Erschienen in: International Interactions ; 48 (2022), 3. - S. 471-491
https://doi.org/10.1080/03050629.2022.2025365

Race to the bottom: Spatial aggregation and event data

Scott J. Cook[a] and Nils B. Weidmann[b]

[a]Texas A&M University; [b]University of Konstanz

ABSTRACT

Researchers now have greater access to granular georeferenced (i.e., spatial) data on social and political phenomena than ever before. Such data have seen wide use, as they offer the potential for researchers to analyze local phenomena, test mechanisms, and better understand micro-level behavior. With these political event data, it has become increasingly common for researchers to select the smallest spatial scale permitted by the data. We argue that this practice requires greater scrutiny, as smaller spatial or temporal scales do not necessarily improve the quality of inferences. While highly disaggregated data reduce some threats to inference (e.g., aggregation bias), they increase the risk of others (e.g., outcome misclassification). Therefore, we argue that researchers should adopt a more principled approach when selecting the spatial scale for their analysis. To help inform this choice, we characterize the aggregation problem for spatial data, discuss the consequences of too much (or too little) aggregation, and provide some guidance for applied researchers. We demonstrate these issues using both simulated experiments and an analysis of spatial patterns of violence in Afghanistan.

Los investigadores tienen ahora un acceso como nunca antes a datos georreferenciados granulares (es decir, espaciales) sobre fenómenos sociales y políticos. Estos datos se han utilizado ampliamente, ya que ofrecen a los investigadores la posibilidad de analizar fenómenos locales, probar mecanismos y comprender mejor el comportamiento a nivel micro. Con estos datos sobre acontecimientos políticos, es cada vez más frecuente que los investigadores seleccionen la escala espacial más pequeña que permitan los datos. Sostenemos que esta práctica requiere un mayor escrutinio, ya que las escalas espaciales o temporales no necesariamente mejoran la calidad de las inferencias. Si bien los datos altamente desagregados reducen algunas amenazas para la inferencia (por ejemplo, el sesgo de agregación), aumentan el riesgo de otras (por ejemplo, la clasificación errónea de los resultados). Por lo tanto, sostenemos que los investigadores deberían adoptar un enfoque basándose más en principios a la hora de seleccionar la

KEYWORDS

Ecological inference; event data; measurement error; spatial scale

CONTACT Scott J. Cook sjcook@tamu.edu Department of Political Science, Texas A&M University, College Station, TX 77843, USA

This paper benefited from comments provided by Benjamin Bagozzi and John Freeman, and also audience feed back at ETH Zurich, the University of Mannheim, and the 2019 Annual Meeting of the Peace Science Society International. All remaining errors are ours alone. For further questions contact Scott J. Cook.

escala espacial para su análisis. Para contribuir a realizar esta elección, caracterizamos el problema de la agregación de los datos espaciales, analizamos las consecuencias de una agregación excesiva (o insuficiente) y ofrecemos algunas orientaciones para la investigación aplicada. Demostramos estas cuestiones utilizando tanto experimentos simulados como un análisis de los patrones de violencia en Afganistán.

Les chercheurs ont maintenant un meilleur accès à des données granulaires géoréférencées (c-à-d, spatiales) sur les phénomènes politiques et sociaux que jamais auparavant. Ces données ont été largement utilisées, car elles offrent aux chercheurs le potentiel d'analyser des phénomènes locaux, de tester des mécanismes et de mieux comprendre les comportements au niveau micro. Avec ces données sur les événements politiques, il est devenu de plus en plus courant pour les chercheurs de sélectionner la plus petite échelle spatiale permise par les données. Nous soutenons que cette pratique exige un examen plus approfondi, car des échelles spatiales ou temporelles plus petites n'améliorent pas nécessairement la qualité des déductions. Bien que les données très désagrégées réduisent certains risques pour les déductions (p. ex. biais d'agrégation), elles accroissent le risque d'autres facteurs (p. ex. mauvaise classification des résultats). Par conséquent, nous soutenons que les chercheurs devraient adopter une approche plus raisonnée lorsqu'ils choisissent l'échelle spatiale pour leur analyse. Afin d'éclairer ce choix, nous caractérisons le problème de l'agrégation des données spatiales, nous discutons des conséquences d'une trop grande (ou trop faible) agrégations des données et nous fournissons quelques conseils aux chercheurs appliqués. Nous démontrons ces problèmes en utilisant à la fois des expérimentations simulées et une analyse des schémas spatiaux de la violence en Afghanistan.

Introduction

Observational data in political science are often wedded to a particular unit of analysis, with measurement and analysis occurring on the same scale. For example, electoral results are usually measured at the precinct level, unemployment insurance information is often reported by federal states, and GDP data is typically available for entire countries. In these instances, the unit of analysis is a (or *the*) political actor of interest, offering a natural correspondence between the data and analysis. In other issue areas, however, there is no "natural" unit of analysis, and researchers have considerable flexibility when selecting the spatial scale. For example, in conflict studies, researchers often use fine-grained event data on political violence and social movements as for example the UCDP's Geo-referenced Event Dataset (Högbladh 2019; Sundberg and Melander 2013) or the Social Conflict Analysis Database (Salehyan et al. 2012), which contain unique spatial coordinates for each event. Since each

event is linked to a single point, researchers can conduct analyses at various (theoretically infinite) levels of analytical resolution. They simply (pre-)select the scale of the units and then summarize the event-level information contained within.

While this flexibility affords the possibility for greater insights into political processes, it also risks researcher-induced error in these evaluations. Historically, data were often only available at highly aggregated levels (e.g., state-level), thereby preventing researchers from "drawing meaningful inferences about the underlying political relationships" (Freeman 1989, 61). With granular event data researchers are freed from these constraints and applied for work at sub-national scales has become commonplace. For example, several studies analyze political violence at sub-national administrative regions (e.g., municipalities, see Weidmann 2011) or geospatial units (e.g., grid cells, see Buhaug et al. 2011). The rapid proliferation of these studies reflects the belief that sub-national political processes *cannot* be effectively analyzed with highly aggregated data. However, it is not obvious to us that researchers are giving equal attention to whether sub-national processes *can* be effectively analyzed with existing sub-national data. More generally, it is not clear whether in considering the benefits of disaggregated data analysis, researchers are also considering its costs.

In this paper, we consider one problem related to the choice of spatial resolution that has received little attention in applied research: geolocation accuracy. Since event data are usually drawn from secondary sources, location information on events is often imprecise or inaccurate.[1] As such, even though testing a particular theoretical mechanism oftentimes implies a fine-grained analytical resolution, our data are often not sufficiently accurate to allow for effective analysis of processes at these levels. For example, in the UCDP Geo-referenced Event Dataset, data only about half of the events in Afghanistan can be located precisely within a 25 km radius; an area of almost 2,000 km^2 (Weidmann 2015). As a consequence, in their attempt to resolve bias due to aggregation, researchers induce bias due to measurement error in using smaller spatial scales.

