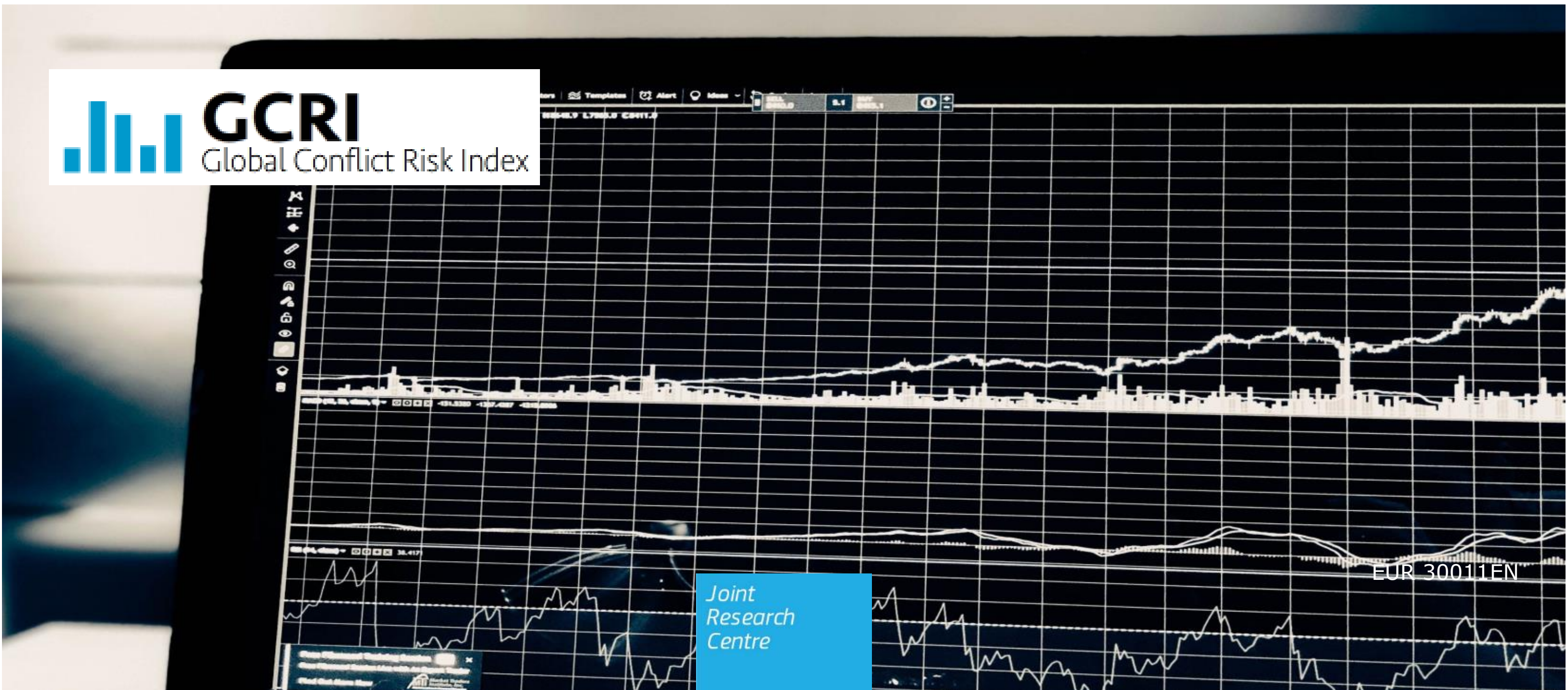
European Commission

# JRC TECHNICAL REPORTS

# Dynamic Global Conflict Risk Index

HALKIA Matina, FERRI Stefano, PAPAZOGLOU Michail, VAN DAMME Marie-Sophie, JENKINSON Gabriel, BAUMANN Kathrin-Manuela, THOMAKOS Dimitrios

2019




EUR 30011 EN

**Contact information**
Name: Matina Halkia
Address: Joint Research Centre, Via Enrico Fermi 2749, TP 480, 21027 Ispra (VA), Italy
Email: matina.halkia@ec.europa.eu
Tel.: +39 0332786242

**JRC Science Hub**
https://ec.europa.eu/jrc

JRC118701

EUR 30011 EN

PDF ISBN 978-92-76-14133-4 ISSN 1831-9424 doi:10.2760/846412

Luxembourg: Publications Office of the European Union, 2019

## Contents

## Acknowledgments

JRC would like to thank FPI.2, Marc Fiedrich and Giovanni Squadrito, as well as EEAS/CSDPCR.PRISM, Rene Van Nes, Gosia Sendrowska and Pavla Danisova for their unwavering support to the development of the Global Conflict Risk Index.

***Authors***

HALKIA Matina[1], FERRI Stefano[1], PAPAZOGLOU Michail[2], VAN DAMME Marie-Sophie[3], JEKINSON Gabriel[1], BAUMANN Kathrin Manuela[1], THOMAKOS Dimitrios[4]

[1] European Commission, Joint Research Centre (JRC), Ispra, Italy

[2] Unisystems S.A, Via Michelangelo Buonarroti 39, 20145 Milano, Italy

[3] Piksel Ltd Italian Branch, Via Breda 176, Milano (MI), Italy

[4] Unisystems S.A, Rue Edward Steichen 26, Luxembourg, LU L-2540

## Abstract

This report presents a dynamic model of the Global Conflict Risk Index (GCRI), a conflict risk model supporting the design of European Union's (EU) conflict prevention strategies developed by the Joint Research Centre (JRC) of the European Commission (EC) in collaboration with an expert panel of researchers and policy-makers.

While most studies as well as the regression GCRI measure conflict intensity by counting the number of causalities, the proposed dynamic GCRI integrates and identifies every stage of the conflict development or de-escalation in its entire complexity. The emergence of conflict related event data sets offers researchers new ways to quantify and predict conflicts through big data.

Using country-level actor-based event data sets that signal potential triggers to violent conflict such as demonstrations, strikes, or elections-related violence, the model aims at estimating the occurrence of material conflict events, under the assumption that an increase in material conflict events goes along with a decrease in material and verbal cooperation.

Three potential datasets are tested in this report following a political event coding classification: (i) the Global Data on Events Location and Tone (GDELT) project, (ii) the Integrated Crisis Early Warning System (ICEWS) Dataverse dataset and (iii) the Phoenix - Open Event Data Alliance (OEDA)-Phoenix Dataset.

The Artificial Intelligence (AI) methodology adopted to model the dynamic GCRI is built upon a Long-Short Term Memory (LSTM) Cell Recurrent Neural Network (RNN). These models are well-suited to classify, process and make predictions based on time series data and forecast near future events.

Besides this AI model, we have set up an early warning alarm system to signal abnormal social unrest upheavals. The dynamic GCRI, through the AI and early warning alarm, seems to be able to predict the materialization of a conflict on a monthly basis.

This new tool gives policy makers the possibility to observe the situation in a country on a monthly base, taking into consideration both the current and the predicted available information, and to implement preventive actions more rapidly to mitigate conflict exacerbations at an earlier stage of the conflict development cycle.

## 1. Introduction

Most of the studies that are trying to predict future conflicts consider the number of casualties as a proxy to measure conflict intensity, using datasets such as the Armed Conflict Location & Event Data (ACLED) and Uppsala Conflict Data Program/Peace Research Institute Oslo (UCDP/PRIO) (Hegre et al., 2013; Szayna et al., 2017; Halkia et al., 2019). While a certain degree of correlation between conflict intensity and the number of battle casualties may be identified, it does not consider conflict development stages (Qiao et al., 2017) and the complexity of events like demonstrations, protests, election violence, or even tension relief events such as diplomatic cooperation.

The Peace and Stability team at the European Commission's Joint Research Centre proposes a new type of Global Conflict Risk Index (GCRI), which unlike the original GCRI, would integrate and disentangle every stage of the conflict development or de-escalation cycle.

Using the proprietary GCRI in combination with big data and neural networks, the team is working to produce a model that monitors conflict risk worldwide in real time, and follows the movements and activities of terrorists, insurgents, illegal trade, and organized crime syndicates.

In order to do so, however, the model must draw from a comprehensive, accurate, and reliable big data source.

In this paper, we test three different available news media datasets based on a political event coding classification. At the moment, two different event-coding classifications exist, the Conflict and Mediation Event Observations Event and Actor Codebook (CAMEO) classification and the Political Language Ontology for Verifiable Event Records (PLOVER)[5]. Due to data availability, we rely on the CAMEO framework for classifying event data in four primary classes (called QuadClass): verbal cooperation (Q1), material cooperation (Q2), verbal conflict (Q3), and material conflict (Q4). Additionally, 20 major subcategories are identified, which are then divided into several sections, so as to create a detailed classification scale (Schrodt, 2012), following the typical evolution stages of social unrest; appeal, accusation, refuse, escalation, and finally protests/riots (Qiao et al., 2017). Most of the social unrest events initially start as a demonstration to the public or the government, and often escalate afterwards into general chaos, resulting in violent, riots, sabotage, and other forms of crime and social disorder.

The datasets which are tested in this report are: (i) the Global Data on Events Location and Tone (GDELT) project, (ii) the Integrated Crisis Early Warning System (ICEWS) Dataverse dataset and (iii) the Phoenix - Open Event Data Alliance (OEDA)-Phoenix Dataset.

To our knowledge, one paper has used the GDELT data to measure conflict intensity (Levin, Ali, & Crandall, 2018). In their article, however, Levin, Ali and Crandall only consider the monthly time series of the absolute number of events occurring in the CAMEO Q4 subclasses or take a MaxMin normalization over their time series (i.e. normalization between zero and one based on the minimum and maximum values in the time series of each country). The event-based GCRI goes one step further in its conflict analysis, as it additionally evaluates the increase in the proportion of the various QuadClasses over the total number of events. Although the absolute and normalized number of events under each CAMEO QuadClass are giving important information, the conflict cycle development is better captured in its entire complexity when considering the analysis in proportions.

The Artificial Intelligence (AI) methodology adopted to model the dynamic GCRI is built upon a Long-Short Term Memory (LSTM) Cell Recurrent Neural Network (RNN). This model treats the social unrest event prediction as a sequence classification problem and is well-suited to classify, process, and make predictions based on time series data and forecast near future events.

[5] PLOVER is intended to replace the earlier CAMEO classification of events and adds a new category for criminal behaviour. Unfortunately, it is not yet available to potential users.

Besides this AI model, we have set up an early warning alarm system to signal abnormal social unrest upheavals.

Although the dynamic GCRI, through the AI and early warning alarm, is able to predict the materialization of a conflict on a monthly basis, we should further assess the different signals and various peaks as there is no correct model as yet to predict social unrest. This new tool gives policy makers the possibility to implement preventive actions more rapidly to mitigate conflict exacerbations at an earlier stage of the conflict development cycle.

Section 2 presents the various datasets and Section 3 explains the model and methodology proposed for the dynamic GCRI, whereas Section 4 presents the results. Eventually, Section 5 concludes the report with a feasibility discussion.

## 2. Event Datasets Using the Conflict and Mediation Event Observations Event and Actor Codebook (CAMEO) Classification

The Global Database of Events, Language, and Tone Project (GDELT), the Integrated Crisis Early Warning System (ICEWS) Dataverse (ICEWS)[6] and the Phoenix - Open Event Data Alliance (OEDA) Dataset are arguably the largest event data collections in social science at the moment, which gather huge amounts of news items from various sources around the world. During their brief existence they have been among the most influential datasets in terms of their impact on academic research and policy advice.

Due to data availability, we classify the event news databases with the Conflict and Mediation Event Observations (CAMEO) code, a framework for classifying event data in four primary classes (called QuadClass): verbal cooperation (Q1), material cooperation (Q2), verbal conflict (Q3), and material conflict (Q4). In addition, the 20 major subcategories are then divided into several sections, so as to create a detailed classification scale (Schrodt, 2012) following the typical evolution stages of social unrest; appeal, accusation, refuse, escalation, and finally protests/riots (Qiao et al., 2017). In order to fulfil the purposes of this report, we investigate the use of these three news media datasets as possible inputs for a dynamic GCRI[7].

### 2.1. Global Database of Events, Language, and Tone (GDELT) Project

The GDELT project is an on-going attempt to monitor print, broadcast, and web news media in over 100 languages from all over the world, so as to continuously be updated on breaking news worldwide. According to their website the 'GDELT moves beyond the focus of the Western media towards a far more global perspective on what is happening and how the world is feeling about it' (GDELT, 2019).

It has been created by Kalev Leetaru and Georgetown University in cooperation with Google, BBC Monitoring, National Academies Keck Futures Program, LexisNExis Group, JSTOR, DTIC, and the Internet Archive (GDELT, 2019).

The GDELT Event Database records over 300 categories of physical activities around the world, from riots and protests to peace appeals and diplomatic exchanges. This means that the existent database consists of more than 2.5TB of available information per year. In absolute numbers, it contains more than a quarter billion of records. Supported by Google Jigsaw, the platform includes data from 1979 till present, and it is updated every 15 minutes as of April 1, 2013.

It is worth emphasizing that the GDELT dataset is updated every 15 minutes, meaning that there is a continuous flow of records inserted in the database. That tremendous amount of event records - greater than any other event dataset - opens up a new perspective in this research area.

So far, works rarely aim at utilizing GDELT to make predictions about social unrest and only a few scholars have conducted predictions using GDELT. Alikhani attempted to use GDELT with linear regressions in 2014, while Yonamine studied the dataset for time series forecasting in 2013. More recent papers use GDELT for frequent subgraphs mining (Qiao et al., 2017) and apply artificial intelligence (Smith et al., 2018).

### 2.2. Integrated Crisis Early Warning System (ICEWS) Dataverse

The ICEWS program is a comprehensive, integrated, automated, generalizable, and validated system to monitor, assess, and forecast national, sub-national, and internal crises (Lockheed Martin, 2019).

It is an early warning system designed to help US policy analysts predict a variety of international crises to which the US might have to respond (Ward et al., 2013).

[6] ICEWS is not updated since June, 2019.

[7] For the purposes of this report we investigate the use of three news media datasets as input datasets. There are more available datasets but there are also major limitations that we cannot overcome i.e. limited time series (Phoenix_RT from Oct. 2017 to today).

The ICEWS events are automatically identified and extracted from news articles by the BBN ACCENT event coder, following the above-mentioned classification (i.e. CAMEO). These events consist of a source actor, an event type (according to the CAMEO taxonomy of events), and a target actor.

As stated on their official webpage, ICEWS is 'unable to provide the story text from which events are extracted due to copyright reasons' and is 'also unable to provide URLs for the stories' as they 'purchase the content and receive it directly as opposed to via a website.' (ICEWS Automated Daily Event Data - Integrated Crisis Early Warning System (ICEWS) Dataverse, 2019). The ICEWS dataset has been discussed in the conflict prediction research literature (Tikuisis, Carment & Samy, 2013; Ward et al., 2013; Yonamine, 2013) as well as in relation to the coding of political events (Schrodt & Van Brackle, 2013).

## 2.3. Cline Center Historical Phoenix Event Data (Phoenix)

The Cline Center Historical Phoenix Event Data is a daily updated event dataset with a global coverage, developed with the support of Linowes Fellow and Faculty Affiliate Prof. Dov Cohen and help from academic and private sector collaborators in the Open Event Data Alliance (OEDA). The OEDA is a consortium of for-profit organizations, not-for-profit organizations, and individuals committed to facilitating the development, adoption, and analysis of social and political event data (OEDA Datasets, 2019).

It is updated daily and documents the agents, locations, and issues at stake in a wide variety of conflict, cooperation and communicative events in the CAMEO ontology framework.

The Phoenix dataset is coded using the PETRARCH coder, a Python-based program using full parsing of news text, replacing the earlier and less sophisticated TABARI software (Schrodt, 2001). Phoenix's software, like GDELT's, is designed to extract major international political and social events from wire service reports (Halterman & Beieler, n.d.).

## 2.4. Data Comparison and Evaluation

All the above-mentioned datasets are based on the CAMEO classification, which is crucial for the modelling phase of this analysis. In Table 1, one can see the time, geographical and language coverage of each dataset as well as their updated frequency.

**Table 1.** Event-based datasets coverage

| **Dataset** | **Time coverage** | **Geographical coverage** | **Language Coverage** | **Update frequency** | **Source availability** |
|---|---|---|---|---|---|
| **GDELT** | 1979-now | Global | 100 languages | 15 min | Publicly available source URL |
| **ICEWS** | 1995-June 2019 | Global | English, Spanish, Portuguese and Arabic | Monthly, daily as of October 2018 | Unavailable source URL due to copyright reason |
| **OEDA-Phoenix** | 1945-2005<br>1995-2004<br>1979-2015 | Global | English | Daily | Unavailable source URL (only the news provider name) |

There are limitations to all the three datasets. The major limitation of the GDELT project is the fact that the monitoring is based on simple keywords, which may lead to a collection of irrelevant records (noise). However, as the source URL is given, we can undertake sample validation tests[8]. The ICEWS program scans news on a daily basis solely after October 2018 and on a monthly basis since 1995. Before that date, there is no data available for this particular dataset. Another limitation is that we are not able to identify the source of the records and hence, we cannot validate the output of the dataset. According to Ward et al. (2013), who compared the GDELT and ICEWS data sets, 'it is clear that both databases pick up major events remarkably well. The volume of GDELT data is very much larger than the corresponding ICEWS data [...] It seems clear, however, that GDELT over-states the number of events by a substantial margin, but ICEWS misses some events as well.'

The Phoenix-OEDA dataset uses 14 million articles but only from the following news sources: the New York Times (NYT), the British Broadcasting Corporation's (BBC) Summary of World Broadcasts (SWB), and the Central Intelligence Agency's (CIA) Foreign Broadcast Information Service (FBIS) (Althaus et al., 2017). Each news source has a different time coverage, generating three databases instead of one. Moreover, the Phoenix-OEDA data has not published data after 2015.

The CAMEO classifies these event data bases in four primary classes: verbal cooperation (Q1), material cooperation (Q2), verbal conflict (Q3), and material conflict (Q4). Each of the four classes is further subdivided in five subcategories, which consist of several sections[9]. In this way, one creates a detailed classification scale (Schrodt, 2012) following the typical evolution stages of social unrest: appeal, accusation, refuse, escalation, and finally protests/riots (Qiao et al., 2017).

Most of the social unrest events initially start as a demonstration to the public or the government, and often escalate afterwards into general chaos, resulting in riots, sabotage, and other forms of crime and social disorder.

Eventually, the social unrest event prediction is formulated as a sequence classification problem that identifies any possible sequence or stage of events that potentially lead to social unrest.

The proposed event-based model built upon the CAMEO classification to predict social unrest assumes that an increase in material and verbal conflict events goes along with a decrease in material and verbal cooperation.

Observing that there is a general increasing trend in the amount of news articles reported in Q3 and/or Q4 with respect to the total number of articles, the model is able to measure an increase in conflict related tensions.

In order to evaluate how appropriate each one of the data sets is in predicting conflict, the following subsections make a visual assessment on how CAMEO classifies the news events in each of the databases respectively and contrast the results with ACLED data on the number of fatalities when available[10] for Libya (2.4.1.), Sudan (2.4.2.), Egypt (2.4.3.), Maldives (2.4.4.) and Nicaragua (2.4.5.). We selected these cases as all of them are typical cases where the existing models were unable to predict the escalation of violence[11].

### 2.4.1. Case Study 1: Libya

A highly interesting case study is Libya as almost none of the current models were able to predict the start of Arab Spring in the country in February 2011. Both the GDELT and the ICEWS datasets show an increase in the proportion of events categorized as material conflict, Q4, in February 2011 (Figure 1 and

---

[8] For more details about the GDELT sample validation please see Appendix I. The GDELT dataset qualitative validation.

[9] For more information about the subdivision of the CAMEO, please see Appendix III.

[10] While ACLED is widely used, its time and geographical coverage is limited. It covers Africa since 1997, Asia since 2010, the Middle East since 2016, and Europe since 2018.

[11] In particular, Libya and Egypt were selected to check the model's capacity to predict the start of the Arab spring. Additionally, Sudan was selected as the level of violence has been high for a long period of time. Finally, Nicaragua and Maldives are countries where ACLED data is not available, however, both countries have an increasing geo-strategic importance.

*Figure* **2** on the next page), while an increase in the total amount of events reported can be observed in the OEDA-phoenix, but the increase in the events mentioned as material conflict is not significant (see Figure 3). The red vertical line in the following plots refers to the start of the Arab Spring in Libya in February 2011.

