

# Conflict Event Modelling: Research Experiment and Event Data Limitations

## Matina Halkia, Stefano Ferri, Michail Papazoglou, Marie-Sophie Van Damme, Dimitrios Thomakos

European Commission, Joint Research Centre (JRC), European Commission, Joint Research Centre (JRC), Unisystems S.A, Pikel Ltd, Unisystems S.A  
Via E. Fermi 2749 Ispra Italy, Via E. Fermi 2749 Ispra Italy, Via Michelangelo Buonarroti 39 Milano Italy, Via Breda 176 Milano Italy, Rue Edward Steichen 26, Luxembourg  
{Matina.HALKIA, Stefano.FERRI}@ec.europa.eu  
{Michail.PAPAZOGLU, Marie-Sophie.VAN-DAMME, Dimitrios.THOMAKOS}@ext.ec.europa.eu

### Abstract

This paper presents the conflict event modelling experiment, conducted at the Joint Research Centre of the European Commission, particularly focusing on the limitations of the input data. This model is under evaluation as to potentially complement the Global Conflict Risk Index (GCRI), a conflict risk model supporting the design of European Union’s conflict prevention strategies. 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. It adopts a Long-Short Term Memory Cell Recurrent Neural Network on country-level actor-based event datasets that indicate potential triggers to violent conflict such as demonstrations, strikes, or elections-related violence. The observed data and the outcome of the model predictions consecutively, consolidate an early warning alarm system that signals abnormal social unrest upheavals, and appears promising as an approach towards a conflict trigger model. However, event-based systems still require overcoming certain obstacles related to the quality of the input data and the event classification method.

**Keywords:** Early Warning System, actor-based event datasets, conflict prevention

## 1. Introduction

Quantitative modelling studies aiming to predict future conflicts, consider the number of casualties as a proxy for 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., 2020). While the correlation between conflict intensity and the number of battle casualties is plausible, it does not consider conflict development stages (Qiao et al., 2017), as well as the complexity of events like protests, demonstrations, election violence, or even tension relief events such as diplomatic cooperation.

The Peace and Stability team at the European Commission’s Joint Research Centre has developed a conflict event modelling algorithm (Halkia et al., 2019), which unlike the original structural conflict risk model based on statistical regressions (Halkia et al., 2017b, a, 2020), integrates and disentangles every stage of the conflict development or de-escalation cycle.

In this paper, we discuss two available news media datasets tested for this experimental conflict event model and their limitations: (i) the Global Data on Events Location and Tone (GDELT) project and (ii) the Integrated Crisis Early Warning System (ICEWS) Dataverse dataset. Both are based on the Conflict and Mediation Event Observations Event and Actor Codebook (CAMEO) classification.

The CAMEO codebook classifies event data in four primary classes, called QuadClass, i.e. verbal cooperation ( $Q_1$ ), material cooperation ( $Q_2$ ), verbal conflict ( $Q_3$ ), and material conflict ( $Q_4$ ). These primary classes are

subdivided into 20 major categories and 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). Social unrest events, that initially start as a demonstration to the public or the government, often escalate into general chaos, resulting in violence, riots, sabotage, and other forms of crime and social disorder. The Deep Learning (DL) methodology adopted to model the actor-based conflict events is a Long-Short Term Memory (LSTM) Cell Recurrent Neural Network (RNN). Besides this DL model, we have set up an early warning alarm system to signal abnormal social unrest upheavals.

Although the experimental conflict event model, through the DL and early warning alarm, is able to predict the materialization of a conflict on a monthly basis, event-based systems require supplementary research to offset the databases’ shortcomings, such as automated data validation, new classifiers and dictionaries.

Section 2 presents the various datasets and their limitations; Section 3 explains the model and methodology proposed for the experimental conflict event model, whereas Section 4 presents the results. Finally, Section 5 and 6 respectively discuss the model’s feasibility and steps forward.

## 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) and the Integrated Crisis Early Warning

System (ICEWS) Dataverse (ICEWS) are arguably the largest currently available event data collections in social science, which gather broad 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. In order to fulfil the purposes of this paper, we investigate the use of these two news media datasets as possible inputs for the experimental conflict event model and discuss their limitations<sup>1</sup>.