Adjudicating the relative costs of this inferential trade-off in applied empirical research is not possible without ground truth data. Instead, we utilize Monte Carlo experiments to examine how data accuracy and the level of aggregation jointly affect the quality of our inferences. In our experiments, we contrast "local" (i.e., high resolution) and "regional" (i.e., low resolution) models, under different levels of local-level variation and random measurement error in both the outcome and predictor. As expected, local-level (i.e., disaggregated) analysis performs well when there

[1]Researchers constructing event data have considered this issue more directly, see Lee, Liu, and Ward (2019) for a recent discussion on improving the extraction of event locations from media reports.

is local-level variation in the process and no measurement error in the data. However, any deviation from these ideal conditions causes the performance of this model to diminish dramatically. For example, even under relatively low levels of random measurement error, regional-level analysis of aggregated data often provides estimates closer to the true values. These results were obtained without considering even more severe threats to inference with disaggregated event data, such as systematic geolocation error (Hammond and Weidmann 2014) and underreporting of events (Cook et al. 2017).

To be clear, we are not arguing that researchers should never analyze data at disaggregated levels. Rather, we recommend that extreme disaggregation should not be adopted as the default without further consideration of possible trade-offs. Much of the literature has focused primarily on the consequences of *over*-aggregation, without considering the additional challenges that arise with imperfect real-world spatially disaggregated data. Our simulation studies demonstrate that under plausible (and relatively benign) conditions, analysis of disaggregated data can actually decrease the quality of our inferences. Therefore, instead of always preferring the highest level of resolution, both theory and data quality—and not simply data availability—should inform one's choice of spatial scale.

Analytical Resolution in Political Science Research

The question of the "right" degree of analytical resolution for empirical work has a long history in geography (Openshaw and Taylor 1979; Openshaw 1984) and the social sciences (Achen and Shively 1995; King 1997). Much of this research has highlighted the shortcomings—ex. aggregation bias and ecological fallacy—of analyzing aggregate data to assess individual behavior (Robinson 1950). For example, a researcher seeking to understand the micro-level behavior of individuals may draw false conclusions when using country-level data (Achen and Shively 1995). Recognizing these limitations, applied researchers interested in political behavior now increasingly use individual-level data and multilevel models (Gelman et al. 2007; Glynn and Wakefield 2010).

More broadly, concerns of aggregate data analysis have inspired a shift toward smaller spatial units in research across the social sciences. This includes quantitative work in the study of contentious politics and political violence, which has increasingly relied on "disaggregated" approaches (Cederman and Gleditsch 2009). Disaggregated studies try to increase the resolution of the empirical data and tests used in a study, such that it better matches the causal process under examination. Most of the work required for this has gone into the creation of new event-level datasets (Schrodt

2012). For example, rather than using aggregate indicators of political events at the country level, researchers have collected geographically disaggregated data on civil war violence (Sundberg and Melander 2013), protest (Salehyan et al. 2012; Klein and Regan 2018; Weidmann and Rød 2019), and militarized interstate disputes (Braithwaite 2010).

These datasets provide incident-level information on discrete events, including time, location, actors, and other characteristics. As such, each observation corresponds to a specific spatial point, often identified by geographic coordinates (often latitude and longitude). Research in political science, however, is often not concerned with the precise location of specific events, but instead in accounting for the occurrence or frequency of events within an area. To analyze that, researchers often aggregate these individual events into area summaries and then include these in their models. In this aggregation step, researchers using event data have considerable flexibility in selecting the resolution for their subsequent analysis. This is because these individual events can be aggregated in literally infinite ways, ranging from very small (such as grid cells or small administrative units), to much larger spatial units (such as provinces or entire countries).

Reflecting this flexibility, we have seen event data aggregated in a variety of ways. While some studies use real-world administrative boundaries at low levels (Weidmann and Ward 2010; Weidmann 2011), most researchers have preferred artificial grid cells as units of analysis, which allows them to freely choose the size of these cells. The preferred dimensions of grids can vary, with common use ranging from quite small cells ($<$10 km, see Raleigh and Hegre 2009; Hegre, Østby, and Raleigh 2009) to relatively large cells with a size of 100 by 100 km (Buhaug and Rød 2006). Yet, the vast majority of spatial work on conflict and contention uses the default dimensions of the PRIO-Grid dataset (Tollefsen, Strand, and Buhaug 2012). With a resolution of 0.5° (55 km at the equator) and several pre-computed variables, this dataset has become a new standard for small-scale, spatially disaggregated analyses. Work using the PRIO-Grid has been highly influential, including highly-cited research by Pierskalla and Hollenbach (2013) on cellphones and violence, Fjelde and Hultman (2014) on violence against civilians, and von Uexkull (2014) on drought and violence.

While this degree of flexibility can be advantageous, it is not clear to us whether careful enough attention is being paid when making these aggregation choices. Failing to meaningfully consider the level of resolution can be problematic for at least two reasons. First, as is well-known from the vast literature on the Modifiable Areal Unit Problem (MAUP), different aggregation and zoning choices can dramatically affect results (Openshaw 1984). In short, Openshaw (1984) points out that there are an infinite number of ways to aggregate "non-modifiable" units (i.e., individual people) into

Table 1. Event data: promise *vs.* practice.

	Ideal type event data	Real world event data
Completeness	All events included	Incomplete list of events
Time and location of event	Accurately coded	Coding often inaccurate
Covariate data availability	Available at high resolution	Missing or poorly measured

higher-order units (e.g., households, counties, states, etc.). This is problematic since, like widely-known issues from ecological inference (King 1997), relationships found at aggregate levels of analysis (ex. county-level demographics and vote totals) differ from relationships found at individual levels of analysis (ex. individual characteristics and vote choice). Relationships of interest in event data analysis (ex. population and conflict) are also scale-dependent. Schrodt (2012) even highlights this as one of the open issues in event data research, since aggregation choices influence the results we obtain.[2]

Second, even if researchers knew the optimal spatial scale for their analysis—enabling them to evade MAUP-related concerns—it may be that available data do not support credible analysis at this level of resolution. For example, expert knowledge may suggest that a disaggregated research design would help us get closer to micro-level mechanisms driving violent behavior. However, if data at that level are missing or of low quality, then little would be gained from such an analysis. In extreme cases, low data quality may even cause us to draw incorrect inferences on these relationships of interest. This is especially problematic if researchers believe that higher-resolution analysis (on poor-quality data) necessarily improves upon prior lower-resolution analyses (on higher-quality data) since it would cause us to incorrectly update our prior understanding. In sum, researchers need to consider both theory and data quality when selecting their level of analysis.

As we summarize in Table 1, the event data utilized in applied research is often limited in different ways. First, event data sets do not often contain an exhaustive list of all real-world events (Earl et al. 2004). Moreover, this underreporting is often systematic, with the severity or location of the event affecting the probability of reporting (Salehyan 2015; Weidmann 2016; Cook et al. 2017). Secondly, even when events are included in our data, the characteristics of these events may be inaccurate. For example, the timing and location of the events reported in a given data set may be incorrect due to imprecise (i.e., events only located in a broad area) or inaccurate (i.e., events located in the incorrect area) media reports on the

[2]While much of the existing discussion on aggregation choices has focused on the frequency (i.e., temporal aggregation) of the data (Freeman 1989; Shellman 2004), the spatial aggregation choices discussed here present related challenges. Even in a single time series, the selection of the *spatial* scale is important. For example, patterns of unemployment over time will be different for a county *vs.* a state *vs.* a country.

event. As noted above, Weidmann (2015) finds that in the UCDP GED for Afghanistan, only about half of the events can be located precisely within a 25 km radius. Finally, even when the event data itself is accurate, researchers may not have high-quality data on the correlates of these events. Since we are often interested in understanding the detriments of political events, having high-quality input data (to use as covariates) is as important as high-quality output data.

