Communication technology and reports on political violence

Cross-national evidence using African events data

June 22, 2016

Abstract

The spread of internet and mobile phone access around the world has implications for both the processes of contentious politics and subsequent reporting of protest, terrorism, and war. In this paper, we explore whether political violent events that occur close to modern communication networks are systematically better reported than others. Our analysis approximates information availability by the level of detail provided about the date of each political violent event in Africa 2008-2010 and find that while access to communication technology improve reporting, the size of the effect is very small. Additional investigation find that the effect can be attributed to the ability of journalists to access more diverse primary sources in remote areas due to increased local access to modern communication technology.

1 Introduction

Recent technological and methodological innovations offer improved access to information from conflict zones that previously was hidden from view. The ability to collect and analyze event data provide contemporary scholars' with opportunities to explore micro-level mechanisms of repression, mobilization, and strategies of violence (Gleditsch, Metternich and Ruggeri 2014).

Yet, we know little about possible bias in the data provided by projects such as the Uppsala Conflict Data Program (UCDP), Armed Conflict Location & Event Data Project (ACLED), or the Political Instability Task Force Worldwide Atrocities Dataset. In contrast to the extensive literature on bias in *newspaper*-sourced data (Galtung and Ruge 1965; Snyder and Kelly 1977; Franzosi 1987; Woolley 2000; Fleeson 2003; Earl et al. 2004), there have been few efforts to explore the quality of "Big Data" in the internet age (notable exceptions include Price and Ball (2014), Weidmann (2015) and Weidmann (2016)).

In this paper, we investigate whether access to communication technology can account for spatial variation with regards to the quality in conflict data. Drawing on the media studies literature (Fenton 2010; Domingo and Paterson 2011), we expect journalists that directly can be in contact with primary sources through internet and mobile phones will be able to provide more detailed reports about political violence. Considering such *details* are essential for data collection projects to identify perpetrators, severity, and targets for political violence (Kreutz 2015*b*), we contend that even a marginal improvement in quality may substantially influence information used in much contemporary conflict scholarship. In particular, our study may be important for the growing interest in whether modern communication technology assists organized crime, terrorism,

or insurgency (Andreas 2002; Weimann 2006; Pierskalla and Hollenbach 2013; Shapiro and Weidmann 2015). If, as we expect, information about violence is better reported in areas with developed communication structures, then we cannot know whether technological advancement actually does increase violence or if such correlations are spurious.

Empirically, we focus on the quality of reporting about political violence in Africa 2008-2010. There are three reasons for this. First, Africa is the region which, together with Asia, has experienced the most armed conflicts in the post-Cold War era.¹ Second, Africa is becoming known as "the mobile continent" due to its embrace of digital media over (previously underdeveloped) infrastructure suggesting that communication technology may be particularly important in this region (Hersman 2013). Third, most published research on spatial variation in armed conflict is focusing on Africa, as the early version of UCDP Georeferenced Conflict Event Data (UCDP-GED)(Sundberg and Melander 2013) and projects such as ACLED (Raleigh et al. 2010) and Social Conflict in Africa Database (SCAD) (Salehyan et al. 2012) primarily provide data from this continent.

This paper differs from earlier work on event data quality as we are neither comparing information from different datasets (Restrepo, Spagat and Vargas 2006; Eck 2012) nor data collected with competing methodologies (Davenport and Ball 2002; Price and Ball 2014; Weidmann 2015). Instead, we use the *precision scores* assigned to each event in the UCDP-GED (Sundberg and Melander 2013) which indicates the level of detail of available information. This measure is not produced following some estimation technique but represents the specific information in the coded material about when and where an event occurs. We focus on the temporal precision, approximating that events with information

¹In 1990-2013, there were 372 conflict-years in Asia, 308 in Africa, 115 in the Middle East, 67 in the Americas and 62 in Europe (Themnér and Wallensteen 2014).

about the specific date are better reported than those reported as only within a given week, month, or year.

The next section outline how new communication technology should facilitate more detailed reporting on political violence events, before we describe our research design. Following an analysis of 2,369 events in Africa 2008-2010, we find a statistically significant and robust correlation between reporting quality and access to communication technology although the size of the effect is relatively small. We then extend the analysis by exploring the original source that offer information about political violence and find that contemporary reporting is less dependent on official statements and instead rely on eyewitness accounts more than in the pre-internet era. The final section concludes and discusses the implications of these findings for future scholarship.

2 Spatial bias in political violence data

Existing research on reporting contentious politics has identified two sources of bias. The first, which is the focus for this article, relates to the ability of media to access information about a given event while the second relates to the deliberate-strategic selection of which events are reported and how these are described (Galtung and Ruge 1965; Earl et al. 2004). The spatial location may influence both the ability and the willingness to report about a particular event.

Most of the information provided by international media from conflict zones is collected by news bureaus with limited resources. This means that events that occur closer to major political centers are likely to receive more coverage simply because reporters have better access to witnesses which should improve

reporting both in terms of output and quality (Fleeson 2003; Weidmann 2015). This differ from "distant sources (.. who ..) are less able to navigate the local terrain (physically but also politically and socially). Outsiders are less able to identify events, less able to understand who the combatants are, and less able to know where the best informants can be found. Distant sources may find themselves relying on the ones most readily available but farthest from the events of interest" (Davenport 2010, p. 70).

It has also been argued that access to information should be influenced by government censorship and other restrictions on free movement although existing empirical evidence about this factor so far have been inconclusive. On the one hand, studies show that terrorism is probably underreported in countries with limited press freedom (Drakos and Gofas 2006) and threats and violence on journalists reduced the coverage of human rights abuses in Guatemala (Davenport and Ball 2002). On the other hand, other findings indicate that dangerous security environments in general does not reduce news coverage (Urlacher 2009) and media in both Mexico and Uganda has refused to bow to government intimidation (Lawson and Lawson 2002; Ocitti 2005).

