Journal of
Peace Research

Journal of Peace Research
1–11

DOI: 10.1177/0022343319836697
journals.sagepub.com/home/jpr
SAGE

Introducing xSub: A new portal for cross-national data on subnational violence

Yuri M Zhukov, Christian Davenport & Nadiya Kostyuk

Department of Political Science, University of Michigan

Abstract

Researchers today have access to an unprecedented amount of geo-referenced, disaggregated data on political conflict. Because these new data sources use disparate event typologies and units of analysis, findings are rarely comparable across studies. As a result, we are unable to answer basic questions like 'what does conflict A tell us about conflict B?' This article introduces xSub – a 'database of databases' for disaggregated research on political conflict (www.x-sub.org). xSub reduces barriers to comparative subnational research, by empowering researchers to quickly construct custom, analysis-ready datasets. xSub currently features subnational data on conflict in 156 countries, from 21 sources, including large data collections and data from individual scholars. To facilitate comparisons across countries and sources, xSub organizes these data into consistent event categories, actors, spatial units (country, province, district, grid cell, electoral constituency), and time units (year, month, week, and day). This article introduces xSub and illustrates its potential, by investigating the impact of repression on dissent across thousands of subnational datasets.

Keywords

conflict, contention, disaggregation, event data, micro-foundations, protest, repression, subnational, violence

In the last two decades, social scientists have produced a tremendous amount of disaggregated data on political conflict and violence.[1] Large-scale collection projects (e.g. Schrodt, Davis & Weddle, 1994; Raleigh et al., 2010; Salehyan et al., 2012; Sundberg & Melander, 2013) and specialized studies of individual countries (e.g. Sullivan, Loyle & Davenport, 2012; Verpoorten, 2012; Osorio, 2015) have extracted georeferenced information on political events from press reports, social media, and state archives, manually or with automated techniques. These data have fueled new waves of subnational research, employing novel research designs at granular levels of analysis.

Disaggregation has advanced scholarship in numerous ways, but five interrelated problems have impeded progress: (1) most studies that use subnational data nevertheless conduct analysis at a highly aggregated, macro level; (2) most micro-level studies focus on one or few countries; (3) cross-dataset comparisons are rare; (4) operational definitions vary; and (5) there are no consistent units of analysis, which might otherwise enable direct comparisons. As a result, idiosyncratic and contradictory findings have become common in subnational research, leaving unanswered basic questions about the causes, dynamics, and consequences of conflict.

The barrier to generalizability is not a lack of data – in many cases these data exist and are in the public domain. Rather, it is that no one has undertaken the entrepreneurial effort to merge and combine disparate subnational conflict datasets into a unified, analysis-ready format, with consistent definitions, measures, and units. This is no small task: it involves geo-locating events, classifying them by type, assigning them to

[1] We define political conflict as a dispute between two or more political actors (e.g. governments, challengers, third parties) over the pursuit, maintenance or distribution of power. Actions to resolve the dispute may include both physical force and nonviolent measures (e.g. demonstrations, strikes, civil disobedience).

Corresponding author:
zhukov.yuri@gmail.com

administrative and geographic units, aggregating over time, and repeating for each country, dataset, and variable. Without such infrastructure-building efforts, the field cannot move forward.

With these thoughts in mind, we present xSub, a 'database of databases' for cross-national research on subnational conflict. xSub currently features subnational datasets from 21 sources, covering 156 countries, organized into consistent categories and units, by *space* (country, province, district, grid cell, electoral constituency), *time* (year, month, week, day), *actors* (government, opposition, civilian, unaffiliated), and *actions*.[2] xSub also provides data on local demographics, geography, ethnicity, weather, and other covariates. These data are freely available through x-sub.org or the xSub R package.

xSub has applications across a wide range of substantive areas. Scholars may use it to examine relationships between climate variability and conflict, terrain and insurgency, ethnicity and communal fighting, elections and protests, spillover effects, conflict duration, recurrence, micro-dynamics of non-state activities, and – as we demonstrate below – the relationship between repression and dissent. Users may also contribute original data to the platform, making their work available to new researchers asking new questions.

Why xSub?

Traditionally, research on political conflict – including genocide, (counter-)revolution, (counter-)insurgency, (counter-)terrorism, protest (policing), (violations of) civil liberties and human rights – has maintained a cross-national focus, tracking macro-level variation between nation-states. This work has produced innumerable insights about why conflicts begin (Collier & Hoeffler, 2004), become increasingly lethal (Poe & Tate, 1994), employ particular tactics (Kalyvas & Balcells, 2010), respond to diverse factors of contention (Davenport, 1995), endure (Fearon, 2004), and reoccur (Walter, 2004) – in some countries but not others. It has been less informative about variation within conflicts, and why contentious events occur at particular places and times, within the nation-state.

[2] A subnational dataset is an organized collection of information on the location, timing, participants, and properties of conflict events in a single country. A data source is an entity that collects or distributes data. One source may produce multiple datasets (e.g. ACLED), and multiple datasets may cover the same country.

To fill this gap, a growing movement of subnational research has disaggregated actors and actions to a more micro level. This movement has helped illuminate the local and short-term dynamics of conflict, advancing our understanding of when, where, and why unrest will likely emerge (Dube & Vargas, 2013), spread (Schutte & Weidmann, 2011), and vary in response to government action (Lyall, 2009).

While resolving some problems, subnational research has created others. To take stock of this rapidly growing literature, we surveyed the universe of topically related studies published between 2006 and 2017 in top disciplinary and general scientific journals.[3] We reviewed 392 articles, organized them by topic, geographic and temporal scope, unit of analysis, and methodology (Online appendix A1). Our survey revealed five common problems.

