A Morden Computational Approach to Conflict Prediction

Simon Polichinel von der Maase

January 2019

Character count: 43.950/44.000

Contents

Abstract

This paper present a modern computational approach to conflict prediction at the sub-national level. Using various features from the literature on civil war and internal conflicts and an ensemble of modern boosting algorithms I create a predictive framework. The framework is capable of estimate the probability of conflict across a global grid of cells - with each cell measuring 0.5×0.5 decimal degrees. The approach shows good potential, but while the framework can successfully captures most conflict zones the price is a large number false positives compared to true positives. More crucial, however, it is not capable of capturing sudden political developments such as ceasefires or identify new onset far removed in time and space from any present or past conflicts. Three recommendations are given pertaining future endeavours; better handling of the imbalanced data, better modeling of temporal/spatial patterns; and the inclusion features derived from large scale text analysis.

1 Introduction

Humans kill humans - a lot of humans. In 2009 an estimated 46,772 people were killed in internal conflicts alone[1]. The number of lives ruined as a consequence of these killings no doubt exceeds the number of deaths by magnitudes. Even far away from the conflict zones people experiences the consequences of conflict - not least in the form of large influx of refugees and the ensuring political turmoil. The long term solution to this senseless onslaught probably entails some fundamental changes to the global power structures involving the cooperation of the global political, economic and cultural elites alike. However, while we wait for this eventuality to ensue, another, less Utopian approach is be to mitigate the ongoing catastrophe by creating a reliable "early warning system". A predictive framework capable of estimating the probability that a given geographic location will experience conflict during a given time period. Using this information governments, NGOs or other responsive actors can act to prevent the conflict, mitigate the fallout, or secure civilians. Naturally, such interventions are both disruptive and expensive. Thus, for any predictive framework to work as a practical early-warning-system a high level of accuracy is needed; we need to ensure that the geographic locations, which are being appointed the highest risk by the framework, are indeed where future conflicts are most likely to occur.

This paper presents a modern computational approach to conflict prediction and forecasting. It is constructed as a unified framework to asses the probability that a given sub-national geographical location will experience fatalities as a direct consequence of intra-state conflict. The challenge is handled as a forecasting problem, thus the prediction target is whether or not the given location experiences any battle-related deaths *"next year"*.

This is a preliminary project, thus the aim is as much to explore fruitful approaches and methods for future research as it is making good predictions. As a consequence, the larger part of this paper is dedicated to choosing what goes into the model and not least evaluating what comes out.

What goes in to the model is a roster of features derived from both the theoretical and empirical literature regarding civil wars and intra-state conflict. What comes out of the model, is first of all probabilities of conflict deaths pertaining to a given geographical cell in a given year; which are then evaluated against the actual observations. But just as importantly, the framework allows for evaluation of the individual feature's importance in the prediction process. This insight is paramount in deciding how to improve the framework and where to invest resources in future endeavours.

The framework constructed shows great promise; out-of-sample prediction generally estimates high probability of conflicts where conflicts actually ensues. One challenge is that the framework

[1]according to the best estimate from the Uppsala Conflict Data Program

still has a hard time separating very close and similar cells from each other leading to some lack in precision. That is, if we where to classify all cells with a probability of conflict above 10% as future conflict zones, we capture almost 80% percent of all future conflicts; however, we would also classify two false positives for each true positive. Furthermore, the model struggles to capture sudden political developments such as peace agreements, and would not fare well predicting new conflict-onsets far removed from any other conflicts in time and space.

Going forward with future endeavours I recommend allocating resources to create a unified and systematic framework modeling the temporal-spatial evolution of conflicts as a function of it self. For one thing, it seems the most productive way to improve prediction power as features pertaining these dimensions contributed with nearly all the prediction power in the present framework. Secondly, but just as important, the conflict data is more frequently updated and the last entry is closer to the present year then the contextual features used in this project.

Another source of more up-to-data and topical data could be text data from news sources or political outlets. This might also amend the framework's inability to capture novel political developments.

The following subsections will first present the motivation followed by the research questions and design. Then, a short presentation of the included features. Next I present the predictive framework and a thorough analysis of the derived results. This is proceeded by a discussion regarding future challenges and suggestions for improvements. Lastly a conclusion sums up the main findings.

