Introducing the Sierra Leone Local - Location Event Dataset (SLL-LED)

Kars de Bruijne¹

University of Groningen

***

This data-note introduces the Sierra Leone Local – Location Event Dataset (SLL-LED) – an event dataset of the armed conflict in Sierra Leone. These data includes an unrivalled number of events, actors and allies. The set is based entirely on local sources, and each event is geo-coded. This note first explains the process of the data construction and potential limitations. Second, it compares SLL-LED to other data collection efforts; UCDP-GED, the TRC, Bellows & Miguel, 2009 and ACLED-v4. It finds that these efforts have potential biases, inconsistencies and weak spots whereas the SLL-LED data is unrivalled in scope, depth and precision. A thorough analysis of the civil war from 1991-2001 in Sierra Leone is best achieved using SLL-LED. Finally, this data-note gives descriptive characteristics of the conflict, describes four distinct conflict environments and indicates potential contribution of the data to theoretical debates.

***

When using SLL-LED data please cite: “De Bruijne, Kars. 2014, “Introducing the Sierra Leone Local - Location Event Dataset (SL-LED).” Armed Conflict Location Event Dataset. www.acleddata.com”²

¹ The author likes to thank Alison Smith, Nicollo Giovanni, Nicollo Figa Talamanca of NPJW for their help in constructing coding rules. Clionadh Raleigh, Caitriona Dowd, Gabrielle Daoost for ACLED support. Edward Miguel for sharing data and Paul Richards for highlighting limitations to SLL-LED.

² Data was prepared for De Bruijne, Kars. Bullets and Bargaining: rationalist decision-making during the civil conflicts in Sierra Leone and Slovenia, 2015, Unpublished PhD Thesis.
**Introduction**

The conflict in Sierra Leone (1991-2001) figures prominently in the conflict literature and continues to serve as an exemplar for many explanations of civil conflict. However, to date, the lack of thorough and verifiable quantitative information on the number, nature, locations and agents of the conflict has hindered analysis of this conflict. To my knowledge, the Sierra Leone Local-Location Event Dataset (SLL-LED) introduces the first comprehensive and entirely local-sources based event dataset of the Sierra Leone conflict. The data is based on a public report by No Peace without Justice (NPWJ. Smith et. al, 2004). The collected events have an unrivalled level of disaggregation and geographical specificity of fighting patterns amongst the multiple fighting groups in Sierra Leone. This data-note introduces the data, presents the coding process and assesses the reliability and the potential biases in the data. Subsequently, it compares SLL-LED to four different datasets on the conflict to assess the reliability of this data. Finally, it presents descriptive statistics for the data.

Data presented here fits into the general trend of producing disaggregated within-country conflict information (Cederman & Gleditsch, 2009). Disaggregated conflict data is the gold standard of political violence information: it enables the assessment of specific causal and correlated relationships, spatial and temporal trends and patterns. Existing data projects have variable levels of similarity in coding procedures, sources and conflict categories. This event dataset generally follows established categories as proposed by the Armed Conflict Location Event Dataset- ACLED (Raleigh, 2010; Raleigh, Linke, Dowd, 2014), where the fundamental unit of analysis in this set-up is the *event* defined as “the interaction of designated actors at a specific point location and on a specific day (Raleigh, et.al. 2014).” Table I highlights some minor differences in the coding of specific events: SLL-LED, for example, does not include ‘riots’ as a specific category. Further, these data concern the civil war in Sierra Leone, and supplement the ongoing real-time ACLED collection of that state: SLL-LED consists of over 8700 events: the first event taking place in March 1991 and the last event in December 2001.

---

2. Local-source data and the coding process

The Sierra Leone local source – Location Event Dataset is based on a preliminary report published by No Peace Without Justice report (Smith et. al, 2004). The preliminary report is in turn based on an undisclosed dataset. This dataset has not been made available as the data was admitted into evidence by the Special Court of Sierra Leone e.g. in the trial vs. the Revolutionary United Front commanders Sesay, Gbao and Kallon (Smith, 2014: 57).

Local sources

SLL-LED is gathered from reports by No Peace Without Justice (henceforth, NPWJ), an NGO who engaged in extensive interviews and data collection in 2002 and 2003, following the announced end of the conflict. The conflict mapping exercise of NPWJ comprised of two stages. During the first stage, information was acquired from over four hundred ‘key persons’, while information was crosschecked and expanded by open source material in the second stage. The second phase allowed for verification of dates, location and actors. Furthermore, various local and international human rights experts were involved in triangulating the findings, most notably field monitors from the local NGO Campaign for Good Governance. These robust procedures suggest that the NPWJ collection was the most thorough and verified collection of information on the Sierra Leone civil war available.

