On the Beaten Path: Violence Against Civilians and Simulated Conflict Along Road Networks

Andrea Salvi, Mark Williamson, and Jessica Draper

Abstract Why do some conflict zones exhibit more violence against civilians than others? In answering this question, the literature has emphasized ethnic fractionalization, territorial control and strategic incentives, while overlooking the consequences of armed conflict itself. This oversight is partly due to the methodological hurdles of finding an appropriate counterfactual for observed battle events. In this chapter, we aim to test empirically the effect of instances of armed clashes between rebels and the government in civil wars on violence against civilians. Battles between belligerents may create conditions that lead to surges in civilian killings as combatants seek to consolidate civilian control or inflict punishment against populations residing near areas of contestation. Since there is no relevant counterfactual for these battles, we utilize road networks to help build a synthetic risk-set of plausible locations for conflict. Road networks are crucial for the logistical operations of a civil war and are thus the main conduit for conflict diffusion. As such, the majority of battles should take place in the proximity of road networks; by simulating events in the same geographic area, we are able to better approximate locations where battles hypothetically could have occurred but did not. We test this simulation approach using a case study of the Democratic Republic of the Congo (1998–2000) and model the causal effect of battles using a spatially disaggregated framework. This work contributes both substantively and methodologically to the literature on micro-foundations of civil war and reactive violence in two main ways: (1) It offers a tentative framework for crafting synthetic counterfactuals with event data. (2) It proposes an empirical test for explaining the variation of violence against civilians as a result of battle events.

A. Salvi (✉)
Trinity College Dublin, Dublin, Ireland
e-mail: salvia@tcd.ie

M. Williamson
McGill University, Montreal, Canada
e-mail: mark.williamson@mail.mcgill.ca

J. Draper
University of Mannheim, Mannheim, Germany
e-mail: jdraper@mail.uni-mannheim.de

1 Introduction

The consequences of violence against civilians (VAC) within the context of civil wars continue to be a debated topic with, at times, contradictory findings. Some argue that the use of violence against civilians is counterproductive to incumbents' goals (Kalyvas 2006; Kocher et al. 2011; Lyall 2017) while others find the opposite (Lyall 2009; Stoll 1993). Departing from this debate, a growing literature has sought to better understand VAC as a dependent rather than independent variable. Consequently, numerous studies have examined factors that may explain VAC, including ethnic fractionalization, territorial control, strategic incentives, and various geographic variables (Fjelde and Hultman 2014; Schwartz and Straus 2018; Raleigh 2012; Wood 2010; Schutte 2015, 2017). This chapter aims to contribute to this body of literature by using a geographic event-based approach to further investigate the occurrence of violence against civilians.

Building on theories of VAC being used as a tactic by warring parties, we proceed to ask the following question: what effect do instances of armed conflict actually have on violence against civilians? We suspect that geographical factors, specifically road networks, are crucial for the logistical operations of a civil war, thus making areas around these road networks more prone to conflict in general. Are actual, observed battles between incumbents and insurgents causing the variation in violence against civilians or does indiscriminate violence simply occur in these more conflict-prone areas, even where battles do not necessarily take place? By addressing this question using georeferenced micro-level data, we hope to increase our understanding about the relationship between conflict waged between combatants and violence experienced by civilians.

This chapter begins with a brief overview of the literature that is both theoretically and methodologically relevant to the present study. Next, we offer our new approach for achieving causal identification in spatial models by using simulated conflict events around road networks. Finally, we present results to demonstrate the feasibility of our approach and then conclude with recommendations for future researchers similarly seeking to use simulation techniques.

2 Conflict and Violence Against Civilians

The existing literature has not reached a consensus on whether armed conflict has any direct effect on the prevalence of violence against civilians. Sullivan (2012) found evidence that insurgent violence may increase the probability of massacres

carried out by the state in order to inflict punishment and remove insurgent threats. Fjelde and Hultman (2014) showed that warring parties are more likely to inflict violence against civilians in geographical areas inhabited by the enemy's "ethnic constituency" for a similar reason: to undermine and weaken the enemy and their potential ethnic support base (p. 1233). This follows the findings of Valentino et al. (2004) demonstrating that, particularly in guerrilla wars, combatants target civilians in order to increase their own control and reduce collaboration between local populations and their adversary. The role of rebel strength and capacity has also been investigated with results suggesting that weaker insurgents are more likely to engage in violence against civilians as a means to raise the enemy's cost of fighting (Hultman 2007) or to compel support from the population (Wood 2010). On the whole, existing theories regarding the relationship between conflict and violence against civilians are largely driven by military strategy, whether that be reducing the enemy's base of support, coercing one's own support, or inflicting fighting costs to achieve concessions.

While this literature offers compelling theoretical contributions with regard to the relationship between conflict and VAC, only recently have empirical studies thoughtfully examined the occurrence of armed conflict as a main independent variable. The introduction of georeferenced event-level data has been central in this emerging research agenda. Most notably, Raleigh (2012) utilized a time-location-actor-action model using event data from the Armed Conflict Location & Event Data Project (ACLED) and found a lack of co-occurrence of armed conflict events and instances of violence against civilians in time and space. Instead, a pattern of VAC events emerged around areas occupied by several active groups, suggesting that violence against civilians is not, in fact, "a strategy to gain civilian support or punish civilians" but rather a strategy more often used as a means for competition among violent actors (Raleigh 2012, p. 478). That these results run counter to the prevailing discourse highlights the importance of using event-level data to study the effect of armed conflict occurrence on VAC.

