Territorial control in civil wars: Theory and measurement using machine learning

Therese Anders*

21 April 2019

Comments on the main manuscript and the methodology of measuring conflict exposure upon using spatial and temporal decay functions to aggregate conflict events to grid cells (Appendix A) are highly appreciated.

Abstract

Territorial control is a central variable in the study of civil wars — yet, we lack data that are fine-grained enough to capture subnational dynamics and offer cross-country coverage. The paper advances an innovative measurement strategy for territorial control in asymmetric civil wars. Territorial control is conceptualized as an unobserved latent variable that can be estimated via observed variation in rebel tactics. The measurement strategy builds on a theoretical model of rebel tactics by which rebels use terrorism less when they control a given area — preferring conventional tactics, which require higher levels of territorial control. The latent variable territorial control is estimated via a Hidden Markov Model (HMM). I leverage geo-coded event data and use a function of the relative frequency of terrorist attacks and conventional war acts, weighted by time and distance, as an observable indicator for rebel tactics. The model yields monthly estimates of territorial control for asymmetric civil wars at a resolution of 0.25 degree minimum diameter hexagonal grid cells. The validation of the estimates for the Colombian and Nigerian civil wars suggests HMMs as a fruitful avenue to estimate spatiotemporal variation of territorial control.

*Ph.D. candidate, School of International Relations, University of Southern California (tanders@usc.edu).

Securing control over territory is a key objective for actors in violent conflict. Those who exert control over an area have the opportunity to extract resources (Carter, 2015), pursue the collaboration of the population (Arjona, 2016; Kalyvas, 2006), and increase their mobilization base (Stewart and Liou, 2017; de la Calle and Sánchez-Cuenca, 2015). Areas of consolidated control can serve as safe havens for combatants and a home base from which future offensives can be coordinated and launched (Arjona, 2016). Gaining control over an area is a pre-condition for the establishment of non-violent political order. It gives actors the ability to govern non-coercively, for example via the provision of public goods (Stewart, 2018). Territorial control is thus a central variable of interest for the study of intrastate conflict dynamics (Staniland, 2012; Sambanis, 2004).

Many intrastate conflicts, in particular civil wars that feature a high power asymmetry between rebels and the government, do not exhibit clearly defined front lines. Territorial control in these asymmetric civil wars tends to be spatially fragmented and difficult to measure. To date, we lack data on territorial control that offer wide temporal and cross-sectional coverage and are sufficiently fine-grained to account for variation on the subnational level. This shortage of information causes territorial control to be astonishingly absent as a variable in micro-level studies of civil war — despite its centrality as a theoretical concept. Given the difficulty of measuring territorial control in asymmetric conflicts via direct observation, *how can we estimate changes in territorial control across time and space?* I advance a novel measurement strategy for territorial control in asymmetric intrastate conflict. I show that we can estimate territorial control by translating a theory of actor behavior in civil war into a machine learning model and leveraging information on variation in rebel tactical choice based on geo-coded event data.

Building on existing work regarding the relationship between territorial control and tactical choices of insurgents, I develop a theoretical model that links the relative frequency of terrorist attacks and conventional war acts between government forces and the rebels to patterns of territorial control. The measurement strategy builds on two empirical relationships: 1) rebels use terrorism predominantly outside their strongholds; 2) preferring conventional guerrilla tactics when they command higher levels of control. Hence, we observe more insurgent terrorist attacks relative to conventional fighting in areas exhibiting a higher level of government control, and vice versa. Translating this theoretical relationship into measurement, I employ a function of an area's spatially and temporally weighted exposure to terrorism and combat events as observable emissions from the latent variable territorial control.

I treat territorial control as a latent variable that can modeled and estimated. Following Kalyvas (2006), territorial control is conceptualized as a categorical variable ranging from complete rebel control to complete government control, with levels of contestation in between. The latent variable territorial control is estimated via a Hidden Markov Model (HMM). HMMs compute the most likely evolution of territorial control in an area over time, conditional on observed rebel tactics and model priors. These priors specify beliefs about how likely a cell changes from one state of territorial control to another (transition probabilities) and how accurately observed rebel tactics measure the latent state of control (emission probabilities). Contrary to dominant applications in computer science, in this application, priors are not learned from training data but instead informed by theoretical arguments and out-of-sample empirical observations. As a proof of concept, I present estimates of territorial control for the conflict between the *Fuerzas Armadas Revolucionarias de Colombia - Ejército del Pueblo* (FARC) rebels and the Colombian government from 2006 to 2017 and the Boko Haram insurgency in Nigeria from 2008 to 2017.

The project yields three main contributions. First, I provide fine-grained data on territorial control that accommodate high levels of spatial and temporal variation but are produced with a methodology that can be applied cross-nationally. Second, I show how conflict scholars can use their rich theoretical knowledge to inform priors in machine learning applications. Finally, I develop a new approach toward the measurement of an area's exposure to conflict events. Rather than discretely assigning conflict events to grid cells, I compute a cell's exposure as the spatially and temporally weighted sum of conflict events. Not only does this approach account for the spatial dependence of conflict events in the estimation of the HMM

— it presents a valuable methodology for event-based subnational conflict analyses beyond the measurement of territorial control.

The paper proceeds as follows. First, I establish the centrality of territorial control for the study of intrastate conflict and discuss previous measurement attempts. I then formulate a theoretical account of actor behavior in civil war through which I relate territorial control to tactical choices of insurgents, specifically the use of domestic terrorism versus conventional fighting. Leveraging conflict event data on the observed relative frequency of terrorist attacks and violent incidents that are indicative of guerrilla warfare, I obtain estimates of territorial control in a machine learning framework. Finally, I conduct out-of-sample validation to show that the methodology can recover fine-grained spatiotemporal patterns of territorial control.

Territorial control in civil war

Territorial control is a crucial variable for understanding violent conflict, as it shapes armed actors' tactics and aspirations, in particular "the dynamics of bargaining, recruitment, and lethality" (de la Calle and Sánchez-Cuenca, 2012, 583). Who commands what level of control has also been shown to condition civilian behavior in conflict zones, such as voting (García-Sánchez, 2016) and information sharing (Arjona, 2016). I define territorial control as the "extent to which actors are able to establish exclusive rule on a territory." (Kalyvas, 2006, 111) For the purpose of measurement, I conceptualize territorial control as a five-category variable ranging from full rebel to full government control. It is assumed to be zero-sum: In any given area, increases in the level of territorial control exercised by one actor equal decreases for competing actors.

In the existing literature, territorial control is most prominently studied as a factor determining selective versus indiscriminate victimization of civilians by government or non-state actors (e.g. Stewart and Liou 2017; Quinn 2015; Kalyvas and Kocher 2009). Generally speaking, the less territorial control actors command, the more indiscriminate the violence they inflict will be, and vice versa. Other work considers the interplay between civilian cooperation and armed actor coercion in the establishment and consolidation of territorial control (Arjona, 2016).

In cross-national studies, scholars frequently rely on a binary indicator from the Non-State Actors in Armed Conflict Dataset (NSA) to code whether a group exercised control over territory (Cunningham et al., 2013). The NSA data additionally rate how effectively non-state armed actors exercise control. However, the information is supplied at the group-level and is unable to capture temporal and subnational variation. Alternative approaches that operationalize government territorial control as a function of the distance from a country's capital vary subnationally, but do not allow for temporal variation at a given location (Schutte, 2017).

One of the key characteristics of asymmetric civil war is the absence of clearly defined front lines. Rebels that are weak compared to the government tend to avoid direct contact with state forces and "try to disperse as much as possible so that the state cannot respond to the multipronged challenge." (Arjona, 2016, 43) Neither the operationalization of territorial control via binary actor-level indicators, nor distance to the capital city, can account for the "messy patchwork" patterns (Kalyvas, 2006, 88) that are observed in empirical studies of asymmetric conflict. Published examples of country-specific accounts documenting this fragmentation include recent efforts to create estimates of territorial control using conflict event data for Liberia (Tao et al., 2016), post- or in-conflict surveys (Kalyvas and Kocher, 2009; Arjona, 2016), or the study of historic military records and interviews (Kalyvas, 2006). The approach by Tao et al. (2016) and Aronson et al. (2017) to hand-code assault initiators and the status of territorial control after attacks based on media reports underlying the Georeferenced Event Dataset (GED) produces fine-grained estimates that can be constructed for a cross-section of conflict zones. However, it is extremely labor intensive and, at the time of writing, no data has been publicly released.

While recent years have seen increased interest in studying territorial control in asymmetric civil war, the field suffers from a shortage of available data that vary subnationally and temporally, can recover "patchy" patterns of control, and are produced with a methodology that accommodates cross-country comparison. I improve upon existing approaches by conceptualizing territorial control as a latent variable that can be estimated for small subnational spatial and temporal units using publicly available event data. The methodology can feasibly be applied cross-nationally but provides sufficient detail for within-country analyses. The model set-up is based on a theoretical account that links observable variation in rebel tactics — their choice between irregular terrorist and more conventional guerrilla fighting strategies — to territorial control in the context of asymmetric civil wars.

Tactical choice in asymmetric civil war

"*...[I]nsurgency is almost entirely terroristic.*" (Schelling, 1960, 27)

Terrorism and civil war are not separate phenomena and instead co-occur frequently (Bakker et al., 2016; Fortna, 2015). Only a small share of terrorist incidents are perpetrated by groups that are specialists in this tactic and terrorism is commonly used as one among many possible forms of violence within home territories (Tilly, 2004). In fact, it has been shown that"most incidents of terrorism take place in the geographic regions where civil war is occurring and during the ongoing war." (Findley and Young, 2012, 286) Within conflict zones, we observe large variation in the degree of overlap between terrorist attacks and events that are indicative of conventional insurgent tactics. The micro-level separation of terrorist attacks and conventional war fighting can be explained by tactical choices of rebels.[1]

Once a dissident group has made the strategic choice between violent versus non-violent resistance, the question of terrorism versus non-terror violence is a matter of tactics (Bakker et al., 2016). The tactical choice for insurgents is to either attack states' armed forces directly or indirectly target the government via coercive action intended to spread fear among the public (Carter, 2015, 117). Polo and Gleditsch (2016, 816) state that while definitionally, the two concepts are not mutually exclusive and hard to delineate, "[t]errorism [. . .] differs from conventional attacks in civil conflicts in that the immediate targets or victims are typically

[1] I use the term *conventional* war fighting with respect to tactics that are conventionally used by rebels in asymmetric civil wars, such as small battles with the government, ambushes, and hit-and-run attacks, not with regard to the usage of the term in international humanitarian law.

non-combatants, and each individual victim is normally less important than the purpose of conveying a message to the intended audience.” I follow the previous literature in stipulating three conditions that have to be met for a violent event to be coded as terrorism, rather than non-terror violence. To qualify as terrorism, the violent action must seek to convey a political message to an audience broader than the immediate targets of the attack, does not directly target the military capability of the state, and lie outside the realm of “legitimate warfare activities,” including but not limited to, the targeting of noncombatants (START, 2016, Bakker et al. 2016; Chenoweth 2013).

Territorial control is a key factor in explaining insurgents’ use of terrorism as opposed to more conventional guerrilla tactics (Carter, 2015; de la Calle and Sánchez-Cuenca, 2015). Terrorism arises from insurgents’ inability to control territory (Asal et al., 2012). Rebel territorial control is associated with guerrilla tactics such as “hit-and-run attacks, ambushes, raids, and small-scale battles,” however, when forced to remain underground, those same groups rely predominantly on bombings and assassinations (de la Calle and Sánchez-Cuenca, 2015, 810).

Tactical choices in civil war echo actors’ maximization of benefits and minimization of costs, subject to resource constraints and the actions of the opponent. All else equal, rebels prefer conventional tactics over terrorism for two reasons. First, in a quest to indirectly pressure the government by inflicting pain and fear among the population, terrorist campaigns run a high risk of alienating civilians whose support rebels depend upon. Second, terrorism does not aid insurgents’ immediate goal of securing territorial gains (Carter, 2015, 130). Terrorism is therefore a second-best choice of tactics for rebels that do not command control over a given area and are unable to engage in direct fighting with government forces. Terrorism “allows dissidents to avoid direct, costly strikes on government forces that are typically superior in numbers and weaponry.” (Hendrix and Young, 2014, 335)

Territorial control is qualitatively different from rebel strength. Territorial control measures the degree to which rebels or the government rule over an area without interference from opposing actors. Rebel strength refers to the size of a group or its material capability. However, a group's military power and territorial control are related. Military power, the power to destroy, can be distinguished from coercive power, that is the power to hurt (Schelling, 1960). The less territory a group controls, the more it will rely on coercive, as opposed to military power (de la Calle and Sánchez-Cuenca, 2015, 797). In environments characterized by low state capacity, armed actors are more likely to adopt conventional tactics, while groups facing more capable governments are likely to resort to terrorism (see de la Calle and Sánchez-Cuenca in Asal et al. 2012, 482).