The NPWJ report is based on the use of ‘key persons’. NPWJ obtained information at the chiefdom level and covered in total 146 chiefdoms (out of 149). An average of three ‘key persons’ covered each chiefdom. Key persons were selected depending on their knowledge of the conflict, presence in the chiefdom during the conflict, respect by the local community and often a degree of (higher) education. Often they were farmers or held a position of authority during the conflict. A total of 6.7% of the ‘key persons’ joined willingly one of the fighting factions and 10.7% were captured. Given these criteria – especially the requirement of being present in the chiefdom during the war - many ‘key persons’ have probably been members of or associated with, the Civil Defense Forces (CDF). As a result, a bias towards the activities of CDF may be present in the data. At the same time, a high-level membership in the Revolutionary United Front (RUF) provided detailed information, which may counterbalance a bias. There is no (clear) empirical evidence for a CDF bias (see below).

---

4 NPWJ provides an extensive justification and explanation for the chosen methodology (Smith et. al., 9-18). It may that information from the CGG human rights violation mapping in 2001 ended up in the report, p 11.
5 For a full list of criteria see, Smith et. al, 14. The conflict mapping recorders (CMRs) were responsible for the initial selection of ‘key persons’.
Statements from these ‘key persons’ were taken by 136 Sierra Leonian conflict mapping recorders. The intention was that the recorders were from same chiefdom as the key persons. Yet, NPWJ’s preliminary report admits excluding chiefdoms. Although the location of these chiefdoms is confidential, inspection shows that five chiefdoms are absent altogether and 16 chiefdoms have ≤10 events (outside 99% CI per district). Most of these were the least affected chiefdoms – often in the Northern Area. As a result, 136 conflict mapping recorders covered 128 chiefdoms (149-[16-5]) implying that each recorder covered one chiefdom.

**Coding the report**

For NPWJ publishing the report required a balance between gleaning significant information and protecting the identities of ‘key persons’. NPJW offered key persons anonymity when reporting on the events they had participated in or witnessed. NPWJ ensured protection for identities by vague details on frequency, location and dates of violence. As a result, SLL-LEDs systematic coding procedure also had to strike a similar balance between obtaining significant information and not revealing the identity of information. Fifteen coding rules were developed for consistent and reproducible inclusion of events and ensured that the confidentiality agreement was not violated. The NPWJ’s authors crosschecked these rules (using more than thirty examples) against the undisclosed database in their possession. Based on this, NPWJ offered guidance on how to best to code the frequencies, dates and some miscellaneous issues from the report (see annex 1 for the codebook).

The codebook specifies all of the coding choices the author made while the FAQ reports on common questions coders did encounter. The systematic coding diverged from the NPWJ report in clarifying the identity of actors from May 1997 onwards. In the NPWJ report, from May 1997 onwards the AFRC (Armed Forces Revolutionary Council) and RUF (Revolutionary United Front) are presented as one actor (the AFRC/RUF)). By now, it is well known that the relation between the AFRC and RUF was highly strained during the junta period (they kept separate lines of commands), broke down after the withdrawal from Freetown in February 1998 and was ad-hoc and often of an individual nature from December 1998 onwards. Moreover, both the AFRC and the RUF increasingly factionalized. I separated both spatially and temporally AFRC and RUF violence and identified factions within. This was motivated by the idea that users of this disaggregated dataset are likely to be interested in individual actor behavior, association of fighting groups with geographical areas and

---

6 Plus an additional 23 for the Western Area
7 Users should account for this exclusion when relevant to their analysis.
primarily as I aimed for doing justice to the mounting evidence. Separation was based on evidence provided by the Special Court. A justification is in annex 2.

Construction of the dataset was carried out in 2013 and 2014. To prevent coding-bias, the main coder included page-numbers for each entry twice. Moreover, various signals to identify each type of event characterized the ‘working copy’ of SLL-LED. This ‘working’ set is made available online to enhance transparency and encourage users to identify and corrected possible coding errors. After coding the main part of the data, two ACLED-coders reviewed the data and paid particular attention to the spatial dimensions of the data. Subsequently, two other coders checked and improved geo-coding.

The total amount of events in SLL-LED is 8786 entries. By contrast, the NPWJ preliminary report compiled approximately 5500 separate incidents from the ‘key persons’ (excluding information from the RUF source) and used 8500 entries from open sources to fill in gaps and ‘refresh the memory’ of key persons. Some open-source events ended up in the preliminary report. SLL-LEDS 8786 events on the one hand exceed the 5500 reported incidents by key persons but on the other hand are less than the reported number of both key person and open source data (14000 in total). Hence, the question arises as to how the number of events in SLL-LED relate to the NPWJ report. I claim that a direct comparison between the source and the SLL-LED data cannot be made. First, ACLED has a strict definition of what constitutes ‘political violence’ whereas it is unclear what constitutes an ‘incident’ or ‘entry’ in the NPWJ definition. Events in SLL-LED are for example ‘atomic’ - occurring between agents on a specific day and in a particular place while an ‘incident’ for NPWJ may include multiple agents and multiple days. Similarly, some events are coded as two separate events in SLL-LED while probably coded as one ‘incident’ by NPWJ, such as a non-violent take-over of a village while simultaneously killing civilians. Second, direct comparison between the SLL-LED events and NPWJs ‘incidents’ is impossible as NPWJ added ‘other data’ (p.17) of which the nature is unspecified. SLL-LED took a conservative interpretation of discrete events and crosscheck coding rules with NPJW. This dataset is, therefore, likely to be a reasonable approximation of the war’s intensity and character as compiled by NPWJ’s conflict mapping program.