The final factor that determines media content is decisions by news editors about what the audience is likely to be interested in. This is partly influenced by the nature of the story, where violent and unexpected developments usually are preferred, but also by the location of the event. The threshold for what is considered newsworthy increases with distance, meaning that minor protests close to the publication outlet may be given as much attention as exceptionally dramatic events far away (Smith et al. 2001; Myers and Caniglia 2004).

2.1 New technology brings new news reporting

Communication scholars have suggested that the development of modern communication technology has fundamentally changed the nature of news media (Severin and Tankard 2010). What is important to remember, though, is that the "news media" is not a unified and coherent entity with consistent output across time and space, but a diverse set of actors and practices sensitive to competition and technological change (Pavlik 2000; Fleeson 2003; Mitchelstein and Boczkowski 2009). As in any competitive market, one of the most influential instigators of change are innovations that facilitates high quality news gathering to a lower cost, such as the introduction of new communication technology. This influence both the means of news gathering (the input of information) and the means of publishing (the output). Therefore, as shown in Figure 1, it is not surprising that the amount of reports of violence does not perfectly correlate with the actual fluctuations of violent events.

Journalistic practices has undergone a substantial shift following the development of internet and mobile phone networks. Using modern communication technology, reporters can now faster and easier gather information through *direct* contact with witnesses rather than having to physically travel to the location of the event after-the-fact. While this has influenced journalists everywhere, the impact of new technologies on reporting has been particularly profound in areas where access to information previously was restricted and difficult, such as in states characterized by lower economic development where also political violence is more likely. Anecdotal evidence from Zambia and South Africa suggest that internet access provide ordinary people with new channels to improve communication with centers of power, including the mainstream media (Spitulnik 2002; Goldfain and Van der Merwe 2006).

Figure 1: Newswire articles on political violence in Africa (1989-2010) compared to total number of UCDP GED events. The 1999 increase in number of articles is partly due to the inclusion of AFP reporting for Africa in Factiva.

New technology also offer journalists access to new sources of information, as outlets such as twitter, youtube, wikis, and blogs provide opportunities for sources to anonymously provide documentation about events. This approach has, for example, been extensively used by civilians reporting atrocities by criminal gangs and government agents in Mexico in recent years (Kirchner 2014).

Increasing globalization and the spread of the internet has not only influenced the ways that reporters collect information, it has also had a substantial impact

on the process of publishing. The previous practice where stories were sold to and published by set-format media (newspapers, radio, and television), has been superseded in the era of internet publishing by outlets without space constraints (Domingo and Paterson 2011). This has removed one of the most influential sources for systematic bias on whether political violence is reported, as the role of the news editor as a "gatekeeper" has been reduced (Schudson 1989).

Indeed, news agencies in the internet age are no longer forced to exclude reports but on the contrary encouraged to provide more output. In the contemporary news cycle, news bureaus compete about being the first to offer "breaking stories" and journalists are expected to provide multiple versions of the same story where the updates add details when these become available. This has led to an increased use of the internet for information gathering from, for example, tweets, blogs, and social media as this may provide more unique details than official press conferences (Farhi 2009).

We contend that the combined effect of all these different effects from the development of new communication technology has created variation in the quality of information available about political violence events. Reporting will be substantively better in areas where journalists easily can seek out information through internet and mobile phone networks.

3 Empirical investigation

Figure 2 visualizes the data we employ for our empirical analysis. It is worth noting that the use of modern communication technology in Africa is rarely limited by individuals' *ownership* of computers or mobile phones. In addition to

Figure 2: Map showing internet access, UCDP GED data and road distances

commercial options for getting online, studies have shown that mobile phones and computers often are shared among members in the local community (Atton and Mabweazara 2011).

In this paper, we use information from events of all different types of violence covered by UCDP-GED (Sundberg and Melander 2013). This means that we are exploring the reporting of events regardless of whether these constitute part of an armed conflict between states and/or rebels (Gleditsch et al. 2002), non-state conflict (including communal violence) (Sundberg, Eck and Kreutz 2012), or one-sided violence against civilians (Eck and Hultman 2007).² Since we are interested in the spatial variation in reporting quality, we need to focus on events for which the location is confidently reported. Thus, our analysis is restricted to the observations where we know that the report contains sufficient information to locate the event confidently at an exact town/village or within a 25 km radius from the exact location.

3.1 Dependent variable

The dependent variable for our analysis consist of a previously underutilized facet of the UCDP-GED, namely the precision score given to the quality of information provided about each event. The coding of this score is straightforward and directly based on the actual information provided in the news material. Table 1 summarizes the criteria for coding precision scores (Sundberg, Lindgren and Padskocimaite 2011).

For our analysis, we recode the summary temporal precision score as 6, giving

²We include events from conflicts below the aggregate 25 deaths/year threshold, and "unclear" armed conflicts where the incompatibility criteria is loosened (see Kreutz (2015b) for the benefits of this).

Table 1: UCDP Precision scores

Temporal	Information
0	Summary event
1	Exact day of the event known
2	Event can be located within a 2-6 day period
3	Event can be located within a given week
4	Event can be located within a given month
5	Event can be located within a given year
Spatial	Information
1	Exact location of the event is known
2	Event occurred within a ca. 25 km radius around a known point
3	Event occurred in a given second order administrative division
4	Event occurred in a given first order administrative division
5	Spatial reference for the event is a linear/polygon reference point
6	Event occurred within a given country
7	Event occurred in international water or airspace

us a scale with 1 as the most detailed information and 6 as the least specific. The information behind these scores comes from the following process. Every year, UCDP extract and collect information from a large amount of news media content, including (for Africa) outlets such as *Africa Confidential* and the *African Research Bulletin*, as well as reports from international and national NGOs and other sources. However, many NGO investigations uses the work of locally based journalists. For example, the sources used for the annual human rights reports by the US State Department and Amnesty International are composed of a combination of stories reported in local media and on-site investigations (Kreutz 2015*a*).

