Data & Policy (2024), 6: e17
doi:10.1017/dap.2024.10

DATA PAPER

Introducing a global dataset on conflict forecasts and news topics

Hannes Mueller[1,2], Christopher Rauh[3,4] and Ben Seimon[1,5]

[1]Institut d'Anàlisi Econòmica, CSIC, Barcelona, Spain
[2]Barcelona School of Economics, Barcelona, Spain
[3]Faculty of Economics, University of Cambridge, Cambridge, UK
[4]PRIO, Oslo, Norway
[5]Fundació d'Economía Analítica, Barcelona, Spain
Corresponding author: Christopher Rauh; Email: cr542@cam.ac.uk

Received: 02 October 2023; **Revised:** 02 February 2024; **Accepted:** 08 February 2024

Keywords: conflict; civil war; forecasting; machine learning; news topics; random forest; topic models

Abstract

This article provides a structured description of openly available news topics and forecasts for armed conflict at the national and grid cell level starting January 2010. The news topics, as well as the forecasts, are updated monthly at conflictforecast.org and provide coverage for more than 170 countries and about 65,000 grid cells of size 55 × 55 km worldwide. The forecasts rely on natural language processing (NLP) and machine learning techniques to leverage a large corpus of newspaper text for predicting sudden onsets of violence in peaceful countries. Our goals are a) to support conflict prevention efforts by making our risk forecasts available to practitioners and research teams worldwide, b) to facilitate additional research that can utilize risk forecasts for causal identification, and c) to provide an overview of the news landscape.

Policy Significance Statement

This article carries profound policy significance by harnessing the power of cutting-edge technology and openly available data sources to fortify global peace endeavors. It offers a systematic account of accessible news topics and predictive forecasts for armed conflict, spanning national and grid cell levels, and are regularly updated at conflictforecast.org for over 170 countries and about 65,000 grid cells globally starting in January 2010. The adoption of natural language processing (NLP) and advanced machine learning techniques empowers these forecasts to anticipate the emergence of violence, even in historically peaceful regions. In terms of NLP, the method relies on the Latent Dirichlet allocation (LDA), which is an unsupervised machine learning algorithm that clusters text. The topic model is used to summarize more than 5 million newspaper articles. For the grid-cell level predictions, we detect specific locations mentioned in our news corpus to compute the local distribution of the topics. The methodology holds the potential to revolutionize conflict prevention strategies. In terms of policy impact, the article serves a tripartite mission: first, by furnishing risk forecasts to practitioners and research teams worldwide, it bolsters conflict prevention efforts; second, it opens doors for further research endeavors, enabling causal identification and case studies through these predictive insights; and third, it delivers a monthly overview of the dynamic news landscape for each country in the entire sample period. This multidimensional approach demonstrates how novel data sources and technology can be pivotal in advancing global peace and security initiatives, making the world a safer place.

1. Background

In this article, we explain the methodology and illustrate the performance of risk forecasts made available dating back to 2010 and updated monthly in real time on conflictforecast.org.[1] The risk forecasts span different forecasting horizons, outcome measures of violence, and geographical units. One of the key predictors included in the prediction models is news topics derived from millions of news articles with global coverage. These news topics form part of the data made available to the public, and we put a particular emphasis on discussing their role in the forecast.

Early warning signals are important because civil wars are a humanitarian and economic disaster. Data from Hegre et al. (2020) shows that over 200,000 people were killed as a direct result of the conflict in 2022, while as of June 2023, over 1 billion people are currently living in countries of active conflict. Besides the deceased, the arguably most affected are the forcibly displaced, for which UNHCR report that there were over 108 million forced displacements in 2022, driven by conflicts, for example, in Ukraine, the Democratic Republic of Congo, and Ethiopia.

Traditionally, the economics literature has focused on understanding the causal mechanisms that drive conflict (Hegre et al., 2017). Ward et al. (2010) demonstrate that variables identified as causal often carry little predictive power in forecasting systems. Mueller and Rauh (2022b) show that commodity price variations and export weights provide little benefit for forecasting conflict outbreaks despite being central to work on the causal drivers of conflict. In light of this, a rich literature has been developed that uses quantitative methods to forecast armed violence. Goldstone et al. (2010) show that countries can be categorized usefully by using standard cross-sectional variables. Bazzi et al. (2022) and Hegre et al. (2022a) produce forecast systems for the subnational level. Chadefaux (2014) and Mueller and Rauh (2018, 2022b, 2022c) show that, by using news text, these systems can be made almost real-time and predict outbreaks in previously peaceful countries. One of the best-known projects in this area is the VIEWS system (Hegre et al., 2019; 2021). The group has recently started to organize friendly forecasting competitions to improve collective scientific knowledge on forecasting (de-)escalation in Africa (Hegre et al., 2022b; Vesco et al., 2022).

A growing focus of the literature has been the formulation of policy recommendations that can support the prevention of conflict (Rohner and Thoenig, 2021). For prevention, the most valuable forecasts are those which identify heightened risks in cases where extended periods of peace have been experienced. Kleinberg et al. (2015) call this a *prediction policy* task. It is in these situations that truly preventative policies can be enacted. In this context, the *hard problem* of prediction becomes pertinent. This can be summarised as follows:

- Conflict prediction is an imbalanced class problem—there are significantly more instances of peace (0's) than onsets of conflict (1's).
- The existence of the conflict trap (where countries are stuck in repeated cycles of violence) leads to conflict history becoming an extremely powerful predictor of risk.
- Hence the *hard problem* of conflict prediction is to predict outbreaks of violence in countries without a recent history of violence. It is in these cases where text data is found to add significant predictive power.

Our aim is to support policymakers facing different contexts. As a result, we provide forecasts for three different target variables at the national level for 3 months and 12 months in advance. These are updated monthly:

1. Any violence: The target variable is a binary indicator of at least one battle-related death. Hence, this is a classification problem.

[1] Given that the website and the monthly updates were not launched until July 2021, only the predictions published as of then are "true" out-of-sample predictions, as in the future is truly unknown. All of the downloadable predictions referring to previous periods are generated by a rolling forecast which is out-of-sample, that is, no future information is used. In other words, for past predictions we pretend like the future is unknown at each time step.

2. Armed conflict: The target variable is a binary indicator that meets our definition of armed conflict. We define armed conflict as a per capita measure – 0.5 deaths per 1 million inhabitants. Hence, this is a classification problem.
3. Violence intensity: The target variable is the number of fatalities per capita. More specifically, when training the model, we predict the log of the average battle deaths per capita in a 3- and 12-month window plus one.[2] The goal of the *violence intensity* model is to capture escalations when a country is already in conflict. Hence, this is a regression problem.

For countries with extended periods of peace, the *any violence* and *armed conflict* forecasts provide an indication of the possible risk of an outbreak. However, this is less valuable for countries currently in violence or stuck in the conflict trap. In these situations, the *violence intensity* forecast can provide an indication of future escalations/de-escalations. At present, we only provide forecasts of *any violence* at the grid cell level (~ 55 km × 55 km or 0.5 × 0.5 decimal degrees) for 12 months ahead. The longitude and latitude of the grid cells are given by the PRIO-GRID (Tollefsen et al., 2012).

Risk forecasts can inform decision makers when allocating resources or prioritizing tasks. However, researchers can also use risk forecasts for the sake of case studies and matching. Mueller and Rauh (2022a) utilize risk forecasts to causally identify the effect of policies. Using difference-in-difference methods, they find that, on average, power-sharing agreements lead to an 8% decrease in the occurrence of violence and an 18% drop in the intensity of armed violence. This highlights that our forecasts can be used to support forward-looking policy decisions, but also retrospectively analyze which policies have been effective in preventing/de-escalating violence.

2. Forecast methodology

Overall, we publish six different sets of forecasts at the national level and one set of forecasts at the grid cell level. The methodology has been proven to be successful at predicting the likelihood of violence (Mueller and Rauh, 2018, 2022b) and the intensity of violence (Mueller and Rauh, 2022c). It relies on detecting patterns of violence in relation to past violence and changes in news reporting. A high-level overview of the methodology is provided in Figure 1. We combine historic information on violence and news topics as predictors into a random forest. The simplified illustration shows how the random forest may consider countries with no recent violence as low risk and those with recent violence and a lot of conflict news as high risk. The trained model is then used to predict the likelihood and intensity of violence in the unknown future. The pipeline is explained in more detail in what follows.

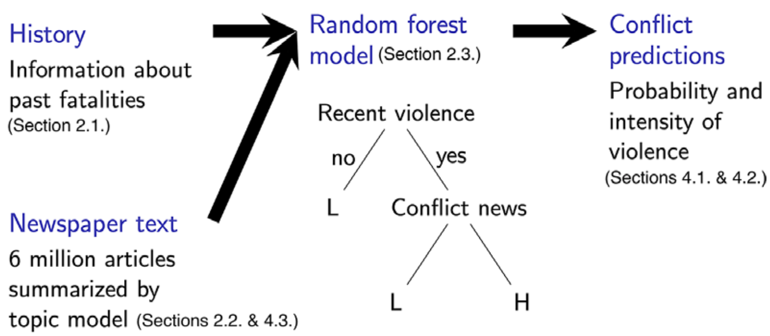

Figure 1. *Schematic illustration of prediction pipeline.*

[2] We transform the number of battle deaths using the log due to the skewed nature of the number of fatalities and add one due to the many zeros.

https://doi.org/10.1017/dap.2024.10 Published online by Cambridge University Press

2.1. Fatalities data

We rely on the UCDP Candidate Events Dataset that makes available monthly releases of violence data with not more than a month's lag globally (Sundberg and Melander, 2013; Hegre et al., 2020; Davies et al., 2023). We are interested in deaths at the country/month level or grid cell/month level resulting from armed force used by an organized actor against another organized actor or against civilians. We include state-based conflict, non-state conflict, and one-sided violence. Since the dataset does not code zeros, we allocate a zero to any unit in which GED data is available, and the country is independent. Importantly, this means we are predicting political violence and escalations into internal armed conflict. We do not predict external wars like invasions of one country by another.