1. **Most studies underutilize subnational data**. While multinational disaggregated datasets like the Uppsala Conflict Data Program's Georeferenced Event Dataset (UCDP GED) and the Armed Conflict Location & Event Data Project (ACLED) have facilitated cross-national comparisons of subnational trends, most studies (94%) continue aggregating events to the country level, eschewing local geographic variation altogether. Absent strong theoretical reasons to focus only on macro-level patterns, this practice indicates an underutilization of subnational data.
2. **Most subnational studies focus on one or few countries**.[4] In-depth subnational analyses may have strong internal validity, but rarely demonstrate the generalizability of their findings. 46% of subnational studies focus on a single country, and 56% on a single region (e.g. sub-Saharan Africa, Eastern Europe). The self-contained study of individual countries and regions has led to geographic fragmentation in research. For instance, 64% of studies on climate and conflict use data on Africa. Yet Africa represents less than 1% of studies on indiscriminate violence,

[3] These include but are not limited to *Science*, *Nature*, *American Political Science Review*, *American Journal of Political Science*, *International Organization*, *World Politics*, *Journal of Peace Research,* and *Journal of Conflict Resolution* (see Scientific Journal Ranking, 2018: https://www.scimagojr.com/journalrank.php).

[4] We define a 'subnational study' as one that explores variation across spatial units of analysis smaller than country (e.g. grid, city, district, etc.).

repression, or occupational politics (Online appendix A2).

3. **Cross-dataset comparisons are rare**. 65% of the literature has used pre-existing conflict datasets, and the rest collected original data. Different datasets employ different sampling and measurement strategies, and rely on different sources (e.g. archives, newswires, NGOs). These choices are consequential for disaggregated and macro-level work (Eck, 2012; Davenport & Moore, 2015). Yet only 22 studies use cross-dataset validation,[5] and just three explore how these inconsistencies affect inference (Eck, 2012; Hammond & Weidmann, 2014; Ward et al., 2013).
4. **Operational definitions vary**. Among the many things that vary across data sources are typologies of actors and actions. Most studies (91%) focus on violence by government agents and challengers, sometimes disaggregating further by tactics. Large data collections like ACLED and the Social Conflict Analysis Database (SCAD) provide detailed descriptions of actors – down to specific protest movements and military commanders. Some, like UCDP GED, also provide short textual descriptions of events. These infinitely customizable choices create some non-trivial challenges. Determining which actors belong to pro-government or opposition factions requires familiarity with individual cases, and extracting information from text is not always straightforward. These challenges have made it difficult to compare subnational results with findings from other countries.
5. **There are no standard units**. Cross-national units of analysis have generally converged on country-years and dyad-years, but there is no similar 'industry standard' in subnational work. Research on some topics (e.g. indiscriminate violence, economic shocks) has been quite diverse in spatial and temporal scales. Other research (e.g. foreign aid, elections) has favored specific units, like country-years, avoiding subnational units commonly used in other domains, like grid cells. These choices are often data-driven, and lack clear theoretical motivations. Without a standard 'menu' of spatio-temporal aggregation options, it is unsurprising that most studies that use subnational data still analyze them at the national level.

xSub addresses these problems directly, by pulling together hundreds of existing subnational datasets, and aggregating conflict events and covariates to consistent units of analysis across countries. As a public good, xSub significantly reduces barriers to comparative subnational research, empowering researchers to quickly construct custom, analysis-ready datasets and compare their findings across countries and sources.

Similar initiatives already exist for macro-level conflict research, like NewGene (Bennett, Poast & Stam, 2017). There is no counterpart at the subnational level, despite several important integration efforts, including PRIO-GRID, a spatio-temporal grid structure for data compilation and analysis (Tollefsen, Strand & Buhaug, 2012), GROWup, a data platform for ethnic settlement patterns and conflict (Girardin et al., 2015), and geomerge, an R package to construct spatial panel datasets (Linke & Donnay, 2017). xSub builds on these efforts by offering barrier-free access to preprocessed event data, from multiple sources, at multiple resolutions. By providing a user-friendly web interface and an R package with more advanced functionality, xSub serves a methodologically diverse user base, from undergraduates to senior researchers. While not eliminating the need to understand the limitations of individual data sources, xSub makes these data more accessible, removing key obstacles to the accumulation of knowledge.

What is xSub?

xSub's online repository and accompanying R package currently feature 25,112 datasets on the location, dynamics, and intensity of conflict events, in 156 countries (1969–2017), from 21 data sources, with consistent categories and customizable spatio-temporal units. These datasets include popular covariates on weather, ethnicity, demographics, and geography. As such, xSub is well suited for single-country and cross-national analyses of conflict onset, (de)escalation, diffusion, termination, recurrence, and legacy.

xSub data are available as individual events and customizable spatial panel datasets, and are designed for integration with standard statistical software (e.g. Stata, R), and geographic information systems.[6] The platform is updated annually, with new data sources, bug fixes,

[5] This number only reflects studies that use multiple datasets for robustness checks (e.g. Buhaug, 2010), not studies that use covariates from multiple datasets (e.g. Ruhe, 2015).

[6] Each file contains spatial unit IDs for merging with GIS spatial geometry files, like ESRI shapefiles (.shp).