1.1 The 'Why'

> *[...] the estate of Man can never be without some incommodity or other; and [...] the greatest, that in any form of Government can possibly happen to the people in generall, is scarce sensible, in respect of the miseries, and horrible calamities, that accompany a Civill Warre;* (Hobbes, 1651, 128)

The perils and miseries of civil war and internal conflict have plagued mankind all throughout history - not least in modern times. Since the conclusion of the Second World War intra-state conflicts have been far more common than inter-state wars (Collier and Hoeffler, 2004, 563); Over five times as many people have died in intra-state conflicts compared to inter-state wars (Collier and Hoeffler, 2004, 563); and since 1960 over one half of all nations have experienced some sort of violent internal conflict leading to fatalities (Blattman and Miguel, 2010, 3-4).

Importantly, Internal conflicts should not be viewed as internal affairs of little concern to other then the inflicted host and its allied. Examples of spillover-effects facilitating the spread of conflicts across boarders a ample. At country level, having a country located in a conflict ridden neighbourhood have been shown to be a robust predictor of internal conflict (Hegre and Sambanis, 2006; Goldstone et al., 2010). Internal conflict is thus a highly destructive and potentially contentious malaise. Understanding how internal conflicts originates and spreads in order to prevent or mitigate the destruction is indeed as crucial as ever.

Encouragingly, developments in statistical techniques, data availability and computational power makes the endeavour slightly more feasible with each passing year [2]. Not least the recent shift from cross country comparison towards disaggregated analyzes on sub-country unites holds great potential (Cederman and Gleditsch, 2009). This disaggregated approach is further enhanced by evermore accessible geospatial software and powerful new machine learning algorithms. As I will show, these developments can aid us in the construction of a reliable early-warning-system.

1.2 The 'How'

The endeavour at hand can be summed up by three research questions:

First research question(Q_1): To what extent[3] is it possible to predict the geographic location of future intra-state conflicts using a modern computational approach.

Second research question(Q_2): What phenomenons and features presents themselves as the most important in this prediction effort.

Third research question(Q_3): Given the conclusions pertaining to Q_1 and Q_2 what can be done to create an even more informative feature space in the future.

To answer these questions I use data from the Uppsala Conflict Data Program (UCDP) (Sundberg and Melander, 2013; Croicu and Sundberg, 2017). This data includes counts of conflict related deaths along with both coordinates of the scene and estimated time of the event. The coordinates are linked up to specific geographical cells of 0.5×0.5 decimal degrees derived from the PRIO grid database (Tollefsen et al., 2012). I aggregate the conflict data to sum up the yearly number of conflict deaths in a given cell. This measure is further dichotomized such that it only indicates whether or not a given cell experienced conflict deaths or not. I "lead" the measure, effectively lagging all explanatory features to come. Or put another way; I shift the target feature one year ahead (t+1) such that the explanatory features of e.g. 2006 will try to predict conflicts

[2] Unless, of course, conflict is inherently shrouded in ontological uncertainty rather than epidemiological uncertainty as implied by Gartzke (1999)

[3] The phrase "To what extent [...]" present in Q_1 implies the caveat: " - Given the scope of the project and the limited computational resources at my disposal. As this is preliminary research I will spend more time evaluating the results and discussing how to improve future frameworks than I will training and fine-tuning the model and optimizing hyper parameters.

in 2007. To further mimic the forecasting nature of the problem, the model is created only on the basis of data from 1990 through to 2005. Data from 2006 to 2010 are reserved for model evaluation through out-of-sample prediction.

To create the explanatory features I borrow from both the Uppsala conflict data itself and from the large number of features available from the PRIO grid database. Due to limitation in the available data from the PRIO grid database, the date only covers 1990 through 2010. This is naturally a big obstacle if the framework is ever to be applied in practice, and as such, I shall return to this challenge in the discussion. Further description of the data sources can be found in the appendix, subsection 7.1.

From here, I construct a predictive framework using an ensemble of xgboost algorithms. I use a large ensemble to generate a more robust result and to facilitate insights into the uncertainty inherent in the prediction effort.

The next section will present the features derived from the data sources briefly introduced above.

2 The Included Features

In this section I will introduce the roster of features used in the endeavour at hand. Some features are readily available from one of the two data sources, other requires some feature engineering before they correspond to the theoretical phenomenon believed to be related to internal conflicts. I draw on insights from both the country aggregated civil war literature and the more resent disaggregated conflict litterateur. As such, some features are cell specific, while others are country specific, and yet others denotes the difference between cell and country values. While it would be satisfying to present the features through a comprehensive literature review, the scope and focus of this endeavour does not allow such academic gluttony[4]. The appendix, subsection 7.2 presents slightly more theoretical context for all features along relevant mathematical definitions.

Wealth and Capacities: Reflecting the various arguments regarding wealth as a conflict inhibitor as put forth in Collier and Hoeffler (1998); Fearon and Laitin (2003); Collier and Hoeffler (2004). Here, however, with *night light emission* as a proxy for wealth, as proposed by Elvidge et al. (2009), Chen and Nordhaus (2011) and Cederman et al. (2013). I have included both a cell-specific and country specific measures, both divided by the population count of the specific country/cell.

