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

Therese Anders (t.na.anders@gmail.com)

Hertie School

Data Science Lab & SCRIPTS Cluster of Excellence

Abstract

Territorial control is a central variable for civil war research — yet, we lack sufficiently detailed data to capture subnational dynamics and offer cross-country coverage. This article advances a new 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 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). As an observable indicator for rebel tactics, I leverage geo-coded event data and a function of the relative frequency of terrorist attacks and conventional war acts, weighted by time and distance. The model yields estimates of territorial control for asymmetric civil wars at a resolution of 0.25 decimal degree minimum diameter hexagonal grid cells. Validation of estimates for the Colombian and Nigerian civil wars suggests HMMs as a fruitful avenue to estimate spatiotemporal variation in territorial control.

Keywords— Territorial control, civil war, latent variables, event data

Territorial control in civil war influences dynamics of violence, civilian victimization, and rebel governance. While territorial control is universally recognized as theoretically important, empirical studies are scarce because we lack data on who controls an area for most conflicts. For example, the majority of research on the Colombian civil war ignores territorial control as a variable,[1] or uses occurrence of rebel violence as a proxy.[2] However, armed actors' presence cannot be equated with the magnitude of their rule. Rebels might attack a town market, only to immediately retreat to remote hideouts without commanding control or preventing access by government forces. Important questions, such as how rebel territorial control affects internal displacement, cannot be adequately answered without systematically produced data on territorial control that vary temporally and subnationally. The difficulty of measuring territorial control limits the data available for analysis, especially for asymmetric civil wars that do not feature clearly defined frontlines and instead exhibit 'messy patchwork' patterns of control (Kalyvas, 2006: 88).

Given the difficulty of directly observing territorial control in asymmetric conflicts, *how can we estimate changes in territorial control across time and space?* To fill the gap in availability of data, I propose 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 into a machine learning model — leveraging information on variation in rebel tactics based on event data.

Building on existing work regarding the relationship between territorial control and insurgent tactics, I develop a model linking the relative frequency of terrorist attacks and conventional war acts to patterns of 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 expect to observe more insurgent terrorism relative to conventional fighting in areas exhibiting higher levels of government control, and vice versa. Translating this theoretical relationship into measurement, I employ a function of an area's spatially and temporally

[1] An exception is Arjona (2016) who collects data on historic patterns of control in Colombia. However, the data are not publicly available and limited to a few villages.

[2] E.g. Prem, Saavedra & Vargas (2019).

weighted exposure to terrorism and combat events as emissions of the latent variable territorial control. The evolution of control is estimated via a Hidden Markov Model (HMM). I validate estimates for conflicts between the *Fuerzas Armadas Revolucionarias de Colombia - Ejército del Pueblo* (FARC) rebels and the Colombian government, and the Boko Haram insurgency in Nigeria.

The article yields three main contributions. First, I provide an approach toward estimating territorial control that fills a gap in available information on patterns of control for asymmetric conflicts. The estimates allow future research to reduce omitted variable bias, and empirically investigate determinants and consequences of changing control patterns. As a proof of concept, I produce estimates of control for conflicts in Colombia and Nigeria that feature spatiotemporal variation and are produced with a methodology that can be applied cross-nationally.[3] Second, I show how conflict scholars can use their rich theoretical knowledge to inform priors in machine learning applications. Third, I advance a new approach to measure areas' exposure to conflict. Rather than discretely assigning conflict events to grid cells, I compute cells' exposure as the spatially and temporally weighted sum of events. This allows me to account for spatial dependence of conflict exposure in the HMM estimation. The approach presents a valuable methodology for event-based subnational conflict research beyond the measurement of territorial control, especially because it reduces bias from the modifiable areal unit problem (Openshaw & Taylor, 1979).[4]

Territorial control in civil war

Territorial control is a crucial variable for understanding violent conflict, as it shapes armed actors' actions and aspirations (de la Calle & Sánchez-Cuenca, 2012). Gaining or maintaining territorial control is a key objective for both rebels and government.[5] Those who exert control may extract resources (Carter, 2015), pursue collaboration with the population (Ru-

[3] Estimates are available as a simple feature data frame.

[4] See online appendix B.

[5] Following Kalyvas (2006: 111), I define territorial control as the 'extent to which actors are able to establish exclusive rule on a territory.'

bin, 2020; Arjona, 2016), and increase their mobilization base (Stewart & Liou, 2017). Areas of consolidated control can serve as safe havens for combatants and a home base from which future offensives can be coordinated (Arjona, 2016). Gaining control is a pre-condition for establishing non-violent political order. It gives actors the ability to govern non-coercively, for instance via public good provision (Stewart, 2018). Who commands what level of control also conditions civilian behavior in conflict zones, such as voting (García-Sánchez, 2016) and information sharing (Arjona, 2016). Existing research most prominently studies territorial control as a factor determining selective versus indiscriminate civilian victimization (Stewart & Liou, 2017; Quinn, 2015; Kalyvas & Kocher, 2009). Generally speaking, actors commanding less territorial control inflict more indiscriminate violence, and vice versa. Others consider the interplay between civilian cooperation and armed actor coercion in establishment and consolidation of territorial control (Arjona, 2016).

In cross-national studies, scholars frequently rely on binary indicators from the Non-State Actors in Armed Conflict Dataset to operationalize territorial control (Cunningham, Gleditsch & Salehyan, 2013). However, the information is supplied at the group-level — limiting the ability to capture temporal and subnational variation. Alternative approaches operationalizing control as a function of distance from countries' capitals vary subnationally, but do not allow for temporal variation (Schutte, 2017).

A key characteristic 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 capital cities, can account for fragmented spatial patterns observed in asymmetric conflict. Examples of country-specific accounts documenting this fragmentation include recent efforts to create territorial control estimates using conflict event data (Tao et al., 2016), post- or in-conflict surveys (Kalyvas & Kocher, 2009; Arjona, 2016), and the study of military records (Rubin, 2020; Kalyvas, 2006). Approaches by Tao et al. (2016) and Aronson et al. (2017) to hand-code control from reports underlying the Georeferenced Event Dataset (GED) produce fine-grained estimates; however, they are labor-intensive and, at time of writing, no data is publicly released.

