CARMA 2024

6th Int. Conf. on Advanced Research Methods and Analytics
Universitat Politècnica de València, València, 2024
DOI: https://doi.org/10.4995/CARMA2024.2024.17831

Violence Index: a new data-driven proposal to conflict monitoring

Luca Macis, Marco Tagliapietra, Elena Siletti, Paola Pisano
Department of Economics and Statistics "Cognetti de Martiis", University of Turin, Italy

How to cite: Macis, L.; Pisano, P.; Siletti, E.; Tagliapietra, M. 2024. Violence Index: a new data-driven proposal to conflict monitoring. In: 6th International Conference on Advanced Research Methods and Analytics (CARMA 2024). Valencia, 26-28 June 2024. https://doi.org/10.4995/CARMA2024.2024.17831

Abstract

In this work, we propose a Violence Index (VI), as a comprehensive indicator of violence related to wars, conflicts, and disorders across different countries worldwide. This index is defined by mashing up different data sources: big data represented by the temporal progression of ACLED (Armed Conflict Location and Event Data) variables and an ad hoc dataset defined by our team documenting wars, armed conflicts, civil wars and violent demonstrations since 2010. The purpose is to encapsulate the intensity and impact of such unrest events in a single measure, the VI, to give a simpler, up-to-date, and manageable tool to practitioners and policymakers, for both prevention and strategic planning to let them behave better in future tragic scenarios.

Keywords: *ACLED; war; indicator; merging data; web scraping*

1. Introduction

Conflict prediction and early warning systems play a crucial role by identifying potential risks and threats, and offering decision-makers timely information to formulate policies for conflict mitigation and prevention. Scholars identified two possible ways of doing conflict and unrest prediction: the identification of potential conflicts and crises heavily relies on individual diplomatic and political knowledge, intuition, and subjective judgment; or you can recognize the potential of current technology, data science, and adopt it in this field (Gleditsch, 2002). In literature, different data have been adopted for this kind of issue, and they can be summarized in “Social Datasets” (i.e. social network data as Twitter, Telegram, etc.), “Disaggregated Datasets” (i.e. ACLED, UCDP (Uppsala Conflict Data Program), GDELT(Global Database of Events, Language, and Tone)) or “Aggregated Dataset” (i.e. World Bank, V-DEM).

Some scholars have enlightened some issues related to the use of “Social Datasets” in this context. The effectiveness of social network—Twitter—data for organizing insurrections has diminished due to prohibitions on violent tweets and the tracking of users by authoritarian regimes (Junior et al., 2021); There are no notable applications in the field of conflict prediction using social network data (Telegram, Khaund et al., 2021). Considering diplomatic datasets –

both disaggregated and aggregated – as ACLED (Raleigh et al., 2023), UCDP (Sundberg et al., 2013; Davies et al., 2023) or GDELT (Leetaru et al., 2013), other issues could occur: they are defined using indirect information (newspapers, etc.) and not by direct sources; moreover, they are mostly disaggregated datasets, since the disaggregation form avoids the use before data manipulation to practitioners or policymakers.

To avoid these issues, in our proposal, we adopted different datasets, with different characteristics in origin, structure, and frequency (Iacus et al, 2020), to define a single indicator (VI), which allows prompt use, without any further analysis, by policymakers.

2. The Data

The VI is defined mashing up different sources of data. In the first step, we use an ad hoc dataset defined by our team collecting in different ways information about the wars and armed conflicts that have occurred from 2010. In the second step we use a huge dataset, that is largely adopted for studies in the wars or conflict context (Hegre et al., 2012; Halkia et al., 2020), defined as the temporal progression of ACLED variables across more than 40 wars and conflicts spanning from the year 2010 to the present. In the following sections a description of the data.

2.1. ACLED Data

The Armed Conflict Location and Event Data Project (ACLED) is a project finalized for data collection, analysis, and crisis mapping, that was created by Clionadh Raleigh in 2005. It is a remote organization, which allows its team to live and work in all countries and contexts, where they collect and analyze instability. The members work within ACLED's executive office, global programs, external engagements, fundraising, and development or operations departments. In this way, ACLED data are derived from a wide range of local, national, and international sources in over 75 languages. The team conducts analysis to describe, explore, and test conflict scenarios, and makes both data and analysis open for free use. Moreover, researchers worldwide collect information on the dates, actors, locations, fatalities, and types of all reported political violence and protest events around the world. All data is updated in real time and published weekly. Years of historical coverage vary across countries and regions.

As detailed in Table 1, ACLED data take into account different event types, they focus on tracking a range of violent and non-violent actions by or affecting political agents, including governments, rebels, militias, identity groups, political parties, external forces, rioters, protesters, and civilians.