For each political violent event coded into the UCDP-GED dataset, coders assign precision scores that reflect on the level of detail in the reports about where (`where_precision`) and when (`date_precision`) the event occurred. If there are multiple reports about the same event, UCDP always use the most detailed and disaggregated information meaning that "poor" confidence scores should only be assigned for events where detailed reports are lacking.

Figure 3: Proportions of events' spatial and temporal precision

Thus, our dependent variable is the confidence score for the temporal precision of the event. We consider reports on when an event occurred constitute a cross-national comparable "hard fact" that we don't expect to be sensitive to political or editorial pressures that otherwise may influence the narrative of an event (Davenport 2010). Figure 3 show the correlation matrix between spatial and temporal precision in our data, indicating substantial variation for the dependent variable in our sample.

3.2 Independent variable: Internet access

Internet access is determined by the local geography and the distance between an eyewitness and the nearest internet node. For this, we use the Maxmind GeoIP database (the version released on December 1, 2010) which constitute a global dataset assigning geographical information to every known internet (IPv4) address in use³. This data is typically used by web-related industries for customising or restricting content and advertising in various geographic areas.

The spatial resolution of the data is the city, while the best data point coarseness claimed is the individual IP address. Independent studies of the accuracy of IP geolocation databases has indicated a 40%-60% accuracy rate in matching individual locations with an area (1:1 matching) within 100 km from the actual location of the assigned IP address. In Africa, Maxmind claims an accuracy of between 38% and 89% for 1:1 matching (MaxMind 2013; Shavitt and Zilberman 2011; Poesse et al. 2011). We don't consider this seemingly low reliability a major concern because of the extremely demanding requirements of such tests, which are modeled on the typical commercial usage, i.e. the ability to precisely identify the exact location of a random, individual IP address. Since we are interested in the internet point-of-presence (i.e. the location of internet access), which is a much coarser measure (approximately 4 orders of magnitude) than the individual IP address, we assume that aggregation mitigates most identified 1:1 errors⁴.

³Maxmind accounts for 3,525,991,153 individual IPv4 addresses out of a maximum possible number of 3,706,452,992 (IANA 2013; ICANN 2011). A more in-depth discussion of the Maxmind dataset is presented in the web appendix.

⁴For robustness tests, we retain the number of identified internet hosts in a single location as a measure of Internet pervasiveness. Another concern is the unknown probability that the dataset fails to identify Internet points-of-presence altogether (i.e. not assigning even one location to such points). While this cannot be determined due to lack of "real world" data outside extremely small survey-based samples in the developed world (Shavitt and Zilberman 2011; Poesse et al. 2011), we estimate that this probability is extremely small, as the active detection techniques employed for gathering the data have a failure function that is inversely

3.3 Calculating distances

To link the location of a political violence event with internet access, we measure the distance between event and internet nodes in two ways. The first is the great circle distance (geodesic distance) calculated using `PostGis 2.0.1` on the WGS84 spheroid and expressed in kilometers (i.e. the shortest possible straight-line route between event and internet access point), while the second is the shortest possible road distance between event and the closest internet node.⁵ The two measures differ substantially, with different closest points of internet access for more than 20% of events in our sample (483 out of 2369).

To calculate road distances, we use `gRoads dataset` version 1 (CIESIN-ITOS-NASA SEDAC 2013), an open-source global road-network dataset. Distances between events and internet nodes are calculated with Dijkstra’s algorithm using `pgRouting 2.0`(`pgRouting Project` 2013) with a tolerance level of 0.01 decimal degrees (approximately 0.8-1.2 km, depending on latitude and longitude). This tolerance level is on the same magnitude as twice the stated standard error of the `gRoads` dataset (i.e. at least 2 times 300 m) to avoid misspecification due to potential `gRoads` coding errors.⁶ For points not located on a road, the nearest road was used as a starting point and the distance to that road added to the calculation. Further, distances were not calculated for events located more than 50 km away from any road (excluding less than 5% of total events).

The `gRoads` data also provide information on the quality of the individual roads,

proportional with the density of active internet connections (Shavitt and Zilberman 2011; Poese et al. 2011) and Africa (the area under study) has by far the lowest internet penetration figures in the world (Kim 2010).

⁵International borders were not taken into consideration given the porous nature of and significant interaction across national boundaries in Africa.

⁶Given a stated error of 300 m, two roads intersecting in real life may be displayed in the dataset as being at most 600 meters apart. Further, contiguous segments of road in real life may not be displayed as contiguous in the dataset, especially at ”breaking points” for data sources such as borders.

which is useful for our purpose to measure individuals access to the internet. We impose a penalty on roads classified as "trails" where we expect travelling speed to be ten times slower than on proper, even poor-quality, roads.⁷

As distance calculations on a dataset as large as gRoads are computationally intensive, we identified potential closest nodes candidates through a sliding window approach with an expanding sub-setting buffer around each data point. The buffer grew by a radius of 1 decimal degree at a time, stopping when 5 suitable internet nodes (to which distances could be calculated) were identified. For analysis purposes, the decimal logarithm was taken from all distances, as we expect the effect follows a logarithmic function rather than a linear one.

3.4 Statistical technique

We model the relationship between the distance to internet access and quality of information about political violence events as a proportional odds ordinal logistic regression (Long and Cheng 2004; Fullerton 2009). The probability of the temporal precision confidence score being a value m , with $1 \leq m \leq 6$ is estimated as follows:

$$Pr(y = m|x) = \begin{cases} cdf_{logistic}(\tau_1 - x\beta), m = 1 \\ cdf_{logistic}(\tau_m - x\beta) - cdf_{logistic}(\tau_{m-1} - x\beta), 1 < m < 6 \\ 1 - cdf_{logistic}(\tau_{m-1} - x\beta), m = 6 \end{cases}$$

⁷The 10x penalty relationship approximate a general walking speed for humans at around 5-6 km/h while a car would on average travel at 50-60 km/h on a non-tarmac road. Our findings are robust using the road data without surface quality specification.

where x is the covariate vector, β is the associated coefficient vector for the covariates, τ is the unknown cutoff point between precision scores and $cdf_{logistic}$ is the cumulative logistic density function (Long and Cheng 2004; Fullerton 2009). As we assume a single process determining the probabilities, the coefficient vector does not vary across the 6 equations, producing proportional slopes (Fullerton 2009).