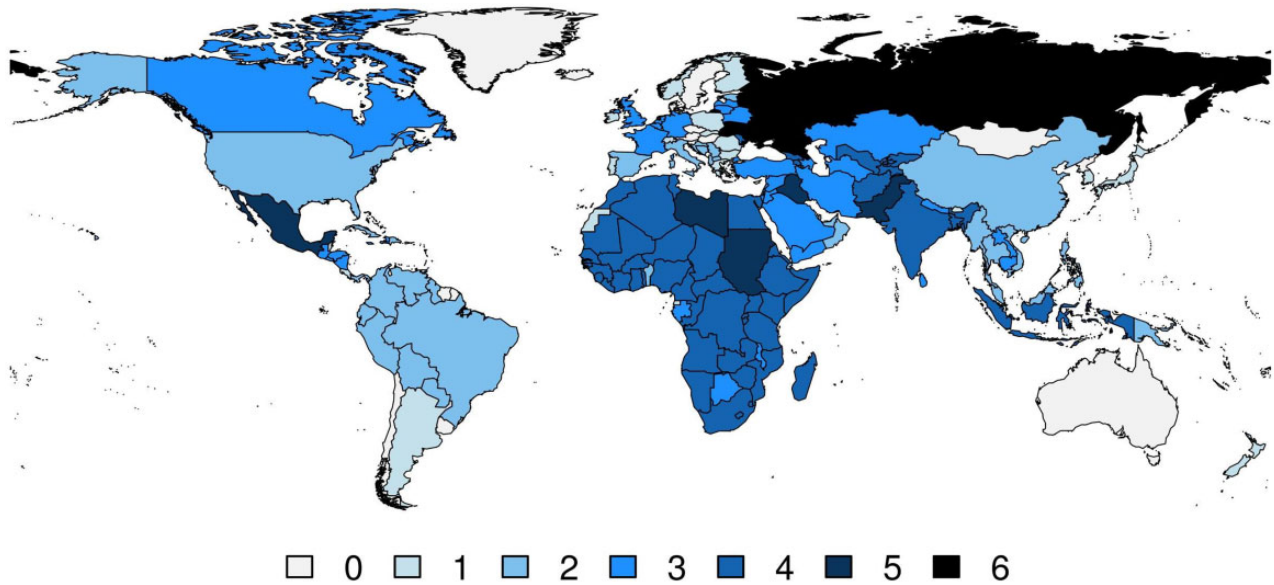


Figure 1. Number of unique data sources per country

and features.[7] All replication code is available through xSub's GitHub page.[8]

Data sources

xSub features event data on political conflict and violence from 21 sources, including widely used large-scale data collections and boutique datasets from scholars who volunteered to be early contributors to the platform.[9] xSub also welcomes uploads of users' original data, with submission guidelines specified at x-sub.org/data-upload.

Figure 1 shows the geographic distribution of xSub data, with darker colors indicating more data sources per country. At the atomic level, the data are individual events, with information on locations, dates, actors, and tactics. We use these to construct local event counts, at various levels of analysis, as well as pooled, multiple-source datasets, integrated with the MELTT software package (Donnay et al., 2019).[10]

Actors

xSub organizes events into *initiator-target* dyads, with four categories of actors: government (Side A), opposition (Side B), civilian (Side C), and unaffiliated (Side D).

Side A: The *government* category includes state security forces, pro-government militias, activists, and third parties acting on the incumbent's behalf (e.g. foreign troops, contractors). It excludes mutinous military factions, and supporters of ousted regimes.[11]

Side B: The *opposition* category includes rebels, dissidents, revolutionaries, anti-government militias, rioters and protesters, third parties

[7] Annual releases of xSub are scheduled for every November.

[8] https://github.com/zhukovyuri/xSub_ReplicationCode.

[9] xSub has secured written consent from each data provider, explicitly allowing proposed usage. Providers currently include Armed Conflict Location and Event Data (ACLED) (version 8) (Raleigh et al., 2010), Empirical Studies of Conflict (Worldwide Incidents Tracking System: Afghanistan, 2010 version), Iraq (version 3), Pakistan (2010 version) (Wigle, 2010), Iraq SIGACT (version 3; Berman, Shapiro & Felter, 2011), drug-related murders (2006–2011) and homicides (1998–2011) in Mexico (Calderón et al., 2015), BFRS data on Pakistan (version 10; Bueno de Mesquita et al., 2013), Political Instability Task Force's Worldwide Atrocities Dataset (version 1.1b1) (Schrodt & Ulfelder, 2016), Social Conflict Analysis Database (SCAD) (1990–2015) (Salehyan et al., 2012), UCDP's Georeferenced Event Dataset (UCDP GED; version 17.1) (Sundberg & Melander, 2013), American Bar Association's Darfur data (Totten, 2006), data from the Northern Ireland Research Initiative (Davenport, Loyle & Sullivan, 2017), National Violence Monitoring System's data on Indonesia (Barron, Jaffrey & Varshney, 2016), Beissinger's and Zhukov's data on the former Soviet Union (Beissinger, 2002; Toft & Zhukov, 2015; Zhukov, 2016).

[10] These datasets combine events from our 21 collections, offsetting underreporting in each source with information from another. We currently provide eight varieties of integrated data, with alternative spatial (1–5 km), temporal (1–2 days), and event taxonomy filters (directed, undirected dyad).

[11] For instance, we code the Taliban in Afghanistan as government 1996–2001, and as opposition post-2001.

Table I. xSub actor typology for all data sources

	Initiator				
Target	*Side A (government)*	*Side B (opposition)*	*Side C (civilians)*	*Side D (unaffiliated)*	*Subtotal*
Side A	DYAD_A_A	DYAD_B_A	DYAD_C_A	DYAD_D_A	TARGET_SIDEA
Side B	DYAD_A_B	DYAD_B_B	DYAD_C_B	DYAD_D_B	TARGET_SIDEB
Side C	DYAD_A_C	DYAD_B_C	DYAD_C_C	DYAD_D_C	TARGET_SIDEC
Side D	DYAD_A_D	DYAD_B_D	DYAD_C_D	DYAD_D_D	TARGET_SIDED
Subtotal	INITIATOR_SIDEA	INITIATOR_SIDEB	INITIATOR_SIDEC	INITIATOR_SIDED	

Table II. xSub event typology for all data sources

	Actor				
Action	*Side A (government)*	*Side B (opposition)*	*Side C (civilians)*	*Side D (unaffiliated)*	Subtotal
Any violence	SIDEA_ANY	SIDEB_ANY	SIDEC_ANY	SIDED_ANY	ACTION_ANY
Indirect violence	SIDEA_IND	SIDEB_IND	SIDEC_IND	SIDED_IND	ACTION_IND
Direct violence	SIDEA_DIR	SIDEB_DIR	SIDEC_DIR	SIDED_DIR	ACTION_DIR
Protests	SIDEA_PRT	SIDEB_PRT	SIDEC_PRT	SIDED_PRT	ACTION_PRT
Subtotal	SIDEA_ANY	SIDEB_ANY	SIDEC_ANY	SIDED_ANY	

acting on rebels' behalf, and other groups directly challenging the government.

