict Management and Peace Science* **27**(2): 107–128.

Buhaug, Halvard, Kristian Skrede Gleditsch, Helge Holtermann, Gudrun Østby, and Andreas Foro Tollefsen. 2011. "It's the local economy, stupid! geographic wealth dispersion and conflict outbreak location." *Journal of Conflict Resolution* **55**(5): 814–840.

Buhaug, Halvard, and Scott Gates. 2002. "The geography of civil war." *Journal of Peace Research* **39**(4): 417–433.

Buhaug, Halvard, and Kristian Skrede Gleditsch. 2008. "Contagion or confusion? Why conflicts cluster in space." *International Studies Quarterly* **52**(2): 215–233.

Buhaug, Halvard, and Jan Ketil Rød. 2006. "Local determinants of african civil wars, 1970–2001." *Political Geography* **25**(3): 315–335.

Chadefaux, Thomas. 2014. "Early Warning Signals for War in the News." *Journal of Peace Research* **51**(1): 5–18.

Collier, Paul, and Anke Hoeffler. 2004. "Greed and grievance in civil war." *Oxford Economic Papers* **56**(4): 563–595.

Danneman, Nathan, and Emily Hencken Ritter. 2013. "Contagious rebellion and preemptive repression." *Journal of Conflict Resolution* **58**(2): 254–279.

Donnay, Karsten, and Vladimir Filimonov. 2014. "Views to a war: Systematic differences in media and military reporting of the war in Iraq." *EPJ Data Science* **3**(1) 25.

Farrell, Theo, and Antonio Giustozzi. 2013. "The Taliban at war: Inside the Helmand insurgency, 2004–2012." *International Affairs* **89**(2013): 845–871.

Fearon, James D., and David D. Laitin. 2003. "Ethnicity, Insurgency, and Civil War." *American Political Science Review* **97**(1): 75–90.

Gall, Carlotta. 2004. "Taliban Leader Vows Return." *The New York Times*, November 13, 2004.

Greenhill, Brian, Michael D. Ward, and Audrey Sacks. 2011. "The separation plot: A new visual method for evaluating the fit of binary models." *American Journal of Political Science* **55**(4): 991–1002.

Grossman, Herschel I. 1991. "A General Equilibrium Model of Insurrections." *American Economic Review* **81**(4): 912–921.

Gruber, Jack. 2007. "The perils of carving a path to the taliban's front door." *USA Today*: June 20, 2007.

Hegre, Håvard, Gudrun Østby, and Cionadh Raleigh. 2009. "Poverty and Civil War Events: A Disaggregated Study of Liberia." *Journal of Conflict Resolution* **53**(4): 598–623.

Herbst, Jeffrey. 2000. *States and power in Africa: Comparative lessons in authority and control*. Princeton, NJ: Princeton University Press.

Hirose, Kentaro, Kosuke Imai, and Jason Lyall. 2014. "Can civilian attitudes predict civil war violence?." Paper prepared for the 1st Asian Political Methodology Meeting, Tokyo Institute of Technology, January 6–7, 2014.

Iqbal, Zaryab, and Harvey Starr. 2008. "Bad Neighbors: Failed States and Their Consequences." *Conflict Management and Peace Science* **25**(4): 315–331.

Ito, Gaku, and Susumu Yamakage. 2014. "From KISS- to TASS-modeling: A preliminary analysis of the segregation model incorporated with GIS data on Chicago." Unpublished Manuscript, University of Tokyo.

Johnson, Neil, Spencer Carran, Joel Botner, Kyle Fontaine, Nathan Laxague, Philip Nuetzel, Jessica Turnley, and Brian Tivnan. 2011. "Pattern in escalations in insurgent and terrorist activity." *Science* **333**(6038): 81–84.

Johnson, Shane D. 2008. "Repeat burglary victimisation: A tale of two theories." *Journal of Experimental Criminology* **4**(3): 215–240.

Johnson, Thomas H., and Matthew C. DuPee. 2012. "Analysing the new Taliban Code of Conduct (Layeha): An assessment of changing perspectives and strategies of the Afghan Taliban." *Central Asian Survey* **31**(December 2014): 77–91.

Johnson, Thomas H., and M. Chris Mason. 2008. "No sign until the burst of fire: Understanding the Pakistan-Afghanistan frontier." *International Security* **32**(4): 41–77.

Kalyvas, Stathis N. 2006. *The Logic of Violence in Civil War*. Cambridge: Cambridge University Press.

———. 2008. "Promises and pitfalls of an emerging research paradigm: The microdynamics of civil war." In Kalyvas, Stathis N., Ian Shapiro, and Tarek Masoud eds. *Order, Conflict, and Violence*. Cambridge: Cambridge University Press.

Lim, May, Richard Metzler, and Yaneer Bar-Yam. 2007. "Global pattern formation and ethnic/cultural violence." *Science* **317**(5844): 1540–1544.

Linke, Andrew M., Frank D. W. Witmer, and John O'Loughlin. 2012. "Space-time Granger analysis of the war in Iraq: A study of coalition and insurgent action-reaction." *International Interactions* **38**(4): 402–425.

Linke, Andrew M., Sebastian Schutte, and Halvard Buhaug. Forthcoming. "Population attitudes and the spread of political violence in sub-saharan africa." *International Studies Review*.