While recent years have seen increased interest in studying territorial control, the field suffers from a shortage of data that vary subnationally and temporally, can recover 'patchy' patterns of control, and accommodate cross-country comparison. I improve upon existing approaches by conceptualizing territorial control as a latent variable that can be estimated for small spatial and temporal units. The model is based on a theory linking observable variation in rebel tactics to territorial control.

Tactical choice in asymmetric civil war

Terrorism and civil war are not separate phenomena, and often co-occur (Fortna, 2015; Findley & Young, 2012). Within conflict zones, we observe large variation in the degree of overlap between terrorist attacks and events indicating conventional insurgent tactics, which can be explained by tactical choices of rebels.[6] Insurgents may attack armed forces directly, or indirectly target the government via coercive action intended to spread fear among the public (Carter, 2015). Polo & Gleditsch (2016: 816) state that while the two concepts are not mutually exclusive and conceptually difficult to delineate, '[t]errorism [...] differs from conventional attacks in civil conflicts in that the immediate targets or victims are typically non-combatants, and each individual victim is normally less important than the purpose of conveying a message to the intended audience.' I follow the literature in stipulating three conditions for violence to be coded as terrorism: 1) convey a political message to an audience broader than immediate attack targets, 2) indirectly target states' military capability, and 3) lie outside the realm of 'legitimate warfare activities,' including targeting of noncombatants (START, 2016, Bakker, Daniel W Hill & Moore 2016; Chenoweth 2013).

Territorial control is a key factor in explaining insurgents' use of terrorism as opposed to conventional guerrilla tactics (Carter, 2015). Terrorism arises from insurgents' inability

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

to control territory. 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 & Sánchez-Cuenca, 2015: 810).

Tactical choices in civil war echo actors' maximization of benefits and minimization of costs, subject to resource constraints and opponent actions. All else equal, rebels prefer conventional tactics over terrorism for two reasons. First, pressuring the government by inflicting fear among the population risks alienating civilians whose support rebels depend upon. Second, terrorism does not aid insurgents' immediate goal of securing territory (Carter, 2015) and is thus a second-best choice for rebels unable to directly combat government forces. Territorial control — the degree to which actors rule over an area without interference from opponents — is qualitatively different from actor strength (the size of a group or its material capability). However, they are related. The less territory a group controls, the more it relies on coercive, as opposed to military, power (de la Calle & Sánchez-Cuenca, 2015). 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 (Asal et al., 2012).

Links between rebel tactical choice and territorial control can be observed empirically. Evidence from Nigeria suggests that, once the government re-captured insurgent strongholds in 2015, Boko Haram moved away from fighting for territory and intensified 'its campaign of suicide bombings against soft targets (Reliefweb, 2017). Figure 1 overlays a map of control in Northeast Nigeria with the location of Boko Haram terrorist attacks and events indicative of conventional fighting within two weeks following the measurement of territorial control.[7] Conventional fighting is clustered in contested areas and along borders of insurgent-held territory. With the exception of isolated events along the border with Cameroon, terrorist events are limited to areas of government control.

[7] Map adapted from Reuters, see online appendix F.3.

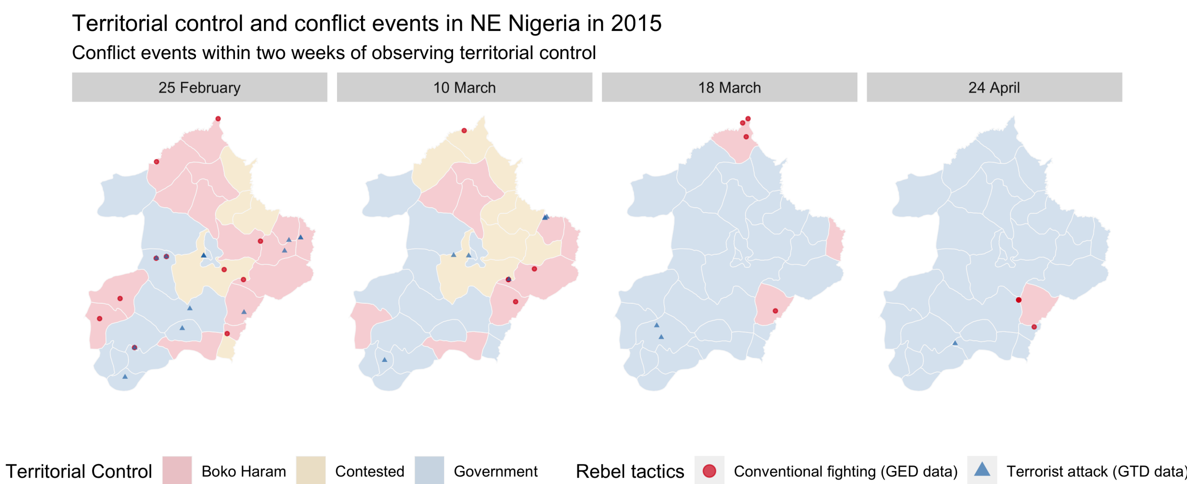

Figure 1. Territorial control and tactical choice

Modeling territorial control

I argue that observed levels of terrorism relative to conventional tactics are indicative of unobserved distributions 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 by either actor, violent events are scarce.

For estimation via a discrete-state HMM, I conceptualize territorial control as a categorical variable with states $\mathbf{Q} = \{R, DR, D, DG, G\}$ ranging from complete rebel (R) to complete government control (G, Table I). These states correspond to zones of territorial control in the literature (Kalyvas & Kocher, 2009; Kalyvas, 2006). For validation, I recode the categorical variable into numerical values from 0 indicating full rebel control to 1 indicating full government control.

Table I here

Table I. States of territorial control

Categorical	Numerical	
R	0	Rebel control
DR	0.25	Disputed, closer to rebel control
D	0.5	Disputed
DG	0.75	Disputed, closer to government control
G	1	Government control

Measuring rebel tactics

I operationalize rebel tactics as a function of areas' relative exposure to terrorist attacks versus events indicating conventional guerrilla fighting. I employ a heuristic that translates a function of the relative frequency of terrorist attacks T_{it} and conventional fighting C_{it} into values of observable emissions o_{it} in area i at time t. Specifically, I compare the probability of observed exposure to terrorist events $T_{it} = P(\lfloor E_{it}^{[T]} \rfloor; \lambda_t^{[T]})$ to the 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 rebel tactic observations $\mathbf{O} = \{O1, O2, O3, O4\}$ (Table II).