Previous micro-level empirical studies investigating the relationship between conflict and civilian victimization have taken varied approaches to causal identification. In some works, the amount of spatial-temporal correlation between conflict events and VAC is taken as evidence for or against a theory of strategic civilian targeting in war (e.g., Raleigh 2012). However, since the location and timing of conflict occurrence is likely to be driven by strategic behavior related to the civilian population, there is a need to consider whether it is (a) armed clashes that are themselves causing civilian targeting or (b) the underlying conditions that provoke armed clashes that are simultaneously driving levels of violence against civilians. Other studies have addressed similar considerations in the case of indiscriminate violence more broadly (Lyall 2009, 2017; Schutte 2015). In seeking to determine the effect of indiscriminate counterinsurgent artillery fire on subsequent insurgent attacks, Lyall (2009) used a matching technique to compare shelled villages ("treatment") and non-shelled villages ("control") with difference-in-difference estimation. Additionally, Schutte (2015) employed a matched wake

analysis (Schutte and Donnay 2014) on spatio-temporal event data to test the "treatment" of indiscriminate violence on civilian collaboration with the incumbent. These studies used improved techniques for causal identification. However, as we describe in the following section, this literature could benefit from designs based around the creation of simulated battle events as a relevant counterfactual to actual occurrence of violence.

3   A New Strategy for Causal Identification: Creating Synthetic Events on the "Beaten Path"

As the foregoing discussion has made clear, existing empirical studies have produced both uncertain and contradictory findings on the effect of armed conflict on civilian victimization. In part, this is a byproduct of the challenges of achieving causal identification using observational data on conflict events. The co-occurrence of armed conflict and violence against civilians in spatial-temporal windows (or lack thereof) cannot provide definitive evidence of a truly causal relationship.

For a more accurate picture, it is necessary to identify and compare a set of counterfactual events against actual observed cases of conflict beyond simple matching or shallow definitions of "plausible areas". These "control" observations would offer an idea of what levels of civilian victimization we should expect in areas that were likely to be sites of conflict, but ultimately did not see any actual battles. In effect, this will allow us to isolate the true effect of the conflict events themselves, while partitioning out the unobservable or unmeasurable variables that are inherently correlated with both the locations of battle events and the likelihood of violence against civilians—such as a location's strategic military importance or its pre- and intra-war social networks. A plethora of studies have applied the logic of simulating conflict events as a way to create control observations with promising results (Lyall 2009; Kocher et al. 2011).

However, the way in which we ought to determine where these hypothetical control events should be located remains up for debate. To study the effect of counterinsurgent violence on rebel responses, Lyall (2009) considers all Chechen villages as plausible control points; similarly, Kocher et al. (2011) examine all hamlets (i.e., small settlements) within the Republic of Vietnam during the Vietnam War. Yet, these approaches make a crucial, unstated assumption about where we should expect to see conflict. In each strategy, it is assumed that battle events can occur in any part of the country, with the likelihood of a given location weighted by relevant spatial covariates.

This assumption has been strained by recent research into the spatial diffusion of conflict. In particular, scholars have identified the significance of road networks in determining where conflict occurs (Zhukov 2012). The availability and quality of roads are among the most crucial logistical constraints for state militaries and insurgents looking to sustain and grow their operations. Moreover, the strategic importance of roads and the settlements that lie along them makes these areas
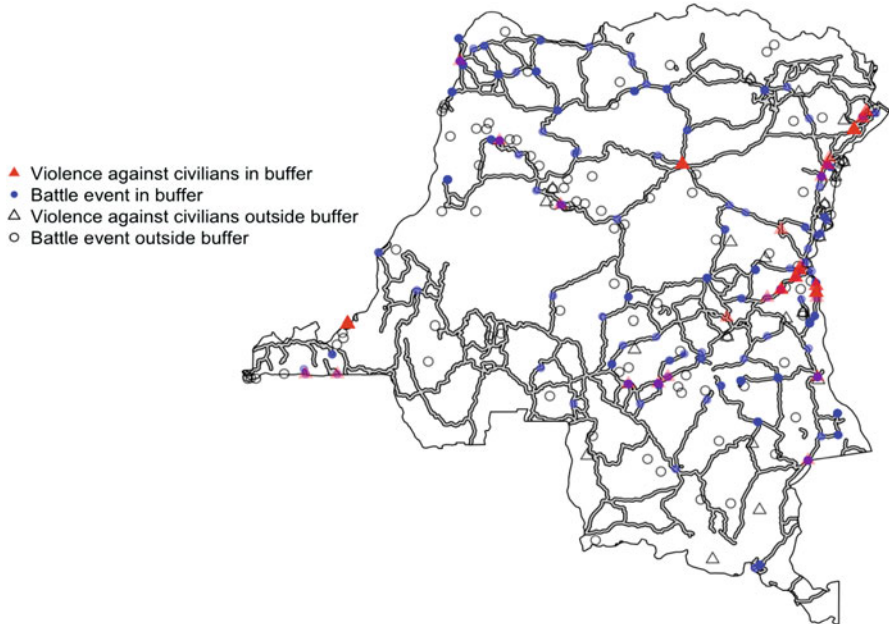
Fig. 1 Observed VAC and battle events in DRC

a hotspot for battles between warring parties seeking to gain an upper hand. Accordingly, the vast majority of armed battles take place in close proximity to roads. Indeed, in the case study of the Democratic Republic of the Congo we describe below, over 62% of all battle events occur within 5 km of a major roadway, despite the actual area of this space capturing less than 14% of the country's total land area.[1] As combatants move further away from the core road network, the costs of sustaining combat operations in more remote areas make it increasingly unlikely that armed actors will have the motivation or capacity to engage in violence (see Fig. 1).