[4]Readings which, together, serve as a rather comprehensive review are Hegre and Sambanis (2006), Kalyvas (2007), Cederman and Gleditsch (2009) and Blattman and Miguel (2010)

Inequality and deprivation: Reflecting the argument of relative deprivation as a facilitator of conflict (Gurr, 1970; Cederman et al., 2013). Again, I utilize night light emission as a proxy for wealth and I included both the specific measure used in Cederman et al. (2013) and a more simple measure simply capturing whether or not a given cell is less wealthy compared to the median cell of the corresponding country.

Ethnicity and Exclusion: A dichotomous feature denoting whether or not the cell is inhabited by one or more "excluded" ethnic groups at any given year. This captures the argumentation put forth in Cederman et al. (2011, 2013)[5].

Population size and density: The correlation between the size of a country's population and civil war has been found rather robust (Collier and Hoeffler, 1998; Fearon and Laitin, 2003; Fearon, 2004; Collier and Hoeffler, 2004; Hegre and Sambanis, 2006). I include measures of absolute[6] population counts for both cell and country.

Geography and Accessibility: Fearon and Laitin (2003) argue that rough terrain and mountains are natural obstacles hindering effective projection of state power. Hegre and Sambanis (2006) conclude, that a feature for rough terrain is robustly correlated with civil war across a large number of model specifications.(Hegre and Sambanis, 2006, 526-529)[7]. The PRIO grid includes a readily available feature measuring the proportion of mountainous terrain within the cell[8] which I utilize.

Distance to Power Center: Another natural hindering for projecting state power is sheer distance (Fearon, 2004; Buhaug et al., 2009; Cederman et al., 2009; Buhaug, 2010). To capture phenomenons like this I have included the distance to the nation's capital[9], the travel time to the nearest major city, and the total size of the country (Tollefsen and Buhaug, 2015).

Trans-boarder Influences: A number of different mechanism have been proposed and explored (Blattman and Miguel, 2010, 29-30), the one explored here is rather simple and follows Hegre and Sambanis (2006). This is simply the distance to the nearest (land) boarder shared with another country.

[5]the measures are originally from the GeoEPR/EPR data by Vogt et al. (2015)

[6]Density and logged was tried but yielded no better results

[7]Though, see Goldstone et al. (2010)

[8]based on Blyth et al. (2002)

[9]The measure utilized by Buhaug (2010)

Prime Commodities, Oil and the Recourse course: A lot have been written on this subject, but the empirical results have varied a lot (Collier and Hoeffler, 1998; Fearon and Laitin, 2003; Fearon, 2004; Ross, 2004; Collier and Hoeffler, 2004; Fearon, 2005; Buhaug, 2010; Hegre et al., 2009). In this endeavour, I include a dichotomous feature denoting whether or not a given cell is known to hold oil deposits. Thus, a cell is not denoted as having oil before oil is actually discovered.

Inertia, dispersion, traps and time trends: Conflict traps and inertia have been modelled or proposed by Collier and Hoeffler (2004), Hegre and Sambanis (2006), Perry (2013) and Cederman et al. (2013) while cross-country dispersion has been used by Goldstone et al. (2010). I include a number of features which aims to capture the pattern of conflict as it moves through space and time. Distance from cell center to nearest conflict, yearly fatalities in cell, yearly fatalities in the country as a whole and the total number of years the given cell has been a conflict zone in the past and .

As already noted, the features are measured at $year = t$, with the target at $year = t+1$. Thus, when I e.g. define "distance from cell center to nearest conflict", the measure concerns the year preceding the actual prediction.

A lot more features and specifications were applied and tested[10] and even more could have been scrutinized. Furthermore some of the included features could be even better specified. Never-the-less, the features presented above will do for now. I now proceed to presenting the model, the estimations and the results.

3 The Predictive Framework and the Results

In the follow sections I present the predictive framework and evaluate the performance of the framework through appropriate metrics. One of the greater challenges pertaining to the project at hand is the imbalance of the data; most of the time, most grid cells do not experience any conflict fatalities. The data as such is imbalanced and only around 1% of the data constitutes "events" - actual conflicts. To handle this challenge I use a suitable estimation procedure, I under-sample the non-events and I utilize the appropriate evaluation metrics. These approaches will all be presented in the sections below.

But first, I will briefly highlight some of the main difference between a dedicated prediction efforts such as this endeavour, and the parameter estimations that many scholars of social and political science are accustomed to.