Side C: *Civilians* are individuals who abstain from willful participation in politically contentious behavior. Civilians generally enter the dataset not as initiators of conflict events, but as unarmed victims of violence by any side.[12]

Side D: The final category (*other*) includes militias, tribes, self-defense units, and other actors *not* directly challenging or supporting the government. This group also includes factions that don't neatly fall into the first three categories due to political non-affiliation (e.g. intercommunal groups, criminal organizations, peacekeepers).

xSub's data sources differ in the actors they include and the typologies they use to organize them. Where data sources provide only detailed descriptions of involved actors, but no government/opposition labels, we construct source- and country-specific dictionaries (available at http://x-sub.org/about/what-is-xsub) to map actors to each category. The GitHub repository also provides code for users to create custom dictionaries and categories.

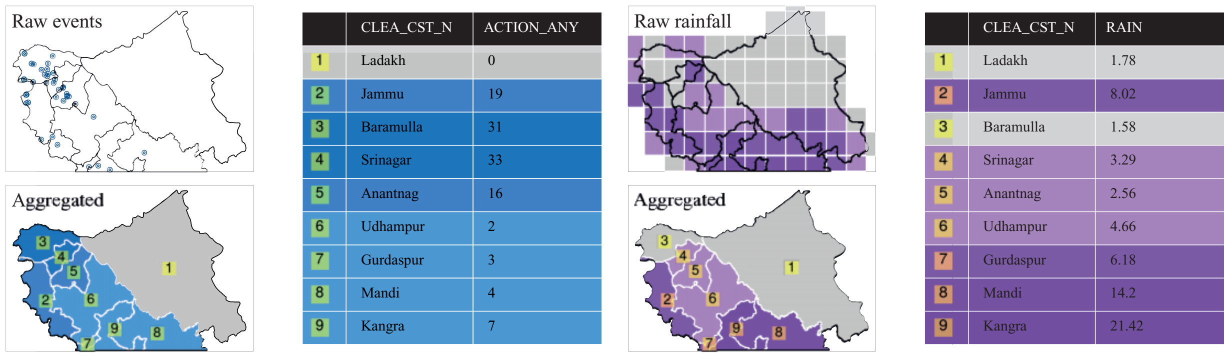
Table I illustrates the actor typology. Where data sources do not explicitly distinguish between the initiators and targets of conflict events (e.g. ACLED, UCDP-GED), the variables DYAD_* are to be interpreted as 'event involving actors X and Y' (undirected dyads). In all other cases, the interpretation is 'action by actor X against actor Y' (directed dyads).

Actions

xSub categorizes events into four categories of actions: (1) *Any* use of force, (2) *Indirect* force, including shelling, air strikes, chemical weapons, (3) *Direct* force, including firefights, arrests, assassinations, and (4) *Protests*, both violent and nonviolent.[13] While some data sources contain detailed information on tactics, most do not. Where there were no details, we coded actions as *Any*, or consulted the original article or author(s). Where data sources instead provided textual descriptions of events, we constructed a custom action dictionary, and used natural language processing to categorize the event into one of these categories. Table II summarizes the event typology. xSub also provides event-level data on specific actions, like air strikes, ambushes, armored offensives, arrests, artillery shelling, bombings, chemical weapons,

[12] For instance, we classify unarmed anti-government protesters as opposition (Side B) and classify armed civilians in local self-defense units as members of a fourth group (Side D), described below.

[13] This typology follows Balcells's (2011: 399–400) distinction between 'indirect' violence, which involves 'heavy weaponry [. . .] and does not require a face-to-face interaction', and 'direct' violence, which involves 'light weaponry [. . .] in a face-to-face interaction'.



(a) Point-to-polygon aggregation

(b) Polygon-to-polygon aggregation

Figure 2. Spatial aggregation

civilian abuses, displacement, firefights, kidnappings, killings, property damage, protests, raids, riots, riot policing, robberies, rockets, sieges, storms, suicide bombings, and terrorism.

Covariates

In addition to conflict, xSub includes other variables frequently used in subnational research: local demographics (e.g. population density), geography (e.g. elevation, roads, land cover), ethnicity (e.g. local nationalities, linguistic groups), and weather (e.g. temperature, rainfall). To make these covariates consistent and internationally comparable, we drew them from publicly available GIS datasets with global coverage. Some of the covariates are time-variant and date back to 1900 (e.g. precipitation). Others are static (e.g. elevation).

Units of analysis

xSub provides event-level and spatial panel datasets, the latter of which aggregate events and covariates to users' preferred spatial and temporal units. Geographic units include *countries*, *provinces* (first-order administrative divisions), *districts* (second-order divisions), *PRIO-GRID cells* (0.5 x 0.5 decimal degree lattice; Tollefsen, Strand & Buhaug, 2012), and *electoral constituencies* (Kollman et al., 2017). Temporal units include *years*, *months*, *weeks*, and *days*.

Figure 2 illustrates the spatial aggregation procedure, where raw data are (a) points, like event locations, and (b) polygons or grid cells, like weather. The upper left overlays raw data with spatial units of interest (here, electoral constituencies, CLEA_CST_N). The lower left displays aggregated measures for each unit, and the right pane shows the same in tabular form. For points (Figure 2a), we identify spatial units that contain each event, and generate local event counts at each time interval.[14] Because polygon borders do not always align (Figure 2b), we aggregate such data with area-weighted means.[15]

Figure 3 illustrates how one's choices of units affect the spatial and temporal distribution of conflict data in India: smaller units yield more variation, but also more sparsity.

User interface

xSub offers two options for customizing and downloading data. The first is an interactive web-based interface, at x-sub.org/data-download, where users can select countries, data sources or units of analysis, preview the data, and download a zipped archive with the requested data and supporting documentation. For example, the selection 'Country: India', 'Source: UCDP GED', 'Space: district', 'Time: week' will generate weekly observations for India's districts, with local event counts for each week (from UCDP GED, broken down by actor and tactics), and local average statistics for weather and other covariates. The second option, for more advanced uses, is the xSub R package, available

[14] If the original source does not include geographic coordinates, but only an address (e.g. name of city or village, like Beissinger, 2002), we geo-code events to the most precise location possible, with mapping APIs from Google, Yandex, and MapQuest.