The link between rebel tactical choice and territorial control can be observed empirically. Evidence from Nigeria suggests that once the government was able to re-capture insurgent strongholds in 2015, Boko Haram moved away from fighting for territory and intensified "its campaign of suicide bombings against soft targets."[2] Figure 1 overlays a coarse map of territorial control in Northeast Nigeria with the location of Boko Haram terrorist attacks (blue triangles) and conflict events that are indicative of conventional fighting (red dots) within the two weeks following the measurement of territorial control.[3] The map shows conventional fighting to be clustered in contested areas and along the borders of insurgent-held territory. With the exception of isolated events in the border region with Cameroon in the West, terrorist events were limited to areas of government control.

[2] https://reliefweb.int/report/nigeria/analysis-scrutinising-boko-haram-resurgence, accessed 18 August 2018.

[3] The map is adapted from Reuters, see http://blogs.reuters.com/data-dive/2015/05/05/mapping-boko-harams-decline-in-nigeria/, accessed 24 October 2018. To the best knowledge of the author, at the time of writing, this is the most detailed information on territorial control that is publicly available for Nigeria at the height of the conflict. The map covers 32 local government areas in the Yobe, Borno, Adamawa states: Abadam, Askira/Uba, Bama, Bayo, Biu, Chibok, Damboa, Dikwa, Geidam, Gubio, Gujba, Gulani, Guzamala, Gwoza, Hawul, Jere, Kaga, Kala/Balge, Konduga, Kukawa, Kwaya Kusar, Madagali, Mafa, Magumeri, Maidugur, Marte, Michika, Mobbar, Monguno, Ngala, Nganzai, Shani. Gwoza is coded as being under rebel control on 24 April 2015 because it contains the Boko Haram stronghold in the Sambisa forest.

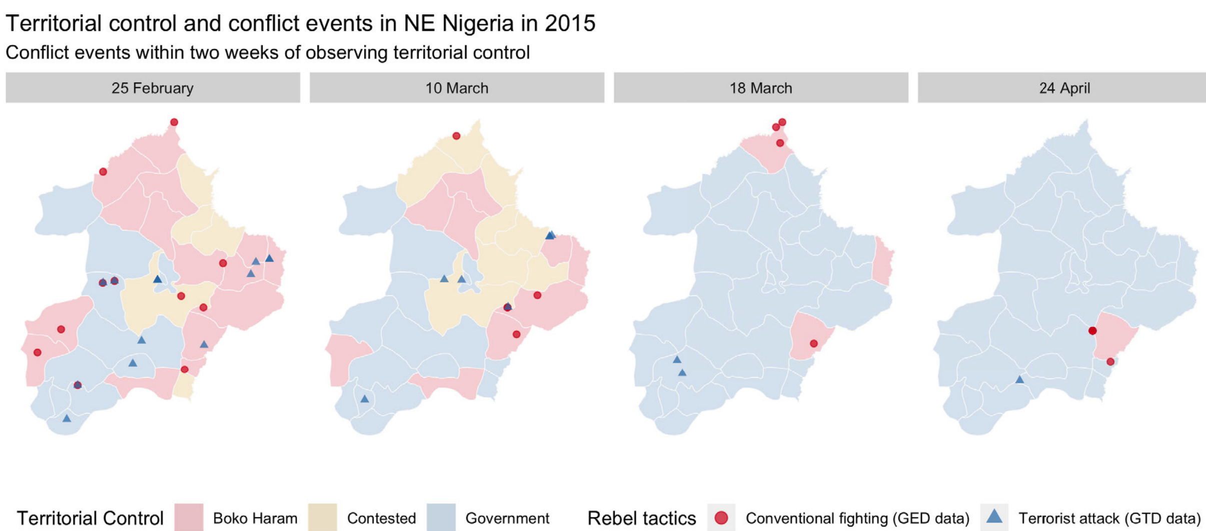


Figure 1: The map illustrates the relationship between territorial control and Boko Haram tactical choice in civil war in Nigeria in 2015.

Modeling territorial control

I argue that the observed level of terrorism relative to conventional tactics is indicative of the unobserved distribution of territorial control in asymmetric civil wars. I translate a theoretical model of the relationship between rebel tactical choice and territorial control into a measurement model. The model rests on the insight that higher levels of rebel territorial control are associated with higher levels of conventional fighting, while higher levels of government control are associated with more terrorism. In areas of complete control of either actor, violent events will be scarce.

In principle, territorial control could be operationalized along a continuum from full rebel to full government control. However, for the purpose of estimation via a discrete-state HMM below, I conceptualize territorial control as a categorical variable ranging from complete rebel ($S1$) to complete government control ($S5$), with levels of contestation in between (see Table I). These states correspond to the categorization of zones of territorial control in the existing literature (Kalyvas and Kocher, 2009; Kalyvas, 2006).

Territorial control Q	Description
$S1$	Full rebel control
$S2$	Contested, closer to rebel control
$S3$	Highly contested
$S4$	Contested, closer to government control
$S5$	Full government control

Table I: Set of possible states $\mathbf{Q} = \{S1, S2, S3, S4, S5\}$ of the latent variable territorial control.

Measuring rebel tactics

Rebel tactics are observable emissions of the unobserved latent variable territorial control. I operationalize rebel tactics in a given area as a function of that area's relative exposure to terrorist attacks versus events that are indicative of conventional guerrilla fighting. This approach is informative regarding the mixture of tactics used in a given area. I employ a heuristic that translates a function of the relative frequency of terrorist attacks T_{it} and conventional war acts C_{it} into values of the variable of observable emissions o_{it} in area i at time t. Specifically, I compare the probability of the observed exposure to terrorist events $T_{it} = P(\lfloor E_{it}^{[T]} \rfloor; \lambda_t^{[T]})$ to the respective probability of observed conventional fighting $C_{it} = P(\lfloor E_{it}^{[C]} \rfloor; \lambda_t^{[C]})$ from a zero-inflated Poisson distribution. $E_{it}^{[T]}$ and $E_{it}^{[C]}$ are continuous measures of an area's exposure to terrorist and conventional conflict events, respectively. $\lambda_t^{[T]}$ and $\lambda_t^{[C]}$ denote the expected number of events for each tactic in a given time period t across all areas i within a country. There are four possible observations $\mathbf{O} = \{A, B, C, D\}$ of the rebel tactic variable O, as outlined in Table II.

Tactics O	Observation	Description	Comments
$o_{it} = A$	$E_{it}^{[T]} \approx E_{it}^{[C]} \approx 0$	Little to no exposure to terrorism and conventional events.	To account for coding errors and small margins, observed exposure values below a threshold xs, in the main specification $xs = 0.1$, are truncated to zero.
$o_{it} = B$	$\lvert C_{it} - T_{it} \rvert \leq m$	Similar *non-zero* exposure to terrorism and conventional fighting.	In the main specification, overlap of zero-inflated Poisson probabilities of conflict exposure is specified as $m = 0.025$.
$o_{it} = C$	$C_{it} < T_{it}$, and $\lvert C_{it} - T_{it} \rvert > m$	More exposure to terrorism than conventional fighting.	
$o_{it} = D$	$C_{it} > T_{it}$, and $\lvert C_{it} - T_{it} \rvert > m$	More exposure to conventional fighting than terrorism.	

Table II: Heuristic to translate the observed exposure to terrorist attacks and conventional war acts into the categorical variable of rebel tactics O.

I develop a continuous measure of areas' exposure to terrorist events $E_{it}^{[T]}$ and conventional war fighting $E_{it}^{[C]}$. The influence of individual conflict events on area i is modeled to dissipate continuously over space and time. I compute exposure as the sum of spatially and temporally weighted event counts for the centroid of area i at time t.[4] While the HMM computes the most likely sequence of territorial control independently for each subnational area, the use of weights allows for spatial dependence in observed rebel tactics between spatial units.

Mapping rebel tactics onto territorial control

Figure 2 illustrates the theoretical model how observed rebel tactics o_{it} relate to unobserved levels of territorial control q_{it}. The prevalence of the use of terrorism by rebels is theorized to be increasing in the level of government control. As previous research posits, "guerrillas resort to terrorist tactics when they act beyond their areas of control" (de la Calle and Sánchez-Cuenca in Asal et al. 2012, 483). Hence, the use of terrorist tactics has an inverse relationship with the level of territorial control of that group in a given area. The use of conventional tactics is increasing in the level of rebel control — suggesting more direct confrontation and hence more conventional war fighting between the two actors.[5]

Based on this theoretical model, zones of full rebel or full government control are associated with the relative absence of violence ($o_{it} = A$). Control is either undisputed, or actors successfully established exclusive rule and prevent opponents from penetrating the area. Areas that are contested, but closer to rebel control, are expected to see relatively higher levels of conventional war fighting, as opposed to terrorism ($o_{it} = D$). Here, rebels will limit their use of terrorism because they seek to reduce the amount of harm inflicted on their constituent population in an effort to minimize the risk of denunciation and maximize popular collaboration (Polo and Gleditsch, 2016). In these areas, I expect a higher level of direct war fighting between insurgents and the government because when rebel control is high but not consolidated, government forces seek to confront rebels conventionally in a quest to regain

[4] $E_{it} = \sum_{j=1}^{J} \left(w_{d_{ij}} \times w_{a_{jt}} \right)$, where $w_{d_{ij}} = 1/(1 + e^{-7+0.35 d_{ij}})$ denotes weighted distances d_{ij} from event j to the centroid of area i in kilometers, and $w_{a_{jt}} = 1/(1 + e^{-8+2.5 a_{jt}})$ weighted event ages a_{jt} in months. Weighted distances or ages below $w < 0.05$ are truncated to zero. For more detail, see section 1 in the online appendix.

[5] The model is based on the simplifying assumption that there are only two parties to the conflict: a state and a non-state armed challenger that are both treated as unitary actors. This is of course a strong, and in many conflict settings unrealistic, assumption. However, because the unit of analysis in this project is a small grid cell, this assumption does not preclude an estimation of territorial control in conflict settings with more than one non-state armed actor, as long as groups' aspirations for control do not overlap significantly.

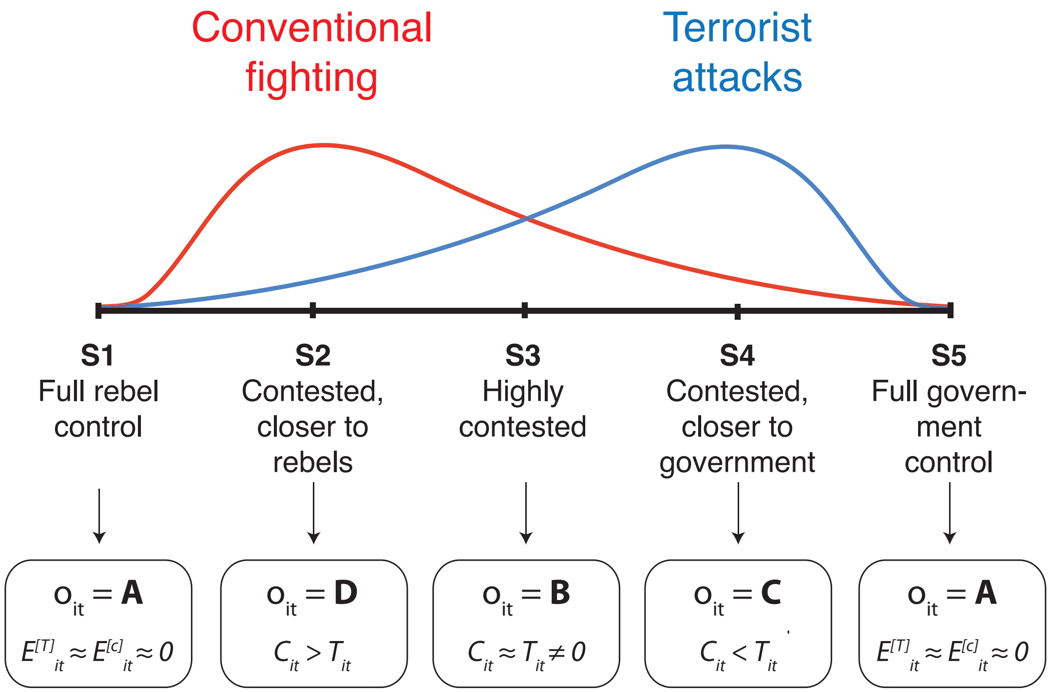


Figure 2: Theorized relationship between observed variation in rebel tactical choice and levels of territorial control.

control. Highly contested areas, that is regions where neither rebels nor the government command a high level of control, are characterized by a relative parity of conventional and terrorist events ($o_{it} = B$). Areas that are contested, but in which the government enjoys a high level of territorial control, are expected to exhibit relatively more terrorist attacks than conventional war fighting ($o_{it} = C$), because insurgents fighting a highly capable government will substitute conventional war acts with terrorism (Carter, 2015).

Estimation

I estimate territorial control via a Hidden Markov Model (HMM).[6] HMMs are graphical models that provide a method for uncovering the most likely sequence of unobserved states of a discrete latent variable given a set of observable outputs, transition, and emission proba-

[6]The model is implemented using the `HMM` package in `R` (Himmelmann, 2010). A computer science conference proceedings think piece discusses difficulties of the estimation via HMM and extension to Hidden Markov Random Field Models (Anders et al., 2017). The present manuscript is the first to develop a thorough theoretical model and compute HMM estimates of territorial control, made possible by accounting for the spatial dependence between grid cells through continuous spatial decay in observable emissions.