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

The GDELT project is an on-going attempt to monitor print, broadcast web news media in over 100 languages from all over the world, almost real-time (GDELT, 2019). It is worth emphasizing that the GDELT dataset is updated every 15 minutes, meaning that there is a continuous flow of records integrating the database.

So far, researchers rarely aim at utilizing GDELT to make predictions about social unrest and only a few scholars have conducted predictions using GDELT, as it has been criticised for abundant duplicate returns. Alikhani (2014) attempted to use GDELT with linear regression models, while Yonamine (2013) studied the dataset for time series forecasting. More recent papers use GDELT for frequent subgraphs mining (Qiao et al., 2017) or artificial intelligence applications (Smith et al., 2018).

To our knowledge, one study has used the GDELT data to measure conflict intensity (Levin, Ali & Crandall, 2018). In their article, however, Levin, Ali and Crandall consider the monthly time series of the absolute number of events occurring in the CAMEO Q<sub>4</sub> 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). This conflict event modelling approach presented here, 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 major shortcoming 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). Furthermore, the automated codebook algorithm is not publicly available, which does not allow investigation on

the source of potential errors in the news classification. However, as the source URL is given, we can undertake sample validation tests in order to detect misclassified events.

### 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). ICEWS 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).

The ICEWS program is temporally restricted as it scans news on a daily basis only since October 2018 and on a monthly basis since 1995. No data is available before 1995. Another limitation is that validating event classification is cumbersome and does not facilitate identifying the source of the record for the analyst. Hence, the cost of the validation effort is disproportionate.

According to Ward et al. (2013), who compared the GDELT and ICEWS datasets, *“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”*.

## 3. Methodology

### 3.1 Deep Learning Event based Modelling

Most of the social unrest events, that initially start as a public demonstration against the government, often escalate into general chaos, resulting in riots, sabotage, and other forms of crime and social disorder (Qiao et al., 2017). Predicting these events can therefore be formulated as a sequence classification problem that identifies any possible 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. CAMEO classifies significant occurrence of supportive statements, requests and engagements in diplomatic cooperation as Verbal Cooperation (Q<sub>1</sub>), while collaborative investigations, engagements in material cooperation, and provision of aid as Material Cooperation

---

<sup>1</sup> There are more available datasets but we cannot overcome their present limitations i.e. limited time series (e.g. Phoenix\_RT from Oct. 2017 to today, Cline Center Historical Phoenix Event Data

(Phoenix) up to 2015), and limited sources (e.g. Cline Center Historical Phoenix Event Data).

(Q<sub>2</sub>). On the other hand, disapprovals, objections, and complaints, threats, rejection of cooperation and civilian demonstrations are assigned to Verbal Conflict (Q<sub>3</sub>). Finally, CAMEO names as Material Conflict (Q<sub>4</sub>) all military or police moves, repression and violence against civilians, use of conventional and unconventional forms of violence as well as mass violence (Schrodt, 2012).

When observing an increasing trend of news articles reported as Verbal Conflict (Q<sub>3</sub>) and/or Material Conflict (Q<sub>4</sub>) with respect to the total number of articles, the model is able to measure an increase in conflict related tensions.

The historical time series of Verbal Cooperation (Q<sub>1</sub>), Material Cooperation (Q<sub>2</sub>), Verbal Conflict (Q<sub>3</sub>), and Material Conflict (Q<sub>4</sub>) quantify and allow us to grasp the direction in which the tensions evolve in order to predict future conflict events. The DL methodology adopted to the experimental conflict event model, is a Long-Short Term Memory (LSTM) Cell Recurrent Neural Network (RNN). LSTM models are well suited to classify, process, and make predictions based on time series data and forecast near-future events (Hochreiter & Schmidhuber, 1997; Chung et al., 2014).

By applying this model to conflict prediction, our implicit assumption is that the current situation depends on the conflict history. We could think of it as a composite function in which the oldest 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 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 appropriate 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.