**Figure 1.** Events per QuadClass in Libya, monthly aggregation (1989-2019 GDELT)

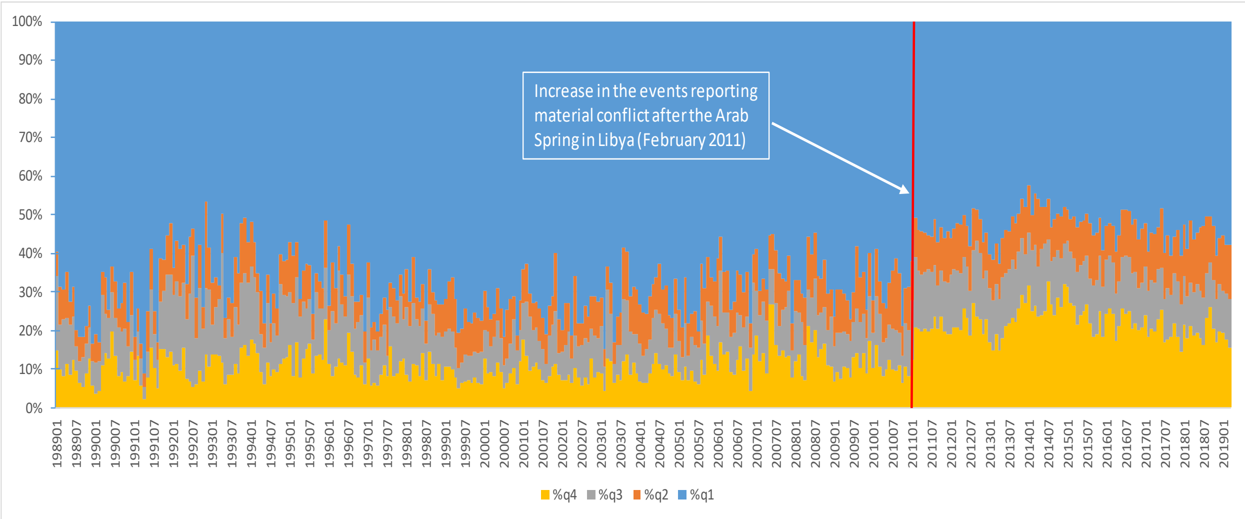



**Figure 2.** Events per QuadClass in Libya, monthly aggregation (1995-2019 ICEWS)

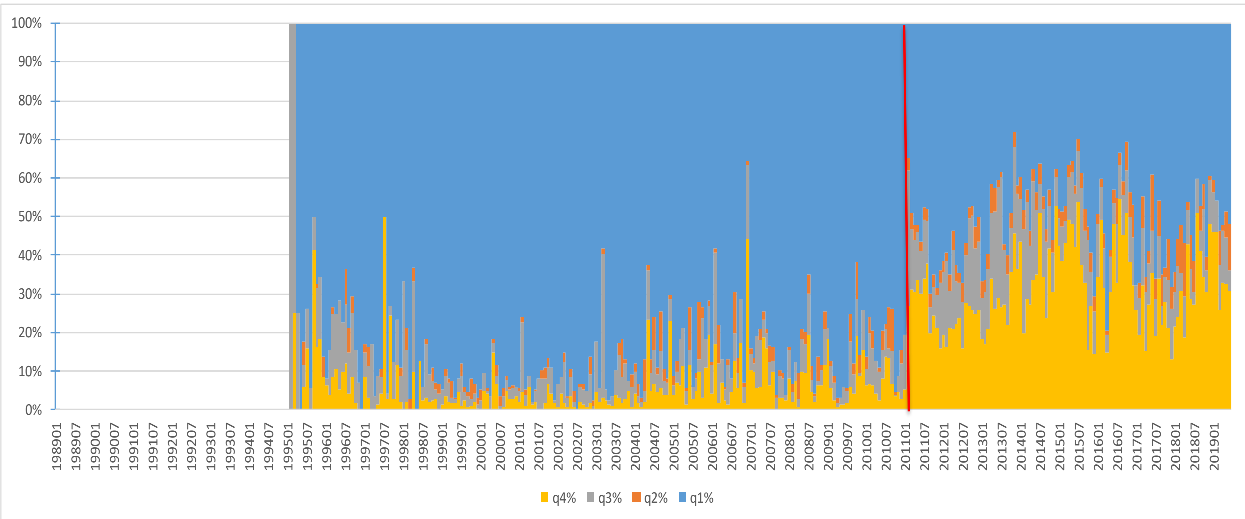
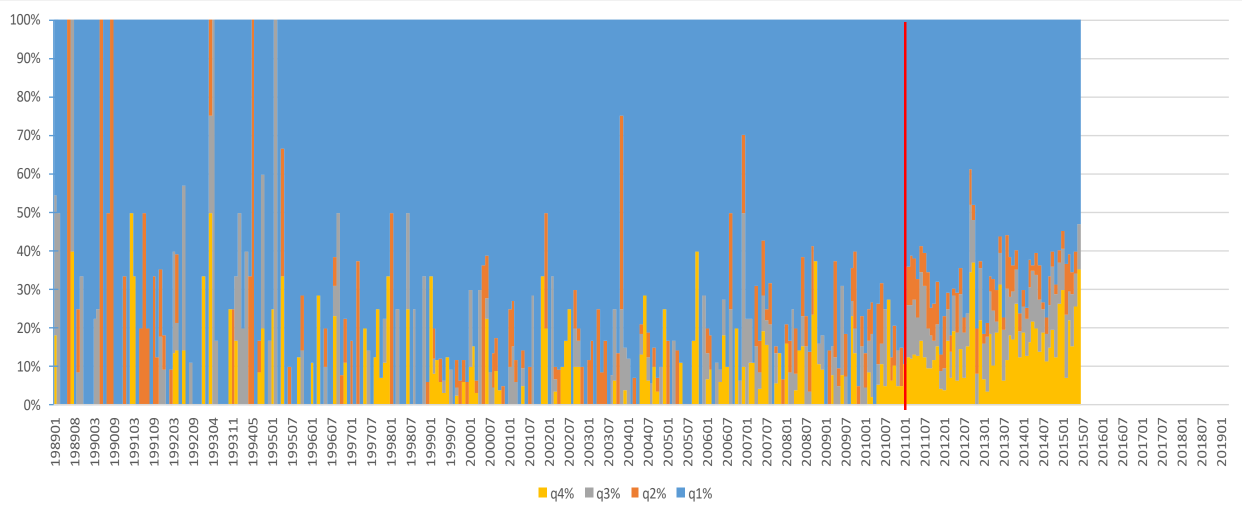
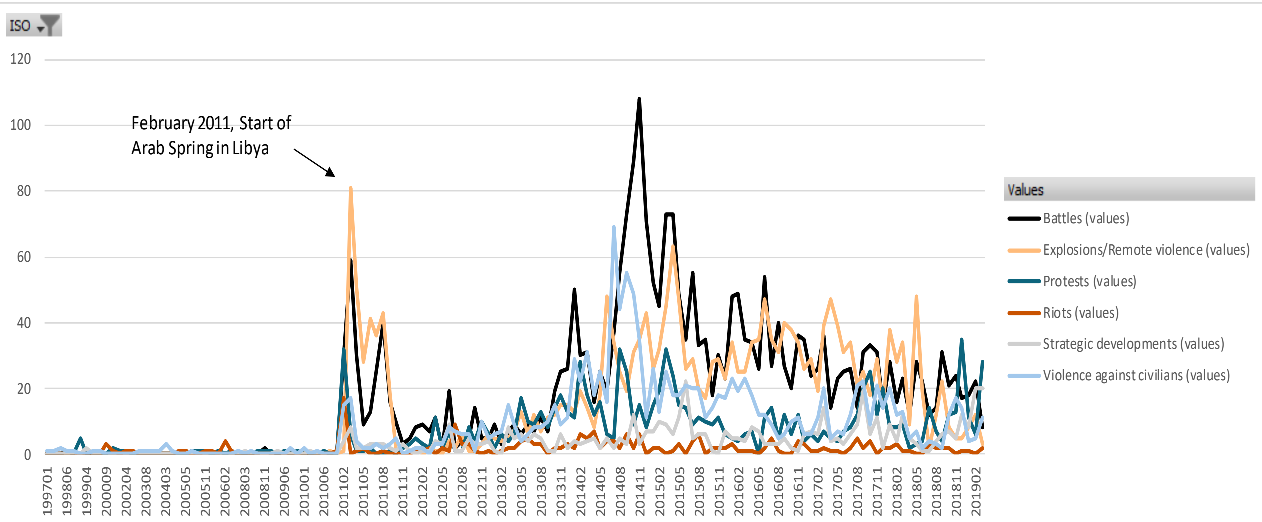


**Figure 3.** Events per QuadClass in Libya, monthly aggregation (1989-2019 OEDA)



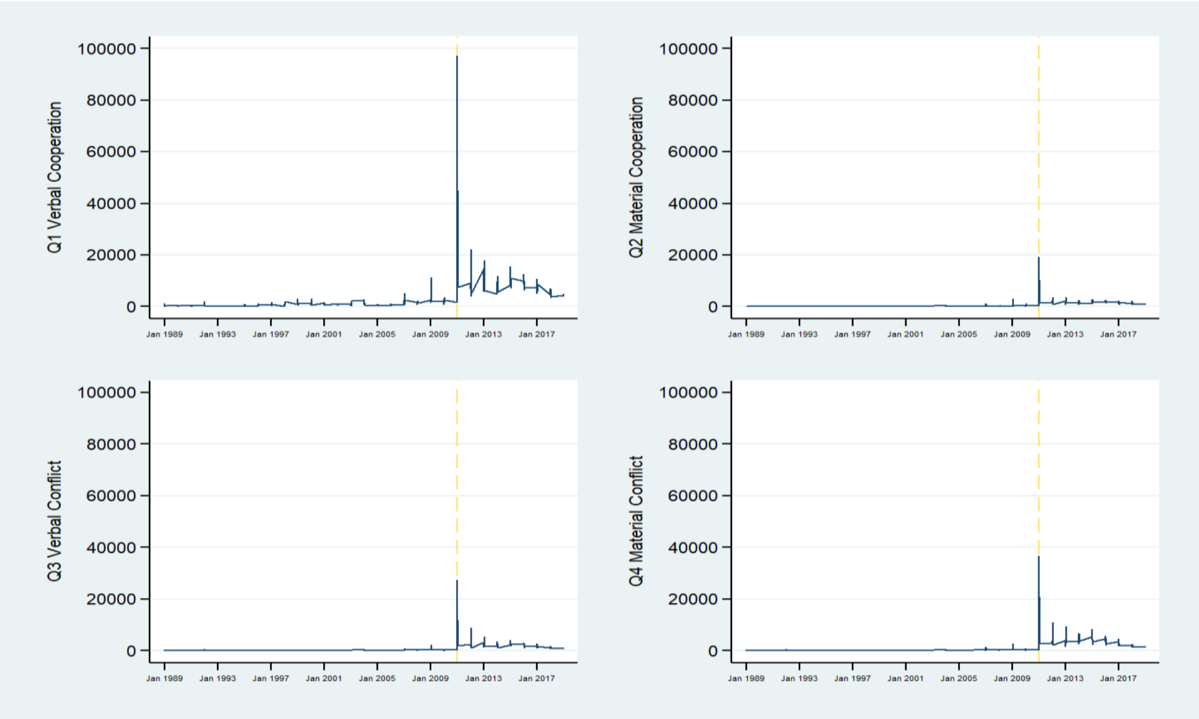

In Figure 4, a major sudden peak can be observed in the reported number of fatalities by ACLED at the beginning of February 2011.

**Figure 4.** Number of fatalities in Libya, monthly aggregation (1997-2019 ACLED)




In Figure 5, an overall increase in the number of news events included in the GDELT dataset and classified under Verbal Cooperation (Q1), Material Cooperation (Q2), Verbal Conflict (Q3), and Material Conflict (Q4) is reported. However, the biggest increase in news is found under Q1 and not under Q3 or Q4. From this information, we solely see that more cooperation than conflict events had taken place in Libya in February 2011.

**Figure 5.** Absolute monthly news frequency per CAMEO QuadClass in Libya - GDELT (Jan 1989 - March 2019)




Normalizing our monthly time series as in Levin, Ali and Crandall (2018) between zero and one, based on the minimum and maximum values, we can identify global and local maxima (see Figure 6). As it can be seen in Figure 6, all four QuadClasses reach a global maximum in February 2011 in Libya. This helps us see whether the increase in the news events under each QuadClass is more or less important over time with respect to preceding events, however, it does not depict the start of the Arab Spring in Libya; the increase in the normalized cooperation events is similar to the increase in normalized conflict events.

If we additionally look at the increase or decrease per QuadClass in the proportion of events classified under Q1, Q2, Q3, and Q4 of the total reported news in Libya before and after February 2011 (see Figure 7), we can observe a significant increase in Q3 and Q4 whereas Q1 and Q2 decreased significantly in February 2011. In this way, we can have a better idea of how a conflict is escalating, stagnating or de-escalating from month to month.

**Figure 6.** MaxMin normalized monthly news per CAMEO QuadClass in Libya - GDELT (Jan 1989 - March 2019)

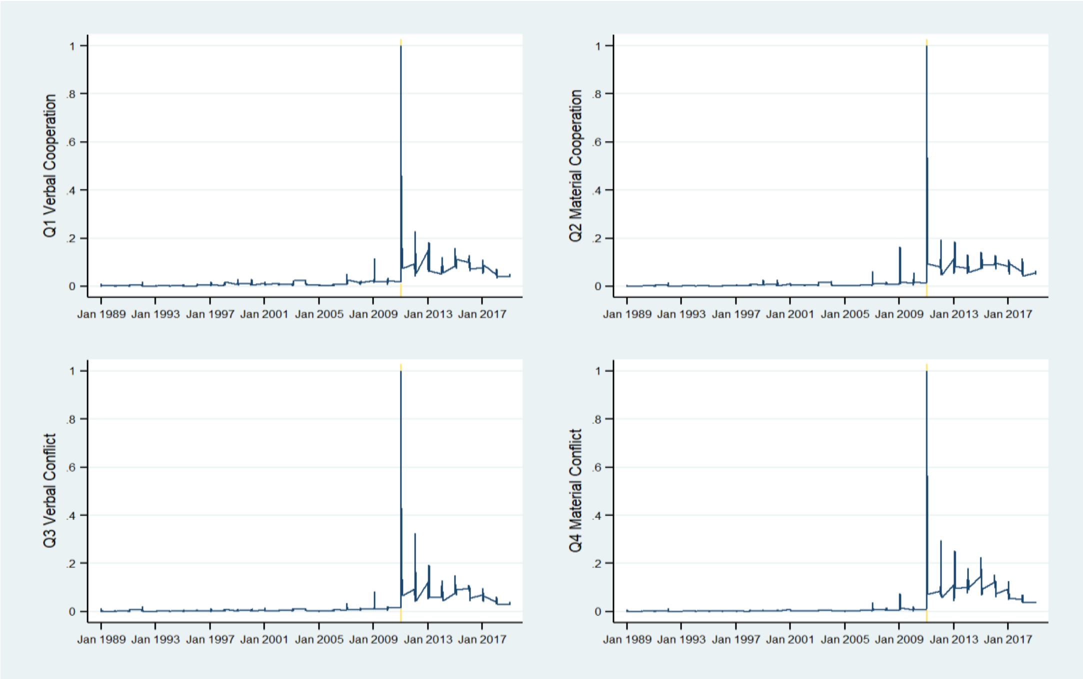
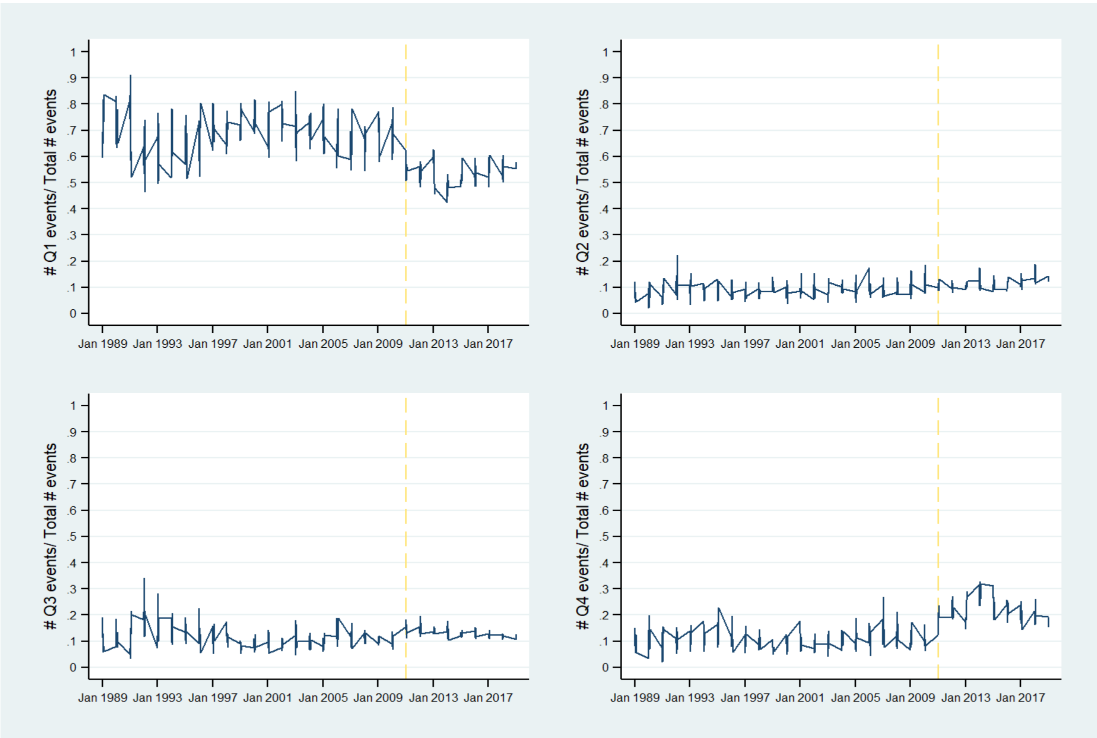
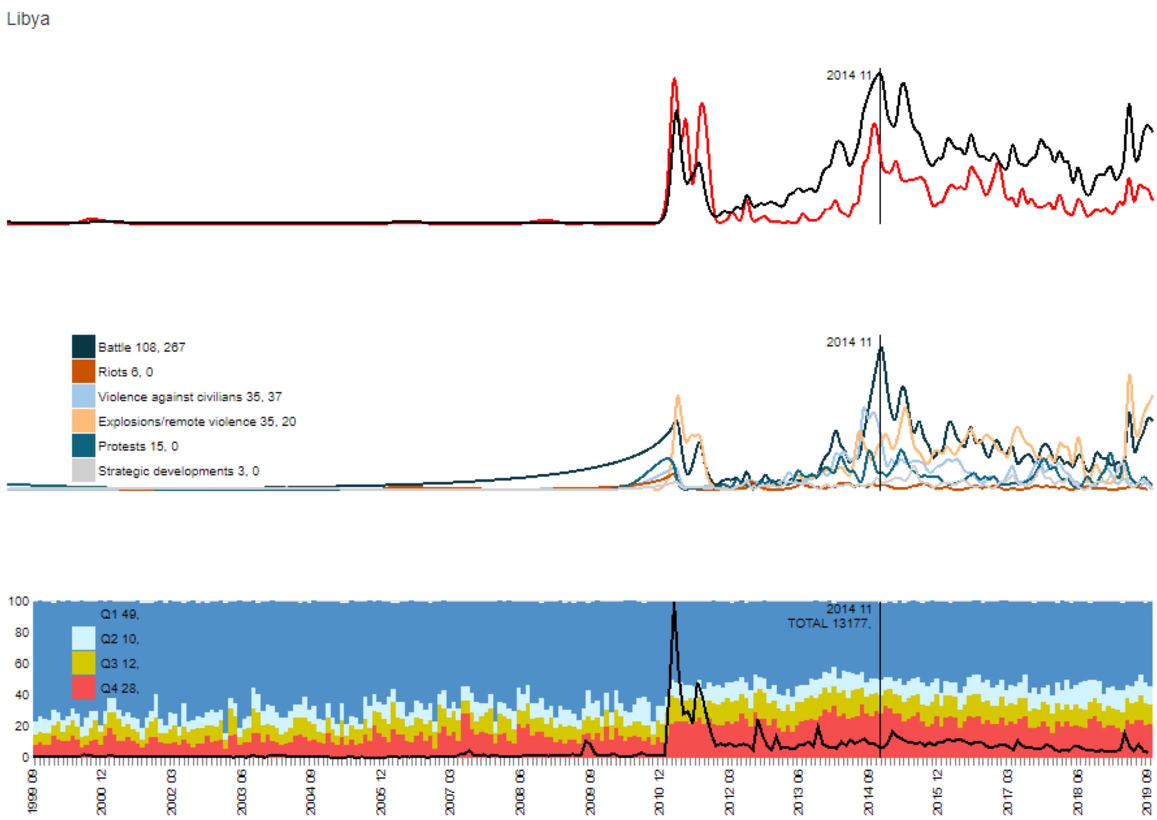



**Figure 7.** Proportion of reported events per QuadClass in Libya - GDELT (Jan 1989 - March 2019)



In the following figures, one can observe that ACLED reports another major peak in November, 2014[12]. However, we are not able to confirm it using other event datasets, i.e. GDELT.

**Figure 8.** ACLED and GDELT comparison for Libya. The black trend in the top chart represents the total number of events mentioned in ACLED while the red one the total number of deaths.




As it can be seen in the top chart, there is a major peak for both the total number of events and the number of deaths mentioned in ACLED. On the other hand, news datasets like GDELT and ICEWS did not report this increase.