Lyall, Jason. 2009. "Does indiscriminate violence incite insurgent attacks? Evidence from Chechnya." *Journal of Conflict Resolution* **53**(3): 331–362.

———. 2014. "Bombing to Lose? Airpower and the Dynamics of Coercion in Counterinsurgency Wars." Unpublished manuscript, Yale University.

Mao, Tse-tung. 1961. *On guerrilla warfare*. Chicago: Praeger.

McColl, Robert W. 1969. "The Insurgent State: Terrirotial Bases of Revolution." *Annals of the Association of American Geographers* **59**(4): 613–631.

Most, Benjamin A., and Harvey Starr. 1980. "Diffusion, reinforcement, geopolitics, and the spread of war." *American Political Science Review* **74**(4): 932–946.

Nordhaus, William D. 2006. "Geography and Macroeconomics: New Data and New Findings." *Proceedings of the National Academy of Sciences of the United States of America* **103**(10): 3510–3517.

O'Loughlin, John, Frank D. W. Witmer, Andrew M. Linke, and Nancy Thorwardson. 2010a. "Peering into the fog of war: The geography of the WikiLeaks Afghanistan War Logs, 2004–2009." *Eurasian Geography and Economics* **51**(4): 472–495.

O'Loughlin, John, Frank D. W. Witmer, and Andrew M. Linke. 2010b. "The Afghanistan-Pakistan Wars, 2008–2009: Micro-geographies, conflict diffusion, and clusters of violence." *Eurasian Geography and Economics* **51**(4): 437–471.

O'Loughlin, John, and Frank D. W. Witmer. 2012. "The diffusion of violence in the North Caucasus of Russia, 1999–2010." *Environment and Planning A* **44**(10): 2379–2396.

Østby, Gudrun, Ragnhild Nordås, and Jan Ketil Rød. 2009. "Regional inequalities and civil conflict in Sub-Saharan Africa." *International Studies Quarterly* **53**(2): 301–324.

Raleigh, Clionadh, and Håvard Hegre. 2009. "Population size, concentration, and civil war: A geographically disaggregated analysis." *Political Geography* **28**(4): 224–238.

Robin, Xavier, Natacha Turck, Alexandre Hainard, Natalia Tiberti, Frédérique Lisacek, Jean-Charles Sanchez, and Markus Müller. 2011. "pROC: An open-source package for R and S+ to analyze and compare ROC curves." *BMC Bioinformatics* **12**(1): 77–84.

Schutte, Sebastian. 2014. "Geography, Outcome, and Casualties: A Unified Model of Insurgency." *Journal of Conflict Resolution.*

Schutte, Sebastian, and Nils B. Weidmann. 2011. "Diffusion patterns of violence in civil wars." *Political Geography* **30**(3): 143–152.

Sieff, Kevin. 2014. "After billions in U.S. investment, Afghan roads are falling apart." *The Washington Post*, January 30, 2014, A3.

Siegel, David A. 2011. "When does repression work? Collective action in social networks." *Journal of Politics* **73**(4): 993–1010.

Toft, Monica Duffy. 2003. *The geography of ethnic violence: Identity, interests, and the indivisuality of territory*. Princeton, NJ: Princeton University Press.

Toft, Monica Duffy, and Yuri M. Zhukov. 2012. "Denial and punishment in the North Caucasus: Evaluating the effectiveness of coercive counter-insurgency." *Journal of Peace Research* **49**(6): 785–800.

Tollefsen, Andreas Forø, and Halvard Buhaug. 2015. "Insurgency and inaccessibility." *International Studies Review*: n/a–n/a.

United States Agency for International Development (USAID). 2014. "Infrastructure factsheet, Aug. 2014." http://www.usaid.gov/afghanistan/infrastructure, accessed February 2, 2015.

Ward, Michael D., Brian D. Greenhill, and Kristin M. Bakke. 2010. "The perils of policy by p-value: Predicting civil conflicts." *Journal of Peace Research* **47**(4): 363–375.

Weidmann, Nils B. 2009. "Geography as motivation and opportunity: Group concentration and ethnic conflict." *Journal of Conflict Resolution* **53**(4): 526–543.

———. 2013. "The higher the better? The limits of analytical resolution in conflict event datasets." *Cooperation and Conflict* **48**(4): 567–576.

———. 2015. "Communication, technology, and political conflict: Introduction to the special issue." *Journal of Peace Research*.

———. Forthcominga. "Micro-cleavages and violence in civil wars: A computational assessment." *Conflict Management and Peace Science*.

———. Forthcomingb. "On the accuracy of media-based conflict event data." *Journal of Conflict Resolution*.

Weidmann, Nils B., and Idean Salehyan. 2013. "Violence and ethnic segregation: A computational model applied to Baghdad." *International Studies Quarterly* **57**(1): 52–64.

Weidmann, Nils B., and Michael D. Ward. 2010. "Predicting Conflict in Space and Time." *Journal of Conflict Resolution* **54**(6): 883–901.

Zammit-Mangion, A., Michael Dewar, Visakan Kadirkamanathan, and Guido Sanguinetti. 2012. "Point Process Modelling of the Afghan War Diary." *Proceedings of the National Academy of Science* **109**(31): 12414–12419.

Zhukov, Yuri M. 2012. "Roads and the diffusion of insurgent violence." *Political Geography* **31**(3): 144–156.