[10]Not least logged variants and various density measures. See the appendix, subsection 7.2

3.1 Parameter Estimation vs. Prediction

Most scholars of political science are familiar with the process of estimating parameters through linear or logistic regression. Often the main concern in these efforts is to ensure that the parameter estimates are unbiased. The endeavour at hand takes a different focus; predicting future events. The framework produces no parameters to evaluate, I do not deal with significance level and I do not overly concern my self with bias. Instead I concern myself mainly with increasing prediction power and preventing "over-fitting". Traditional models for parameter estimates are perfectly capable of making predictions, but these predictions are often rather poor not least do to overfitting (Ward et al., 2010). Overfitting, to put it simply, is the accidental identification of patterns present in the data, yet not present as a phenomenon in the real world (McElreath, 2018, 166-168). In quantitative studies of conflicts scholars have often neglected taking steps to counter overfitting leading to bad estimations and possible erroneous conclusions (King and Zeng, 2001a; Ward et al., 2010).

Preventing overfitting is paramount if I am to generate reliable predictions for real world applications. One tool available is using out-of-sample prediction as evaluation approach (King and Zeng, 2001a; Ward et al., 2010; Perry, 2013; Schrodt, 2014). This simply entails construction the predictive framework on one set of data and testing it on another; a trainingset and a testset. While out-of-sample predictions can give an honest evaluation and reveal overfitting, other tools are needed to mitigate and prevent the problem in the first place. Some of these I shall return to shortly.

I use a ratio of roughly 25/75 for respectively test- and trainingset (Friedman et al., 2001; Ward et al., 2010). To mimic the forecasting element of the project I choose the fist years (1990-2005) as the training set and the last years (2006-2010) of the data as the testset (Goldstone et al., 2010).

I shall return to how exactly to evaluate the out-of-sample performance of the framework further down. In the next section I will first introduce the predictive framework it self.

3.2 Predictive Framework

The predictive framework is made up by a subset of approaches. The subsections below present each step at a time, before moving on to the evaluation effort in the next section.

3.2.1 Extreme Gradient Boosting

I utilize the Extreme Gradient Boosting (xgboost) algorithm as the basis of my predictions (Chen and Guestrin, 2016)[11]. This method is a bit advanced and I shall refrain from digging into the technicalities of the algorithm. However, some fundamentals serve to explain why this framework is especially well suited for the problem at hand. Three characteristics needs presenting; it is a boosting algorithm, it is consisting of regression trees and it is self-regularizing (or self-pruning).

Boosting involves using a lot of weak classifiers to create a strong classifier. Imagine that we begin with one classifier with which we try to classify geographic locations which will experience conflict fatalities next year. The classifier does rather poorly, not least due to the fact that the events are few and and hard to classify. Thus, we decide to run a new classifier, but now we give less weight to the observations which was correctly predicted and more weight to the observations which was incorrectly predicted. We reiterate this procedure until some measurement criteria is met. Then, all our classifiers are weighted according to their performance and used as a weighted ensemble to predict which geographic locations will experience conflict fatalities of the following year (Friedman et al., 2001, 338-339). Since the procedure ensures continuing focus on "hard to classify" observations, it is a particularly fruitfully approach when dealing with imbalanced data.

The Next question is naturally which classifiers to build our boosting framework on. Xgboost utilizes regression tree. These closely resembles decision tress where the features are used to split the observations into categories according to which splits yields the most information according to some predefined criteria. Thus, the first split would here be the split which best sorted conflict events from non-events and so on. But unlike decision trees, each regression tree holds a continuous score on each of the leafs, which can be converted into probabilities rather then binary classifications (Chen and Guestrin, 2016, 2). To put it simply, xgboost utilizes N number of regression tress and iteratively uses the predictions of one tree to update the weights of the next tree, such that 'harder to classify' observations are given continuously more and more attention.

The last question here is then how to decide to number of times we shall allow a given tree to split. Too few times and the tree fails to learn any pattern of the data and we underfit. To many splits and the tree starts to learn idiosyncratic artifacts of the training data and we overfit. The xgboost algorithm handles this by penalizing complicated tress. Thus, when the algorithm evaluates whether or not to make a split, the information gain obtained by the potential split is down-weighted according to the increased complexity induced by new splits (Chen and Guestrin, 2016, 4-7). Thus, the algorithm searches for any patterns which might help the prediction effort, but stops before such pattern becomes overly complicated.

