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

Therese Anders

Ph.D. Candidate, University of Southern California

27 April 2019

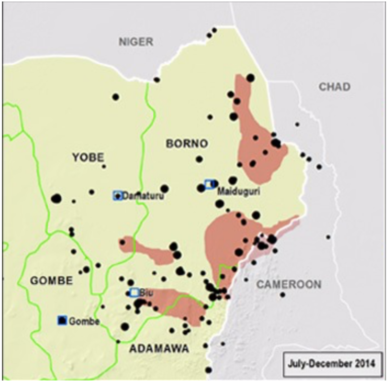
New Faces in Political Methodology XI, Penn State University

Symmetric conflict - Ukraine

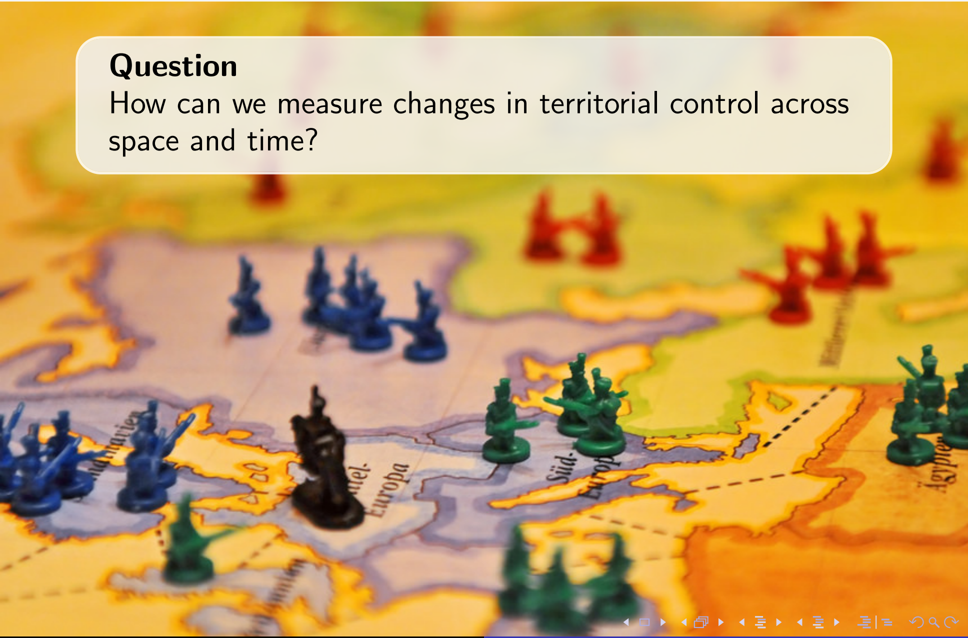
Motivation

Symmetric conflict - Ukraine

Asymmetric conflict - Nigeria



Overview



Question

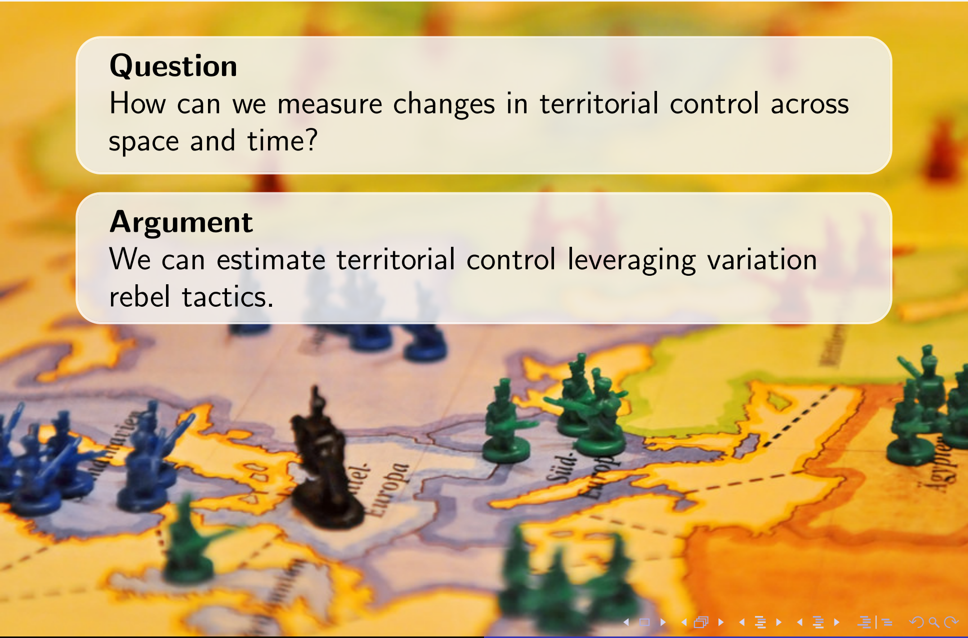
How can we measure changes in territorial control across space and time?

Question

How can we measure changes in territorial control across space and time?

Argument

We can estimate territorial control leveraging variation rebel tactics.



Question
How can we measure changes in territorial control across space and time?

Argument
We can estimate territorial control leveraging variation rebel tactics.

Contribution

1. Measurement model for rebel tactics.
2. Hidden Markov Model (HMM) estimates of territorial control.

Roadmap

1. Model of territorial control & rebel tactics

Roadmap

1. Model of territorial control & rebel tactics
2. Measuring conflict exposure

Roadmap

1. Model of territorial control & rebel tactics
2. Measuring conflict exposure
3. Hidden Markov Model (HMM)

Roadmap

1. Model of territorial control & rebel tactics
2. Measuring conflict exposure
3. Hidden Markov Model (HMM)
4. HMM estimates & validation
    - Nigeria
    - Colombia

Theoretical background

Territorial control influences tactical choices in civil wars.
[Carter 2015, de la Calle and Sánchez-Cuenca 2015, 2012]