### 2.4.2. Case Study 2: Sudan

The second case study analyses the Sudanese protests in 2018-2019. On December 19, 2018, people took to the streets to demonstrate against rising living costs and the deterioration of economic conditions at all levels of society. The protests quickly turned from requests for urgent economic reforms into demands for the removal of President Omar al-Bashir. Both GDELT and ICEWS show an increase in the amount of events categorized as material conflict, Q4, in December 2018 (see Figure 9 and Figure 10). The red vertical line in the following plots refers to the start of the Sudanese protests in December 2018.

[12] In November 2014, Libya's Supreme Court declared the internationally recognized parliament of Libya as unconstitutional. In the aftermath, battles between loyalist forces and militias killed nearly 400 people in three weeks (Laessing et al., 2017).

**Figure 9.** Events per QuadClass in Sudan, monthly aggregation (1989- 2019 GDELT)

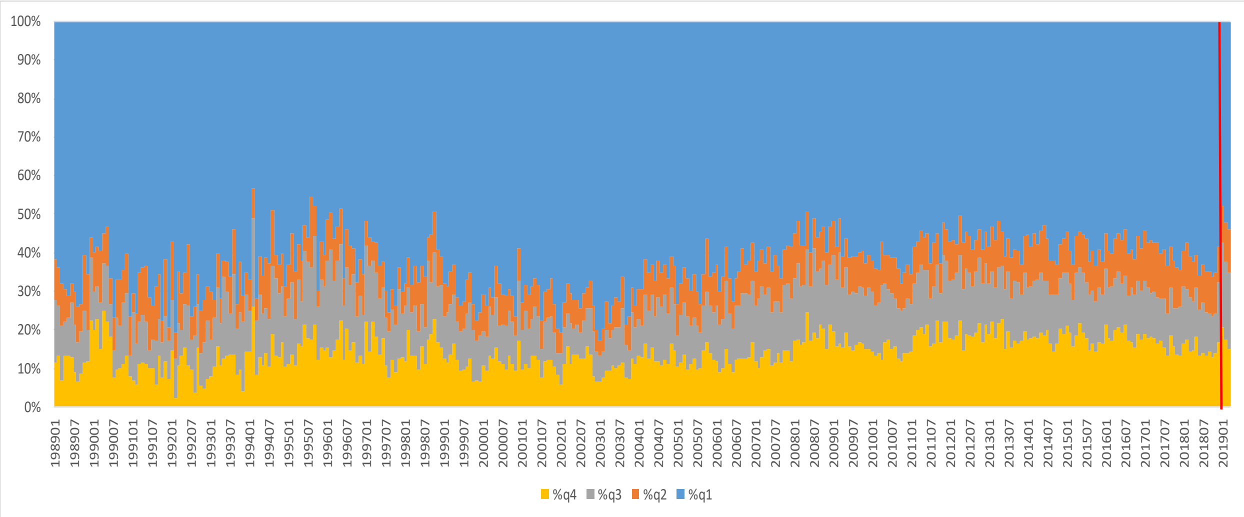
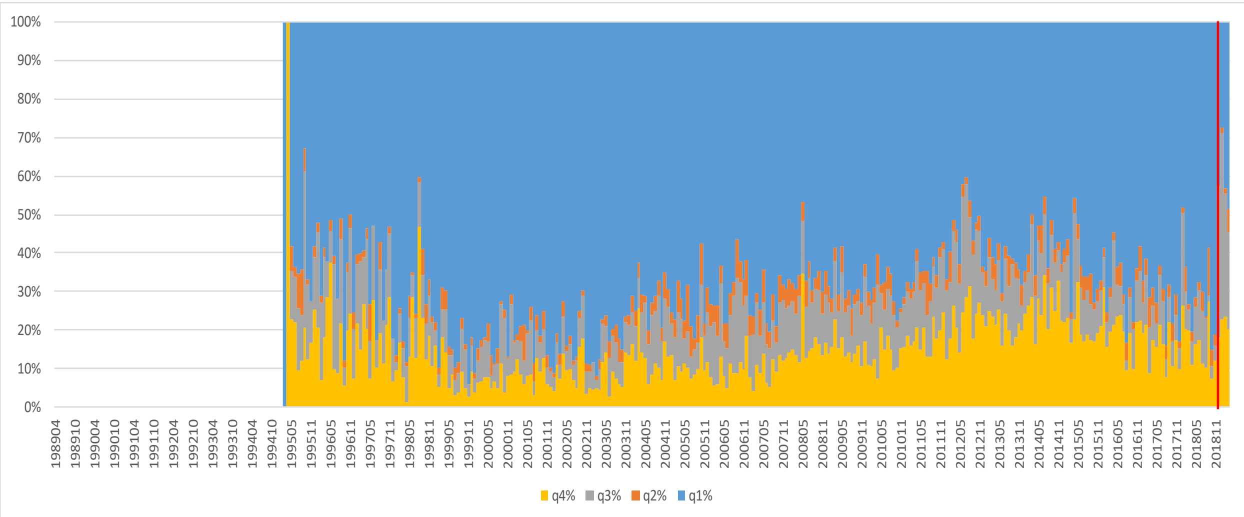



**Figure 10.** Events per QuadClass in Sudan, monthly aggregation (1995-2019 ICEWS)




Given that the OEDA-Phoenix database does not provide data after 2015, we are not able to evaluate and visualize it (see Figure 11). Comparing these observations with ACLED fatalities (see Figure 12), one can verify an important decrease in material collaboration translating in an increase in social tension in the same period of December 2018.

**Figure 11.** Events per QuadClass in Sudan, monthly aggregation (1989-2019 OEDA)

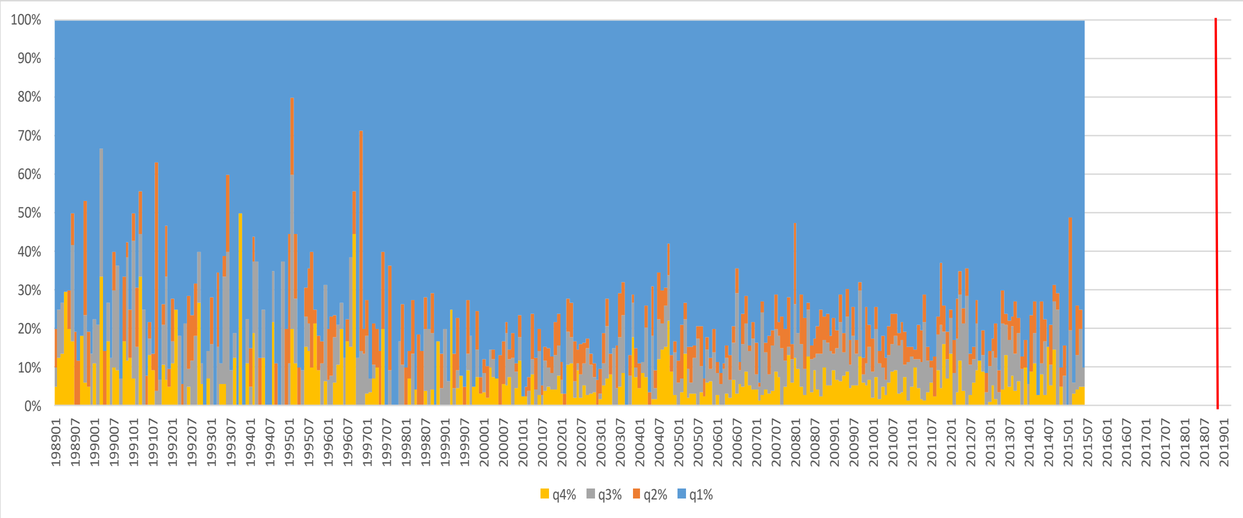
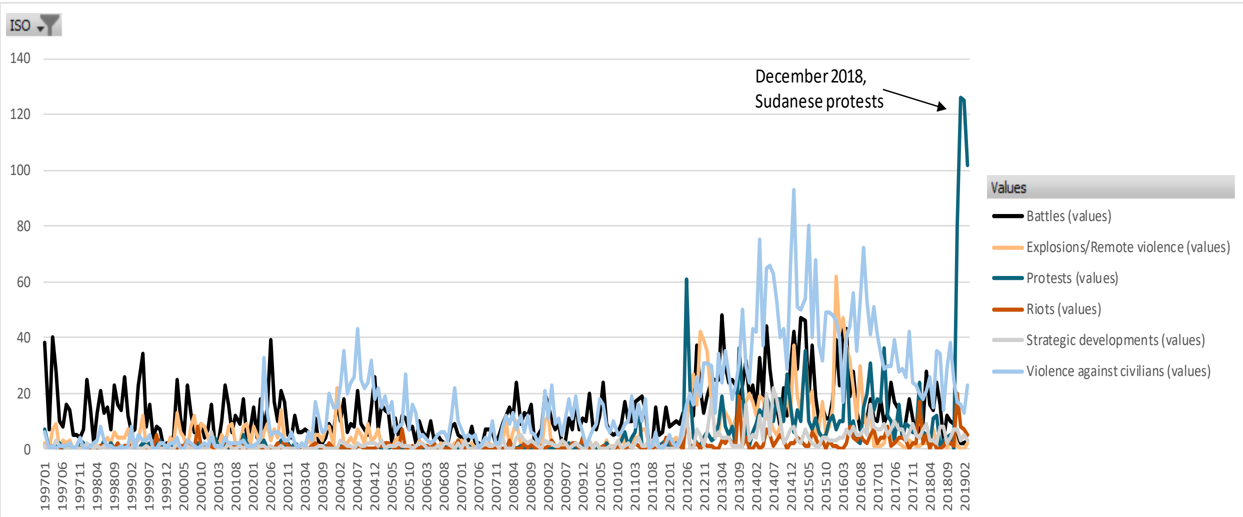
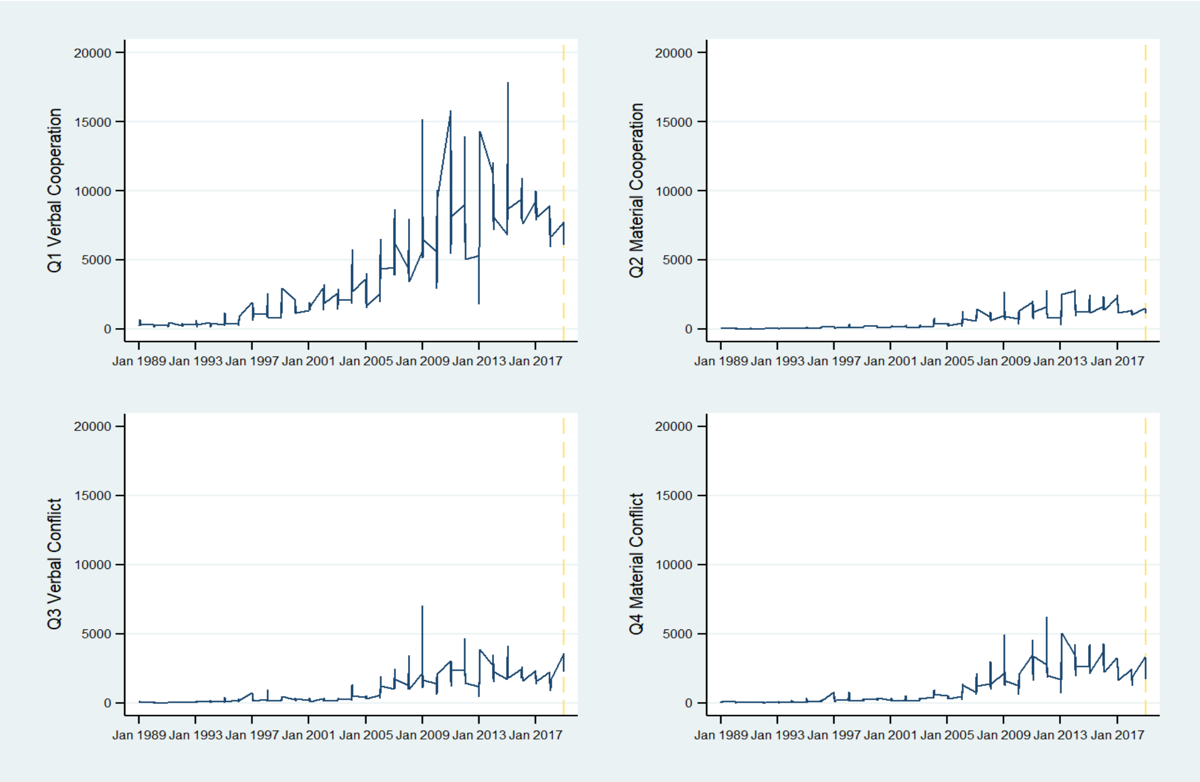



**Figure 12.** Number of fatalities in Sudan, monthly aggregation (1997-2019 ACLED)




In Figure 13, a local increase in the number of news events classified under Verbal Cooperation (Q1), Material Cooperation (Q2), Verbal Conflict (Q3), and Material Conflict (Q4) is reported. Notice that the biggest increase in news for January 2019 is found under Q1 and not under Q3 or Q4. From this information, we solely see that in January 2019, more cooperation than conflict events have taken place in Sudan.

**Figure 13.** Absolute monthly news frequency per CAMEO QuadClass in Sudan - GDELT (Jan 1989 - March 2019)




Normalizing our monthly time series as in Levin, Ali and Crandall (2018) between zero and one based on the minimum and maximum values, we can identify global and local maxima (see Figure 14). In Figure 14, all four QuadClasses reach a local maximum in January 2019 in Sudan.[13] This helps us see whether the increase in the news events under each QuadClass is more or less important over time with respect to preceding events, however, it does not reflect the Sudanese protests from January 2019, as the increase in the normalized cooperation events is similar to the increase in normalized conflict events.

If we additionally look at the increase or decrease per QuadClass in the proportion of events classified under Q1, Q2, Q3, and Q4 of the total reported news in Sudan before and after January 2019 (see Figure 15), we can observe that Q3 and Q4 increased and reached a peak in January 2019, whereas Q1 and Q2 decreased significantly. In this way, we can have a better picture of how a conflict is escalating, stagnating or de-escalating from month to month.

[13] In January 2019, more than 800 protesters were arrested and at least 25 people killed in clashed between anti-government protesters and state forces.

**Figure 14.** MaxMin normalized monthly news per CAMEO QuadClass in Sudan - GDELT (Jan 1989 - March 2019)

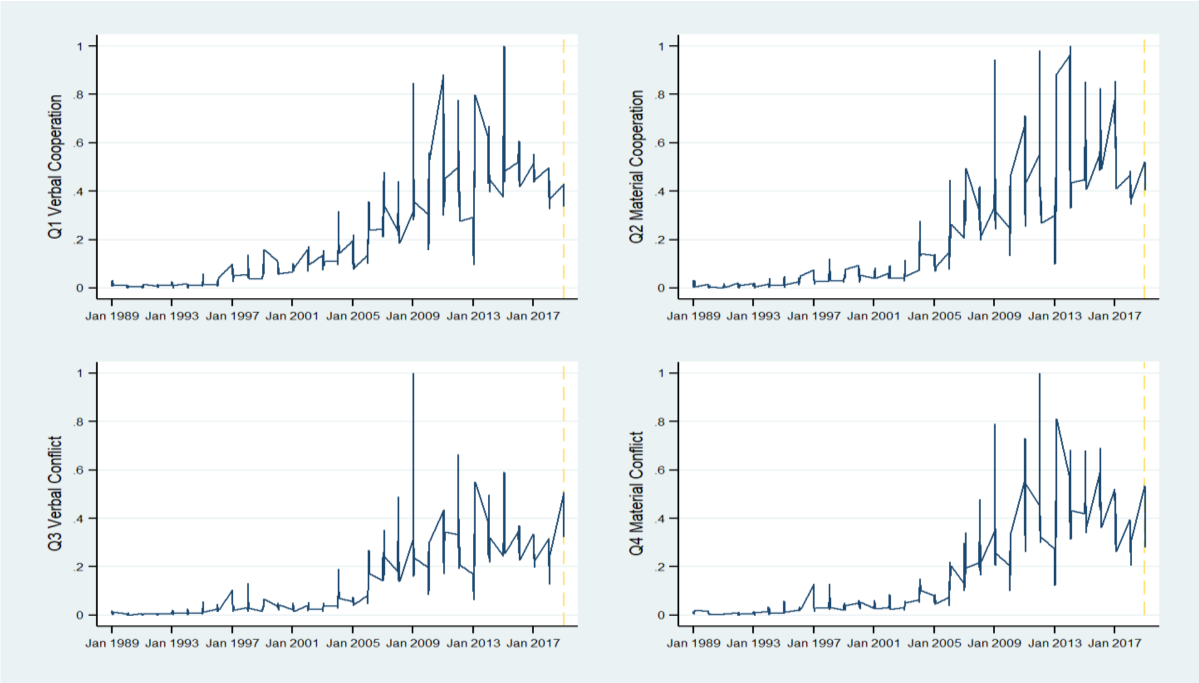
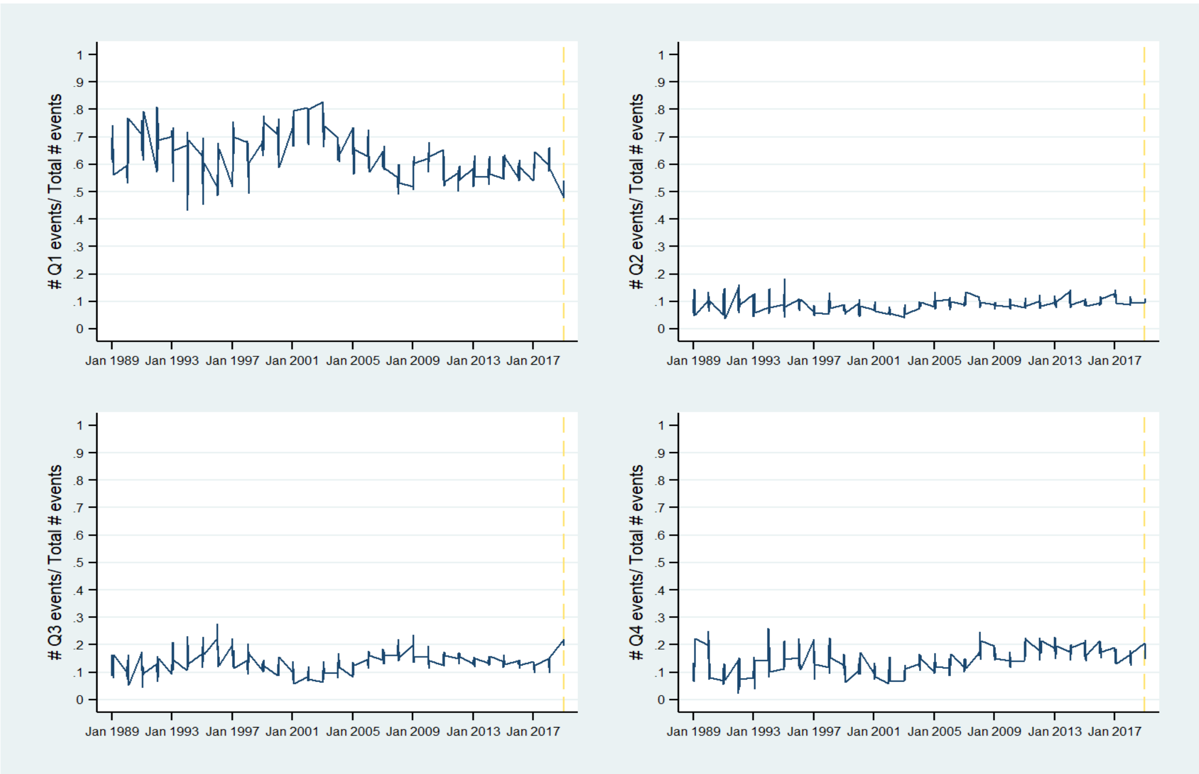
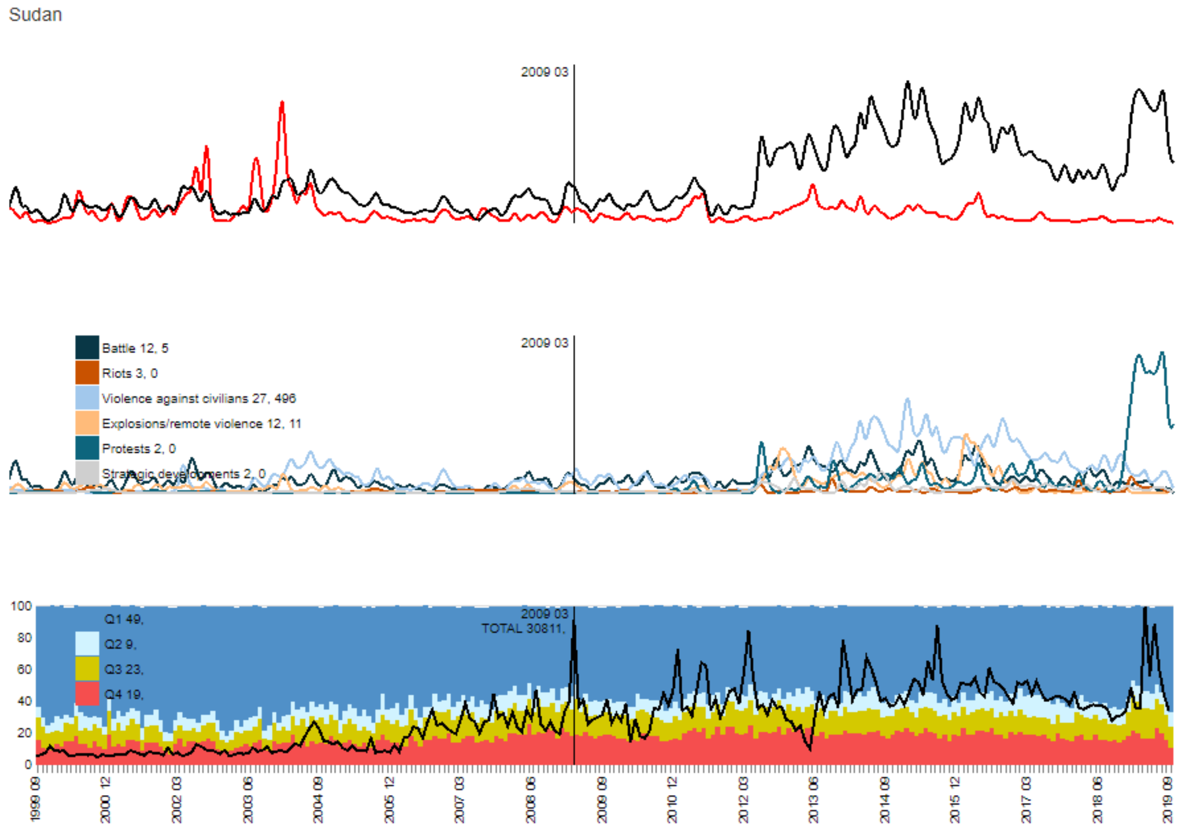



**Figure 15.** Proportion of reported events per QuadClass in Sudan - GDELT (Jan 1989 -March 2019)




On the other hand, there is a major peak in the GDELT dataset in March 2009 which is absent in the ACLED data. Indeed, in March 2009, The International Criminal Court in The Hague issued an arrest warrant for President Bashir on charges of war crimes and crimes against humanity in Darfur.