The total events collected by ACLED since 1997 are more than two millions, indeed from 2010 are more than one and a half million. The structure of the data within this dataset is meticulously designed to capture and organize information essential for comprehensive analysis of conflict dynamics. At its core, ACLED data revolves around detailed event descriptions, encompassing the date, time, and location of each recorded incident. This information is vital for understanding the temporal and spatial dimensions of conflicts. Moreover, ACLED provides in-depth insights into the nature of events, including the actors involved and the characteristics of each incident. This categorization enables researchers and analysts to discern patterns of conflict, identify key stakeholders, and assess the intensity and outcomes of various events. Central to the reliability of ACLED data is its rigorous verification process, which documents the sources of information for each event. This transparency enhances the credibility and trustworthiness of the data, essential for informed decision-making and academic research. Furthermore, ACLED's temporal and geospatial dimensions add depth to its analytical capabilities. By organizing data chronologically and georeferencing event locations, ACLED empowers researchers to conduct temporal and spatial analysis, identifying temporal trends and spatial hotspots in conflict activity.

Indeed, in our study, we have chosen to aggregate these data on a weekly basis, grouping them according to event types. This approach entails tabulating the occurrence of events within a specific week in a given country, leveraging the geospatial information provided by ACLED. Consequently, this process yields a dataset wherein each row corresponds to a particular week in a country, including the count of events of a specific type transpiring during that week within that country.

Simply, we defined a new panel dataset, defined by countries and weekly frequency. The time series consists of reporting the counts of sub-event types that occurred in each country and their respective fatalities. To finalize the creation of our dataset, we opted to exclude data exhibiting more than 94% zeros, reflecting a significant absence of observations. Following this criterion, we retained data from 175 countries for subsequent analysis.

Due to the nature of the original ACLED information, which could be downloaded for free by each user through their API[1], and the dimension of the dataset, it could be defined as a Big Data resource.

2.2. Wars dataset

For the creation of the ad hoc war dataset, we adopted a web scraping technique, downloading data from Wikipedia. In this way, we obtained a .csv file containing information about wars, conflicts, and violent protests from January 1st, 2010 to December 31st, 2022 (from now on, unrest events).

1 Armed Conflict Location and Event Data Project (ACLED); https://acleddata.com

Table 1. ACLED Event Types (https://acleddata.com/)

<table>
<tr><th>Event type</th><th>Su-event type</th><th>Disorder type</th></tr>
<tr><td rowspan="3">Battles</td><td>Government regains territory</td><td rowspan="3">Political violence</td></tr>
<tr><td>Non-state actor overtakes territory</td></tr>
<tr><td>Armed clash</td></tr>
<tr><td rowspan="3">Protests</td><td>Excessive force against protesters</td><td>Political violence; Demonstrations</td></tr>
<tr><td>Protest with intervention</td><td rowspan="3">Demonstrations</td></tr>
<tr><td>Peaceful protest</td></tr>
<tr><td rowspan="2">Riots</td><td>Violent demonstration</td></tr>
<tr><td>Mob violence</td><td rowspan="10">Political violence</td></tr>
<tr><td rowspan="6">Explosions/ Remote violence</td><td>Chemical weapon</td></tr>
<tr><td>Air/drone strike</td></tr>
<tr><td>Suicide bomb</td></tr>
<tr><td>Shelling/artillery/missile attack</td></tr>
<tr><td>Remote explosive/landmine/IED</td></tr>
<tr><td>Grenade</td></tr>
<tr><td rowspan="3">Violence against civilians</td><td>Sexual violence</td></tr>
<tr><td>Attack</td></tr>
<tr><td>Abduction/forced disappearance</td></tr>
<tr><td rowspan="8">Strategic developments</td><td>Agreement</td><td rowspan="8">Strategic developments</td></tr>
<tr><td>Arrests</td></tr>
<tr><td>Change to group/activity</td></tr>
<tr><td>Disrupted weapons use</td></tr>
<tr><td>Headquarters or base established</td></tr>
<tr><td>Looting/property destruction</td></tr>
<tr><td>Non-violent transfer of territory</td></tr>
<tr><td>Other</td></tr>
</table>

We considered this time window since the ACLED data collection from 2010 takes into account not only African countries.

The necessity to use this data is because ACLED data only records pure factual events but does not record political statements such as declarations of wars, revolutions, etc..

For each unrest event, we kept information about:

- Country: the country where the war happened.
- ISO 3166 code: the 3 letters unique code associated with each country.
- Starting date: the starting date of the unrest event.
- Ending date: the ending date of the unrest event. If the war is still ongoing, we use "present".

