Online Appendix

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

1 March 2019

1 Measuring conflict exposure

A simplistic approach to coding areas' exposure to terrorism conflict events would sum the number of events that fall within a given grid cell. This procedure faces the problem that the assignment of conflict events to grid cells is highly dependent on the sampling of centroid locations—an issue recognized in the existing literature as modifiable areal unit problem (MAUP). MAUP describes the discrepancy between real world spatial patterns of events and patterns created via aggregation of events into homogenous spatial units (Openshaw and Taylor, 1979). Shifting the location of the centroids can have a severe influence on the number of events that are assigned to a particular cell. This is particularly concerning when the drawing of grid cell boundaries leads clusters of events to be broken up into smaller groups — causing the relative frequency of terrorist events and conventional war acts to change dramatically. Figure 1 illustrates this issue. Based on the location of grid cell centroids in panel A, we would code the relative frequency of rebel and conventional war fighting to correspond to the values of the variable $\text{Tactics}_{it} = [D, D, A]$. If the centroids were shifted by 25% relative to the location of the events, we would conclude the emissions of these three cells to have values of $\text{Tactics}_{it} = [B, D, C]$.

To alleviate this problem, I propose a novel measurement model for rebel tactics in civil war that uses spatial and temporal weights to associate conflict events with grid cells rather than relying on discrete assignment. The importance of individual violent events for the estimation of

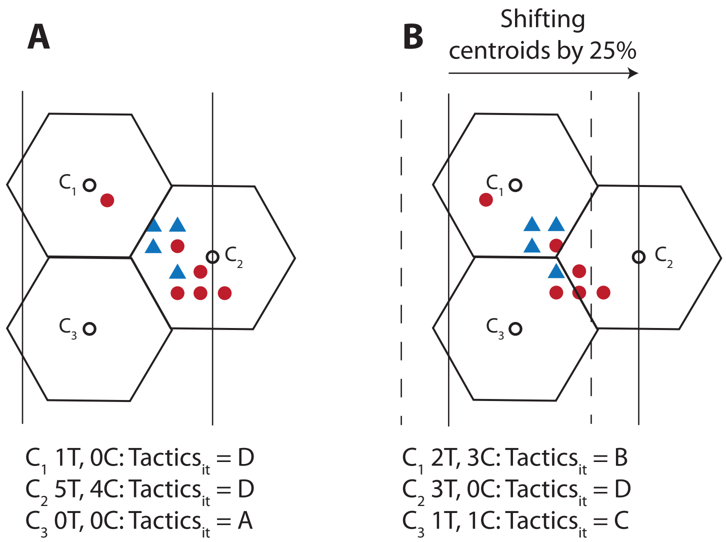


Figure 1: The schematic illustrates how shifting the location of the grid cell centroids from their original (randomly sampled) location (panel **A**) by just 25% (panel **B**) can result in vastly different conclusions about the coding of rebel tactics. Red dots indicate the location of conventional events; blue triangles those of terrorist attacks. This is a simplified example—in the analysis, Tactics_{it} is computed using probabilities from Poisson distributions under application of a margin parameter.

territorial control decreases over time and space. I model this intuition by assigning space- and time-varying weights to each event.

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

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

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

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

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

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

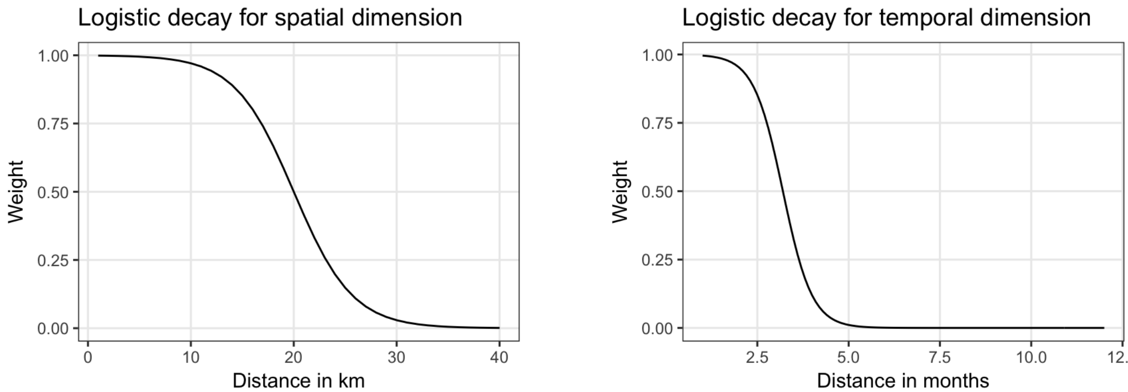
where x denotes the decaying quantity (here event age or distance between the event and a centroid), κ determines the slope of the curve and γ defines its inflection point. To describe a decay function, both the slope parameter κ and he inflection parameter γ have to be positive real numbers. To model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