Naturally there is more to the xgboost algorithm than presented here, not least the fact that it

[11] Implemented through python

has been optimized for sparse data in a number of more technical ways making it even more suited for large and imbalanced data-set then other boosting algorithms (Chen and Guestrin, 2016, 5). Furthermore, the algorithm has a plethora of hyper parameters which can be tuned and optimized to further enhance performance and reduce overfitting[12]. This is beyond the scope of this project however, and while it would probably increase the predictive power of the framework somewhat, it would change nothing in regards to the overall approach or the challenges going forward.

Thus, while the complexity is great and the opportunities many, the small introduction above will suffice to illustrate why the approach works well for the problem at hand but also why, as we shall see, there still are some challenges ahead.

A final note on the xgboost algorithm pertains to an important property induced by the approach being based on regression trees: one can readily asses how "important" each included feature is in the prediction effort. Given the exploratory nature of the present endeavour it is paramount that I am able to evaluate which features might present the most fertile research areas in future prediction efforts.

3.2.2 Undersampling

While the boosting approach presented above is able to handle imbalanced problems, we can help it along by re-sampling our training data to create a more balanced ratio between events and non-events. This is called undersampling (He and Garcia, 2008, 1266-1267). The specific procedure employed here is called *case-cohort sampling*. I utilizes all available events from the trainingset along a random subset of no-events from the trainingset in order to construct the classifiers used (King and Zeng, 2001b, 142). Some trial and error showed that around 20 times as many non-events as events yielded the best prediction results. Now, instead of events being 1% of the training data, they now constitute around 5% of the training data. The actual probabilities estimates are, however, inflated since the model thinks that conflicts are less rare than is the actually the case; this is corrected using Bayesian prior correction closely related to what is presented in King and Zeng (2001b), King and Zeng (2001a) and Goldstone et al. (2010). The procedure is presented in the appendix, subsection 7.3. Suffice it to say it corrects the estimated probabilities according to the true propensity of internal conflicts. As I refrain from setting any final hard threshold regarding when a predicted probability should manifest into conflict, this correction does not change the overall performance of the framework; it merely presents the estimated probabilities in a more intuitively appealing and honest light.

Returning to the task at hand; while most of the information in the data is stored in the events and not the non-events (King and Zeng, 2001b, 139), some information is still lost by not using all

[12] Usually optimized using k-fold cross validation (Friedman et al., 2001, 241-249)

non-events. The chosen solution is a variant of *informed undersampling.* To counter the inefficient use of information, one can use an ensemble of classifiers all utilizing a new independent subset of non-events to combine with the subset of all available events (He and Garcia, 2008, 1267). This approach not only utilizes all information available, but also facilities the estimation of the uncertainty inherent in the prediction effort. Specifically, I run the same xgboost algorithm 1000 times using a new independent subset of non-events each time. Thus, for all results derived from the framework, predictions, evaluations metrics and feature importance alike, I get a distribution of 1000 results effectively constituting the probability distribution of the given results[13]. For both the predictions and the evaluations metrics, the mean of these distributions constitutes a more reliable and honest point estimate of respectively the "true" risk of conflict and the "true" performance of the framework. Furthermore I also get estimates regarding the uncertainty of the predicted risk and the variance of the framework's performance.

This brings us neatly to the evaluation effort.

3.3 Evaluating the Framework

In the following subsections, I evaluate the predictive capabilities of the framework presented above. As mentioned, I use out-of-sample prediction as basis for the evaluation. There exists a plethora of different evaluation metrics designed to convey the performance of a given framework's out-of-sample performance. Given the specific task at hand, some are more appropriate than others. Thus, I reserve the first subsection below to introduce the specific evaluation metrics utilized in the present endeavour. This is followed by the evaluations.

3.4 The Evaluation Metrics

Many metrics are available, but the most commonly known are probably "Accuracy", "Receiver Operating Characteristic" curve and the related measure "Area Under the Curve" (AUC) scores. While accuracy is intuitively meaningful and the ROC/AUC score has been coined as a "gold-standard" (Perry, 2013, 366), these measures also have some inherent limitations in regards to the specific problem at hand. Accuracy is not very suited for imbalanced data (He and Garcia, 2008, 1264). The ROC/AUC score tend to judge performance on highly imbalanced data too favorable (He and Garcia, 2008, 1278). Going forward, I instead report the two metrics "Recall" and "Precision" along two closely related ratio metrics; the "Precision-Recall Curve" and the "Average Precision Rate". These measures are suitable for imbalanced data; have a substantial and conveyable interpretation; judges the present framework appropriately honest and harshly;

[13]Given the data and model specifications utilized of course

and with the ratio metrics we circumvent the problem of a hard threshold (He and Garcia, 2008, 1278).