We argue that future efforts to simulate counterfactual conflict events should acknowledge the strong clustering pattern of battles around road networks. There are at least two substantive reasons for that. Firstly, insurgencies tend to display high degrees of mobility. Insurgents—and opposing forces alike—need to quickly move between strategic objectives in order to maximize their impact. For this reason, strategic infrastructure such as roads and bridges are crucial tactical elements. Secondly, major roads normally connect not only major settlements but also strategic infrastructure (ports, airports, energy plants). While data on settlements might offer a viable alternative given that civil wars tend to occur in more populated

---

[1]Similarly, 70% of events involving violence against civilians occur within this same area around roadways.

areas, they leave out a plethora of other points of interest that can be determinant for warring parties and where civilians tend to "cluster" even when leaving settlements (e.g., after displacement or mass flee). Towards this end, we propose a simple technique for creating relevant "control" battles where none are directly available in the data. Our goal here is to create a set of points that represent locations where armed clashes are likely to occur, but have not actually taken place. Subsequent occurrences of violence against civilians near these hypothetical battle locations can then be compared to the outcomes around locations of actual battles to determine the true causal effect of the battles themselves.

Our proposal involves, first, creating a buffer area around all major roadways in the area under study. Figure 2 suggests that the particular width of the buffer is unlikely to cause any significant difference in our results for widths set at less than 10 km from the roadways. Increasing the buffer width from 1 to 10 km results in a substantial increase in the amount of land area captured in the buffer area while producing only marginal improvements in the number of true battle events captured in that same area. For our analysis, we set the width of the buffer at 5 km on each side of the road (i.e., for a total width of 10 km at any given point).

The buffer is then stored as a vector polygon and overlaid on the overall country polygon using geoprocessing tools that are commonly available today.[2] Then, to create a set of simulated events, we rely on a simple point process model to randomly


Fig. 2 Coverage sensitivity of road buffer width

---
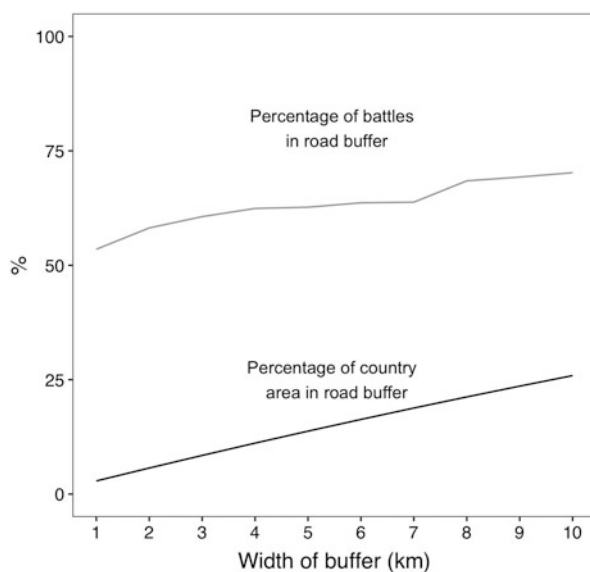
[2]In this specific case, we made use of "gBuffer" from *rgeos* (Bivand et al. 2018) package for R due to its high customizability in creating polygons around point data. Other common tools include the "Buffer" function from ESRI ArcMap.

assign locations within this buffer polygon. This is done using a uniform Poisson process within the road buffer windows with intensity (i.e., points per unit area) equaling that of the observed events. This is, of course, an approximation as we assume that each location on the road network to be a candidate for a battle. While this is in line with our theoretical expectation, we recognize that other factors (e.g., population density) might make a certain area more prone to civilian victimization than another on the same road network. Nonetheless, this simplified approach better suits the aim of this paper and, as mentioned above, places emphasis on the importance of roads. Finally, we assign dates to these synthetic events by randomly sampling the dates of actual battles, in hopes of creating a similar temporal distribution of conflict occurrence.