Figure 2 plots the decay functions using these parameter values. Based on the shape of the logistic decay functions above, an event that occurs at the location of the centroid of a grid cell receives a spatial weight of 1. An event that occurs 10km away from the centroid receives a weight of 0.97 and an event 25km away is weighted by 0.15—after which its influence tends toward 0. The temporal weight features a different rate of decay. In the first month, the event receives a temporal weight of 1, followed by 0.95 in the second, 0.62 in the third, and 0.11 in the fourth month; after which the weight approaches zero.

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

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

Thus, the resulting unit of observation is the grid cell-month, i.e. a vector of exposure values of



(a) Spatial

(b) Temporal

Figure 2: Logistic function that describes the decay of the influence of an event in relation to a centroid in the spatial and temporal dimensions.

size $E = I \times T$.

2 Estimation procedure

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

1. Compute the exposure of the grid cell i in month t to terrorist events E_{it}^{T} and to conventional war acts E_{it}^{C} to all events J, by

   (a) computing the spatial distance d_{ij} of each event j to the centroid of grid cell i in kilometers and weighting it using a logistic decay function,

   (b) computing the temporal distance (each event's age) a^{jt} between the month m when the event occurred and the time of observation of the grid cell t in months and weighting it using a logistic decay function (note that only positive temporal distances a^{jt} are considered), and

   (c) summing the product of the spatial and temporally weighted distances for terrorist and conventional events for each grid cell-month to arrive at E_{it}^{T} and E_{it}^{C}. Note that spatially- and temporally weighted sums of under 0.05 in a grid cell-month are set to zero to avoid later grid-cell months having inflated cumulative event exposures.

2. For each grid cell i in each month t, determine the value of the variable $\mathrm{o}_{it} = f\left(E_{it}^{T}, E_{it}^{C}\right)$.

3. For each grid cell i in each month t, create a sequence of observed outputs $\mathbf{O} \in \{A, B, C, D\}$, where an individual observation o_{it} is determined by Tactics_{it}.

4. For each grid cell i compute the most probable sequence of latent states $\mathbf{Q} \in \{S1, S2, S3, S4, S5\}$ over all time periods t, given the sequence of observed indicator of rebel tactics $\mathbf{O}$ over all time periods t, the time-invariant matrix of transition probabilities Θ, and the time-invariant matrix of emission probabilities Φ via a Hidden Markov Model.

3 Summary statistics for deforestation model

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

Table 1: Summary statistics for the logistic regression model of deforestation in Colombia on changes in territorial control as a result of the 2016 peace agreement. The unit of analysis for territorial control is annual averages of monthly-level estimates for 0.25 degree hexagonal grid cells.