With respect to the grid cell forecasts, uncertainties arise with respect to the exact location of events. These are directly coded in the UCDP data as "where_prec." The lowest precision for "where_prec" is "only the country where the event took place is known." But at the grid cell level, we only include events that have been coded with geographical precision up to the ADM2 level (i.e., an individual grid cell). We also retain events that have been coded as taking place in international waters and air space.[3]

2.2. Text data

Our text data are comprised of over 6 million documents from 1989 to present. These are downloaded from Factiva and are sourced from two newspapers (347,874 articles from the *New York Times* and 142,813 from the *Economist*) and three news aggregators (968,898 articles from the Associated Press, 3,588,489 from the BBC Monitor, and 39,232 from LatinNews).[4] Text is downloaded according to rules set in an extensive query. As a generalization, a document is downloaded if a country or capital name appears in the title or lead paragraph. One limitation relates to the inherent bias of news data, particularly in political regimes where the media is censored or restricted. However, Mueller and Rauh (2022b) show that this bias results in no obvious failure of the model when predicting hard onsets. The inclusion of LatinNews as a source is specifically intended to improve the text signal for Latin America since BBC Monitor generally focuses on Asia and Africa.

Standard natural language preprocessing (NLP) techniques are used, including the removal of punctuation, stop words, and lemmatization. In addition to single words (unigrams), we also consider common combinations of two or three words (bigrams and trigrams). Any token (unigram, bigram, or trigram) that appears in at least half of the documents (too frequent) or in fewer than 200 documents (too infrequent) is also removed.

This results in a corpus whereby documents are assigned to the country/month level. Hence, we have a set of documents that represent the news landscape for every country for every month between January 1989 and present. For the grid-cell level predictions, we additionally use prepositions to detect locations mentioned in news articles. The challenge is to condense this text data into a set of features that improve the forecast.

LDA, developed by Blei et al. (2003), is a probabilistic model used for topic modeling and document clustering in natural language processing. LDA assumes that documents are mixtures of topics, and topics are mixtures of words. It aims to uncover the latent topics within a corpus by iteratively assigning words in documents to topics and estimating the distribution of topics in documents and words in topics. This modeling technique allows LDA to extract meaningful topics from a collection of text data, making it a valuable tool for tasks such as document categorization, content recommendation, and understanding the thematic structure of large text datasets.

For the implementation, we rely on the Python package from Řehřek and Sojka (2010) of the dynamic Latent Dirichlet allocation (LDA) topic model (Hoffman et al., 2010) estimated with

[3] Specifically, we only include observations coded as a 1, 2, 3 or 7 according to the UCDP definition. More details can be found in the UCDP Candidates Event codebook at https://ucdp.uu.se/downloads/.

[4] The number of articles refers to the articles for which we estimate topics in the latest vintage of the topic model including text up to and including August 2023. Some articles are considered duplicates and others contain too little information for the topic model.

https://doi.org/10.1017/dap.2024.10 Published online by Cambridge University Press

15 topics. This enables the reduction of the dimensionality of the text data without using priors on which elements of the text will be most useful for forecasting conflict. The dynamic LDA can be summarized as follows:

1. Topics: Topics are distributions over words. Each topic assigns different probabilities to different words in the full set of tokens. For example, a topic on "economics" might assign high probabilities to words such as "economy," "inflation," "investment," and so forth We estimate topics by first training the model on all text data up until 2010m1. We allow the a priori weight variational hyperparameters for each document to be inferred by the algorithm, and α, the a priori belief for each topics' probability is set to the default of $(1/N)$, where N is the number of topics. The estimated topics and top keywords are discussed in more detail in Section 4.3 and Table 6.
2. Document-topic distribution: This is a matrix of size $D \times N$, where D is the number of documents and N is the number of topics. Each row sums to 1, that is, a document is represented by the relative proportion of various topics in that document.
3. Country-topic shares: Hence, for each country at each time step, we can compute the proportion of the news landscape that is assigned to a given topic by averaging over the document-topic distribution assigned to that country/month. This process is dynamic because we reinterpret the topic distribution of previous months as new documents become available each month.

Note that the purpose of the text analysis here is to generate a broad, all-encompassing dimensionality reduction of the entire news text. This is because the project does not impose what kind of news content or events will be indicative of conflict risk. This is a significant difference from the approach taken in the Political Instability Task Force (PITF), for example, which relies on event coders to produce large databases like the ICEWS database from Boschee et al. (2015) or, more recently, the Polecat global event dataset (Halterman et al., 2023). We expect the NLP approach here to work better if the factors that predict conflict, especially those that are negatively associated with conflict risk, are hard to foresee.

2.3. Prediction method

For the conflict prediction task, we rely on a Random Forest (Breiman, 2001) and implement a rolling forecast methodology. Random Forest is a robust and widely used ensemble machine learning algorithm that excels in predictive modeling tasks. It operates by constructing multiple decision trees during training, each based on a random subset of the data and features. These individual trees are then combined to make predictions. Random Forest offers several advantages, such as handling high-dimensional data and nonlinear relationships, reducing overfitting, and providing feature importance rankings. This algorithm's strength lies in its ability to capture complex interactions within the data, making it a valuable tool for both regression and classification problems. In a rolling forecast, the forecast horizon typically remains fixed (in our case, 3 and 12 months into the future), but the forecast is updated at monthly intervals. As each period passes, we add the most recent actual data and update the forecast for the next prediction horizon.

Our datasets underlying the predictions are set up at the geographic unit/month level. At the national level, we generate forecasts for three different target variables across two different time horizons. We then distinguish between models that rely only on text features (text model) or a combination of historical violence and text features (best model). The text model only relies on the topic shares, and therefore "knows" much less about the situation a country is in. As a consequence, it reacts much stronger to news as this is the only source of its information. On the contrary, the best model is well informed about the history of violence in a country, but therefore, it will put much less weight, and hence react much less, to changes in news topics. This trade-off will be discussed further in Section 3, in which we evaluate the predictive performance of the models.

https://doi.org/10.1017/dap.2024.10 Published online by Cambridge University Press

Table 1. *National forecast target variables*

Target variable group	Target variable	Definition	Forecast output
Any violence	ons_anyviolence3	1 if at least 1 battle-related death in 3 months; N/A if ongoing violence; else 0	Probability of at least 1 battle related death in any of the next 3 months
	ons_anyviolence12	1 if at least 1 battle-related death in 12 months; N/A if ongoing violence; else 0	Probability of at least 1 battle related death in any of the next 12 months
Armed conflict	ons_armedconf3	1 if at least 0.5 fatalities per 1mn inhabitants in 3 months; N/A if ongoing armed conflict; else 0	Probability of at least 0.5 fatalities per 1mn inhabitants in any of the next 3 months
	ons_armedconf12	1 if at least 0.5 fatalities per 1mn inhabitants in 12 months; N/A if ongoing armed conflict; else 0	Probability of at least 0.5 fatalities per 1mn inhabitants in any of the next 12 months
Violence intensity	lnbest_pc3	Log transformation of average fatalities per capita in the next 3 months[a]	Average number of fatalities over the next 3 months[b]
	lnbest_pc12	Log transformation of average fatalities per capita in the next 12 months[a]	Average number of fatalities over the next 12 months[b]

[a]We first compute $z = \sum_{i=1}^{w} \frac{x_i \times 1000}{y_i}$, where w is the forecast horizon, while x_i and y_i represent the number of fatalities and population in month i, respectively. Hence, z represents the sum of fatalities per 1000 inhabitants over the next w months. We then have that $a = \frac{z}{w}$, that is, the average fatalities per 1000 inhabitants over the next w months. The log transformation is then conducted as $\ln(a+1)$.
[b]Strictly speaking, the model outputs the log-transformed average number of fatalities per capita. To convert back to number of fatalities, the prediction is transformed as best $= (e^x - 1) \times \frac{y}{1000}$, where x is the predicted log transformed average number of fatalities per capita and y is population.

Table 1 defines the target variables and interpretation of the forecasting outputs. We define armed conflict as 0.5 deaths per 1 million inhabitants in any given month. Tables 2 and 3 describe the features used in the respective models. Note that the text features across all models and target variables are the same. However, the historical conflict features are the same for *any violence* and *armed conflict* predictions, but differ in the *violence intensity* case. Finally, population data is sourced from the World Bank. Since our unit of analysis is the country/month level, but the data is available annually, we assume that the population of a country is the same for any month in a given year.[5] In the case of missing data, we forward fill using the latest available data.