- War name: a label that recognizes the event.
- Type: a classification describing the event (Violent Demonstration, Armed Conflict, Civil War)
- Precise location: if available, it reports the specific area, such as a city, region, or country.
- War description: a brief note portraying the event and its actors and reasons.
- Link: the Wikipedia link reporting the event.

After the web scraping, the dataset was defined by 69 rows – i.e., events – but after a first human analysis, some events were dropped because there were no recorded battles – by ACLED – in the period under consideration (one week before and one week after the starting date). The final dataset consists of 46 unrest events from 2010 to 2022.

3. Methods and results

The VI integrates all ACLED variables, their weights determined through our analysis of the data within the war ad hoc dataset. Our proposal, for this reason, is defined in different steps.

Initially, we normalized all ACLED variables using the Min-Max method, re-scaling them to a range between 0 and 1 to ensure uniformity across variables (Mazziotta & Pareto, 2020). This normalization facilitated comparisons of variable fluctuations over time. Specifically, we examined the percentage change of variables within a two-week timeframe, spanning one week before to one week after the starting date of an unrest event. This process was defined by:

$$y_{i,j,k} = \frac{x_{i,j,k} - min_i(x_{i,j})}{max_i(x_{i,j}) - min_i(x_{i,j})} \tag{1}$$

where i represents the countries, j denotes the variables, and k indicates the weeks.

Subsequently, we analyzed the behavior of scaled variables within a two-week window surrounding each unrest event. By aggregating these variations across all events, we quantified their contributions and expressed them consistently as percentages.

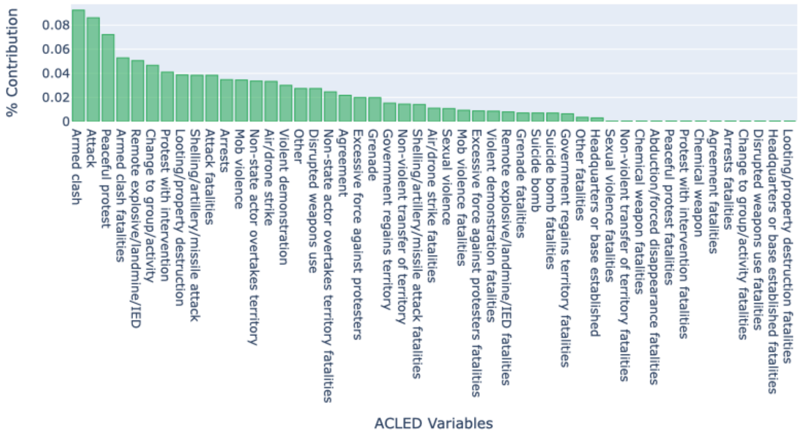

Figure 1. The contributions of each variable.

We then constructed the VI multiplying the scaled value of each ACLED variable by its corresponding weight contribution, determined in the previous step, for each week.

The culmination of our proposal involved summing the weekly contributions to obtain the VI. This index peaks during periods of war or armed conflict, reflecting intensified events among the 46 unrest events analyzed.

The summation of scaled data within the window allowed us to capture overall changes in scaled values during critical periods for each variable. Additionally, calculating the percentage contribution of each scaled variable within the specified time window provided insights into the variables with the most significant impact.

As a result, we can identify the variables that experienced the most substantial variations in intensity at the onset of wars since 2010. Lastly, we defined the VI for each country and day in the dataset based on scaled values of relevant variables and their respective percentage contributions. This calculation, outlined by equation (2), considers the weighted contributions of all ACLED variables, providing a quantitative description of unrest events tailored to each country.

$$VI_{i,j} = av_{i,j}^{(1)} \cdot cav^{(1)} + \quad av_{i,j}^{(2)} \cdot cav^{(2)} + \ldots + av_{i,j}^{(n)} \cdot cav^{(n)} \tag{2}$$

where:

$av^{(n)}$ is the n-th re-scaled ACLED variable, $cav^{(n)}$ is the n-th weighted contribution w.r.t. $av^{(n)}$, i is the i-th week, j is the j-th country, and n is the total number of ACLED variables.