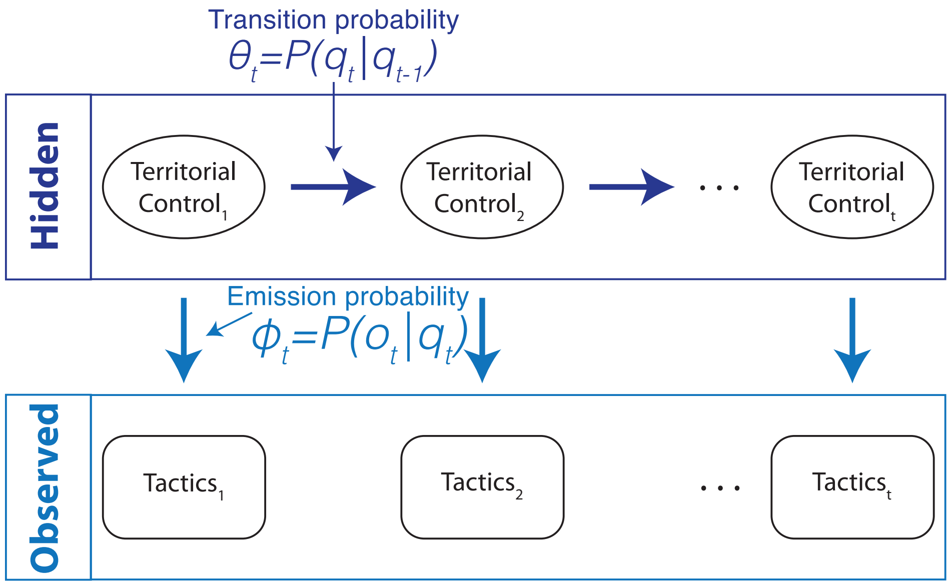


Figure 3: Graphical representation of a HMM as a Bayesian network.

bilities. The ability to map four emission states onto a five-category latent variable distinctly positions HMMs as an effective method to estimate the latent variable territorial control.

Figure 3 illustrates the conditional dependencies between the sequence of hidden states Q and the sequence of observed outputs O over time t. The model makes two assumptions. First, the value of the output o_{it}, that is the emission of cell i at time t, is generated by a process whose true state q_{it} is not observed. Second, the sequence of hidden states follows a Markov process, that is the state of the hidden variable at time t depends only on the state of the variable at $t-1$, but no prior time periods (Ghahramani, 2001). The HMM computes the maximum likelihood path of territorial control over the entire period of observation. The most likely sequence of labels is decoded via the Viterbi algorithm.[7]

While territorial control is operationalized as a five-category variable $\mathbf{Q} = \{S1, S2, S3, S4, S5\}$, the observable emissions only have four possible outcomes $\mathbf{O} = \{A, B, C, D\}$. A situation in which no terrorist attacks or conventional war acts are observed is theorized to be indicative of either full rebel or full government control. Hence,

[7]HMM maximized the function $v_t(h) = \max_{g=1}^{N} v_{t-1}(g)\theta_{gh}\phi_h(o_t)$, where h indexes current state, g indexed previous state, $v_{t-1}(g)$ indicates the path probability of previous time step, θ_{gh} denotes the transition probability from q_g to q_h, and $\phi_h(o_t)$ the emission probability given h. The Viterbi algorithm conducts the maximization step and via recursion returns the label for each unit of the most probable path (Jurafsky and Martin, 2017).

we cannot linearly map the observed indicator of rebel tactics onto territorial control. HMMs provide a solution to this problem via two unique features. First, the model allows me to specify that the latent states of full rebel control ($S1$) and full government control ($S5$) are equally likely to produce an observable output of little to no violence ($\mathrm{o}_{it} = A$). Second, HMMs maximize the most probable path over the entire sequence of observations, not a single instance in time. This allows HMMs to sort out whether an area that experiences little to no violence is more likely to be under the full rebel or full government control, given transition probabilities, emission probabilities, as well as path probability of the previous time step.

Transition probabilities

The transition matrix $\mathbf{\Theta}$ specifies the probabilities of an area transitioning from one latent state to another. Each cell in $\mathbf{\Theta}$ captures the probability θ of an area transitioning to a specific state of the latent variable territorial control q_t, given its instance in the previous period q_{t-1}. Conceptually, transition probabilities capture the volatility of territorial control. They are informative about assumptions on how the latent variable evolves over time, independent of the observed output.

				q_t		
		$S1$	$S2$	$S3$	$S4$	$S5$
	$S1$	0.250	0.500	0.025	0.200	0.025
	$S2$	0.250	0.150	0.075	0.500	0.025
q_{t-1}	$S3$	0.050	0.025	0.050	0.850	0.025
	$S4$	0.025	0.075	0.150	0.125	0.625
	$S5$	0.050	0.075	0.475	0.025	0.375

Table III: Transition matrix $\mathbf{\Theta}$, as inspired by Kalyvas (2006). Rows sum to one. Values across the diagonal indicate the probability of a grid cell remaining in the same state across two periods.

I leverage existing research to construct the matrix of transition probabilities.[8] The transition matrix in Table III is inspired by observed transitions between zones of territorial

[8]In principle, transition (and emission) probabilities could be learned from training data. However, in the present application this is infeasible because of a lack of publicly available fine-grained data on territorial control.

control during the Greek civil war (Kalyvas, 2006, 277). In the original matrix, a number of possible transitions, for example from full rebel to full government control, are never observed. This would indicate a transition probability $P(S5|S1) = 0$. However, while areas are unlikely to transition from one extreme on the spectrum of territorial control to another without at least temporarily experiencing contestation, it is not impossible. Therefore, the transition probabilities presented in Table III are modified from Kalyvas's empirical results to allow for all possible transitions between states to have non-zero probabilities.[9]

Emission probabilities

Emission probabilities guide the translation of observations into hidden states. The emission probabilities $\mathbf{\Phi}$ in Table IV are derived heuristically. Each cell describes the probability ϕ of observing a specific output value o_t given the true state of the latent variable territorial control q_t.

		o_t			
		A	B	C	D
q_t	S1	0.600	0.175	0.050	0.175
	S2	0.050	0.175	0.175	0.600
	S3	0.050	0.600	0.175	0.175
	S4	0.050	0.175	0.600	0.175
	S5	0.600	0.175	0.175	0.050

Table IV: Matrix of emission probabilities used in the estimation of the HMM. Rows sum to one. The probability value in each cell of this matrix answers the following question: "Given that the true state of an area i at time t is, for example, $S1$, what is the probability of observing, for example, A from the data?"

Here, I offer a brief justification for select probabilities in Table IV. If an area was in a state of complete rebel control $S1$, given the theoretical model, I do not expect to see any terrorist incidents or conflict events, hence observing evidence state $\mathrm{o}_{it} = A$ has the highest emission probability with $P(A|S1) = 0.6$. However, since zones of complete

[9] Kalyvas's empirical transition matrix contains a row with transitions to a territorial control zone of value "0." No further explanation is given what this zone entails. Therefore, I spread the relative frequency of observations of a transition to zone 0 proportionately across zones 1 through 5. In addition, I make small adjustments in the numerical values to allow for a minimum transition probability of 2.5% between all possible states of territorial control. The overall patterns of possible transitions remain unchanged, see Figure 3 in the online appendix.

rebel control in asymmetric conflicts are expected to be rare, there is a non-zero chance of observing occasional fighting between the rebels and the government. Hence, the probability of observing similar levels of terrorism and conventional war fighting $P(B|S1) = 0.175$ or more conventional fighting than terrorism $P(D|S1) = 0.175$ when the true underlying state is complete rebel control are small, but not zero. The probability of observing more terrorism than conventional fighting in areas of complete rebel control is conceptualized to be extremely low with $P(C|S1) = 0.05$.

For the latent variable state of contested territory with rebels having an advantage, $S2$, the most likely output to observe is $o_{it} = D$ with $P(D|S2) = 0.6$. If the situation is reversed and the government has the upper hand, the most likely observation is $o_{it} = C$ with $P(C|S4) = 0.6$. In case of high contestation without a clear advantage for either side, I expect to observe similar numbers of terrorist and conventional conflict events, such that $P(B|S3) = 0.6$. Emission probabilities can be thought of as capturing our confidence in the ability of the model to recover the true sequence of states. Here I assume that there is a 60% chance that given the true state of the latent variable territorial control q_t, we observe the expected emission o_t.

Data

The unit of analysis is the grid cell-month. I leverage geo-coded event data to measure each area's monthly exposure to terrorist incidents and conventional war acts in hexagonal grid cells with a minimum diameter of 0.25 degrees (approximately 28km at the equator). Data on conventional war acts come from the Georeferenced Event Dataset (GED) version 17.1 (Croicu and Sundberg, 2017; Sundberg and Melander, 2013). To achieve the highest level of delineation between terrorist attacks and events that are indicative of conventional fighting between rebels and government, only observations that are categorized by the GED as occurring within the realm of "state-based conflict" are considered. This excludes events that are classified as "non-state conflicts" or "one-sided violence" directed against civilians. Further, only events that can be attributed to at least the second order subnational administrative division are included.

Data on terrorist incidents come from the Global Terrorism Database (GTD) (START, 2016). GTD codes whether there is any doubt that an event constitutes terrorism as opposed to other forms of violence, such as conventional war acts or common crime. This variable is available from 1997 onward. I restrict my sample to GTD observations post-1997 that are unambiguously coded as terrorist attacks and that can be attributed to at least a second order administrative division.

Case selection

The measurement strategy outlined above is applicable to cases characterized by a high power asymmetry in favor of the government. As a logical derivative of the theoretical framework, only rebels that are weak compared to the government should resort to terrorist tactics. Hence, I expect to observe significant amounts of terrorism only when a high power asymmetry prevails throughout significant stretches of the conflict. Data on troops ratios between 1997 to 2011 from Polo and Gleditsch (2016) suggest 37 intrastate conflicts in which rebels are at most half as strong as the government — rendering them candidates for an estimation of territorial control via HMM.[10]

Data to assess the validity of the estimates of territorial control in asymmetric civil wars are extremely sparse. In fact, it is the lack of fine-grained data on territorial control that motivates the development the new estimation strategy. As an initial proof of concept, I present estimates of territorial control for the conflicts between the FARC rebels and the Colombian government from 2006 to 2017 and for the Boko Haram insurgency in Northeast Nigeria from 2008 to 2017. Colombia allows for an initial validation of the methodology by assessing the correlation of territorial control with deforestation in the aftermath of the 2016 peace agreement. Nigeria is included in the Armed Conflict Location & Event Data Project (ACLED) database, which allows for the construction of a coarse set of out-of-sample validation data on territorial control.

[10] See Figures 4 and 5 in the online appendix.

HMM estimates of territorial control

Colombia

The 2016 peace agreement between the Colombian government and the FARC ended a conflict that left approximately 250,000 people dead and displaced millions.[11] Colombia experienced decades of violence stemming from clashes between insurgent groups, paramilitaries, and official armed forces, as well as violence linked to and exacerbated by the presence of drug cartels in the country (Cortés and Montolio, 2014; Dube and Vargas, 2013). Figure 4 plots estimated levels of territorial control for the FARC guerrilla and the Colombian government from 2006 to 2017.[12] Graphed are annual averages of monthly territorial control estimates for 851 grid cells. I exclude the Amazon and Orinoco natural regions in the East because the low population density in these areas raises concerns over systematic under-reporting of conflict events. The shading of the grid cells in Figure 4 indicates estimated levels of territorial control from red denoting full rebel to blue denoting full government control.

The maps demonstrate significant spatial and temporal variation in territorial control, with a few persistent hot spots of rebel controlled areas along the border to Venezuela in the Northeast and the Southwestern coastal region. These patterns are corroborated by expert accounts of the conflict. Upon being elected to power in 2002, president Alvaro Uribe ignited a heavy military campaign against the rebels. "From 2002, the state saw a steady recovery of areas that had been under guerrilla control. [...] Certain regions of the country, however, continued to exhibit high levels of violence, especially in the west and near the border with Venezuela." (Arjona, 2016, 92)

The Colombian peace process provides a unique opportunity for out-of-sample validation

[11] The Guardian, 23 June 2016, `https://www.theguardian.com/world/2016/jun/23/colombia-farc-rebel-ceasefire-agreement-havana`, accessed 1 December 2017.

[12] The full model is estimated on data from 1997 to 2017. Existing studies suggests that the guerrillas controlled significant amounts of territory in the 1990s (Arjona, 2016, 91). At the time of writing, no information is available regarding the exact location of FARC strongholds in 1997. If such information became available, grid cells covering these areas could be initiated with a strong prior that suggests rebel control. Absent this information, all grid cells are initiated with flat priors of 0.2 for all possible states of control. Based on my theory of rebel tactics, I expect little to no violence in guerrilla strongholds. This suggests that FARC control is likely underestimated in the late 1990s and early 2000s. The estimates are expected to be more valid after 2005 when the paramilitaries started to demobilize.