The purpose of the rolling forecast is to replicate the information set that would be available to a decision maker. In other words, they would observe features until period T and make a forecast for the aggregate window $T < t \leq T + W$, where W is equal to 3 or 12. In other words, when the window is 3 months, then we are predicting, for instance, the likelihood of any battle death over the entirety of the 3 months. We do not predict outcomes for each of the 3 months separately. Similarly, when predicting the 12 month window, we consider the aggregate outcome over the next 12 months, and not for each of the next 12 months separately. For example, an *armed conflict* forecast in 2015 m1 for a window of 12 months is predicting the likelihood of an armed conflict outbreak in any of the following 12 months. The same applies to the *violence intensity* forecast—for a forecasting window of 12 months, we are

[5] Given that our armed conflict definition is predicated on the population value, we are currently working on an interpolation/extrapolation method to avoid “jumps” as new data becomes available. To date, this has not yet been implemented.

https://doi.org/10.1017/dap.2024.10 Published online by Cambridge University Press

Table 2. *National violence and armed conflict forecast features*

Variable category	Variable	Definition	Included: Text model	Included: History model	Included: Best model
Text features	tokens_stock[a]	Weighted stock of total token count (unigrams, bigrams, and trigrams)	Y		Y
	weighted_topic_stock_k[b]	Weighted stock of topic share, where k can range from 0 to $N-1$ and $N=15$ – the total number of topics in the Dynamic LDA model	Y		Y
Historical conflict features	anyviolence_dp	Number of months since last month with at least one battle-related death		Y	Y
	armedconf_dp	Number of months since the last armed conflict episode		Y	Y
	civilwar_dp	Number of months since the last Civil War episode		Y	Y
	past_x	Total number of fatalities per capita in the past x months, where $x=6$, 12, 60, and 120		Y	Y
Country-specific features	Population	Country population in a given year	Y	Y	Y

[a]The number of tokens represents the flow of unigrams, bigrams, and trigrams mentioned in the documents for each country-month. To smooth out changes over time, instead of using the flow, we use a token stock (W_t), which consists of the present value of the flow of tokens. Let us define w_t as the number of tokens (unigrams, bigrams, trigrams) in all documents of a specific country at month t. For a decay rate of $\delta=0.8$, the token stock for a specific month T is

$$W_{t=T}=\sum\nolimits_{t=1}^{T}\delta^{T-t}w_t.$$

[b]Similarly, the share of topics in the news of each month can be seen as a flow, which we transform into a stock ($X_{k,t}$) to reduce its variability. In this case, to account for the fact that months with a higher volume of news should carry more weight when updating the stock, we weight the flow by the number of tokens in each month. For a specific country and month t, let us define $x_{k,t}$ as the share of topic k, w_t as the token count, and W_t as the stock of total token count. For a decay rate of $\delta=0.8$, the stock of the share of topic k for a specific month T is:

$$X_{k,t=T}=\frac{\sum_{t=1}^{T}\delta^{T-t}w_t x_{k,t}}{W_T}$$

predicting the average number of fatalities per capita per month over the next 12 months. We train a model to learn a functional form using all data from 1989m1 to 2009m12 as follows:

$$y_{i,T<t\leq T+W}=F_T(\mathbf{X}_{i,T}),$$

With the resulting model, we then produce out-of-sample predictions on a rolling basis from 2010m1 onwards:

$$\hat{y}_{i,T<t\leq T+W}=F_T(\mathbf{X}_{i,T}),$$

For *any violence* and *armed conflict*, hyperparameters are chosen by maximizing the area under the curve (AUC) of the receiver operating characteristics (ROCs) curve via pseudo-out-of-sample rolling forecasting on the sample 2010 to 2015. In the case of *violence intensity*, we seek to minimize the mean squared error (MSE). To solidify the rolling forecast methodology, we provide an example of the

https://doi.org/10.1017/dap.2024.10 Published online by Cambridge University Press

algorithm employed where the full data sample ranges from January 1989 to August 2023. This can obviously be modified by updating Aug 2023 to the latest available date for which data is available:

Algorithm 1. Rolling Forecast: Pseudo Out of Sample Forecasting

Require: Full data sample $D = \{d_{1989m1}, d_{1989m2}, \ldots, d_{2023m8}\}$
Require: Window size W
Require: Forecasting model F
Ensure: Forecasts $\hat{Y} = \{\hat{y}_{2010m1 < t \leq 2010m1 + W}, \hat{y}_{2010m2 < t \leq 2010m2 + W}, \ldots, \hat{y}_{2023m8 < t \leq 2023m8 + W}\}$.
Train model F on data $\{d_{1989m1}, d_{1989m2}, \ldots, d_{2009m12}\}$
Optimise and fix hyperparameters of F using cross-validation on data $\{d_{2010m1}, d_{2010m2}, \ldots, d_{2014m12}\}$
for T from 2010m1 to 2023m8 **do**
    $D_{\text{train}} \leftarrow$ Data for all $1989m1 \leq t \leq T - W$
    Retrain model F on D_{train}
    $\hat{y}_t \leftarrow$ Aggregate forecast of F for $T < t \leq T + W$
    Append $\hat{y}_t$ to $\hat{Y}$
end for
    return $\hat{Y}$

Table 3. *National violence intensity forecast features*

Variable category	Variable	Definition	Included: Text model	Included: History model	Included: Best model
Text features	tokens_stock[a]	Stock of total token count (unigrams, bigrams, and trigrams)	Y		Y
	weighted_topic_stock_k[b]	Weighted stock of topic share where k can range from 0 to $N-1$ and $N = 15$ – the total number of topics in the Dynamic LDA model	Y		Y
Historical conflict features	anyviolence_dp	Number of months since last month with at least one battle-related death		Y	Y
	ln_bestpc[c]	Log transformation of fatalities per capita, 1 month lag		Y	Y
	ln_bestpc13[c]	Log transformation of fatalities per capita, 13 month lag		Y	Y
	past_x	Total number of fatalities per capita in the past x months, where $x = 6$, 12, 60 and 120		Y	Y
Country-specific features	Population	Country population in a given year	Y	Y	Y

[a]See note a of Table 2.
[b]See note b of Table 2.
[c]Log transformation is conducted as $\ln(z+1)$, where $z = \frac{x*1000}{y}$, such that x is the number of fatalities and y is population. In other words, z represents fatalities per 1000 inhabitants.

https://doi.org/10.1017/dap.2024.10 Published online by Cambridge University Press

In this way, we are able to generate a full history of predictions via pseudo-out-of-sample forecasts. For example, imagine rewinding back to 2015m1. A policymaker would only have access to data up until 2014m12 to inform a forecast for 2015m1 to 2015m3 (for a time horizon of 3 months ahead). This is exactly the process we seek to simulate via the rolling forecast methodology. This enables a realistic evaluation of what is possible in terms of forecasting power in actual applications, as no data that has been used for training purposes is included in the test set (Mueller and Rauh, 2022b).

3. Performance

In the following section, we outline the performance of our forecast for predictions from January 2010 to August 2023. The latest predictions can be downloaded at conflictforecast.org.

3.1. National

3.1.1. Any violence and armed conflict

To evaluate performance, we compare realizations $y_{i,T<t\leq T+W}$ for all $t \in 2010m1, \ldots, last\_month$ with the predicted values $\hat{y}_{i,T<t\leq T+W}$.[6] For *any violence* and *armed conflict*, we present AUC-ROC and precision–recall curves. The receiver operating characteristic–area under the curve (ROC-AUC) and precision–recall curves serve as essential evaluation metrics for binary classification models. The ROC-AUC curve offers a graphical representation of a classifier's ability to discriminate between positive and negative classes across various threshold values, while its associated AUC quantifies the overall discriminative power of the model. The x-axis of the ROC-AUC curve represents the false positive rate (FPR). The y-axis represents the True Positive Rate (TPR), also known as sensitivity or recall.

Conversely, the precision–recall curve provides insights into the model's trade-off between precision and recall, making it particularly valuable for scenarios involving imbalanced datasets or instances where false positives are of concern. In the precision–recall curve, the x-axis represents recall, which is the same as the TPR on the ROC curve. Precision is the ratio of true positives to the sum of true positives and false positives. In both cases, the curves display how the model's performance changes as the classification threshold is varied.

When evaluating performance, it is important to distinguish the two forecast horizons (3 and 12 months) as well as the difference between the *armed conflict* model that features less onsets (higher imbalance) and *any violence* that features more onsets. We also distinguish performance on hard onsets (after more than 60 months of peace) and all onsets. An important element of evaluations is that they can only be run on outcomes that are realized. When evaluating an onset model forecasting up to 12 months ahead in August 2023, one can only evaluate the model until August 2022 as the onsets in the 12 months following September 2022 are not realized yet in August 2023. The 95% confidence intervals of the ROC-AUCs run on the data from 2010m1 to 2022m8 are presented in Figure 2. The confidence intervals are generated via bootstrapping.[7]

One striking feature of the forecast model is that performance holds up extremely well over time—despite the fact that the period 2010 to 2023 is characterized by dramatic shifts in geopolitical dynamics, types of conflicts, and the news landscape. Figure 3 shows performance of the armed conflict model over time by showing AUCs by year.[8] As before, the best model performs better than the text model. However, neither model shows dramatic fluctuations despite the dramatic aforementioned changes. While not significant, there is even a tendency of improvement visible in the best model.

[6] If we have conflict data until 2023m8 then the *last_month* for $W = 3$ would be 2023m5 and for $W = 12$ would be 2022m8.

[7] We generate 1000 boostrapped samples for January 2010 to August 2023 by drawing instances, with replacement, from the observed onsets of violence. 1000 ROC-AUC scores are then computed using the relevant predictions and we plot the mean, 2.5th and 97.5th percentile of the resulting distribution of scores.