Table II here

Table II. Coding of tactics variable O

Tactics O	Observation	Description	Comments
$\mathrm{o}_{it} = O1$	$E_{it}^{[T]} \approx E_{it}^{[C]} \approx 0$	Little to no terrorism and conventional events	Observed exposure values below threshold $xs = 0.1$ truncated to zero
$\mathrm{o}_{it} = O2$	$\mathrm{C}_{it} > \mathrm{T}_{it}$, and $\lvert C_{it} - T_{it} \rvert > m$	More exposure to conventional fighting than terrorism	
$\mathrm{o}_{it} = O3$	$\lvert C_{it} - T_{it} \rvert \leq m$	Similar *non-zero* exposure to terrorism and conventional fighting	$m = 0.025$ (Colombia) and $m = 0.05$ (Nigeria)
$\mathrm{o}_{it} = O4$	$\mathrm{C}_{it} < \mathrm{T}_{it}$, and $\lvert C_{it} - T_{it} \rvert > m$	More exposure to terrorism than conventional fighting	

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.[8] While 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 how observed rebel tactics relate to unobserved levels of territorial control. Prevalence in rebel terrorism is theorized to correspond to higher levels of government control. As previous research posits, 'guerrillas resort to terrorist tactics when they act beyond their areas of control' (Asal et al., 2012: 483). Use of conventional tactics increases with rebel control, suggesting more direct confrontation and more conventional war fighting between the two actors.[9]

Zones of complete rebel or government control are associated with the relative absence of violence ($\text{o}_{it} = O1$). Control is either undisputed, or actors successfully established exclusive rule and prevent opponents from penetrating the area. Disputed areas, which are closer to rebel control, are expected to see higher levels of conventional war fighting ($\text{o}_{it} = O2$). Rebels limit terrorism to reduce harm inflicted upon constituents and maximize popular collaboration (Polo & Gleditsch, 2016). Highly disputed areas exhibit relative parity of conventional and terrorist events ($\text{o}_{it} = O3$). Areas that are disputed, but closer to government control, are expected to exhibit relatively more terrorist attacks than conventional war fighting ($\text{o}_{it} = O4$), because insurgents fighting a highly capable government will substitute conventional war acts with terrorism (Carter, 2015).

[8] $E_{it} = \sum_{j=1}^{J} \left(w_{d_{ij}} \times w_{a_{jt}} \right)$, 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 online appendix B.

[9] The model is based on the simplifying assumption that there are only two conflict parties: a state and a non-state armed actor. Since the unit of analysis is a small grid cell, this assumption does not preclude estimation of territorial control, as long as groups' aspirations for control do not overlap significantly.

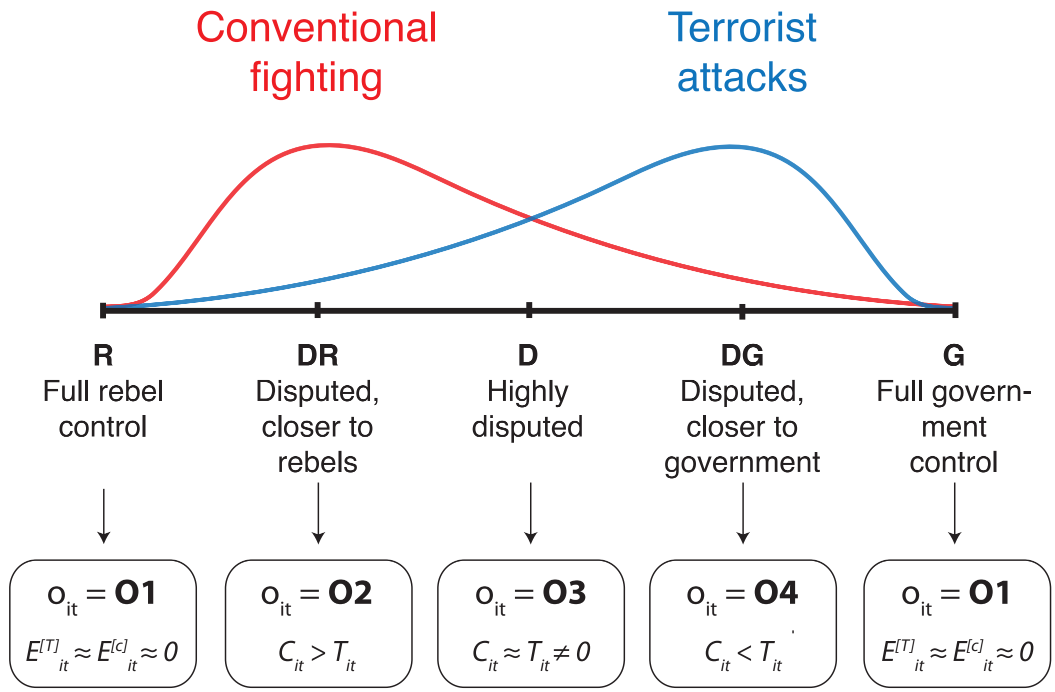

Figure 2. Theorized relationship between rebel tactics and territorial control

Estimation

I estimate territorial control via a Hidden Markov Model (HMM).[10] HMMs are graphical models which uncover the most likely sequence of unobserved states of a discrete latent variable given a set of observable outputs, transition, and emission probabilities.[11] Figure 3 illustrates the conditional dependence between hidden and observed states. HMM computes the maximum likelihood path over the observation period. The most likely sequence of labels is decoded via the Viterbi algorithm.[12]

Territorial control is operationalized as a five-category variable **Q** while observable emis-

[10]A computer science conference proceedings think-piece discusses difficulties of estimation via HMM (Anders et al., 2017). The present article is the first to develop a thorough theoretical model and compute HMM estimates of control, made possible by accounting for spatial dependence between grid cells through continuous spatial decay in observable emissions.

[11] Prominent political science applications of HMMs include forecasting conflict (Schrodt, 2006) or modeling change points (Park, 2012). Existing work uses HMMs to solve a maximization problem: How can we train model parameters to provide the most accurate sequence? The present article solves a decoding problem: What is the sequence of hidden states that best explains observations?