Rebel tactical choice

Higher rebel control

$\Longrightarrow$ Conventional tactics

Rebel tactical choice

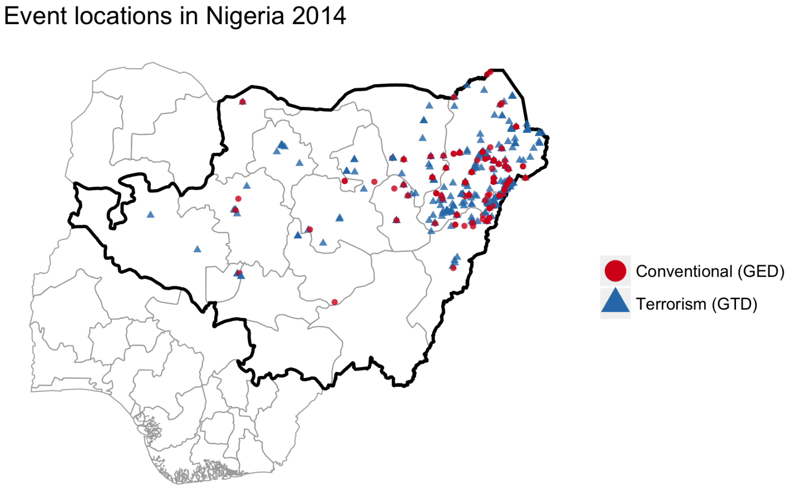
Higher rebel control

$\Longrightarrow$ Conventional tactics

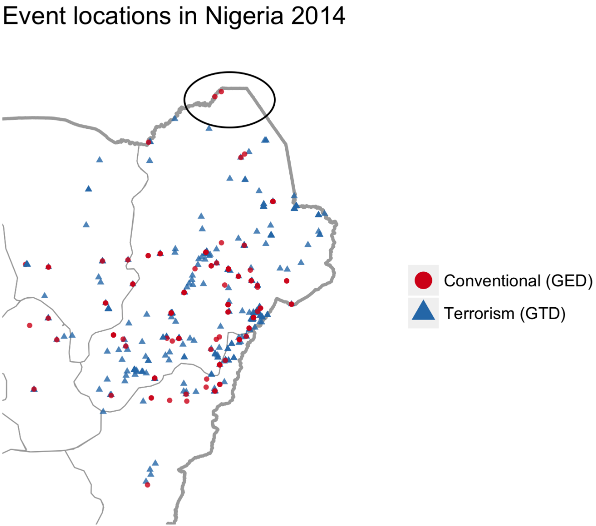
Lower rebel control

$\Longrightarrow$ Terrorist attacks

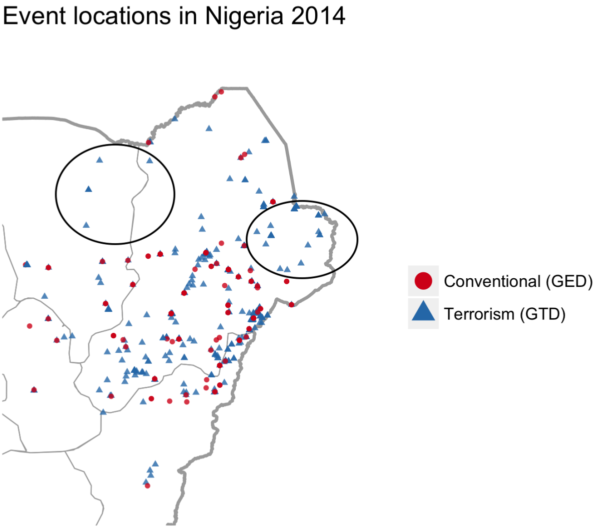
Co-occurrence of conventional and terrorist tactics



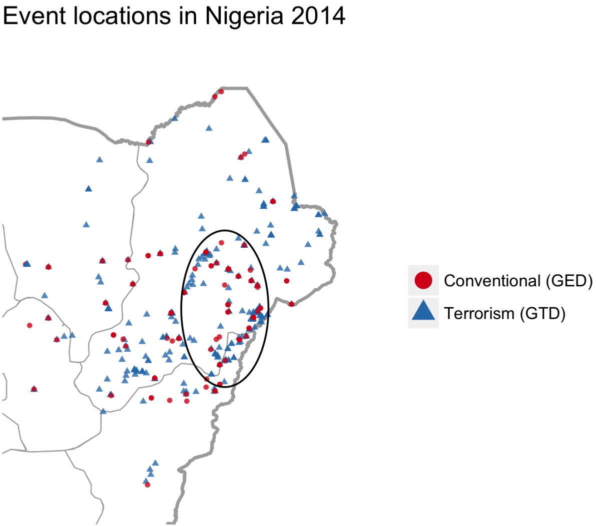
Co-occurrence of conventional and terrorist tactics



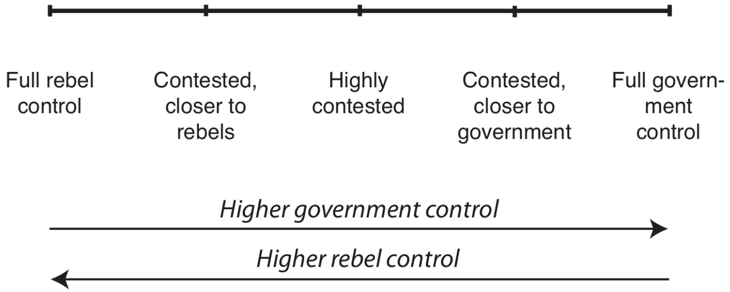
VCo-occurrence of conventional and terrorist tactics



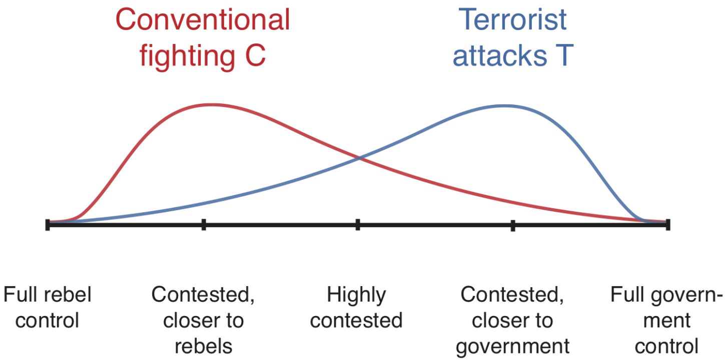
Co-occurrence of conventional and terrorist tactics



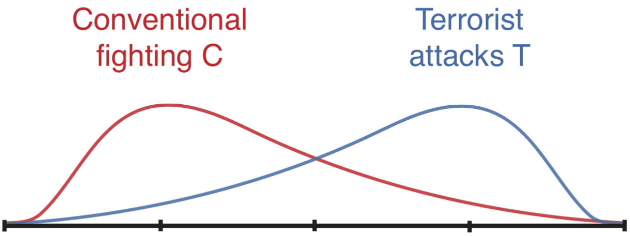
Continuum of territorial control



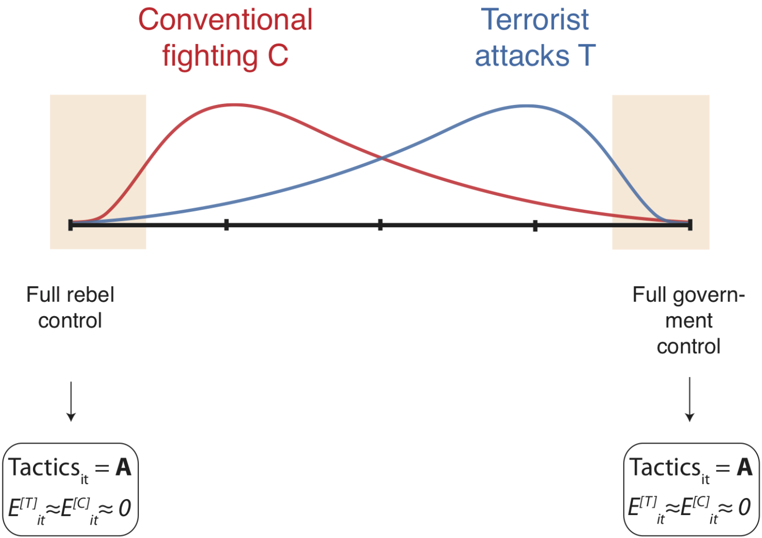
Continuum of territorial control and tactical choice



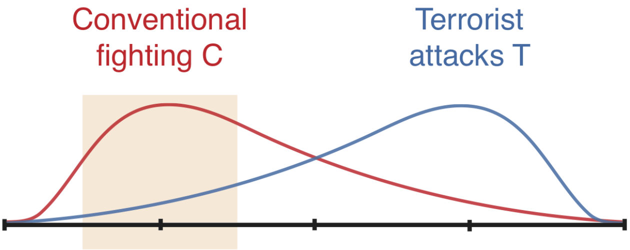
Continuum of territorial control and tactical choice