[8] Note that these statistics are not comparable to the AUC evaluated on the entire sample.

https://doi.org/10.1017/dap.2024.10 Published online by Cambridge University Press

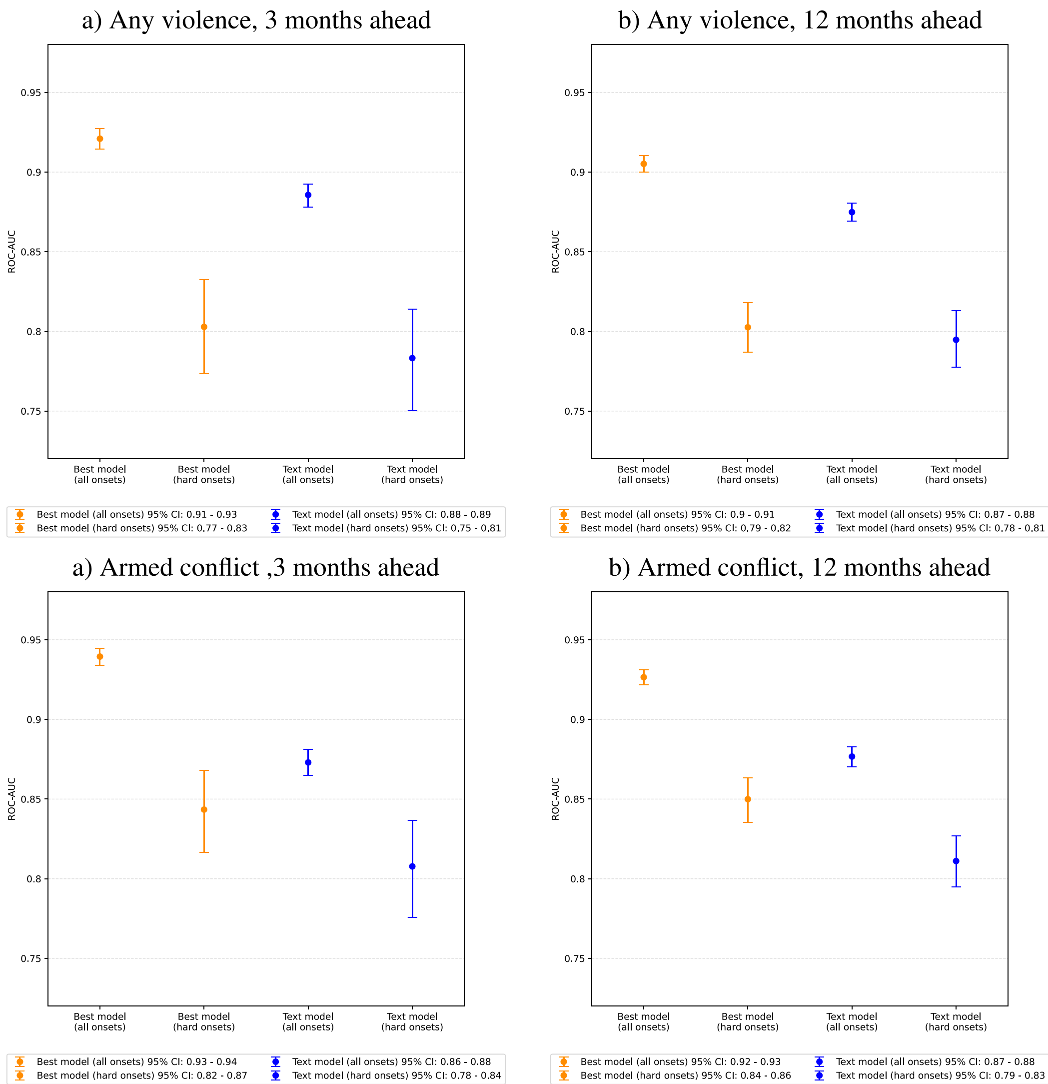

Figure 2. *ROC-AUC 95% confidence intervals.*

Figure 3 shows the performance of the *armed conflict* model, focusing on hard onsets of conflict after 60 months of peace. The AUCs shown are for 2-year periods. The best model now sometimes performs worse than the text model. However, both models show little performance trends over time and do not show a dramatic variation over time in general. This is despite the fact that hard onsets are much rarer and much more specific to the international landscape. The period we look at does, for example, include the Arab Spring period, which was characterized by instability in several countries that were previously stable for long periods.

One important takeaway from these figures is that the COVID-19 period with its dramatic change in reporting patterns did not consistently damage the performance of the models. The trend in the best model is overall positive over time, and there is no clear pattern in the text model. This suggests that over time, performance should stay constant or even improve.

However, as discussed above, the AUC still does not tell the full truth about performance as imbalance does not have an effect on this performance measure. Policymakers should care at least as much about precision. This is also where the issue of increasing imbalance in hard onsets appears most clearly.

https://doi.org/10.1017/dap.2024.10 Published online by Cambridge University Press

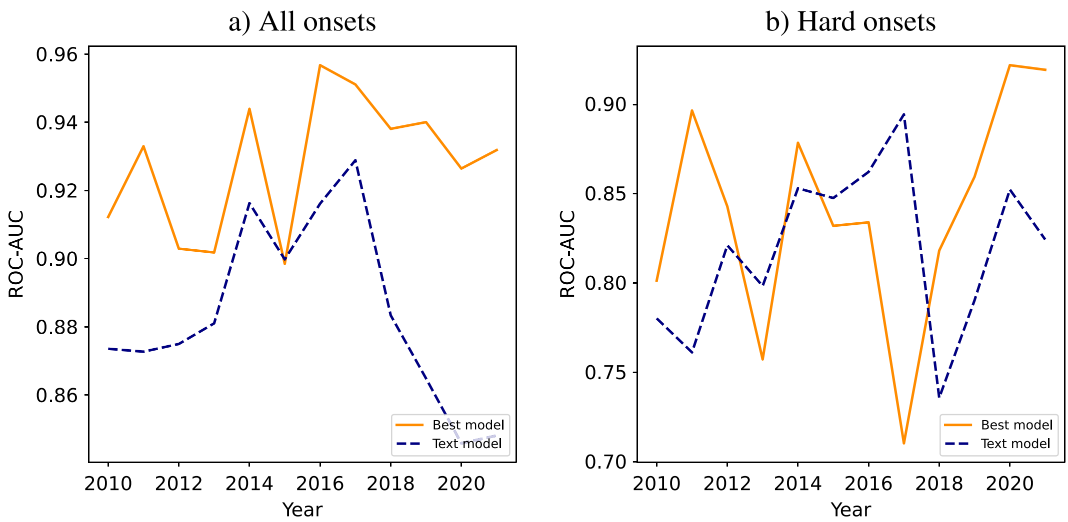

Figure 3. *ROC-AUCs by year for armed conflict, 12 months ahead.*

In Figure 4, we show the precision–recall curves for the *armed conflict* 12 months ahead forecast. Figure 4a shows the pseudo-out-of-sample performance for all onsets. Precision in the best model remains above 80% for a recall rate of over 50%. The text model performs much worse as violence history is a very important driver of risk. However, it is important to keep in mind that armed conflict onsets are rare events. Even for the 12 months ahead forecast, the precision eventually falls far below 20% for all onsets. It is a significant feat to bring overall precision to over 80% for low recall rates.

Figure 4b shows the pseudo-out-of-sample performance for the same model but only for onsets that happened after 60 months of peace. Precision in the best model is now significantly lower and is around 20% for a recall rate of 20%. However, the text model now performs relatively much better. For very low recall rates, the text model can even beat the best model. The imbalance problem is now extreme with a baseline likelihood of only a few percentage points when forecasting *armed conflict* 12 months ahead.

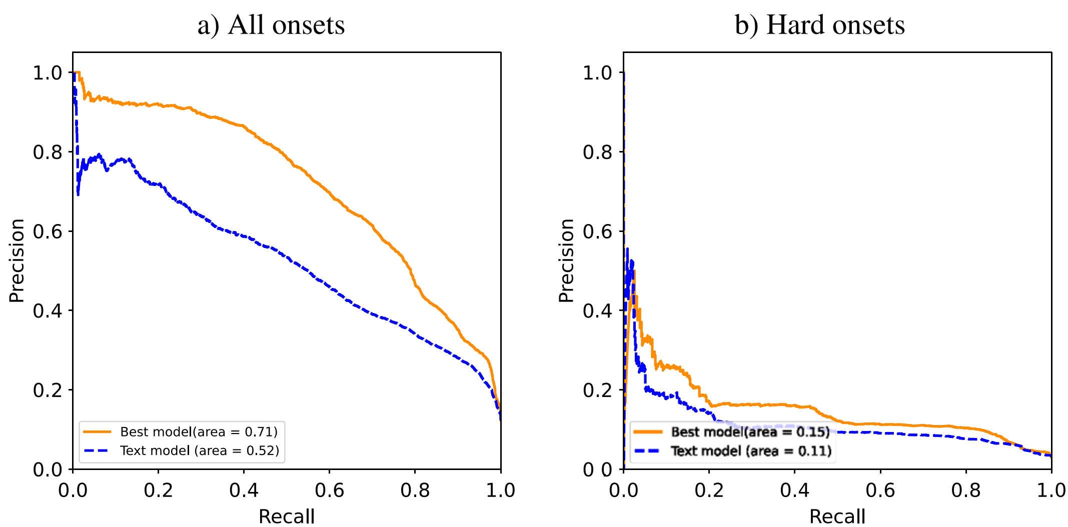

Figure 4. *Precision recall curve for armed conflict forecast, 12 months ahead.*

https://doi.org/10.1017/dap.2024.10 Published online by Cambridge University Press