[15] Formally, area-weighted means are $x_i = \sum_{j_i}^{J_i} w_{j_i} x_{j_i}$, where x is the variable of interest (e.g. weather), $i \in \{1, \ldots, N\}$ indexes destination spatial units (e.g. electoral constituency), $j \in \{1, \ldots, J\}$ indexes source units (e.g. grid cell), and J_i is the subset of source units that overlap with i. Weights $w_{j_i} = \frac{A_{i \cap j}}{A_i}$ represent the proportion of i's area (A_i) covered by the intersection of i and j ($A_{i \cap j}$).

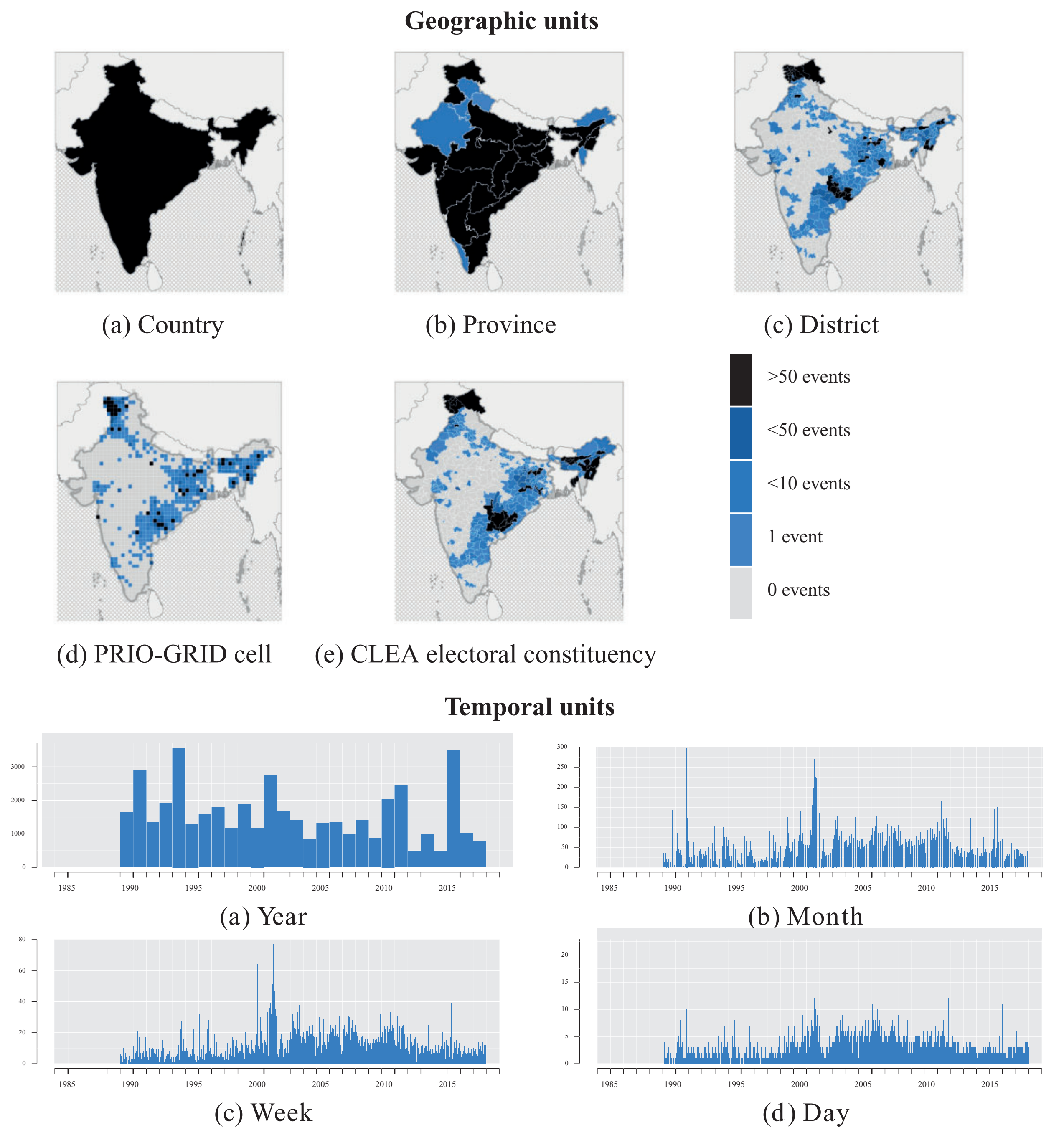


Figure 3. Geographic units of analysis
Country: India. Source: UCDP-GED. Variable: *Action any*.

through the https://cran.r-project.org/package=xSub repository. This package provides additional functionality not supported by the website, including direct import of data into R and merging of datasets across countries.[16]

[16] The website also provides Stata code to download and merge data files.

Illustration of use: Repression and dissent

To demonstrate how scholars might use xSub, we investigate the empirical relationship between overt manifestations of repression and dissent. A longstanding topic in civil conflict research is the contentious interaction between governments and challengers: how the actions and tactics of one side influence the actions and tactics of another, and whether escalation sparks reciprocal steps (Davenport, 2007). The dominant view is that crackdowns on opposition tend to inflame dissent (Gurr & Lichbach, 1986; Mason & Krane, 1989; Sullivan, Loyle & Davenport, 2012). Another school holds that, by raising the costs of behavioral challenges, repression is more likely to deter dissent (Tilly, 1978; Lyall, 2009). Some maintain that repression can have both effects, inflaming dissent at intermediate levels but deterring it at extremes – an 'inverted U' (Gurr, 1970; DeNardo, 2014) – or decreasing dissent at intermediate levels and increasing at extremes – a 'U-shape' (Lichbach & Gurr, 1981). Due to a plurality of idiosyncratic research designs and data sources – cross-national and subnational – the field has produced contradictory findings about which of these patterns is dominant.