Denoting "True Positives" TP, "False Posits" FP and "False Negatives", recall and precision are defined as follows[14]:

$$precision = \frac{TP}{TP + FP} \tag{2}$$

$$recall = \frac{TP}{TP + FN} \tag{3}$$

As such, both scores are ratios which takes a value between 0 and 1. Intuitively, precision is the ratio between correctly predicted events and all predicted events. It tells us how many of the events we predict actually turns out to be events. Recall is the other side of the coin. It is the ratio between the correctly predicted events and all true events. It denotes how many of the true events our framework actually identifies. As seen in equation 2 and 3 above, more false negatives results in a lower recall score. More false positives results in a lower precision score. This is exactly the information needed to accurately evaluate the highly imbalanced data of rare events. Through these metrics we learn how many of the actual rare events we are able to capture, and how many non-events we incorrectly classify as events in the process.

Naturally, deriving false positives and false negatives demands that I binarize the obtained probability according to some hard threshold - which I already deemed imprudent. The predicted probabilities are far more informative (Goldstone et al., 2010). The solution is a "Precision-Recall Curve". This circumvents the problem by generating the ratio between recall and precision at all possible thresholds. This ratio is then easily conveyed visually as a curve representing the tread-off between recall and precision and any possible thresholds (He and Garcia, 2008, 1278). Furthermore the curve itself can be summarized by the "Average Precision" (AP) score. This score denotes the weighted mean of the precision achieved at each possible threshold, where the weighting is the increase in recall from the previous threshold. Denoting recall R, precision P, and the different threshold as n the AP score can be calculated as follows:

$$AP = \sum_{n} (R_n - R_{n-1}) P_n \tag{4}$$

The AP score is a convenient summary and will serve to evaluate the uncertainty associated

[14] All mathematical definitions below are sourced from an article from He and Garcia (2008), but most of them can be found in any standard machine learning text book, such as Friedman et al. (2001).

with the framework's predictive capabilities. Unfortunately, it conveys less intuitive meaning than recall and precision disaggregated. Thus, using these highly related measures in cohort, we get the best of both worlds.

Having presented the backdrop against which the framework will be judged, the next subsection presents the derived results and evaluations.

3.5 The Evaluation

The Precision-Recall Curve is presented in Figure 1. As seen, there is a clear tread-off between recall and precision. Lower thresholds inhibits the framework against generating false positives, but at the expense of an increased number of false negatives. Increasing the threshold will yield less false negatives (more true positives), but also more false positives. Moving out of the x-axis, recall increases as precision drops. The substantial consequence is that we must accept some imprecision in regards to which specific cells will actually experience conflict. As we shall see shortly, the cells potentially misclassified as conflicts are not randomly spread out, but often clustered near actual conflict zones. Thus, the risk estimated might not be far of, even though no conflict manifested.

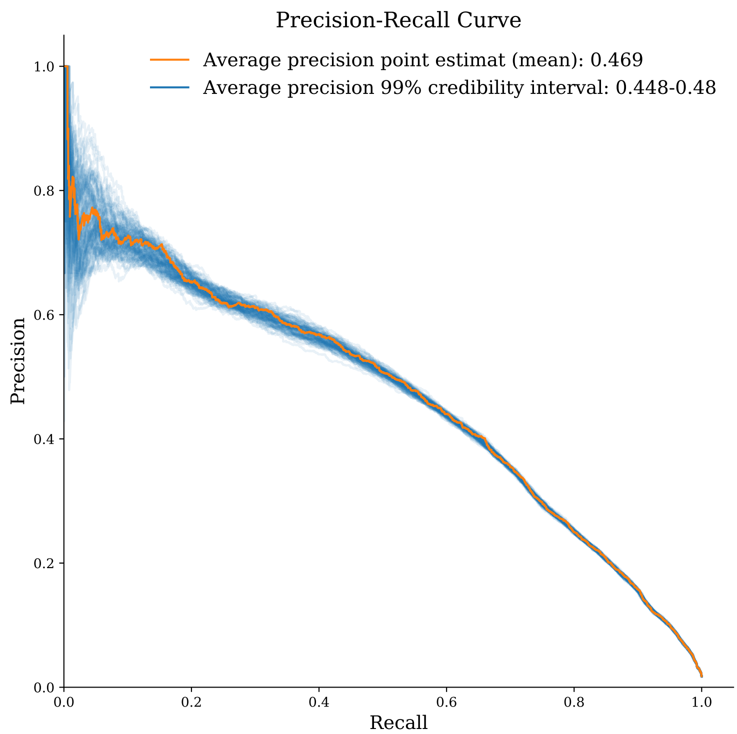

Figure 1: The credibility interval is plotted using only 100 random models for computational efficiency and no substantial changes comes from using all 1000. The reported credibility interval are based on all 1000 models. The point estimate (mean) reported and visualized are both derived on basis of all 1000 models

The same information presented in Figure 1 is conveyed in Figure 2, here merely at different specific marginal thresholds.