Continuum of territorial control and tactical choice



Continuum of territorial control and tactical choice

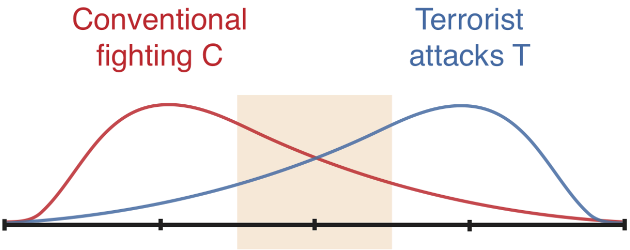


Full rebel control
↓
Tactics$_{it}$ = **A**
$E^{[T]}_{it} \approx E^{[C]}_{it} \approx 0$

Contested, closer to rebels
↓
Tactics$_{it}$ = **D**
$C_{it} > T_{it}$

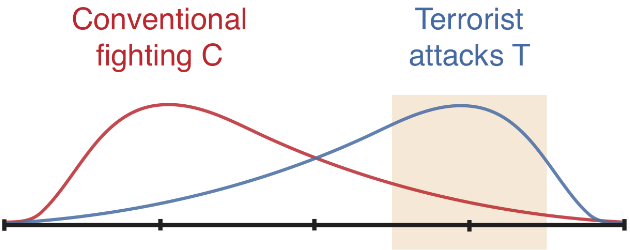
Full government control
↓
Tactics$_{it}$ = **A**
$E^{[T]}_{it} \approx E^{[C]}_{it} \approx 0$

Continuum of territorial control and tactical choice



Full rebel control	Contested, closer to rebels	Highly contested		Full government control
↓	↓	↓		↓
$\text{Tactics}_{it} = \mathbf{A}$	$\text{Tactics}_{it} = \mathbf{D}$	$\text{Tactics}_{it} = \mathbf{B}$		$\text{Tactics}_{it} = \mathbf{A}$
$E^{[T]}_{it} \approx E^{[C]}_{it} \approx 0$	$C_{it} > T_{it}$	$C_{it} \approx T_{it} \neq 0$		$E^{[T]}_{it} \approx E^{[C]}_{it} \approx 0$

Continuum of territorial control and tactical choice



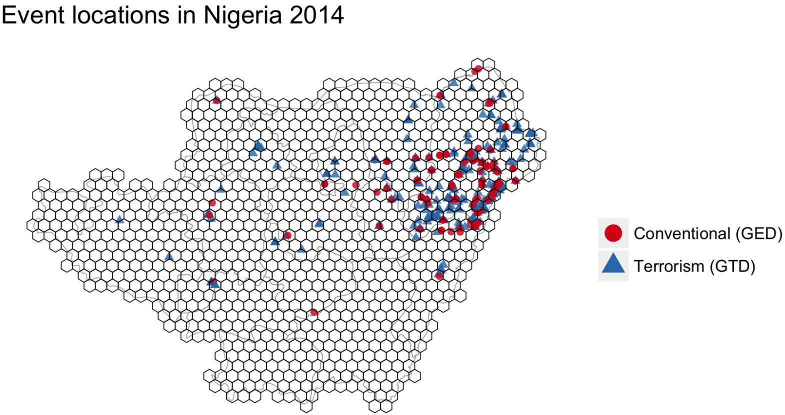
Full rebel control	Contested, closer to rebels	Highly contested	Contested, closer to government	Full government control
↓	↓	↓	↓	↓
$\text{Tactics}_{it} = \mathbf{A}$	$\text{Tactics}_{it} = \mathbf{D}$	$\text{Tactics}_{it} = \mathbf{B}$	$\text{Tactics}_{it} = \mathbf{C}$	$\text{Tactics}_{it} = \mathbf{A}$
$E^{[T]}_{it} \approx E^{[C]}_{it} \approx 0$	$C_{it} > T_{it}$	$C_{it} \approx T_{it} \neq 0$	$C_{it} < T_{it}$	$E^{[T]}_{it} \approx E^{[C]}_{it} \approx 0$

Measuring variation in rebel tactics

$$\begin{aligned} Tactics_{it} &= \{A, B, C, D\} \\ &= f\left(E_{it}^{[T]}, E_{it}^{[C]}\right) \end{aligned}$$

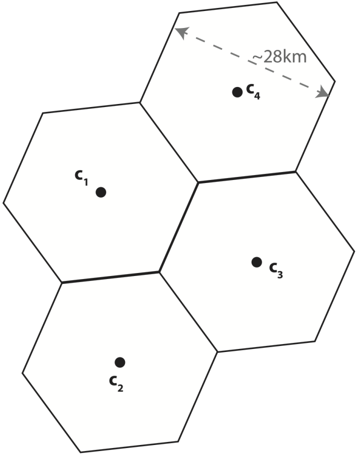
$E_{it}^{[T]}$ area i's exposure to terrorist events at t
$E_{it}^{[C]}$ area i's exposure to conventional events at t

Measuring exposure conflict events



Measuring exposure E to conflict events K

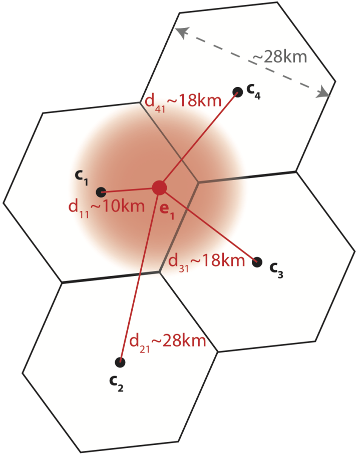
Unit of observation: cell-month C_{it}



Measuring exposure E to conflict events K

Unit of observation: cell-month C_{it}

a) Spatial distance d_{ij} in km, temporal distance (age) a_{jt} in months.



Measuring exposure E to conflict events K

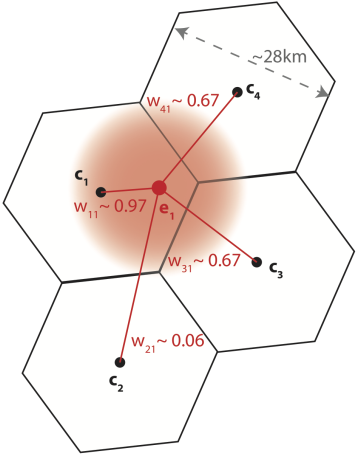
Unit of observation: cell-month C_{it}

a) Spatial distance d_{ij} in km, temporal distance (age) a_{jt} in months.

b) Logistic weights $w_{d_{ij}}$ and $w_{a_{jt}}$

$$w_{d_{ij}} = \frac{1}{1 + e^{-7+0.35 d_{ij}}}$$

$$w_{a_{jt}} = \frac{1}{1 + e^{-8+2.5 a_{jt}}}$$



Measuring exposure E to conflict events K

Unit of observation: cell-month C_{it}

a) Spatial distance d_{ij} in km, temporal distance (age) a_{jt} in months.