To take stock of whether repression (i.e. government violence) increases or decreases opposition activity, we use xSub to conduct a meta-analysis across hundreds of subnational datasets. For each country and data source, we fit the following core model:

$$\begin{aligned} \text{Dissent}_{it} = \text{Repression}_{it-1}\beta_1 + \text{Repression}^2_{it-1}\beta_2 + \alpha_i \\ + \gamma_t + u_{it} \end{aligned} \quad (1)$$

where Dissent_{it} is the number of opposition actions in locality i at time t (SIDEB_ANY on Table II), Repression_{it-1} is the number of government uses of force in i at $t-1$ (SIDEA_ANY). We use locality fixed effects α_i to account for unobserved local factors influencing both repression and dissent, and time fixed effects γ_t, representing common shocks over time. To make estimates maximally comparable, we report standardized coefficients (i.e. impact of a standard deviation increase in repression on standard deviation changes in dissent).[17]

Our interest is in how the β coefficients vary across datasets. The relationship between repression and dissent is strictly inflammatory ('anger' or 'backlash') if $\beta_1 > 0, \beta_2 \geq 0$, with increases in repression followed by linear ($\beta_2 = 0$) or exponential ($\beta_2 > 0$) increases in protests. If $\beta_1 < 0, \beta_2 \leq 0$, the relationship is negative ('fear' or 'deterrence'), with dissent declining after repression. If $\beta_1 > 0, \beta_2 < 0$, the relationship is 'inverted-U'shaped ('anger then fear'), where increases in repression are correlated with increases in protests, but the rate of increase gradually declines. Finally, $\beta_1 < 0, \beta_2 > 0$ indicates the opposite, 'U-shaped' relationship ('fear then anger').

We repeated this analysis across thousands of xSub spatial panel datasets, limiting our inquiry to yearly and monthly datasets with at least ten incidents of government violence and ten protests.[18] This narrowed our empirical domain to 14,299 datasets from 113 countries, at multiple units of analysis.[19] Figure 4 reports the predicted shape of the repression–dissent relationship across spatio-temporal units and data sources, based on a weighted average of β_1, β_2 estimates.[20]

The results show a general tendency toward an 'inverted-U' relationship between repression and dissent: 'anger' at lower levels, 'fear' at higher levels ($\beta_1 > 0, \beta_2 < 0$). Overall (upper-left), a standard deviation increase in repression increases protests by .36 of one standard deviation, but also *decreases* growth in protests by .05 of one standard deviation. Completing the square ($-\frac{\beta_1}{2\beta_2}$), the slope changes from positive to negative when governments escalate violence to 3.7 standard deviations above the mean.

The relationship is highly sensitive to sources and spatio-temporal units: as units become smaller, the slope becomes increasingly positive. The curve's steepness also varies regionally: European states reach the inflection point earlier, on average, than African states – potentially indicating greater coercive capacity, lower opposition resolve, or both.

The goal of this analysis has been purely illustrative, demonstrating how one may use xSub to assess the local relationship between repression and dissent, and how it varies – in direction and magnitude – across conflicts and data sources. These models cannot – and are not intended to – identify a causal effect. A more rigorous analysis with xSub might consider additional sources of variation and bias, like the endogeneity of repression, interdependence, tactical shifts, and spatial autocorrelation.

[17] To account for past levels of dissent, Online appendix A3 considers a dynamic panel data analysis, using a generalized method of moments estimator.

[18] We excluded weekly and daily datasets for computational reasons.

[19] xSub's current collection includes 25,112 files: 6,313 individual-source panel datasets, 17,130 multiple-source panel datasets, 445 event-level individual-source datasets, and 1,224 event-level multiple-source datasets.

[20] Weighted means are $\bar{\beta} = \sum_d w_d \beta_d$, where β_d are coefficients from dataset d, and w_d are model weights, proportional to sample size of d ($N \times T$). We estimated confidence intervals through bootstrapping.

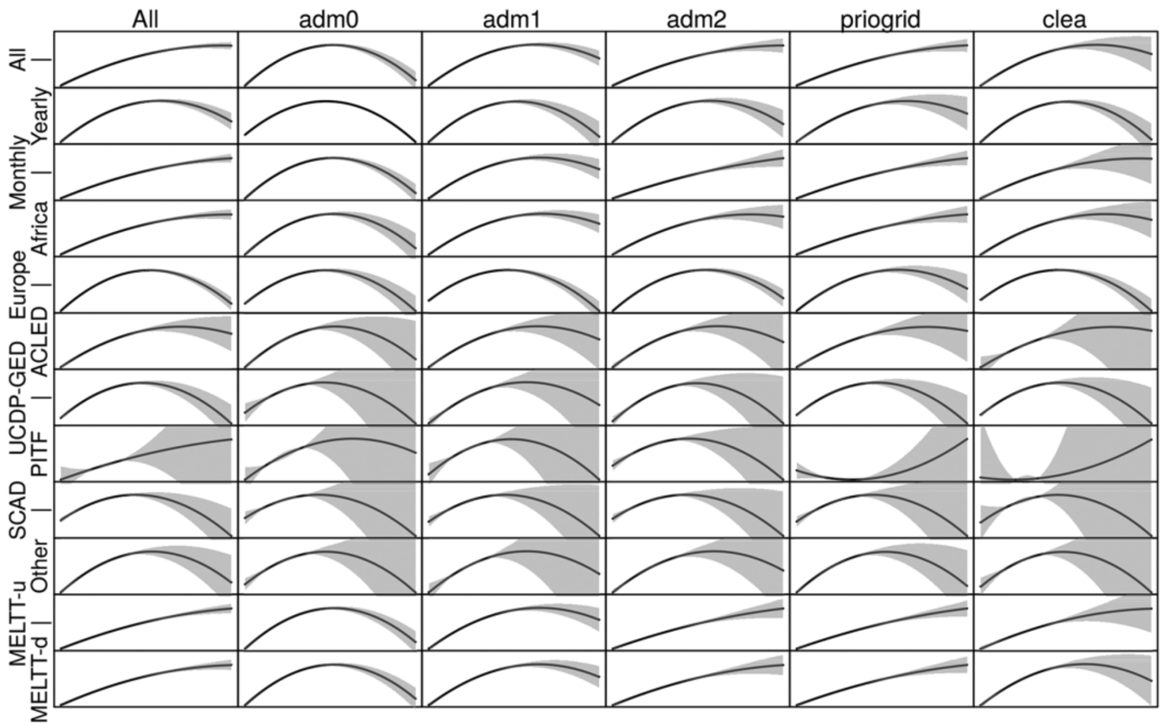


Figure 4. Meta-analysis of repression and dissent

Quantities reported are weighted averages of standardized coefficient estimates and bootstrapped 95% confidence intervals for the model in Equation (1). Spatial units are years [adm0], provinces [adm1], districts [adm2], grid cells [priogrid], and electoral constituencies [clea]. Data sources include ACLED, UCDP GED, PITF, SCAD, and the remaining providers listed in footnote 9 ('Other'). MELTT-u/-d are multiple-source undirected and directed dyadic data, integrated with MELTT (Donnay et al., 2019).