**Figure 16.** ACLED and GDELT comparison for Sudan. The black trend in the top chart represents the total number of events mentioned in ACLED while the red one the total number of deaths.




### 2.4.3. Case Study 3: Egypt

Next, we examine the case of the Arab Spring in Egypt, which started in January 2011. In the GDELT, there is a visible increase in the amount of events categorized as material conflict, Q4 (see Figure 17), and the Arab Spring is also clearly identifiable in the ICEWS (see Figure 18), but no conclusion can be drawn from the OEDA-Phoenix dataset (see Figure 19) as there is no clear picture of the situation at that point in time. The red vertical line in the following plots refers to the start of the Arab Spring in Egypt in January 2011.

**Figure 17.** Events per QuadClass in Egypt, monthly aggregation (1989- 2019 GDELT)

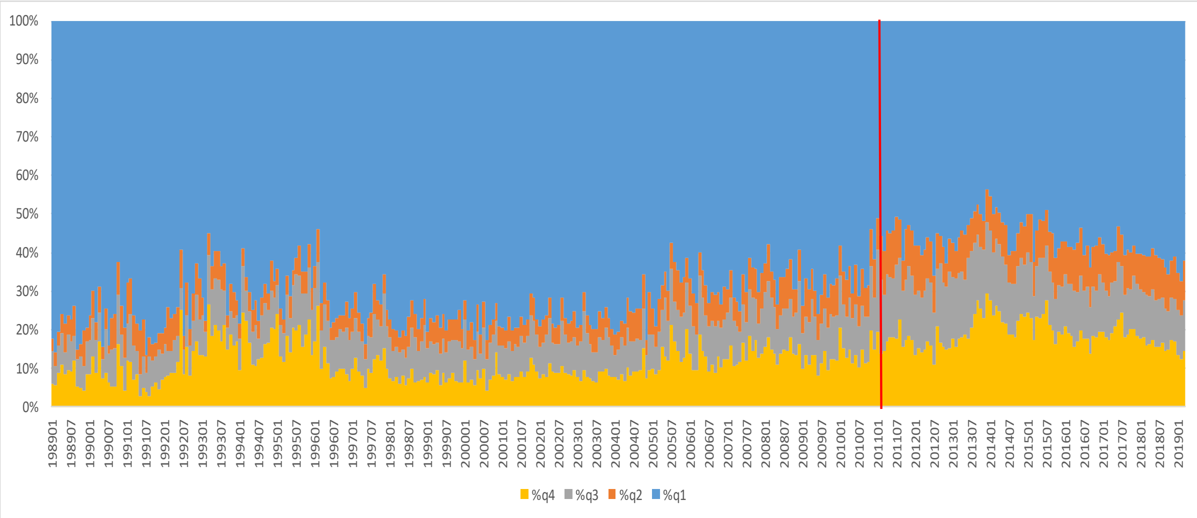



**Figure 18.** Events per QuadClass in Egypt, monthly aggregation (1995-2019 ICEWS)

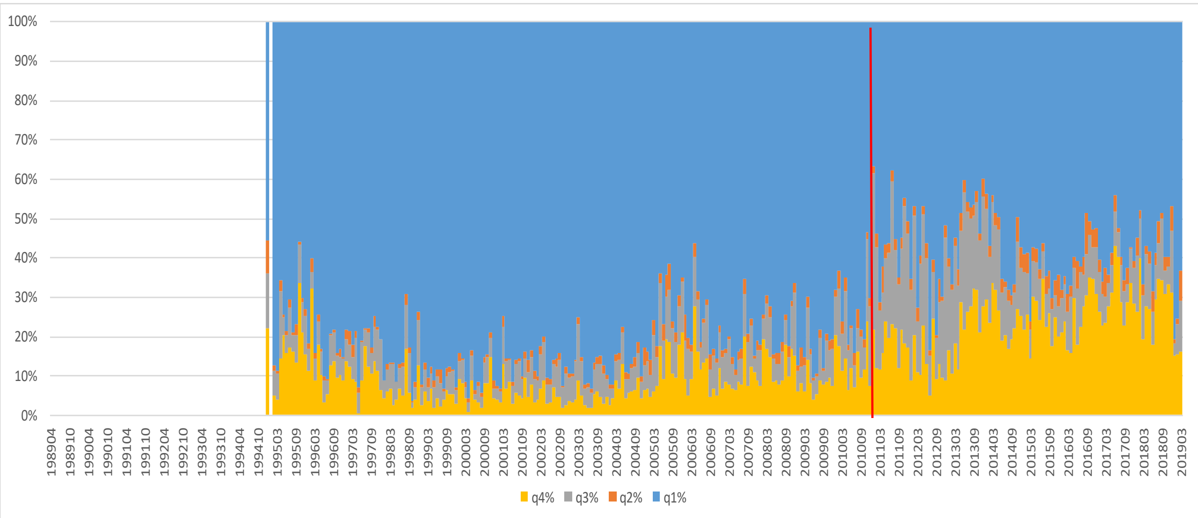
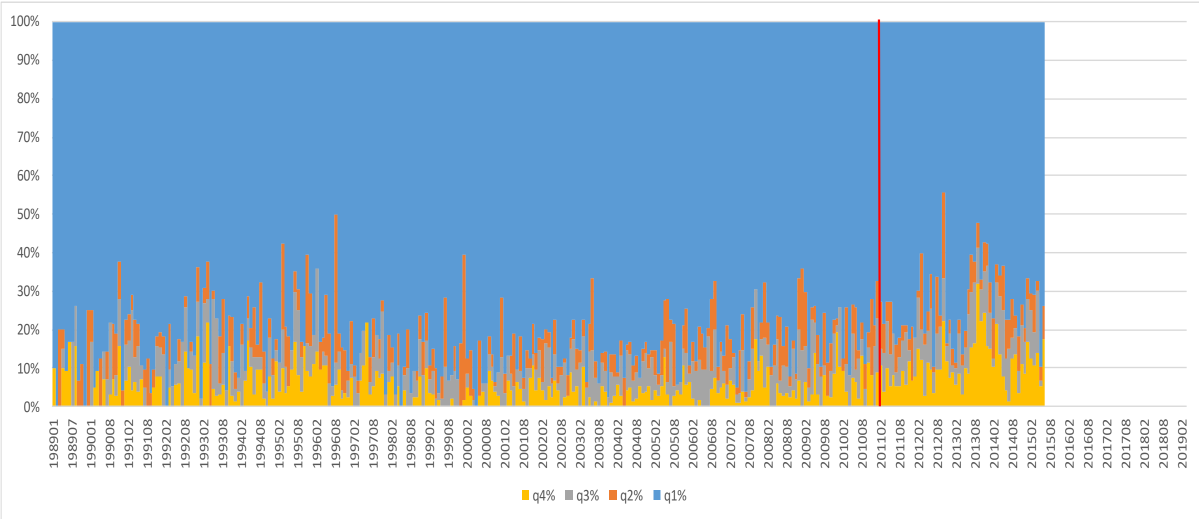
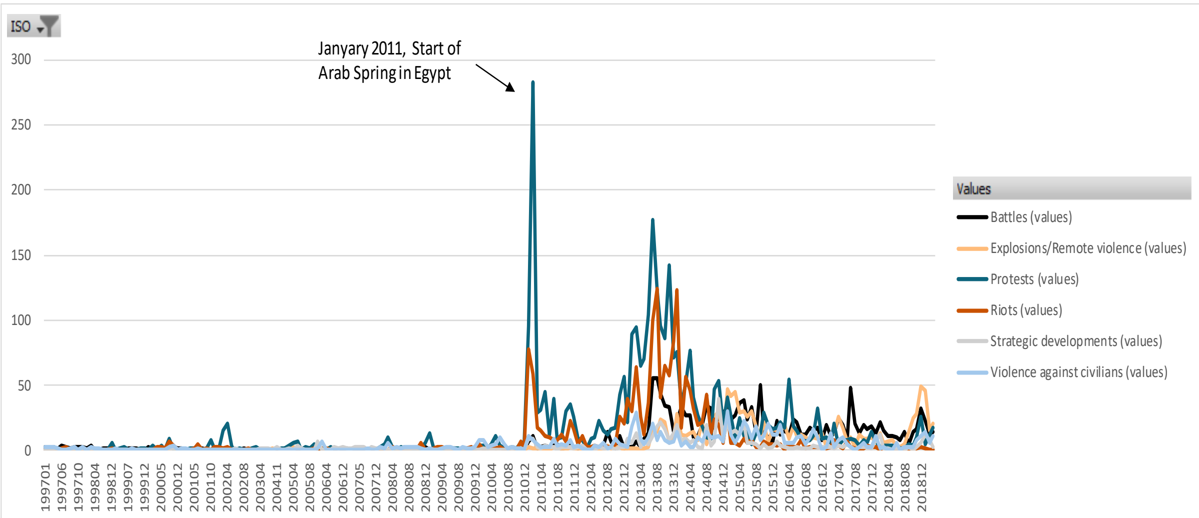


**Figure 19.** Events per QuadClass in Egypt, monthly aggregation (1989-2019 OEDA)



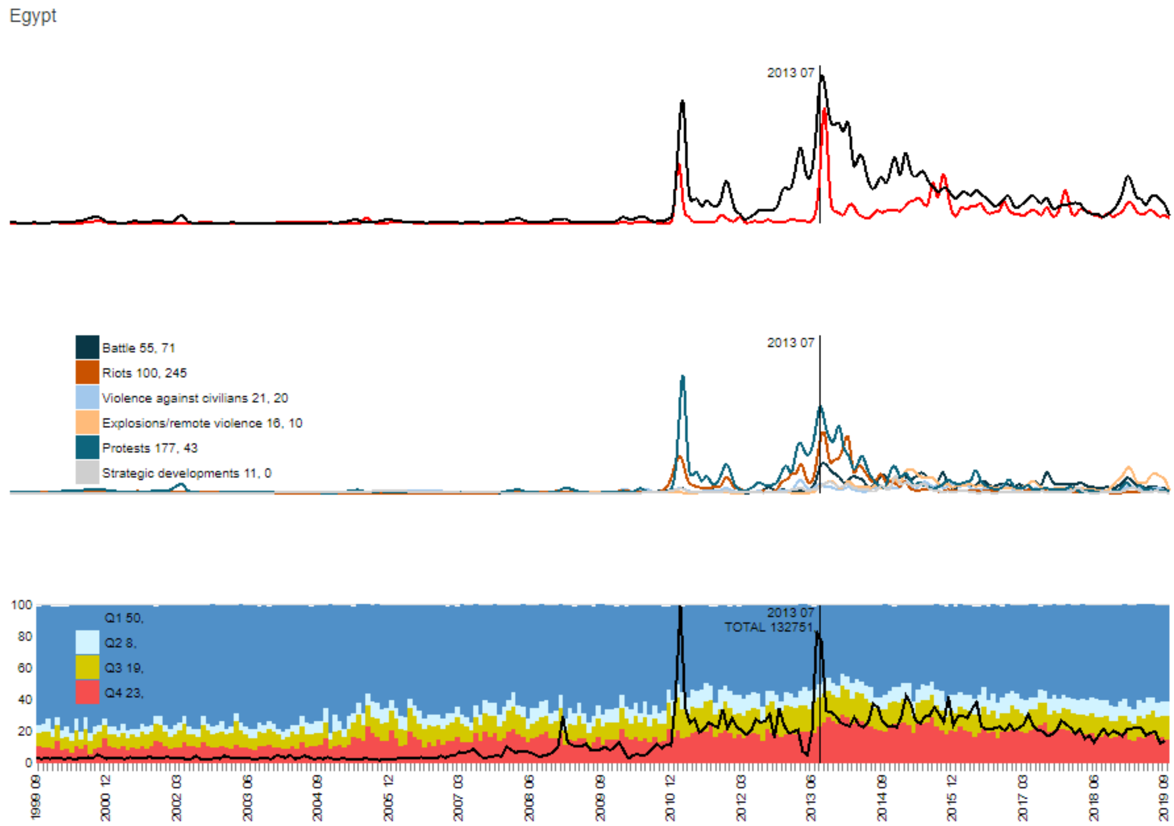

Eventually, the peak in ACLED's reported fatalities data confirms the start of the Arab Spring and the observed social tension soar depicted in the GDELT and ICEWS databases (see Figure 20).

**Figure 20.** Number of fatalities in Egypt, monthly aggregation (1997-2019 ACLED)




For Egypt, ACLED and GDELT time trends seem to be very similar.

**Figure 21.** ACLED and GDELT comparison for Egypt. The black trend in the top chart represents the total number of events mentioned in ACLED while the red one the total number of deaths.




An increase in the number of news events classified under Verbal Cooperation (Q1), Material Cooperation (Q2), Verbal Conflict (Q3), and Material Conflict (Q4) is reported in Figure 22. However, the biggest increase in news for January 2011 is found under Q1 and not under Q3 or Q4. From this information, we solely see that more cooperation than conflict events have taken place in Egypt in January 2011.

Normalizing our monthly time series as in Levin, Ali and Crandall (2018) between zero and one based on the minimum and maximum values, we can identify global and local maxima (see Figure 23). In Figure 23, one can see that all four QuadClasses reach a global maximum in January 2011 in Egypt. This helps us see whether the increase in the news events under each QuadClass is more or less important over time with respect to preceding events. However, it does not reflect the Arab Spring, which started in Egypt in January 2011, because the increase in the normalized cooperation events is similar to the increase in normalized conflict events.

If we additionally look at the increase or decrease per QuadClass in the proportion of events classified under Q1, Q2, Q3, and Q4 of the total reported news in Egypt before and after January 2011 (see

**Figure 24**), we can see Q3 and Q4 increased significantly in January 2011, whereas Q1 and Q2 decreased significantly. In this way, we can have a better picture of how a conflict is escalating, stagnating or de-escalating from month to month.

**Figure 22.** Absolute monthly news frequency per CAMEO QuadClass in Egypt - GDELT (Jan 1989 - March 2019)

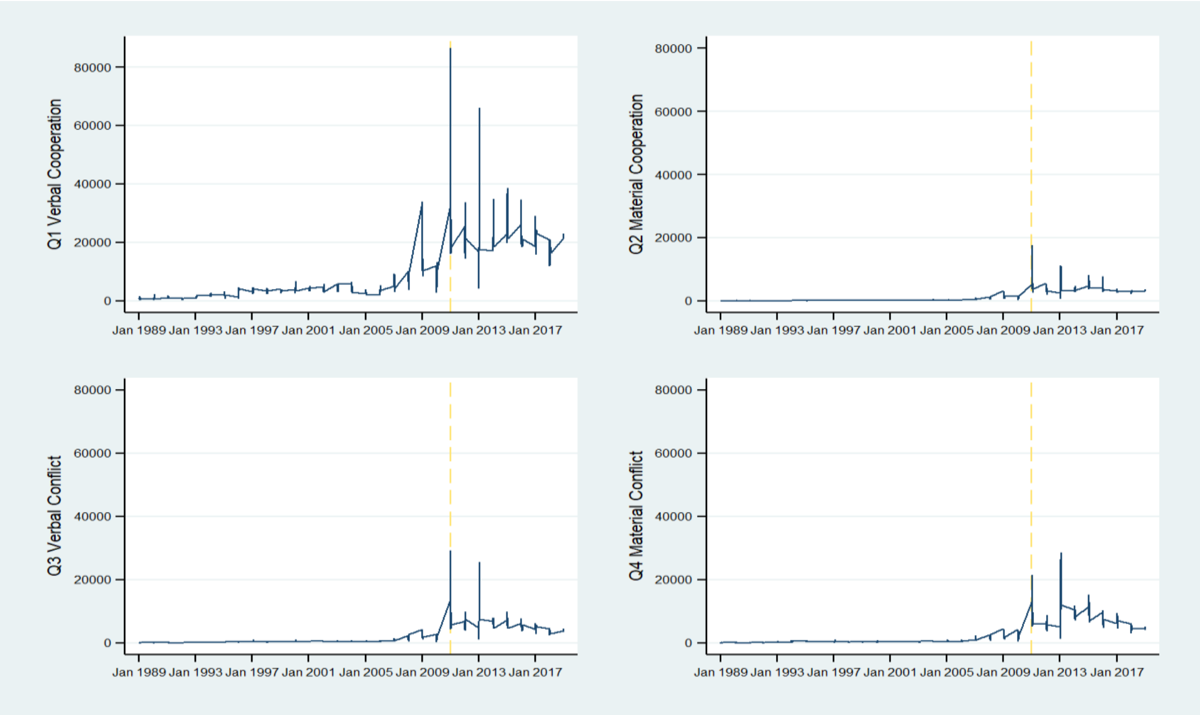



**Figure 23.** MaxMin normalized monthly news per CAMEO QuadClass in Egypt - GDELT (Jan 1989 - March 2019)

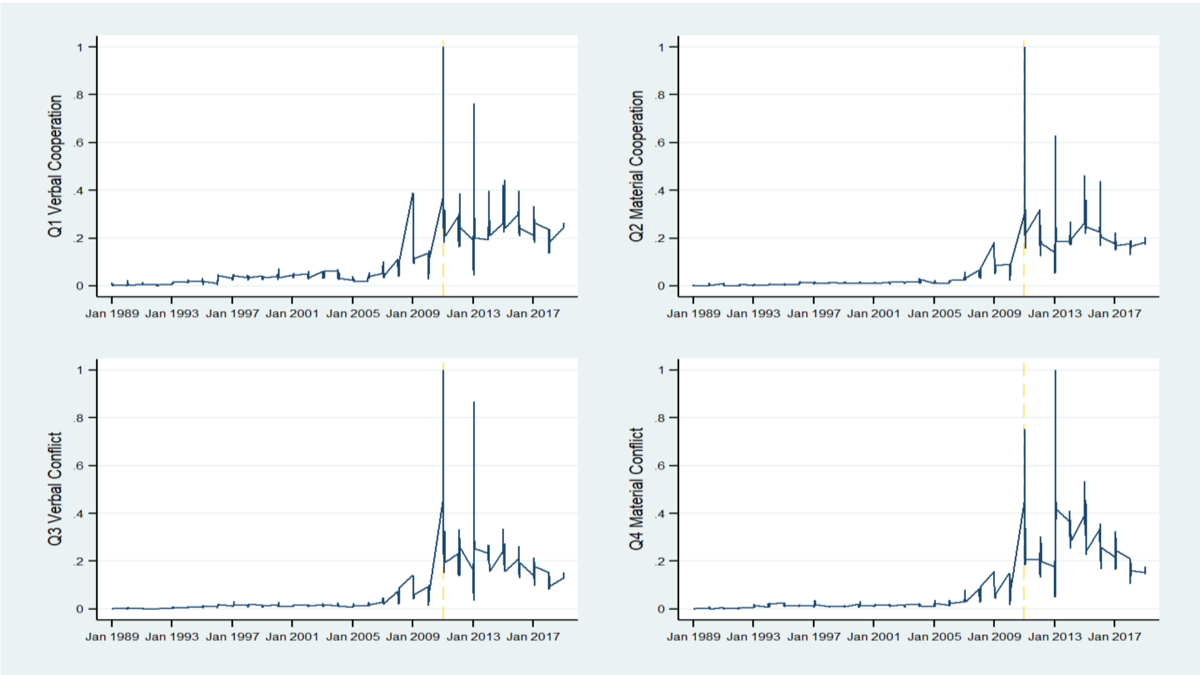
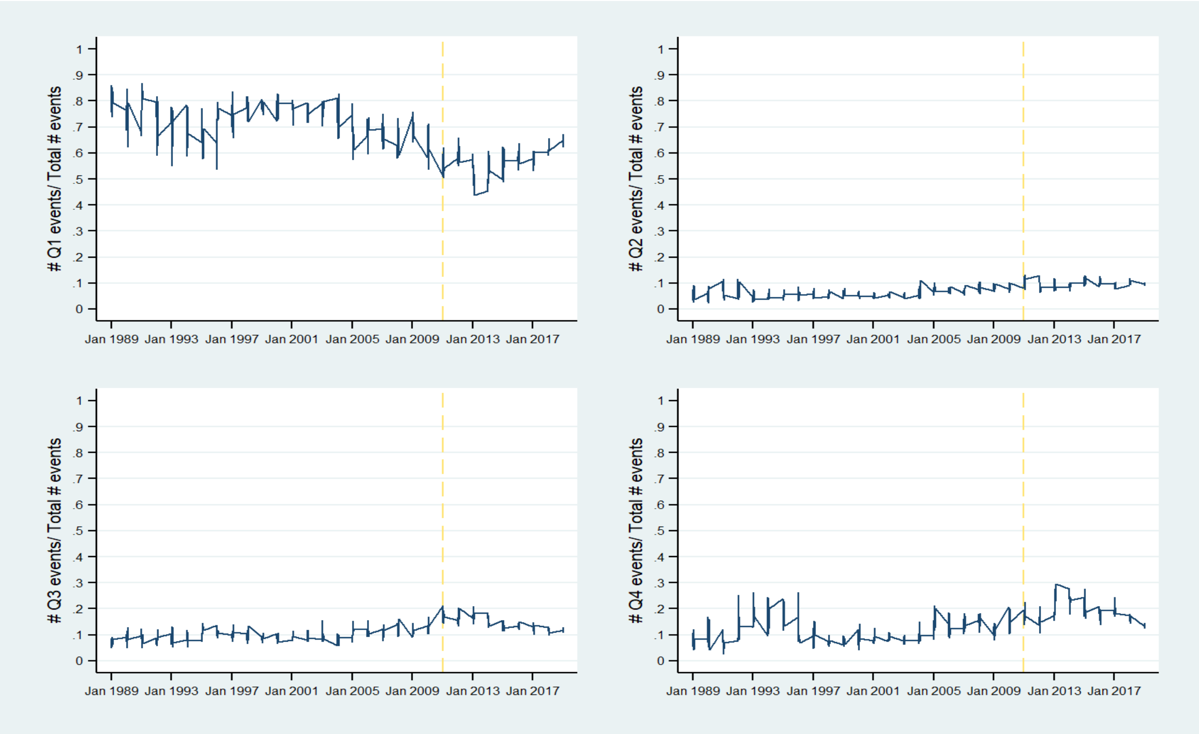