[12]The HMM maximizes $v_t(h) = \max_{g=1}^{N} v_{t-1}(g)\theta_{gh}\phi_h(o_t)$, where h indexes the 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 & Martin, 2019).

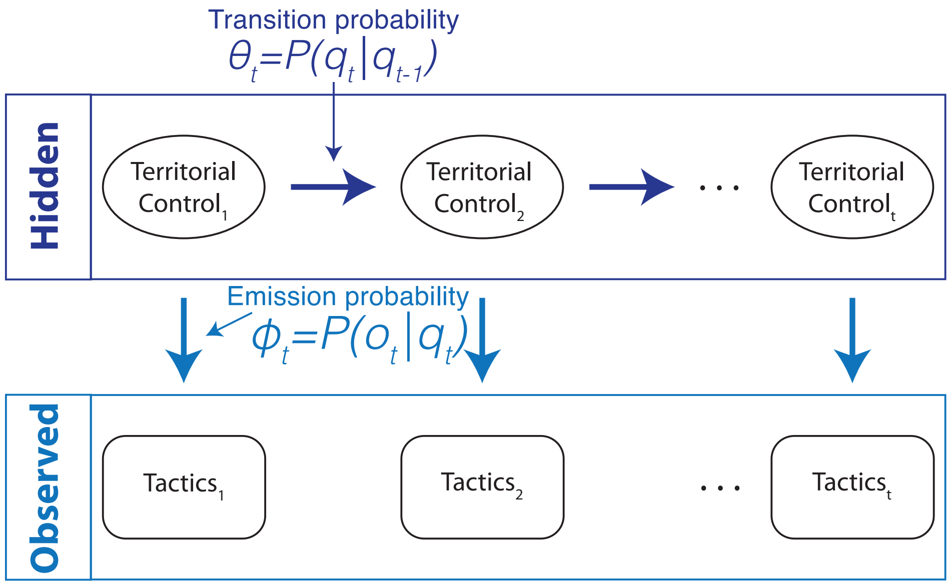

Figure 3. Graphical representation of a HMM

sions **O** have only four possible outcomes. Hence, we cannot linearly map observed rebel tactics onto territorial control. HMMs provide a solution via two features. First, we can specify that full rebel control *and* full government control are equally likely to produce little to no violence ($o_{it} = O1$). Second, HMMs maximize the most probable path over the entire sequence of observations — not a single instance. This allows HMMs to sort out whether an area experiencing little to no violence is more likely to be under full rebel or full government control, given transition probabilities, emission probabilities, and path probability of the previous time step.

Transition probabilities

The transition matrix $\mathbf{\Theta}$ specifies 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}. The transition matrix in Table III is derived from observed transitions between zones of territorial control during the Greek civil war (Kalyvas, 2006: 277).[13]

[13] I modify Kalyvas's matrix to allow for transitions between states of territorial control that empirically have never been observed, see online appendix D.1.

Table III here

Table III. Transition probabilities $\mathbf{\Theta}$ (see Figure 3), based on Kalyvas (2006)

		q_t				
		R	DR	D	DG	G
q_{t-1}	R	0.250	0.500	0.025	0.200	0.025
	DR	0.250	0.150	0.075	0.500	0.025
	D	0.050	0.025	0.050	0.850	0.025
	DG	0.025	0.075	0.150	0.125	0.625
	G	0.050	0.075	0.475	0.025	0.375

Emission probabilities

Emission probabilities specify how observed variation in the co-occurrence of terror and non-terror tactics relates to unobserved levels of territorial control. Each emission probability ϕ in matrix $\mathbf{\Phi}$ (Table IV) answers the following question: Given that the true unobserved state at time t is full rebel control R, for example, what is the probability of observing a signal of no violence $O1$ from the data?

Table IV here

Table IV. Emission probabilities $\mathbf{\Phi}$ (see Figure 3)

		o_t			
		$O1$ $(E^{[T]} \approx E^{[C]} \approx 0)$	$O2$ $(C_{it} > T_{it})$	$O3$ $(C_{it} \approx T_{it} \neq 0)$	$O4$ $(C_{it} < T_{it})$
q_t	R	0.600	0.175	0.175	0.050
	DR	0.050	0.600	0.175	0.175
	D	0.050	0.175	0.600	0.175
	DG	0.050	0.175	0.175	0.600
	G	0.600	0.050	0.175	0.175

Due to a lack of ground truth data on territorial control from which emission probabilities could be learned, they are derived heuristically. For each state of territorial control, the emission that maps onto the theoretical expectation in Figure 2 is expected to be observed with a probability of 0.6, the emission that is theoretically least feasible with a probability of 0.05, and the second and third most feasible with a probability of 0.175.

Consider the first row in Table IV. Based on the theoretical model, I expect an area completely controlled by rebels (R) not to experience terrorist incidents, nor conventional fighting. Hence, I set the probability of observing no events of either type $\mathrm{o}_{it} = O1$, given that the true underlying state is full rebel control, to $P(O1|R) = 0.6$. If the true state is full rebel control, I expect the probability of observing higher levels of conventional fighting than terrorism $O2$, or similar levels of terror and non-terror tactics $O3$, to be much lower at $P(O2|R) = P(O3|R) = 0.175$. The probability of observing more terror than non-terror tactics is expected to be very low $P(O4|R) = 0.05$ if an area is under full rebel control.[14]

Data

The unit of analysis is the grid-cell-month. I leverage geo-coded event data to measure areas' exposure to terrorist and conventional fighting in 0.25 decimal degree minimum diameter hexagonal grid cells. Data on conventional fighting come from GED v18.1 (Croicu & Sundberg, 2017). To achieve the highest level of delineation between terrorist attacks and events indicative of conventional fighting, only GED observations categorized as 'state-based conflict' are considered. Terrorism data come from the Global Terrorism Database (GTD) (START, 2016). GTD codes whether doubt exists that an event constitutes terrorism, rather than combats or crime. I restrict my sample to observations post-1997 — when the 'doubt' variable becomes available — that are unambiguously coded as terrorist attacks.[15] To ensure that events can be accurately related to grid cells for both datasets, only events attributable to at least the second order administrative division are included.[16]

[14] See online appendix G.1 for sensitivity analyses regarding emission probabilities.
[15] Attacks against military targets are excluded. Online appendix E.3 provides robustness checks.
[16] See online appendix A.1.