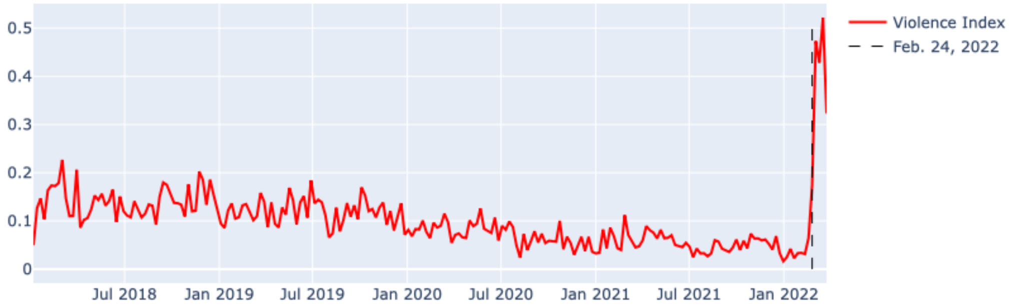

Figure 2.The VI trend from January 2017 to March 2022 in Ukraine.

The methodology offers several advantages. Firstly, it provides a detailed quantitative portrayal of unrest events specific to each country, thereby offering nuanced insights into the complexities of sociopolitical conflicts. Secondly, it allows for the delineation of unrest events tailored to the unique contexts of individual countries or regions. Furthermore, by integrating weighted contributions across all unrest events, the analysis ensures a comprehensive evaluation of variable impacts, enhancing the reliability of the results. However, the methodology also presents some challenges. The varying scales of magnitudes across countries may impede direct comparisons between countries, potentially complicating cross-country assessments. These considerations underscore the importance of cautious interpretation and contextualization of findings within the specific socio-political landscapes of each country.

Further results will be detailed and described during the presentation.

Acknowledgments:

This study was funded by the European Union - NextGenerationEU, in the framework of the GRINS -Growing Resilient, INclusive and Sustainable project (GRINS PE00000018 – CUP D13C22002160001). The views and opinions expressed are solely those of the authors and do not necessarily reflect those of the European Union, nor can the European Union be held responsible for them.

References

ACLED. (2019). *ACLED Codebook, 2019*. Armed Conflict Location & Event Data Project (ACLED). www.acleddata.com

Halkia, M., Ferri, S., Schellens, M. K., Papazoglou, M., & Thomakos, D. (2020). The Global Conflict Risk Index: A quantitative tool for policy support on conflict prevention. *Progress in Disaster Science*, *6*, 100069. https://doi.org/10.1016/j.pdisas.2020.100069

Hegre, H., Karlsen, J., Nygård, H. M., Strand, H., & Urdal, H. (2012). Predicting Armed Conflict, 2010-20501. *International Studies Quarterly*, *57*(2), 250–270. https://doi.org/10.1111/isqu.12007

Davies, Shawn, et al. *Organized Violence 1989–2022, and the Return of Conflict between States.* Journal of Peace Research, vol. 60, no. 4, 13 July 2023, https://doi.org/10.1177/00223433231185169.

Gleditsch, K. S. (2002). Expanded Trade and GDP Data. Journal of Conflict Resolution, 46(5), 712–724. https://doi.org/10.1177/0022002702046005006

Iacus, S.M., Porro, G., Salini, S., & Siletti, E. (2020) Controlling for Selection Bias in Social Media Indicators through Official Statistics: a Proposal, *Journal of Official Statistics*, ISSN: 2001-7367 (0282-423X); doi: https://doi.org/10.2478/jos-2020-0017

Junior, M., Melo, P., Couto da Silva, A.P., Benevenuto, F. & Almeida, J. (2021). *Towards understanding the use of telegram by political groups in Brazil*. In: Proceedings of the Brazilian Symposium on Multimedia and the Web WebMedia ’21 , 237–244. New York, NY, USA. Association for Computing Machinery (https://doi.org/10.1145/3470482.3479640).

Leetaru, K., & Schrodt, P. A. (2013, April). Gdelt: Global data on events, location, and tone, 1979–2012. In ISA annual convention (Vol. 2, No. 4, pp. 1-49). Citeseer.

Khaund, T., Hussain, M.N., Shaik, M. & Agarwal, N. (2021). Telegram: Data collection, opportunities and challenges. In: Lossio-Ventura, J.A., Valverde-Rebaza, J.C., Diaz, E. & Alatrista-Salas, H. (eds) *Information Management and Big Data* , 513–526. Cham. Springer International Publishing.

Mazziotta M., Pareto A. (2020) *Gli indici sintetici*, Giappichelli Ed., Turin.

Raleigh, C., Kishi, R. & Linke, A. (2023) Political instability patterns are obscured by conflict dataset scope conditions, sources, and coding choices. *Humanit Soc Sci Commun* 10, 74. https://doi.org/10.1057/s41599-023-01559-4

Sundberg, Ralph, and Erik Melander. *Introducing the UCDP Georeferenced Event Dataset.* Journal of Peace Research, vol. 50, no. 4, July 2013, pp. 523–532, ucdp.uu.se/downloads/ged/ged172.pdf, https://doi.org/10.1177/0022343313484347.