8 www.acleddata.com/geopv/
Potential biases

SLL-LED potentially has three potential biases users may want to take into account: 1) the nature and aims of NPWJ; 2) the coverage of the report and; 3) coding choices that may influence results. Starting with the first, NPWJ works to advance an “international criminal justice system” and human rights. Already during the war (June 1998), NPWJ started assisting one side in the conflict, the recently reinstalled SLPP government (Sierra Leone Peoples Party) in the negotiations for the establishment for the International Criminal Court (ICC). Moreover, in 2000 the lead author of the NPWJ conflict mapping report acted as chief legal advisor to Solomon Berewa, Vice-President in the SLPP government - for the purpose of setting up the Special Court.9 As the SLPP was for a large part of the war in allegiance with the CDF, the relation between NPWJ and the SLPP may potentially have led to biases in the data (for example through NPWJs use of SLPP patrimonial networks and the selection of ‘key persons’). There are ‘noteworthy’ patterns in the data in relation to the CDF: after July 1999 there is suddenly no CDF violence against civilians recorded and there are suspected low violence levels for the most violent CDF ‘Black December’ campaign. The selection-criterion of ‘being present in the chiefdom’ rather than the relation between the SLPP and the CDF is likely to explain potential biases. Another concern may be the European Commission financed that NPWJs program while the Special Court for Sierra Leone co-financed it during the analysis phase. The general post-conflict climate in Sierra Leone at the time of recording was one of persistent misinformation over the Courts’ mandate. It might have led informants to report a lower number of recorded events to NPJW due to possible ties with the Court.10

Secondly, there are potential biases in the coverage of conflict events. In quite a number of situations, NPWJ required ‘key persons’ to recollect events more than ten years back. Such recollection is a clear problem even when taking place right after the conflict. Yet, the problem is attenuated by NPWJs use of open source material. Another bias may be the NPWJs focus on ‘human rights violations’ rather than battles. However, this bias is toned down by the simultaneous desire for ‘chain of events’ and ‘battle and command’ structure reconstruction, to facilitate commander accountability. Finally, three ‘key persons’ reporting (on average) per chiefdom may have led to underreporting – in particular of conflict events in forested chiefdoms in the rainy season. ‘Key persons’ are likely to have resided in one village for most of the conflict (due to the criterion of ‘position of authority’).11 Yet, most chiefdoms

---

10 E.g. the trials versus Taylor, AFRC, CDF and RUF include various instances that are not recorded.
11 The authors thank Paul Richards for pointing this out.
contain of over 100 villages and towns and it is very well possible that ‘key persons’ were unable to keep track of all events. This problem of underreporting is likely to have been particularly severe in the Eastern district where towns and villages are tucked away in forested terrain and inaccessible for much of the rainy season. Figure 1a depicts the number of events per month, showing a steep drop of events in the rainy season (May-Nov) in particular during the wettest months (July/August). Whereas the drop highlights the inability of the fighting forces to engage in fighting (a common feature of all violence in West Africa), figure 1b indicates that in densely forested areas the drop is more pronounced. Possibly, these figures display the effects of underreporting.

Finally, coding choices may have influenced results. In particular, the precision of dates, the geographical information and the inclusion of bases warrants caution. A weakness of the preliminary NPWJ report is the generality of dates (e.g. ‘from time to time’). NPWJ did provide advice on how to be more precise but the data have event clusters in the mid points of months. Secondly, SLL-LED includes an indication for geographical specificity (1-precise, 2-chiefdom level, 3-district level). Yet, despite NPWJ precision of listing even the smallest villages, the data has a large number of the more general ‘2’ codes which may strike observers as strange. However, a number of villages either disappeared or were too small to be located. Moreover, events often occurred in the neighborhood (but not in the direct vicinity) of small villages and towns and in those cases, a ‘2’ was assigned for geo-precision. Hence, the large number of ‘2’s. Finally, the NPWJ report did not provide start and end dates for headquarters and bases. To retain as much information as possible, I included any reference to a base in the dataset. While it is up to users to make decisions on how to model start and end dates of bases, I provided a list of main bases of the fighting factions in annex 3.
3. Comparing SLL-LED to other collected data

Four other datasets contain conflict information on the Sierra Leone Conflict.\(^{12}\) Two sets are general data collection efforts: ACLED - version 4 (2014) and the Uppsala Conflict Data Program Geo-referenced Event Dataset (UCDP-GED). Both event-datasets rely on secondary source material, although UCDP data does include some NPWJ data (>50%). Two other datasets are local sets that include local information but are not geo-referenced: the Truth and Reconciliation Committee’s dataset (TRC) and data from Bellows & Miguel, 2009 (see table II). The TRC-data used local testimonies to the commission right in 2003. Bellows & Miguel use the NPWJ report. Overall, each of these four datasets provides a different, and sometimes opposing, narrative. At the same time, SLL-LED is similar to the other datasets in terms of frequency and (rank) correlation of the nature of the fighting, actor culpability, timing and intensity and spatial characteristics. This section ends by discussing some (major) shortcomings of each set and points out that SLL-LED avoids most of these.