While the first problem (completeness) is an issue no matter the level of resolution, the severity of the second (time and geolocation accuracy) and the third (covariate data availability) are often increasing in the resolution of the data. That is, as we analyze increasingly disaggregated units, the likelihood of inferential errors due to these data limitations increases.[3] Consider geolocation accuracy: it is widely appreciated that we have limited spatial accuracy for a subset of the incidents in most event datasets. For example, we find that roughly 20–40% of events cannot be spatially referenced more precisely than second-order administrative units:

- Social Conflict in Africa Database (SCAD): 24% of events (4,172/17,644) can only be located at "Province/region", "Nationwide", or "location unknown,"
- UCDP Geo-referenced Event Dataset (GED): 41% of events (58,353/142,902) are only known at the "second-order administrative division" or lower,
- Global Terrorism Database (GTD): 20% of events (36,689/181,685) can only be located at the second-order administrative unit or lower.

Attempting to locate events into smaller units than those supported by the original report necessarily increases the magnitude of geolocation error. This induces a form of misclassification in discrete outcomes, which biases coefficient estimates even when the error is random (Carroll et al. 2006). As these errors are likely not random—since the accuracy of reporting accuracy increases for events near cities, violent events, etc.—this bias is not even necessarily attenuating.

The magnitude of these biases grows further still when we consider that the quality of covariate information also diminishes as we further disaggregate the data. For example, Nordhaus (2006) provides a widely-used measure of sub-national economic activity. However, the quality of these data varies considerably both within and across countries, ranging from

[3]Some researchers instead choose to subset the event data and focus only on high precision incidents using geolocation accuracy indicators. There are two problems with this. First, as shown by Weidmann (2015) there is still substantial geolocation error even among these incidents. Second, this trades off measurement error for selection bias by actively omitting incidents from the analysis.

government-supplied measures of economic production for second-order administrative units (as for the US states) to no regional economic data at all (as in Nigeria). At best, this measurement error in the covariates will attenuate the coefficient estimates. However, given that the sub-national data quality does not vary at random and correlates with the quality of outcome data (e.g., both covariates and outcomes will be measured less precisely in low-income countries), the direction of the bias cannot be assumed.

Taken together, it cannot be assumed that disaggregating data always improves the quality of our inferences. When increasing the resolution of an empirical analysis, researchers instead face an unavoidable trade-off. On the one hand, higher-resolution analysis can help them get closer to the causal mechanism under study and alleviate problems of ecological inference. On the other, this requires increasingly high levels of precision in measurement, which is something we often do not have in social science data. As a result, researchers often introduce measurement error into outcomes and covariates as they disaggregate their data. Therefore, it is important for researchers to recognize the *costs* of undertaking disaggregated data analysis with imperfect data as well as its *benefits*. In the next section, we undertake a series of simulation experiments that demonstrate the magnitude of these biases under different levels of data quality.

Simulation Studies

To demonstrate the small sample consequences of different aggregation choices, we undertake a series of simulated experiments. For clarity, we focus on a stylized example where we have 900 observations (N), each located in 100 low-resolution areas (such as regions), that are further subdivided into nine high-resolution units (such as cities, see Figure 1).[4] This enables us to generate both outcome and predictor data at both levels, with varying degrees of measurement error. Since the terms low resolution and high resolution may not be familiar to readers—and could be confusing given that higher-resolution data corresponds to a lower scale in mapping—we use the terms "local" (or L) for high-resolution data and "regional" (or R) for low-resolution data in the following sections.

One thousand realizations (i.e., samples) of the data are generated at the local (i.e., high resolution) level as:

$$\Pr(Y_{L,R} \quad 1) \quad \text{logit}^{\ 1}(\beta_0 + \beta_1 X_{L,R}), \tag{1}$$

[4]These dimensions reflect a hypothetical single county analysis. For context, Nigeria has 36 states (first order administrative units) and 327 PRIO grid cells.

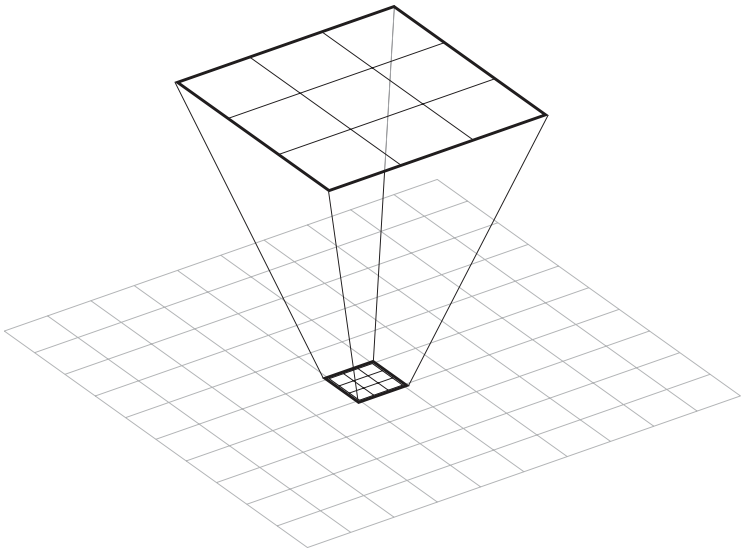


Figure 1. Simulation sample, 100 low-resolution (or regional) areas each sub-divided into 9 high-resolution (or local) areas.

where we set β_0 to −4, β_1 to 3, and the input

$$X_{L,R} \quad \alpha_R + u_{L,R}, \tag{2}$$

varies at both the regional, $\alpha_R \sim \mathcal{N}(0,1)$, and local level, $u_{L,R} \sim \mathcal{N}(0,\gamma)$. Differences in γ vary the extent of the local-level variation with a region, with results for $\gamma \quad 0$ (no local variation) and $\gamma \quad 1$ (high local variation) presented below.[5]

In our simulations, $Y_{L,R}$ and $X_{L,R}$ represent the true data. Were these fully observed by the researcher, standard estimators for model coefficients and marginal effects would be unbiased and efficient (up to usual sampling error). However, as we elaborate above, in the Analytical Resolution in Political Science Research section, this is rarely the case with event data, either because the researchers do not possess local data (only regional realization) or the local data they have is measured with error. To better approximate these real-world considerations, we aggregate and/or induce measurement error into these data.