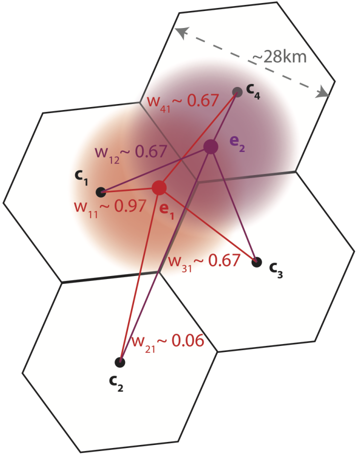
b) Logistic weights $w_{d_{ij}}$ and $w_{a_{jt}}$

$$w_{d_{ij}} = \frac{1}{1 + e^{-7+0.35d_{ij}}}$$

$$w_{a_{jt}} = \frac{1}{1 + e^{-8+2.5a_{jt}}}$$

c) Exposure E

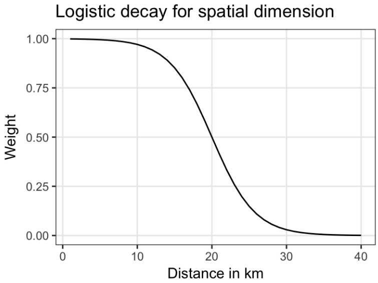
$$E_{it} = \sum_{j=1}^{J} \left(w_{d_{ij}} \times w_{a_{tm}} \right)$$



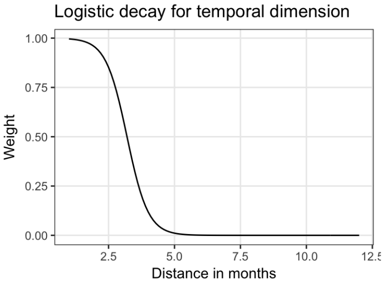
Logistic decay functions

$$f(d) = \frac{1}{1 + e^{-\kappa + \lambda d}}$$

Space: $\kappa = 7, \lambda = 0.35$

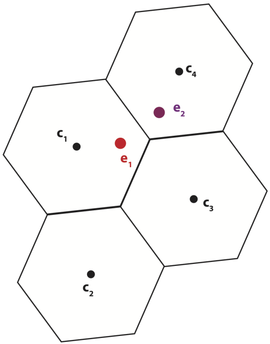


Time: $\kappa = 8, \lambda = 2.5$

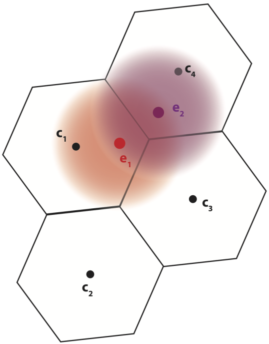


Advantage of using decay functions

Discrete assignment

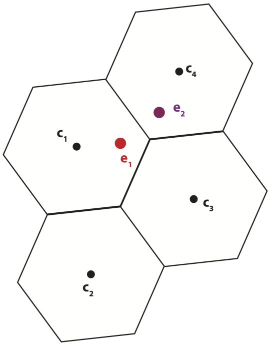


Continuous assignment

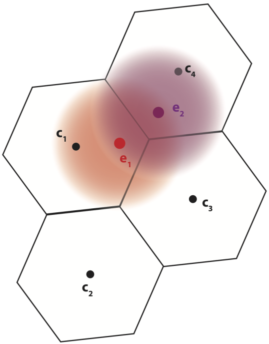


Advantage of using decay functions

Discrete assignment



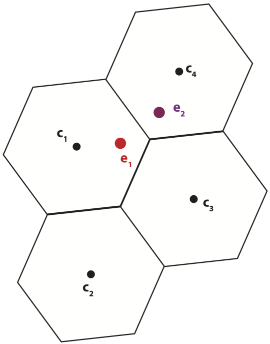
Continuous assignment



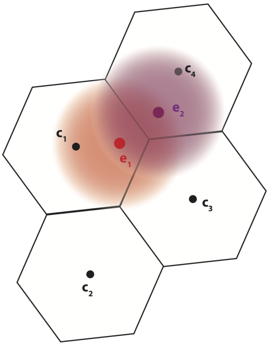
- Reduce MAUP

Advantage of using decay functions

Discrete assignment

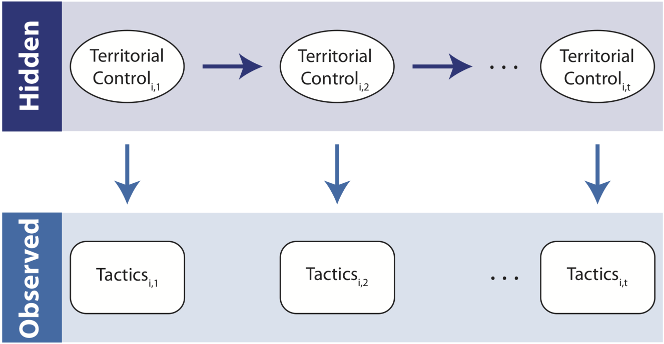


Continuous assignment

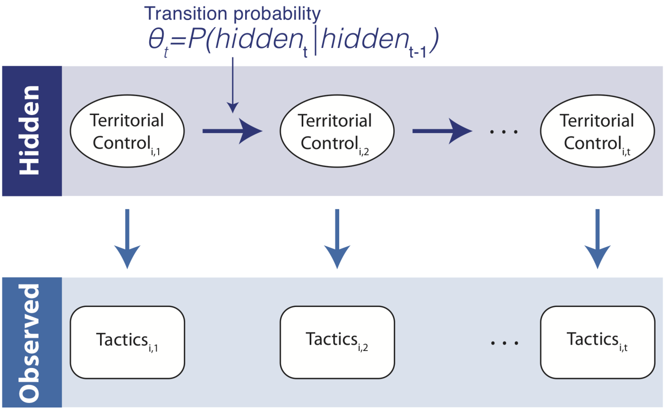


- Reduce MAUP
- Spatial dependency in HMM

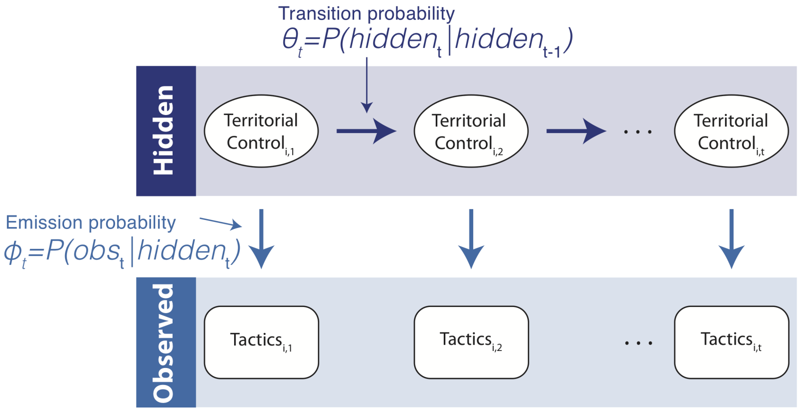
Hidden Markov Model



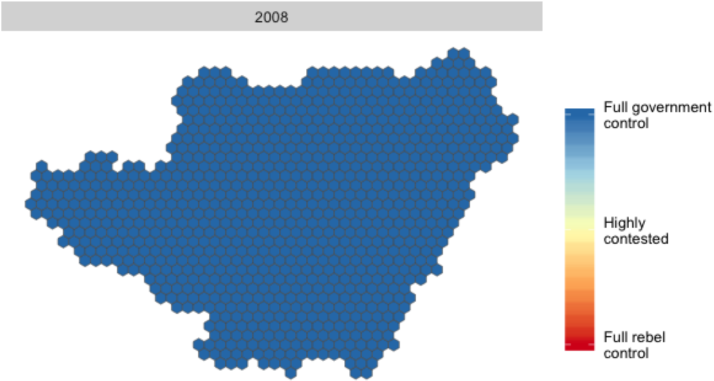
Hidden Markov Model



Hidden Markov Model

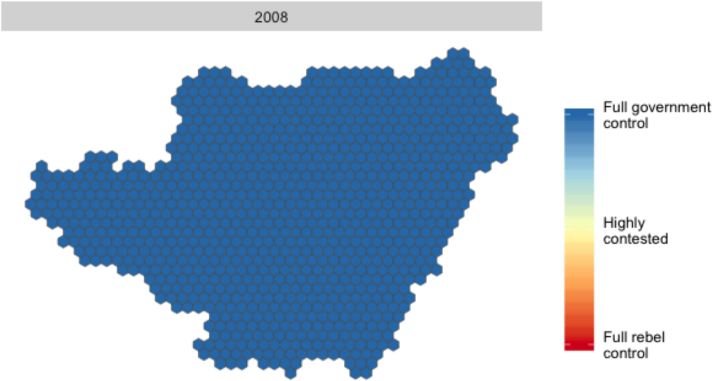


HMM estimates for Nigeria



[Discrete monthly-level estimates averaged for each year, N = 942]