4 Additional figures

4.1 Transition probabilities

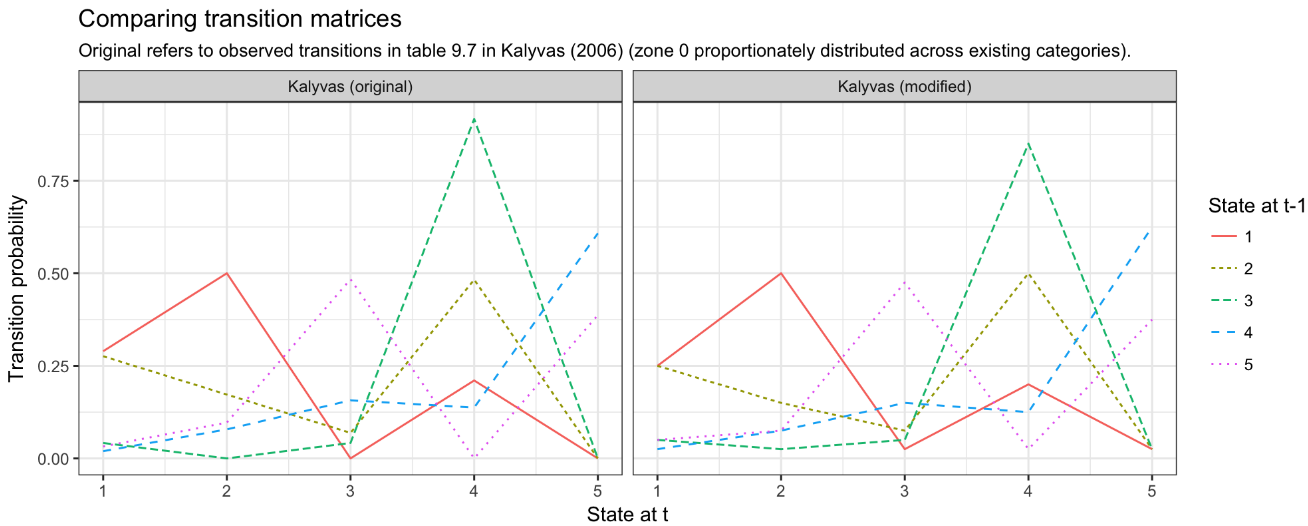


Figure 3: The figure compares the distribution of transition probabilities between the empirical observations from Kalyvas (2006) and the modified transition probabilities in this paper. The graph shows that while the transition probabilities differ slightly, the patterns of transitions between states from $t-1$ to t remain unchanged.

4.2 Case selection

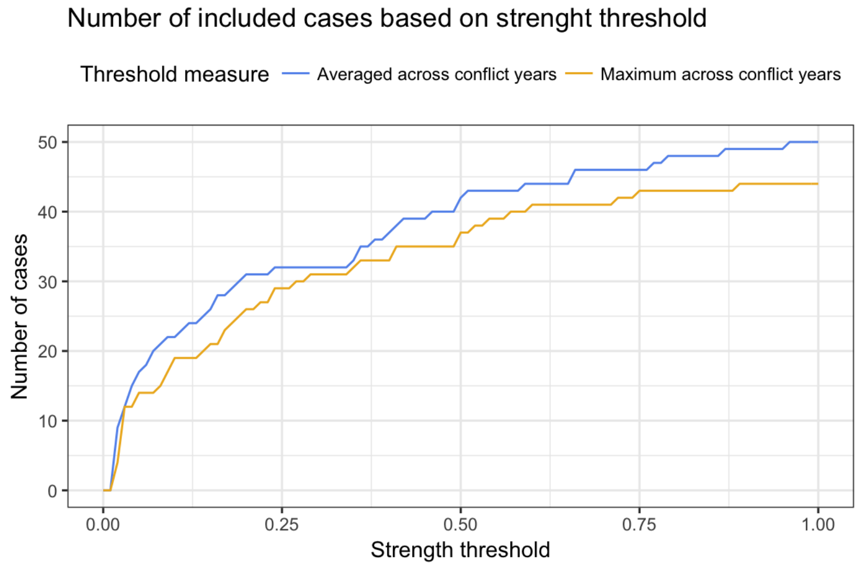


Figure 4: Number of cases that the measurement strategy can be applied to based on different thresholds of power asymmetry between the rebels and the government. Data on power asymmetry come from Polo and Gleditsch (2016).