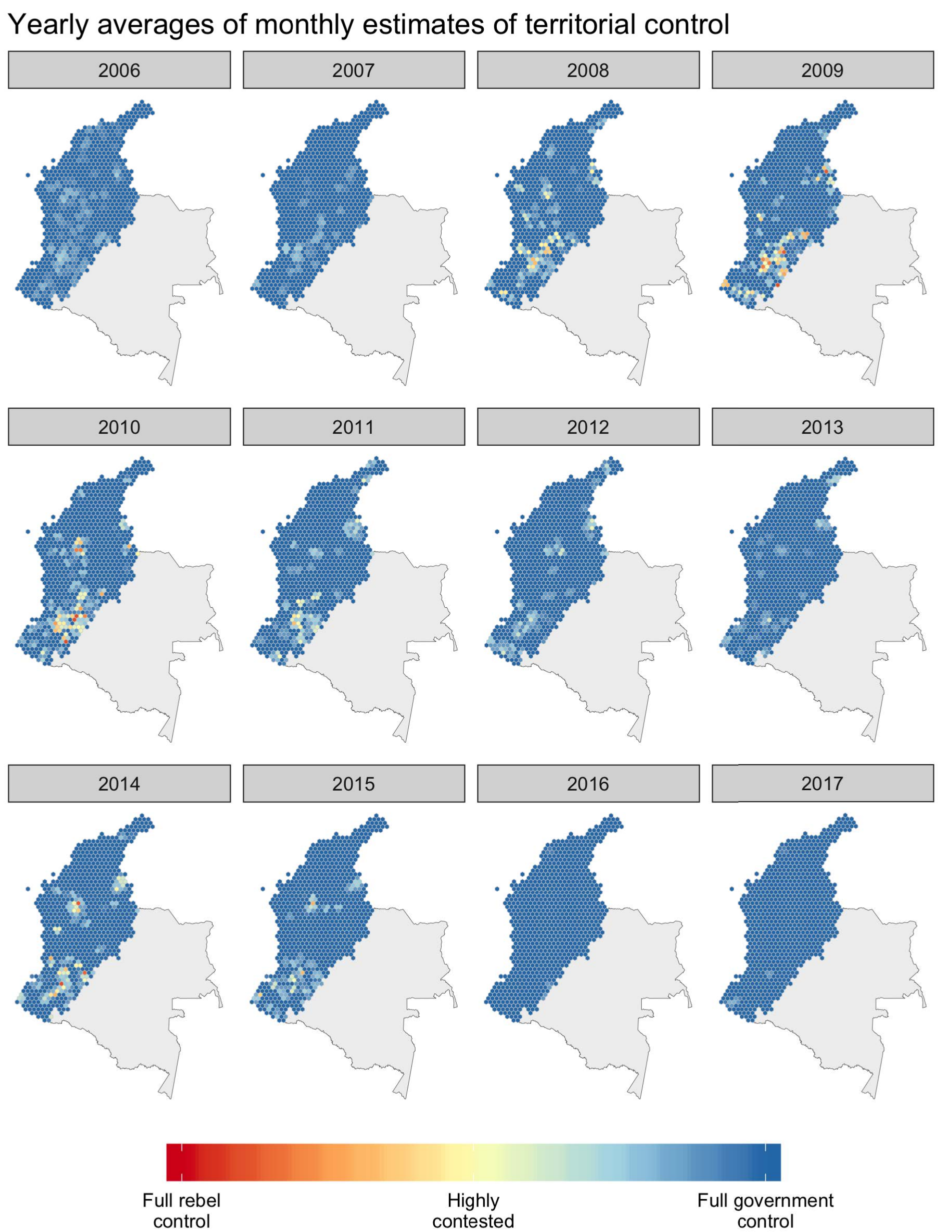


Figure 4: Estimated levels of territorial control in Colombia for hexagonal grid cells with a minimum diameter of 0.25 degrees, excluding the Amazon and Orinoco natural regions (N = 851).

of my estimates of territorial control. The signing of the peace agreement between the government and the FARC in 2016 introduced a sudden change in territorial control, because it forced rebels to disarm and abandon their strongholds. The timing of the eventual signing of the peace agreement in 2016 was plausibly unexpected, given the long history of failed peace negotiations between the government and the FARC. The unanticipated timing minimizes endogeneity bias when relating the pre- and post-peace differences in proxy variables to the differences in FARC territorial control.

As a major source of revenue for the FARC, gold mining and coca production were heavily regulated within their territorial strongholds. It has also been reported that "guerrillas enforced strict limits on logging by civilians – in part to protect their cover from air raids by government warplanes."[13] Criminal organizations (*bandas criminales* or BACRIM) were quick to move into the power vacuum that the FARC left when it laid down its arms following the peace agreement. The BACRIM are reportedly less inclined to regulate mining, logging, and coca production and are instead intensifying these operations. This lead to a rise in rates of deforestation, in particular in areas that were previously controlled by the FARC. Relating pre- and post-peace levels of forest cover to estimated changes in territorial control allows me to corroborate my estimates of territorial control.

If my estimates of territorial control are valid, deforestation should be more likely in areas that saw larger changes in territorial control as a result of the peace accord. To test this, I estimate the following model.

$$Deforestation_{i,t} = \beta_0 + \beta_1 \Delta Control_{i,t} + \beta_2 Peace_t + \beta_3(\Delta Control_{i,t} \times Peace_t) + \epsilon_i$$

$Deforestation_{i,t}$ is a dummy that equals one if a grid cell i experienced deforestation over the course of year t. $\Delta Control$ denotes the change in the annual level of territorial control (averaged over monthly HMM estimates) between year t and $t-1$, with positive values indicating changes toward more government control.[14] $Peace_t$ is a dummy for the year

[13] `https://www.theguardian.com/world/2017/jul/11/colombia-deforestation-farc`, accessed 1 December 2017.

[14] To compute annual averages and changes in territorial, I recode the discrete territorial control variable

2016 in which the peace agreement was signed and FARC forces started to disarm. A positive coefficient on the interaction between changes in territorial control and the the peace dummy $\Delta Control_{i,t} \times Peace_t$ indicates a higher probability of deforestation in 2016 in areas that saw changes in territorial control as a result of the demobilization of rebel forces. Errors are clustered by grid cell. The probability of deforestation is estimated via logistic regression. Raster data on deforestation from 2013 to 2016 are obtained via the forest monitoring system from the Colombian *Instituto de Hidrología, Meteorología y Estudios Ambientales* (IDEAM). For each grid cell in each year, I code a binary indicator whether the cell experienced deforestation. In 2013, 26 grid cells saw their forest cover reduced. This number jumped to 51 cells in 2016.

	Deforestation$_{i,t}$		
	(1)	(2)	(3)
ΔControl$_{i,t}$ × Peace$_t$		3.59*	3.29*
		(1.68)	(1.63)
ΔControl$_{i,t}$	−0.63	−1.51	−1.56
	(1.00)	(0.97)	(0.86)
Peace$_t$	0.29	0.22	0.07
	(0.16)	(0.16)	(0.18)
Deforestation$_{i,t-1}$			0.86**
			(0.28)
Constant	−3.03***	−3.04***	−2.94***
	(0.10)	(0.10)	(0.11)
Observations	3,404	3,404	2,553
Log Likelihood	−670.74	−668.87	−545.38
Akaike Inf. Crit.	1,347.48	1,345.73	1,100.75
Note:	*p<0.05; **p<0.01; ***p<0.001 Logistic regression coefficients with bootstrapped clustered standard errors by grid cell in parentheses.		

Table V: Relationship between rebel territorial control and deforestation in Colombia.

The results in Table V support the hypothesis that changes in territorial control as a result of the peace agreement are associated with a higher probability of deforestation. The

with 0 indicating full rebel control, 0.25 indicating zones that are contested but closer to rebel control, 0.5 denoting highly contested areas, 0.75 indicating contestation with the government having the upper hand, and 1 signifying full government control.

coefficient on the interaction between changes in territorial control and the timing of the 2016 peace agreement in model 2 is positive and statistically significant at the minimum 5% level. A change of 0.25 in the annual level of territorial control, for example that from an area that is on average contested but closer to government control to one that is on average fully controlled by the state, is associated with an increase in the predicted probability of deforestation from 5.6% to 9.1%. The coefficient for the interaction remains statistically significant upon including a lagged dependent variable in model 3. Thus, HMM estimates of territorial control produce results in line with observed empirical relationships between changes in territorial control and deforestation in Colombia.

Nigeria

To demonstrate the applicability of the approach across country contexts, I provide monthly-level estimates of territorial control for Northeast Nigeria. The area has seen sustained fighting between the Nigerian government and Boko Haram since 2009.[15] In late 2014 and early 2015, reports suggest that the insurgents controlled 15 localities in the northeast border region with Cameroon, and had partial control over additional 15 local government areas.[16] In early to mid 2015, the Nigerian government and African Union multilateral troops launched attacks against Boko Haram that caused the insurgents to loose a majority of the territory they previously controlled.[17] However, recent reports cast doubt over the government's claim that it drove the insurgents out of the region, with an account from May 2018 suggesting that Boko Haram continues to control parts of Borno state via roadblocks, stop and search operations, and the collection of "taxes" for protection.[18]

[15] See https://www.nytimes.com/2014/11/11/world/africa/boko-haram-in-nigeria.html accessed 10 September 2018.

[16] See https://www.amnesty.org/en/latest/news/2015/01/boko-haram-glance/, accessed 27 September 2018. Estimates suggests that the total area controlled by the insurgents amounted to approximately 20,000 square miles — approximately the area of Belgium (see https://www.telegraph.co.uk/news/worldnews/africaandindianocean/nigeria/11337722/Boko-Haram-is-now-a-mini-Islamic-State-with-its-own-territory.html, accessed 27 September 2018).

[17] http://web.stanford.edu/group/mappingmilitants/cgi-bin/groups/view/553?highlight=boko%2Bharam, accessed 27 September 2018.

[18] https://www.dw.com/en/boko-haram-islamists-still-control-parts-of-northeastern-nigeria/a-43851013, accessed 27 September 2018.

Figure 5 illustrates yearly averages for monthly-level HMM estimates for 15 Northeast Nigerian states from 2009 to 2017 for grid cells whose centroids are 0.25 degrees apart (N = 942).[19] The estimates illustrate the onset of the Boko Haram insurgency in 2009.[20] Starting in 2011, Boko Haram is estimated to gain temporary areas of control in Borno state. The insurgents subsequently establish persistent strongholds that are at first scattered throughout the study region. By 2014, the estimates show the consolidation of Boko Haram control in the Northeast. Following the offensive of the Nigerian government and the African Union, Boko Haram is estimated to have lost significant amounts of territory between 2015 and 2016. By 2017, its strongholds are limited to a few areas in the border region with Cameroon.

The inclusion of Nigeria in the Armed Conflict Location & Event Data Project (ACLED) data version 8.0 allows me to construct an out-of-sample validation data set (Raleigh et al., 2010). Through event type labels, ACLED contains information on whether an event resulted in rebels gaining control or establishing a base (coded as S1, continuous value 0), battles with no changes in control (S3, continuous value 0.5), the government gaining control or establishing a base (S5, continuous value 1), and instances of remote violence (S2 for government remote violence with a continuous value 0.25; S4 for insurgent remote violence mapped to continuous value of 0.75).[21] The events are aggregated to grid cells on a monthly level. Territorial control is assigned based on the occurrence of control-related events within a grid cell. New events cause a cell to update the coded level of territorial control based on event type. In the case of multiple events occurring in the same grid cell month, I average across the continuous values that are indicative of the status of territorial control associated with each event. Cells that experience zero or one event over the entire period of observation 2008 to 2017 are assumed to be under full government control. Similarly, if a cell does not experience any violence in the previous six months, it is assumed to be under government

[19] The coverage of the HMM results mirrors the 15 states included in the study by Aronson et al. (2017) that are most subjected to Boko Haram violence, namely Adamawa, Bauchi, Benue, Borno, Gombe, Jigawa, Kaduna, Kano, Katsina, Nassarawa, Niger, Plateau, Taraba, Yobe, and the Federal Capital Territory.

[20] The model is estimated starting in 2008 and initiated with a strong prior of full government control in the first month of observation.

[21] ACLED contains a small number of events for which manual coding is necessary to determine the actor gaining control, in particular for occurrences of remote violence. The respective documentation is available upon request.