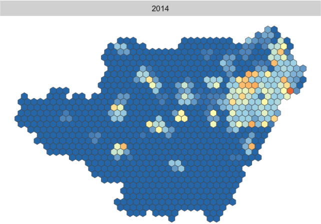
HMM estimates for Nigeria



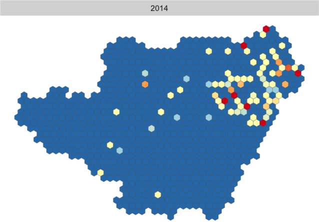
[Discrete monthly-level estimates averaged for each year, N = 942]

Validation

HMM estimates

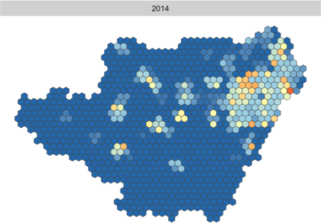


ACLED validation data

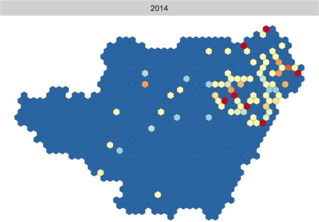


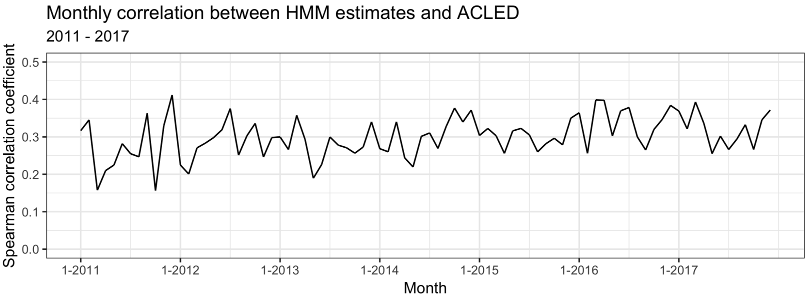
Validation

HMM estimates

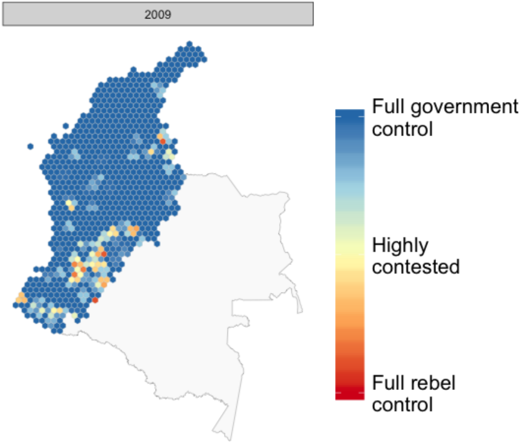


ACLED validation data



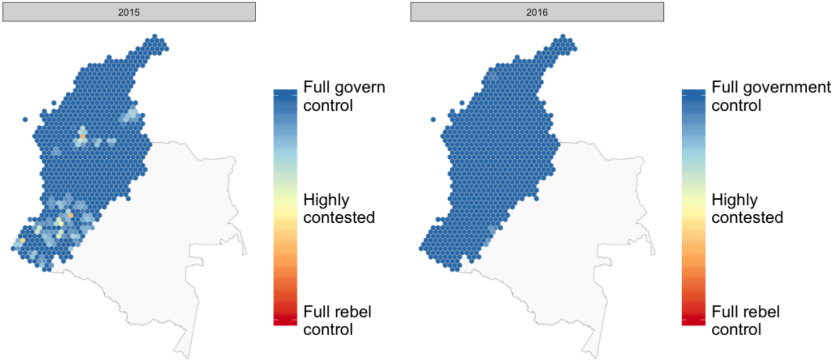


HMM estimates for Colombia



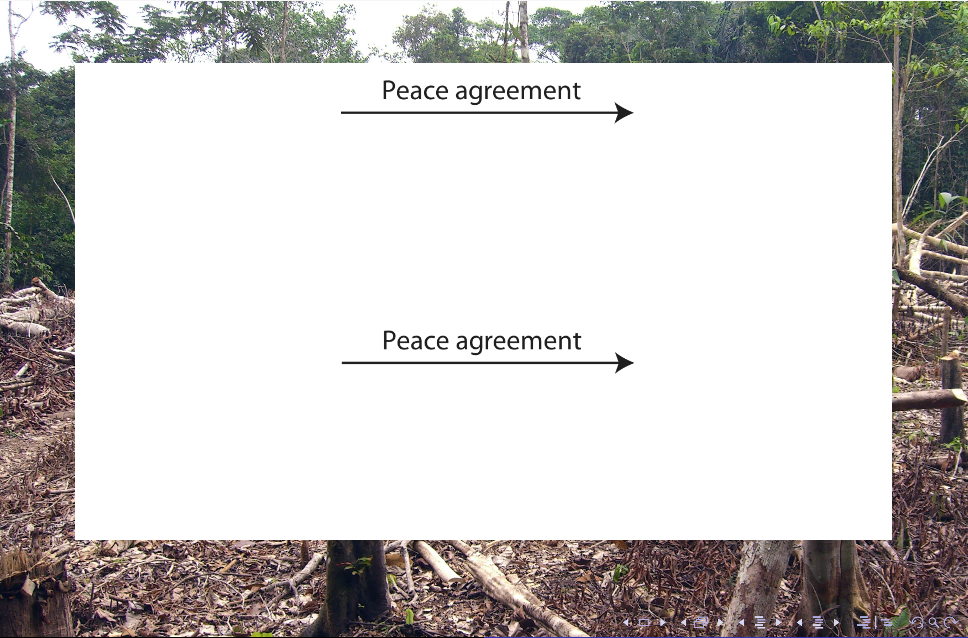
[Discrete monthly-level estimates averaged for each year, N = 851]