**Figure 24.** Proportion of reported events per QuadClass in Egypt before and after January 2011 - GDELT (Jan 1989 - March 2019)




### 2.4.4. Case Study 4: Maldives

In the case of the Maldives, a political crisis arose in February 2018 after President Abdulla Yameen decided to disobey the Supreme Court's order to release nine political prisoners and reinstating 12 parliament members. Additionally, the government declared a 15-day state of emergency, which suspended constitutional protections, banned public assemblies, and granted security forces powers to arrest and detain protestors, opposition figures, and activists.

As shown in Figure 25 and Figure 26, the increase in political tensions through an accession in the amount of events categorized as material conflict, Q4, is not visible in both the GDELT and the ICEWS databases; no data is available for that period in the OEDA-Phoenix dataset and the Maldives are not covered by the ACLED data. The red vertical line in the following plots refers to the start of the political crisis in the Maldives in February 2018.

**Figure 25.** Events per QuadClass in Maldives, monthly aggregation (1989-2019 GDELT)

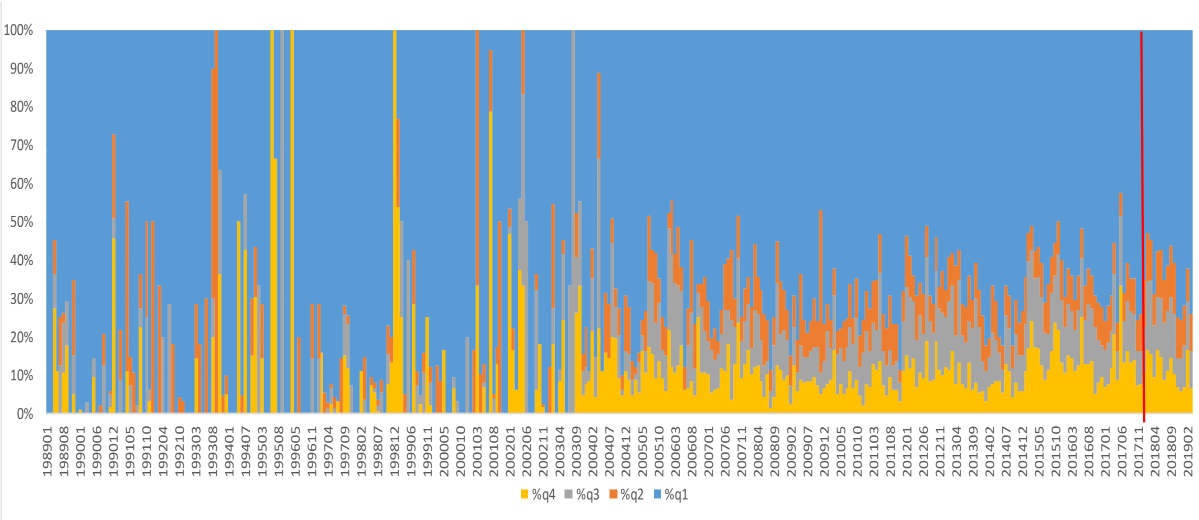
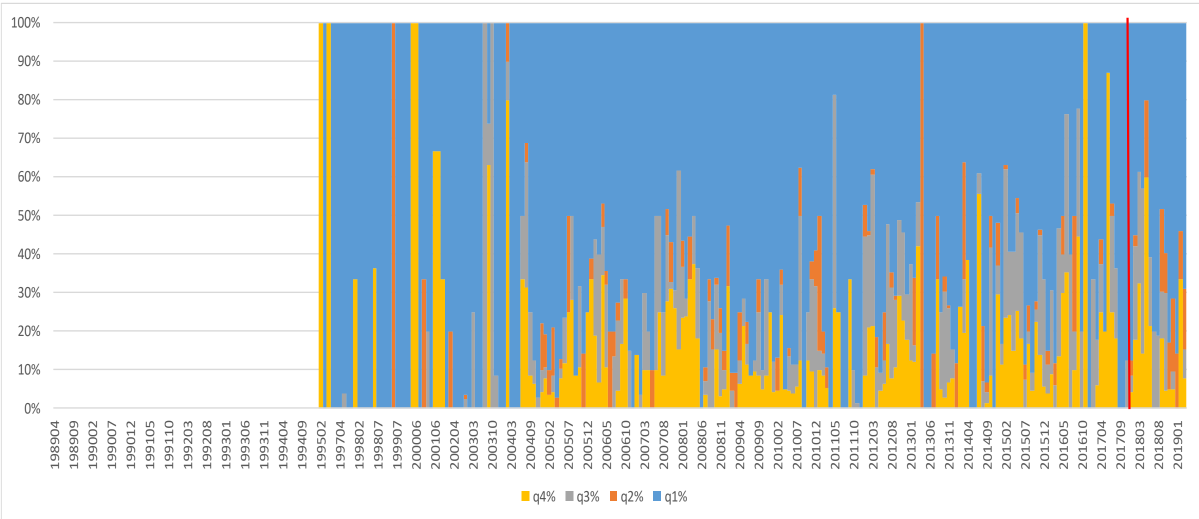



**Figure 26.** Events per QuadClass in Maldives, monthly aggregation (1995- 2019 ICEWS)




The Maldives political crisis is reported in Figure 27 by an increase in the number of news events classified under Verbal Cooperation (Q1), Material Cooperation (Q2), Verbal Conflict (Q3) and Material Conflict (Q4). Notice that the biggest increase in news for February 2018, is found under Q1 and not under Q3 or Q4. From this information, we solely see that in February 2018, more cooperation than conflict events have taken place in the Maldives.

Normalizing our monthly time series as in Levin, Ali and Crandall (2018) between zero and one based on the minimum and maximum values, we can identify global and local maxima (see Figure 28). In Figure 28, all four QuadClasses reach a global maximum in February 2018 in the Maldives. This helps us see whether the increase in the news events under each QuadClass is more or less important over time with respect to preceding events, but it does not reflect the political crisis that arose in February 2018; the increase in the normalized cooperation events is similar to the increase in normalized conflict events.

**Figure 27.** Absolute monthly news frequency per CAMEO QuadClass in the Maldives - GDELT (Jan 1989 - March 2019)

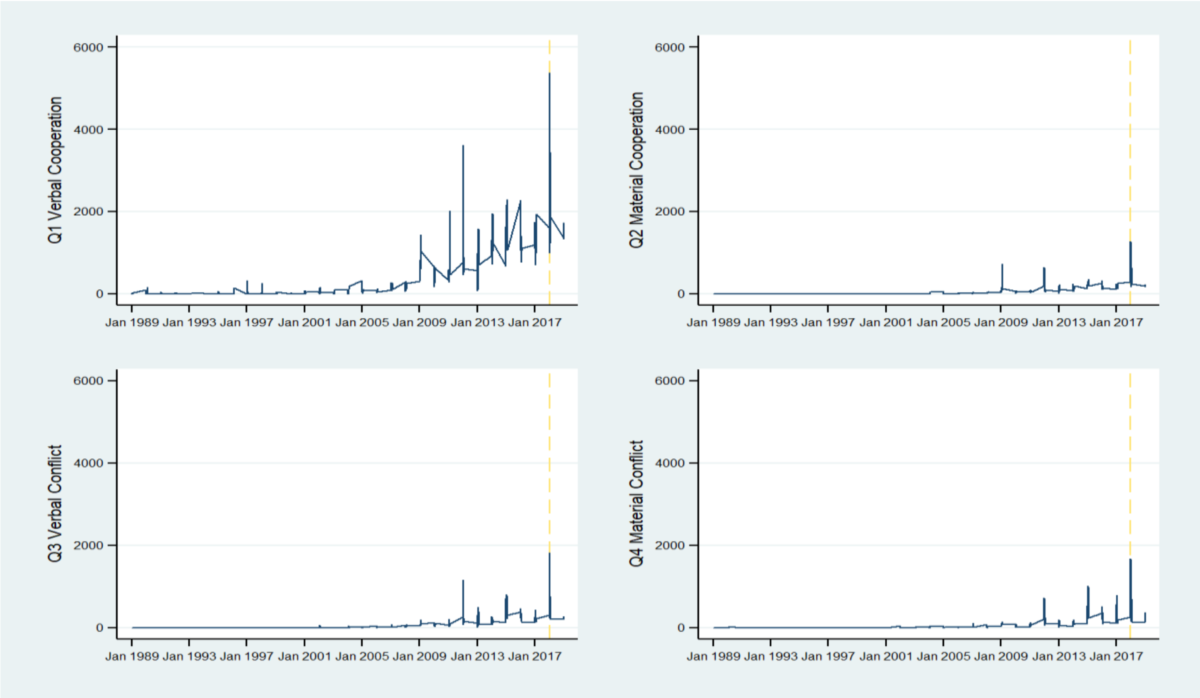
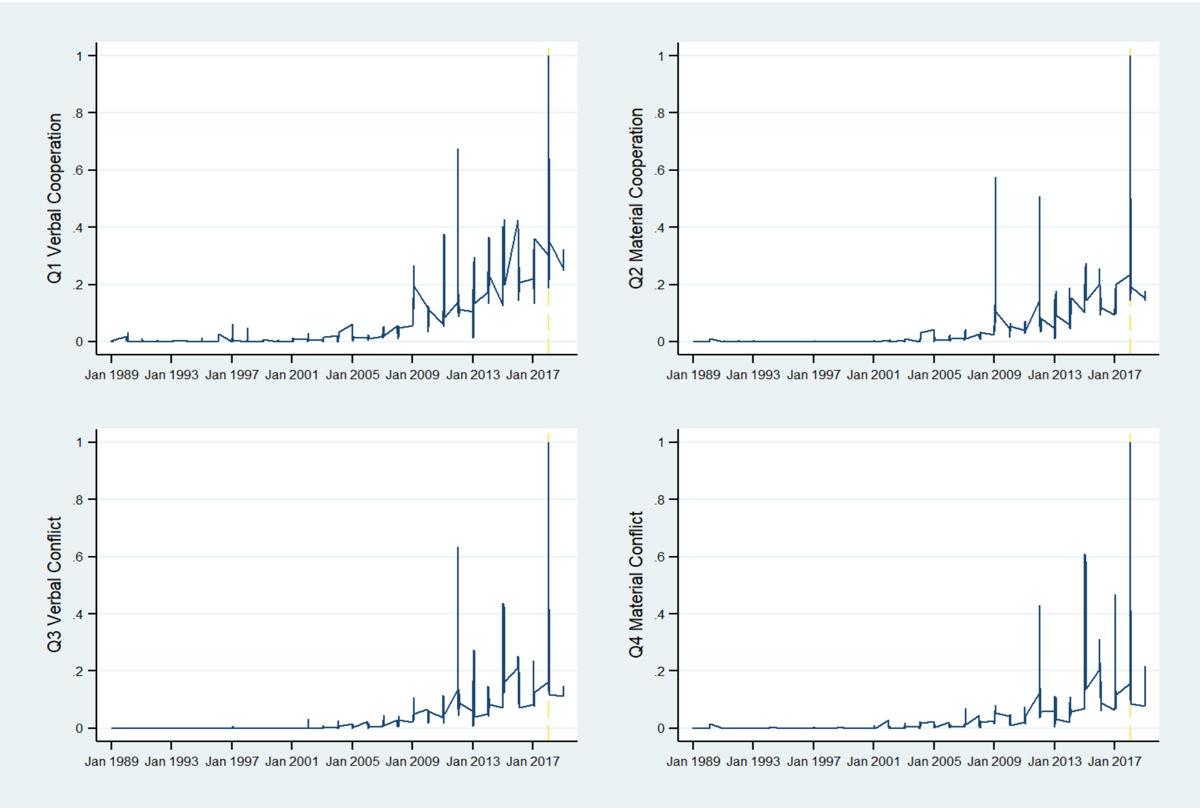
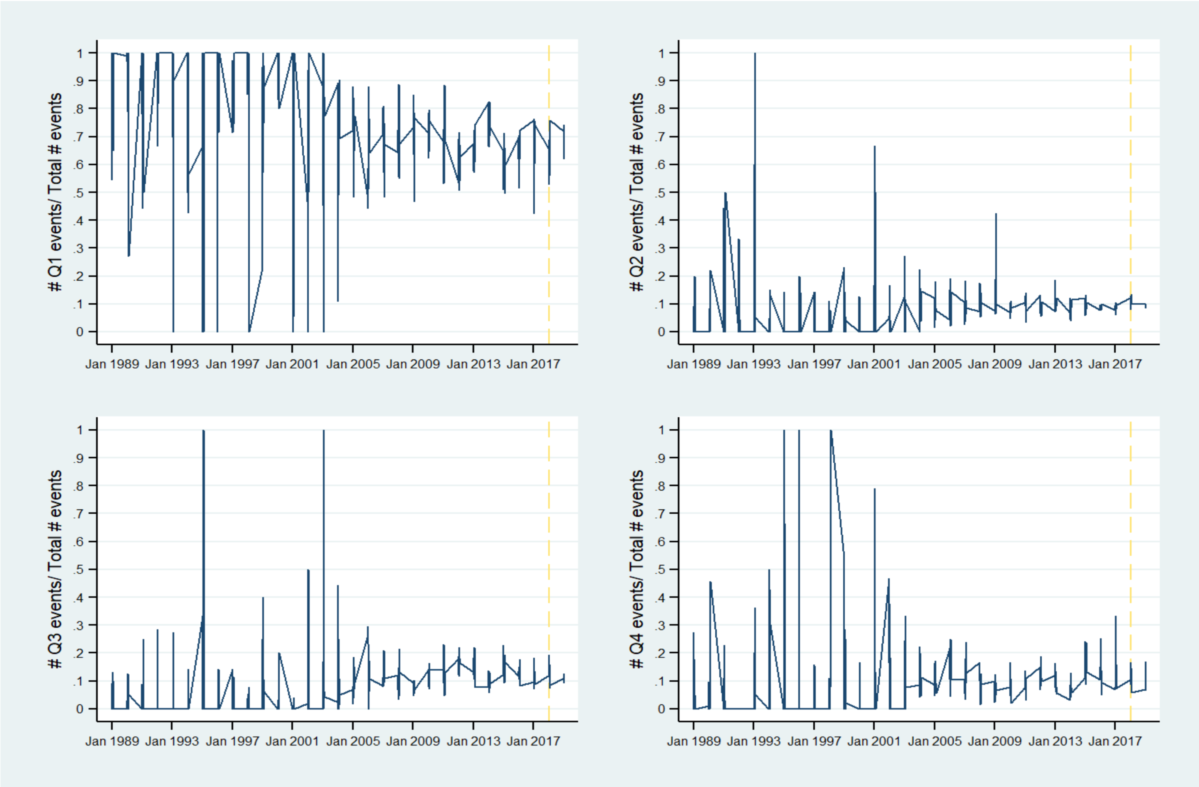



**Figure 28.** MaxMin normalized monthly news per CAMEO QuadClass in the Maldives - GDELT (Jan 1989 - March 2019)




If we additionally look at the increase or decrease per QuadClass in the proportion of events classified under Q1, Q2, Q3, and Q4 of the total reported news in the Maldives before and after February 2018 (see Figure 29), we can see that Q3 and Q4 increased significantly in February 2018, whereas Q1 and Q2 decreased significantly. In this way, we can have a better picture of how a conflict is escalating, stagnating or de-escalating from month to month.

**Figure 29.** Proportion of reported events per QuadClass in the Maldives - GDELT (Jan 1989 - March 2019)




### 2.4.5. Case Study 5: Nicaragua

The last case study deals with the Nicaraguan protests in April 2018. The 2018 - 2019 Nicaraguan protests began on April 18, 2018 when demonstrators in several cities of Nicaragua initiated protests against the social security reforms decreed by President Daniel Ortega that increased taxes and decreased benefits. After five days of unrest in which nearly thirty people were killed, Ortega announced the cancellation of the reforms. However, Nicaraguans continued to protest and demanded the resignation of Ortega. On September 29, 2018, political demonstrations were declared illegal by President Ortega. As a result of the unrest, the European Parliament called for an early election, despite the fact that Ortega was unconstitutionally re-elected for the third consecutive time in 2016 in an election without the presence of international election observers. As of October 2019, the protests are still on-going with more than 400 causalities.

Analyzing the GDELT and ICEWS dataset in Figure 30 and Figure 31, an increase in the amount of events categorized as material conflict, Q4, can be observed. Unfortunately, no data is available in the OEDA-Phoenix dataset for 2018 (see Figure 32) and Nicaragua is not covered by ACLED data. The red vertical line in the plots refers to the start of the Nicaraguan protests in April 2018.

In Figure 33, the protests are reported by an increase in the number of news events classified under Verbal Cooperation (Q1), Material Cooperation (Q2), Verbal Conflict (Q3), and Material Conflict (Q4). Yet, the biggest increase in news for April 2018 is found under Q1 and not under Q3 or Q4. From this information, we solely see that more cooperation than conflict events have taken place in Nicaragua in April 2018.