Since we only have one data-point for Internet access locations, we subset the UCDP GED to only include data for the 2008-2010 period, treating it as fully cross-sectional data.⁸

3.5 Controls

A consistent finding in existing literature on media selection bias is that more violent events are given more attention (Price and Ball 2014). We therefore include a variable indicating the total annual intensity of the specific armed conflict, non-state conflict, one-sided violence interaction (or *dyad*) that the event belong to as well as the fatality estimate for the specific event.

We are also interested in whether internet access overlap with other forms of modern communication technology, including mobile phones, which feature more prominently in existing research (Dafoe and Lyall 2015).⁹ The data on mobile phone coverage is obtained from a high-quality print map produced by the GSM Association and Europa Technologies in January 2009 (GSM Association and Europa Technologies 2009), extracted through both GIS specific digitiza-

⁸Our findings are robust for the use of only 2010 UCDP-GED data. Further, Maxmind data is slow-changing in nature. Comparing the Maxmind data version we use with the version released on September 10, 2013 (almost three years later) indicate less than 0.975% change in locations coded.

⁹For individuals to access the internet with their mobile phones, they must obviously be close to an internet access point.

tion and vectorization techniques (zones of coverage and lack of coverage) as well as a support vector machine based algorithm. The support vector machine was used for categorization of pixels in buckets corresponding to coverage and lack of coverage.¹⁰

Our dependent variable, temporal precision scores, exhibit a small degree of geographic auto-correlation with a clustering tendency (Moran’s I of 0.054***¹¹), motivating the inclusion of a simple spatio-temporal lagged term consisting of the number of previously reported fatalities from events in the past 7 days within a 25 km radius.¹² To control for local economic development, we include information on local domestic product (regional GDP)(Nordhaus 2006), collected on a 1 degree by 1 degree cell (extracted from PrioGrid v.1.01(Tollefsen, Strand and Buhaug 2012)). We also control for country-level media censorship using the annual Freedom House freedom of the press score (FH 2012).

Finally, to control for the possibility that communication technology simply is a proxy for urban areas, we measure geographic features in two ways. The first is the distance in minutes to the nearest location with 50,000 inhabitants or more, using data provided by the European Commission(Nelson 2008), and the second is the proportion of mountainous terrain in a 0.5 by 0.5 degrees cell where the political violence occur (Tollefsen, Strand and Buhaug 2012). Not

¹⁰The model was trained on both the pixel itself and neighboring pixels, and both the unprocessed map and the processed data are available with our replication material.

¹¹Moran’s I indicates the level of global spatial dependency of a variable - i.e. the tendency of values of a point to be correlated with values situated nearby. Moran’s I can take values on a scale of -1 to +1, with 0 indicating no spatial correlation (random disposition) and ± 1 indicating perfect negative respectively positive correlation(Tiefelsdorf 2006).

¹²We choose 25 km based on the UCDP definitions for precision scores, but our results are robust for the use of 50km and 30 days, as well as for an alternative specification consisting of the number of events inside the same spatio-temporal window. We also explored the inclusion of a thin plate smoothing spline (Wood 2003; Zhukov 2012) or dynamic spatial ordered models (Wang and Kockelman 2009). However, they proved to be difficult to adapt to the event as the unit of analysis rather than to the typical spatial location (i.e. village, area, grid-cell, administrative unit) as very frequently multiple events, with different precision scores, share a single location, leading to a problem of under-fitting the models.

surprisingly, we find a strong negative correlation between urbanization and mountainous terrain so to avoid multicollinearity, we include these variables in different estimations.¹³

4 Results

Our expectation is that better access to communication technology correlates with more detailed reports of political violence. The dependent variable in all models in Table 2 is the quality in reporting the temporal location of an event, with 1 being the best and 6 being the worst. The explanatory variable (distance to closest internet node) is measured as road distance in Models 1-5 and as geodesic distance in Models 6-10.

Across all models we find that that quality of information, i.e. the precision about events, decreases with distance from internet nodes in line with our expectations. Results are similar regardless of how we calculate distance and consistently statistically significant on at least 95 % confidence level. One benefit of the ordered logit is the possibility to interpret information about whether the correlation is statistically significant only in some part of the scale (i.e. potentially the best or worst reported events). We find, however, that the distance to internet node his statistically significant for each single step. Our findings are robust when controlling for the severity of violence, both measured on a yearly basis and for the specific event, the local level of preceding violence, urbanisation, mountainous terrain, local economic development, and press freedom.

In Models 4 and 9, we include the dichotomous measure of mobile phone cover-

¹³Our findings are robust for the use of a variable of local (spot) population density, see web appendix.