Figure 3 presents an illustration of this technique. In the upper left panel, the locations of observed instances of violence against civilians and battles are plotted in the case study area. The upper right pane adds major roadways throughout the
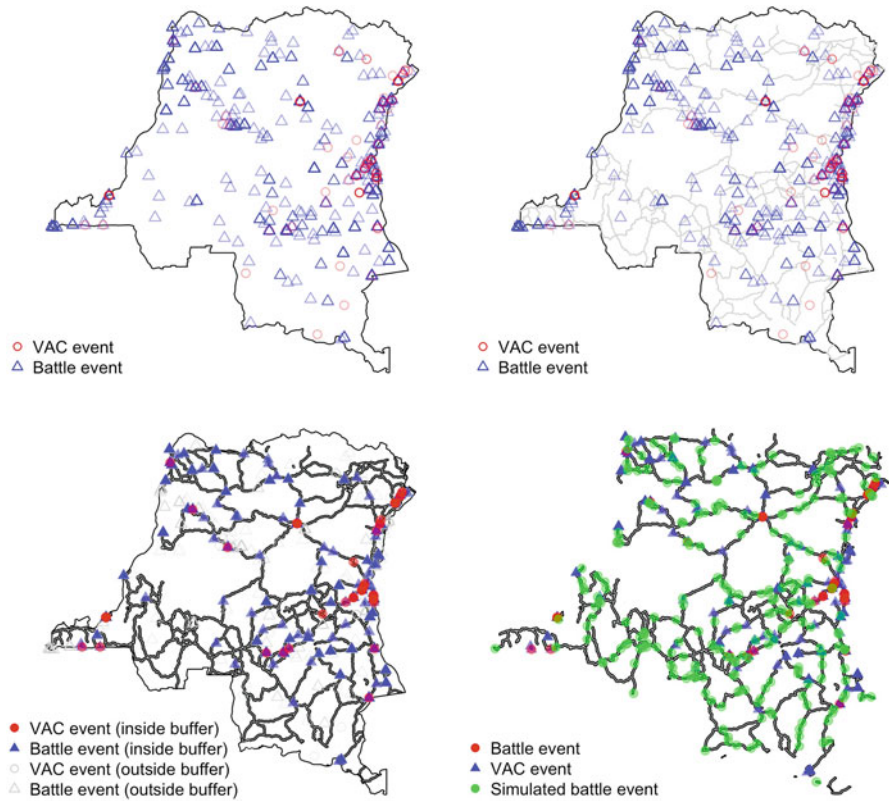
Fig. 3 Demonstration of buffer creation and conflict event simulation. The different shades of red and blue indicate the concentration of events in the same area. Darker colors correspond to areas where more events took place

country. The bottom left panel adds a 5 km buffer around those roadways and then excludes the observed battle and VAC events that occur outside this buffer area. This idea is formalized in the bottom right panel, where those remote events are dropped entirely and the relevant polygon becomes the road buffer area. Finally, in the same bottom right panel, a number of point locations are added within the buffer area to serve as simulated battles events for the modelling stage.

After creating the simulated events within the buffer area and dropping observed events outside this polygon, we proceed to model civilian victimization by taking counts of these two classes of events as our predictors. Here, the number of simulated events act as a control group while the observed events represent a treatment condition.

At this point, since our proposed approach does not fundamentally alter either the dependent variable or the nature of the point-location independent variables, there are numerous spatial models suitable for estimating the causal effect. We discuss our chosen modelling approach, matched wake analysis, in Sect. 5.

Ultimately, we argue that by simulating conflict events only in close proximity to roadways, we are able to offer a more plausible counterfactual when assessing the effects of actual battle events.

4   Data and Case Selection

To test the method described above, we carry out an in-depth analysis of conflict processes in the Democratic Republic of the Congo (DRC) from 1998 to 2000—a period capturing the earlier portion of the Second Congo War—using data from the Armed Conflict Location and Event Data Project (ACLED) (Raleigh et al. 2010). ACLED includes spatially-tagged observations of both battle events and instances of violence against civilians. Each observation is coded using press reports from a range of local and national sources. Previous studies into the relationship between conflict and civilian victimization have made similar use of ACLED data (Raleigh 2012).

In the analysis below, we focus our attention on battle events as an independent variable explaining the occurrence of violence against civilians. In the ACLED data, battle events include any "violent interaction between two politically organized armed groups at a particular time and location" that may or may not result in a change of territorial control (ACLED 2017, p. 8). Here, we exclude what ACLED terms "remote violence," or conflict events where the combatants are not physically present at the location of the violence as a result of using remote technologies such as improvised explosive devices (IEDs) or missile attacks. We removed this category as identifying the planned target of remote violence is not immediate due to the presence of collateral victims (most often civilians). Furthermore, identifying the perpetrator is quite an undertaking and data on this is far from complete. Violence against civilians is defined as "a deliberate violent act perpetrated by an organized armed group against unarmed non-combatants" (ACLED 2017, p. 10). For both of

these event types, we remain agnostic about the actors that are engaged in the battles or perpetrating the violence against civilians and include all cases of both rebel and state conflict.

