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

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 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 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 estimates of territorial control for asymmetric civil wars at a resolution of 0.25 decimal 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.

Territorial control in civil war influences dynamics of violence, civilian victimization, and rebel governance, and the financing of rebellion. 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 the abundant research on the Colombian civil war ignores territorial control as a variable,[1] or uses the occurrence of rebel violence as a poor proxy.[2] However, the presence of an armed actor cannot be equated with the magnitude of their rule. Rebels might perpetrate an attack on a town market only to immediately retreat to remote hideouts without commanding control or being capable of preventing access to the area by government forces. Important questions, such as how rebel territorial control affects patterns of internal displacement or economic life, cannot be adequately answered without systematically produced data on territorial control that vary temporally and sub-nationally. The dearth of data is due to territorial control being extremely difficult to measure, in particular 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 measuring territorial control in asymmetric conflicts via direct observation, *how can we estimate changes in territorial control across time and space?* To fill the gap in the 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 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 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 expect to observe more insurgent terrorist attacks rel-

[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 et al. (2019).

ative 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 of the latent variable territorial control. The most likely evolution of territorial control is estimated via a Hidden Markov Model (HMM). I validate estimates for the conflict between the *Fuerzas Armadas Revolucionarias de Colombia - Ejército del Pueblo* (FARC) rebels and the Colombian government and the Boko Haram insurgency in Nigeria.

The project yields three main contributions. First, I provide an approach toward the estimation of territorial control that has the potential to fill the gap in the availability of information on patterns of territorial control for asymmetric conflicts. Estimates of territorial control will allow future research to reduce omitted variable bias and empirically investigate the determinants and consequences of changing patterns of control. As a proof of concept, I produce estimates of territorial control for the Nigerian and Colombian conflicts that accommodate high levels of spatio-temporal 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 an area's exposure to conflict. 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. 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, in particular because it reduces bias due to the modifiable areal unit problem (Openshaw and Taylor, 1979).[4]

[3] Estimates are available as a simple feature data frame and shapefile on GitHub [link upon publication].
[4] See online appendix B.

Territorial control in civil war

Territorial control is a crucial variable for understanding violent conflict, as it shapes armed actors' actions and aspirations, in particular "the dynamics of bargaining, recruitment, and lethality" (de la Calle and Sánchez-Cuenca, 2012, 583). Maintaining or gaining territorial control is a key objective for both rebels and government.[5] Those who exert control over an area have the opportunity to extract resources (Carter, 2015), pursue the collaboration of the population (Rubin, 2019; Arjona, 2016), and increase their mobilization base (Stewart and Liou, 2017). 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 also 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). 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). 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 (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 to code whether a group exercised control over territory (Cunningham et al., 2013). 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).

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

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 fragmented spatial patterns observed in empirical studies of asymmetric conflict. 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 military records (Rubin, 2019; Kalyvas, 2006). The approach by Tao et al. (2016) and Aronson et al. (2017) to hand-code 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 data that vary sub-nationally 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 model set-up is based on a theoretical account that links observable variation in rebel tactics to territorial control.

Tactical choice in asymmetric civil war

Terrorism and civil war are not separate phenomena and instead co-occur frequently (Fortna, 2015; Findley and Young, 2012). Within conflict zones, we observe large variation in the degree of overlap between terrorist attacks and events that are indicative of conventional insurgent tactics, which can be explained by tactical choices of rebels.[6]

Insurgents can either attack states' armed forces directly, or indirectly target the government via coercive action intended to spread fear among the public (Carter, 2015). 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 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, not directly target the state's military capability, 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). Terrorism arises from insurgents' inability 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 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 risk of alienating civilians whose support rebels depend upon. Second, terrorism does not aid insurgents' immediate goal of securing territorial gains (Carter, 2015). Terrorism is thus a second-best choice for rebels that do not command control over a given area and are unable

[6] 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.

to engage in direct fighting with government forces.

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, the two are related. 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). 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).