Nature of the fighting

Of the five datasets under consideration, the TRC provides by far the most extensive set with over 40,000 events in total and over 30,000 unique events.\(^{13}\) ACLED-v4, UCDP-GED and the Bellows & Miguel (henceforth B&M) include between the 1000 and 2000 events contrasting starkly with the 8787 events in SLL-LED (figure 2). Yet, the marked differences in the types of violence recorded by each set are more striking. TRC includes only violence against civilians (VAC), looting and a large category of events (45%) that is not recorded in any of the others sets (such as arbitrary detention, beatings and displacements). It narrates a story of civilian targeting and exploitation. ACLED-v4 offers an opposing and more traditional narrative.\(^ {14}\) In ACLED-v4, the conflict in Sierra Leone is predominantly a conflict characterized by battles between the state and non-state actors over territory (67%). VAC levels stand at about 20% and looting at roughly 8% of all events. The UCDP data strikes a middle ground between ACLED-v4 and the TRC data. In UCDPs reading, the Sierra Leone conflict has elements of a traditional conflict between military and non-state actors (42% battles) but also reports high levels of one-sided violence (58%).

---

\(^{12}\) Two other datasets on human rights violations in Sierra Leone - ABA/Benetech Sierra Leone War Crimes Documentation Survey (SLWCD) and data from the ‘Campaign for Good Governance - are confidential and unavailable.

\(^{13}\) Unique events occur on a different data or a different location.

\(^{14}\) ACLED includes only violence from 1997 onwards. While all sets report an increase in battles as opposed to VAC from 1997 onwards the dominant narratives are unaffected by this change.
SLL-LED however narrates a story of a different and far more complex conflict. It presents the story of a largely absent state as manifested by low levels of direct military engagements (20%). The absence of a state offered plenty of opportunity for unopposed take-over of territory by non-state actors (30%). Moreover, the absence of state authority allowed for looting and VAC on a large scale (45%). SLL-LED does not actually include fewer battles than UCDP-GED (1500 vs. 600). Rather, the open coding categories derived from ACLED combined with a comprehensive local source allow for the identification of additional types of conflict activity that put battles and one-sided violence in perspective. Perhaps, Sierra Leone is a ‘special case’ as what was left of government had little authority across the country, allowing for an unprecedented number of unopposed take-overs. However, it might be the relative importance of unopposed take-overs bears greater application across the continent and requires local source data for identification (see for example ‘troop presence’ in Humprey & Van der Windt, forthcoming: 11).

**Actor culpability**

Who are the main actors according each dataset? Table III provides rank-correlations for the main actors in the conflict from 1997 to 2001. Apart from ACLED-v4 there is general agreement on the relative contribution of each type of actor to the conflict (RUF, Government, CDF, AFRC and International Forces). All position the RUF at the top of the hierarchy except for ACLED. At the same time, there are differences in the relative magnitude each group (see figure 3 for details). Of particularly importance, is that CDF violence is far more common in SLL-LED than in other datasets. This is especially noteworthy in the light of the potential CDF-bias I highlighted before. This CDF activity may have two potential explanations; either CDF-associated ‘key persons’ overstated their relative importance or they told a story previously unnoted in the relative comfort of anonymity. If the first explanation holds, it is, however, also likely that they would downplay their role in human rights violations. However, the data highlights that about 30% of all CDF violence was either looting or killing. Again, this is much higher than the other datasets. It suggests that CDF ‘key persons’ narrated a story not present in other datasets. This may also imply that the bias I identified above may not be present. If most ‘key persons’ have a CDF background, it seem that these CDF associates have been forthcoming with information without censoring their own role. Perhaps this also applies to recollection of non-CDF violence.

---

15 The B&M dataset tells a similar story. The very high level of non-violent take-over of territory in their data (labeled ‘attacks’) is a result of a choice not to code VAC and/or looting seperately.
As SLL-LED reports a similar conflict hierarchy as the other datasets, it seems that its data is reliable. This however, should not imply that any dataset should be used. SLL-LED provides unprecedented specific information on the exact identity of the actors and their alliances. SLL-LED identifies – partly due to separation of AFRC/RUF violence after February 1998 – a total of nearly 100 actors and factions (as opposed to less than 30 in other sets). Moreover, SLL-LED reports that actors cooperated with other actors in 13% of all events. UCDP-GED and the TRC report absent or much lower alliance levels (Table 2).  