In each set of simulated conditions, we assume that the researcher has access to two types of data and, therefore, two possible models at their disposal: a local model and a regional model. In the local model, researchers

[5]Rather than generating the data via a logistic function, we could also generate these as "presence only" data from a continuous Poisson process, locate the individual points within each area, and summarize the results. We do not believe that this would produce meaningfully different results over those we present here. Since it would also require the introduction of a technology (i.e., continuous Poisson processes) unfamiliar to some readers, we fear it would distract readers from our main point: spatial aggregation.

have local-level measures of $\tilde{Y}_{L,R}$ and $\tilde{X}_{L,R}$, which are proxies for $Y_{L,R}$ and $X_{L,R}$ with varying degrees of error (described below). In the regional model, researchers have regional-level aggregates of the true local-level data:

$$\begin{aligned} &\text{Outcome Aggregation: } Y_R \quad 1\left(\sum_{1}^{9} Y_{L,R} > 0\right) \\ &\text{Predictor Aggregation: } X_R \quad \alpha_R, \end{aligned}$$

where Y_R assumes a value of 1 if any of the local units contained within that region is a 1, and X_R reflects the common mean of its local units.[6] Notice that there is no additional error in these data, only that induced by aggregation.[7] This reflects situations where researchers are interested in answering questions on local-level processes, but only possess, or are confident in, regional, aggregated data. This trades off against possible error induced from moving to local data, which better reflect the true (local-level) process, but may do so with noise (as described below). All of the models reported below are estimated using logistic regression.

Random Measurement Error in X

In our first set of experiments, we assume that the local outcome ($Y_{L,R}$) is perfectly observed, confining attention to variation in the predictor. First, we vary the extent of the local-level variation in our input $X_{L,R}$. As noted above, we can manipulate this via γ in Equation (2), which is the variance of the local component of $X_{L,R}$. Second, we vary whether the predictor used in estimation is measured with error, by inducing non-differential (i.e., random) measurement error as:

$$\tilde{X}_{L,R} \quad X_{L,R} + e_{L,R},$$

where $e_{L,R} \sim \mathcal{N}(0, \delta)$. Differences in δ vary the magnitude of the measurement error in $\tilde{X}_{L,R}$.

In our simulations, we allow γ and δ to be either 0 or 1. This gives us four possible combinations of $\tilde{X}_{L,R}$: no true variance, no error variance ($\gamma \quad 0$, $\delta \quad 0$); true variance, no error variance ($\gamma \quad 1$, $\delta \quad 0$); no true variance, error variance ($\gamma \quad 0$, $\delta \quad 1$); true variance, error variance ($\gamma \quad 1$, $\delta \quad 1$). While this is a stylized setup since real-world data often have combinations of both in different magnitudes, focusing on these limiting cases

[6] While there are alternative ways of aggregating *Y* (as for example by summing the events in each region), this is the usual strategy employed in empirical work and therefore the one we consider here.

[7] As always with simulated data, we could generalize the model further still and consider scenarios that would increase (e.g., systematic measurement error, missing outcomes) or decrease (e.g., regional level measurement error) the differences between the local and regional level models. It is not clear to us, however, that this additional complexity would help to better clarify our main point: when the magnitude of measurement errors varies across spatial scales this can have a dramatic effect on inferences.

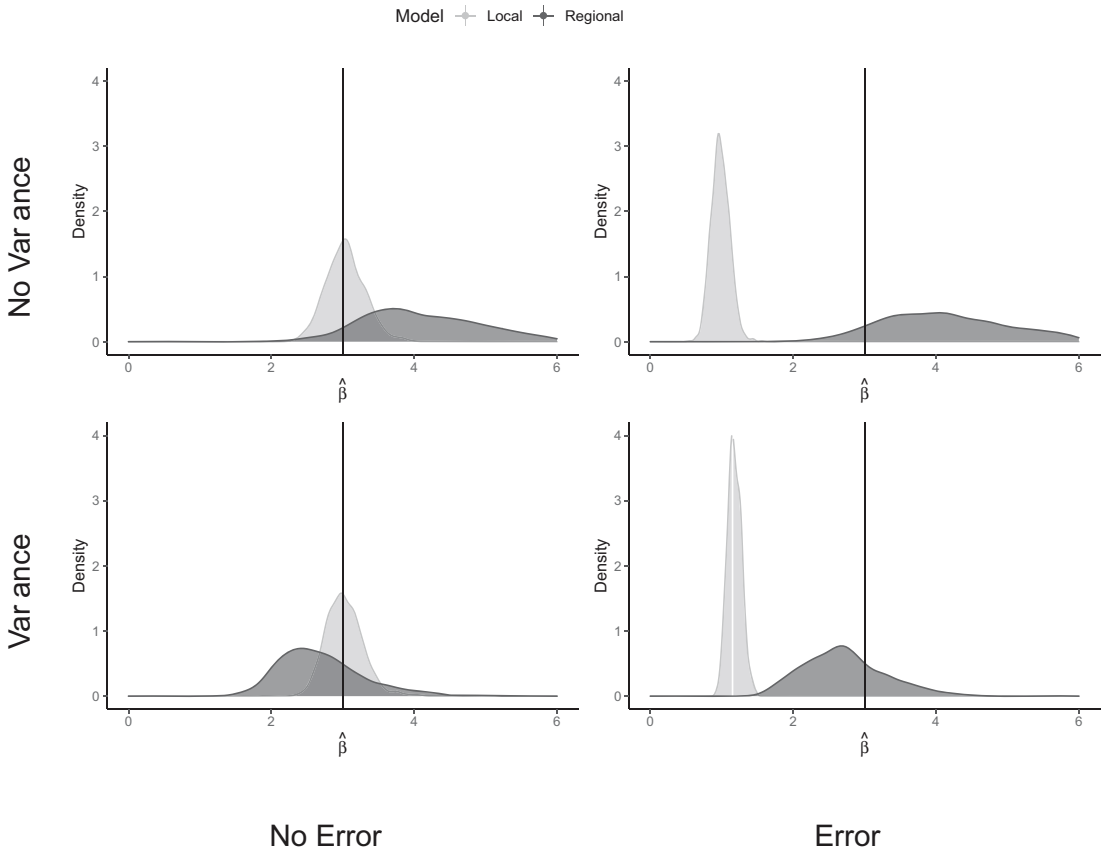


Figure 2. Varying High-Resolution *X*: no true variance, no error variance (upper left panel); true variance, no error variance (lower left panel); no true variance, error variance (upper right panel), true variance, error variance (lower right panel). Vertical black line indicates the true value for β_1. There is no geo-coding error in *Y* in this experiment.

both allows for clearer discussion of the problems confronted and easier extension to geo-coding error in the following section.[8]

The results are presented in Figure 2.[9] Each panel corresponds to a different set of simulated conditions for $\tilde{X}_{L,R}$, where Variance (*vs.* No Variance) indicates whether there is local-level variation in the predictor, and Error (*vs.* No error) indicates whether there is geo-location error in the predictor. Within each panel, there are two densities of the estimates of β_1 for the local model (green) and the regional model (orange) across 1,000 trials, and a vertical black line indicating the true value of β_1 (3) used in the simulations. The results are consistent with our expectations. We observe that the local model does well (is unbiased) when there is no

[8]Moreover, modifying the parameters to reflect alternative conditions (e.g., γ 0.9, δ 0.1) would obviously produce results on the interior of those given here and, therefore, not offer much additional value.