The DRC offers a useful case study to demonstrate our approach for several reasons. First, violence against civilians in the DRC exhibits a strong clustering pattern around road networks. The DRC also has a large country area, which helps to illustrate how small areas around roads are crucial to conflict occurrence. Moreover, it is a case that has experienced persistent conflict over many years, with an ample number of observations of both battle events and violence against civilians. Figure 4 shows a simple comparison between battle events and VAC events in the full ACLED sample (1997–2018) across the whole African continent with the DRC, Sudan, and Somalia having the highest levels of both measures.[3] In the DRC, the magnitude of violence peaked during our temporal window between 1998 and 2000, the first two years of the Second Congo War. Finally, this case is especially informative given the relative mobility with which the conflict unfolded. In particular, the first phase of the conflict was highly "road-intensive." Belligerents sought to move quickly throughout the country to seize strategic locations, movement which frequently occurred via major roadways.
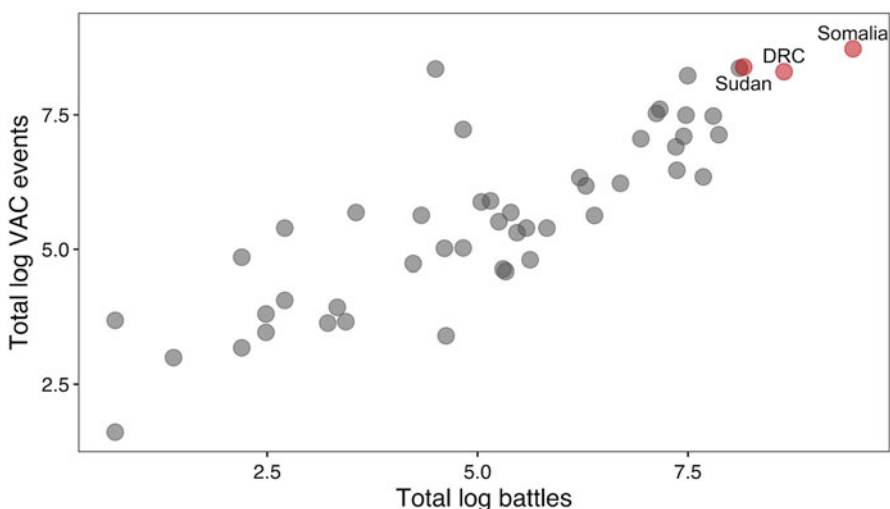
**Fig. 4** Battle versus VAC events by country in ACLED Africa data, 1997–2018

---

[3]For an analysis based on ACLED data on South Sudan, see Kelling and Lin in the chapter "Analysis of Conflict Diffusion Over Continuous Space" of this volume; for an agent-based model simulating the conflict in Somalia, see Duffy et al. in the chapter "Rebel Group Protection Rackets: Simulating the Effects of Economic Support on Civil War Violence".

5 Modeling and Results

In order to test the feasibility of our approach and re-evaluate the causal effect of armed battle events, we rely on Matched Wake Analysis (Schutte and Donnay 2014). This modelling framework combines techniques for causal inference that allow us to evaluate the impact of our treatment on the dependent variable against a control group in a continuous temporal and spatial window. Matched Wake Analysis (MWA) relies on a combination of a Sliding Windows Design, Statistical Matching and Difference-in-Differences approach. This technique has been successfully applied in previous conflict-related empirical studies (e.g., Schutte 2017).

In the MWA framework, all events are first classified as either "treatment" or "control". In our case, these correspond to the observed and synthetic battles, respectively. These georeferenced data are then linked to any number of geospatial covariates through nearest neighbor mapping. A balanced sample is then generated by matching on both the covariates and on pre-treatment trends in the dependent variable using Coarsened Exact Matching (CEM). Finally, a difference-in-differences design is used to estimate the treatment effect (Schutte and Donnay 2014).

Figure 5 provides a graphical representation of this approach showing two units (depicted as cylinders), one with a treatment and one with a control event. The square in the left-side cylinder represents the occurrence of an observed battle event, while the triangle in the right-side cylinder represents a simulated battle. The stars depict single occurrences of VAC, our dependent variable, both as "prior activity" (before the observed or simulated battle, pictured in the lower end of each cylinder) and as "posterior activity" (after the battle, portrayed in the upper end of each cylinder). At the bottom are the relevant spatial covariates on which units are matched.
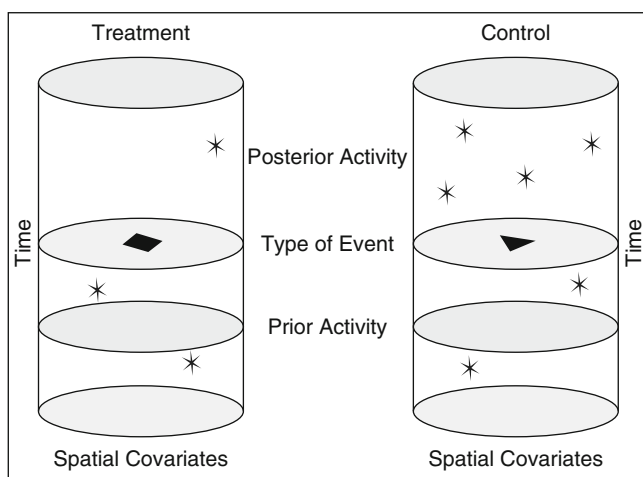
Fig. 5 Demonstration of matched wake analysis. Figure from Schutte and Donnay (2014), reprinted with permission from the authors

In our study, we estimate the following model:

$$n_{\text{post}} = \beta_0 + \beta_1 n_{\text{pre}} + \beta_2 \text{observed battles} + u$$

Here, n_{post} represents our dependent variable: a count of observed instances of post treatment VAC. β_2 represents the average treatment effect of observed battles, while β_1 is the coefficient associated to n_{pre}, which accounts for the effect of pre-treatment levels of violence against civilians. As discussed above, CEM matches samples on the pre-treatment trends in VAC and on other spatial covariates. These control variables have been selected in accordance with the relevant literature on civil conflict discussed above. We include data on population (2000) from Gridded Population of the World (GPW4 2018) and the number of ethnic groups in the area from GeoEPR (Wucherpfennig et al. 2011; Vogt et al. 2015). Furthermore, we compute the distance from the capital city and the elevation[4] associated to each point.