Case selection

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 high power asymmetry favoring the government prevails. Data on troops ratios suggest 37 conflicts in which rebels are at most half as strong as the government (Polo & Gleditsch, 2016), rendering them candidates for an estimation of territorial control via HMM.[17]

Data to assess the validity of territorial control estimates in asymmetric wars are extremely scarce. In fact, it is the lack of fine-grained data on territorial control that motivates development of the new estimation strategy. Here, I present control estimates for conflicts between FARC and the Colombian government, and the Boko Haram insurgency. Colombia allows for an initial validation via the correlation of territorial control with deforestation after the 2016 peace agreement. Nigeria is included in the Armed Conflict Location & Event Data Project (ACLED) database, which allows for the construction of coarse out-of-sample validation data.

HMM estimates

Colombia

Figure 4 plots territorial control estimates for the FARC and the Colombian government 2006–2017.[18] I exclude the Amazon and Orinoco natural regions, because their low population density raises concerns regarding underreporting conflict events. The maps demonstrate significant spatiotemporal variation, with a few persistent hot spots of rebel controlled areas along the border to Venezuela and the Southwestern coastal region. These patterns are corroborated by existing research. When elected in 2002, president Uribe ignited a military campaign against the FARC. 'From 2002, the state saw a steady recovery of areas that had been under guerrilla control. [...] Certain regions of the country, however, continued to

[17] See online appendix Figures 20 and 21.

[18] The full model is estimated 1997–2017. Figure 4 presents estimates from 2006, after paramilitaries demobilized.

exhibit high levels of violence, especially in the west and near the border with Venezuela' (Arjona, 2016: 92).

The Colombian peace process provides an opportunity for validation. The 2016 peace agreement caused a sudden change in control — forcing rebels to disarm and abandon strongholds. Given the long history of failed peace negotiations, the timing of accord signing was plausibly unexpected. The unanticipated timing minimizes endogeneity bias when relating pre- and post-peace differences in proxy variables to changes in FARC control.

Reports document increased deforestation in the post-peace period, especially in former FARC strongholds. FARC regulated activities causing deforestation, including mining, cattle ranching, and coca production. They 'enforced strict limits on logging by civilians – in part to protect their cover from air raids by government warplanes (Brodzinsky, 2017). Criminal organizations (BACRIM) quickly moved into the power vacuum left by FARC disarmament. BACRIM are reportedly less inclined to regulate mining, logging, and coca production and instead intensified operations — thus increasing deforestation (Prem, Saavedra & Vargas, 2019; Reardon, 2018).

If territorial control estimates are valid, deforestation should be more likely in areas experiencing larger changes in control after 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 indicating whether grid cell i experienced deforestation in year t. $\Delta Control$ denotes changes in average annual control between t and $t-1$. Positive values indicate changes toward government control. $Peace_t$ is a dummy for 2016. A positive coefficient on the interaction $\Delta Control_{i,t} \times Peace_t$ indicates a higher probability of deforestation in areas experiencing changes in control as a result of the agreement. The model is estimated via logistic regression. Data on deforestation from 2013 to 2016 come from satellite images via the forest monitoring system from the *Instituto de Hidrología, Meteorología y Estudios Ambientales.*