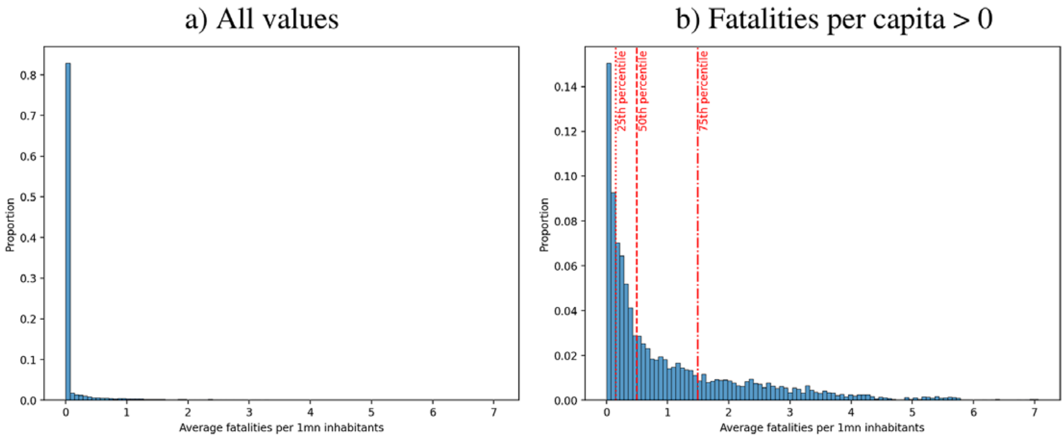

Figure 5. Distribution of average fatalities *per capita.*

For onsets in countries with even longer histories of peace, the relative performance changes further. The best model relies heavily on information about past violence, since due to the conflict trap, past violence is an excellent predictor of future violence. Therefore, this type of model will not detect risk emerging in previously peaceful countries. In contrast, the text model reacts strongly to news and might thereby generate spikes in risk even if a country has experienced a sustained period of peace. On average, it is clear that the text model will perform worse as it is missing important information. Nonetheless, it can provide valuable complementary information to raise red flags in unsuspected settings. In summary, if average performance is a user's main criterion, then the best model is the indicator of choice. However, if the user wants to potentially spot black swan events, then he/she should consider the text model as well.

3.1.2. Violence intensity

For *violence intensity*, we rely on the mean squared error (MSE) as our metric for evaluation and focus on the performance of the 3 months ahead forecast. The MSE is computed as $(\ln(y+1)-\ln(\hat{y}+1))^2$, where y is the observed average fatalities per capita over the next 3 months and $\hat{y}$ is the predicted average fatalities per capita over the next 3 months.[9] The same principles for evaluation apply—for a 3 months ahead *violence intensity* forecast in August 2023, one can only evaluate the model until May 2023 as the violence in the 3 months from June 2023 onwards are not realized yet in August 2023.

First we must define a benchmark model. In this case, we define a "no-change" model as our benchmark—average fatalities per capita for the previous 3 months are used as the prediction for the average fatalities per capita for the next 3 months. Across the full sample, our best model outperforms the no-change model. Next, we undertake a closer inspection of performance by grouping fatalities per capita into bins. Figure 5 shows that the distribution is heavily right skewed—in other words, most observations in our sample are country/months without any battle-related deaths. Hence, we define our bins according to the distribution of average fatalities per capita, where they exceed 0. In total, we have four bins, which align with the 25th, 50th, and 75th percentile as per the distribution shown in Figure 5b.

Table 4 shows the average MSE for the best model and no-change model for the observations where the true average fatalities per capita in the next 3 months fall into the respective bins. The normalized column represents the best model divided by the no-change model—this gives an indication of how well our model performs relative to the benchmark.

We see that for very low-intensity violence, our model performs worse than the no-change model. This is a product of our intentions to capture escalations, which leads to "over-predicting" in situations where 0 fatalities occur in the future. We also see that, on average, the error is almost approximately half of the

[9] See note a of Table 1 for detail on the computation of average fatalities per capita.

https://doi.org/10.1017/dap.2024.10 Published online by Cambridge University Press

Table 4. *Average MSE by bin*

Bin (deaths per 1 mn inhabitants)	Best model	No change model	Normalized
0–0.15	0.00000101	0.00000026	3.86
0.15–0.5	0.00001006	0.00001821	0.55
0.5–1.5	0.00002535	0.00005227	0.49
1.5+	0.00125427	0.00140813	0.89

no-change model for medium-intensity violence (bins 0.15–0.5 and 0.5–1.5), while we observe marginally better performance in cases of extreme violence.

So far, our intensity forecasting model has been less geared towards predicting the exact number of fatalities, but rather is meant to capture conflict dynamics. These dynamics, by definition, can never be predicted by a no-change model. We highlight the performance of our forecast in this context by analyzing its ability to capture escalations/de-escalations. We first define escalations/de-escalations in accordance with the bins outlined above. For example, if a country has experienced average fatalities per 1mn inhabitants in the range 0.5 to 1.5 for the previous 3 months, then we code an escalation if violence exceeds 1.5 fatalities per 1 mn inhabitants on average in the next 3 months. The reverse logic applies to de-escalations. We then observe whether the forecast prediction tracks the realized escalation/de-escalation by comparing the bin of the prediction for the next 3 months with the realizations of the past 3 months.

Across our sample there are a total of 1345 escalations (5.6%), 1201 de-escalations (5.0%) and 21,344 instances (89.3%) of no escalation. Figure 6 presents confusion matrices for the 3 ahead *violence intensity* best and text models. In the no-escalation cases, the best model significantly outperforms the text model, correctly classifying 87% of all cases compared to 52%. Interestingly, the text model better captures escalations, it classifies 82% of cases correctly compared to 71% for the best model, and de-escalations whereby it correctly classifies 19% compared to 7% for the best model. This suggests that our text features are able to identify signals of future changes in violence intensity levels that are not characterized by recent histories of violence. These results also highlight the value of the intensity model for policymakers seeking a data-driven perspective on the future dynamics of conflict in already violent countries. Since the *any violence* and *armed conflict* forecasts are likely to already be elevated for these situations, we

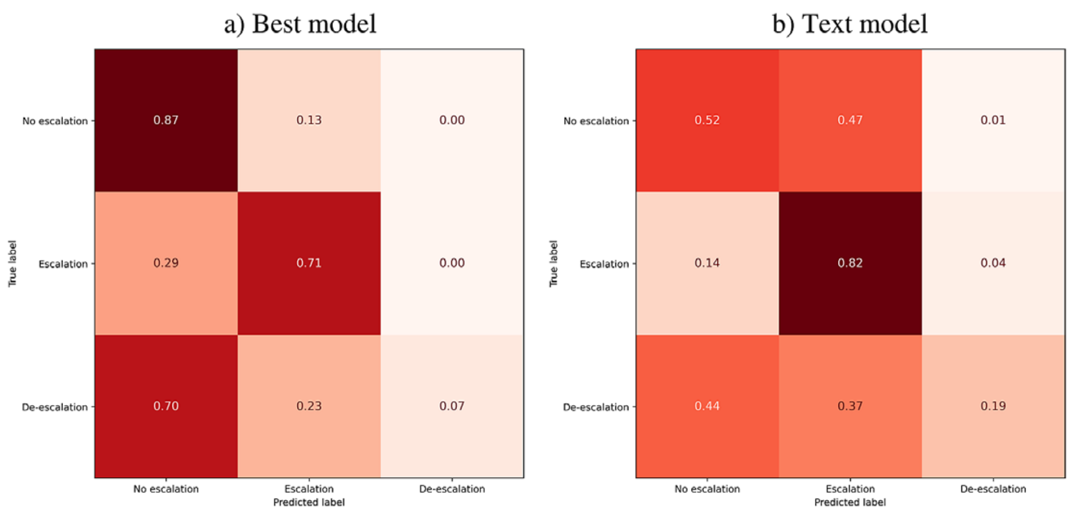

Figure 6. *Violence intensity performance, escalations, and de-escalations 3 months ahead.*

https://doi.org/10.1017/dap.2024.10 Published online by Cambridge University Press

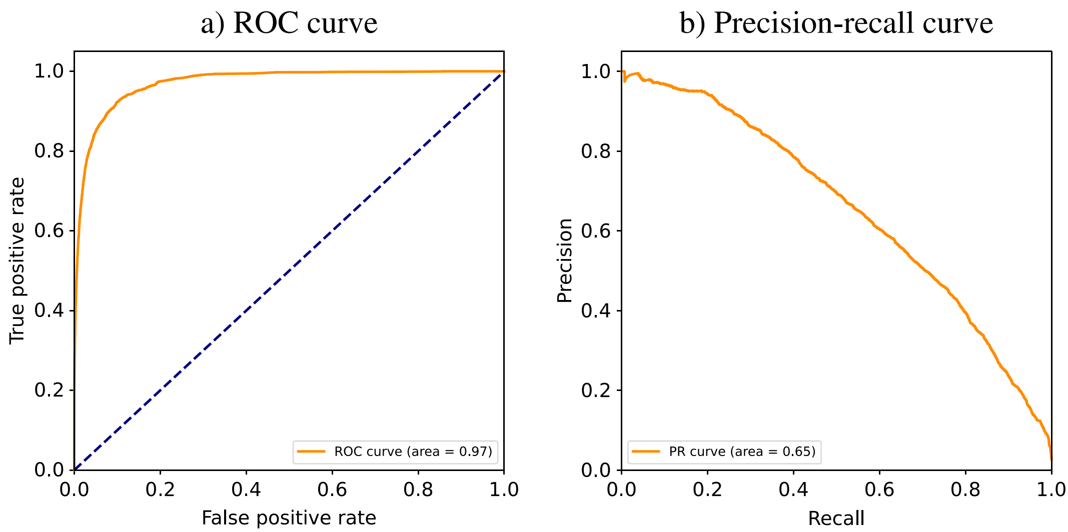

Figure 7. *ROC curve and precision–recall curve at subnational level.*

encourage users of our data to look towards our intensity model to gain insight into whether further escalations might occur.