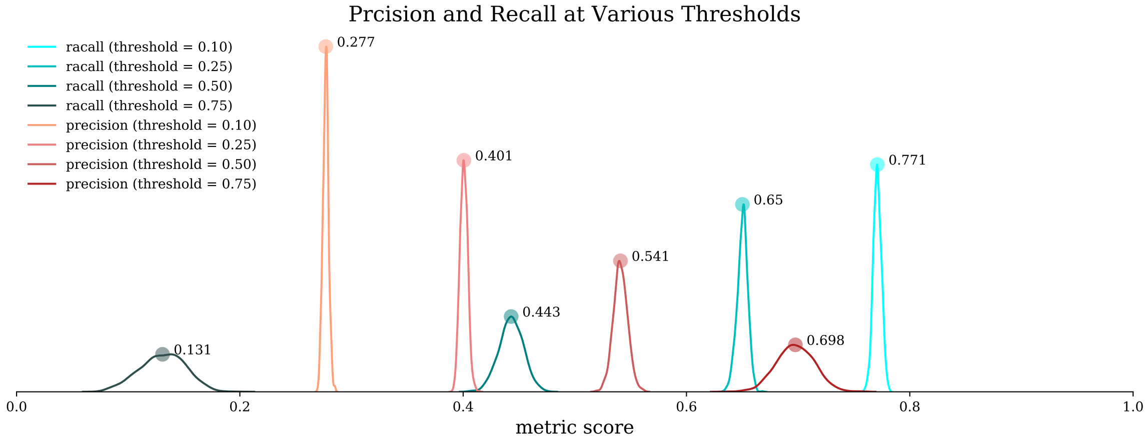

Figure 2: Marginal distributions of recall and precision at various thresholds. The point estimate is marked and annoted while the specific threshold is given in the legend. The distributions are based on all 1000 xgboost models. As these sampling distributions can be interpreted as probability distributions the height of these does not convey any other meaning then the certainty pertaining to the estimates.

Again, the inverse relationship and the increased variance with higher thresholds, is apparent. Furthermore, this plot gives us some substantial answers. For example; if we where to set a threshold denoting any grid cell with a probability of conflict above 10% as a potential conflict zone, we would correctly identify 77.1% of next years present conflicts. The trade-of is that these correctly identified events only constitute 27.7% of the events predicted to have above 10% risk of conflict. That is; our framework here generates around two false positives for each true positive.

Now, it is worth remembering that the grid cells in question are relatively small and does not correspond to any salient political, cultural or geographical distinction. It is merely squares. The point is, that predicting the exact cell is a hard case, and we can hardly hope to know exactly which of three neighbouring cells with high risk of conflict will actually experience conflict. And even if one experience conflict and the others do not, it does not mean that the risk was not real in all three cells. To illustrate my point Figure 3 visualizes all conflicts observed in 2007 versus the predicted probabilities for the same year in Figure 4.

The maps presented in Figure 3 and Figure 4 clearly illustrates the potential of the framework - but also some of its challenges. It is clear to see that the areas deemed most probable of experiencing conflict, were also generally where conflict ensued. Indeed the predicted risk zones encompasses larger areas than the actual conflict zones, but this merely captures the uncertainty inherent in the task.

This is the good news. More problematic are the systematic mistakes. Looking at the year 2007

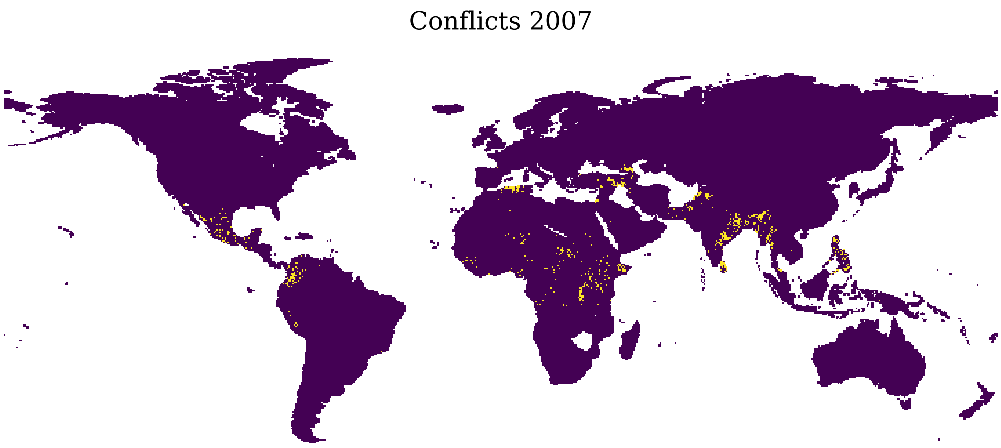

Figure 3: Conflicts observed in 2007 by UCDP aggregated at PRIO grid cell level. The measure is binary, with yellow denoting one or more conflicts in the given cell. Afghanistan, Iraq, Turkmenistan, Georgia and Zimbabwe are missing due to the coding rules of UCDP.