Figure 6 depicts the estimated values of β_2—the average treatment effect of observed battles—for each space and time window.[5] We set the spatial window as a range from 0 to 50 km and the time window as a range from 0 to 50 days. These intervals were chosen in order to capture the effect of battles in a fairly immediate time frame and within a "local" spatial domain. Larger distances or temporal periods would begin to introduce the possibility that variables that are either included in our model or impossible to measure will begin to bias our results. In the 50 day–50 km design, we are still able to test our model over a set of spatial-temporal windows that can help answer our main research questions and produce important policy implications.

As shown in Fig. 6, the treatment effect is positive across the whole window, suggesting that increases in instances of VAC occurred after observed battles took place, as compared to our synthetic battles between belligerents. While this is not surprising from a theoretical standpoint, the distribution of statistically significant effects reveals where we can be more confident about the effect of battle events.

In particular, Fig. 6 reveals a positive and significant effect that depends equally on time and distance from the battle event. The earliest period we see a significant effect on VAC is in areas relatively nearby the battle event. Crucially, however, the effect of battles increases as the distance from the battle location grows. At shorter distances from battles (from 0 to 22.5 km), VAC is likely to increase between roughly 22.5–37.5 days after a battle event. As the distance from the treatment increases (i.e., the interval between 22.5 and 42.5 km), we see an increase in the temporal range in which we would expect to see an increase in VAC (roughly 27.5–42.5 days). After 42.5 km, all time-intervals from 27.5 to 50 days are significant.

---

[4]Computed and subset from terrain data by USGS and NGA: GMTED2010 (Danielson and Gesch 2011).

[5]The precise values of each cell, along with their associated p-values can be found in Table 1 at the end of this document.
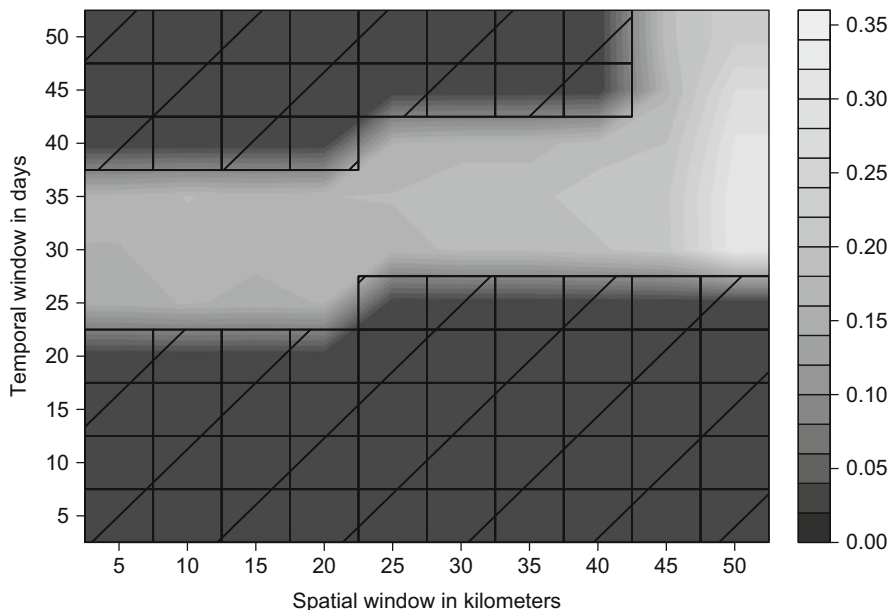
Fig. 6 Results of the MWA analysis of VAC in DRC (1998–2000). The contour plot represents the effect of observed battles on the occurrence of violence against civilians (with the scale on the right indicating the magnitude of the effect). The area not covered with crossed lines correspond to $p \leq 0.05$

This positive correlation between the timing and distance of the effect accords with an intuitive story of conflict diffusion. The results suggest that, after a battle event occurs, belligerents do not engage in VAC in the immediate aftermath. As the existing literature suggests, after taking control over a battle area combatants tend to invest resources in reinforcing their positions rather than seeking out civilians for immediate retribution. Conversely, the losing side will often either retreat entirely—thus not being able to engage in VAC—or adopt a 'wait and see' behavior to determine the stability of the new local power arrangement. Therefore, it is only after roughly 25 days that we see a surge in civilian victimization and only in an area fairly close to the battle location. The belligerents are most likely starting to secure the area with 'pseudo-policing' operations to tighten their clutch on the territory. Quite naturally, as the days pass, violence then spreads spatially as combatants seek to expand their base of control to surrounding populations. It is only at this point that we begin to detect a statistically significant effect in the areas more than 40 km from the battle location.

As for the magnitude of these effects, the average over all significant combinations is 0.21 (Table 1). That is, when comparing the treatment and control groups on average, for every battle event, we should expect to observe 0.21 additional instances of violence against civilians in the particular significant spatial-temporal areas identified above. While actual battles seem to yield more subsequent civilian victimization, the buffer space represents an area of high-risk, where most of the

clashes between belligerents and VAC events take place. The moderate, yet positive, effect of the treatment as compared to the control group shows that the synthetic counterfactuals are "plausible" candidate locations for conflict, thus confirming our prior expectations.