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."[7] Figure 1 overlays a map of territorial control in Northeast Nigeria with the location of Boko Haram terrorist attacks and events indicative of conventional fighting within the two weeks following the measurement of territorial control.[8] Conventional fighting is clustered in contested areas and along the borders of insurgent-held territory. With the exception of isolated events in the border region with Cameroon, terrorist events are limited to areas of government control.

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

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

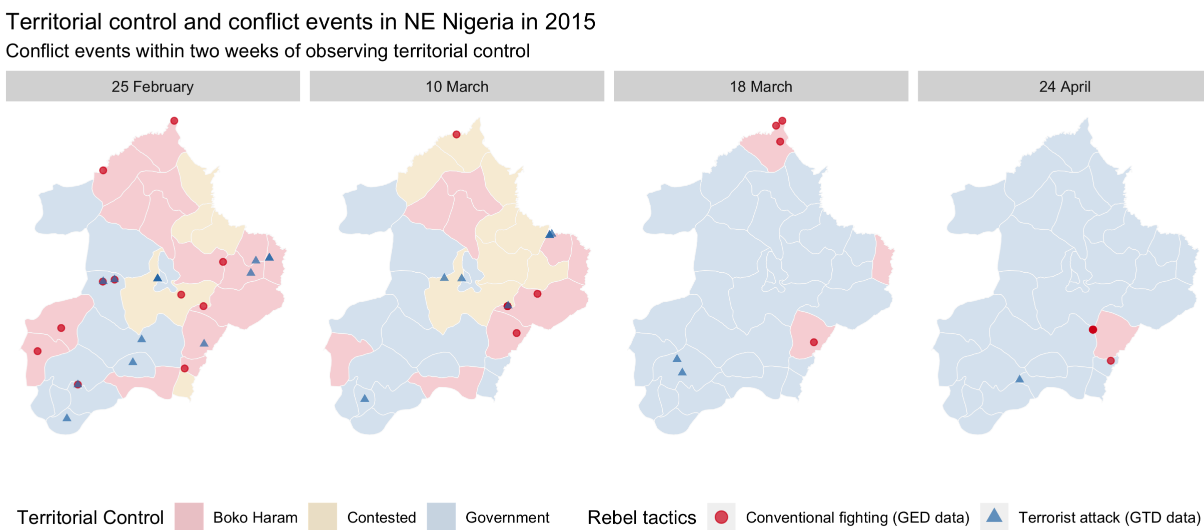

Figure 1: Territorial control and tactical choice.

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.

For the purpose of 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), with levels of contestation in between (Table I). These states correspond to the categorization of zones of territorial control in the existing literature (Kalyvas and Kocher, 2009; Kalyvas, 2006). For the validation exercises, I recode the categorical variable into numerical values from 0 indicating full rebel control to 1 indicating full government control (Table I, column 2).

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

Table I: States of territorial control.

Measuring rebel tactics

Rebel tactics are observable emissions of the unobserved latent variable territorial control. I operationalize rebel tactics area as a function of an area's relative exposure to terrorist attacks versus events indicative of conventional guerrilla fighting. 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 $\mathbf{O} = \{O1, O2, O3, O4\}$ of the rebel tactic variable O, as outlined in Table II.

Tactics O	Observation	Description	Comments
$o_{it} = O1$	$E_{it}^{[T]} \approx E_{it}^{[C]} \approx 0$	Little to no terrorism and conventional events.	Observed exposure values below a threshold $xs = 0.1$ are truncated to zero.
$o_{it} = O2$	$C_{it} > T_{it}$, and $\lvert C_{it} - T_{it} \rvert > m$	More exposure to conventional fighting than terrorism.	
$o_{it} = O3$	$\lvert C_{it} - T_{it} \rvert \leq m$	Similar *non-zero* exposure to terrorism and conventional fighting.	Overlap of zero-inflated Poisson probabilities specified as $m = 0.025$.
$o_{it} = O4$	$C_{it} < T_{it}$, and $\lvert C_{it} - T_{it} \rvert > m$	More exposure to terrorism than conventional fighting.	

Table II: Coding of tactics variable 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.[9] 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.