HMM estimates for Colombia 2015 vs. 2015

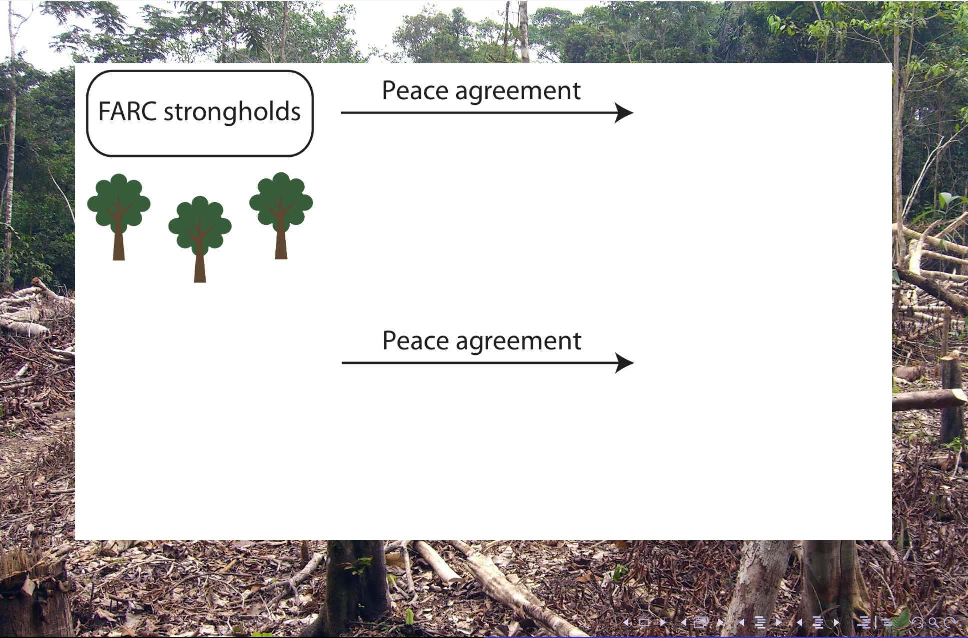


Validation: Deforestation and territorial control

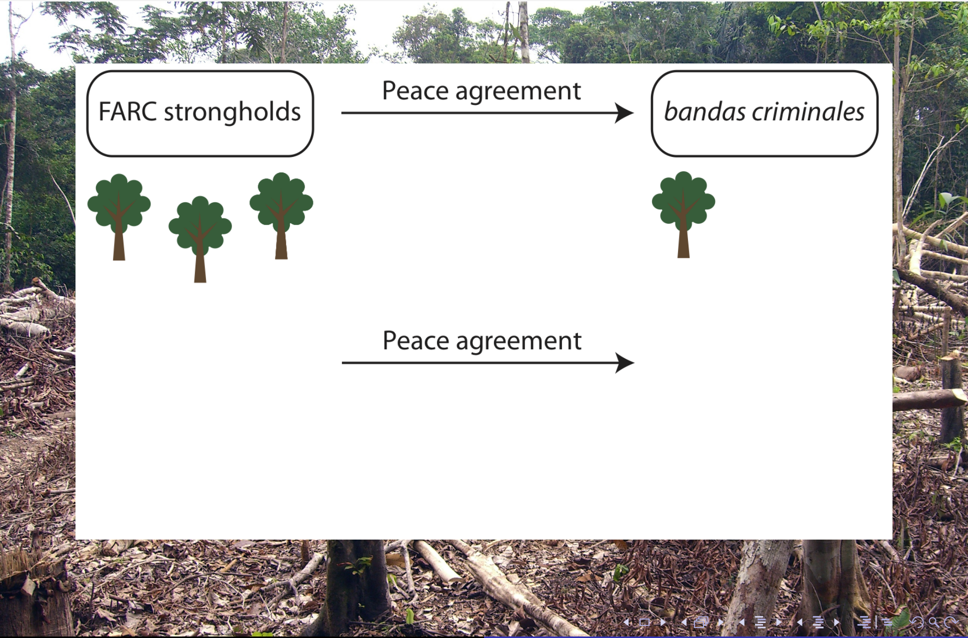
Validation: Deforestation and territorial control



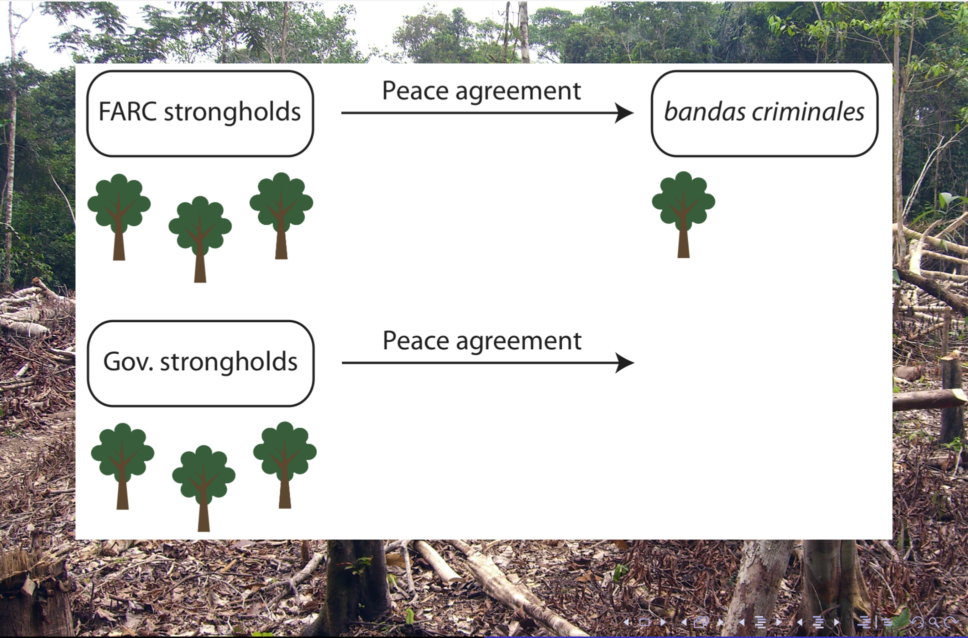
Validation: Deforestation and territorial control



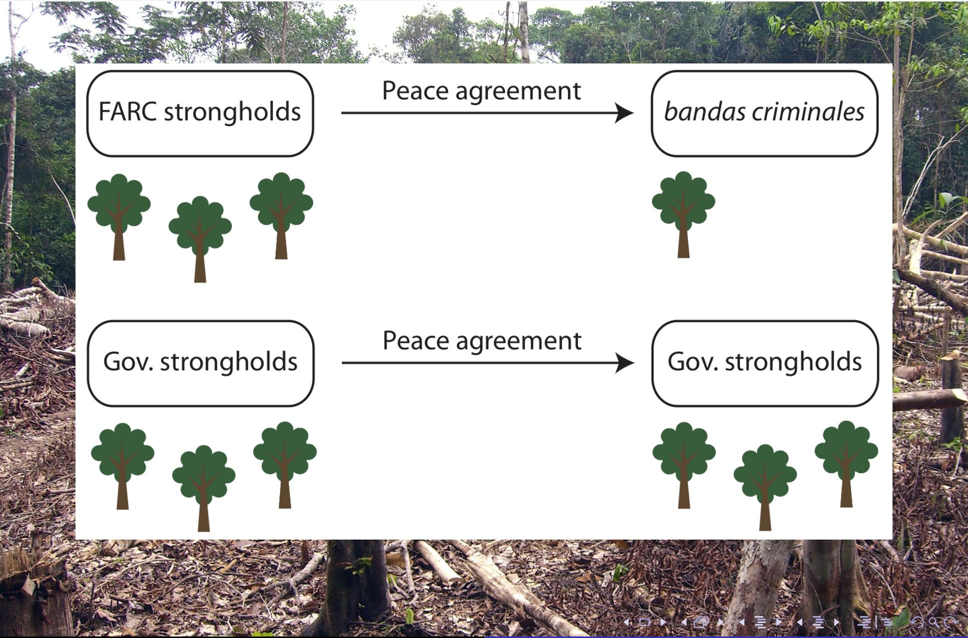
Validation: Deforestation and territorial control



Validation: Deforestation and territorial control



Validation: Deforestation and territorial control



Validation: Deforestation and territorial control

Hypothesis

Rebel-controlled areas have a higher probability of deforestation post-peace than government-controlled areas.