Further building on this example, xSub allows scholars to uncover many sources of heterogeneity. For instance, we may expect indirect artillery shelling to have a very different effect on dissent than arrests and detentions. Protesters may react to repression differently than armed rebels. Political actors may also behave differently in locations closer or farther from their primary bases of support. While space constraints limit the scope of our inquiry here, xSub users can easily assess the generalizability of empirical relationships, switching from aggregated (country-year) to disaggregated (district-month) scales, replicating results across countries and datasets, with consistent definitions and units.

Conclusion

xSub reduces the barriers to comparative subnational research, by empowering researchers to quickly construct custom, analysis-ready datasets, pre-loaded with several popular covariates. xSub also offers a platform for scholars to contribute and distribute their own, original data. One of the reasons for fragmentation in subnational research is that many individual data collection efforts are project-specific: scholars assemble a new dataset for a one-off paper, and – apart from posting a replication archive – never re-use those data again. Rather than allow a dataset to 'die' with a paper, xSub enables researchers to give their data a second life, in the hands of new researchers, asking a new set of questions. Let a thousand flowers bloom.

Replication data

The dataset, codebook, and R-files for the empirical analysis in this article, along with an Online appendix, can be found at http://www.prio.org/jpr/datasets.

Acknowledgments

We thank David Backer, Alex Braithwaite, David Cunningham, Karsten Donnay, Eric Dunford, Kathleen Gallagher Cunningham, Gary Goertz, Benjamin Graham, Macartan Humphreys, Volodymyr Ishchenko, Sebastian Kraus, Andrew Linke, Yonatan Lupu, Erin McGrath, Thomas O'Mealia, Andy Owsiak, Paul Staniland, Sebastian Schutte, Christopher Sullivan, Andreas Tollefsen, Henrik Urdal, Camber Warren, Nils Weidmann, Julian

Wucherpfennig, and Thomas Zeitzoff for helpful comments, and Yioryos Nardis and Andrew Versalle for technical support.

Funding

We are grateful for financial support from the Research Council of Norway (Davenport; Project 250441). Nadiya Kostyuk's work was partially supported by the William and Flora Hewlett Foundation.

ORCID iD

Yuri M Zhukov https://orcid.org/0000-0003-0436-0209
Christian Davenport https://orcid.org/0000-0001-5652-2914

References

Balcells, Laia (2011) Continuation of politics by two means: Direct and indirect violence in civil war. *Journal of Conflict Resolution* 55(3): 397–422.

Barron, Patrick; Sana Jaffrey & Ashutosh Varshney (2016) When large conflicts subside: The ebbs and flows of violence in post-Suharto Indonesia. *Journal of East Asian Studies* 16(2): 191–217.

Beissinger, Mark R (2002) *Nationalist Mobilization and the Collapse of the Soviet State*. Cambridge: Cambridge University Press.

Bennett, D Scott; Paul Poast & Allan C Stam (2017) NewGene: A conceptual manual. SSRN (https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2596297).

Berman, Eli; Jacob N Shapiro & Joseph H Felter (2011) Can hearts and minds be bought? The economics of counterinsurgency in Iraq. *Journal of Political Economy* 119(4): 766–819.

Bueno de Mesquita, Ethan; C Christine Fair, Jenna Jordan, Rasul Bakhsh Rais & Jacob N Shapiro (2013) The BFRS Political Violence in Pakistan dataset (https://esoc.princeton.edu/files/bfrs-political-violence-pakistan-dataset).

Buhaug, Halvard (2010) Climate not to blame for African civil wars. *Proceedings of the National Academy of Sciences* 107(38): 16477–16482.

Calderón, Gabriela; Gustavo Robles, Alberto Daz-Cayeros & Beatriz Magaloni (2015) The beheading of criminal organizations and the dynamics of violence in Mexico. *Journal of Conflict Resolution* 59(8): 1455–1485.

Collier, Paul & Anke Hoeffler (2004) Greed and grievance in civil war. *Oxford Economic Papers New Series* 56: 563–595.

Davenport, Christian (1995) Multi-dimensional threat perception and state repression: An inquiry into why states apply negative sanctions. *American Journal of Political Science* 39(3): 683–713.

Davenport, Christian (2007) State repression and political order. *Annual Review of Political Science* 10: 1–23.

Davenport, Christian & Will Moore (2015) Conflict consortium: Data standards and practices for observational data (http://conflictconsortium.weebly.com/uploads/1/8/3/5/18359923/cc-datastandardspractices7apr2015.pdf).

Davenport, Christian; Cyanne Loyle & Christopher Sullivan (2017) Northern Ireland research initiative (https://niresearchinitiative.weebly.com/#).

DeNardo, James (2014) *Power in Numbers: The Political Strategy of Protest and Rebellion*. Princeton, NJ: Princeton University Press.

Donnay, Karsten; Eric Dunford, Erin C McGrath, David Backer & David E Cunningham (2019) Integrating conflict event data. *Journal of Conflict Resolution* 63(5): 1337–1364.

Dube, Oeindrila & Juan F Vargas (2013) Commodity price shocks and civil conflict: Evidence from Colombia. *Review of Economic Studies* 80(4): 1384–1421.

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

Fearon, James D (2004) Why do some civil wars last so much longer than others? *Journal of Peace Research* 41(3): 275–301.

Girardin, Luc; Philipp Hunziker, Lars-Erik Cederman, Nils-Christian Bormann & Manuel Vogt (2015) *GROWup: Geographical Research On War, Unified Platform*. ETH Zurich.