Table 1 Combinations of temporal and spatial areas showing the associated cut-points in days and kilometers

Time (days)	Distance (km)	Effect size	P-value	Adjusted R-squared
25	5	0.189	≤ 0.05	0.0313
25	10	0.202	≤ 0.05	0.0352
25	15	0.195	≤ 0.05	0.0335
25	20	0.201	≤ 0.05	0.0337
30	5	0.199	≤ 0.05	0.0336
30	10	0.212	≤ 0.05	0.0379
30	15	0.204	≤ 0.05	0.0357
30	20	0.21	≤ 0.05	0.036
30	25	0.211	≤ 0.05	0.0363
30	30	0.224	≤ 0.05	0.0394
30	35	0.224	≤ 0.05	0.0394
30	40	0.237	≤ 0.05	0.0435
30	45	0.254	≤ 0.05	0.049
30	50	0.356	≤ 0.05	0.0583
35	5	0.207	≤ 0.05	0.035
35	10	0.221	≤ 0.05	0.0394
35	15	0.213	≤ 0.05	0.0374
35	20	0.219	≤ 0.05	0.0376
35	25	0.222	≤ 0.05	0.0382
35	30	0.235	≤ 0.05	0.0415
35	35	0.235	≤ 0.05	0.0415
35	40	0.249	≤ 0.05	0.0458
35	45	0.257	≤ 0.05	0.0488
35	50	0.356	≤ 0.05	0.0588
40	25	0.2	≤ 0.05	0.0436
40	30	0.211	≤ 0.05	0.0466
40	35	0.211	≤ 0.05	0.0466
40	40	0.232	≤ 0.05	0.0547
40	45	0.24	≤ 0.05	0.0586
40	50	0.344	≤ 0.05	0.0638
45	45	0.215	≤ 0.05	0.0779
45	50	0.321	≤ 0.05	0.0697
50	45	0.235	≤ 0.05	0.0835
50	50	0.274	≤ 0.05	0.1002

Each row records a single window along with the associated size of the effect, p-values, and R-squared

It is important to note that we also tested our buffer model against an "any-where/anytime goes" model where simulated points were generated across the whole country and conflict timespan. This approach yielded very different results in terms of statistical significance as a result of less efficient artificial counterfactual events. In particular, this baseline model showed an overabundance of significant areas across the whole spatio-temporal window with a less clear pattern in both space and time. This suggests that a less restrictive approach to counterfactual simulation could yield an overidentification of significant results, rather than the more nuanced picture that emerges from our analysis.

6 Conclusion

This chapter has offered a preliminary step towards a new method for identifying the causal effect of armed conflict on civilian victimization. We have proposed the use of simulated conflict events as a method for creating relevant counterfactual events in cases where we can only observe actual instances of violent clashes between armed groups. By leveraging the strong clustering behavior of armed conflict around road networks, our strategy for simulating these "control" events acknowledges the underlying drivers of where true conflict activity is most likely to occur.

Using such an approach on a case study analysis of the Democratic Republic of the Congo, we found that armed battles tend to result in increased levels of violence against civilians across certain spatio-temporal windows. This increase can be observed in the immediate spatial proximity of the battle and reverberates across larger distances as well. Regarding the time component, belligerents seem to con-sistently engage in VAC after roughly 25 days from the occurrence of a battle event.

While our attention has focused on the consequences of battle events, the frame-work presented here should extend to a broader class of conflict-related independent variables that tend to diffuse and cluster along road networks. Protests, military base establishments, and remote violence, such as missile or bomb attacks, all exhibit this behavior. The analysis of the causal effects of these events could similarly benefit from a simulation-based approach to identifying relevant counterfactual observations.

There are, of course, limitations to our approach. First, our simulation strategy assumes that the coders of data on battles and VAC events are correctly assigning those events to their exact locations. When conflict events occur in remote areas, it may be the case that either media sources or dataset coders elect to record the location of those events as a nearby settlement rather than the true coordinates. Normally, such a decision would represent a small amount of measurement error, but, in our framework, it is crucial to modelling outcomes. Since most settlements lie along road networks, the assignment of observations to nearest settlements could inflate the amount of conflict occurring within our road buffer polygons, thus biasing estimates of the true causal effect in these areas. Future approaches may encompass a "hybrid buffer" weighted on the basis on existing settlement.

It is unclear whether this type of error occurs in commonly used event datasets. As Eck (2012) has shown, these datasets, when they are created from media sources, tend to be biased towards greater coverage of urban areas in general. The ACLED coding guidelines are fairly robust against coders misassigning observations and the dataset does include a variable indicating the precision of the coordinates (ACLED 2017, p. 25–26). In future applications, this variable could be used to filter out low-certainty observations. However, these checks at the coder-level would not prevent misreporting by the media sources further upstream (i.e., non-local, national, and international media). Thus, our approach is not immune from the literature's ongoing concerns about the precision of spatial conflict data.

Second, our proposed framework makes a strong assumption about where simulated events should be located. A simple implementation of our approach, as shown above, ensures that no simulated events will fall outside the roadway buffer area. A more nuanced approach, which would remain methodologically consistent with our general framework, could involve a probabilistic model to simulate events. Using this strategy, the likelihood that a simulated event is tagged at a given point would decline exponentially as the distance of that point from a major roadway increases. Here, the general principle of roadways as crucial to explaining the spatial distribution of conflict is maintained, while allowing for greater variance in the simulated locations.