**Figure 30.** Events per QuadClass in Nicaragua, monthly aggregation (1989-2019 GDELT)

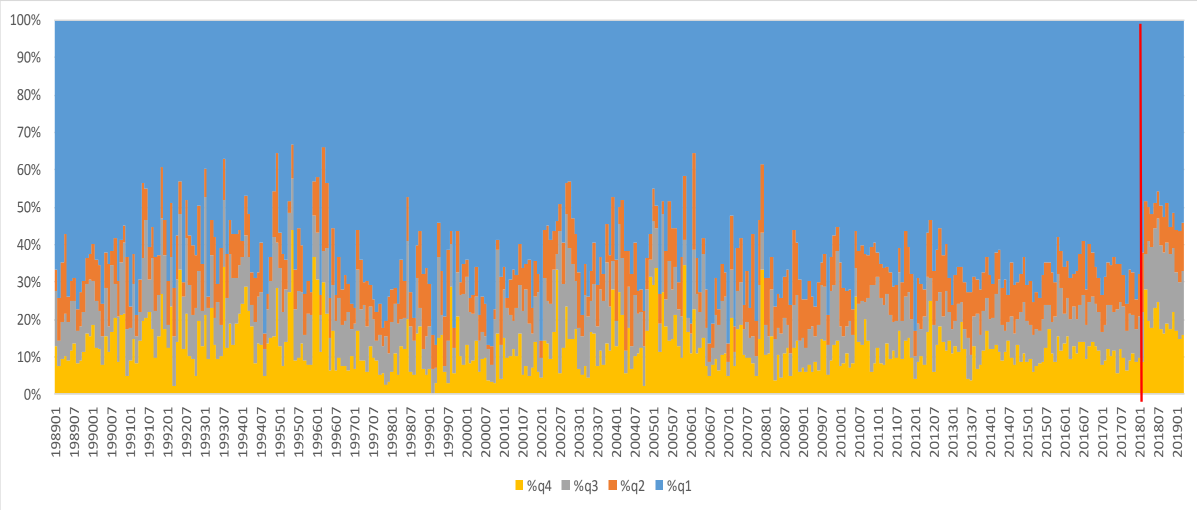



**Figure 31.** Events per QuadClass in Nicaragua, monthly aggregation (1995- 2019 ICEWS)

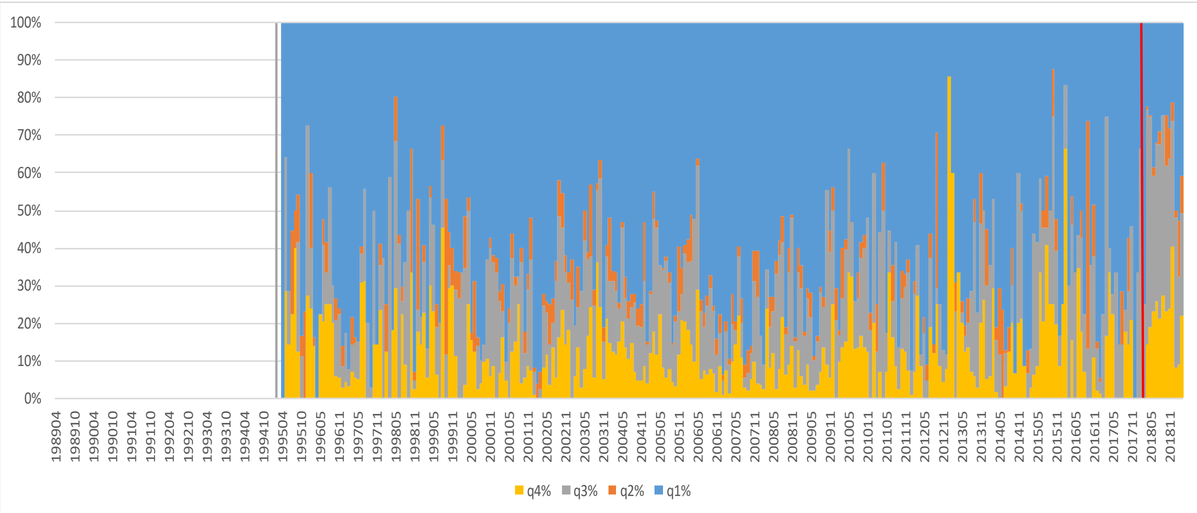


**Figure 32.** Events per QuadClass in Nicaragua, monthly aggregation (1989-2019 OEDA)

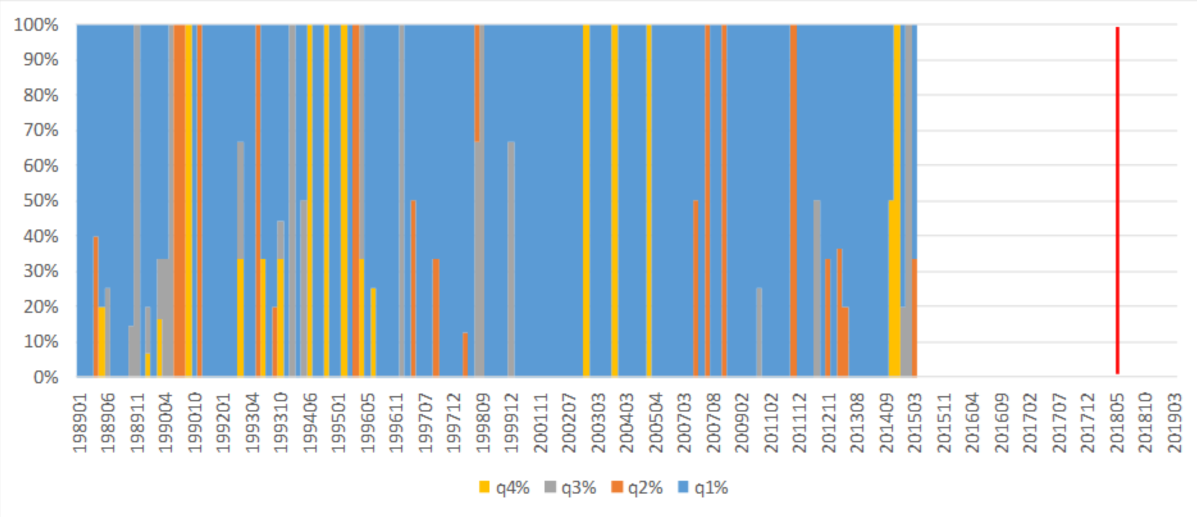
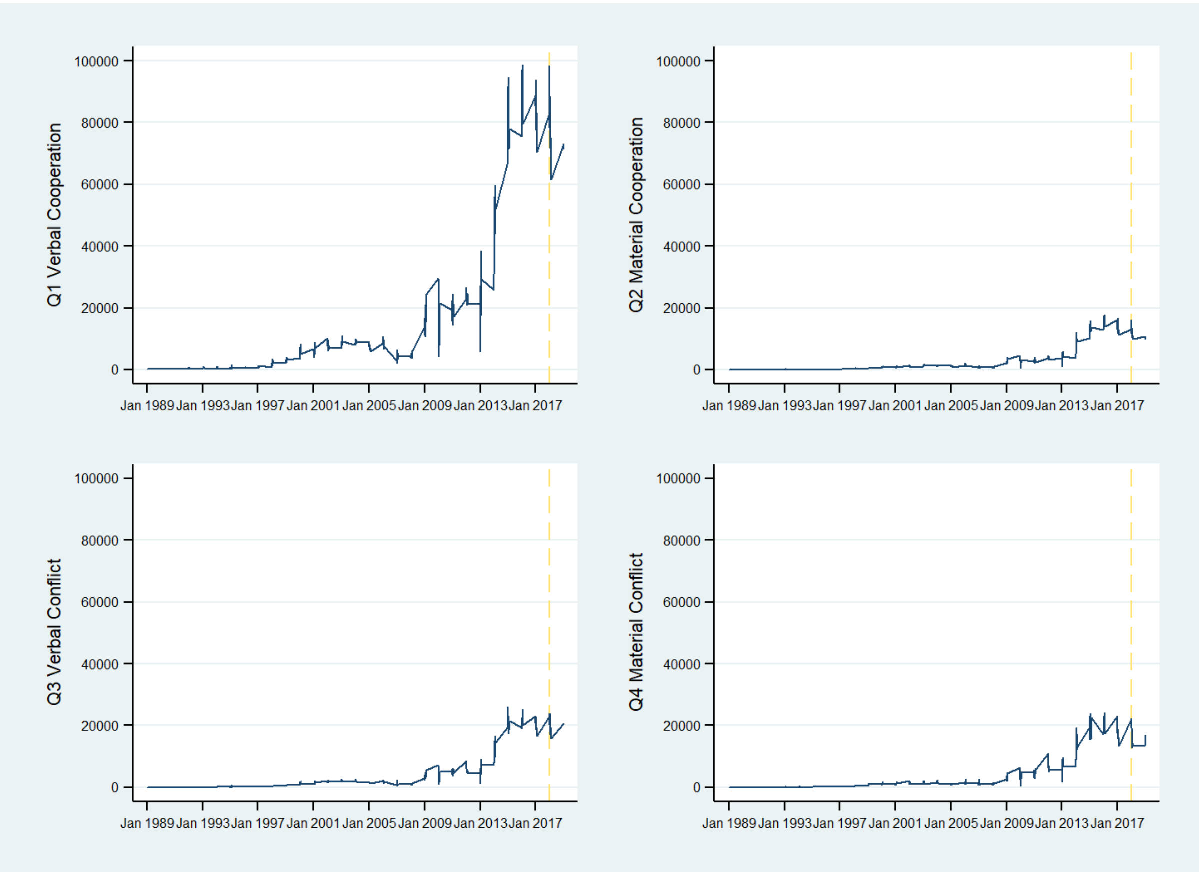



**Figure 33.** Absolute monthly news frequency per CAMEO QuadClass in Nicaragua - GDELT (Jan 1989 - March 2019)




Normalizing our monthly time series as in Levin, Ali and Crandall (2018) between zero and one based on the minimum and maximum values, we can identify global and local maxima (see Figure 34). In Figure 34, all four QuadClasses reach a local maximum in April 2018 in Nicaragua. This helps us see whether the increase in the news events under each QuadClass is more or less important over time with respect to preceding events, but it does not reflect the Nicaraguan protests, which started in April 2018. The increase in the normalized cooperation events is similar to the increase in normalized conflict events.

If we additionally look at the increase or decrease per QuadClass in the proportion of events classified under Q1, Q2, Q3, and Q4 of the total reported news in the Maldives before and after April 2018 (see Figure 35), we can see that Q3 and Q4 increased significantly in February 2018, whereas Q1 and Q2 decreased significantly. In this way, we can have a better picture of how a conflict is escalating, stagnating or de-escalating from month to month.

**Figure 34.** MaxMin normalized monthly news per CAMEO QuadClass in Nicaragua - GDELT (Jan 1989 - March 2019)

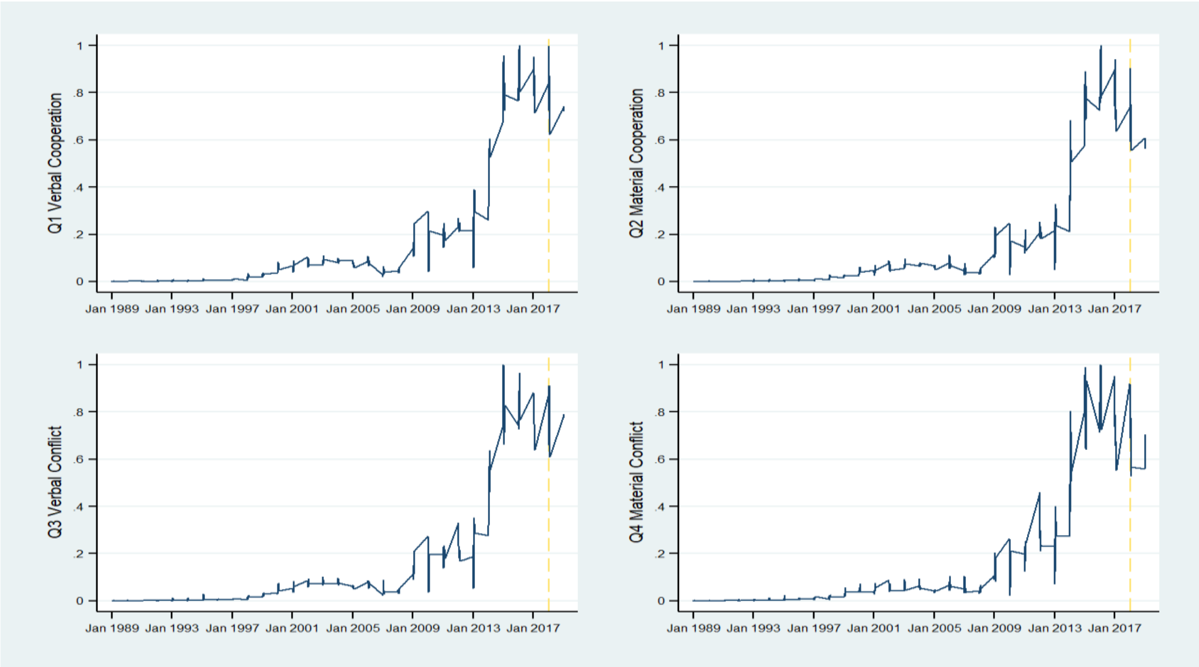
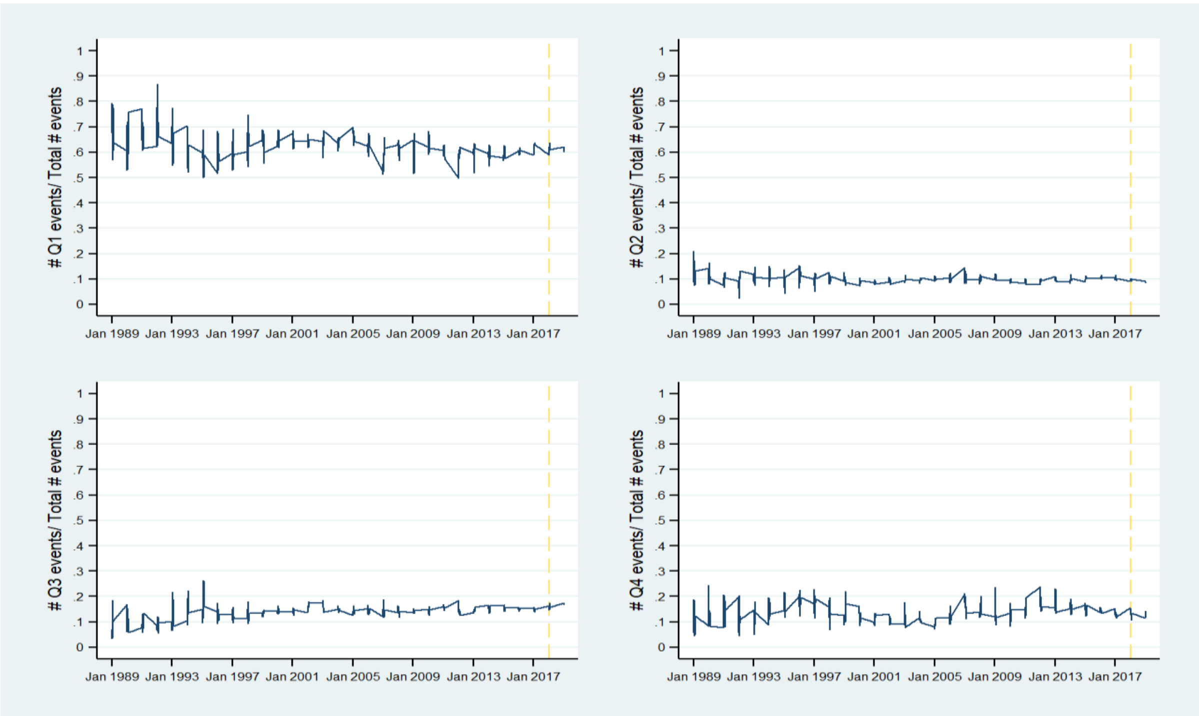



**Figure 35.** Proportion of reported events per QuadClass in Nicaragua - GDELT (Jan 1989 - March 2019)



# 3. Methodology

## 3.1. Artificial Intelligence Event based Modelling

The case studies reported in section 2 highlight the existence of social stress which can escalate to conflict. The historical time series of Verbal Cooperation (Q1), Material Cooperation (Q2), Verbal Conflict (Q3), and Material Conflict (Q4) quantify and allow us to grasp the direction in which the tensions evolve in order to predict future conflict events. The Artificial Intelligence (AI) methodology adopted to model the dynamic GCRI is built upon a Long-Short Term Memory (LSTM) Cell Recurrent Neural Network (RNN). These models are well-suited to classify, process, and make predictions based on time series data and forecast near future events.

Recurrent neural networks let us learn and extract patterns from sequential data (time series, music, audio, video frames, etc.) in which the current hidden state is a function of the previous hidden state and the current input. In the example of stock prices, the current stock price depends on past stock prices and today's market situation.

Applying this model to conflict prediction, our implicit assumption is that the conflict state today depends on a history of conflict states. We could think of it as a composite function in which the eldest events are nested within the more recent events.

Mathematically speaking, past events receive in this way a smaller weight than the more recent events. To avoid this, and to equally 'reweight' all events in the model's memory, we apply the Long Short-Term Memory (LSTM) to our RNN.

In the same way a linear regression model is solved through an optimization problem of minimizing the squared errors between the prediction and the actual value of a dependent variable, the LSTM RNN is an optimization of a gradient while minimizing the model's errors.

Neural networks like LSTM RNNs are able to almost seamlessly model problems with multiple input variables and as mentioned by Sak et al. (2014), 'LSTM is a specific recurrent neural network (RNN) architecture that was designed to model temporal sequences and their long-range dependencies more accurately than conventional RNNs' (Sak, Senior & Beaufays, 2014).

As previously said, LSTM RNNs are well-suited to classify, process, and make predictions based on time series data, since there can be lags of unknown duration between important events in a time series. This is a great benefit in time series forecasting and supervised time series learning (Bakker, 2002), areas in which classical linear methods fail to adapt to multivariate or multiple input forecasting problems.

The main difference between the original GCRI (Halkia et al., 2017a, b, 2019) and the dynamic GCRI proposed in this report is the nature of the input data. While the original GCRI uses yearly data time series, which allow us to estimate the conflict risk of a country in the next one to four years, the use of a 15-minute or a daily updated dataset based on event data gives us the opportunity to have a(n) (almost) real time conflict risk estimation.

Taking into account the update frequency of both the GDELT and the ICEWS datasets, we have aggregated the data by month for both datasets to be able to compare the results of the probability estimates of a conflict event for the same period[14] (see 4.1).

Using the CAMEO classification, we are able to observe changes in the trends which are crucial for the training of the model. We have filtered the GDELT dataset to only have the available information after 1989, which is the starting year of the original GCRI input values, enabling juxtaposition of the results obtained in the respective models.

[14] Considering the lack of available time series for the Phoenix dataset, we are not able to have an estimation of the probability of a conflict event based on it. For the ICEWS, the data captures events on a monthly basis until October 2018, and becomes daily as of October 2018.

In a second step, we created our model by using a random sample consisting out of 50% of the available dataset as a training set and the remaining 50% as the testing/controlling one. Having a testing and a training sample makes it possible to control and validate the accuracy of the model[15].

## 3.2. Conflict Risk Alarm System (CRA-S) Configuration

Through the AI model's capacity to predict the future proportion of conflict or cooperation related events in a country, we have set up a Conflict Risk Alarm System (CRA-S), which signal social unrest upheavals (an abnormal increase in the proportion of the Q4x, Q4x+1 or in the total number of events mentioned as Q3x or Q4x, where x stands for point in time). This allows policy makers to implement preventive actions more rapidly to mitigate conflict exacerbations at an earlier stage of the conflict development cycle. First, we aggregate the events recorded in each of the three databases per month.

Next, we compute the monthly amount of news articles reported in each QuadClass (Q1, Q2, Q3, and Q4) with respect to the total number of articles per month.

Finally, we compute a 95% Confidence Interval (CI) to estimate the significance of the local maxima in the increase of the number of events in Q3 and Q4 with respect to the other events. The CI was computed setting a 3 and 6 month moving window. In the case of a 3-month moving window, we only take the events of the past 3 months into consideration to calculate the local maxima in conflict events and the CI.

Doing so, we can have an alarm for the cases in which the prediction of our model is out of the bound of the 95% CI, which means that the prediction is a real local max and the increase in the tension in a given country is significant.

## 3.3. Ranking of Countries based on CRA-S

In order to rank the countries in the most appropriate way, we compute the rate of change between the Q4 of the current month and the Q4 of the previous one (here called delta classification). Hence, the rate of change is

$$\Delta Q_4 = \frac{Q_{4x} - Q_{4x-1}}{Q_{4x-1}}$$

where $Q_{4x}$ is the proportion of the Q4 for the current month and $Q_{4x-1}$ is the proportion of the Q4 for the previous month. Based on this rate, we rank the countries so as the country with the highest increase in the Q4 ($\Delta Q_4$) will be first and the country with the highest decrease will be last. In case the Q4x is a local max, meaning that the increase in the current value of the Q4 is significant, we have set an alarm following the same methodology as described in the section 3.2.

Beside this, we set another alarm when the total absolute number of the events mentioned in the Q3x and Q4x (current values) QuadClasses is a local max[16] (significant increase in the absolute number of the events).

---

[15] We define the LSTM model with 50 neurons in the first hidden layer and with 1 neuron in the output layer for predicting the risk. We then use the Root Mean Square Error (RMSE) to validate the accuracy of the model. The model will be fit for 500 training epochs with a batch size of 72 (in the neural network terminology: one epoch = one forward pass and one backward pass of all the training examples; batch size = the number of training examples in one forward/backward pass).

[16] In mathematics, the maxima and minima (the respective plurals of maximum and minimum) of a function, known collectively as extrema (the plural of extremum), are the largest and smallest value of the function, either within a given range (the local or relative extrema) or on the entire domain of a function (the global or absolute extrema).

Cumulatively, the countries' ranking and the CRA-S consist of the following parameters:

- **Initial ranking:** The initial ranking is based on the $\Delta Q_4$.
- **Alarm 1:** The proportion of the $Q_4$ ($Q_{4x}$) for the current month is a local max, meaning that the increase is significant and out of the 95% Confidence Interval that we have calculated for the x-month moving window.
- **Alarm 2:** The total absolute number of the events mentioned in the $Q_{3x}$ and $Q_{4x}$ (current values) is a local max.
- **Alarm 3:** The proportion of the predicted values of the $Q_4$ ($Q_{4x+1}$) for the next month is a local max.