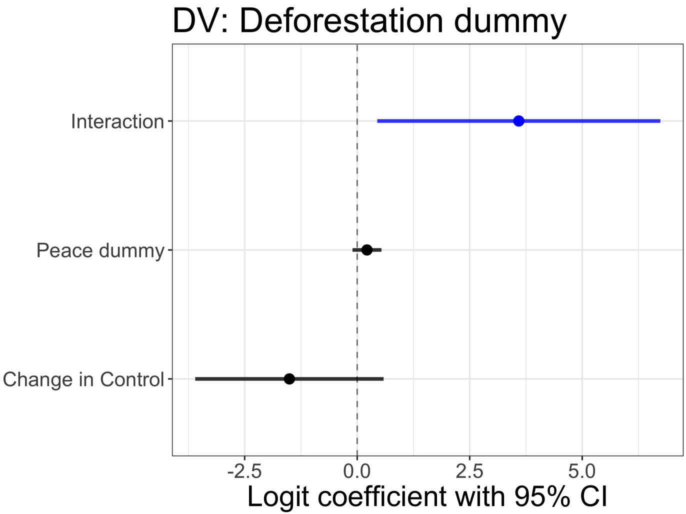
Validation: Deforestation and territorial control

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

$Deforest_{i,t}$	Deforestation in grid cell (0/1, from IDEAM)
$Control_{i,t}$	HMM Territorial control estimate (1 = gov. control)
$Peace_t$	Peace agreement 2016 (0/1)

N $= 3{,}404$ ($I = 851, 2013 - 2016$)

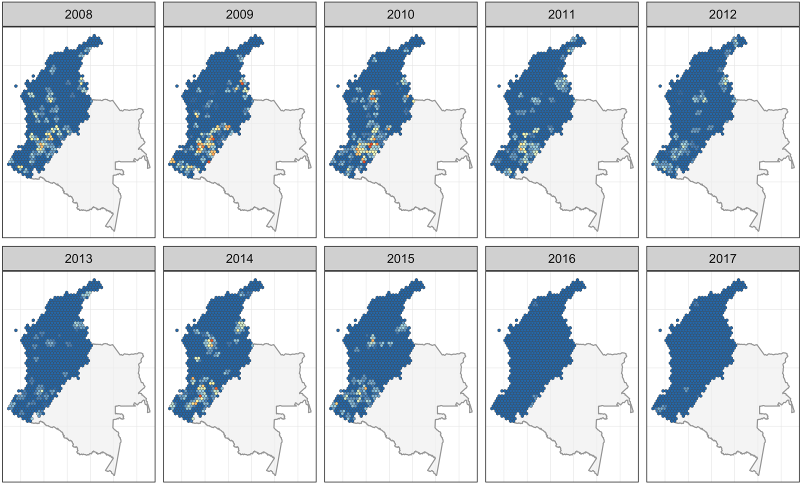
Regression results



[Bootstrapped standard errors clustered by grid cell.]

Discussion

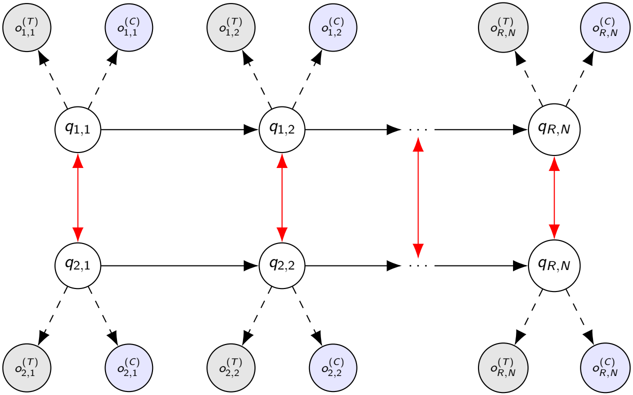
HMMs are promising method to estimate territorial control.



tanders@usc.edu

Appendix

Hidden Markov Random Field (HMRF) model



Picture sources (in the order they appear)

1. Risk board game https://www.flickr.com/photos/tambako/4598642399, licensed under CC BY-ND 2.0, accessed 16 July 2016.
2. Nigeria map https://reliefweb.int/sites/reliefweb.int/files/resources/Nigeria_Boko_Haram_threat_dec%202015.pdf.
3. Ukraine http://www.kas.de/wf/en/33.48639/.
4. Deforestation https://www.flickr.com/photos/16725630@N00/1524189000/.

Decoding via Viterbi algorithm

Most probable path for hidden state q_h at t:

$$v_t(h) = \max_{g=1}^{N} v_{t-1}(g)\theta_{gh}\phi_h(o_t)$$

[Maximum likelihood sequence of labels via argmax]

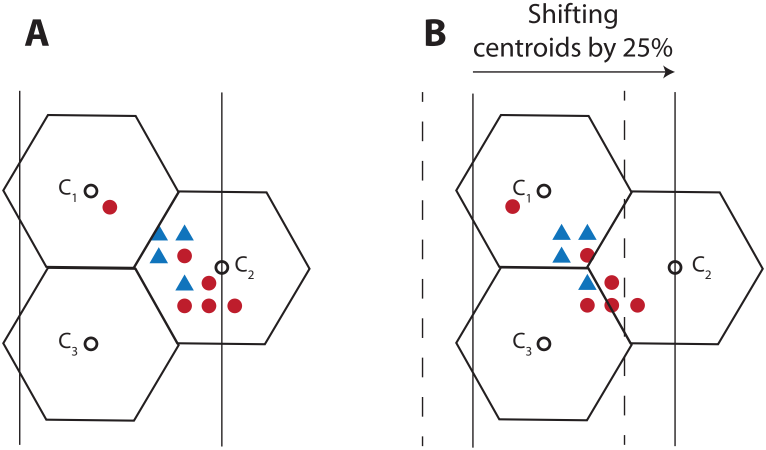
h . . . indexes current state
g . . . indexed previous state
$v_{t-1}(g)$ . . . path probability of previous time step
θ_{gh} . . . transition probability from q_g to q_h
$\phi_h(o_t)$ . . . emission probability given h HMM returns path labels for maximum likelihood of observation sequence.

Inference

Observation	Latent variable
$E_{it}^{[T]} \approx E_{it}^{[C]} \approx 0$	Full rebel control
$C_{it} > T_{it}$, and $\|C_{it} - T_{it}\| > m$	Incomplete rebel control
$\|C_{it} - T_{it}\| < m$	Highly contested
$C_{it} < T_{it}$, and $\|C_{it} - T_{it}\| > m$	Incomplete government control
$E_{it}^{[T]} \approx E_{it}^{[C]} \approx 0$	Full government control

- $m = 0.025$
- $T_{it} = P(\lfloor E_{it}^{[T]} \rfloor; \lambda_t^{[T]})$ and $C_{it} = P(\lfloor E_{it}^{[C]} \rfloor; \lambda_t^{[C]})$ are probabilities from a zero-inflated Poisson distribution.
- $\lambda_t^{[T]}$ and $\lambda_t^{[C]}$ denote expected number of events for each tactic in a given time period t across all areas I within a country.