[9]In the Appendix we provide numerical results for the bias, RMSE, and overconfidence. Note that across the 297,000 simulated samples we consider, a small number (fewer than 500 cases) produce perfect prediction in one or more of our models. This is not unexpected given that we are simulating small samples of rare events. However, since these instances are infrequent and it is not the focus of our study, we simply remove these results from our analysis. Should researchers encounter this issue in their observed data analysis, we recommend corrections such as those discussed in Zorn (2005).

measurement error in $\tilde{X}_{L,R}$ (the left two panels), and suffers from traditional attenuation bias when there is measurement error (the right two panels). We also see that true local-level variance in $\tilde{X}_{L,R}$ (the lower two panels) produces expected efficiency gains (as compared to the respective upper panels). In short, the presence of local-level variance in $\tilde{X}_{L,R}$ produces efficiency gains in the local-level model because the variance of the estimator is inversely related to the variance in the covariate.

On the other hand, we see that the regional model suffers from inflationary bias when there is no variance in $\tilde{X}_{L,R}$. While this may seem odd given that this is the scenario where the values of $\tilde{X}_{L,R}$ are constant within the region, it is a consequence of the aggregation of Y_R. In the regional model (with Y_R), there is only one outcome to classify in each region, as opposed to nine varied outcomes in the local-level model. As a consequence, higher values of $\bar{X}_{L,R}$ predict 1 of 1 event perfectly in the regional model, as opposed to, say, 3 of 9 in the local model. Local variance in $X_{L,R}$ offsets this bias by introducing measurement error into the regional model, as all sub-regional variation goes unmodeled.

In mean-square error terms (see Supplementary Appendix), the local model performs better when there is no error in local X (the left two panels), and the regional model does better when there is (the right two panels). Under error in local X, the local models also perform poorly in terms of coverage rates: it never overlaps the true population parameter. This suggests that for applied data, we cannot know with certainty which of these models will perform best; as we do not know the true level of measurement error in X. We demonstrate above that under error in our predictor variables, researchers face a bias-variance trade-off when selecting the spatial scale: higher-resolution models are more efficient and more biased, whereas lower-resolution models are less efficient and less biased. Since neither approach generally dominates, researchers instead need to carefully consider the quality of their covariate data when selecting the spatial scale. With error in covariates, we show that the bias of the estimator is increasing in the spatial resolution of the sample, even when we have perfect information on our outcome.

Geo-Coding Error in Y

Often, however, we cannot be this confident in our event data. We now add measurement error to the local outcome data $\tilde{Y}_{L,R}$. To induce this geo-coding error, we select a sub-sample of the regional areas and then randomly reshuffle the 0 and 1s in its local units (as described in Algorithm 1 and illustrated in Figure 3). That is, the number of events (1s) in the region remains fixed, for example, there is no outcome

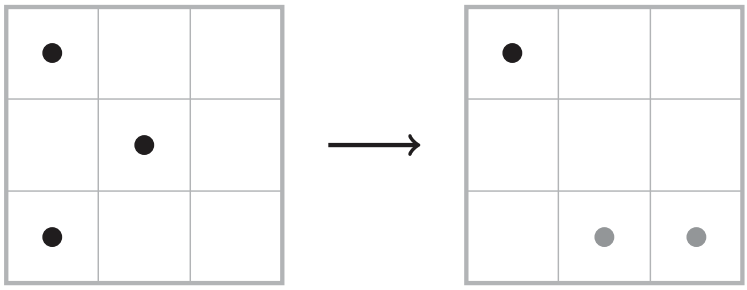


Figure 3. Illustrating geo-coding error. True events in the left grid are re-shuffled and assigned to new locations in the right grid. As this is random, events can be re-located into local areas which truly had an event (the black dot) or events that did not have an event (the light gray dots).

misclassification at the regional level—but the events are randomly reassigned to new local units.[10] This is a relatively benign form of error in event data: it reflects situations in which we are confident that an event occurs in a region, but cannot accurately locate it in the higher-dimensional units (e.g., cities). Across the simulations we allow K to vary from 0 to 100—obviously the more regions re-sampled, the greater the error.

Algorithm 1: Geo-coding error in Y

```
[
  Input: Y_{L,R} as given in Equation (1)
  Output: Ỹ_{L,R}
Randomly select K regions
if region R is selected then
  | re-sample Ỹ_{L,R} from Y_{L,R} in R without replacement      //Re-shuffling stage
else
  | Y_{L,R}
end if
```

The results are presented in Figure 4, where the distribution of $\hat{\beta}$ across the 1,000 simulations is presented on the y-axis, under different levels of geo-coding error (the x-axis). These experiments reveal several notable patterns. First, if there is no local-level variation in X (the top two panels), then the extent of the geo-coding error in Y does not matter. Since the value of the predictor is constant within each region, it does not matter whether an event is located in one local area or another. Instead, the differences that we do observe between the local and regional models are entirely a function of either aggregation or local-level error in X. When there is no

[10] As always, we could induce greater error still: allow for outcome misclassification (ex. underreporting of events), allow for non random geo coding error, etc. While this may better reflect some real world data, it would go beyond the purpose of our analysis.

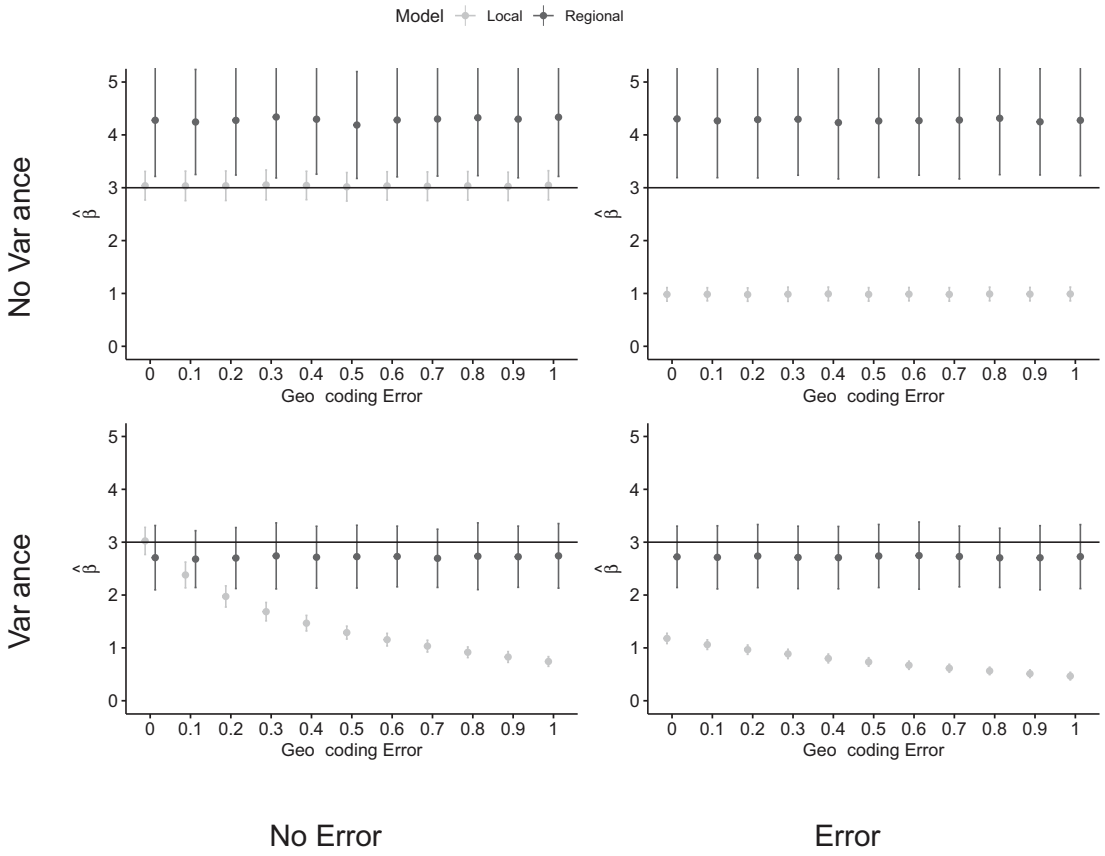


Figure 4. Varying High-Resolution *X*: no true variance, no error variance (upper left panel); true variance, no error variance (lower left panel); no true variance, error variance (upper right panel), true variance, error variance (lower right panel).

error, the local model dominates (it is unbiased and efficient). However, in the presence of local-level error, the performance is dramatically reduced, returning estimates less than one-third of the true value. The regional model, on the other hand, is not affected by local-level error—as it does not use local-level covariates—and instead overestimates the true value, due to aggregation (i.e., ecological) bias under either condition.