Gurr, Ted Robert (1970) *Why Men Rebel*. Princeton, NJ: Princeton University.

Gurr, Ted Robert & Mark Irving Lichbach (1986) Forecasting internal conflict: A competitive evaluation of empirical theories. *Comparative Political Studies* 19(1): 3–38.

Hammond, Jesse & Nils B Weidmann (2014) Using machine-coded event data for the micro-level study of political violence. *Research & Politics* 1(2). https://doi.org/10.1177/2053168014539924.

Kalyvas, Stathis N & Laia Balcells (2010) International system and technologies of rebellion: How the end of the Cold War shaped internal conflict. *American Political Science Review* 104(3): 415–429.

Kollman, Ken; Allen Hicken, Daniele Caramani, David Backer, David Lublin, Joel Selway & Fabricio Vasselai (2017) *GeoReferenced Electoral Districts Datasets (Beta)*. Ann Arbor, MI: University of Michigan Press.

Lichbach, Mark Irving & Ted Robert Gurr (1981) The conflict process: A formal model. *Journal of Conflict Resolution* 25(1): 3–29.

Linke, Andrew M & Karsten Donnay (2017) Scale variability misclassification: The impact of spatial resolution on effect estimates in the geographic analysis of foreign aid and conflict. Paper presented at ther annual meeting of the International Studies Association, 24 March, Baltimore, MD.

Lyall, Jason (2009) Does indiscriminate violence incite insurgent attacks? Evidence from Chechnya. *Journal of Conflict Resolution* 53(3): 331–362.

Mason, T David & Dale A Krane (1989) The political economy of death squads: Toward a theory of the impact of state-sanctioned terror. *International Studies Quarterly* 33(2): 175–198.

Osorio, Javier (2015) The contagion of drug violence: Spatiotemporal dynamics of the Mexican war on drugs. *Journal of Conflict Resolution* 59(8): 1403–1432.

Poe, Steven & C Neal Tate (1994) Repression of human rights to personal integrity in the 1980s: A global analysis. *American Political Science Review* 88(4): 853–872.

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

Ruhe, Constantin (2015) Anticipating mediated talks: Predicting the timing of mediation with disaggregated conflict dynamics. *Journal of Peace Research* 52(2): 243–257.

Salehyan, Idean; Cullen S Hendrix, Jesse Hamner, Christina Case, Christopher Linebarger, Emily Stull & Jennifer Williams (2012) Social conflict in Africa: A new database. *International Interactions* 38(4): 503–511.

Schrodt, Philip A & Jay Ulfelder (2016) Political Instability Task Force Worldwide Atrocities event data collection, version 1.1b1 (http://eventdata.parusanalytics.com/data.dir/atrocities.html).

Schrodt, Philip A; Shannon G Davis & Judith L Weddle (1994) Political science: KEDS – a program for the machine coding of event data. *Social Science Computer Review* 12(4): 561–587.

Schutte, Sebastian & Nils B Weidmann (2011) Diffusion patterns of violence in civil wars. *Political Geography* 30(3): 143–152.

Sullivan, Christopher M; Cyanne E Loyle & Christian Davenport (2012) The coercive weight of the past: Temporal dependence and the conflict–repression nexus in the Northern Ireland 'troubles'. *International Interactions* 38(4): 426–442.

Sundberg, Ralph & Erik Melander (2013) Introducing the UCDP Georeferenced Event Dataset. *Journal of Peace Research* 50(4): 523–532.

Tilly, Charles (1978) *From Mobilization to Revolution*. Boston, MA: Addison-Wesley.

Toft, Monica Duffy & Yuri M Zhukov (2015) Islamists and nationalists: Rebel motivation and counterinsurgency in Russia's North Caucasus. *American Political Science Review* 109(2): 222–238.

Tollefsen, Andreas Forø; Håvard Strand & Halvard Buhaug (2012) PRIO-GRID: A unified spatial data structure. *Journal of Peace Research* 49(2): 363–374.

Totten, Samuel (2006) The US investigation into the Darfur crisis and the US government's determination of genocide. *Genocide Studies and Prevention* 1(1): 57–78.

Verpoorten, Marijke (2012) Leave none to claim the land: A Malthusian catastrophe in Rwanda? *Journal of Peace Research* 49(4): 547–563.

Walter, Barbara F (2004) Does conflict beget conflict? Explaining recurring civil war. *Journal of Peace Research* 41(3): 371–388.

Ward, Michael D; Andreas Beger, Josh Cutler, Matt Dickenson, Cassy Dorff & Ben Radford (2013) Comparing GDELT and ICEWS event data (https://benradford.github.io/images/publications/GDELTICEWS.pdf).

Wigle, John (2010) Introducing the Worldwide Incidents Tracking System (WITS). *Perspectives on Terrorism* 4(1).

Zhukov, Yuri M (2016) Trading hard hats for combat helmets: The economics of rebellion in eastern Ukraine. *Journal of Comparative Economics* 44(1): 1–15.

YURI M ZHUKOV, b. 1980, PhD in Government (Harvard University, 2014); Assistant Professor of Political Science, University of Michigan (2014–); research interests: international security, civil conflict.

CHRISTIAN DAVENPORT, b. 1965, PhD in Political Science (Binghamton University, 1992); Professor of Political Science, University of Michigan (2012–); current research interests: repression, human rights, political conflict, contentious politics; recent books: *The Peace Continuum: What It Is and How to Study It* (Oxford University Press, 2018) and *How Social Movements Die: Repression and Demobilization of the Republic of New Africa* (Cambridge University Press, 2015).

NADIYA KOSTYUK, b. 1987, MA (University of Michigan, Ann Arbor, 2016); PhD candidate in Political Science and Public Policy, University of Michigan, Ann Arbor; Pre-doctoral Fellow, Belfer Center for Science and International Affairs, Harvard Kennedy School (2017–18); Pre-doctoral Fellow, Computer Science Department and Fletcher School of Law and Diplomacy, Tufts University (2019–20); current main interests: international security, cyber conflict and coercion.