3.2. Subnational

Figure 7 shows the ROC and precision–recall curve for the grid cell predictions published at conflictforecast.org. We began publishing the subnational predictions on the website in March 2022 and then had some interruptions, while we were restructuring the data pipeline. Since the predictions are for the next 12 months, we, therefore, are only evaluating the 3 months that we predicted out-of-sample. The ROC-AUC is very high, with 0.97, and the precision–recall curve indicates that if we want to capture half of all conflicts, our list of grid cells would contain three-quarter of cases that were correctly classified as conflict.

4. Datasets

4.1. National level forecasts

Table 5 provides a description of the variables in the datasets we make freely available at conflictforecast.org. Note that these are available to download as separate data files according to the target variable and time horizon. We publish our forecasting outputs from the text and best (i.e., combination of text and historical conflict features) models. The text model prediction tends to be updated during the first week of a month and the best model after the 20th day of a given month. This delay is due to the fact that the best model relies on updated input from the UCDP data on fatalities, which tend to be published on the 20th of a given month.

The website provides a graphic illustration of current conflict risk levels across the globe in terms of a map. Figure 8 shows the risk of an outbreak of violence in the next 12 months across the globe, estimated with all the information available at the end of August 2023. The datasets that can be downloaded contain the information underlying the map as well as all past predictions.

4.2. Subnational level forecasts

The subnational-level forecasts are still in their beta version and are not derived from published academic research. In their current form, the prediction models are an extension of the national predictions and were

https://doi.org/10.1017/dap.2024.10 Published online by Cambridge University Press

Table 5. *National and subnational forecast dataset.*

Variable category	Variable	Definition
National		
Identifiers	isocode	Unique three-digit identifier of a country
	year	Observation year, ranging from 2010 to present
	month	Observation month, ranging from 1 to 12
Forecast outputs	text_model	Forecast generated by the text model for the given target variable
	best_model	Forecast generated by the best model for the given target variable
Features	topic_k	The weighted stock of topic share k, where k can range from 0 to $N-1$, $N = 15$—the total number of topics in the dynamic LDA model
	anyviolence_dp	Number of months since last month with at least one battle-related death
	armedconf_dp	Number of months since the last armed conflict episode
	populationwb	Country population in a given year
Country/month violence	anyviolence	1 if at least 1 battle-related death, 0 otherwise
	armedconf	1 if at least 0.5 fatalities per 1 mn inhabitants, 0 otherwise
	lnbest_pc[a]	Log transformation of fatalities per capita
Target variables	ons_anyviolence3	1 if at least 1 battle-related death in 3 months; N/A if any violence = 1; else 0
	ons_anyviolence12	1 if at least 1 battle-related death in 12 months time; N/A if any violence = 1; else 0
	ons_armedconf3	1 if at least 0.5 fatalities per 1mn inhabitants in 3 months; N/A if armedconf = 1; else 0
	ons_armedconf12	1 if at least 0.5 fatalities per 1mn inhabitants in 12 months; N/A if armedconf = 1; else 0
	lnbest_pc_3[b]	Log transformation of average fatalities per capita in 3 months
	lnbest_pc_12[b]	Log transformation of average fatalities per capita in 12 months
Subnational		
Identifiers	Gid	Unique PRIO grid cell identifier
	isocode	Three-digit identifier of a country
	countryname	Full country name
	Year	Observation year
	month	Observation month
Forecast outputs	Risk	Likelihood of any battle death in next 12 months
Coordinates	Longitude	East–west position on Earth's surface
	Latitude	North–south position on Earth's surface

[a]The log transformation is conducted as $\ln(x + 1)$, where x is equal to the number of fatalities divided by population.
[b]See note a of Table 1.

developed as part of a project for the Foreign, Commonwealth, and Development Office of the UK (Mueller et al., 2022).

In terms of target variables, we only provide the likelihood of a battle death in a grid cell. In terms of prediction windows, we only provide a forecast for the next 12 months. In other words, we predict how likely it is that a given grid cell will experience any battle death within the next 12 months. We train six separate random forest classification models depending on whether there currently is conflict in a grid cell, whether a neighboring grid cell is experiencing conflict, and whether the grid cell has

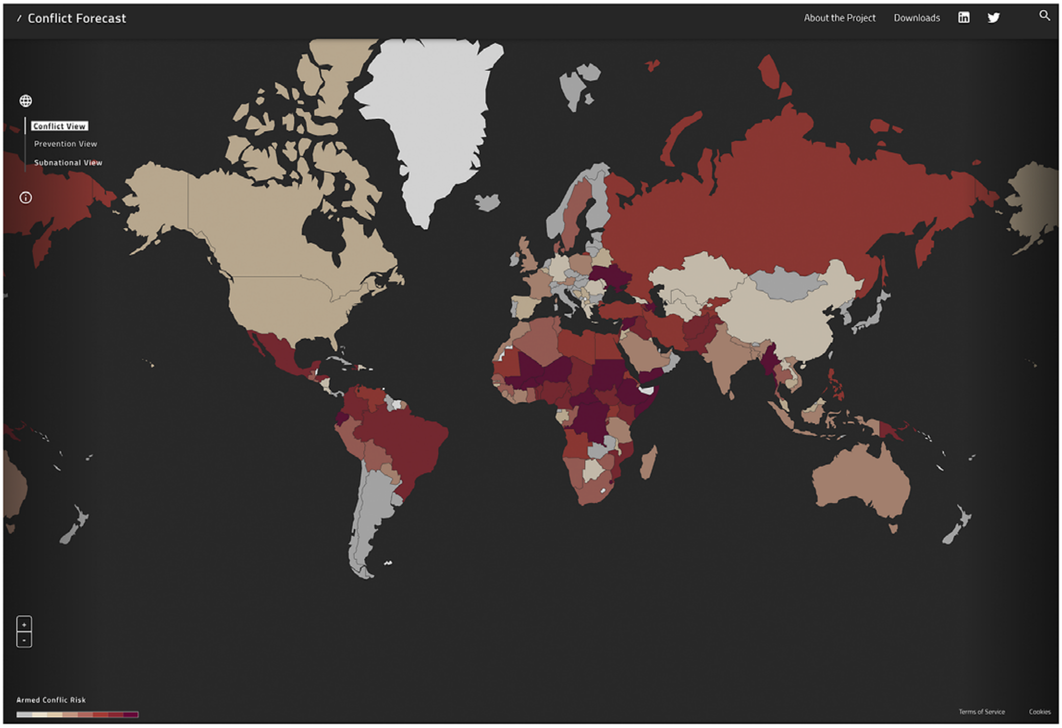

Figure 8. *National conflict risk across the globe at the end of August 2023.*

experienced any battle death in the past 5 years. More specifically, the following six models are trained on:

1. Regions that had no battle deaths in the last 5 years and have no battle deaths in their immediate neighboring cells.
2. Regions that had at least one battle death in the last 5 years but no ongoing violence and have no battle deaths in their immediate neighboring cells.
3. Regions that had no battle deaths in the last 5 years and have battle deaths in their immediate neighboring cells.
4. Regions that had at least one battle death in the last 5 years but no ongoing violence and have battle deaths in their immediate neighboring cells.
5. Regions experiencing ongoing violence and no battle deaths in their immediate neighboring cells.
6. Regions experiencing ongoing violence and battle deaths in their immediate neighboring cells.

The reasoning for this segmentation is twofold. First, we can derive predictions based on the conditional distribution of grid cells in a similar situation, and second, we can tailor the predictors to the current situation. For instance, for all regions without violence, the number of months of ongoing violence will be zero, and, therefore, this predictor need not be included for grid cells not experiencing violence. Again, the hyperparameters are chosen through threefold cross-validation on the sample until the year 2010. The resulting depth of trees varies between 3 and 5 nodes and the number of trees is either 300 or 400. In terms of grid-cell level predictors, we include

- the discounted past deaths,
- time since the last battle death,
- current battle deaths,

https://doi.org/10.1017/dap.2024.10 Published online by Cambridge University Press

- discounted and current neighboring battle deaths,
- local news topics,
- neighboring news topics,
- distance to capital,
- population,
- discounted and current number of riots and protests obtained from ACLED, the Armed Conflict Location & Event Data Project (Raleigh et al., 2010, 2023),
- time since the last riot and protest,
- consecutive months of battle deaths.

From the country level, we include

- discounted and current battle deaths and
- the time since the last battle death.