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. Next, we transformed the absolute number of articles of each major category and QuadClass into proportion of the total number of articles. We consider as independent variables all the 20 major categories and the 4 QuadClasses as defined in the CAMEO and applied on the GGELT and ICEWS datasets. By doing so, we provided as input to the model 24 independent variables as time-month records per country.

We have filtered the GDELT dataset to include only the available information after 1989, which is the starting year of the original GCRI input values, enabling a comparative analysis of the results obtained in the respective models.

In a fourth step, we create 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. 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. The model consists of 191 separate models, one per each country included in the analysis. We then use the Root Mean Square Error (RMSE) to validate the accuracy of the model<sup>2</sup>. The model will be fit for 500 training epochs with a batch size of 72.

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

Through the DL 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). The CRA-S signals social unrest upheavals (an abnormal increase in the proportion of the Q<sub>4x</sub>, Q<sub>4x+1</sub>) or the media pressure variations (increase or decrease in the total number of events mentioned as Q<sub>1x</sub>, Q<sub>2x</sub>, Q<sub>3x</sub> or Q<sub>4x</sub>, where x stands for the point in time). This allows policy makers to implement short-term preventive actions to mitigate conflict exacerbations at an earlier stage of the conflict development cycle.

The CRA-S functions can be detailed as follows: First, we aggregate the events recorded in each of the two databases per month. Next, we compute the monthly amount of news

---

<sup>2</sup> The root mean square error (RMSE) has been widely used as a standard statistical metric to measure model performance in various studies (Chai, Chai & Draxler, 2014). In order to validate

the accuracy of the model we lagged the dataset by a month, so the prediction refers to the last available month.

articles reported in each QuadClass ( $Q_1$ ,  $Q_2$ ,  $Q_3$ , and  $Q_4$ ) 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 total number of 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.

The calculation of different time windows gives the opportunity to evaluate the predictions performance under different timeframes (see results in 4.2). 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  $Q_4$  of the current month and the  $Q_4$  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  $Q_4$  for the current month and  $Q_{4x-1}$  is the proportion of the  $Q_4$  for the previous month. Based on this rate, we rank the countries so as the country with the highest increase in the  $Q_4$  ( $\Delta Q_4$ ) will be first and the country with the highest decrease will be the last. In case the  $Q_{4x}$  is a local max, meaning that the increase in the current value of the  $Q_4$  is significant, we have set an alarm following the same methodology as described in the section 3.2.

To the initial country ranking, we further add a set alarms (value 0 or 1 if true) that consist of the following parameters:

- **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% CI that we have calculated for the x-month moving window.
- **Alarm 2:** The total absolute number of the events mentioned (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:

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

Taking into account the presence or absence of the abovementioned alarms reflecting different time windows the analyst has a choice in either summarising the news, having long term predictions or a more detailed overview of the situation. As a result, 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.

## 4. Results

### 4.1 Deep Learning Root Mean Square Error (RMSE) and Model Predictions

The results of running the LSTM RNN model for five study cases and two databases are listed in the last column of Table 1 for March 2019. To see how accurate our model predictions are in estimating the percentage of material conflict events ( $Q_4$  of the CAMEO classification) in percentage of the overall events, we measure the Root Mean Square Error (RMSE)<sup>3</sup> of the model, which is the standard deviation of the residuals (prediction errors).

Country	Dataset	RMSE
Libya	ICEWS	0.215
	GDELT	0.089
Sudan	ICEWS	0.097
	GDELT	0.041
Egypt	ICEWS	0.119
	GDELT	0.059
Maldives	ICEWS	0.210
	GDELT	0.071
Nicaragua	ICEWS	0.147
	GDELT	0.063

<sup>3</sup> RMSE is a measure of how spread out the residuals are. In other words, it tells you how concentrated the data is around the line of best fit.

Table 1: RMSE for March 2019 per dataset

As reported in Table 1, the RMSE using the GDELT data is the lowest in all the case studies. In other words, the predictions based on this dataset are closer to the observed values. We postulate that due to the limited data availability (1995-2019) in the ICEWS database, the model using the GDELT data is more precise and accurate.