**Timing and conflict-intensity**

Figure 2 shows the relative distribution of conflict intensity. As is clear from this figure there is strong agreement between the datasets as regard to conflict intensity. Statistically the null hypotheses of no (rank) correlation for year, quarter and month between the sets are all rejected (see tables IV & V). This suggests that regardless of the reliance on secondary or local sources, all data collection efforts capture the intensity of the conflict over time. However, time-agreement between the sets over the intensity of fighting is weaker when disaggregating to VAC only. Table VI shows a weaker (month) correlation between SLL-LED, the TRC and UCPD-GED. The table shows that ACLED-v4 presents very different results even when the analysis is limited to 1997 to 2001 only. The consistent inconsistency of ACLED-v4 (deviations in identification main actors and the nature of the fighting) provide the main reason for replacing ACLED-v4 data on Sierra Leone with SLL-LED. At the same time, it is noteworthy that ACLED-v4 is the only dataset that relies exclusively on publicly available information mainly from newspaper sources. Perhaps, the usage of local sources allows for a more ‘truthful’ identification of nature, intensity and location of conflict. SLL-LED for example reports 1394 unique dates (1394). By contrast, all other sets report a lower number: TRC 945 (about 68%), UCDP-GED (48%), B&M (44%) and ACLED-v4 (42%).

**Spatial characteristics**

Spatial factors are critical in advancing and testing explanations of conflict. How do the various datasets relate on this dimension? A comparison of conflict intensity per district warrants two conclusions. First, it finds agreement between UCDP-GED, B&M and SLL-LED (Table VII) which underscores the reliability of SLL-LED. Second, it is possible that

---

16 The high level of alliances in B&M is due to the merging of AFRC and RUF violence.  
17 To compare conflict-intensity over time I excluded all TRC data without precise dates (68% of the data).  
18 Based on comparison of the 97-01 - to allow for ACLED inclusion - the results are very similar for the 91-01 period.
there is an urban bias in datasets relying on secondary sources. To see this, I compared all datasets by only using data from 1997 to 2001 and VAC and looting only. This had two effects. On the one hand, the only source relying exclusively on secondary sources (ACLED-v4) ranked the Western urban area as the most affected area in the conflict. On the other hand, a dataset partly based on secondary sources and partly on the NPWJ report (UCDP-GED) is subject to a similar problem. When excluding all events based on NPWJ report, UCDP-GED ranks the Western Urban area in the top three of most affected areas (table VIII). This suggests caution in using data based on secondary sources; such data tends to stress urban over rural violence and by extension, urban over rural explanations. This potentially can lead to incorrect conclusions as conflict-researchers are likely to use UCDP-GED and ACLED data to find spatial correlation between location and hypothesized conflict-causing variables (e.g. conflict and resources or conflict with distance to capitals).

Again, the correlation between datasets on location should not lead to indifference over choice of data. SLL-LED outperforms other data collection efforts. First, SLL-LED is likely to ameliorate the effects of urban bias. Secondly, it allows for testing of small geospatial units such as chiefdoms and villages. Increasingly, conflict researchers aim for microrelations beyond the district level using as units of analysis grids or chiefdoms (Raleigh & de Bruijne, 2014; Bellows & Miguel, 2009; Voors, Bulte van der Windt, forthcoming; Mokuwa et. al. 2012, Toleffesen, Strand and Buhaug, 2012, Wucherpfenning, et. al. 2011). While all datasets - except the TRC data - allow for analysis beyond the district level, none of datasets has SLL-LEDs detail for specific locations and spread within chiefdoms. As local data and analyzing techniques will become increasingly available, the benefits of using SLL-LED for within country analysis outweigh the benefits of using other types of data.

Problems in UCDP-GED, B&M and the TRC data

In addition, in being less comprehensive all dataset have also major individual problems: unclear coding for Bellows & Miguel, unclear usage of sources for UCDP-GED and a likely looting bias in the TRC data. First, there are large discrepancies between SLL-LED and the B&M data despite the use of the same underlying source. Partly this is due to different coding rules. All B&M plural events are coded as singular events and their coding as not been crosschecked with NPWJ. Users have to take into account that the resulting set, therefore, has major discrepancies with the conflict data recorded by NPWJ conflict mappers. Moreover,

The TRC data also ranks Freetown as first. However, this positions is driven by the Freetown attack (January 1999). The area ranks as 10th when excluding this attack. Exclusion does not affect ACLED-v4 ranking.
coding choices are different. Bellows and Miguel coded categories deductively and did not follow established data-collecting efforts. As a result, some categories merge various types of violence. For example, their category ‘attack’ includes VAC, looting and taking over territory unopposed. All events not attributable to chiefdoms were excluded and data is not geo-coded. SLL-LED is a major improvement of the Bellows and Miguel coding of the NPWJ report.