Table 2: Quality of reporting and internet access

	<i>DV: Temporal precision</i>									
	Model									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Internet node (road)	0.293*** (0.082)	0.255*** (0.085)	0.430*** (0.113)	0.427*** (0.113)	0.281*** (0.089)					
Internet node (geodesic)						0.246*** (0.088)	0.210** (0.091)	0.343*** (0.120)	0.344*** (0.120)	0.246*** (0.095)
Dyad severity (total)	0.460*** (0.111)	0.475*** (0.113)	0.471*** (0.116)	0.453*** (0.117)	0.426*** (0.117)	0.490*** (0.111)	0.501*** (0.113)	0.495*** (0.116)	0.474*** (0.117)	0.443*** (0.117)
Event severity		0.019*** (0.002)	0.018*** (0.002)	0.018*** (0.002)	0.019*** (0.002)		0.019*** (0.002)	0.019*** (0.002)	0.019*** (0.002)	0.019*** (0.002)
Mobile phone coverage				0.155 (0.127)					0.164 (0.127)	
Severity prior week		0.002 (0.002)	0.001 (0.002)	0.002 (0.002)	0.002 (0.002)		0.001 (0.002)	0.001 (0.002)	0.001 (0.002)	0.002 (0.002)
Local GDP			-0.043 (0.073)	-0.045 (0.073)	0.040 (0.071)			-0.047 (0.073)	-0.050 (0.073)	0.031 (0.071)
Press censorship			0.003 (0.006)	0.000 (0.006)	-0.003 (0.006)			0.004 (0.006)	0.001 (0.006)	-0.002 (0.006)
Distance to city			-0.001** (0.000)	-0.001** (0.000)				-0.001* (0.000)	-0.001 (0.000)	
Mountains					0.640*** (0.158)					0.649*** (0.159)
Observations	2,118	2,118	2,111	2,111	2,111	2,118	2,118	2,111	2,111	2,111

*p<0.1; **p<0.05; ***p<0.01

Note:

age and find that the internet distance remains statistically significant. However, a separate regression (see web appendix) where we replace internet information with mobile phone coverage also correlate with better reporting, suggesting that our finding indeed show the effect of the communication process rather than the particular *means* used.

While our study identify a robust statistically significant correlation, the size of the effect is relatively small. To estimate the size of the effect, we build on Model 4 in Table 1 and run 1000 simulations for each 0.1 increase of logged road distance between 1.0 (10 km, the cut-off point in the data) and 3.2 (approx. 1585 km, close to the maximum observed value in the data), giving us a total of 32000 simulations¹⁴. The dyad severity is set to low (the most common observation type), with all other values in the model held at their observed means.¹⁵

¹⁴Simulations were performed using the Zelig(Imai, King and Lau 2008) R package.

¹⁵Our findings are robust for excluding "summary" events from the analysis, and regardless of which "precision"-step we choose for the post-estimation we find similar results.

Figure 4: Predicted probabilities based on road distance

Figure 4 show the simulated predicted probabilities of obtaining the best (single-day specified) and worst (summary event) precision confidence scores as a function of road distance to the closest internet node. The blue (top) lines indicate the predicted probability that a given event is coded with the best temporal confidence, while the red lines show the predicted probability of the event given the least detailed precision. The reason that the predicted probability is much higher for getting the "best" precision is because our data consist of already coded and scrutinized events rather than all news articles. This means that our

findings should be interpreted in light of the knowledge that even the "worst" reported data is still reports deemed sufficiently reliable to be coded into UCDP-GED

For events that occur at 10 road-kilometers from an internet node, the predicted probability that reports identify the day of the event (highest precision) is 0.774 (CI: 0.735; 0.811). However, for events 100 road-kms away from an internet node, the predicted probability of such detail reporting decreases to 0.727 (CI: 0.699; 0.753). For events with the least precision, we identify the opposite trend as distance increases from the internet nodes. The predicted probability of an event being reported in a summary (lowest possible precision) is 0.059 (CI: 0.029; 0.099) close to internet nodes but increases to 0.74 (CI: 0.041; 0.111) at 100 kms distance.¹⁶

Turning to the control variables, some findings warrant discussion. First, there has - to our knowledge - not before been any systematic studies whether violence in more urban areas actually *is* better reported than in the countryside. There are claims of an consistent "urban bias" in identifying instances of political violence (Kalyvas 2004) although it has also been pointed out that insurgent activity in cities may be difficult to parse out from surrounding noise (Staniland 2010). Our study cover more forms of political violence than civil strife, but the findings in Table 2 provide mixes support regarding the effect of urbanisation. Violence closer to major cities is reported with lower precision, but this is not consistently statistically significant.

The second notable finding is with regards to severity of violence and reporting

¹⁶Results are similar for geodesic distances: the probability of a "single-day" event based on Model 9 decreases from 0.763 (CI: 0.721; 0.800) at 10 km distance to 0.723 (CI: 0.691; 0.752) at 100 km. The probability of a summary event increases from 0.063 (CI 0.030; 0.105) to 0.074 (CI: 0.042 - 0.112).

quality; a factor regularly argued as making events more newsworthy and hence better reported (Galtung and Ruge 1965; Price and Ball 2014). In both of our Tables, however, we find the opposite relationship - the precision of reporting decreases for more violent events as well as for conflicts where the overall interaction is more violent. We suspect that this may be caused by our focus solely on lethal violence, in contrast to much of the literature on newsworthiness that focus on the size of protests (Oliver and Myers 1999; Smith et al. 2001; Earl et al. 2004; Herkenrath and Knoll 2011).

5 Is communication technology the reason?

Our statistical analysis find a small but statistically significant spatial variation regarding the quality of reporting of political violence, and that this correlates with distance to internet access. Whether this variation can be explained by the suggested mechanism of better information provision through modern communication technology, we now take a closer, qualitative, look at which sources are attributed to the information in the actual reports.

We revisited the background text of the UCDP GED events and coded the collected information about original sources. To systematize this data, we group the sources into four broad categories. First, we refer to "official sources" when the original source was the government (e.g. military spokesperson, police, minister, local administration etc.) or a dissident organization (e.g. rebel group or a media outlet controlled by a rebel group); second, "journalists" are reporters with unclear, neutral or unknown allegiance (e.g. a national, private TV or radio station; a Reuters correspondent etc.); third, "other" sources include international organizations, NGOs, or foreign governments; and, finally, "eyewitnesses"

(for example a local bystander).

We basically expect a greater risk of political bias when media reports are based on "official sources" while the use of "eyewitnesses" should improve the quality of reporting. To see if there has been a change over time that can be attributed to improved communication technology, we combine this information from 2008-2010 with UCDP GED events from 1992-1993. In this period, access to the world wide web was basically nonexistent in Africa (or, for that matter, in most of the world). An additional advantage for our purposes is that the event data covering 1992-93 was collected by UCDP-GED during 2008-2010, meaning the use of the same human coders, definitions, sources and methodology which means that inter-coder reliability issues are unlikely to affect the comparison.