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

Mapping rebel tactics onto territorial control

Figure 2 illustrates how observed rebel tactics relate to unobserved levels of territorial control. 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" (Asal et al., 2012, 483). The use of terrorism has an inverse relationship with the level of rebel territorial control. 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.[10]

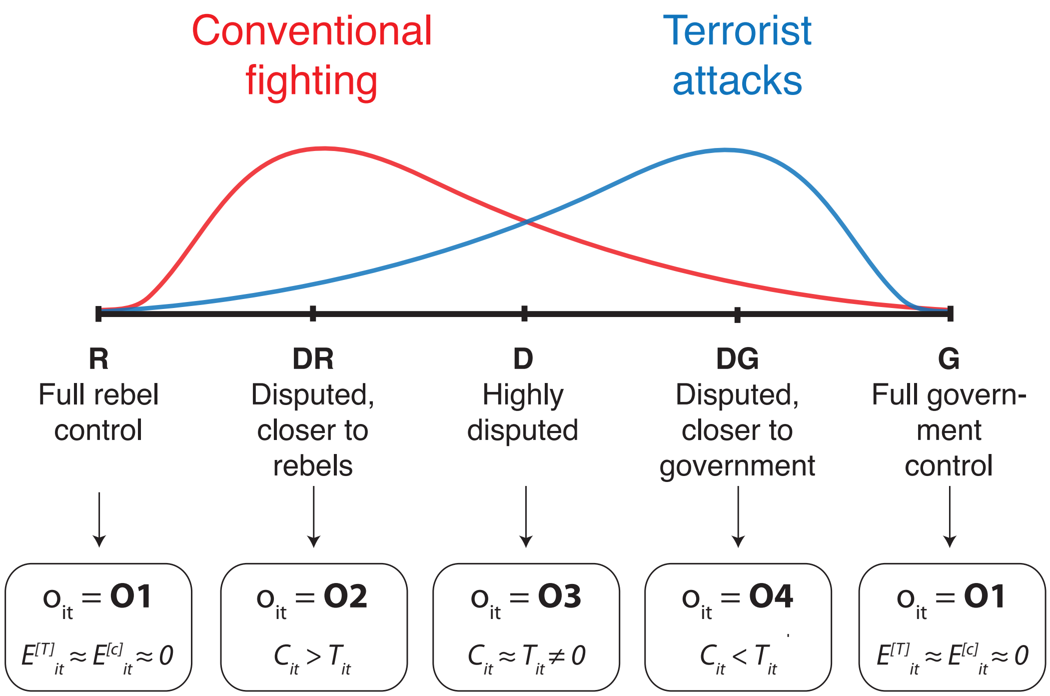

Figure 2: Theorized relationship between rebel tactics and territorial control.

Based on this theoretical model, zones of full rebel or full government control are associated with the relative absence of violence ($o_{it} = O1$). Control is either undisputed, or actors successfully established exclusive rule and prevent opponents from penetrating the area. Areas that are disputed, but closer to rebel control, are expected to see higher levels of

[10]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. This assumption may not hold in all conflicts. However, because the unit of analysis is a small grid cell, this assumption does not preclude an estimation of territorial control, as long as groups' aspirations for control do not overlap significantly.

conventional war fighting than terrorism ($o_{it} = O2$). Rebels limit terrorism to reduce harm inflicted on their constituent population to maximize popular collaboration (Polo and Gleditsch, 2016). Highly disputed areas are characterized by a relative parity of conventional and terrorist events ($o_{it} = O3$). Areas that are disputed, but closer to government control, are expected to exhibit relatively more terrorist attacks than conventional war fighting ($o_{it} = O4$), because insurgents fighting a highly capable government will substitute conventional war acts with terrorism (Carter, 2015).