Second, UCDP-GED depends too much on non-justified source-usage. Generally, the UCDP-GED uses conservative categories that cannot capture unopposed take-overs, win-loss ratios, and does not include information on alliances or smaller violent groups. However, analysis of the depiction of Sierra Leone in UCDP-GED data points to a particular severe problem. The overall narrative of UCDP-GED is that battles exceed the incidence of one-sided violence in civil conflict. Sierra Leone, however, is depicted a conflict with more one-sided violence than battles. It finds itself in a class with Congo, Liberia, CAR and Rwanda - all conflicts where one-sided violence outweighs battles (table XI). As these conflicts are often viewed as ‘new wars’ this is unsurprising. But do these conflicts indeed have a different conflict environment? To assess this, I focussed specifically on the usage of source; what types of sources were used to determine what types of violence? For Sierra Leone, UCDP-GED relied for its one-sided violence reporting on the NPWJ report: 68% of the one-sided violence in the UCDP data was based on the NPWJ report. At the same time, only 31% of the state-based violence in UCDP-GED came from NPWJ. This implies that UCDP's characterization of the conflict may derive its logic from using the report rather than an inherent different conflict environment.

To assess how UCDP depiction of Sierra Leone relateds to other conflicts, I categorized the sources UCDP-GED used in constructed the data: Human Rights sources (Amnesty, Human Rights Watch, NPWJ), News sources (Reuters, AFP, Xinhua, AP) and an undefined category with other sources (books, UN reports, local Human Rights sources). Unsurprisingly, a strong relation exists between the type of source used and type of violence reported; news sources identify battles while human rights sources detect one-sided violence. Subsequently, I compared Sierra Leone and the countries with higher one-sided violence to those with more battles and focused on source use. Quite surprisingly, UCDP-GED included substantively more HR-sources in all a-typical countries (> one-sided violence - see Table XII). UCDP-GED characterization of the conflicts in Sierra, therefore, derives it logic from a different (and not justified) usage of underlying sources. It is likely that a-typical countries like Sierra Leone were typical countries had same proportion of news sources been used.
Third, the TRC data may be subject to looting bias. A first indication came from comparing rank-agreement on the location of violence when focusing on both looting and VAC (Table IX). This was particularly surprising, since for all other indicators (timing, actors and all violence) a relation between the datasets was established. Moreover, one would in particular expect a relation between TRC and SLL-LED data since both used local source data. Even more puzzling was that a correlation between TRC and other datasets re-appeared after I restricted the sample further to VAC only (Table X). As noted in a comparison of the TRC data to other (undisclosed) local source datasets, there may be a looting bias in the TRC data. Possibly testimonies before the committee were given in expectation of receiving compensation for losses suffered during the war (Ghodes, 2010).

Indeed the incidence of looting dominates VAC in the TRC data contradicting all other datasets where VAC consistently dominates looting. For a looting bias to exist the following empirical predictions could hold: 1) less precise date indication for looting reports; 2) precise date-indications at well-known conflict peaks; 3) absence of correlation between TRC looting report and those in other datasets and; 4) a bias towards perpetrators without ties to current regime. In fact, each empirical prediction is confirmed by the data (Figure 4). Firstly, when TRC statements lack precise dates, looting incidence is much higher than VAC. Secondly, when focusing on precise indications of dates, looting and VAC have similar levels over time, unless one focuses on commonly known peaks of the conflict (March, 1991, February 1995 and January 1999). In those cases, reports of looting exceed by far reported VAC levels. Thirdly, there is no spatial correlation between the TRC data and other datasets when focusing on looting only (but the relation appears again when I focus on VAC only). Fourthly, it was common knowledge that the RUF had lost the war and did not have ties to the SLPP government. Many AFRC, WSB (and CDF) soldiers were recruited back in the retrained army. The TRC records a relative high number of RUF-activity (70%). This may suggest that people reported higher RUF looting levels as to avoid loyalty issues. A clear example is the Port-Loko district. The West Side Boys were notorious for looting activity in Port Loko. In all other datasets, Port-Loko is one of the most affected districts in large part because it was home to the West Side Boys.20 Surprisingly, in the TRC data, Port Loko is the least affected district. This contradicts common knowledge about the conflict and the behavior of the West Side Boys in particular (Utas, 2008). This example and the preceding four points point to a likely looting bias in the TRC data.

---

20 But also because it was a transit to Freetown from Kono and back.
4. Descriptive Statistics: four conflict environments

What does SLL-LED tells users substantively about the Sierra Leone civil war? I provide descriptive statistics for general characteristics of the war and identify four conflict environments. RUF as the primary actor is likely to set the stage for each conflict environment. The subsequent section also outlines potential implications for debate.

Civil war in Sierra Leone

SLL-LED conflict-narrative is that of a largely absent Sierra Leone state that enabled its challenger to advance unopposed. This permitted a conflict environment of unprecedented levels of VAC and looting by any conflict actor that entered the battlefield (RUF, Government, CDF, AFRC and International Actors). From March 1991 until December 2001, SLL-LED gives an average of 68 events per month, comprising of 14 battle-events, 30 VAC/looting events and 24 unopposed take-overs of towns and villages. Battle success rates are at parity with on average seven draws, four wins for government and its allies (CDF, International Forces) and three for RUF. Yet for unopposed takeovers RUF has the advantage, taking on average seventeen villages a month compared to seven for Government (2), CDF (3) and International Forces (2) combined (Table XVI). These numbers testify to the formidable challenge RUF posed to the various governments throughout the conflict.