#### 4.2 Early Warning Alarm System Predictions and Accuracy

In this part of the paper, the results of the early warning alarm system (an abnormal increase in the proportion of  $Q_4$  – material conflict events) are presented for five case study scenarios and two databases. As we described above, an alarm is given in case there is an extraordinary increase in the predictions out of the 95% 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 2, 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 in 2018, and the Nicaraguan protests the same year. For this experiment, we have pre-filtered the GDELT dataset in order to obtain more reliable input data. Based on our GDELT validation (Halkia et al. 2019), we have set a filter on 100 mentions per article. In other words, when the filter is applied, only articles that have been mentioned more than 100 times are included in that part of the analysis (GDELT100 in Table 2). This has been done to remove information noise in the GDELT database, hypothesizing that if an event really happens, it should be reproduced by more than one media source and in more than one article. 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 inclusion of all available information within the GDELT database affects significantly the results<sup>4</sup>.

ISO3	Date	Dataset	3-month local max	6-month local max
LBY	Feb 2011	GDELT	<b>ALARM</b>	<b>ALARM</b>
		GDELT100	<b>ALARM</b>	<b>ALARM</b>
		ICEWS	NO ALARM	NO ALARM
SDN	Dec 2018	GDELT	NO ALARM	NO ALARM
		GDELT100	<b>ALARM</b>	<b>ALARM</b>
		ICEWS	NO ALARM	NO ALARM
EGY	Jan 2011	GDELT	NO ALARM	NO ALARM
		GDELT100	<b>ALARM</b>	<b>ALARM</b>
		ICEWS	NO ALARM	NO ALARM

<sup>4</sup> We are not able to do the same filtering for ICEWS due to dataset limitations (no available URL).

MDV	Feb 2018	GDELT	NO ALARM	NO ALARM
		GDELT100	NO ALARM	NO ALARM
		ICEWS	NO ALARM	NO ALARM
NIC	Apr 2018	GDELT	NO ALARM	NO ALARM
		GDELT100	NO ALARM	NO ALARM
		ICEWS	<b>ALARM</b>	<b>ALARM</b>

Table 2: Sample validation of the predictions based on past events.

In Table 2, 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. In contrast, the ICEWS dataset could only predict the 2018–2019 Nicaraguan protests. Overall, using the filter improves the reliability of the GDELT database and enhances the model towards more accurate predictions. In addition, the 3-month and 6-month window estimations render similar predictions, demonstrating that both time frames report equal results.

## 5. Discussion

In this paper, the experimental conflict event model, which is based on country-level actor-based event data, signalling a potential trigger (including demonstrations, strikes, election violence, etc.) to violent conflict, has been discussed, along with its present shortcomings.

The experiment presented here, using an LSTM RNN model to predict the materialization of a conflict, demonstrates that the GDELT is potentially the most comprehensive database, probably due to the amount of available information, despite all its limitations, including information noise. However, in some cases (islands and small countries), where the reporting is limited, social upheaval prediction may be challenging.

Nevertheless, many provisions must be added to any method using the GDELT database in order to render it accurate and effective. The LSTM RNN model we propose, one of the most advanced neural networks for modelling temporal sequences and their long-range dependencies, performs well and is able to handle time series data and classify each event based on historical information. While the absolute number of events informs of a significant escalation or de-escalation of tension in a given country, the normalized number of events provides information on the relative significance of the occurrence to preceding ones. The proportions taken by different types of events complete 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.

The results indicate that the model is able to correctly predict social upheaval in countries where there is available information on news (Libya, Sudan, Egypt). In contrast, for the cases where there is very little or no available information (Nicaragua, Maldives), the model fails to detect upheavals. The next steps to further evaluate the performance and robustness of the model will consider a k-fold cross validation and an ANOVA (ANalysis Of Variances) analysis.

## 6. Conclusions

This paper presented the data driven limitations of the experimental conflict event model, developed at the Joint Research Centre of the European Commission.

The proposed model 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 datasets 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 DL methodology used to model the conflict events is 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 DL model, we have set up an early warning alarm system to signal abnormal social unrest upheavals.