Estimation

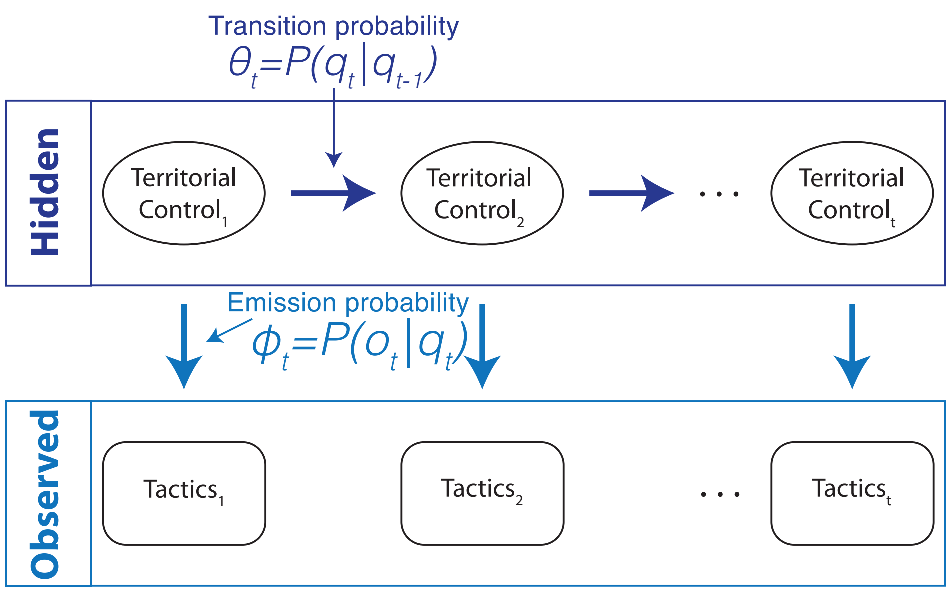

Figure 3: Graphical representation of a HMM.

I estimate territorial control via a Hidden Markov Model (HMM).[11] HMMs are graphical models to uncover the most likely sequence of unobserved states of a discrete latent variable given a set of observable outputs, transition, and emission probabilities.[12] Figure 3 illus-

[11] A computer science conference proceedings thinkpiece discusses difficulties of the estimation via HMM (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.

[12] 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 paper solves a decoding problem: What is the sequence of hidden states that best explains observations?

trates the conditional dependence between hidden and observed states. 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.[13]

Territorial control is operationalized as a five-category variable $\mathbf{Q}$ while observable emissions $\mathbf{O}$ have only four possible outcomes. Hence, we cannot linearly map observed rebel tactics onto territorial control. HMMs provide a solution to this problem via two unique features. First, we can specify that both 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 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

				q_t		
		R	DR	D	DG	G
	R	0.250	0.500	0.025	0.200	0.025
	DR	0.250	0.150	0.075	0.500	0.025
q_{t-1}	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

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

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}. The transition matrix in Table III is derived from observed transitions between zones of territorial control during the Greek civil war (Kalyvas, 2006, 277).[14]

[13]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 and Martin, 2017).

[14]I modify Kalyvas's matrix to allow for transitions between states of territorial control that empirically

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, for example, full rebel control R, what is the probability of observing, for example, a signal of no violence $O1$ from the data?"

		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

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

Due to a lack of ground truth data on territorial control from which the emission probabilities could be learned, in this application, 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, respectively.

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 conventional fighting. Hence, I set the probability of observing no events of either type $o_{it} = O1$, 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

have never been observed, see online appendix D.1.

tactics is expected to be very low $P(O4|R) = 0.05$ if an area is under full rebel control.[15]

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 decimal degrees. Data on conventional war acts come from the Georeferenced Event Dataset (GED) version 17.1 (Croicu and Sundberg, 2017). To achieve the highest level of delineation between terrorist attacks and events that are indicative of conventional fighting, only GED observations categorized as occurring within the realm of "state-based conflict" are considered. Data on terrorist incidents come from the Global Terrorism Database (GTD) (START, 2016). GTD codes whether there is 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.[16] To ensure that events can accurately be related to grid cells, for both GED and GTD, only events attributable to at least the second order administrative division are included.[17]

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 a 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 and Gleditsch, 2016) — rendering them candidates for an estimation of territorial control via HMM.[18]

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

[15] See online appendix G.1 for a sensitivity analysis of estimates to the choice of emission probabilities.
[16] Attacks against military targets are excluded. Online appendix E.3 relaxes this assumption.
[17] See online appendix A.1.
[18] See online appendix Figures 19 and 20.

present estimates of territorial control for the 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 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 data on territorial control.