Second, when there is local-level variation in *X* (the bottom two panels), geo-coding error in *Y* dramatically affects the performance of the local model. We see that even with no local-level error in *X* (the best case scenario), the performance of the local model degrades rapidly as the extent of the geo-coding error in *Y* increases. If we also add local-level error in *X* (the bottom right panel) the performance of the local model is worse still. Conversely, the regional model does fairly well under these conditions—with the distribution bounding the true value—since it is unaffected by the geo-coding error in *Y*.[11] Moreover, the introduction of the true error

[11]As before, additional quantities of interest from the simulation are reported in the Appendix. These are consistent with the results presented earlier, in the Random Measurement Error in *X* subsection, except the local model now performs worse still given the geolocation error in *Y*. Importantly, under the vast majority of

variance in local X now offsets the inflationary aggregation bias we observed previously since the measurement error in the regional X attenuates the coefficient estimate. To be clear, our argument is not that researches should rely on one form of bias to offset another, as in applied settings we could not be confident that this would bound the true value as it has here. Instead, we aim to demonstrate that only addressing one form of bias (e.g., aggregation bias) does not strictly increase the quality of our results, and can often make them worse.[12]

Illustration: Spatial Patterns of Violence in Afghanistan

To demonstrate that our insights from the simulations above are also relevant when analyzing real-world data, we provide a simple study of violence in Afghanistan. Specifically, we analyze how population density affects the risk of violence. The outcome variable (*Conflict*) is measured using event data from the Uppsala Conflict Data Program's *Geo-referenced Event Dataset* (GED), Version 19.1 (Högbladh 2019; Sundberg and Melander 2013). For our predictor, we rely on the Landscan dataset (Bright et al. 2009), a spatial population dataset with an original resolution of 1 arc-second (∼1 km). To keep our illustration simple, we restrict the sample to 2011, one of the most violent time periods in Afghanistan. According to the GED, there were 630 violent events in Afghanistan during 2011.[13]

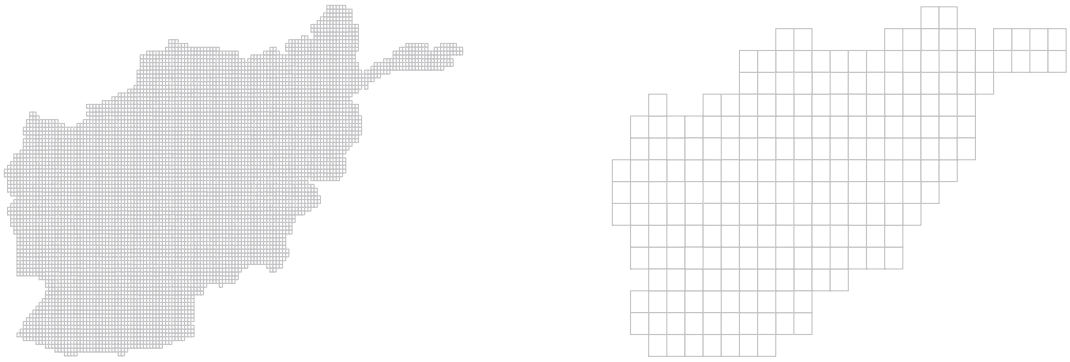
As we are using real-world data, we are not able to manipulate all of the experimental dimensions we explored earlier in our simulation studies. Fortunately, however, we are able to consider the most important ones: (i) the geo-coding error in the outcome, and (ii) the spatial resolution of the analysis. First, to vary the extent of geo-coding error in the outcome, we compare results from a sample with no geo-coding error to those from a sample where the additional error has been intentionally introduced. For our "no geo-coding error" sample, we use only those events from the GED with the highest level of geo-coding precision (1), indicating that the "exact location of the event [is] known and coded" (Högbladh 2019).[14] To generate our sample "with geo-coding error" we exploit the fact that additional

scenarios considered, the coverage rates of the local model are 0, that is, they never recover the true population parameter.

[12]Betz, Cook, and Hollenbach (2020) demonstrate this in a different context, indicating the costs of addressing one type of endogeneity while neglecting others.

[13]As detailed later in the paper, this includes only those events classified as a "1" on the *where_prec* variable indicating that "exact location of the event [is] known and coded" in the GED.

[14]While there may be some uncertainty even in these data (Weidmann 2015), this is the subset we should have the most confidence in. Put differently, if researchers cannot rely on this subset of data, then it suggests high resolution analysis should not even be considered.



(a) Local model, 0.1 degree resolution (b) Regional model, 0.6 degree resolution

Figure 5. Two different grids were used in the Afghanistan analysis.

(simulated) noise can always be added to a given data set.[15] Specifically, we apply jittering to the coordinates of the original set of events, where the original event location is shifted by a uniform random value between ±0.3° both horizontally and vertically.[16] Here we simply focus on this single comparison (no error *vs.* induced error), however, more or less geo-locational uncertainty could be further explored by manipulating the variance used in the jittering.

Second, to vary the spatial resolution used in the analysis of these data, we simply vary the grid size. Keeping with existing approaches in spatial analysis (Tollefsen, Strand, and Buhaug 2012), we create models of different resolutions by dividing up the study area into grid cells of different sizes. Our local model uses cells with a resolution of 0.1°, with a total of N 6,582 cells. The regional model has a cell size of 0.6°, which results in N 221 cells. These two grids are depicted in Figure 5. We aggregate both our predictor and the outcome to the grid cell level in the same way as in our simulations above. The population values for each cell are computed as the average of all LandScan values that fall within a grid cell. The dependent variable *Conflict* is coded as 1 if at least one event from the GED occurs within a cell, and 0 otherwise. Not surprisingly, the distribution of conflict varies considerably between the two resolutions: in the local analysis, about 5% of all cells experienced conflict, while the same applies to about 40% of all cells in the regional dataset.

As in our simulations above, we estimate the model parameters by applying logistic regression to the local and regional samples, using both the original event locations (i.e., no geo-coding error) and the locations with

[15]Adding simulated uncertainty is used in other measurement error methods, such as simulation extrapolation (SIMEX).

[16]In several instances, violent events occur at exactly the same location according to the GED. Shifting each of these independently would lead to major differences in the distribution of events between the jittered and the non jittered data. As such, for these events apply a common jitter to ensure that all events sharing a location in the original sample also share a (new) location in the new sample.