Using these parameters, we re-rank the countries according to the following rules:

- **Rule 1:** If all the alarms in a country signal at the same time, this country will be re-ranked as first. In case there is more than one country, we keep the delta classification as described above.
- **Rule 2**: If two of the alarms in a country signal at the same time, this country will be re-ranked just below. In case there is more than one country, we keep the delta classification as described above.
- **Rule 3**: If one of the alarms in a country signals, that country will be re-ranked just after the countries of the previous case.
- **Rule 4**: The remaining countries with no alarm signals are ranked thereafter by keeping their initial ranking ($\Delta Q_4$).

Doing so, we create a classification method based on a system of three different alarms, taking into account both the absolute and the relative number of events per country. In addition, we take not only the material conflict events but also the verbal conflict mentions into consideration since the escalation of a verbal conflict to a material conflict is always possible and this information is definitely important for every policy maker. Furthermore, we stress the importance of both the alarms and the classification method, taking into consideration their internal relation, by re-ranking the countries in the aforementioned way.

## 4. Results

### 4.1. Artificial intelligence Root Mean Square Error (RMSE) and Model Predictions

The results of running the LSTM RNN model for the five case study countries and three databases are depicted in the last column of Table 2 for the month of March 2019. To see how accurate our model predictions are in estimating the percentage of material conflict events (Q4 of the CAMEO classification) in percentage of the overall events, we measure the Root Mean Square Error (RMSE) of the model, which is the standard deviation of the residuals (prediction errors). The smaller the RMSE, the more precise and accurate the model is.

**Table 2.** RMSE for March 2019 per dataset[17]

| **Country** | **Dataset** | **RMSE** |
|---|---|---|
| **Libya** | ICEWS | 0.215 |
| | GDELT | 0.089 |
| | OEDA | **No data** |
| **Sudan** | ICEWS | 0.097 |
| | GDELT | 0.041 |
| | OEDA | **No data** |
| **Egypt** | ICEWS | 0.119 |
| | GDELT | 0.059 |
| | OEDA | **No data** |
| **Maldives** | ICEWS | 0.210 |
| | GDELT | 0.071 |
| | OEDA | **No data** |
| **Nicaragua** | ICEWS | 0.147 |
| | GDELT | 0.063 |
| | OEDA | **No data** |

As reported in Table 2, the RMSE using the GDELT data is the lowest in all the case studies. This means that the predictions based on this dataset are closer to the observed values. This is most probably due to the limited data availability (1995-2019) in the ICEWS database. Overall, this implies that the model using the GDELT data is most precise and accurate.

[17] The Phoenix-OEDA dataset does not provide time series after 2015.

## 4.2. Early Warning Alarm System Predictions and Accuracy

In this part of the report, the results of the early warning alarm system (an abnormal increase in the proportion of Q4 – material conflict events) are presented in Table 3 for the five case study scenarios and three databases. To validate the results, we compare them with ACLED's total number of fatalities in the last column of Table 3.

The ALARM is given in case there is an extraordinary increase in the predictions out of the 95% confidence interval (CI) we have set for two local maxima in respectively a 3-month and a 6-month window from the case study event.

In Table 3, we report whether or not the model gives us an alarm for the Arab spring in Libya and Egypt, the Sudanese protests, the political crisis in the Maldives, and the Nicaraguan protests. It is worth mentioning that we have also filtered the GDELT dataset to have more reliable input data. Based on our GDELT validation (see Annexes

Annex 1. The GDELT Dataset Qualitative Validation. ), we have set a filter on 100 mentions per article. In other words, when the filter is applied we include only the articles that have been mentioned more than 100 times in that part of the analyses (GDELT_100 in Table 3). This has been done to remove possible noisy information in the GDELT database, taking into consideration the fact that if an event really happens, it should be reproduced by more than one media source and in more than one article. In addition, we are aware that this may lead to the exclusion of important information in countries where local press is being repressed and international media has only a limited interest. However, the number of such countries is limited and the inclusion of all available information within the GDELT database would lead to greater bias than the exclusion of some information.

**Table 3.** Sample validation of the predictions based on past events.

| **Country** | **Date** | **Dataset** | **3-month local max** | **6-month local max** | **ACLED number of fatalities (Total)**[18] |
|---|---|---|---|---|---|
| **Libya (LY-LBY)** | February 2011 (Start of Arab spring) | GDELT | ALARM[19] | ALARM | 99 (+98) |
| | | GDELT_100 | ALARM[20] | ALARM | |
| | | ICEWS | NO ALARM | NO ALARM | |
| | | OEDA-Phoenix | NO ALARM | NO ALARM | |
| **Sudan (SU - SDN)** | December 2018 (2018–19 Sudanese protests) | GDELT | NO ALARM | NO ALARM | 130 (+82) |
| | | GDELT_100 | ALARM[21] | ALARM | |
| | | ICEWS | NO ALARM | NO ALARM | |
| | | OEDA-Phoenix | NA | NA | |
| **Egypt (EGY-EG)** | | GDELT | NO ALARM | NO ALARM | 203 (+197) |

[18] The increase compared with the previous month number of fatalities.

[19] The percentage of predicted material conflict for LY is 14.59 %, out of the 95% CI.

[20] The percentage of predicted material conflict for LY is 14.59 %, out of the 95% CI.

[21] The percentage of predicted material conflict for SU is 41.81 %, out of the 95% CI.

| | | | | | |
|---|---|---|---|---|---|
| | January 2011 (Start of Arab spring) | GDELT_100 | ALARM[22] | ALARM | |
| | | ICEWS | NO ALARM | NO ALARM | |
| | | OEDA-Phoenix | NO ALARM | NO ALARM | |
| **Maldives (MDV-MV)** | February 2018 (2018 Maldives political crisis) | GDELT | NO ALARM | NO ALARM | **NA** |
| | | GDELT_100 | NO ALARM | NO ALARM | |
| | | ICEWS | NO ALARM | NO ALARM | |
| | | OEDA-Phoenix | **NA** | **NA** | |
| **Nicaragua (NIC-NU)** | April 2018 (2018–2019 Nicaraguan protests) | GDELT* | NO ALARM | NO ALARM | **NA** |
| | | GDELT_100 | NO ALARM | NO ALARM | |
| | | **ICEWS** | **ALARM[23]** | **ALARM** | |
| | | OEDA-Phoenix | **NA** | **NA** | |

In Table 3, we can observe that the ALARM rang for Libya's Arab Spring using the GDELT dataset (either filtered or unfiltered), while the model predicted the 2018-19 Sudanese protests and the Start of Arab spring in Egypt using the filtered GDELT data. Using the ICEWS data, we could solely predict the 2018–2019 Nicaraguan protests. As one can observe, using the filter improves the reliability of the GDELT database and forces the model make more accurate predictions.

[22] The percentage of predicted material conflict for EG is 42.23 %, out of the 95% CI.

[23] The percentage of predicted material conflict for NIC is 17.66 %, out of the 95% CI.

## 5. Discussion

In this report, we presented our proposal for a dynamic GGCRI based on actor-based event data at a country level, signalling a potential trigger to violent conflict, including demonstrations, strikes, election violence, etc. We also presented potential datasets based on the CAMEO classification system, i.e. the GDELT, ICEWS, and the OEDA-Phoenix datasets.

The dynamic GCRI, using an LSTM RNN model to predict the materialization of a conflict, demonstrates that the GDELT is potentially the most comprehensive database. Nevertheless, many provisions must be added to the fundaments of any program using the GDELT database in order to render it accurate and effective. The same applies for the other two datasets, as they are similar to the GDELT with regard to their general performance. However, we were only able to validate the GDELT dataset as there is not available information for the news source for the ICEWS and OEDA-Phoenix.

Moreover, the LSTM RNN model we propose, which is one of the most advanced models nowadays, performs well and it is able to handle historical time series data and "understand" each event based on its previous knowledge. In addition, while the absolute number of events informs us whether something is happening in a given country and the normalized number of events gives us the importance of an event with respect to the preceding ones, the additional analysis in proportions completes the picture on how a conflict is escalating, stagnating or de-escalating from month to month. Finally, the local maxima modelling gives us the possibility to have an early warning system, which informs the policy makers in case of an abnormal increase in the tensions in a given country.

# 6. Conclusions

This report presented a dynamic model of the GCRI, a conflict risk model supporting the design of EU's conflict prevention strategies developed by the JRC in collaboration with an expert panel of researchers and policy-makers.

The proposed dynamic GCRI integrates and identifies every stage of the conflict development or de-escalation in its entire complexity, including internationalized contentious action. Using country-level actor-based event data sets that signal potential triggers to violent conflict such as demonstrations, strikes, or elections-related violence, the model aims at estimating the occurrence of material conflict events, under the assumption that an increase in material conflict events goes along with a decrease in material and verbal cooperation.

The AI methodology adopted to model the dynamic GCRI is built upon a Long-Short Term Memory (LSTM) Cell Recurrent Neural Network (RNN). These models are well-suited to classify, process and make predictions based on time series data and forecast near future events. Besides this AI model, we have set up an early warning alarm system to signal abnormal social unrest upheavals.

Three potential datasets were tested in this report following the CAMEO political event coding classification: (i) the Global Data on Events Location and Tone (GDELT) project, (ii) the Integrated Crisis Early Warning System (ICEWS) Dataverse dataset and (iii) the Phoenix - Open Event Data Alliance (OEDA)-Phoenix Dataset.

Even though the AI and early warning alarm seem to be able to predict the materialization of a conflict in the near future, the analysis of the results conveys that implementing the GDELT, ICEWS or OEDA-Phoenix as an input to the dynamic GCRI requires overcoming certain obstacles. First, the automated codebook algorithm is not publically available for GDELT, which does not allow investigation on the source of potential errors in the news classification. Second, the ICEWS data sources are not publically available so validation is not readily possible. Third, the Phoenix dataset has been discontinued after 2015, so it cannot be used currently for forecasting. Moreover, common issues need to be resolved in all three datasets: false positive rates, duplication rates, urban and other geographical/socioeconomic biases, “media fatigue”, particularly in conflict zones as well as biases in existing event data sets primarily focusing on violent conflict.

The Europe Media Monitor (EMM) event data set could be a promising alternative in the near future, but it could not be tested at this stage because it is not based on the CAMEO classification methods.

To conclude, the dynamic GCRI methodology applied on the GDELT dataset already gives policy makers the possibility to observe the situation in a country on a monthly basis, but event – based systems will require supplementary research to offset the databases' shortcomings, such as automated data validation and custom result parameters.

## References


Althaus, S, J Bajjalieh, JF Carter, B Peyton & DA Shalmon (2017) Cline Center Historical Phoenix Event Data v1.0.0 (http://www.clinecenter.illinois.edu/data/event/phoenix/).

Bakker, Bram (2002) Reinforcement Learning Memory. International Conference on Neural Information Processing Systems 1475–1482 (https://papers.nips.cc/paper/1953-reinforcement-learning-with-long-short-term-memory.pdf).

Fillieule, Olivier & Manuel Jiménez (n.d.) Appendix A The Methodology of Protest Event Analysis and the Media Politics of Reporting Environmental Protest Events (https://serval.unil.ch/resource/serval:BIB_22172.P001/REF).

GDELT (2019) The GDELT Project (https://www.gdeltproject.org/).

Halkia, Matina, Stefano Ferri, Inès Joubert-Boitat & Francesca Saporiti (2017a) Conflict Risk Indicators: Significance and Data Management in the GCRI. Luxembourg: Publications Office of the European Union.

Halkia, Matina, Stefano Ferri, Inès Joubert-Boitat, Francesca Saporiti & Kauffmann Mayeul (2017b) The Global Conflict Risk Index ( GCRI ) Regression Model : Data Ingestion , Processing , and Output Methods. Luxembourg: Publications Office of the European Union.

Halkia, Matina, Stefano Ferri, Marie Schellens, Michail Papazoglou, Dimitrios Thomakos & Francesca Saporiti (2019) The Global Conflict Risk Index: A quantitative tool for policy support on conflict prevention. Progress in Disaster Science (Under review).

Halterman, Andrew & John Beieler (n.d.) A New, Near-Real-Time Event Dataset and the Role of Versioning [Draft] (http://openeventdata.org/).

Hegre, Håvard, Joakim Karlsen, Håvard Mokleiv Nygård, Håvard Strand & Henrik Urdal (2013) Predicting Armed Conflict , 2010 – 2050. International Studies Quarterly 57(2): 250–270.

ICEWS | Lockheed Martin (2019) (https://www.lockheedmartin.com/en-us/capabilities/research-labs/advanced-technology-labs/icews.html#MoreInfo).

ICEWS Automated Daily Event Data - Integrated Crisis Early Warning System (ICEWS) Dataverse (2019) (https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/QI2T9A).

Laessing, Ulf, Mustafa Hashem Omar Fahmy, Ahmed Elumami & Ayman Al-Warfalli (2017) Libya faces chaos as top court rejects elected assembly - Reuters (https://www.reuters.com/article/us-libya-security-parliament/libya-faces-chaos-as-top-court-rejects-elected-assembly-idUSKBN0IQ0YF20141106).

OEDA Datasets (2019) (http://openeventdata.org/datasets.html).

Qiao, Fengcai, Pei Li, Xin Zhang, Zhaoyun Ding, Jiajun Cheng & Hui Wang (2017) Predicting Social Unrest Events with Hidden Markov Models Using GDELT. Discrete Dynamics in Nature and Society (https://doi.org/10.1155/2017/8180272).

Sak, Hasim, Andrew Senior & Francoise Beaufays (2014) Long Short-Term Memory Recurrent Neural Network Architectures for Large Scale Acoustic Modeling.

Schrodt, Philip A (2012) CAMEO Event (http://eventdata.psu.edu/).

Schrodt, Philip A & David Van Brackle (2013) Automated Coding of Political Event Data. In: Handbook of Computational Approaches to Counterterrorism. New York, NY: Springer New York, 23–49 (http://link.springer.com/10.1007/978-1-4614-5311-6_2).

Smith, Emmanuel M, Jim Smith, Phil Legg & Simon Francis (2018) Predicting the Occurrence of World News Events Using Recurrent Neural Networks and Auto-Regressive Moving Average Models. Advances in Intelligent Systems and Computing 650(Ml): 191–202.

Szayna, Thomas, Stephen Watts, Angela O'Mahony, Bryan Frederick & Jennifer Kavanagh (2017) What Are the Trends in Armed Conflicts, and What Do They Mean for U.S. Defense Policy? (https://www.rand.org/pubs/research_reports/RR1904.html).

Tikuisis, Peter, David Carment & Yiagadeesen Samy (2013) Prediction of Intrastate Conflict Using State Structural Factors and Events Data. Journal of Conflict Resolution Vol. 57.

Ward, Michael D, Andreas Beger, Josh Cutler, Matthew Dickenson, Cassy Dorff & Ben Radford (2013) Comparing GDELT and ICEWS Event Data. Analysis Vol. 21 (http://mdwardlab.com/biblio/comparing-gdelt-and-icews-event-data).

Yonamine, James E (2013) Predicting Future Levels of Violence in Afghanistan Districts. The 3rd Annual Meeting Of The European Political Science Association (http://jayyonamine.com/wp-content/uploads/2013/03/Forecasting_Afghanistan.pdf).

## List of figures

## List of tables

## Annexes

### Annex 1. The GDELT Dataset Qualitative Validation.

This in depth GDELT validation-analysis of a sample of eight countries uses the three most severe CAMEO event root codes representing actual instances of violence (18, 19, and 20). This validation can only be carried out for the GDELT since the ICEWS and Phoenix do not provide the URL sources of their event records.

The eight selected countries (see Table 4) are: Bangladesh, Botswana, Chad, Colombia, Guatemala, Jordan, Myanmar, and Tajikistan. The Marshall Islands, a very small country belonging to the MD region "Other" and with a GCRI conflict probability and intensity of zero, was used as a control country, given that a query for a country with these characteristics would be expected to return zero results. These selected nations cover all six MD regions, all deciles of the GCRI conflict probability, and a wide variety of the GCRI conflict intensity level, population size, and economic development (aggregate PPP GDP and per capita GDP).

**Table 4.** Structural conditions and GCRI probability and Intensity

| Country | Region | GCRI % | GCRI Intensity | Population | GDP (Billion $) | GDP Per Cap ($) | English |
|---|---|---|---|---|---|---|---|
| **Bangladesh** | MD I | 47 | 6 | 159,453,001 | 690.300 | 4,200 | Yes |
| **Botswana** | MD II | 1 | 4 | 2,249,104 | 39.010 | 17,000 | Yes |
| **Chad** | MD II | 95 | 10 | 15,833,116 | 28.620 | 2,300 | No |
| **Colombia** | MD V | 76 | 7 | 48,168,996 | 711.600 | 14,400 | No |
| **Guatemala** | MD V | 33 | 7 | 16,581,273 | 138.100 | 8,200 | No |
| **Jordan** | MD IV | 19 | 6 | 10,458,413 | 89.000 | 9,200 | Yes |
| **Myanmar** | MD I | 80 | 8 | 55,622,506 | 329.800 | 6,300 | No |
| **Tajikistan** | MD III | 65 | 7 | 8,604,882 | 28.430 | 3,200 | No |
| **Marshall Islands** | Other | 0 | 0 | 75,684 | 0.196 | 3,600 | Yes |

At least 700 returns per country and a minimum of one complete month's results were evaluated, with additional months considered if the control month of January 2019 did not return enough results. Each article was individually read and categorized into one of five types: usable, possibly usable, unreliable source or invalid link, useless, and duplicates (see Table 5).

"Usable" returns clearly contained information relevant to the query imputed: violent events which took place in the specified location and time period. "Possibly usable" returns either contained relevant information about events which took place outside of the specified time and location, or detailed events whose relevancy can be questioned but not discredited. Returns categorized under "unreliable source or invalid link" contained articles hosted on platforms known to harbor fake news (such as Infowars) or which cannot be accessed by users in the European Union.

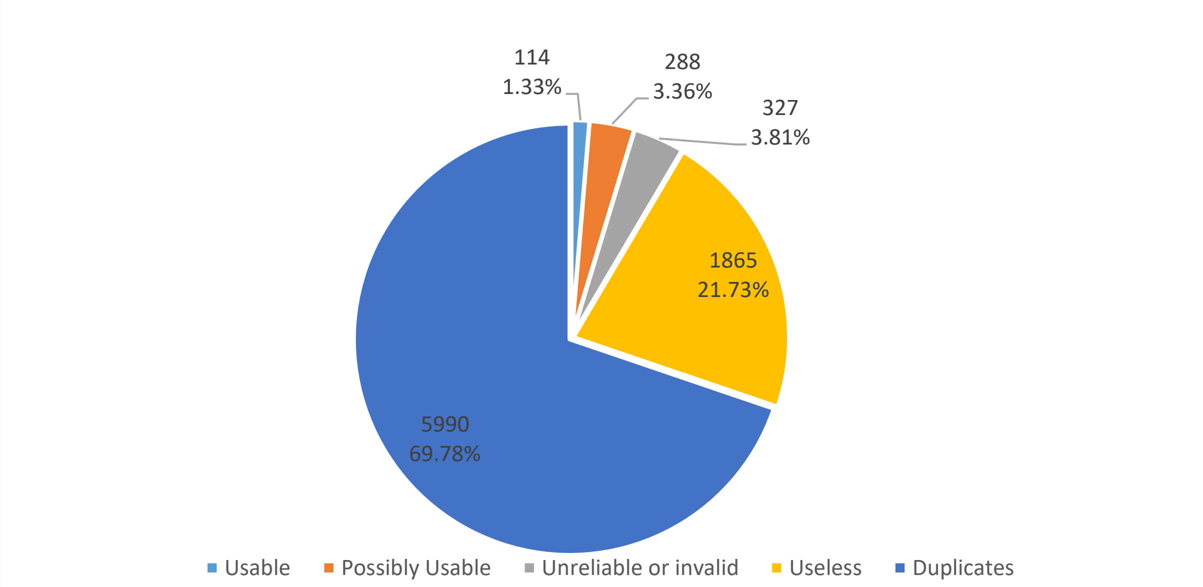
"Usable" returns were then evaluated on the basis of whether or not they received the proper event code and whether or not they pertained to the correct country.