Figure 5: Original sources for reporting in the 1992-93 sample and the 2008-10 sample

Figure 5 shows the distribution of original sources for reports on political violence 1993-94 and 2008-10. In the earlier period, the vast majority of events (67.9%) were reported by "official sources" directly linked to the belligerent parties. This contrast with the paucity of information collected from eyewitnesses or locals, which only contribute to 16.4% of reports. In the post-internet time period, we find a telling difference. In 2008-2010, the number of reports originating with eye witnesses is almost equal to that originating from official sources (41.4% vs. 41.7%). This finding is consistent with a common claim with

regards to the spread of communication technology across Africa: that it will offer opportunities for a wider range of citizens to provide information about local conditions (Spitulnik 2002; Ocitti 2005; Mudhai, Tettey and Banda 2009; Aker and Mbiti 2010). We find a similar trend towards more detailed reporting over time in the UCDP GED dataset overall as an increasing proportion of events are coded with higher precision scores. In 1989, only 57.4% of events are attributed to an individual day, while this was possible for 75.14% events in 2010.¹⁷

If the spread of communication technology over time leads to data improvements, then we should expect a similar variation with regards to different types of original sources and data quality also within the modern data. We find this is the case. The data points situated near internet nodes are almost exclusively reported using two types of primary sources. The first is extremely brief official notes and communiques from actors involved in the violence such as for example the military, the police, or rebel groups. The second, though, consist of more detailed, highly descriptive narratives that provide in depth insights regarding the actions by different actors and the temporal ordering of the violence. Much of this latter type of information (over 50% in areas under 150 km) is reported by sources identified as residents, protest participants, interviewees, local journalists writing opinion pieces, local community leaders, anonymous officials interviewed directly by the media and even blogs, i.e. informal, mostly independent organizations and individuals.

As distance increases from internet nodes, these types of in depth narratives about specific events become less common, and the original sources for information are almost exclusively spokespersons of warring organizations, official

¹⁷For the full trend of the relative distribution in precision scores 1989-2010, see web appendix.

communiques, police/army officials etc., i.e. actors directly involved in violence. When local narratives disappear, these "official" versions dominate available information. As a consequence of this lack of information from local sources, the quality of information decreases including the ability to code such "hard facts" as the day of a given event.

6 Conclusion

Our study show that there is variation in how well political violence is reported across space. Events that occur in areas where journalists with ease can receive information is better reported than events in the periphery. As modern communication technology has spread across the world, reporters are now able to easily access information directly from eyewitnesses and locals rather than rely solely on governmental press briefings. This means that media-based publications now provide richer, more detailed, narratives of events which offers a better understanding of the processes of political violence.

What implications do our findings have for interpreting existing scholarship or the design of new research projects? First, the heterogeneous nature of political violence data should be taken under serious consideration for analyses of event data with a long time span, as the quality of information is markedly different in the "before internet" and "after (with) internet" time periods. We therefore advise researchers to proceed with caution when using longitudinal samples of event data and to always account for temporal dependency in analyses. Our findings also suggest that studies exploring whether internet connections or mobile phone networks facilitate violence need to acknowledge the possibility of selection bias.

Second, our investigation also provide some good news for the emerging field of cross-national micro-level studies of political violence and particularly for users of the UCDP data collection effort. The small effect size between reports where information is readily available and where it is not suggest that findings from inter-spatial (panel) studies using contemporary data generally should not be overly influenced by reporting bias. Considering that many studies aggregate violent events into district or grid-cell structures to merge with explanatory variables, a take-away from this exercise is in line with the recommendation of Weidmann (2015) that this data can be trusted as accurate at district level or within a 50 x 50 km radius.

Thirdly, our study provide support for the claim that news media over time has improved in its capacity of capturing political violence. This may be relevant for the debate on a global decline of conflict and other forms of political violence. Our findings suggest that contemporary news data – at least in the last decade – capture sufficient information about minor instances of violence, which means that we can be relatively confident that conflict data sources are providing a good overview of current instances of global armed conflict. For earlier years, even just a few decades ago, then information is more uncertain and it is likely that even more cases of low-level conflict may be missing as we move further back in time (Kreutz 2015*b*).

The fourth, and more worrying, implication of our findings is that we find that the quality of information declines in more violent conflicts. A possible reason for this is that in excessively violent settings, there are too many events to report leaving less time to investigate the details of the violence. It could also be that the high risk of reporting and the destruction of infrastructure in such situations means that fewer reporters are in a position to even seek information. Case

studies has alluded to this, including that news reports are particularly poor when violence escalates quickly (Davenport and Ball 2002; Restrepo, Spagat and Vargas 2006), and influenced by which actor controls a certain territory (Price and Ball 2014). The current conflict in Syria has drawn attention to the important role of reporting for scholars' access to conflict data (Powers and O'Loughlin 2015), and we hope that our findings can inspire further advances in this research field.

Fifth, and finally, this paper has added to what is starting to become compelling evidence in favor of treating data quality as equally important to theory and methodology in contemporary scholarship. This includes a continued attention towards identifying bias in the data employed for analysis, both through case-specific inquiries and in cross-national settings. We think it is of particular importance that such studies explore countries more at risk for conflict as well as censorship and poor working conditions for journalists rather than only the US or Western Europe.