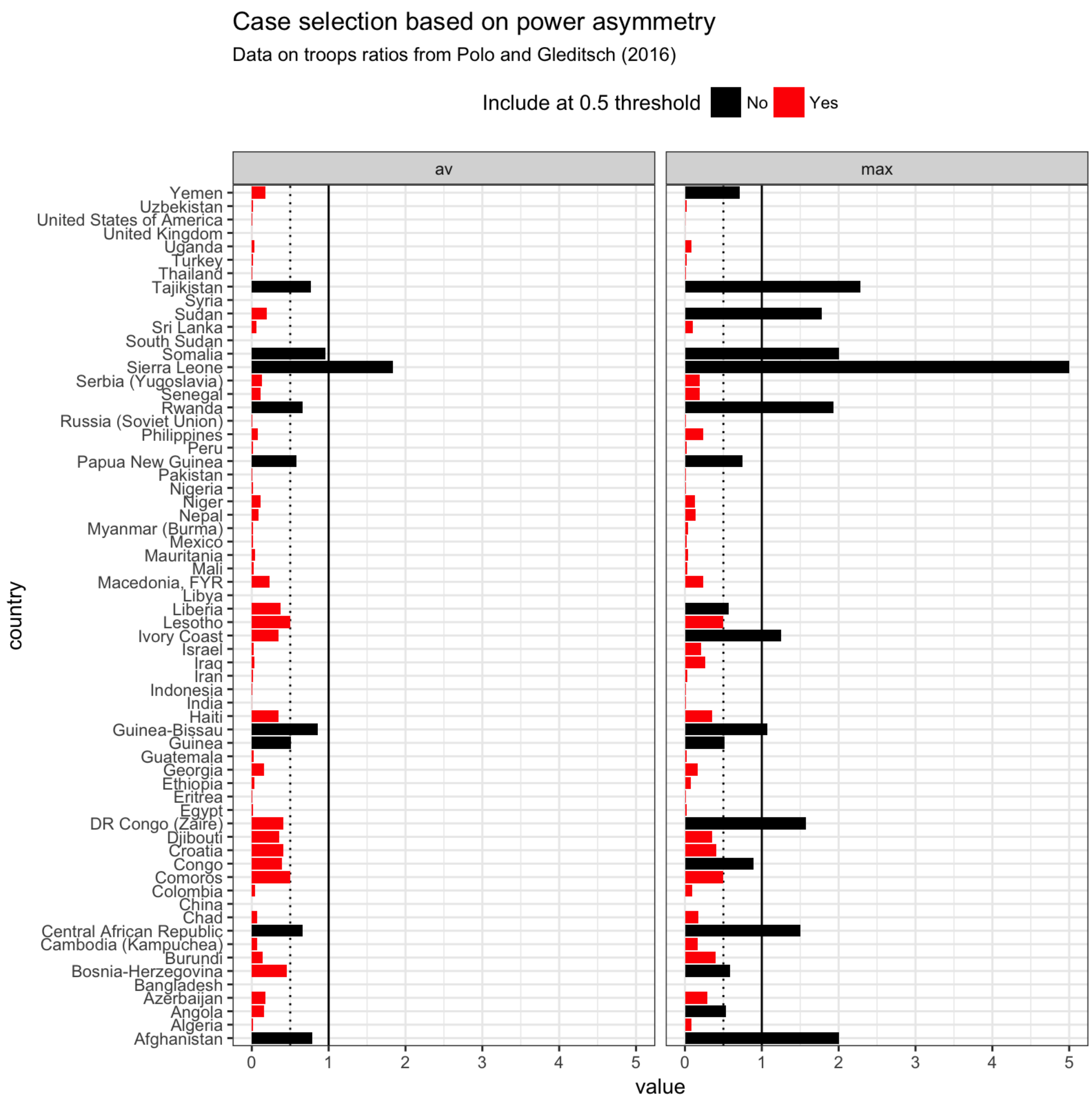
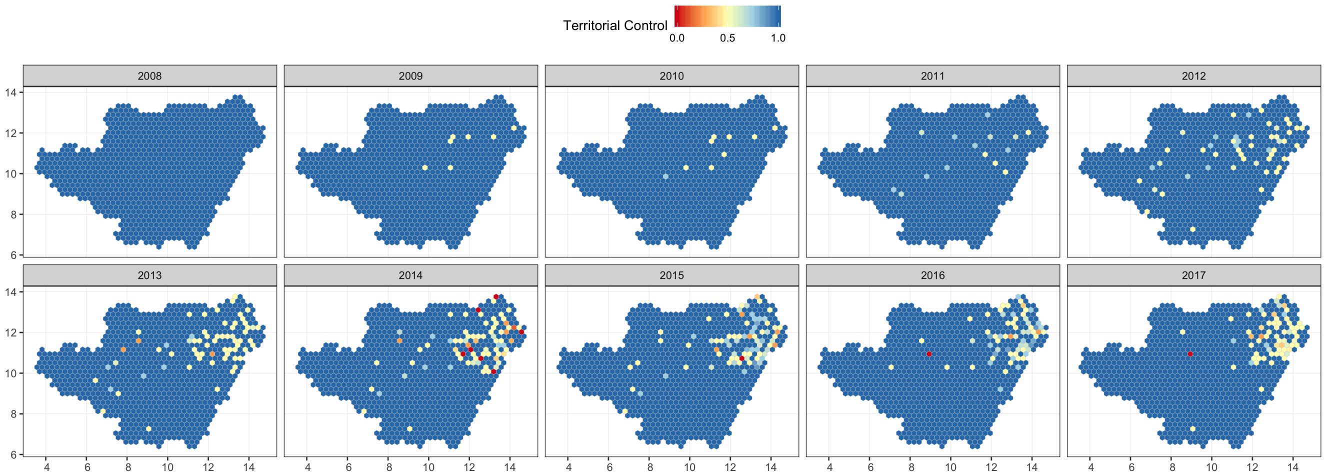
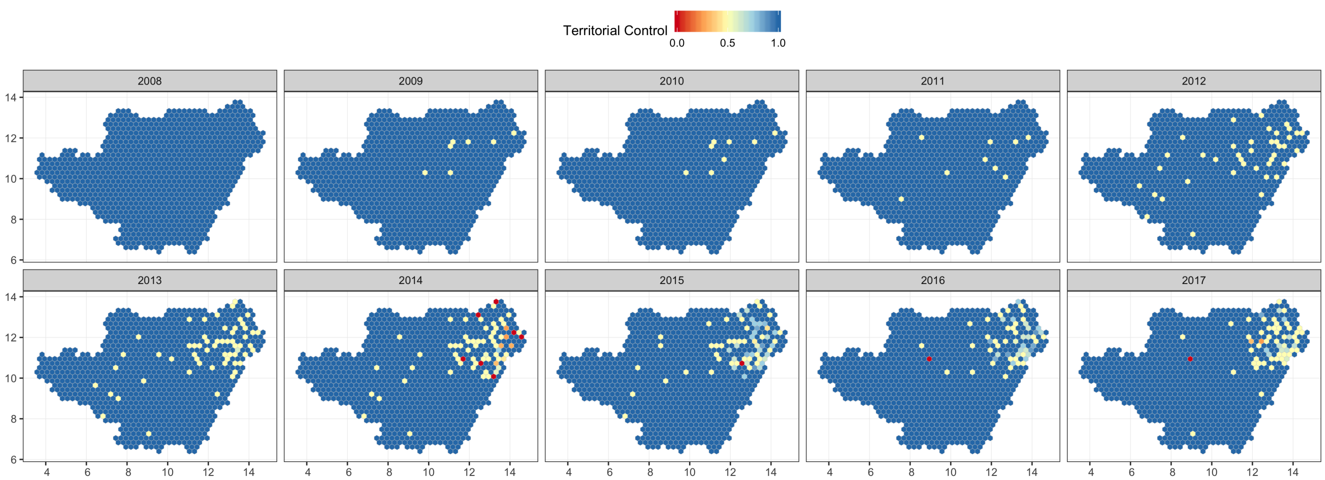


Figure 5: The graph illustrates the selection of cases for which the measurement strategy is applicable based on thresholds in average and maximum rebel-to-government troop ratios over the course of the conflict. Plotted in red are cases that would be included based on a 0.5 threshold indicating rebels that are half as strong as the government forces. Future work will investigate the determination of the most appropriate threshold.

4.3 ACLED validation data for NE Nigeria



(a) Full set of ACLED event types



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

Figure 6: Yearly averages of monthly-level ACLED validation data values. Values that are closer to 0 indicate full rebel control; values closer to 1 full government control. 0.5 indicates cells that are highly contested.

References

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

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

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