Two potential datasets and their limitations, that follow the CAMEO political event coding classification, were discussed in this paper: (i) the Global Data on Events Location and Tone (GDELT) project and (ii) the Integrated Crisis Early Warning System (ICEWS) Dataverse dataset.

Even though the DL 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 or ICEWS as an input to the experimental conflict event model requires overcoming certain obstacles. Firstly, the automated codebook algorithm is not publicly available for GDELT, which does not allow investigation on the source of potential errors in the news classification. Secondly, the ICEWS data sources are not publicly available, so validation is not facilitated.

Common issues need to be resolved in both datasets: false positive rates, duplication rates, geographical or socioeconomic biases, “media fatigue”, particularly in conflict zones.

The Europe Media Monitor (EMM) event dataset 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. The Political Language Ontology for Verifiable Event Records (PLOVER) dictionary could replace the existing CAMEO codebook and provide new categories such as elections. However, it is not available yet. A new automated codebook algorithm could be a potential solution to overcome both obstacles created to the GDELT dataset and the present classifiers.

To conclude, the experimental conflict event modelling methodology applied on the GDELT dataset presently gives policy makers the possibility to observe on escalating or de-escalating situations in a country on a monthly basis. However, event-based systems will require supplementary research to offset the databases’ shortcomings, such as automated data validation, new classifiers and dictionaries reflecting the changing nature of conflict and most importantly evidence on the pathways between social unrest and violent conflict.

## 7. Bibliographical References

- Alikhani, Ehsan (2014) *Computational Social Analysis: Social Unrest Prediction Using Textual Analysis of News* (<https://www.bps.go.id/dynamictable/2018/05/18/1337/persentase-panjang-jalan-tol-yang-beroperasi-menurut-operatornya-2014.html>).
- 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>).
- Chai, Tianfeng, T Chai & RR Draxler (2014) Root mean square error (RMSE) or mean absolute error (MAE)?-Arguments against avoiding RMSE in the literature. *Geosci. Model Dev* 7: 1247–1250 ([www.geosci-model-dev.net/7/1247/2014/](http://www.geosci-model-dev.net/7/1247/2014/)).
- Chung, Junyoung, Caglar Gulcehre, KyungHyun Cho & Yoshua Bengio (2014) Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. 1–9 (<http://arxiv.org/abs/1412.3555>).
- GDELT (2019) The GDELT Project (<https://www.gdelproject.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 & Mayeul Kauffmann (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, Michail Papazoglou, Marie-sophie VaN Damme, Gabriel Jenkinson, Kathrin-Manuela Baumann & Dimitrios Thomakos (2019) *Dynamic Global Conflict Risk Index*.
- Halkia, Matina, Stefano Ferri, Marie K Schellens, Michail Papazoglou & Dimitrios Thomakos (2020) The Global Conflict Risk Index: A quantitative tool for policy support on conflict prevention. *Progress in Disaster Science* (March): 100069 (<https://linkinghub.elsevier.com/retrieve/pii/S2590061720300065>).
- Hegre, Håvard, Joakim Karlsen, Håvard Mogleiv Nygård, Håvard Strand & Henrik Urdal (2013) Predicting Armed Conflict , 2010 – 2050. *International Studies Quarterly* 57(2): 250–270.
- Hochreiter, Sepp & Jurgen Schmidhuber (1997) Long short-term memory. *Neural Computation* 9(8): 1735–1780.
- ICEWS | Lockheed Martin (2019) (<https://www.lockheedmartin.com/en-us/capabilities/research-labs/advanced-technology-labs/icews.html#MoreInfo>).
- Levin, Noam, Saleem Ali & David Crandall (2018) Utilizing remote sensing and big data to quantify conflict intensity: The Arab Spring as a case study. *Applied Geography* 94(May): 1–17.
- 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](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(MI): 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](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.
- Wang, Wei, Ryan Kennedy, David Lazer & Naren Ramakrishnan (2016) Growing pains for global monitoring of societal events. Automated event coding raises promise and concerns. *Science* 353(6307): 1502–1504.
- 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](http://jayyonamine.com/wp-content/uploads/2013/03/Forecasting_Afghanistan.pdf)).