The downloadable dataset is summarized in Table 5. Besides providing the predicted risk of experiencing a battle death within the next 12 months, we also provide geographic identifiers for a given grid cell. This includes the PRIO id, the country the grid cell is (mostly) located, and the longitude and latitude of the center of the grid cell. Figure 9 shows a snapshot of risk at the grid-cell level from the website, which displays the risk data from the downloadable dataset.

4.3. National news-topic shares

Using relatively few topics helps to prevent topics from adapting to specific events, conflicts, regions or countries that dominate the news landscape. We manually label the topics for the sake of illustration, but these labels do not influence the predictions. Table 6 below lists the manual labels for our 15 topics and the top 10 keywords. Actually, topics are probability distributions across the entire dictionary of words in the corpus.

The algorithm provides the topic shares for each individual article. We aggregate these at the country/month level and make them available in the risk datasets downloaded from conflictforecast.org. Topic shares are provided for all 15 topics for each country and the period 2010m1 to the latest update. The topic model changes every month but each update produces a consistent interpretation of the entire news

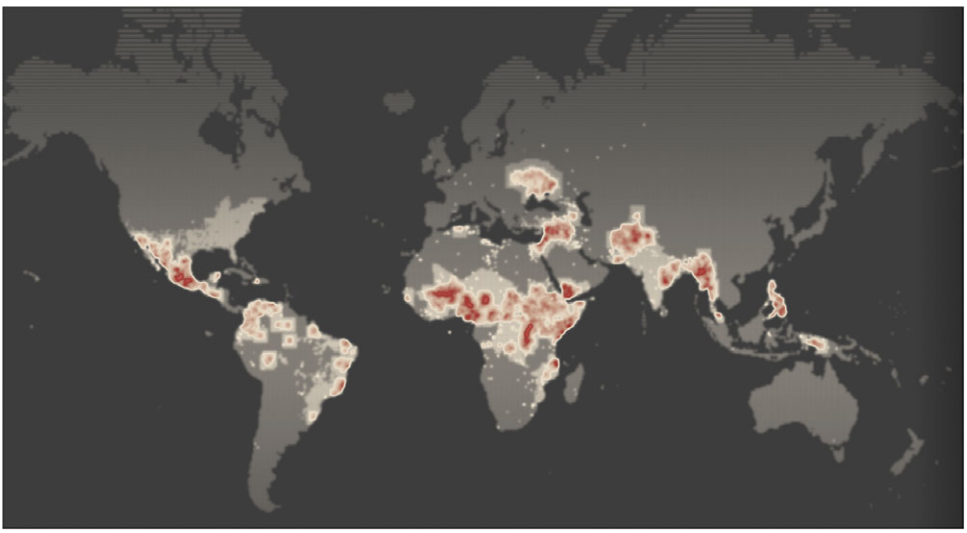

Figure 9. *Subnational conflict risk across the globe at the end of August 2023.*

https://doi.org/10.1017/dap.2024.10 Published online by Cambridge University Press

Table 6. *Top 10 keywords of the 15 topics and suggested labels*

Suggested title	Topic keywords
Change	Think long way little change big want look turn far
Cooperation	Cooperation foreign visit meet meeting discuss countries support security international
Defense	Western war military invasion security defense nuclear foreign include ally
Politics	Party election opposition political prime parliament rule vote politician leader
Economy	Economy price market economic company high increase bank billion global
Legal proceedings	Arrest court case charge law investigation police criminal security information
Conflict zones	Eastern-southern attack kills northern group forces militant province area
Development projects	Project company provides work plant service supply development power water
Sports	Team win play game second cup player match lead score
Rights	Group force capital violence kills armed military human protest war
Military operations	Military force air region defense telegram armed service missile near
Family life	Home city family live child die life know include come
War	War situation foreign political support pro speak want fact republic
Federal discussions	Federal discuss regional meeting national region meet town hold security
Health	Central pandemic communist communist_party health case leaders control party number

landscape for all countries and for the entire sample period. In Figure 10, we show an example of how the topic relating to politics evolved in the USA and the UK over time. The topic share exhibits clear spikes around political events such as elections of the Brexit referendum. While elections worldwide could be coded into a dataset, topics are able to pick up more subtle movements and events across the entire world.

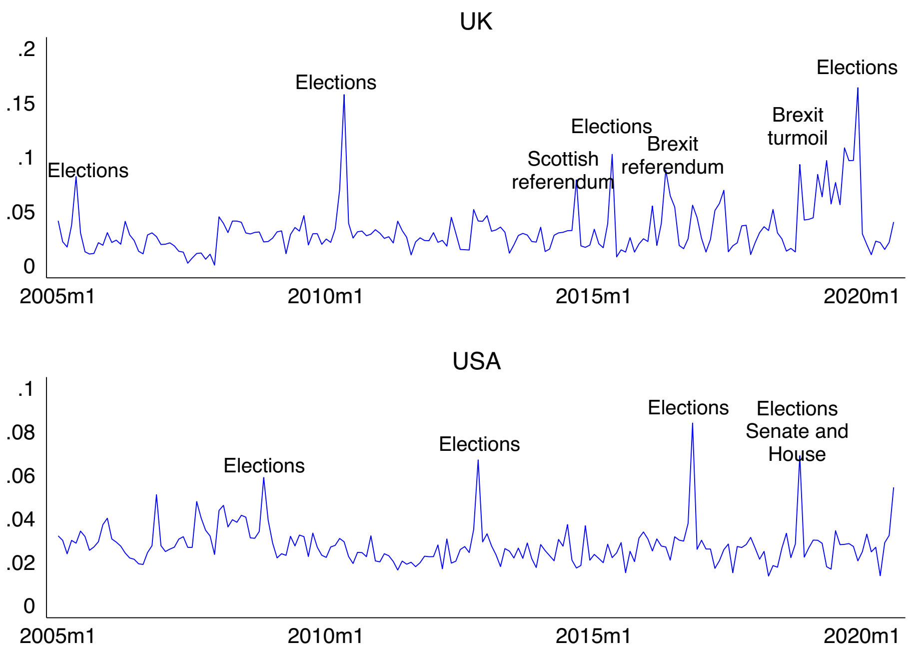

Figure 10. *Politics topics in the USA and the UK over time.*

https://doi.org/10.1017/dap.2024.10 Published online by Cambridge University Press

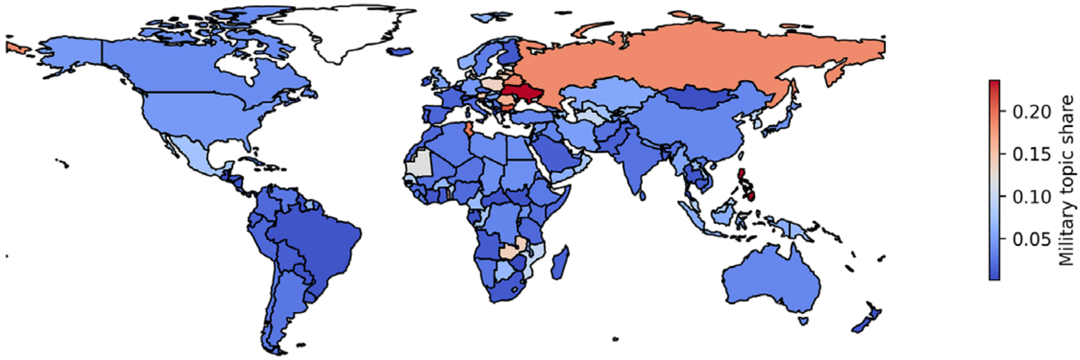

Figure 11. *Military topic share in August 2023.*

Figure 11 provides a global snapshot of the military topic share for August 2023. Unsurprisingly, reporting about themes related to the military features heavily in Russia and Ukraine. Besides being available in the downloadable datasets on the website, further information about the topics can be obtained by clicking on countries. Here, we provide information about the current distribution of topics within a country and whether these topics tend to, on average, be negatively or positively related to conflict risk.

As described in Table 5, the topics are attached to the downloadable risk forecast datasets rather than as separate datasets. We have not yet provided local news information, given that the grid-cell level data and predictions are at an earlier development stage and will be subject to change.

5. Usage

The available datasets can provide policymakers and non-governmental organizations with critical information about where to allocate resources and which potential catastrophes to make contingency plans for. Conflict risk estimates are invaluable tools for policymakers, offering multifaceted applications in the realm of peace and security governance. These estimates provide policymakers with a structured understanding of the likelihood of conflict outbreaks, enabling them to craft more informed and effective strategies for conflict prevention and management. One primary application lies in conflict prevention, where early identification of regions and situations at heightened risk allows for proactive measures to avert conflicts or diminish their severity. Additionally, policymakers can judiciously allocate resources, directing development aid, security forces, and humanitarian assistance to areas exhibiting higher conflict risk. Conflict risk assessments underpin diplomatic efforts, providing insights into potential triggers and involved actors, facilitating preventive diplomacy and mediation. They also help to track risks in countries during phases of stabilization, that is, post-conflict. This means, for example, that conflict risk data can guide the deployment of peacekeeping forces and inform the design of peacebuilding initiatives. Policymakers can leverage this information for crafting security policies, adjusting military deployments and intelligence activities to enhance national security. In the context of development planning, conflict risk estimates are indispensable, enabling policymakers to identify building conflict risks to develop targeted strategies. Furthermore, these assessments are instrumental in policy evaluation, allowing policymakers to gauge the effectiveness of interventions and make necessary adjustments. In essence, conflict risk estimates serve as critical instruments for policymakers, guiding their decisions and actions in the pursuit of peace, stability, and development both domestically and on the international stage.