However, as Figure 5 highlights, distribution of violence over time indicates major differences between years and months. This variation in violence levels suggests distinct temporal patterns. Peaks in the data centered around commonly acknowledged highpoints of the conflict.\textsuperscript{21} Noteworthy, is that SLL-LED (and Bellows & Miguel) finds little support for the idea that “J6” (the attacks on Freetown in 1999) constituted the high point in the conflict. Contrary to popular, human rights and academic sources, the highpoint of the conflict is located in March 1998 when multiple actors engaged in large-scale violence.

Implications for debate

SLL-LED has implications for debates on the location of violence and the occurrence of VAC. Many believe that the civil war in Sierra Leone affected mainly the Southern and Eastern area and in particular Kailahun district. For example, Bellows and Miguel (2009: 42) present a self-reported victimization index that singles out Kailahun. Similarly, most RUF

recruits the South and East had been most affected. However, SLL-LED shows a far more complex picture. The first observation is that Kailahun is only an average district with substantially less violence than Port Loko, Kenema and Kono (the most affected chiefdoms – table XIII). Moreover, the Northern (and Western) district became the most affected districts as the conflict unfolded. SLL-LED, therefore, suggests that violence was much more equally spread over the country than often believed.

In relation to debates about VAC in Sierra Leone, the reference article remains Humphreys and Weinstein (2006) using self-reported abuse levels from more than 1000 random combatants. They conclude that internal characteristics of the fighting units explain VAC variation best. They also argue that contestation – VAC in retaliation after lost battles - and the association of fighting units to territory, do both not explain variation in VAC. Rather than relying on self-reported abuse levels by perpetrators, SLL-LED is based on local source data from (mainly) victims. Perhaps due to this, there are differences between the datasets. After 1992, Humphreys and Weinstein (henceforth H&W) report a gradual increase in RUF abusiveness and a “fairly consistent” abuse rate of the CDF. SLL-LED, however, finds that RUF abuse levels decrease (from 1993-2001) whereas CDF abuse levels increase (Figure 6: 1993-2001).

More important than different descriptive statistics is that SLL-LED potentially undermines H&W core findings. A simple OLS model (chiefdom) and a simple time-series model (chiefdom/month) with VAC as dependent variable, gives significant and robust results for battles and territorial take-overs as explanatory variables (table XIV). This implies that VAC occurred most often in the vicinity of the battlefront which directly impacts H&W findings that contestation does not explain variation in VAC. A similar effect is present in relation to their other result on inability to explain VAC variation by the relation of fighting units to home territory. The largest discrepancy in this relation between VAC and battles is Kailahun district. It tops the list of districts with most battles but is below average for VAC (eight out of thirteen). If internal characteristics explain violence on would in particular expect Kailahun to have high VAC levels. Yet VAC levels are ‘low’ in Kailahun suggesting a different logic for this area and perhaps more broadly. None of the above sufficiently challenges the H&W findings. However, data contained in SLL-LED allows for an alternative means of testing some of the core claims made in the literature on VAC in Sierra Leone.

22 Their identification of district conflict-exposure is similar to SLL-LED.
23 SLL-LED findings square with Keen, 2005. He notes a feedback patter between battles and VAC as fighting groups punished population after lost battles. H&W measure contestation indirectly through dominance of groups.
Related to this debate on VAC is the claim that RUF's initial usage of VAC was a function of NPFL involvement. As the NPFL forces withdrew in 1992 comparison is possible between RUF with and without NPFL fighters. SLL-LED finds no real support. Admittedly, the ratio of VAC per battle for RUF alone is about \( \frac{1}{4} \) of the RUF/NPFL ratio (1.4 vs. 5.7 – per month/weighted – see Table XV). However, this difference is driven completely by the first two months of the war. If removed, RUF is *more violent* that the NPFL.

*Four conflict environments*

A final observation is that SLL-LED empirically supports the idea forwarded by conflict specialists that the Sierra Leone conflict took place in different phases (e.g. Peters, 2010; Hoffman, 2011; Benetech, 2006; TRC Vol. 3A). SLL-LED identifies *four* phases with a distinct conflict environment (Table XVI) and accompanying geographic imprint (Figure 7). Since RUF consistently tops the conflict-hierarchy, it may drive changes between phases.

Phase one (March 1991-November 1993) represents the inception phase of the war. Violence is restricted to the South and East. Battles account for about 19% of all violence, VAC levels are high (48%) and unopposed-takeovers stand at 33%. Phase one presents a mixed picture in terms of overall success. On the one hand, battle success for government more than doubles RUFs. On the other, RUF more than doubles governmental success in reasserting control over territory. Phase one, surprisingly, has the largest percentage of joined operations (actors teaming up with other forces). Both RUF and government benefitted from cooperation as battle success rates for both far surpassed solitary action. Especially for government, military success was driven by outside assistance.