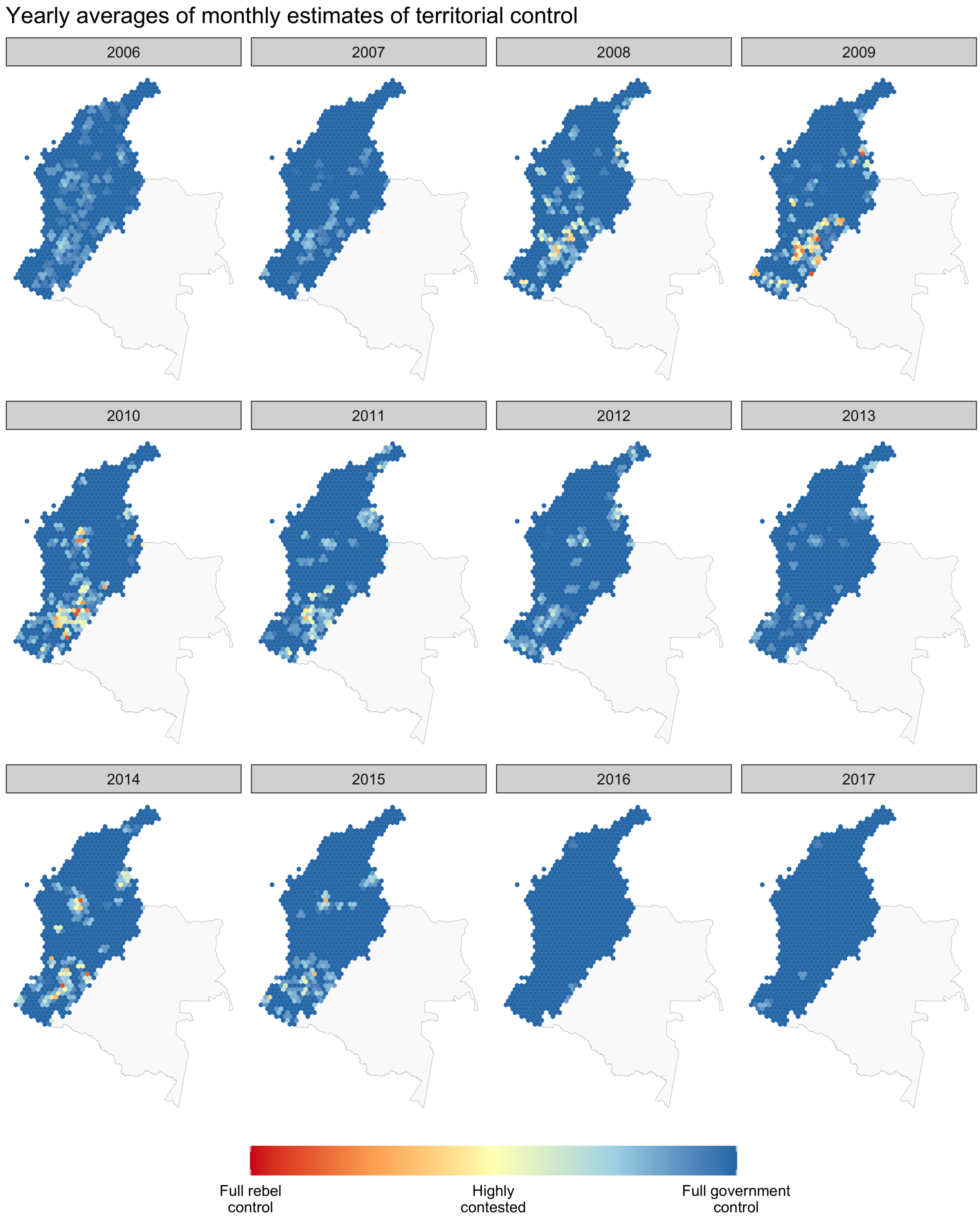

Figure 4. Estimated territorial control in Colombia (N = 851)

Table V here

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

	Deforestation$_{i,t}$		
	(1)	(2)	(3)
$\Delta\text{Control}_{i,t} \times \text{Peace}_t$		3.59^{*}	3.29^{*}
		(1.68)	(1.63)
$\Delta\text{Control}_{i,t}$	-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)
$\text{Deforestation}_{i,t-1}$			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 supports the hypothesis that changes in territorial control resulting from the peace agreement are associated with a higher probability of deforestation. The interaction's coefficient in model 2 is positive and statistically significant at the minimum 5% level. A change of 0.25 in territorial control, for example from DG to G, is associated with an increase in the predicted probability of deforestation from 5.6% to 9.1%. The interaction's coefficient remains statistically significant upon including a lagged dependent variable in model 3. Thus, HMM estimates produce results in line with observed empirical relationships between changes in territorial control and deforestation.[19]

Nigeria

Figure 5 plots territorial control estimates for Boko Haram and the government for 15 Northeast Nigerian states 2009–2017. In late 2014 and early 2015, reports suggest that insurgents controlled 15 localities in the border region with Cameroon, and had partial control over additional 15 local government areas (AI, 2015). In early to mid 2015, Nigerian government and African Union troops launched attacks against Boko Haram that caused insurgents to lose a majority of their territory (CISAC, 2018). However, reports from 2018 cast doubt over the government's claim that it drove out insurgents, suggesting that Boko Haram controlled parts of Borno state via roadblocks, stop and search operations, and the collection of 'taxes' for protection (Al-Amin, 2018).

The estimates illustrate the onset of the Boko Haram insurgency in 2009.[20] Starting in 2011, Boko Haram is estimated to gain control in parts of Borno. Insurgents subsequently established strongholds that are at first scattered throughout the region. By 2014, the estimates show consolidation of insurgent control in the Northeast. Following the Nigerian government and African Union offensive, Boko Haram is estimated to have lost significant amounts of territory in 2015 and 2016. By 2017, its strongholds are limited to the border region with Cameroon.

The inclusion of Nigeria in the Armed Conflict Location & Event Data Project (ACLED,

[19] Robustness checks in online appendix E.3.

[20] The model is initiated in 2008 with a strong prior of full government control.

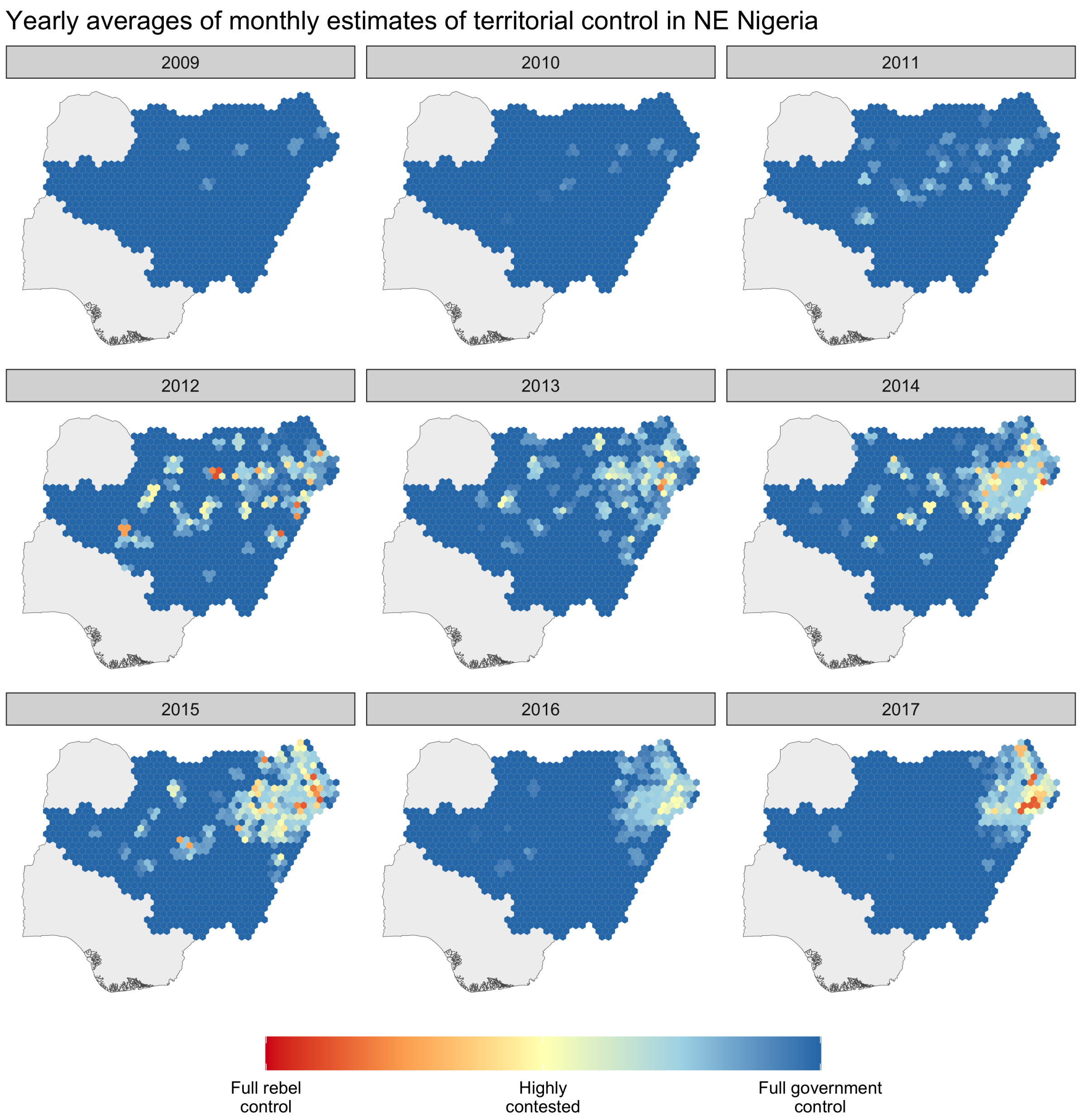

Figure 5. Territorial control in Northeast Nigeria (N = 943)