References

- Aker, Jenny C and Isaac M Mbiti. 2010. "Mobile phones and economic development in Africa." *The Journal of Economic Perspectives* 24(3):207–232.
- Andreas, Peter. 2002. Transnational crime and economic globalization. In *Transnational Organized Crime and International Security*, ed. Mats Berdal and Mónica Serrano. London and Boulder: Lynne Rienner pp. 37–52.
- Atton, Chris and Hayes Mabweazara. 2011. "New media and journalism practice in Africa: An agenda for research." *Journalism* 12(6):667–673.
- CIESIN-ITOS-NASA SEDAC. 2013. "Global Roads Open Access Data Set, Version 1 (gROADSv1).".
- Dafoe, Allan and Jason Lyall. 2015. "From cell phones to conflict? Reflections on the emerging ICT–political conflict research agenda." *Journal of Peace Research* 52(3):401–413.
- Davenport, Christian. 2010. *Media bias, perspective, and state repression: The black panther party*. New York: Cambridge University Press.
- Davenport, Christian and Patrick Ball. 2002. "Views to a Kill Exploring the Implications of Source Selection in the Case of Guatemalan State Terror, 1977-1995." *Journal of Conflict Resolution* 46(3):427–450.
- Domingo, D and C Paterson. 2011. *Vol. 2: Newsroom ethnographies in the second decade of Internet journalism*. Vol. 67 New York [etc.]: Lang.
- Drakos, Konstantinos and Andreas Gofas. 2006. "The Devil You Know but Are Afraid to Face Underreporting Bias and its Distorting Effects on the Study of Terrorism." *Journal of Conflict Resolution* 50(5):714–735.

- 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* 30:65–80.
- Eck, Kristine. 2012. "In data we trust? A comparison of UCDP GED and ACLED conflict events datasets." *Cooperation and Conflict* 47(1):124–141.
- Eck, Kristine and Lisa Hultman. 2007. "One-Sided Violence Against Civilians in War Insights from New Fatality Data." *Journal of Peace Research* 44(2):233–246.
- Farhi, Paul. 2009. "The twitter explosion." *American Journalism Review* 31(3):26–31.
- Fenton, Natalie. 2010. *New media, old news: Journalism and democracy in the digital age*. London: Sage.
- FH, Freedom House,. 2012. "Freedom of the Press." <http://www.freedomhouse.org/report/freedom-press/2012/>. Accessed: January 23, 2013.
- Fleeson, Lucinda. 2003. "Bureau of Missing Bureaus." *American Journalism Review* (25):32–40.
- Franzosi, Roberto. 1987. "The press as a source of socio-historical data: issues in the methodology of data collection from newspapers." *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 20(1):5–16.
- Fullerton, Andrew S. 2009. "A conceptual framework for ordered logistic regression models." *Sociological methods & research* 38(2):306–347.
- Galtung, Johan and Mari Holmboe Ruge. 1965. "The Structure of Foreign News The Presentation of the Congo, Cuba and Cyprus Crises in Four Norwegian Newspapers." *Journal of peace research* 2(1):64–90.

- Gleditsch, Kristian Skrede, Nils W Metternich and Andrea Ruggeri. 2014. "Data and progress in peace and conflict research." *Journal of Peace Research* 51(2):301–314.
- Gleditsch, Nils Petter, Peter Wallensteen, Mikael Eriksson, Margareta Sollenberg and Håvard Strand. 2002. "Armed conflict 1946-2001: A new dataset." *Journal of peace research* 39(5):615–637.
- Goldfain, K and N Van der Merwe. 2006. "The role of a political blog: the case of www. commentary. co. za." *Communicare: Journal for Communication Sciences in Southern Africa= Communicare: Tydskrif vir Kommunikasiewetenskappe in Suider-Afrika* 25(1):p–103.
- GSM Association and Europa Technologies. 2009. GSM World Coverage 2009. Technical report Mobile World Congress 16-19 February 2009.
- Herkenrath, Mark and Alex Knoll. 2011. "Protest events in international press coverage: An empirical critique of cross-national conflict databases." *International Journal of Comparative Sociology* 52(3):163–180.
- Hersman, Eric. 2013. "The mobile continent." *Stanford Social Innovation Review* 11(2):30–31.
- IANA, Internet Assigned Numbers Authority,. 2013. "RFC6890. IANA IPv4 Special-Purpose Address Registry."
- ICANN, Internet Corporation for Assigned Names and Numbers,. 2011. "Available Pool of Unallocated IPv4 Internet Addresses Now Completely Emp-tied." <http://www.icann.org/en/news/press/releases/release-03feb11-en.pdf>. Accessed: 2013-11-20.
- Imai, Kosuke, Gary King and Olivia Lau. 2008. "Toward a common frame-

- work for statistical analysis and development.” *Journal of Computational and Graphical Statistics* 17(4):892–913.
- Kalyvas, Stathis N. 2004. “The urban bias in research on civil wars.” *Security Studies* 13(3):160–190.
- Kim, Chaiho. 2010. A study of Internet Penetration Percents of Africa using digital divide models. In *Technology Management for Global Economic Growth (PICMET), 2010 Proceedings of PICMET’10*:. IEEE pp. 1–11.
- Kirchner, Lauren. 2014. “Media as both weapon and defense in the Mexican drug war.”. Pacific Standard (Online; posted 11-March-2014).
URL: <http://www.psmag.com/navigation/health-and-behavior/media-weapon-defense-mexican-drug-war-76243/>
- Kreutz, Joakim. 2015a. “Separating dirty war from dirty peace: Revisiting the conceptualization of state repression in quantitative data.” *European Political Science* 14(4):458–472.
- Kreutz, Joakim. 2015b. “The war that wasn’t there: Managing unclear cases in conflict data.” *Journal of Peace Research* 52(1):120–124.
- Lawson, Chappell H and Joseph Chappell H Lawson. 2002. *Building the fourth estate: Democratization and the rise of a free press in Mexico*. University of California Press.
- Long, J Scott and Simon Cheng. 2004. Regression models for categorical outcomes. In *Handbook of data analysis*, ed. Melissa A Hardy and Alan Bryman. Thousand Oaks, CA: .
- MaxMind. 2013. “MaxMind GeoIP.” <http://www.maxmind.com/en/city>. Accessed: 2013-11-20.