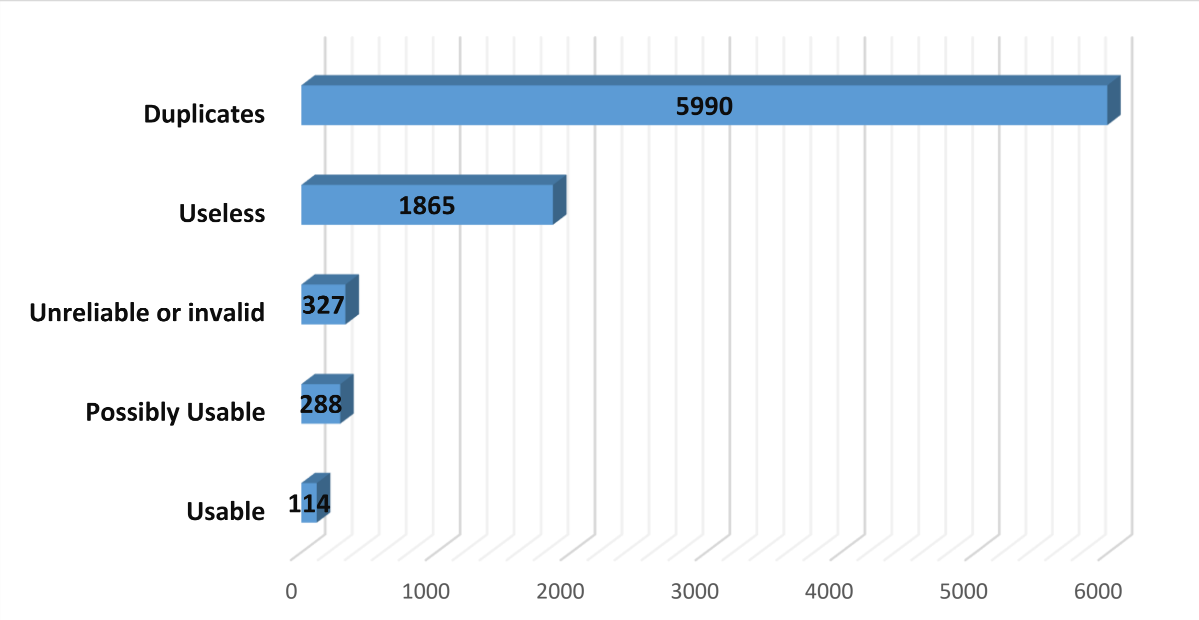
**Table 5.** GDELT data validation - degree of usability

| Country | Region | GCRI % | GCRI Intensity | Date | Total | Usable | Possible Usable | Unreliable or invalid | Useless | Duplicates |
|---|---|---|---|---|---|---|---|---|---|---|
| **Bangladesh** | MD I | 47 | 5.9 | 201901 | 2414 | 27 | 76 | 42 | 390 | 1879 |
| **Botswana** | MD II | 1 | 4.3 | 201901 | 53 | 3 | 3 | 1 | 27 | 18 |
| **Botswana** | MD II | 1 | 4.3 | 201809 | 176 | 1 | 1 | 5 | 45 | 124 |
| **Botswana** | MD II | 1 | 4.3 | 201703 | 186 | 3 | 5 | 5 | 71 | 102 |
| **Botswana** | MD II | 1 | 4.3 | 201604 | 177 | 1 | 4 | 32 | 54 | 86 |
| **Botswana** | MD II | 1 | 4.3 | 201308 | 82 | 0 | 1 | 21 | 16 | 44 |
| **Chad** | MD II | 95 | 10 | 201901 | 514 | 6 | 17 | 7 | 143 | 341 |
| **Chad** | MD II | 95 | 10 | 201710 | 464 | 4 | 16 | 22 | 97 | 325 |
| **Chad** | MD II | 95 | 10 | 201701 | 498 | 4 | 15 | 17 | 44 | 341 |
| **Colombia** | MD V | 76 | 6.7 | 201901 | 947 | 8 | 26 | 27 | 203 | 682 |
| **Guatemala** | MD V | 33 | 7.1 | 201901 | 318 | 5 | 13 | 21 | 61 | 217 |
| **Guatemala** | MD V | 33 | 7.1 | 201708 | 340 | 8 | 10 | 11 | 85 | 226 |
| **Guatemala** | MD V | 33 | 7.1 | 201611 | 172 | 4 | 5 | 12 | 58 | 93 |
| **Jordan** | MD IV | 19 | 5.7 | 201901 | 681 | 0 | 24 | 18 | 314 | 324 |
| **Myanmar** | MD I | 80 | 7.6 | 201901 | 776 | 18 | 21 | 25 | 111 | 601 |
| **Tajikistan** | MD III | 65 | 6.8 | 201901 | 67 | 3 | 9 | 1 | 12 | 42 |
| **Tajikistan** | MD III | 65 | 6.8 | 201807 | 287 | 4 | 10 | 12 | 45 | 213 |
| **Tajikistan** | MD III | 65 | 6.8 | 201802 | 69 | 3 | 4 | 1 | 15 | 46 |
| **Tajikistan** | MD III | 65 | 6.8 | 201708 | 113 | 3 | 4 | 12 | 14 | 80 |
| **Tajikistan** | MD III | 65 | 6.8 | 201703 | 81 | 7 | 6 | 10 | 17 | 41 |
| **Tajikistan** | MD III | 65 | 6.8 | 201609 | 241 | 2 | 17 | 24 | 37 | 161 |
| **Marshall Islands** | Other | 0 | 0 | 201901 | 12 | 0 | 1 | 1 | 6 | 4 |
| **Total:** | | | | | **8668** | **114** | **288** | **327** | **1865** | **5990** |

The most immediate takeaway was the number of articles that were just duplicates of others previously listed. Duplicates accounted for 5,990 out of the 8,668 total returns, or nearly 70% (see Figure 36). This includes returns that display the same article, reposted identical articles, and articles chronicling the same event.

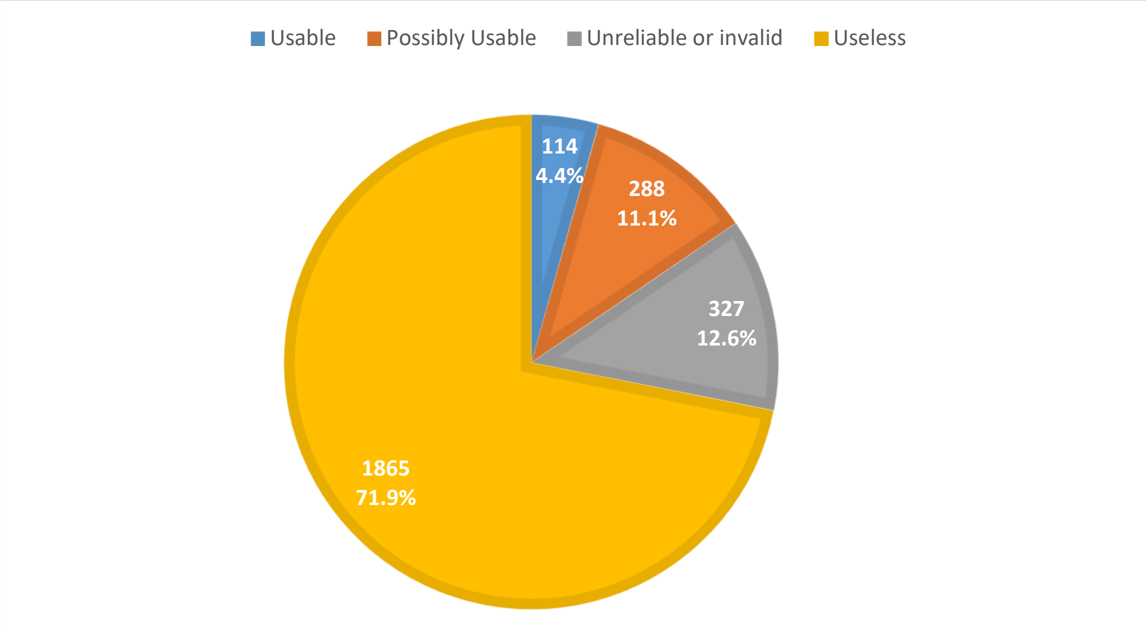
**Figure 36.** GDELT data validation - degree of usability




Of the non-duplicate returns, 1,865 – or nearly 72% – were found to be completely useless (see Figure 37). These returns were entirely irrelevant to the query, but often included individual keywords that reasonably may have caused them to be included in the results.

**Figure 37.** GDELT data validation - degree of usability without duplicates




Another noticeable trend was the inclusion of completely irrelevant articles written by authors or naming subjects having the same name as the country specified in the query, such as Jordan or Chad. Returns in the aforementioned camp often displayed Action Geographical Location Names, Latitudes, and Longitudes of the queried country even though the articles tended to concern matters elsewhere – typically the United States where these names are common.

From the results and the validation, it is apparent that GDELT has the potential to be the comprehensive database necessary to bring the dynamic GCRI, but many provisions must be added to the fundaments of any program using this database in order to render it accurate and effective. The first priority should be to eliminate the abundant duplicate returns. These redundancies clutter the dataset and make it fairly difficult for a human to check and more time consuming for an algorithm to process. Additionally, these empty data points could become a source of confusion and contaminate the data. Eliminating returns with the exact same source is as simple and straightforward as inputting a line of code that filters them out automatically. Weeding out repetitive and reposted articles may be slightly more of a challenge, but one way of doing so is having a program automatically follow the provided source URLs and run a simple plagiarism checking script. If the percentage of similarity reaches a certain threshold, for instance 65%, the article is categorized as a duplicate and filtered out. These methods can retain just the first instance of an event or can give priority to exceptionally reputable sights such as Reuters and Al Jazeera.

Reducing the amount of useless articles will pose a challenge due to the inherent nature of GDELT's search algorithm, but doing so is certainly possible. The easiest type of false return to address is the country name / popular name issue, which could be done by having a program check for the source author name and automatically remove those written by people with the same name as the queried country. Restricting the number of websites and publications considered could cut down the number of irrelevant results, particularly as it is unlikely to find pertinent event information on sports sites or tabloids, for example.

These site filtering provisions could also include sites of dubious quality or ill reputation. Adding known fake news sites and automatically discrediting things like private blogs could reduce the number of unreliable sources and prevent results from being negatively affected. The issue of invalid links can be entirely eliminated by inserting a ping function into the program, which automatically checks if the link works. The ping function can also serve as the first step of the automated source evaluation, saving processing power by not having to check sites that are inutile.

Possibly usable sources are two-edged; they can provide useful information but also negatively impact results. Using a program to double check where and when the event described in a GDELT return took place could assist in this matter, particularly as GDELT's location identification cannot be relied upon. In fact, GDELT's ActionGeo latitude and longitude categories should be disregarded entirely as they are almost never accurate. This was exemplified by the fact that entirely irrelevant returns about other countries' local affairs written by foreign authors with the same name as the country were still sorted into the country's geographical identifiers.

Refining GDELT's results in order to obtain a ratio greater than 1.33% of usable results will require the implementation of a comprehensive system to ensure the quality and value of the sources GDELT provides in addition to the aforementioned provisions. First, there must be a method of verifying the event codes with which GDELT classifies sources, as 54% of the usable results from this experiment were listed under the wrong code. More specifically, any return with the 190 CAMEO event code, i.e. unspecified acts of violence, should come under extra scrutiny due to the tendency for this classification to be incorrectly prescribed. Inherently, a code representing unspecified acts of violence will nearly always coincide with inutile returns, as any article chronicling a military altercation will likely be more specific about the nature of the confrontation. This means that it may be worth considering to eliminate consideration of returns marked with the 190 event code entirely.

Finally, a visible shortcoming of the GDELT system is that it does not deliberately identify or classify organized criminal activity. In order to attain the project goals of the Peace and Stability team, a program to identify and include organized criminal rackets such as drug trafficking, smuggling, poaching, piracy, and human trafficking in datasets must be developed as well; in an ever more globalized world, conflict and criminality are increasingly interwined.

## Annex 2. Countries Ranking based on CRA-S

Here is an example of the CRA-S classification system based on the GDELT data. The current month here is August 2019 and the predictions refer to the next month.

| | Initial Ranking | ISO3C | Official name | RegionCode | Region | EEAS.MD.3 | delta% | 3_alarms | 2_alarms | 1_alarms |
|---|---|---|---|---|---|---|---|---|---|---|
| 2 | 108 | MDV | the Republic of Maldives | AP | Asia and the Pacific | MD I Asia & the Pacific | 35.01015114 | 0 | 1 | 0 |
| 3 | 88 | KGZ | the Kyrgyz Republic | EUCA | Western Balkans, Eastern Europe and Central Asia | MD III Europe & Central Asia | 21.4546102 | 0 | 1 | 0 |
| 4 | 95 | LBN | the Lebanese Republic | MENA | North Africa and Middle East | MD IV North Africa, Middle East, Arabian Peninsula, Iran & Iraq | 12.2198515 | 0 | 1 | 0 |
| 5 | 187 | ZMB | the Republic of Zambia | ESAF | East and Southern Africa | MD II Africa | 9.601475834 | 0 | 1 | 0 |
| 7 | 179 | UZB | the Republic of Uzbekistan | EUCA | Western Balkans, Eastern Europe and Central Asia | MD III Europe & Central Asia | 4.866291943 | 0 | 1 | 0 |
| 9 | 37 | COM | the Union of the Comoros | ESAF | East and Southern Africa | MD II Africa | 250 | 0 | 0 | 1 |
| 10 | 23 | BRB | Barbados | Other | Other | Other | 71.95968645 | 0 | 0 | 1 |
| 11 | 65 | GNB | the Republic of Guinea-Bissau | WCA | West and Central Africa | MD II Africa | 67.70833333 | 0 | 0 | 1 |
| 12 | 38 | CPV | Republic of Cabo Verde | Other | Other | Other | 57.14285714 | 0 | 0 | 1 |
| 13 | 91 | KNA | Saint Kitts and Nevis | Other | Other | Other | 51.55555556 | 0 | 0 | 1 |
| 14 | 135 | PLW | the Republic of Palau | Other | Other | Other | 33.33333333 | 0 | 0 | 1 |
| 15 | 89 | KHM | the Kingdom of Cambodia | AP | Asia and the Pacific | MD I Asia & the Pacific | 29.00835021 | 0 | 0 | 1 |
| 16 | 61 | GEO | Georgia | EUCA | Western Balkans, Eastern Europe and Central Asia | MD III Europe & Central Asia | 28.93937325 | 0 | 0 | 1 |
| 19 | 101 | LSO | the Kingdom of Lesotho | ESAF | East and Southern Africa | MD II Africa | 26.59136546 | 0 | 0 | 1 |
| 20 | 130 | OMN | the Sultanate of Oman | MENA | North Africa and Middle East | MD IV North Africa, Middle East, Arabian Peninsula, Iran & Iraq | 25.59458969 | 0 | 0 | 1 |
| 21 | 147 | SEN | the Republic of Senegal | WCA | West and Central Africa | MD II Africa | 24.88888889 | 0 | 0 | 1 |
| 22 | 10 | BDI | the Republic of Burundi | WCA | West and Central Africa | MD II Africa | 24.83563445 | 0 | 0 | 1 |
| 23 | 68 | GRD | Grenada | Other | Other | Other | 23.65451389 | 0 | 0 | 1 |
| 25 | 114 | MMR | the Republic of the Union of Myanmar | AP | Asia and the Pacific | MD I Asia & the Pacific | 18.87538914 | 0 | 0 | 1 |
| 28 | 174 | TZA | the United Republic of Tanzania | ESAF | East and Southern Africa | MD II Africa | 17.13044974 | 0 | 0 | 1 |
| 29 | 6 | ATG | Antigua and Barbuda | Other | Other | Other | 16.07142857 | 0 | 0 | 1 |
| 30 | 51 | ERI | the State of Eritrea | ESAF | East and Southern Africa | MD II Africa | 14.68555226 | 0 | 0 | 1 |
| 31 | 160 | SWZ | the Kingdom of Eswatini | ESAF | East and Southern Africa | MD II Africa | 12.16931217 | 0 | 0 | 1 |
| 32 | 145 | SAU | the Kingdom of Saudi Arabia | MENA | North Africa and Middle East | MD IV North Africa, Middle East, Arabian Peninsula, Iran & Iraq | 11.28800793 | 0 | 0 | 1 |
| 33 | 185 | YEM | the Republic of Yemen | MENA | North Africa and Middle East | MD IV North Africa, Middle East, Arabian Peninsula, Iran & Iraq | 11.20586519 | 0 | 0 | 1 |
| 34 | 151 | SLV | the Republic of El Salvador | LAC | Latin America and the Caribbean | MD V Americas | 10.74379703 | 0 | 0 | 1 |
| 35 | 3 | ARE | the United Arab Emirates | MENA | North Africa and Middle East | MD IV North Africa, Middle East, Arabian Peninsula, Iran & Iraq | 10.48097117 | 0 | 0 | 1 |
| 38 | 109 | MEX | the United Mexican States | LAC | Latin America and the Caribbean | MD V Americas | 8.366376917 | 0 | 0 | 1 |
| 39 | 143 | RUS | the Russian Federation | EUCA | Western Balkans, Eastern Europe and Central Asia | MD III Europe & Central Asia | 8.321799603 | 0 | 0 | 1 |
| 40 | 81 | ISR | the State of Israel | MENA | North Africa and Middle East | MD IV North Africa, Middle East, Arabian Peninsula, Iran & Iraq | 8.229859265 | 0 | 0 | 1 |
| 42 | 141 | QAT | the State of Qatar | MENA | North Africa and Middle East | MD IV North Africa, Middle East, Arabian Peninsula, Iran & Iraq | 7.716706208 | 0 | 0 | 1 |
| 43 | 84 | JOR | the Hashemite Kingdom of Jordan | MENA | North Africa and Middle East | MD IV North Africa, Middle East, Arabian Peninsula, Iran & Iraq | 7.32244916 | 0 | 0 | 1 |
| 44 | 123 | NGA | the Federal Republic of Nigeria | WCA | West and Central Africa | MD II Africa | 6.953574581 | 0 | 0 | 1 |
| 46 | 128 | NRU | the Republic of Nauru | Other | Other | Other | 6.25 | 0 | 0 | 1 |
| 47 | 64 | GMB | the Republic of the Gambia | WCA | West and Central Africa | MD II Africa | 6.092906509 | 0 | 0 | 1 |
| 48 | 127 | NPL | the Federal Democratic Republic of Nepal | AP | Asia and the Pacific | MD I Asia & the Pacific | 5.872956584 | 0 | 0 | 1 |
| 49 | 117 | MRT | the Islamic Republic of Mauritania | WCA | West and Central Africa | MD II Africa | 5.594832853 | 0 | 0 | 1 |
| 50 | 79 | IRQ | the Republic of Iraq | MENA | North Africa and Middle East | MD IV North Africa, Middle East, Arabian Peninsula, Iran & Iraq | 5.57233085 | 0 | 0 | 1 |
| 51 | 121 | NAM | the Republic of Namibia | ESAF | East and Southern Africa | MD II Africa | 5.362974741 | 0 | 0 | 1 |
| 53 | 134 | PHL | the Republic of the Philippines | AP | Asia and the Pacific | MD I Asia & the Pacific | 3.827596974 | 0 | 0 | 1 |
| 54 | 78 | IRN | the Islamic Republic of Iran | MENA | North Africa and Middle East | MD IV North Africa, Middle East, Arabian Peninsula, Iran & Iraq | 3.659534219 | 0 | 0 | 1 |
| 55 | 163 | TCD | the Republic of Chad | WCA | West and Central Africa | MD II Africa | 3.245917063 | 0 | 0 | 1 |
| 56 | 5 | ARM | the Republic of Armenia | EUCA | Western Balkans, Eastern Europe and Central Asia | MD III Europe & Central Asia | 2.964278747 | 0 | 0 | 1 |
| 57 | 36 | COL | the Republic of Colombia | LAC | Latin America and the Caribbean | MD V Americas | 2.137828035 | 0 | 0 | 1 |

**Annex 3. Conflict and Mediation Event Observations (CAMEO) Classification**

The entire CAMEO event taxonomy (Schrodt, 2012) is broken into four primary classifications. This primary classification allows an analysis at the highest level of aggregation. Additionally, there are 20 major subcategories, each one subdivided into several ones, to create a detailed classification scale.

| Verbal Cooperation | Material Cooperation | Verbal Conflict | Material Conflict |
|---|---|---|---|
| 1-MAKE PUBLIC STATEMENT | 6-ENGAGE IN MATERIAL COOPERATION | 10-DEMAND | 15-EXHIBIT MILITARY POSTURE |
| 2-APPEAL | 7-PROVIDE AID | 11-DISAPPROVE | 16-REDUCE RELATIONS |
| 3-EXPRESS INTENT TO COOPERATE | 8-YIELD | 12-REJECT | 17-COERCE |
| 4-CONSULT | 9-INVESTIGATE | 13-THREATEN | 18-ASSAULT |
| 5-ENGAGE IN DIPLOMATIC COOPERATION | | 14-PROTEST | 19-FIGHT |
| | | | 20-ENGAGE IN UNCONVENTIONAL MASS VIOLENCE |

**GETTING IN TOUCH WITH THE EU**

**In person**

All over the European Union there are hundreds of Europe Direct information centres. You can find the address of the centre nearest you at: https://europa.eu/european-union/contact_en

**On the phone or by email**

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

- by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),

- at the following standard number: +32 22999696, or

- by electronic mail via: https://europa.eu/european-union/contact_en

**FINDING INFORMATION ABOUT THE EU**

**Online**

Information about the European Union in all the official languages of the EU is available on the Europa website at: https://europa.eu/european-union/index_en

**EU publications**
You can download or order free and priced EU publications from EU Bookshop at: https://publications.europa.eu/en/publications. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see https://europa.eu/european-union/contact_en).

KJ-NA-30011-EN-N

## The European Commission's science and knowledge service

Joint Research Centre

### JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.

**EU Science Hub**
ec.europa.eu/jrc

@EU_ScienceHub

EU Science Hub - Joint Research Centre

EU Science, Research and Innovation

EU Science Hub

Publications Office
of the European Union

doi:10.2760/846412
ISBN 978-92-76-14133-4