Table 2. Population density and violence in Afghanistan, 2011.

	Dependent variable: Conflict			
	No geo coding error		With geo coding error	
	Local	Regional	Local	Regional
Population (log)	2.719*** (0.124)	2.509*** (0.373)	1.158*** (0.080)	1.987*** (0.313)
Constant	6.979*** (0.234)	3.652*** (0.525)	4.198*** (0.127)	2.623*** (0.420)
Observations	6,582	221	6,582	221
Log likelihood	891.096	110.292	1,332.669	123.519
Akaike Inf. Crit.	1,786.193	224.585	2,669.339	251.038

***$p < 0.01$.

induced geo-coding error. The results from these four models are reported in Table 2. The main conclusion from our simulations above was that in the presence of geo-coding error, the quality of inferences from the local model can degrade quickly. While we do not know the true values here, the results from our applied example exhibit similar patterns. To see this, first compare the estimates for the samples without geo-coding error (the first two results columns of Table 2). Here we observe very little difference between the local and regional models in terms of the coefficient of population, implying roughly the same effect of population density on violence.[17] This changes considerably once we introduce error into the location of our events (right two columns in Table 2). First, there is a greater divergence between the local and regional models' results, as the coefficient on population is considerably larger (roughly 1.7 times) in the regional sample. Second, the results from the regional model (with error) are also more consistent with the results from the samples without error. Specifically, the estimate from the regional model decreases by only 20% due to random geo-coding error, whereas for the local model there is a drop of about 60% as compared to the same model without geo-coding error.

In sum, the results indicate that the level of spatial aggregation selected by the researcher is more important when there is geo-location error in the data. They also suggest that in the presence of geo-location error, a model with a high resolution may perform worse as compared to one with a lower resolution, consistent with the results from our earlier simulation study.

Conclusion

The production of georeferenced, sub-national data on political and social events has transformed social science research. Researchers are no longer forced to analyze processes at highly-aggregated levels due to data

[17]As anticipated earlier, the constant changes noticeably across the two models due to the different base rates: 5% of all cells experience conflict in the local sample, where 40% of all cells experience conflict in the regional dataset.

availability. Now researchers are freed to choose the level of analysis that best suits their theory. However, as we have shown here, the move to disaggregated data analysis is not cost-free. Because the locations (in addition to other characteristics) of events are often measured imperfectly, the quality of inferences can quickly degrade as one move to lower levels of analysis. Specifically, we demonstrate that even when geo-coding errors are random and geographically confined, there is substantial bias in coefficient estimates. Therefore, we recommend that researchers should not pursue disaggregated data analysis as a default strategy.

Instead, we encourage researchers to recognize that the flexibility afforded by events data is more profound than simply analyzing fixed-dimension grid cells. While data-specific variation precludes us from offering a single ready-made strategy, we suggest that researchers consider the following when selecting their spatial scale. First, researchers should carefully consider the spatial scale of the political process *and* the quality of the data when selecting the scale for analysis. For the latter, comparisons to ground truth data (e.g., Weidmann 2015) allow us to gauge the overall magnitude of geo-location accuracy. When researchers have multiple reports or multiple data sets, they can also use this additional information to reduce measurement errors in event data (Cook et al. 2017; Cook and Weidmann 2019). Second, in the absence of a "natural" spatial scale, researchers should evaluate models at different spatial resolutions. Beyond just looking for consistency across levels—which is not guaranteed for even high-quality data—researchers should evaluate cross-resolution differences for evidence of potential biases. For example, do coefficient results attenuate at higher levels of resolution, as expected under aggregation bias, or do we observe some other pattern? Finally, researchers should be conservative about inferences from disaggregated data. As we have shown here, these are not inherently more sound than analysis of aggregate data. When different inferences obtain across these different levels, it is not clear that the micro-level analysis is closer to the truth. Instead, all the evidence should be taken into account when evaluating support for theories.

While beyond the scope of this note, an alternative strategy for researchers to consider is simply analyzing events data using point process models (Johnson et al. 2018; Schutte et al. 2017), as is commonly done with spatial data in other fields (Renner et al. 2015). Zhu, Cook, and Jun (2021) elaborate on the benefits of applying these methods to political science events data. First, researchers are no longer forced to pre-select the spatial scale. Second, because point process models analyze individual events, it is easier to accommodate event-specific measurement errors in locations. Taken together, this would help researchers to address

concerns about aggregation bias and measurement error directly in a unified framework.

Funding

This work was supported by the NSF (DMS 1925119) and German Research Foundation (402127652).

ORCID

Scott J. Cook http://orcid.org/0000 0003 0869 5137
Nils B. Weidmann http://orcid.org/0000 0002 4791 4913

References

Achen, Christopher H., and W. Phillips Shively. 1995. *Cross Level Inference*. University of Chicago Press.

Betz, Timm, Scott J. Cook, and Florian M. Hollenbach. 2020. "Spatial Interdependence and Instrumental Variable Models." *Political Science Research and Methods*. 8(4): 646 661.

Braithwaite, Alex. 2010. "MIDLOC: Introducing the Militarized Interstate Dispute Location Dataset." *Journal of Peace Research* 47 (1): 91 98. doi:10.1177/0022343309350008.

Bright, Eddie A., Phil R. Coleman, Amy L. King, Amy N. Rose, and Marie L. Urban. 2009. "LandScan 2008." Oak Ridge National Laboratory. https://landscan.ornl.gov

Buhaug, Halvard, and Jan Ketil Rød. 2006. "Local Determinants of African Civil Wars, 1970 2001." *Political Geography* 25 (3): 315 335. doi:10.1016/j.polgeo.2006.02.005.

Buhaug, Halvard, Kristian Skrede Gleditsch, Helge Holtermann, Gudrun Østby, and Andreas Forø Tollefsen. 2011. "It's the Local Economy, Stupid! Geographic Wealth Dispersion and Conflict Outbreak Location." *Journal of Conflict Resolution* 55 (5): 814 840. doi:10.1177/0022002711408011.

Carroll, Raymond J., David Ruppert, Leonard A. Stefanski, and Ciprian M. Crainiceanu. 2006. *Measurement Error in Nonlinear Models: A Modern Perspective*. CRC Press.

Cederman, Lars Erik, and Kristian Skrede Gleditsch. 2009. "Introduction to Special Issue on 'Disaggregating Civil War." *Journal of Conflict Resolution* 53 (4): 487 495. doi:10.1177/0022002709336454.

Cook, Scott J., Betsabe Blas, Raymond J. Carroll, and Samiran Sinha. 2017. "Two Wrongs Make a Right: Addressing Underreporting in Binary Data from Multiple Sources." *Political Analysis: An Annual Publication of the Methodology Section of the American Political Science Association* 25 (2): 223 240. doi:10.1017/pan.2016.13.

Cook, Scott J., and Nils B. Weidmann. 2019. "Lost in Aggregation: Improving Event Analysis with Report Level Data." *American Journal of Political Science* 63 (1): 250 264. doi:10.1111/ajps.12398.