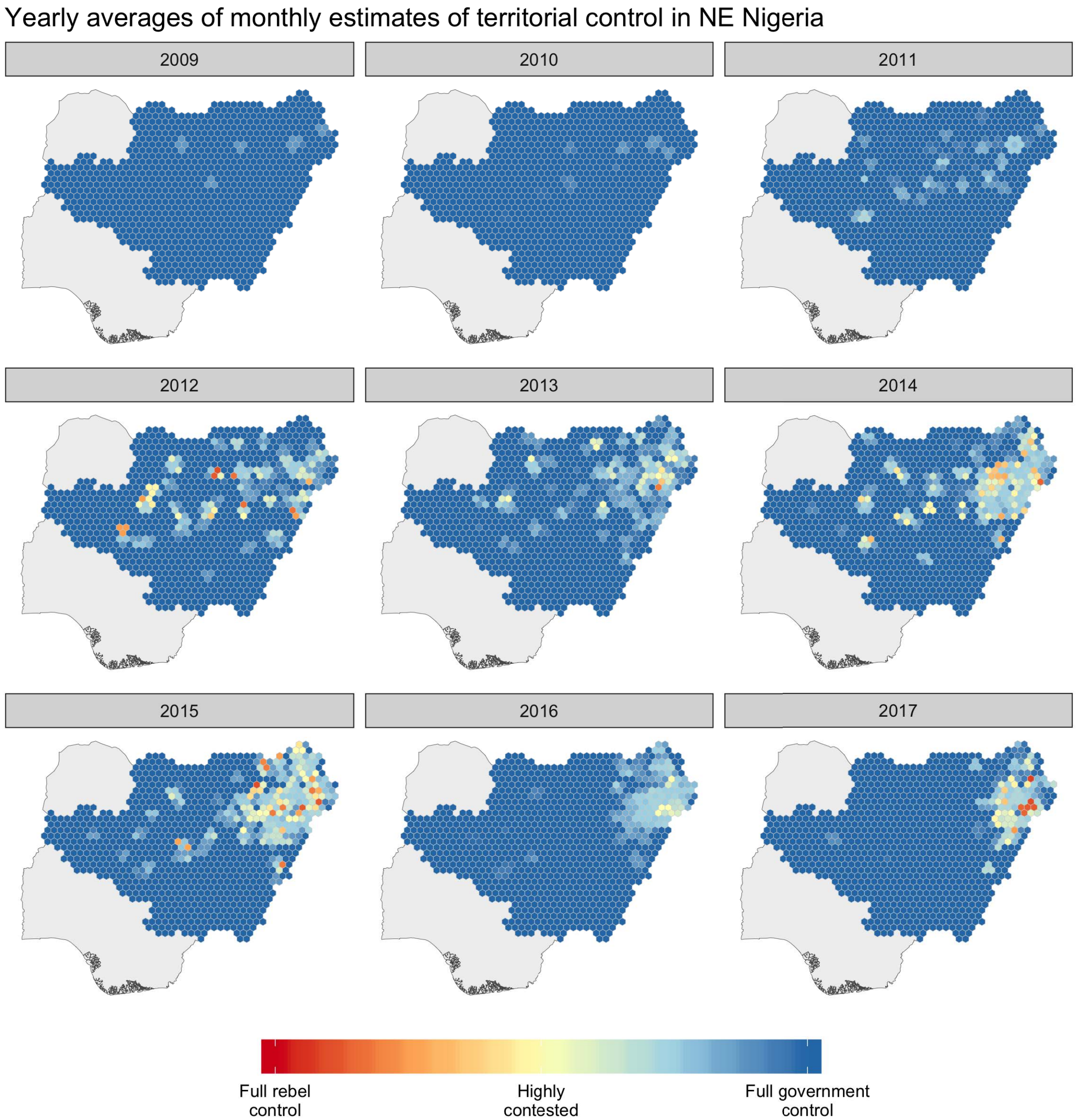


Figure 5: Estimated levels of territorial control in Northeast Nigeria for hexagonal grid cells with a minimum diameter of 0.25 degrees (N = 942).

control.[22]

I create two sets of validation data. The first validation set adopts the assumption that remote violence is indicative of areas that are contested but closer to either rebel or government control, depending on the perpetrator (denoted "full sample" below). The second data set drops this assumption and considers only ACLED events that make explicit reference to changes in territorial control (denoted "restricted sample" below).[23] Due to the strong assumptions necessary to construct a testing set from ACLED as well as concerns regarding reporting error in these data (Eck, 2012), conclusions from a comparison with the HMM estimates should be taken with a grain of salt. For example, based the media reports above, the validation data likely underestimate the extent of Boko Haram control in 2014. However, the testing set constructed from ACLED data offers the best opportunity for out-of-sample validation of territorial control in Nigeria available to date.

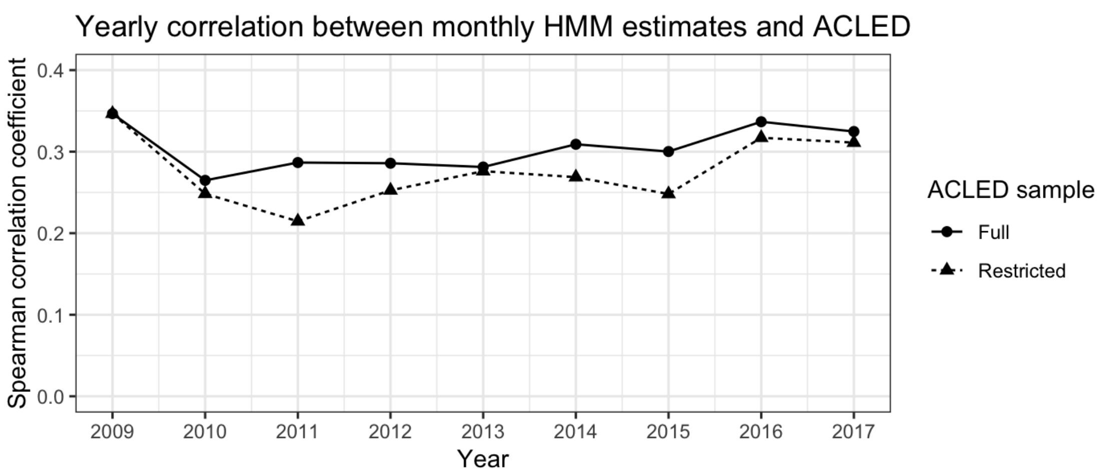


Figure 6: Spearman's rank-order correlation coefficients for a annual-level correlations between HMM estimates and the ACLED validation data.

Figure 6 plots annual Spearman's rank-order correlation coefficients between monthly territorial control estimates and the ACLED validation data. The correlation ranges between 0.21 and 0.35. The validation data that relies on instances of remote violence to indicate

[22]The conflict in northeast Nigeria is highly active during the period between 2009 and 2017 — rendering six months a reasonable upper bound for the stationarity of control in a given cell.

[23]Event types in the restricted sample are the government gaining control or establishing a base, Boko Haram gaining control or establishing a base, and battles with no changes in control. Figure 6 in the online appendix plots yearly averages for both validation data sets.

areas of partial control of either Boko Haram (S2) or the government (S4), on average shows a higher correlation with the HMM estimates. Compared with the validation data, the HMM appears to be overestimating the level of rebel complete control and to be underestimating the level of contestation between the government and Boko Haram. However, the HMM estimates in Figure 5 uncover general spatial patterns and major temporal trends in the distribution of territorial control for the government and Boko Haram in the Northeast of the country. In particular the ability of the model to capture the severe reduction of insurgent territorial between 2015 and 2016 is striking.

Conclusion

I propose a novel measurement approach for the estimation of territorial control in asymmetric civil wars. I leverage observed variation in the co-occurrence of terrorist attacks and conventional fighting within a machine learning framework to obtain estimates of the latent variable territorial control. As a proof of concept, I present estimates of actors' control over territory for the fight between the FARC and the Colombian government and the Boko Haram insurgency in Nigeria. A validation of the Colombia estimates using patterns of deforestation in the aftermath of the 2016 peace agreement suggests that the model is able to recover general trends in the evolution of territorial control across time and space. Newly developed validation data for the Boko Haram insurgency in Nigeria show that the estimates correlate reasonably with alternative measures of territorial control.

The results demonstrate that Hidden Markov Models (HMM) are a fruitful approach to address the lack of data on territorial control in asymmetric civil wars. The methodology outlined in this paper allows for the generation of territorial control estimates that utilize publicly available conflict event data, are easy and fast to estimate, capture fine-grained spatiotemporal subnational patterns, and can be computed for a cross-section of countries. The estimates thus yield a valuable source of information for subnational analyses of civil war dynamics both for within-country and comparative studies. As an example, cross-nationally available estimates of territorial control are crucial for enhancing our understanding of how belligerents' local provision of public goods interacts with their level or territorial control. The estimates will also be instrumental for investigating cross-border contingencies in asymmetric civil wars. Rebel groups like Boko Haram operate across international borders and often leverage the remoteness of border regions to their advantage upon gaining strongholds beyond the reach of their primary enemy's armed forces. HMM estimates are produced with a methodology allows for such localized cross-country analyses.

The derivation of the transition probabilities that specify prior beliefs regarding the temporal evolution of territorial control illustrates how existing empirical or theoretical research can be leveraged to inform model parameters in machine learning applications. All parameters can be adapted based on additional cross-sectional information or case-specific knowledge. As an example, the Nigeria model is initiated with a strong prior for complete government control because the start of the Boko Haram insurgency in 2009 is captured by the data. In other conflict settings, one could initiate the model with a prior of full rebel control for grid cells that are known insurgent strongholds in the first period of observation. Future work should also explore the inclusion of covariates such as terrain or landcover to introduce heterogeneity across grid cells regarding our expectation how likely rebels are able to capture and sustain a territorial base.

The measurement strategy incorporates a new method to gauge subnational areas' exposure to conflict events. Rather than discretely assigning events to grid cells, I allow the impact of violent events for a given area to dissipate continuously over space and time. Conflict exposure is computed as the sum of spatially and temporally weighted events. HMMs estimate territorial control for each grid cell individually. They explicitly model time dependencies but do not account for the correlation of the latent variable between close spatial units. The continuous measure of areas' exposure to conflict events captures spatial dependency in observed rebel tactics and thus overcomes a major limitation of standard HMMs. This methodological innovation renders HMMs a suitable tool to estimate latent constructs that feature spatial *and* temporal variation in subnational-level research.

References

Anders, T., H. Xu, C. Cheng, and T. S. Kumar (2017). Measuring Territorial Control in Civil Wars Using Hidden Markov Models: A Data Informatics-Based Approach. Proceedings of the NIPS 2017 Workshop on Machine Learning for the Developing World. arXiv:1711.06786.

Arjona, A. (2016). *Rebelocracy. Social Order in the Colombian Civil War.* New York: Cambridge University Press.

Aronson, J., D. Ciland, P. K. Huth, and J. I. Walsh (2017, September). An enhanced dataset of territorial control by conflict actors. Paper presented at the 2017 Meeting of the American Political Science Association.

Asal, V., L. De La Calle, M. Findley, and J. Young (2012). Killing civilians or holding territory? How to think about terrorism. *International Studies Review 14*(3), 475–497.

Bakker, R., J. Daniel W Hill, and W. H. Moore (2016). How much terror? Dissidents, governments, institutions, and the cross-national study of terror attacks. *Journal of Peace Research 53*(5), 711–726.

Carter, D. B. (2015). When terrorism is evidence of state success: securing the state against territorial groups. *Oxford Economic Papers 67*(1), 116.

Chenoweth, E. (2013). Terrorism and democracy. *Annual Review of Political Science 16*(1), 355–378.

Cortés, D. and D. Montolio (2014). Provision of public goods and violent conflict: Evidence from Colombia. *Peace Economics, Peace Science and Public Policy 20*(1), 143–167.

Croicu, M. and R. Sundberg (2017). UCDP GED Codebook version 17.1. Technical report, Department of Peace and Conflict Research, Uppsala University.

Cunningham, D. E., K. S. Gleditsch, and I. Salehyan (2013). Non-state actors in civil wars: A new dataset. *Conflict Management and Peace Science 30*(5), 516–531.

de la Calle, L. and I. Sánchez-Cuenca (2012). Rebels without a Territory: An Analysis of Nonterritorial Conflicts in the World, 1970–1997. *Journal of Conflict Resolution 56*(4), 580–603.

de la Calle, L. and I. Sánchez-Cuenca (2015). How armed groups fight: Territorial control and violent tactics. *Studies in Conflict & Terrorism 38*(10), 795–813.

Dube, O. and J. Vargas (2013). Commodity Price Shocks and Civil Conflict: Evidence from Colombia. *The Review of Economic Studies 80*, 1384–1421.

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

Findley, M. G. and J. K. Young (2012). Terrorism and civil war: A spatial and temporal approach to a conceptual problem. *Perspectives on Politics 10*, 285–305.

Fortna, V. P. (2015). Do Terrorists Win? Rebels' Use of Terrorism and Civil War Outcomes. *International Organization 69*, 519–556.

García-Sánchez, M. (2016). Territorial control and vote choice in Colombia. A multilevel approach. *Política y gobierno 23*(1), 53–96.

Ghahramani, Z. (2001). An introduction to hidden markov models and bayesian networks. *Journal of Pattern Recognition and Artificial Intelligence 15*(1), 9–42.

Hendrix, C. S. and J. K. Young (2014). State capacity and terrorism: A two-dimensional approach. *Security Studies 23*(2), 329–363.

Himmelmann, L. (2010). HMM - Hidden Markov Models. R Package version 1.0, `https://CRAN.R-project.org/package=HMM`.