v8.0) allows me to construct an out-of-sample validation set (Raleigh et al., 2010). ACLED codes whether an event resulted in rebels gaining control or establishing a base (coded as R), battles with no changes in control (D), the government gaining control or establishing a base (G), and instances of remote violence (DR for government remote violence; DG for insurgent remote violence).[21] The validation set adopts the assumption that remote violence indicates areas that are contested but closer to either rebel or government control, depending on the perpetrator.[22]

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 their status of territorial control. If multiple events occur in the same grid-cell-month, I average across them. Cells with no events in a given month are imputed to mirror the previous month's control up to a duration of six months, unless a new event is observed. If a cell does not experience any violence in the previous six months, it is assumed to be under government control.[23] Cells experiencing zero events are assumed to be under full government control.

Due to strong assumptions necessary for constructing validation data from ACLED, as well as concerns over reporting bias (Eck, 2012), conclusions from comparison with 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, ACLED validation data offers the best opportunity for out-of-sample validation of territorial control in Nigeria available to date.

Figure 6 plots the correlation between territorial control estimates and ACLED validation data — ranging between 0.14 and 0.46 and, on average, increasing over time. Figure 6 reveals high volatility in monthly-level estimates. Aggregating monthly estimates to yearly averages yields a more stable agreement with the validation data. I recommend future research to

[21] Categorical values are mapped to their continuous expressions in Table I. Sauter (2017) and Wimmer & Miner (2019) employ similar coding procedures. ACLED contains a small number of events for which manual coding is necessary to determine which actor gained control; documentation available upon request.

[22] I discuss a stricter coding in online appendix F.

[23] The conflict is highly active 2009–2017, rendering 6 months a reasonable upper bound. Online appendix Figures 9 and 10 plot yearly averages for 6- and 12-month thresholds. Online appendix figure 11 shows that using a 3-month or 12-month window yields similar results.

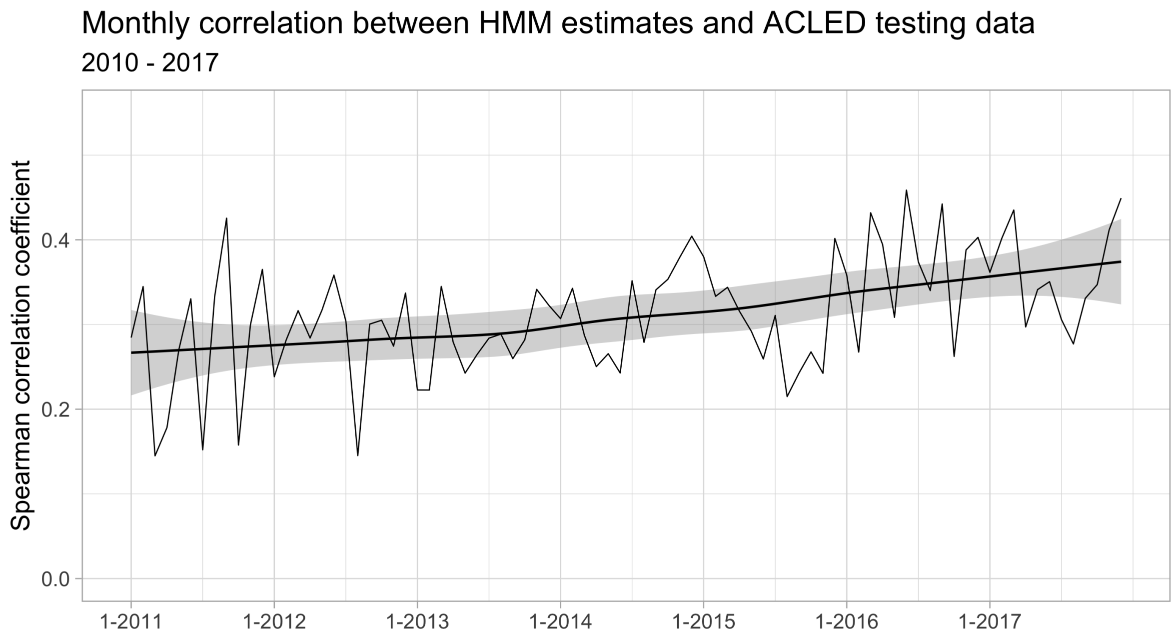

Figure 6. Correlation between HMM estimates and ACLED validation data, including loess smoothed average

incorporate the variation of monthly estimates as uncertainty into future models of territorial control.

Compared to the validation data, the HMM estimates appear to slightly overstate levels of rebel control and understate levels of contestation.[24] However, the annual averages over monthly HMM estimates in Figure 5 uncover spatial patterns and major temporal trends in the distribution of territorial control. In particular, the model captures the reduction of insurgent territorial control from 2015 to 2016.

Conclusion

I propose a novel measurement approach for estimating 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 territorial control estimates for

[24]See online appendix F.

conflict between the FARC and the Colombian government, and the Boko Haram insurgency. Validation of the Colombia estimates using patterns of deforestation after 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 methodology allows for generation of territorial control estimates that utilize publicly available event data, capture fine-grained spatiotemporal subnational patterns, and can be computed for a cross-section of countries. The estimates yield a valuable source of information for analyses of civil war for both within-country and comparative studies. The estimates will also be instrumental for investigating cross-border dynamics. Rebel groups like Boko Haram operate across international frontiers and often leverage the remoteness of border regions to their advantage. The use of a continuous measure of areas' exposure to conflict events overcomes a major limitation of standard HMMs by allowing for spatial dependency in emissions, rendering the model suitable to estimate latent constructs that feature spatial and temporal variation.