HMM estimates of territorial control

Colombia

Figure 4 plots territorial control estimates for the FARC and the Colombian government 2006–2017.[19] I exclude the Amazon and Orinoco natural regions, because their low population density raises concerns over underreporting of 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 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 out-of-sample validation. The peace agreement in 2016 caused a sudden change in territorial control — forcing rebels to disarm and abandon strongholds. The timing of the signing of the accord was plausibly unexpected, given the long history of failed peace negotiations. The unanticipated timing minimizes endogeneity bias when relating the 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

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

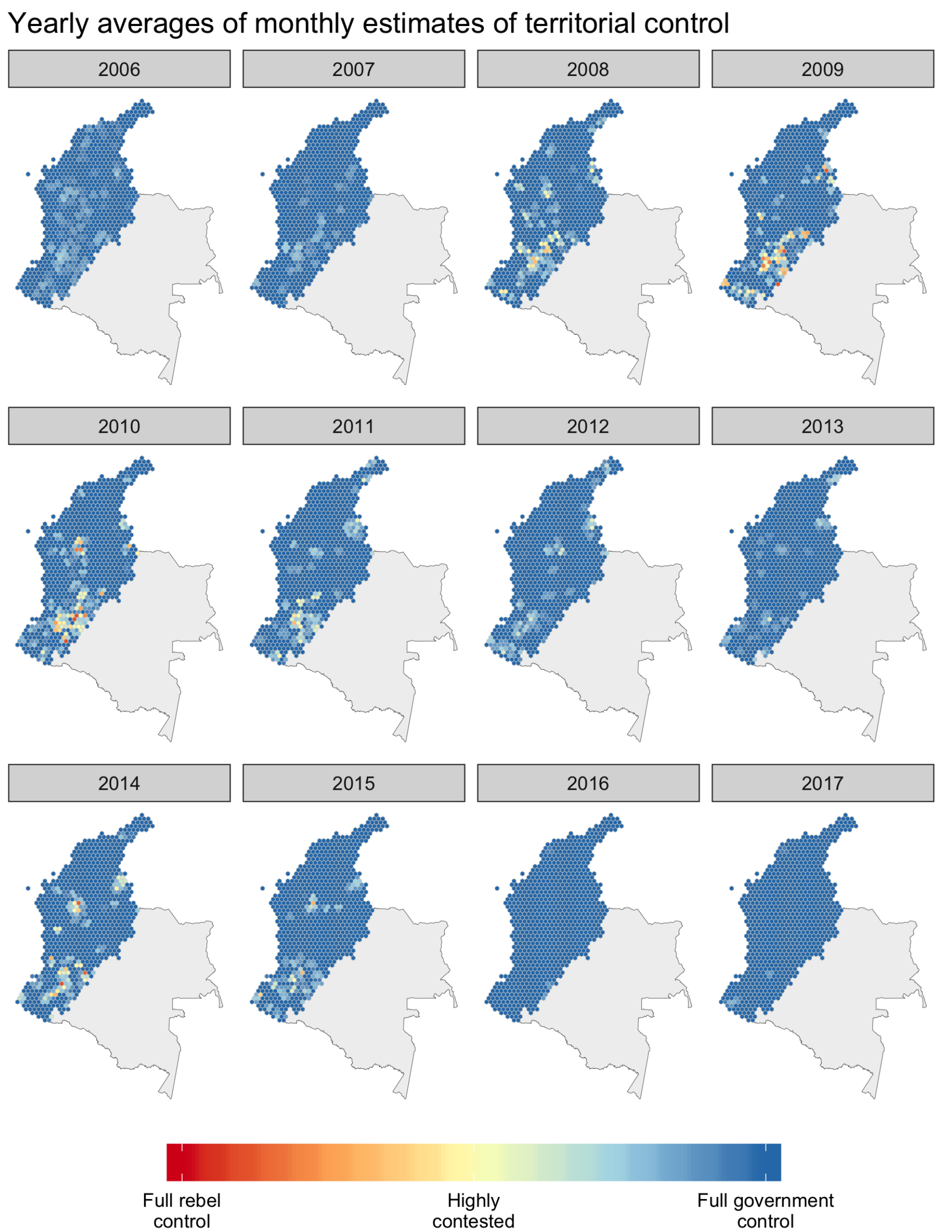

Figure 4: Estimated territorial control in Colombia (N = 851).