- Mitchelstein, Eugenia and Pablo J Boczkowski. 2009. "Between tradition and change A review of recent research on online news production." *Journalism* 10(5):562–586.
- Mudhai, Okoth Fred, Wisdom Tettey and Fackson Banda. 2009. *African media and the digital public sphere*. Palgrave Macmillan.
- Myers, Daniel J and Beth Schaefer Caniglia. 2004. "All the rioting that's fit to print: Selection effects in national newspaper coverage of civil disorders, 1968-1969." *American Sociological Review* 69(4):519–543.
- Nelson, Andrew. 2008. "Estimated travel time to the nearest city of 50,000 or more people in year 2000." *Global Environment Monitoring Unit-Joint Research Centre of the European Commission, Ispra, Italy. Disponible sur <http://bioval.jrc.ec.europa.eu/products/gam/index.htm> (visit le 10/06/2014)* 183.
- 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.
- Ocitti, Jim. 2005. *Press, Politics and Public Policy in Uganda: The Role of Journalism in Democratization*. Edwin Mellen Press.
- Oliver, Pamela E and Daniel J Myers. 1999. "How Events Enter the Public Sphere: Conflict, Location, and Sponsorship in Local Newspaper Coverage of Public Events 1." *American Journal of Sociology* 105(1):38–87.
- Pavlik, John. 2000. "The impact of technology on journalism." *Journalism Studies* 1(2):229–237.
- pgRouting Project. 2013. "pgRouting Extensions for PostGIS."

- 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.
- Poese, Ingmar, Steve Uhlig, Mohamed Ali Kaafar, Benoit Donnet and Bamba Gueye. 2011. "IP geolocation databases: unreliable?" *ACM SIGCOMM Computer Communication Review* 41(2):53–56.
- Powers, Shawn and Ben O'Loughlin. 2015. "The Syrian data glut: Rethinking the role of information in conflict." *Media, War & Conflict* 8(2):172–180.
- Price, Megan and Patrick Ball. 2014. "Big data, selection bias, and the statistical patterns of mortality in conflict." *SAIS Review of International Affairs* 34(1):9–20.
- Raleigh, Clionadh, Andrew Linke, Håvard Hegre and Joakim Karlsen. 2010. "Introducing acled: An armed conflict location and event dataset special data feature." *Journal of peace Research* 47(5):651–660.
- Restrepo, Jorge A, Michael Spagat and Juan F Vargas. 2006. "The Severity of the Colombian Conflict: Cross-Country Datasets Versus New Micro-Data." *Journal of Peace research* 43(1):99–115.
- 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.
- Schudson, Michael. 1989. "The sociology of news production." *Media, culture and society* 11(3):263–282.
- Severin, Werner J and James W Tankard. 2010. *Communication theories: Origins, methods, and uses in the mass media*. Longman.

- Shapiro, Jacob N and Nils B Weidmann. 2015. "Is the phone mightier than the sword? Cell phones and insurgent violence in Iraq." *International Organization* 69(2):247–274.
- Shavitt, Yuval and Noa Zilberman. 2011. "A geolocation databases study." *Selected Areas in Communications, IEEE Journal on* 29(10):2044–2056.
- Smith, Jackie, John D McCarthy, Clark McPhail and Boguslaw Augustyn. 2001. "From protest to agenda building: Description bias in media coverage of protest events in Washington, DC." *Social Forces* 79(4):1397–1423.
- Snyder, David and William R Kelly. 1977. "Conflict intensity, media sensitivity and the validity of newspaper data." *American Sociological Review* pp. 105–123.
- Spitulnik, Debra. 2002. "Mobile machines and fluid audiences: Rethinking reception through Zambian radio culture." *Media worlds: Anthropology on new terrain* pp. 337–354.
- Staniland, Paul. 2010. "Cities on fire: social mobilization, state policy, and urban insurgency." *Comparative Political Studies* 43(12):1623–1649.
- Sundberg, Ralph and Erik Melander. 2013. "Introducing the UCDP georeferenced event dataset." *Journal of Peace Research* 50(4):523–532.
- Sundberg, Ralph, Kristine Eck and Joakim Kreutz. 2012. "Introducing the UCDP Non-State Conflict Dataset." *Journal of Peace Research* 49(2):351–362.
- Sundberg, Ralph, Mathilda Lindgren and Ausra Padskocimaite. 2011. "UCDP Geo-referenced Event Dataset (GED) Codebook version 1.5".
- Themnér, Lotta and Peter Wallensteen. 2014. "Armed conflicts, 1946–2013." *Journal of Peace Research* 51(4):541–554.

- Tiefelsdorf, Michael. 2006. *Modelling spatial processes: the identification and analysis of spatial relationships in regression residuals by means of Moran's I*. Vol. 87 Springer.
- Tollefsen, Andreas Forø, Håvard Strand and Halvard Buhaug. 2012. "PRIO-GRID: A unified spatial data structure." *Journal of Peace Research* 49(2):363–374.
- Urlacher, Brian R. 2009. "Wolfowitz conjecture: a research note on civil war and news coverage." *International Studies Perspectives* 10(2):186–197.
- Wang, Xiaokun and Kara M Kockelman. 2009. "Application of the dynamic spatial ordered probit model: Patterns of land development change in Austin, Texas." *Papers in regional science* 88(2):345–365.
- Weidmann, Nils B. 2015. "On the Accuracy of Media Based Event Data." *Journal of Conflict Resolution* 59(6):1129–1149.
- Weidmann, Nils B. 2016. "A closer look at reporting bias in conflict event data." *American Journal of Political Science* 60(1):206–218.
- Weimann, Gabriel. 2006. "Virtual disputes: The use of the internet for terrorist debates." *Studies in conflict & terrorism* 29(7):623–639.
- Wood, Simon N. 2003. "Thin plate regression splines." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 65(1):95–114.
- Woolley, John T. 2000. "Using media-based data in studies of politics." *American Journal of Political Science* pp. 156–173.
- Zhukov, Yuri M. 2012. "Roads and the diffusion of insurgent violence: The logistics of conflict in Russia's North Caucasus." *Political Geography* 31(3):144–156.