The results demonstrate that HMMs are a fruitful approach to address the lack of data on territorial control in asymmetric civil wars. However, the approach has limitations, offering opportunities for improvement in future research. Currently, the applicability of the model is limited to asymmetric conflicts fought between two actors and future research should explore the extension to more non-state actors. Since the model relies on observed instances of violence, it is susceptible to reporting biases in conflict event data and non-violent changes of territorial cannot be captured.

A 'ground truth' of territorial control information does not exist for most conflicts, especially asymmetric civil wars. This presents a limitation and an opportunity.[25] It likely reduces validity and reliability of the estimates because model parameters cannot be learned from data. However, the derivation of transition probabilities that specify prior beliefs re-

[25] For a discussion of limitations, underreporting bias, and sensitivity to parameter choices, see online appendix G.

garding evolution of territorial control illustrates how domain-specific knowledge can be leveraged to inform model parameters in machine learning applications. For 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 in 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. As more data becomes available, future work should incorporate such information to improve the accuracy of estimates and facilitate more empirical research on territorial control in civil war.

Data replication— Data and `R` scripts for the empirical analyses in this article can be found at http://www.prio.org/jpr/datasets and https://github.com/thereseanders/territorialcontrol.

Acknowledgements— I am grateful to the audiences at meetings of the Peace Science Society, the Society for Political Methodology, International Studies Association, American Political Science Association, the USC Networked Democracy Lab, and CIS Working Paper Series. Special thanks to Lisa Argyle, Adam Badawy, Pablo Barberá, Sam Bell, Charity Butcher, Christopher Fariss, Michael Findley, Benjamin Graham, Joseph Huddleston, Kyosuke Kikuta, Satish Kumar, Brett Ashley Leeds, James Lo, Jonathan Markowitz, Paul Poast, Ashly Townsen, James Walsh, Hong Xu, and three anonymous reviewers for insightful comments and discussion.

References

Al-Amin, Muhammad (2018) Boko Haram Islamists still control parts of northeastern Nigeria. *DW*.

Amnesty International (2015) Boko Haram at a glance (https://www.amnesty.org/en/latest/news/2015/01/boko-haram-glance/).

Anders, Therese; Hong Xu, Cheng Cheng & TK 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*.

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

Aronson, Jacob; Deniz Ciland, Paul K Huth & James I Walsh (2017) An enhanced dataset of territorial control by conflict actors. Paper presented at the 2017 Meeting of the American Political Science Association.

Asal, Victor; Luis De La Calle, Michael Findley & Joseph Young (2012) Killing civilians or holding territory? How to think about terrorism. *International Studies Review* 14(3): 475–497.

Bakker, Ryan; Jr Daniel W Hill & Will 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.

Brodzinsky, Sibylla (2017) Deforestation soars in Colombia after Farc rebels' demobilization. *The Guardian*.

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

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

CISAC (Stanford University) (2018) Boko Haram (https://cisac.fsi.stanford.edu/mappingmilitants/profiles/boko-haram?highlight=boko%2Bharam).

Croicu, Mihai & Ralph Sundberg (2017) UCDP GED codebook version 17.1.

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

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

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

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

Fortna, Virginia P (2015) Do terrorists win? Rebels' use of terrorism and civil war outcomes. *International Organization 69*: 519–556.

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

Jurafsky, Daniel & James H Martin (2019) Speech and Language Processing (3 ed.). chapter Hidden Markov Models, , 548–563.

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

Kalyvas, Stathis N & Matthew 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].

Openshaw, Stanley & Peter J Taylor (1979) A millon or so correlation coefficients: Three experiments on the modifiable areal unit problem. In: Neil Wrigley (ed.) *Statistical applications in the spatial sciences.* London: Pionchapter 5, , 127–144.

Park, Jong H (2012) A unified method for dynamic and cross-sectional heterogeneity: Introducing hidden markov panel models. *American Journal of Political Science* 56(4): 1040–1054.

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

Prem, Mounu; Santiago Saavedra & Juan F Vargas (2019) End-of-Conflict Deforestation: Evidence From Colombia's Peace Agreement. Working Paper 288. Households in Conflict Network (HiCN).

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

Raleigh, Clionadh; Andrew Linke, Håvard Hegre & Joakim Karlsen (2010) Introducing ACLED: An armed conflict location and event dataset: Special data feature. *Journal of Peace Research* 47(5): 651–660.

Reardon, Sara (2018) FARC and the forest: Peace is destroying Colombia's jungle – and opening it to science. *Nature 558*: 169–170.

Reliefweb (2017) Analysis: Scrutinising the Boko Haram resurgence (https://reliefweb.int/report/nigeria/analysis-scrutinising-boko-haram-resurgence).

Rubin, Michael A (2020) Rebel territorial control and civilian collective action in civil war: Evidence from the Communist insurgency in the Philippines. *Journal of Conflict Resolution* 64(2–3): 459–489.

Sauter, Melanie S (2017). Dangerous Territory. A subnational study on how territorial control affects violence against aid workers in Africa. Master's thesis University of Oslo.

Schrodt, Philip A (2006) Forecasting conflict in the Balkans using Hidden Markov Models. In: Robert Trappl (ed.) *Programming for Peace: Computer-Aided Methods for International Conflict Resolution and Prevention.* Dordrecht: Springer Netherlands, 161–184.

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

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

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

Tao, Ran; Daniel Strandow, Michael Findley, Jean-Claude Thill & James I Walsh (2016) A hybrid approach to modeling territorial control in violent armed conflicts. *Transactions in GIS* 20(3): 413–425.

Wimmer, Andreas & Chris Miner (2019) The strategic logic of ethnoterritorial competition: Violence against civilians in Africa's civil wars. *Journal of Global Security Studies*.

Biography— Therese Anders, PhD in Political Science & International Relations (University of Southern California, 2020); Postdoctoral research fellow at the Hertie School Data Science Lab and academic coordinator of the SCRIPTS Cluster of Excellence Data and Methodology Center (2019–2020).