FARC strongholds. FARC regulated activities that cause deforestation, such as mining, cattle ranching, and coca production within their strongholds. They "enforced strict limits on logging by civilians – in part to protect their cover from air raids by government warplanes."[20] 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 are instead intensify these operations — thus increasing deforestation (Prem et al., 2019; Reardon, 2018).

If territorial control estimates are valid, deforestation should be more likely in areas that saw larger changes in 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 indicating whether a grid cell i experienced deforestation in year t. $\Delta Control$ denotes the change in the average annual level of control between year 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 that experienced changes in control as a result of the agreement. The model is estimated via logistic regression. Data on deforestation from 2013 to 2016 are obtained from satellite images via the forest monitoring system from the *Instituto de Hidrología, Meteorología y Estudios Ambientales.*

Table V supports the hypothesis that changes in territorial control as a result of the peace agreement are associated with a higher probability of deforestation. The coefficient on the interaction 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 from *DG* to *G*, 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

[20] `https://www.theguardian.com/world/2017/jul/11/colombia-deforestation-farc`.

	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.

results in line with observed empirical relationships between changes in territorial control and deforestation.[21]

Nigeria

Figure 5 plots territorial control estimates for Boko Haram and the government for 15 North-east Nigerian states 2009–2017. In late 2014 and early 2015, reports suggest that the insurgents controlled 15 localities in the border region with Cameroon, and had partial control over additional 15 local government areas.[22] In early to mid 2015, Nigerian government and African Union troops launched attacks against Boko Haram that caused the insurgents to loose a majority of their territory.[23] However, reports from 2018 cast doubt over the government's claim that it drove the insurgents out of the region, suggesting that Boko Haram controlled parts of Borno state via roadblocks, stop and search operations, and the collection

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

[22] See `https://www.amnesty.org/en/latest/news/2015/01/boko-haram-glance/`.

[23] `http://web.stanford.edu/group/mappingmilitants/cgi-bin/groups/view/553?highlight=boko%2Bharam`.

of "taxes" for protection.[24]

The estimates illustrate the onset of the Boko Haram insurgency in 2009.[25] Starting in 2011, Boko Haram is estimated to gain control in parts of Borno. The insurgents subsequently establish strongholds that are at first scattered throughout the region. By 2014, the estimates show the 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 a few areas in the border region with Cameroon.

The inclusion of Nigeria in the Armed Conflict Location & Event Data Project (ACLED, 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).[26] The 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.[27]

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.[28] Cells experiencing zero events are assumed to be under full

[24] `https://www.dw.com/en/boko-haram-islamists-still-control-parts-of-northeastern-nigeria/a-43851013`.

[25] The model is initiated in 2008 with a strong prior of full government control in the first month.

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

[27] I discuss a more conservative coding in online appendix F.

[28] 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.