Earl, Jennifer, Andrew Martin, John D. McCarthy, and Sarah A. Soule. 2004. "The Use of Newspaper Data in the Study of Collective Action." *Annual Review of Sociology* 30 (1): 65 80. doi:10.1146/annurev.soc.30.012703.110603.

Fjelde, Hanne, and Lisa Hultman. 2014. "Weakening the Enemy: A Disaggregated Study of Violence against Civilians in Africa." *Journal of Conflict Resolution* 58 (7): 1230 1257. doi:10.1177/0022002713492648.

Freeman, John R. 1989. "Systematic Sampling, Temporal Aggregation, and the Study of Political Relationships." *Political Analysis* 1: 61 98. doi:10.1093/pan/1.1.61.

Gelman, Andrew, Boris Shor, Joseph Bafumi, and David Park. 2007. "Rich State, Poor State, Red State, Blue State: What's the Matter with Connecticut?" *Quarterly Journal of Political Science* 2 (4): 345 367. doi:10.1561/100.00006026.

Glynn, Adam N., and Jon Wakefield. 2010. "Ecological Inference in the Social Sciences." *Statistical Methodology* 7 (3): 307 322. doi:10.1016/j.stamet.2009.09.003.

Hammond, Jesse, and Nils B. Weidmann. 2014. "Using Machine Coded Event Data for the Micro Level Study of Political Violence." *Research & Politics* 1 (2): 1 8. doi:10.1177/2053168014539924.

Hegre, Håvard, Gudrun Østby, and Clionadh Raleigh. 2009. "Poverty and Civil War Events: A Disaggregated Study of Liberia." *Journal of Conflict Resolution* 53 (4): 598 623. doi:10.1177/0022002709336459.

Högbladh, Stina. 2019. "UCDP GED Codebook version 19.1." Department of Peace and Conflict Research, Uppsala University. https://ucdp.uu.se/downloads/ged/ged191.pdf.

Johnson, N., A. Hitchman, D. Phan, and L. Smith. 2018. "Self Exciting Point Process Models for Political Conflict Forecasting." *European Journal of Applied Mathematics* 29 (4): 685 707. doi:10.1017/S095679251700033X.

King, Gary. 1997. *A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data*. Princeton University Press.

Klein, Graig R., and Patrick M. Regan. 2018. "Dynamics of Political Protests." *International Organization* 72 (2): 485 521. doi:10.1017/S0020818318000061.

Lee, Sophie J., Howard Liu, and Michael D. Ward. 2019. "Lost in Space: Geolocation in Event Data." *Political Science Research and Methods* 7 (4): 871 888. doi:10.1017/psrm.2018.23.

Nordhaus, William D. 2006. "Geography and Macroeconomics: New Data and New Findings." *Proceedings of the National Academy of Sciences of the United States of America* 103 (10): 3510 3517. doi:10.1073/pnas.0509842103.

Openshaw, Stan. 1984. "The Modifiable Areal Unit Problem." In *Concepts and Techniques in Modern Geography*. Vol. 32. Norwich: Geo Books.

Openshaw, Stan, and Pete Taylor. 1979. "A Million or so Correlations Coefficients: Three Experiments on the Modifiable Areal Unit Problem." In *Statistical Applications in the Spatial Sciences*, edited by N. Wrigley, 127 144. London: Pion.

Pierskalla, Jan H., and Florian M. Hollenbach. 2013. "Technology and Collective Action: The Effect of Cell Phone Coverage on Political Violence in Africa." *American Political Science Review* 107 (2): 207 224. doi:10.1017/S0003055413000075.

Raleigh, Clionadh, and Håvard Hegre. 2009. "Population Size, Concentration and Civil War: A Geographically Disaggregated Analysis." *Political Geography* 28 (4): 224 238. doi:10.1016/j.polgeo.2009.05.007.

Renner, Ian W., Jane Elith, Adrian Baddeley, William Fithian, Trevor Hastie, Steven J. Phillips, Gordana Popovic, and David I. Warton. 2015. "Point Process Models for Presence Only Analysis." *Methods in Ecology and Evolution* 6 (4): 366 379. doi:10.1111/2041 210X.12352.

Robinson, W. S. 1950. "Ecological Correlations and the Behavior of Individuals." *American Sociological Review* 15 (3): 351 357. doi:10.2307/2087176.

Salehyan, Idean. 2015. "Best Practices in the Collection of Conflict Data." *Journal of Peace Research* 52 (1): 105 109. doi:10.1177/0022343314551563.

Salehyan, Idean, Cullen S. Hendrix, Jesse Hamner, Christina Case, Christopher Linebarger, Emily Stull, and Jennifer Williams. 2012. "Social Conflict in Africa: A New Database." *International Interactions* 38 (4): 503 511. doi:10.1080/03050629.2012.697426.

Schrodt, Philip A. 2012. "Precedents, Progress, and Prospects in Political Event Data." *International Interactions* 38 (4): 546 569. doi:10.1080/03050629.2012.697430.

Schutte, Sebastian, Howard Liu, Spencer Dorsey, and Michael D. Ward. 2017. "Finding Duplicates in Machine coded Event Data." Working Paper, University of Konstanz and Duke University.

Shellman, Stephen. 2004. "Measuring the Intensity of Intranational Political Events Data: two Interval like Scales." *International Interactions* 30 (2): 109 141. doi:10.1080/03050620490462603.

Sundberg, Ralph, and Erik Melander. 2013. "Introducing the UCDP Georeferenced Event Dataset." *Journal of Peace Research* 50 (4): 523 532. doi:10.1177/0022343313484347.

Tollefsen, Andreas Forø, Håvard Strand, and Halvard Buhaug. 2012. "PRIO GRID: A Unified Spatial Data Structure." *Journal of Peace Research* 49 (2): 363 374. doi:10.1177/0022343311431287.

von Uexkull, Nina. 2014. "Sustained Drought, Vulnerability and Civil Conflict in Sub Saharan Africa." *Political Geography* 43: 16 26. doi:10.1016/j.polgeo.2014.10.003.

Weidmann, Nils B. 2011. "Violence From Above or From Below? The Role of Ethnicity in Bosnia's Civil War." *The Journal of Politics* 73 (4): 1178 1190. doi:10.1017/S0022381611000831.

Weidmann, Nils B. 2015. "On the Accuracy of Media Based Conflict Event Data." *Journal of Conflict Resolution* 59 (6): 1129 1149. doi:10.1177/0022002714530431.

Weidmann, Nils B. 2016. "A Closer Look at Reporting Bias in Conflict Event Data." *American Journal of Political Science* 60 (1): 206 218. doi:10.1111/ajps.12196.

Weidmann, Nils B., and Espen Geelmuyden Rød. 2019. *The Internet and Political Protest in Autocracies*. New York, NY: Oxford University Press.

Weidmann, Nils B., and Michael D. Ward. 2010. "Predicting Conflict in Space and Time." *Journal of Conflict Resolution* 54 (6): 883 901. doi:10.1177/0022002710371669.

Zhu, Lin, Scott Cook, and Mikyoung Jun. 2021. "Point Process Models for Political Events." arXiv preprint arXiv:2108.12566. doi:10.48550/arXiv.2108.12566.

Zorn, Christopher. 2005. "A Solution to Separation in Binary Response Models." *Political Analysis* 13 (2): 157 170. doi:10.1093/pan/mpi009.