Jurafsky, D. and J. H. Martin (2017, August). Hidden markov models. Chapter 9. In *Speech and Language Processing*. Online `https://www.cs.jhu.edu/~jason/papers/jurafsky+martin.slp3draft.ch9.pdf`, accessed 8 July 2018.

Kalyvas, S. N. (2006). *The Logic of Violence in Civil War*. Cambridge, UK: Cambridge University Press.

Kalyvas, S. N. and M. A. Kocher (2009). The dynamics of violence in vietnam: An analysis of the hamlet evaluation system (hes). *Journal of Peace Research 46*(3), 335–355.

National Consortium for the Study of Terrorism and Responses to Terrorism (START) (2016). Global terrorism database [data file]. Retrieved from http://www.start.umd.edu/gtd, accessed 29 June 2017.

Openshaw, S. and P. J. Taylor (1979). A millon or so correlation coefficients: three experiments on the modifiable areal unit problem. In N. Wrigley (Ed.), *Statistical applications in the spatial sciences*, Chapter 5, pp. 127–144. London: Pion.

Polo, S. M. and K. S. Gleditsch (2016). Twisting arms and sending messages: Terrorist tactics in civil war. *Journal of Peace Research 53*(6), 815–829.

Quinn, J. M. (2015). Territorial contestation and repressive violence in civil war. *Defence and Peace Economics 26*(5), 536–554.

Raleigh, C., A. Linke, H. Hegre, and J. Karlsen (2010). Introducing ACLED: An Armed Conflict Location and Event Dataset: Special Data Feature. *Journal of Peace Research 47*(5), 651–660.

Sambanis, N. (2004). What is civil war? Conceptual and empirical complexities of an operational definition. *The Journal of Conflict Resolution 48*(6), 814–858.

Schelling, T. (1960). *Arms and Influence*. Hartford: Yale University Press.

Schutte, S. (2017). Geographic determinants of indiscriminate violence in civil wars. *Conflict Management and Peace Science 34*(4), 380–405.

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

Staniland, P. (2012). States, insurgents, and wartime political orders. *Perspectives on Politics 10*(2), 243–264.

Stewart, M. A. (2018). Civil war as state-making: Strategic governance in civil war. *International Organization 72*(1), 205–226.

Stewart, M. A. and Y.-M. Liou (2017). Do good borders make good rebels? Territorial control and civilian casualties. *The Journal of Politics 79*(1), 284–301.

Sundberg, R. and E. Melander (2013). Introducing the UCDP Georeferenced Event Dataset. *Journal of Peace Research 50*(4), 523–532.

Tao, R., D. Strandow, M. Findley, J.-C. Thill, and J. Walsh (2016). A hybrid approach to modeling territorial control in violent armed conflicts. *Transactions in GIS 20*(3), 413–425.

Tilly, C. (2004). Terror, terrorism, terrorists. *Sociological Theory 22*(1), 5–13.

Appendix A Measuring conflict exposure

Spatially and temporally disaggregated conflict event data is a key source of information in the study of subnational violence. However, many covariates of interest operate on an areal level, for example economic wealth, terrain, the ethnic composition of the population, the availability of natural resources, or the provision of public services. Individual conflict events are thus typically aggregated to grid cells or administrative units such as districts or municipalities to match the unit of measurement of the covariates and to measure the exposure of these subnational areas to conflict events. Each conflict event is commonly assigned *discretely* to the subnational area within which it is located. This standard practice is problematic for two main reasons. First, scholars are frequently only accounting for events that fall within the boundaries of a chosen subnational unit and do not account for events that happen in the vicinity. Second, inferences regarding an area's exposure to conflict are highly sensitive to the drawing of boundaries — widely cited in the literature as the modifiable areal unit problem (MAUP). A similar issue arises in the temporal domain when conflict events are discretely assigned to the calendar month or year in which they occurred. An event's influence on local conflict dynamics is unlikely to abruptly stop at the chosen spatial or temporal boundaries; nor will it homogeneously affect the entirety of the space. Rather, its impact dissipates *continuously* over space and time.

A simplistic approach to coding areas' exposure to terrorism conflict events would sum the number of events that fall within a given grid cell. This procedure faces the problem that the assignment of conflict events to grid cells is highly dependent on the sampling of centroid locations. MAUP describes the discrepancy between real world spatial patterns of events and patterns created via aggregation of events into homogenous spatial units (Openshaw and Taylor, 1979). Shifting the location of the centroids can have a severe influence on the number of events that are assigned to a particular cell. This is particularly concerning when the drawing of grid cell boundaries leads clusters of events to be broken up into smaller groups — causing the relative frequency of terrorist events and conventional war acts to change dramatically. Figure 7 illustrates this issue. Based on the location of grid cell centroids in panel A, we would code the relative frequency of rebel and conventional war fighting to correspond to the values of the variable $\text{Tactics}_{it} = [D, D, A]$. If the centroids were shifted by 25% relative to the location of the events, we would conclude the emissions of these three cells to have values of $\text{Tactics}_{it} = [B, D, C]$.

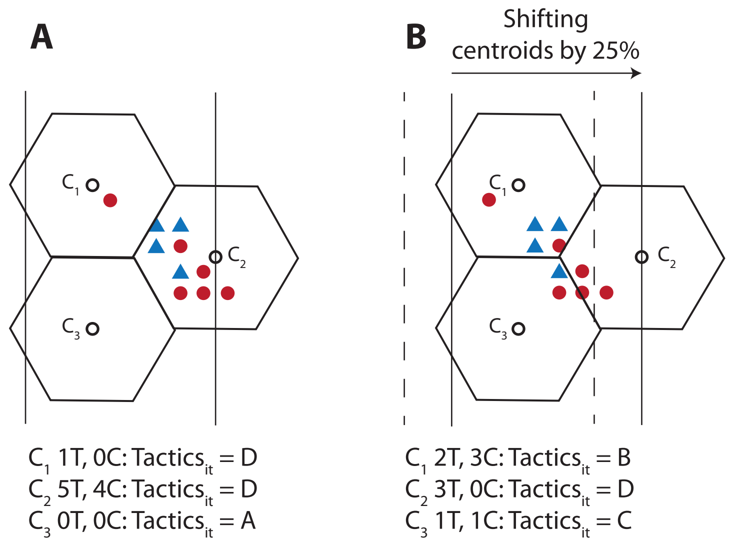


Figure 7: The schematic illustrates how shifting the location of the grid cell centroids from their original (randomly sampled) location (panel **A**) by just 25% (panel **B**) can result in vastly different conclusions about the coding of rebel tactics. Red dots indicate the location of conventional events; blue triangles those of terrorist attacks. This is a simplified example — in the analysis, Tactics_{it} is computed using probabilities from Poisson distributions under application of a margin parameter.

To alleviate this problem, I propose a novel measurement model for rebel tactics in civil war that uses spatial and temporal weights to associate conflict events with grid cells rather than relying on discrete assignment. The importance of individual violent events for the estimation of territorial control decreases over time and space. I model this intuition by assigning space- and time-varying weights to each event.

A.1 Spatial and temporal decay functions

For each grid cell centroid-month c_{it}, $i = 1, \dots I$ indexes centroids and $t = 1, \dots T$ indexes months. For each conflict event e_{jm} , $j = 1, \dots J$ indexes individual events and $m = 0, \dots M$ indexes the calendar month in which the event occurred. Let lon_i and lat_i denote the longitude and latitude of each grid cell centroid c_i in radians, respectively. Similarly, let lon_j and lat_j denote the longitude and latitude of each conflict event e_j in radians. Then the spatial distance d_{ij} in km between centroid c_i and event e_j is computed as the geodesic distance between two points using the Haversine formula,

$$d_{ij} = 2r \arcsin\left(\sqrt{\sin^2\left(\frac{lat_j - lat_i}{2}\right) + \cos(lat_i)\cos(lat_j)\sin^2\left(\frac{lon_j - lon_i}{2}\right)}\right),$$

where $r \approx 6371$ denotes the earth mean radius in kilometers. The temporal distance (in the following called age) $a_{tm} = t - m$ measures the months between when event e_{jm} occurred and the time of observation of the grid cell-month $c_i t$. An event occurring in the month of observation has an age of $a_{tm} = 0$, while an event that occurred four months ago has an age of $a_{tm} = 3$.

For each centroid-month unit c_{it} I measure the spatial distance d_{ij} in km and the temporal distance a_{tm} to each conflict event e_{jm}, resulting in a total number of centroid-event-month observations of size $K = I \times J \times T$. Specifically, for each grid cell c_i in each month t, I create a vector D of spatial distances and a vector A of temporal distances to each event. Events that occur in the future from the time of observation t (i.e. where $u > t$) receive a missing value. I then weight both vectors to allow the impact of conflict events on grid cells to dissipate over space and time.

I assume the impact of an event to dissipate following a logistic decay function of the general form

$$w = \frac{1}{1 + e^{-\kappa + \gamma x}},$$

where x denotes the decaying quantity (here event age or distance between the event and a centroid), κ determines the slope of the curve and γ defines its inflection point. To describe a decay function, both the slope parameter κ and he inflection parameter γ have to be positive real numbers. To model spatial decay, assume the slope parameter to be $\kappa_d = 7$ and the inflection parameter to be $\gamma_d = 0.35$. To model temporal decay in months, I use a steeper sigmoid curve. I assume the temporal slope parameter to be $\kappa_a = 8$ and the inflection parameter to be $\gamma_a = 2.5$.[24]

Figure 8 plots the decay functions using the parameter values above. Based on the shape of the logistic decay functions above, an event that occurs at the location of the centroid of a grid cell receives a spatial weight of 1. An event that occurs 10km away from the centroid receives a weight of 0.97 and an event 25km away is weighted by 0.15 — after which its influence tends toward 0. The temporal weight features a different rate of decay. In the first month, the event receives a temporal weight of 1, followed by 0.95 in the second, 0.62 in the third, and 0.11 in the fourth month; after which the weight approaches zero.

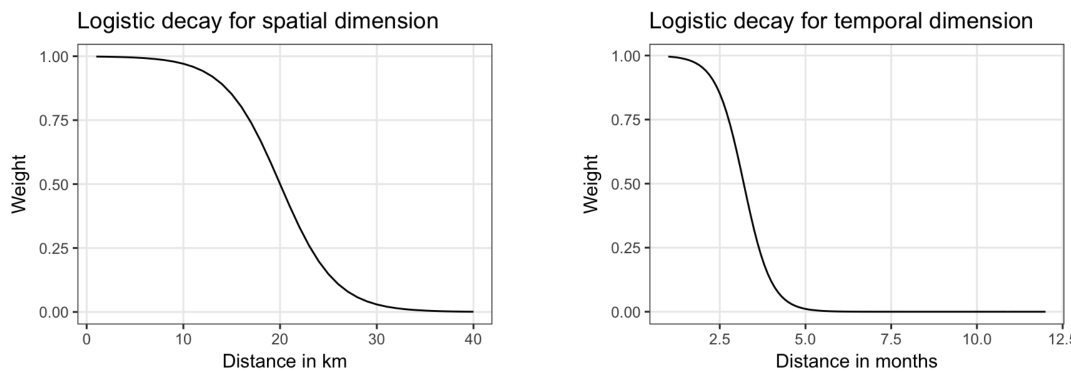
The exposure of grid cell c_{it} to conflict events E_{it} is computed as the sum over all temporally

[24] Figures 9 and 10 illustrate the shape of the logistic decay functions for alternative parameter values.

and spatially weighted events J.

$$E_{it} = \sum_{j=1}^{J} \left(w_{d_{ij}} \times w_{a_{jt}} \right) \tag{1}$$

Thus, the resulting unit of observation is the grid cell-month, i.e. a vector of exposure values of size $E = I \times T$.



(a) Spatial (b) Temporal

Figure 8: Logistic function that describes the decay of the influence of an event in relation to a centroid in the spatial and temporal dimensions.