Phase two is named after a RUF change in strategy from semi-conventional warfare to jungle-guerilla warfare. Phase two has a distinct spatial imprint. Violence moved through the middle of the country via Bo and Kenema district reaching the Northern areas. The conflict actors started to operate in different localities: no CDF violence in the North and hardly external forces in the South and East. Also the type of violence changed. Conflict intensity doubled due to a massive increase in VAC and territorial take-over (16% battles, 45% VAC, 25 Many Sierra Leonean RUF commanders (and civilians) blame NPFL fighting forces assisting RUF Peters, 143.
26 If the two months are removed RUF/NPFL VAC event per battle drops to 1. Moreover, it remains unclear whether RUF vanguards engaged in VAC during RUF/NPFL operations and if RUF held any sway over NPFL.
27 Although settled by now, the starting date of the war has been debated. Officially the starting date has been March 23 Gberi, 2005: 60; TRC, 2004: 115. Yet, Keen, 2005: 49 and Wai, 2010: 93 and Peters, 2011 all point out the initial attack was a cross-border raid of Liberian forces connected to Charles Taylor. RUF sources are split over the starting date: some point to March 27 as the first dispatch of RUF trained insurgents others single out April 1/2. SLL-LED highlights to March 27 as the first dispatch in Kailahun & March 28 in Pujehun. The beginning of April sees an additional dispatch of troops.
28 Government is active in the North as well.
39% take-over). Correlation between battles and VAC dropped from .57 to .26 implying that VAC became a goal in itself rather than accompanying battles. Success in phase two is RUFs: the absolute number of RUF take-overs and relative battle success skyrockets.

A third transition occurs after the May 1997 coup, as the AFRC and RUF start cooperating officially. Phase three is by far the most violent conflict phase. Violence levels double phase two and triple those of phase one. Attesting to the intensity of the conflict is the large number of actors and active conflict dyads (nearly 100). Another difference with the previous phases is a change in the conflict hierarchy. The CDF becomes the most active actor (after RUF) and the primary battle dyad becomes RUF - ECOMOG violence. Battle success for government (& allies) drops further to 37%. But also RUF becomes less successful in taking over territory and fighting battles. The success the movement has is strongly driven by the AFRCs performance. Phase three has a distinct spatial pattern: CDF forces are uncontested in the South, External forces cluster in the North as do AFRC operations.

Starting right after the Lomé agreement (July 1999) and extending to the end of the conflict, phase four is by all accounts the least violence phase of the war. Violence levels are less than a quarter of the monthly violence of phase three and lower than all other phases. The relative contribution of battles to total violence is high (28%). SLL-LED finds major changes in the ability of actors to win battles and take-over territory. RUF and government success in battle drops dramatically but at different times. For RUF the drop is most pronounced post October 2000 (after the Abuja agreement). Battle success is only 2%. Government & allies quadruples RUFs success. What’s more, government & allies also double RUF territorial takeovers (67 to 33%). The spatial imprint of phase four is: low levels of violence in the Southern area, clustering of events in border areas with Guinea and CDF forces operating increasingly in the East and North of the country.
Conclusion

This data note introduced the Sierra Leone Local – Location Event Dataset (SLL-LED). The data is entirely based on local-sources and is unrivalled in detail, precision and comprehensiveness. Data construction used the conflict-mapping program from No Peace Without Justice (NPWJ). NPWJs lead author cross-checked the rules by which their conflict-mapping report was coded into data. Potential biases include underreporting from remote areas, a lack of precision of dates and reliance on informants that may have been liaised with one of the warring factions, the CDF. At the same time, this note found not empirical evidence for a CDF bias.

This note also compared publicly available datasets with SLL-LED. It found rank correlations (from .5 to .9) between SLL-LED and the other datasets in relation to nature of fighting, timing, location, actors. It implies that SLL-LED is likely to be a reliable approximation of existing knowledge on the Sierra Leone conflict. Yet the note also outline major issues in existing data-sets: ACLED-v4 presented different results than all other datasets, data from Bellows and Miguel (2009) is based upon unclear coding, a non-justified and asymmetric use of sources in UCDP-GED depiction of the Sierra Leone conflict and a potential looting bias in the TRC data. As a result, SLL-LED will replace ACLED-v4 data. SLL-LED also includes more types of violence, a greater amount of actors and alliances, a superior amount of geographical information. It also separates AFRC from RUF violence.

Overall, SLL-LED provides a different conflict narrative that the other datasets. A largely absent state allowed challengers to take a lot of territory unopposed leading to high levels of Violence Against Civilians (VAC) and looting. However, the narrative is different over time as SLL-LED identifies four distinct conflict environments in Sierra Leone. There are major differences between these phases in relation to conflict intensity, activity of actors, and overall success. SLL-LED may potentially impact debates on the Sierra Leone war – such as the debate on VAC. Yet, the detail of the data allows for a host of applications such as analysis of the determining factors for battle success, the analysis of alliance patterns (e.g. the currently undetected cooperation between guinea and ULIMO in 1993) and impact assessments of new strategies (such as the deployment Executive Outcomes, ECOMOG, UNAMSIL and the UK). Overall, the data presented here - despite inherent biases - is likely to provide a fuller and more truthful account of the fighting that took place in Sierra Leone than any of the public datasets currently available.
Literature