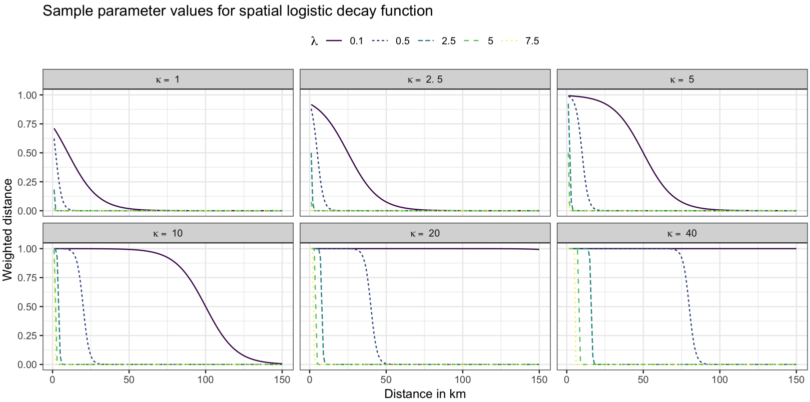
Resampling grid cell locations



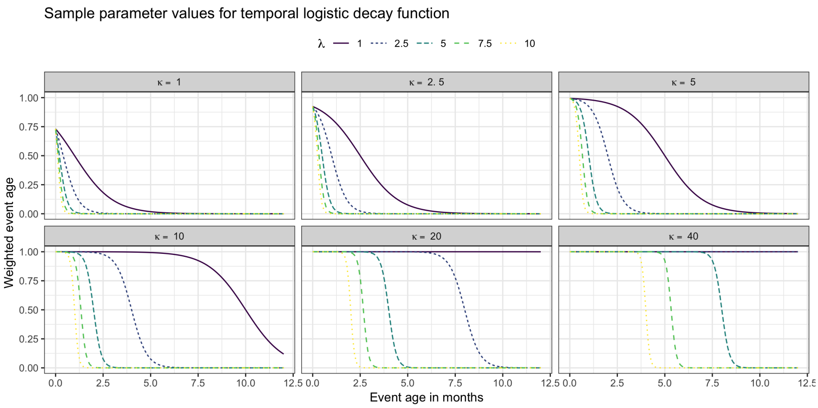
C_1 1T, 0C: $\text{Tactics}_{it} = D$
C_2 5T, 4C: $\text{Tactics}_{it} = D$
C_3 0T, 0C: $\text{Tactics}_{it} = A$

C_1 2T, 3C: $\text{Tactics}_{it} = B$
C_2 3T, 0C: $\text{Tactics}_{it} = D$
C_3 1T, 1C: $\text{Tactics}_{it} = C$

Spatial decay functions



Temporal decay functions



Estimation procedure

For each grid cell

1. Compute terrorist exposure $E_{it}^{[T]}$ and conventional exposure $E_{it}^{[C]}$ to events K_j.
   1. Spatial distance $\longrightarrow$ spatial weight $w_{d_{ij}}$.
   2. Temporal distance $\longrightarrow$ temporal weight $w_{a_{jt}}$.
   3. $E_{it} = \sum_{j=1}^{J} \left(w_{d_{ij}} \times w_{a_{jt}} \right)$
2. Compute $Tactics_{it} = f\left(E_{it}^{T}, E_{it}^{C}\right)$, $Tactics_{it} = \{A, B, C, D\}$.
3. Estimate $Q = \{S1, S2, S3, S4, S5\}$ via HMM.

Transition matrix

Transition probabilities are inspired by Kalyvas (2006).

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

Emission matrix

The probability value in each cell of this matrix answers the following question: "Given that the true state of the grid cell at time t is, for example, $S1$, what is the probability of observing, for example, A from the data?"

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

Three fundamental problems of HMMs

1. **Likelihood** Given transition probabilities θ and emission probabilities ϕ, what is the likelihood of the observation sequence O?

$$P(O) = \sum_{Q} P(O,Q) = \sum_{Q} P(O|Q)P(Q)$$

2. **Decoding** Given transition probabilities θ and emission probabilities ϕ as well as the observation sequence O, what are the most likely sequence of labels for hidden states Q?
3. **Learning** Given the observation sequence O and the structure of the HMM, what are the best transition probabilities θ and emission probabilities ϕ?

See Jurafsky, Daniel & James H. Martin (2017): *Speech and Language Processing*.

Why actors seek territorial control

- Collaboration of population

  [Arjona 2016, Kalyvas 2006]
- Extract resources

  [Stewart and Liou 2017, Carter 2015]
- Increase mobilization

  [Stewart and Liou 2017, de la Calle and Sánchez-Cuenca 2015]
- Governance

  [Stewart Forthcoming]

Data sources

ACLED 8 Raleigh, Clionadh, Andrew Linke, Håvard Hegre and Joakim Karlsen. (2010). "Introducing ACLED-Armed Conflict Location and Event Data." *Journal of Peace Research 47*(5), 651–660.

GED 17.1 Sundberg, Ralph, and Erik Melander, 2013, "Introducing the UCDP Georeferenced Event Dataset." *Journal of Peace Research 50*(4), 523-532.

GTD 2016 National Consortium for the Study of Terrorism and Responses to Terrorism (START), 2016. Global Terrorism Database [Data file]. Retrieved from `https://www.start.umd.edu/gtd`.

IDEAM Instituto de Hidrología, Meteorología y Estudios Ambientales de Colombia. 2017. Cambio cobertura de bosque [Data file]. Retrieved from `http://visor.ideam.gov.co`.

Notes on data

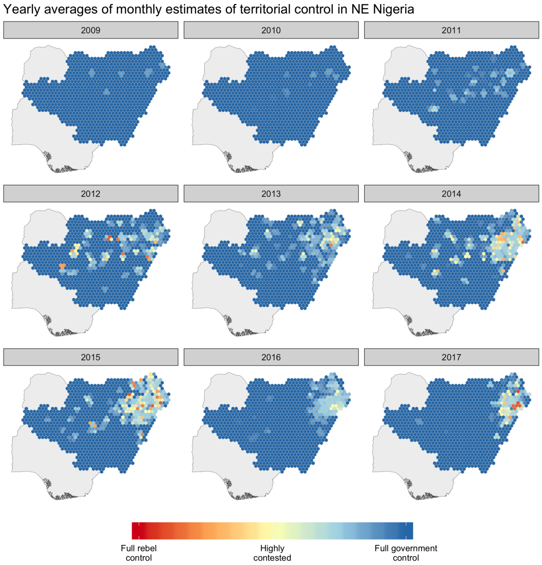
GED

- Considering only state-based conflict (disregarding violence between NSA and violence against civilians).
- Considering only events with sufficient accuracy (levels 1–3).

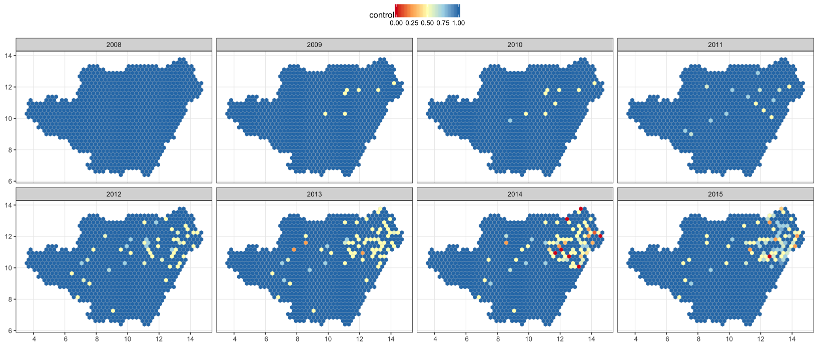
GTD

- Excluding events that are coded as `doubtterr == 1`.
- Considering only events with sufficient accuracy (levels 1–3).

Nigeria estimates



ACLED validation



Summary statistics for deforestation model

Table: The unit of analysis for territorial control is annual averages of monthly-level estimates for 0.25 degree hexagonal grid cells.

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