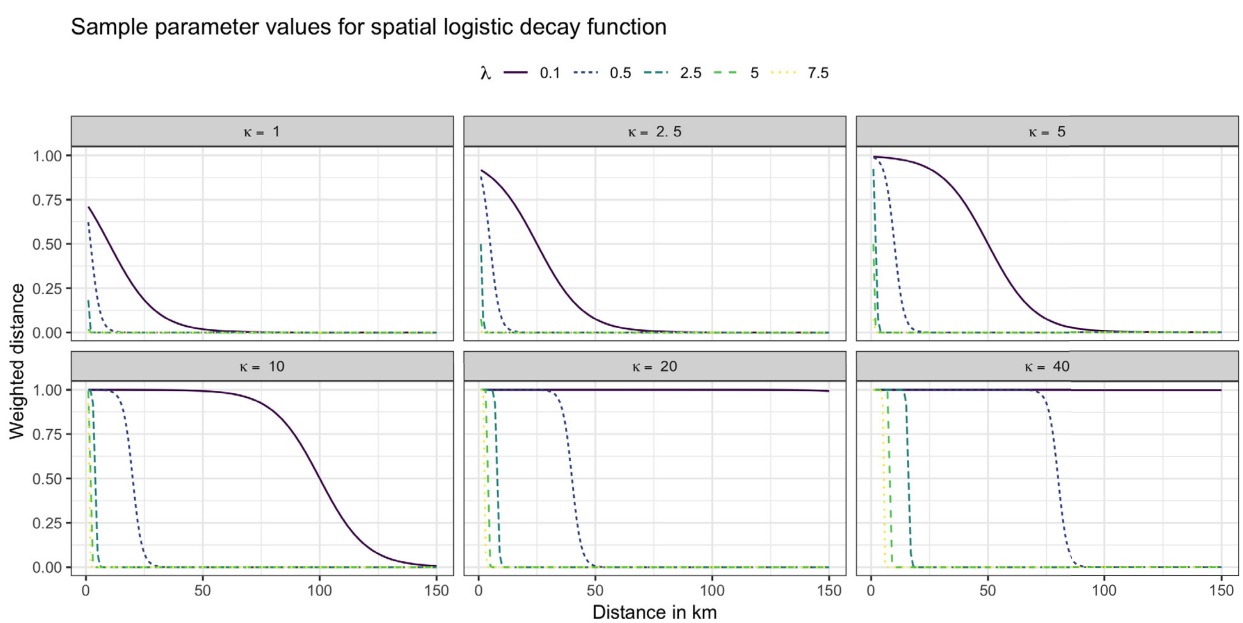


Figure 9: Influence of the slope parameter κ and inflection parameter λ on the shape of the logistic decay curve for an example of distances varying from 1km to 150km.

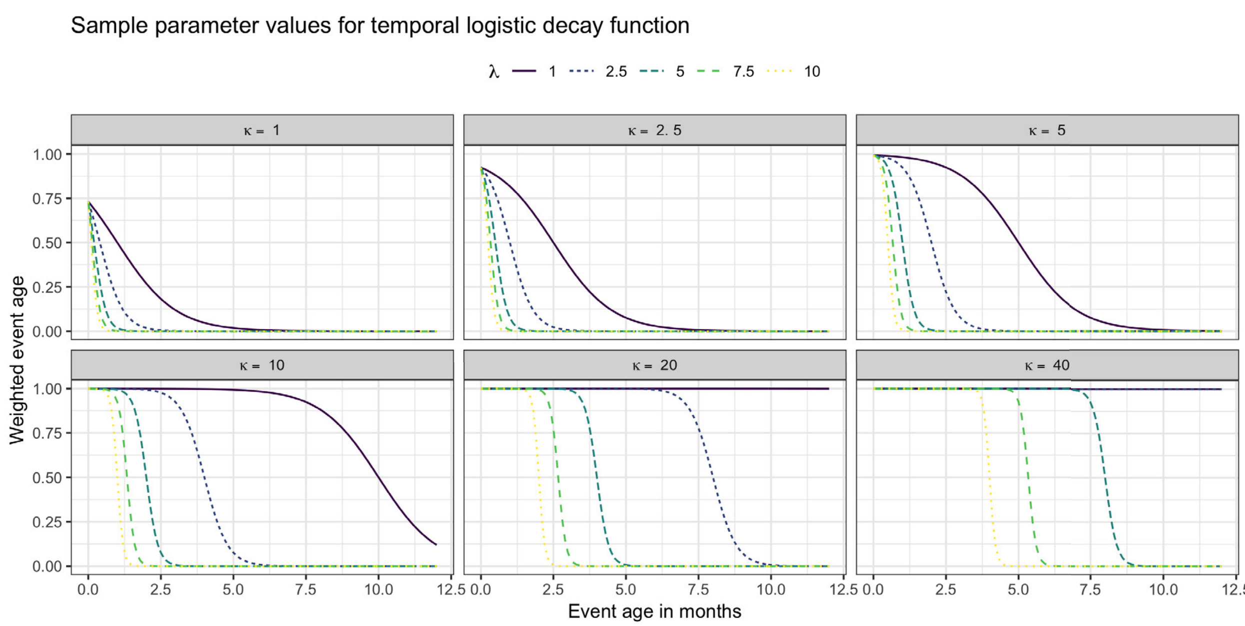


Figure 10: Influence of the slope parameter κ and inflection parameter λ on the shape of the logistic decay curve for an example of event ages from 0 to 12 months.

A.2 Comparison between discrete and continuous aggregation

Figure 11 presents a visual comparison between the discrete and continuous aggregation of events to hexagonal grid cells with a minimum diameter of 0.25 and 0.5 degrees, using simulated data.[25] Note that for illustration, the aggregation of events is performed only in the spatial dimension. No temporal weight is applied. Dots denote the location of events. The shading of cells indicates the computed conflict exposure value for the discrete assignment of events to cells (first row), the continuous assignment with a logistic decay curve that falls below a weight of 0.9 at a distance of approximately 14km (second row), and continuous assignment that dips below 0.9 at approximately 36km (third row).

The plot in Figure 11 highlights the shortcomings of the discrete assignment approach. For grid cells of size 0.25 degree and 0.5 degree, discrete aggregation leads to cells that are neighboring conflict hotspots to receive a zero exposure value. From the plot it is also evident that the same underlying locations would have received a vastly different value of conflict exposure if the centroids of the cells were moved a little, or the cell size was increased or decreased.

It is upon the researcher to decide the appropriate rate of decay of conflict events — a choice that is likely context and application specific. Figure 11 illustrates that the choice the rate of decay is highly influential for the resulting computed conflict exposure value. Applying a rate of decay in that dips below 0.9 at approximately 36km (third row) leads to a sum of weighted events that is significantly higher than the maximum count in the discrete assignment case.[26] Moreover, the maximum conflict exposure $\max(E_i) = 13.1$ for the 0.25 degree resolution grid is computed for a cell that itself does not experience any conflict events within its borders. This cell's high conflict exposure value is caused by the abundance of conflict events in neighboring cells that, given the slow rate of spatial decay, accumulate in the centering cell. The ability of a conflict event to contribute to the total conflict exposure beyond grid cell borders is in principle a desired feature of the methodology to overcome the MAUP. However, as shown in Figure 11, a slow rate of decay leads to a potential over-aggregation of conflict events.

Once the appropriate rate of decay is determined, the cells size appears to be somewhat less influential for the computed conflict exposure of underlying locations. This is not to say that cell

[25] Event locations are a subset of the simulated data from Schutte and Donnay's (2014) `mwa` `R` package.

[26] The maximum values for the discrete aggregation in this example are $\max(E_i) = 5$ for the 0.25 degree grid and $\max(E_i) = 7$ for the 0.5 degree resolution.

size does not matter when using spatial or temporal weights to compute conflict exposure. The fact that cells feature homogenous conflict exposure within their boundaries means that larger cells will mask heterogeneity in the intensity of violence in the underlying locations. However, cell size appears to introduce a smaller bias when impact of events on grid cell centroids is allowed to dissipate continuously, as compared to the discrete aggregation case.

For the sample of event locations in Figure 11, the slope parameter $\kappa = 7$ and inflection parameter $\lambda = 0.35$, combined with the 0.25 degree grid resolution, appear to be an appropriate choice. The conservative rate of decay prevents an over-accumulation of events. Since events can impact locations that are not contained within the bounds of the grid cells within which they occur, cells that do not themselves experience events but are in the vicinity of conflict hot spots receive a non-zero conflict exposure value. These parameter and resolution choices thus strike a balance between preventing an over-accumulation of events, yet minimizing the bias that a re-sampling of grid cell centroids would cause on the resulting conflict exposure values.

Future research should explore a relaxing of the assumption of a homogenous rate of decay for all conflict events and in all locations. The methodology could be extended by incorporating covariates that operate at the event-level or the level of the underlying subnational areas. For example, one could assume a different rate of decay for large-scale terrorist attacks versus instances of communal violence. Similarly, the rates of decay in the spatial and/or temporal dimensions could be computed as a function of population density or terrain.

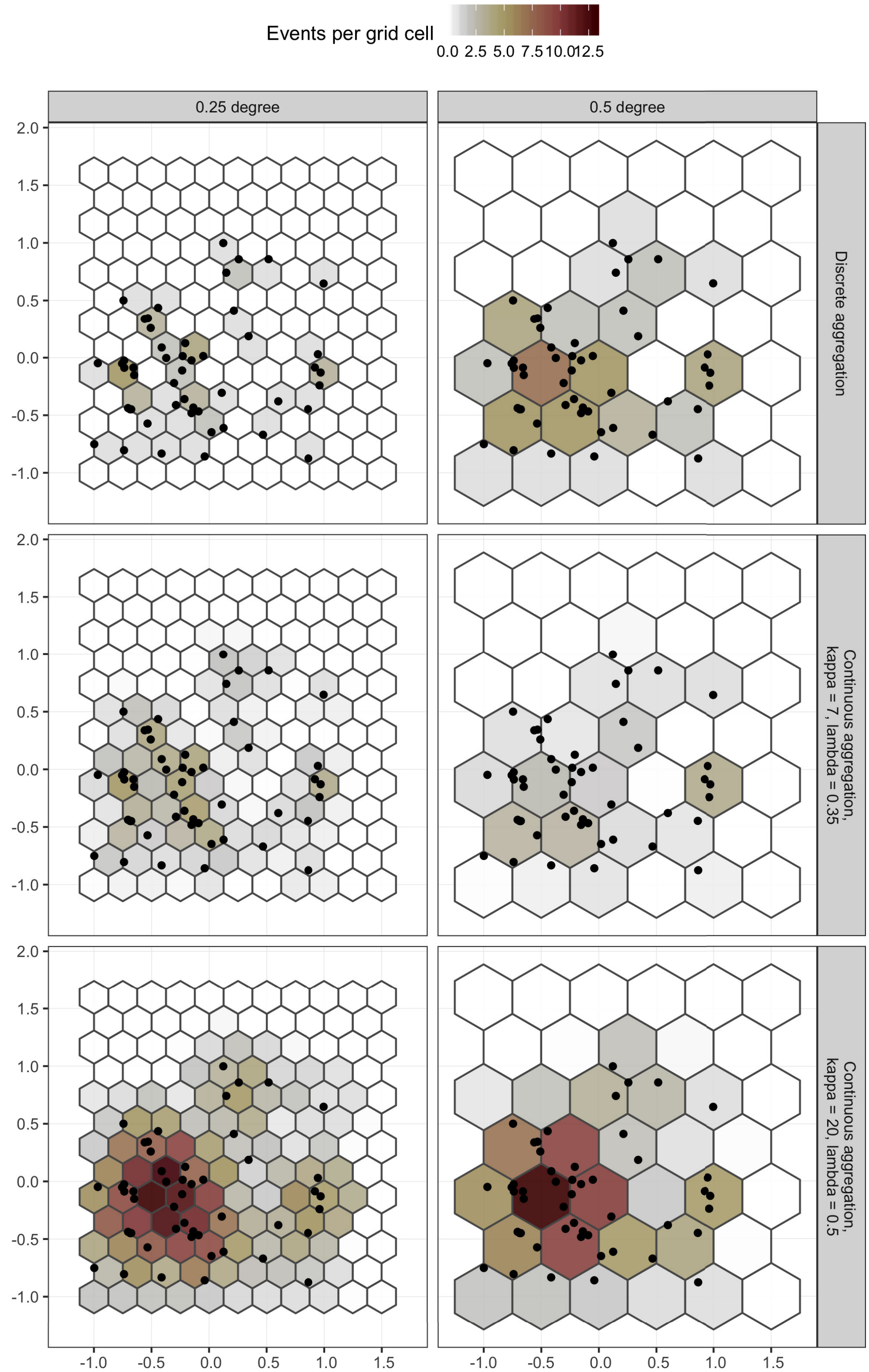


Figure 11: Comparing the dynamics of continuous versus discrete aggregation of events in the spatial dimension using simulated data.

Appendix B Estimation procedure

For each grid cell i, the following procedure is used to estimate the most probable sequence of territorial control over all time periods t.

1. Compute the exposure of the grid cell i in month t to terrorist events E_{it}^{T} and to conventional war acts E_{it}^{C} to all events J, by
   (a) computing the spatial distance d_{ij} of each event j to the centroid of grid cell i in kilometers and weighting it using a logistic decay function,
   (b) computing the temporal distance (each event's age) a^{jt} between the month m when the event occurred and the time of observation of the grid cell t in months and weighting it using a logistic decay function (note that only positive temporal distances a^{jt} are considered), and
   (c) summing the product of the spatial and temporally weighted distances for terrorist and conventional events for each grid cell-month to arrive at E_{it}^{T} and E_{it}^{C}. Note that spatially- and temporally weighted sums of under 0.05 in a grid cell-month are set to zero to avoid later grid-cell months having inflated cumulative event exposures.
2. For each grid cell i in each month t, determine the value of the variable $\mathrm{o}_{it} = f\left(E_{it}^{T}, E_{it}^{C}\right)$.
3. For each grid cell i in each month t, create a sequence of observed outputs $\mathbf{O} \in \{A, B, C, D\}$, where an individual observation o_{it} is determined by Tactics_{it}.
4. For each grid cell i compute the most probable sequence of latent states $\mathbf{Q} \in \{S1, S2, S3, S4, S5\}$ over all time periods t, given the sequence of observed indicator of rebel tactics $\mathbf{O}$ over all time periods t, the time-invariant matrix of transition probabilities Θ, and the time-invariant matrix of emission probabilities Φ via a Hidden Markov Model.