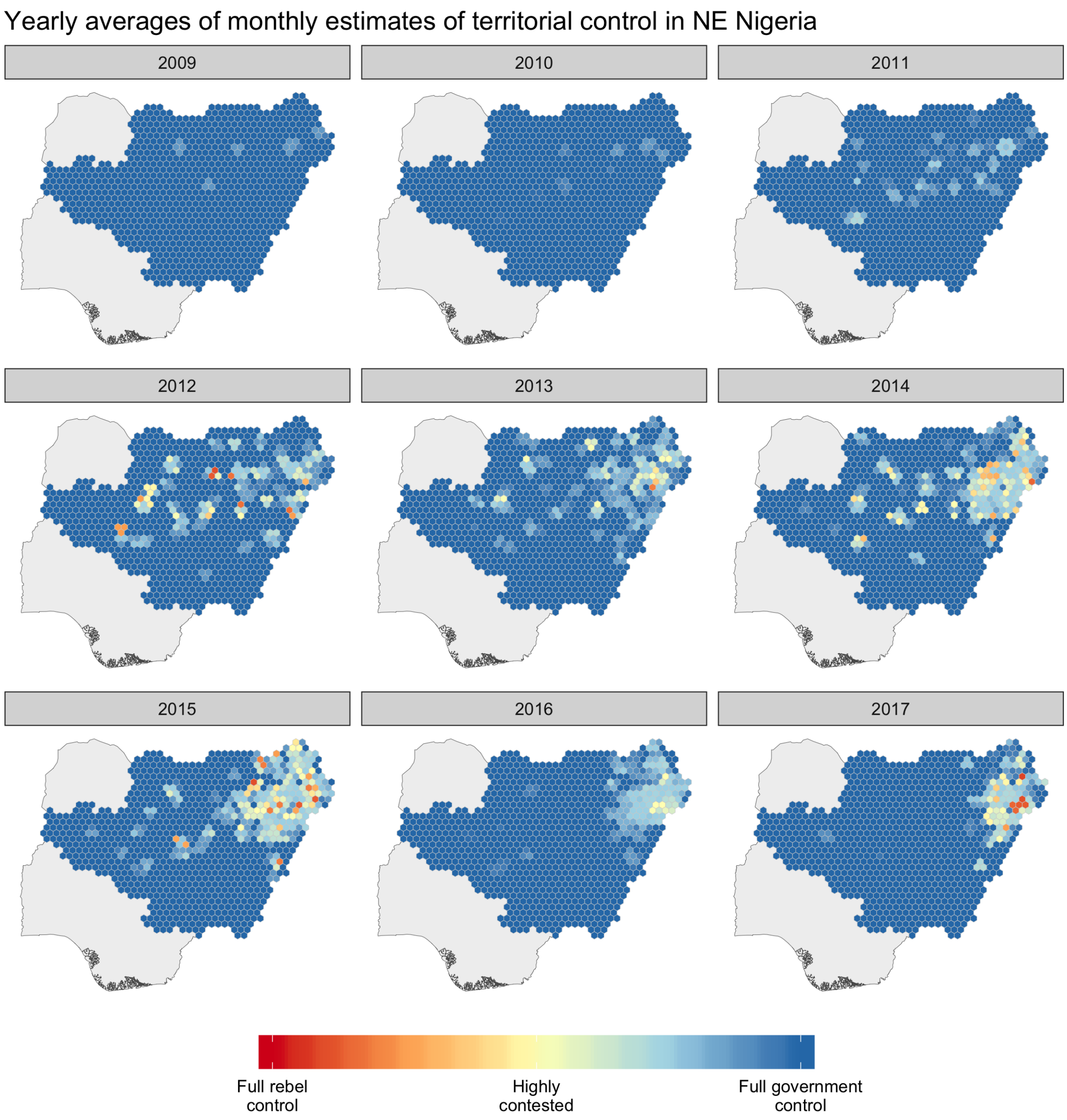

Figure 5: Territorial control estimates in Northeast Nigeria (N = 943).

government control.