For the academic community, the risk forecasts can be used both as an outcome variable as well as for matching purposes. For instance, Mueller and Rauh (2022a) use risk forecasts to match countries that introduce power-sharing agreements to those that do not. The underlying idea is that both countries, absent power-sharing agreements, are predicted to have the same trajectory. This allows for a fair comparison with a valid counterfactual. Similarly, the risk forecast can be used as an outcome variable. Policies may not only reduce violence, which is a rare outcome, but also latent risk. Having reliable risk forecasts shines a light on unobserved risk levels.

The dataset containing news topic shares can be used by policymakers and researcher for their own prediction models. Moreover, they can be studied by the academic community that is interested in media reporting more generally. The news topics are not inherently limited to the study of conflict. They provide a general picture of the reporting landscape across the world over time.

6. Conclusion

This dataset description illustrates that forecasts of *armed conflict* are possible even with long forecast horizons and even of onsets that are occurring in countries that have previously been peaceful. The risk data provided in this way can be useful for a large number of applications and should be able to inform preventative policies around the world.

While the available datasets already provide reliable and rich sources of information, the project continues to expand and improve. We will refine models and their inputs, and add target variables in the future. The grid-cell level predictions will, at some point, include predictions of the level of violence. Moreover, regions will be summarized into administrative units, such as states or provinces, which may be preferred by some stakeholders due to their interpretable nature. The *violence intensity* prediction is still in its beta version and is geared towards providing an indicator for potential escalations where conflict is already taking place. As a consequence, the model performs poorly where there currently is no violence. We aim to improve this forecast to provide a comprehensive intensity measure, including uncertainty estimates in the form of prediction intervals. Moreover, all of the up-to-date codes will also be made publicly available in the future. When these extensions will be completed depends on funding, capacity constraints, and the success of the models.

Data availability statement. The conflict predictions and topics are updated monthly on https://conflictforecast.org. The prediction code is being refined continuously. A previous version of the replication data and code can be found in Harvard Dataverse: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/UX8GUZ.

Acknowledgments. We are grateful for research assistance from Bruno Conte Leite, Luis Ignacio, and Alexandra Malaga. A working paper version of this paper is available at Mueller et al. (2024).

Author contribution. Conceptualization: B.S., H.M., C.R.; methodology: H.M., C.R.; data curation: B.S., H.M., C.R.; data visualization: B.S., H.M., C.R.; writing original draft: B.S., H.M., C.R. All authors approved the final submitted draft.

Funding statement. This research was supported by a grant from the Keynes Fund. We acknowledge financial support under the FCDO agreement *Conflict Forecast: Evaluating Intervention Options for Nigeria* and the ERC Advanced grant, ANTICIPATE (ID: 101055176). Mueller acknowledges support from the Spanish Government and the Spanish Ministry of Economy and Competitiveness through the Severo Ochoa Programme for Centres of Excellence in R&D (CEX2019-000915-S) and the Ministry of Science Innovation and Universities (PGC2018-096133-B-100).

Competing interest. The research does not reflect the opinions of the aforementioned funding institutions.

Ethical standard. The research meets all ethical guidelines and legal requirements.

References

Bazzi S, **Blair RA**, **Blattman C**, **Dube O**, **Gudgeon M and Peck R** (2022) The promise and pitfalls of conflict prediction: evidence from Colombia and Indonesia. *Review of Economics and Statistics 104*(4), 764–779.
Blei DM, **Ng AY and Jordan MI** (2003) Latent Dirichlet allocation. *Journal of Machine Learning Research 3*(Jan), 993–1022.
Boschee E, **Lautenschlager J**, **O'Brien S**, **Shellman S**, **Starz J and Ward M** (2015) ICEWS Coded Event Data. https://doi.org/10.7910/DVN/28075.
Breiman L (2001) Random forests. *Machine Learning 45*, 5–32.
Chadefaux T (2014) Early warning signals for war in the news. *Journal of Peace Research 51*(1), 5–18.
Davies S, **Pettersson T and Öberg M** (2023) Organized violence 1989–2022, and the return of conflict between states. *Journal of Peace Research 60*(4), 691–708.

Goldstone JA, **Bates RH**, **Epstein DL**, **Gurr TR**, **Lustik MB**, **Marshall MG**, **Ulfelder J and Woodward M** (2010) A global model for forecasting political instability. *American Journal of Political Science 54*(1), 190–208.
Halterman A, **Schrodt PA**, **Beger A**, **Bagozzi BE**, **Scarborough GI** (2023) Creating custom event data without dictionaries: A bag-of-tricks. Preprint, arXiv:2304.01331.
Hegre H, **Allansson M**, **Basedau M**, **Colaresi M**, **Croicu M**, **Fjelde H**, **Hoyles F**, **Hultman L**, **Högbladh S**, **Jansen R**, **Mouhleb N**, **Muhammad SA**, **Nilsson D**, **Nygård HM**, **Olafsdottir G**, **Petrova K**, **Randahl D**, **Rød EG**, **Schneider G**, **von Uexkull N and Vestby J** (2019) ViEWS: A political violence early-warning system. *Journal of Peace Research 56*(2), 155–174.
Hegre H, **Bell C**, **Colaresi M**, **Croicu M**, **Hoyles F**, **Jansen R**, **Leis MR**, **Lindqvist-McGowan A**, **Randahl D**, **Rød EG**, **Akbari F**, **Croicu M**, **Dale J**, **Gåsste T**, **Jansen R**, **Landsverk P**, **Leis M**, **Lindqvist-McGowan A**, **Mueller H** *et al.* (2022a) Forecasting fatalities. Mimeo. Available at https://www.diva-portal.org/smash/get/diva2:1667048/FULLTEXT01.pdf
Hegre H, **Bell C**, **Colaresi M**, **Croicu M**, **Hoyles F**, **Jansen R**, **Leis MR**, **Lindqvist-McGowan A**, **Randahl D**, **Rød EG and Vesco P** (2021) ViEWS2020: revising and evaluating the ViEWS political violence early-warning system. *Journal of Peace Research 58*(3), 599–611.
Hegre H, **Croicu M**, **Eck K and Högbladh S** (2020) Introducing the UCDP candidate events dataset. *Research & Politics 7*(3). https://doi.org/10.1177/2053168020935.
Hegre H, **Metternich NW**, **Nygård HM and Wucherpfennig J** (2017) Introduction: Forecasting in peace research. *Journal of Peace Research 54*(2), 113–124.
Hegre H, **Vesco P and Colaresi M** (2022b) Lessons from an escalation prediction competition. *International Interactions 48*(4), 521–554.
Hoffman M, **Bach F and Blei D** (2010) Online learning for latent dirichlet allocation. *Advances in Neural Information Processing Systems 23*, 1–9.
Kleinberg J, **Ludwig J**, **Mullainathan S and Obermeyer Z** (2015) Prediction policy problems. *American Economic Review 105* (5), 491–495.
Mueller H and Rauh C (2018) Reading between the lines: Prediction of political violence using newspaper text. *American Political Science Review 112*(2), 358–375.
Mueller H and Rauh C (2022a) Building bridges to peace: A quantitative evaluation of power-sharing agreements. Technical Report, Working paper, Barcelona School of Economics.
Mueller H and Rauh C (2022b) The hard problem of prediction for conflict prevention. *Journal of the European Economic Association 20*(6), 2440–2467.
Mueller H and Rauh C (2022c) Using past violence and current news to predict changes in violence. *International Interactions 48* (4), 579–596.
Mueller H, **Rauh C and Ruggieri A** (2022) Dynamic early warning and action model. Technical Report, Working Paper, Barcelona School of Economics.
Mueller H, **Rauh C and Seimon B** (2024) Introducing a global dataset on conflict forecasts and news topics. Technical Report, Working paper number 2402, Janeway Institute.
Raleigh C, **Linke A**, **Hegre H and Karlsen J** (2010) Introducing ACLED: An armed conflict location and event dataset. *Journal of Peace Research 47*(5), 651–660.
Raleigh C, **Linke A**, **Hegre H**, **Karlsen J**, **Kishi R and Linke A** (2023) Political instability patterns are obscured by conflict dataset scope conditions, sources, and coding choices. *Humanities and Social Sciences Communications 10*(1), 1–17.
Řehřek R and Sojka P (2010) Software framework for topic modelling with large corpora. Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks, 45–50.
Rohner D and Thoenig M (2021) The elusive peace dividend of development policy: From war traps to macro complementarities. *Annual Review of Economics 13*, 111–131.
Sundberg R and Melander E (2013) Introducing the UCDP georeferenced event dataset. *Journal of Peace Research 50*(4), 523–532.
Tollefsen AF, **Strand H and Buhaug H** (2012) PRIO-GRID: A unified spatial data structure. *Journal of Peace Research 49*(2), 363–374.
Vesco P, **Hegre H**, **Colaresi M**, **Jansen RB**, **Lo A**, **Reisch G and Weidmann NB** (2022) United they stand: Findings from an escalation prediction competition. *International Interactions 48*(4), 860–896.
Ward MD, **Greenhill BD and Bakke KM** (2010) The perils of policy by p-value: Predicting civil conflicts. *Journal of Peace Research 47*(4), 363–375.

Cite this article: Mueller H, Rauh C and Seimon B (2024). Introducing a global dataset on conflict forecasts and news topics. *Data & Policy*, 6: e17. doi:10.1017/dap.2024.10