Appendix C Summary statistics for deforestation model

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
$\text{Control}_{i,t}$	3,404	0.9783	0.0755	0	1	1	1
$\Delta\text{Control}_{i,t}$	3,404	0.0062	0.0895	-1	0	0	1
Peace_{t}	3,404	0.2500	0.4331	0	0	0.2	1
$\text{Deforestation}_{i,t}$	3,404	0.0496	0.2172	0	0	0	1

Table VI: Summary statistics for the logistic regression model of deforestation in Colombia on changes in territorial control as a result of the 2016 peace agreement. The unit of analysis for territorial control is annual averages of monthly-level estimates for 0.25 degree hexagonal grid cells.

Appendix D Additional figures

D.1 Transition probabilities

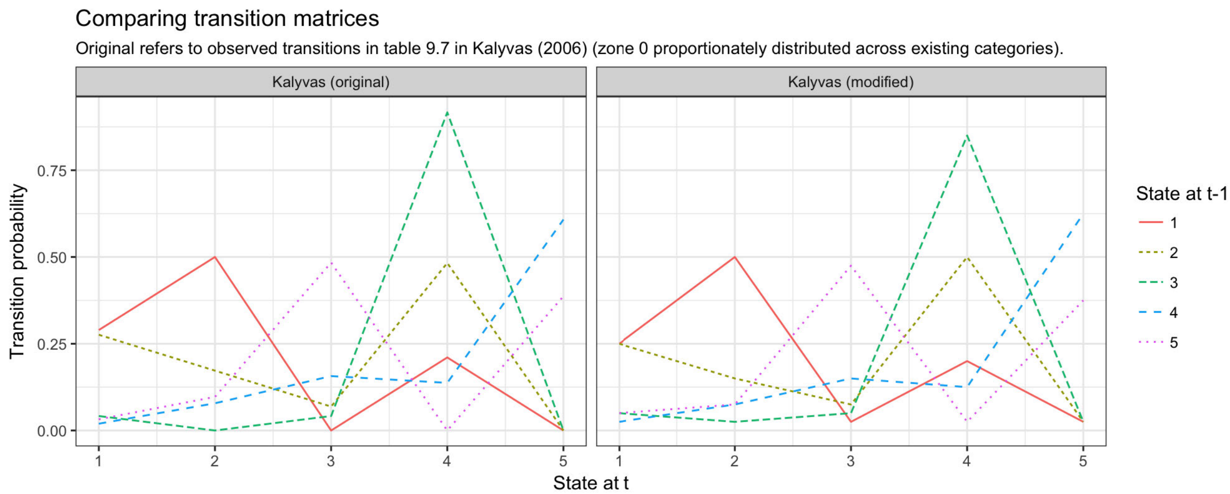


Figure 12: The figure compares the distribution of transition probabilities between the empirical observations from Kalyvas (2006) and the modified transition probabilities in this paper. The graph shows that while the transition probabilities differ slightly, the patterns of transitions between states from $t-1$ to t remain unchanged.

D.2 Case selection

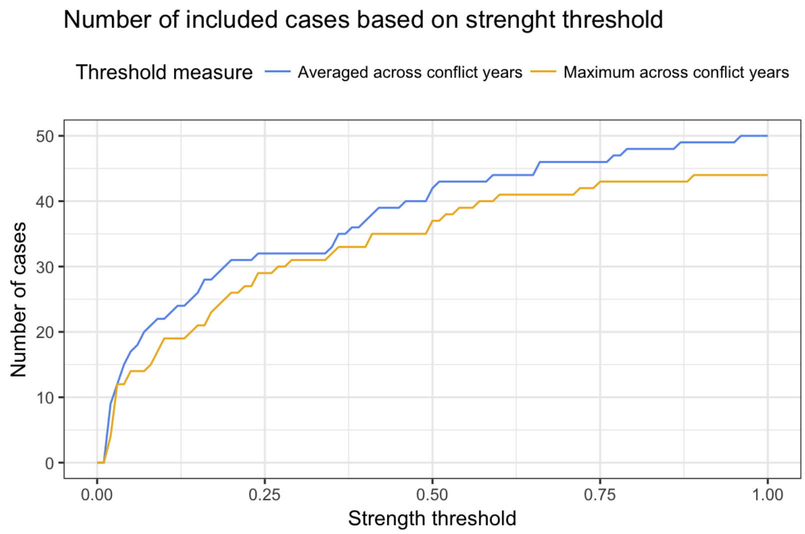


Figure 13: Number of cases that the measurement strategy can be applied to based on different thresholds of power asymmetry between the rebels and the government. Data on power asymmetry come from Polo and Gleditsch (2016).

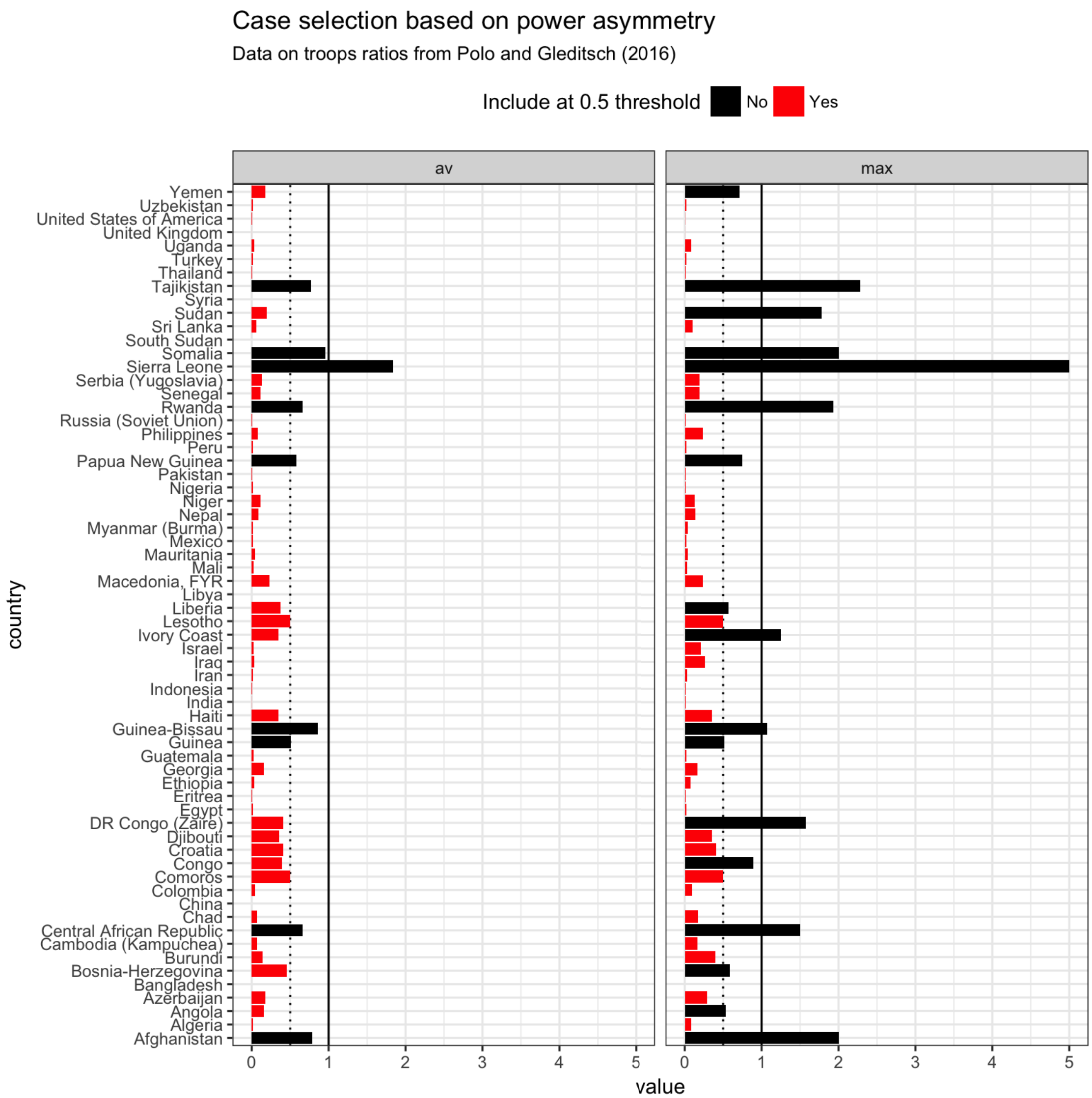
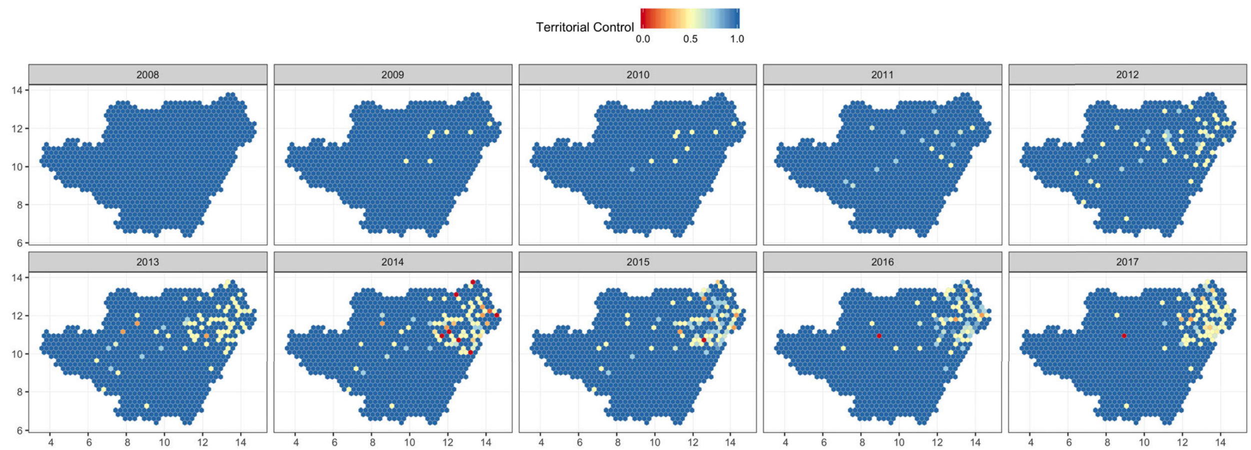
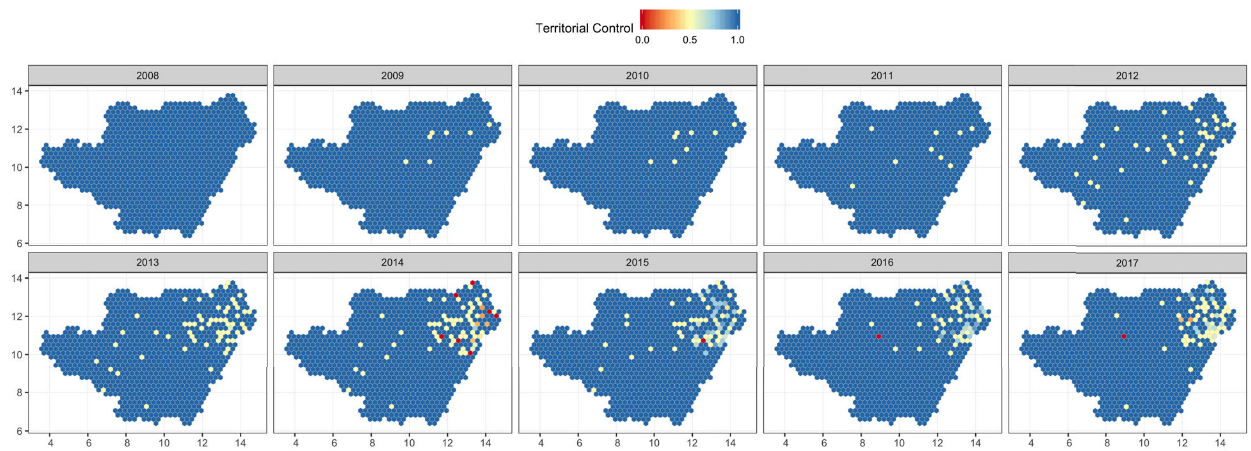


Figure 14: The graph illustrates the selection of cases for which the measurement strategy is applicable based on thresholds in average and maximum rebel-to-government troop ratios over the course of the conflict. Plotted in red are cases that would be included based on a 0.5 threshold indicating rebels that are half as strong as the government forces. Future work will investigate the determination of the most appropriate threshold.

D.3 ACLED validation data for NE Nigeria



(a) Full set of ACLED event types



(b) Subset of ACLED event types (more restrictive)

Figure 15: Yearly averages of monthly-level ACLED validation data values. Values that are closer to 0 indicate full rebel control; values closer to 1 full government control. 0.5 indicates cells that are highly contested.