Third, by excluding battle events far from roadways (or, downweighting their likelihood of occurring in probabilistic simulations), we run the risk of overlooking a possible interaction effect between armed battle events and their distance from population centers. That is, there is a possibility that battles may provoke greater civilian victimization when those battles occur in remote areas. Combatants may assume that attacks on civilians will be less publicized or less likely to result in retribution if they are done in the hinterland. To our knowledge, this scenario remains untested in the existing literature, but given its relevance to the present research design, we recommend future studies to investigate the possibility of such an interaction effect. Under our current approach, excluding these remote observations precludes the possibility of detecting this type of causal relationship.

With these limitations in mind, we argue that simulating counterfactual conflict events along road networks offers a defensible and advantageous strategy for causal identification in observational events data. Simply put, we want future researchers to understand that they can and should try to find better counterfactuals for battle events. Doing so will require creative thinking about where conflict is likely to occur. Extending this approach to other cases would also help to establish whether these effects are representative of civil conflict dynamics in general or are instead case-specific.

In this chapter, we have proposed one solution of using road buffers as a simplifying condition to simulate counterfactual conflict events; infinitely more options also exist that could fit within our framework. We hope conflict scholars will continue to improve upon our approach and take up with greater zeal the potential offered by synthetic event simulation in conflict studies.

References

ACLED. (2017). *Armed conflict location & event data project (ACLED) codebook*, 2017, version 8.

Bivand, R., Rundel, C., Pebesma, E., Stuetz, R., Hufthammer, K. O., Giraudoux, P., et al. (2018). *rgeos: Interface to geometry engine - open source ('GEOS')*. R package version 0.3–24. Retrieved from https://CRAN.R-project.org/package=rgeos

Danielson, J. J. & Gesch, D. B. (2011). *Global multi-resolution terrain elevation data 2010 (GMTED2010)*. U.S. Geological Survey Open-File Report 2011–1073.

Eck, K. (2012). In data we trust? A comparison of UCDP GED and ACLED conflict events datasets. *Cooperation and Conflict, 47*(1), 124–141.

Fjelde, H., & Hultman, L. (2014). Weakening the enemy: A disaggregated study of violence against civilians in Africa. *Journal of Conflict Resolution, 58*(7), 1230–1257.

Gridded Population of the World. (2018). *Version 4 (GPWv4): Population count, revision 11*. Center for International Earth Science Information Network - CIESIN - Columbia University. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC).

Hultman, L. (2007). Battle losses and rebel violence: Raising the costs for fighting, terrorism, and political violence. *Terrorism and Political Violence, 19*(2), 205–222.

Kalyvas, S. N. (2006). *The logic of violence in civil war*. Cambridge: Cambridge University Press.

Kocher, M. A., Pepinsky, T. B., & Kalyvas, S. N. (2011). Aerial bombing and counterinsurgency in the Vietnam War. *American Journal of Political Science, 55*(2), 201–218.

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

Lyall, J. (2017). *Bombing to lose? Airpower, civilian casualties, and the dynamics of violence in counterinsurgency wars*. Retrieved at SSRN: https://ssrn.com/abstract=2422170. https://doi.org/10.2139/ssrn.2422170

Raleigh, C. (2012). Violence against civilians: A disaggregated analysis. *International Interactions, 38*, 462–481.

Raleigh, C., Linke, A., Hegre, H., & Karlsen, J. (2010). Introducing ACLED—Armed conflict location and event data. *Journal of Peace Research, 47*(5), 651–660.

Schutte, S. (2015). Geographic determinants of indiscriminate violence in civil wars. *Conflict Management and Peace Science, 34*(4), 380–405.

Schutte, S. (2017). Violence and civilians loyalties: Evidence from Afghanistan. *Journal of Conflict Resolution, 61*(8), 1595–1625.

Schutte, S., & Donnay, K. (2014). Matched wake analysis: Finding causal relationships in spatiotemporal event data. *Political Geography, 41*, 1–10.

Schwartz, R. A., & Straus, S. (2018). What drives violence against civilians in civil war? Evidence from Guatemala's conflict archives. *Journal of Peace Research, 55*(2), 222–235.

Stoll, D. (1993). *Between two armies in the Ixil towns of Guatemala*. New York: Columbia University Press.

Sullivan, C. M. (2012). Blood in the village: A local-level investigation of state massacres. *Conflict Management and Peace Science, 29*(4), 373–396.

Valentino, B., Huth, P., & Balch-Lindsay, D. (2004). "Draining the sea": Mass killing and guerrilla warfare. *International Organization, 58*, 375–407.

Vogt, M., Bormann, N. C., Rüegger, S., Cederman, L. E., Hunziker, P., & Girardin, L. (2015). Integrating data on ethnicity, geography, and conflict: The ethnic power relations data set family. *Journal of Conflict Resolution, 59*(7), 1327–1342.

Wood, R. M. (2010). Rebel capacity and strategic violence against civilians. *Journal of Peace Research, 47*(5), 601–614.

Wucherpfennig, J., Weidmann, N. B., Girardin, L., Cederman, L. E., & Wimmer, A. (2011). Politically relevant ethnic groups across space and time: Introducing the GeoEPR dataset. *Conflict Management and Peace Science, 28*(5), 423–437.

Zhukov, Y. M. (2012). Roads and the diffusion of insurgent violence: The logistics of conflict in Russia's North Caucasus. *Political Geography, 31*, 144–156.