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

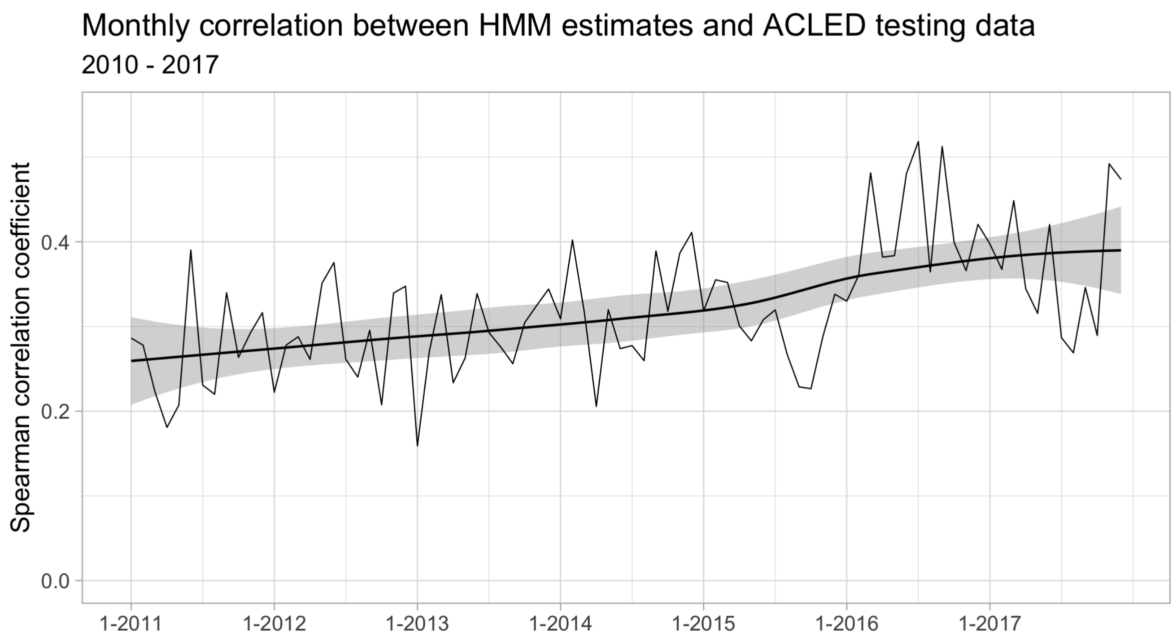

Figure 6: Correlation between HMM estimates and ACLED validation data, including loess smoothed average.

Figure 6 plots the correlation between territorial control estimates and the ACLED validation data. The correlation ranges between 0.14 and 0.48 and is, 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 incorporate the variation of monthly estimates as uncertainty into future models of territorial control. Compared with the validation data, the HMM estimates appear to slightly overstate levels of rebel control and understate levels of contestation.[29] However, the annual averages over monthly HMM estimates in Figure 5 uncover general spatial patterns and major temporal trends in the distribution of territorial control. In particular, the model captures the reduction of insurgent territorial from 2015 to 2016.

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 territorial control estimates for the fight between the FARC and the Colombian government, and the Boko Haram insurgency. A 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 the generation of territorial control estimates that utilize publicly available conflict 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 subnational analyses of civil war for both 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 dynamics. Rebel groups like Boko Haram operate across international frontiers and often leverage the remoteness of border regions to their advantage — establishing strongholds beyond the reach of their primary enemy's armed forces.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 and rendering the model

[29]See online appendix F for more details.

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 and offers 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.[30]

A "ground truth" of territorial control information does not exist for most conflicts, in particular for asymmetric civil wars. This presents a limitation and an opportunity. 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 regarding the evolution of territorial control illustrates how domain-specific knowledge can be leveraged to inform model parameters in machine learning applications. 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 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, for example in former ceasefire areas in Burma/Myanmar. As more data becomes available, future work should seek to incorporate such information to improve the accuracy of the estimates and facilitate more empirical research on territorial control in civil war.

[30] See online appendix G for more detail.

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.

Croicu, M. and R. Sundberg (2017). UCDP GED codebook version 17.1. 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.

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.

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`.

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].

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.

Park, J. 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, 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.

Prem, M., S. Saavedra, and J. F. Vargas (2019). End-of-Conflict Deforestation: Evidence From Colombia's Peace Agreement. Working Paper 288, Households in Conflict Network (HiCN), Brighton, UK.

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.

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

Rubin, M. A. (2019). Rebel territorial control and civilian collective action in civil wars: Evidence from the communist insurgency in the Philippines. *Journal of Conflict Resolution*.

Sauter, M. 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, P. A. (2006). Forecasting conflict in the Balkans using Hidden Markov Models. In R. Trappl (Ed.), *Programming for Peace: Computer-Aided Methods for International Conflict Resolution and Prevention*, pp. 161–184. Dordrecht: Springer Netherlands.

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

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.

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.

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