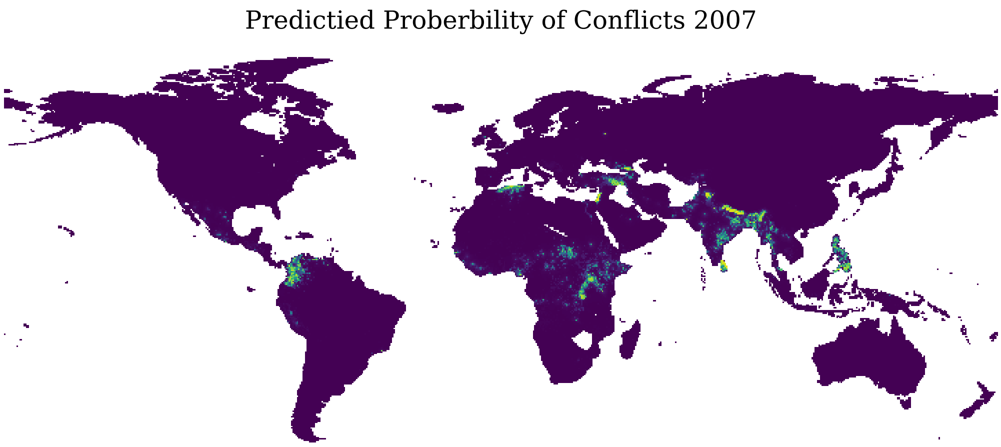

Figure 4: The estimate probability (prior corrected) of conflicts in 2007 using data from 2006 and the model trained on data from 1990 through 2005. The measure is between 0 and 1, where a 1 would denote certain conflict and be colored bright yellow as in Figure 3. Afghanistan, Iraq, Turkmenistan, Georgia and Zimbabwe are missing due to the coding rules of UCDP.

as an illustrative example, the high predicted probability of conflicts throughout Nepal constitutes a clear mistake; there was no conflict deaths in Nepal in 2007. Indeed, there where many conflict deaths from 1996 to 2006 as a result of the Nepalese Civil war (OHCHR, 2012), but non in 2007. The mistake is a product of how I constructed the framework. I ask of my framework to present the probability of conflicts given the features presented earlier and that is what it does. However, none of the included features are able to capture sudden peace agreements or ceasefires - which is exactly what happened on 21th of November 2006 in Nepal (OHCHR, 2012). The point is, that the framework is insensitive to sudden changes and developments in the political processes surrounding the conflicts and the sentiment among the warring parties. As I will discuss further down, this is crucial insight going forward towards a more accurate framework. However, before we come to that, the following section further explores why the framework is doing well in many instances, but also why it fails in e.g. the Nepalese case. This is done by analyzing which features the framework relies most heavily on generation its predictions.

3.6 Feature Importance

Looking at the top five most important features in Figure 5, four are directly related to present or past conflicts. Furthermore, the feature denoting distance to nearest conflict is by far the most influential predictor. More theoretical features did not contribute with much prediction power once the temporal/spatial patterns was accounted for. Intuitively this might not be surprising, yet the conventional literature does not devote overly much passion to the modelling the pattern of conflict dispersion through time and space. To be sure, the literature does sometimes include measures of temporal/spatial dependency and dispersion, but these measures are often crude and ad hoc - in which regard the present endeavour is no different. I shall return to this matter in the discussion of future improvements.

The question regarding Nepal is also neatly answered be the results in Figure 5. As mentioned above, the framework gets almost all of its predictive power through past and present conflict patterns. To capture political events such as peace treaties we would need to include an altogether different type of features; able of capturing sudden changes in the political development and sentiment in the given region or country. From this we can also deduce that the framework would be hard pressed, if it where to predict conflict onsets far removed in time or space from any other conflicts. Indeed a robustness test only including new conflict onsets, did perform markedly worse, but the hierarchy of feature importance did not change.

A proposal regarding handling of this issue is naturally presented in the discussion, to which I turn to now.

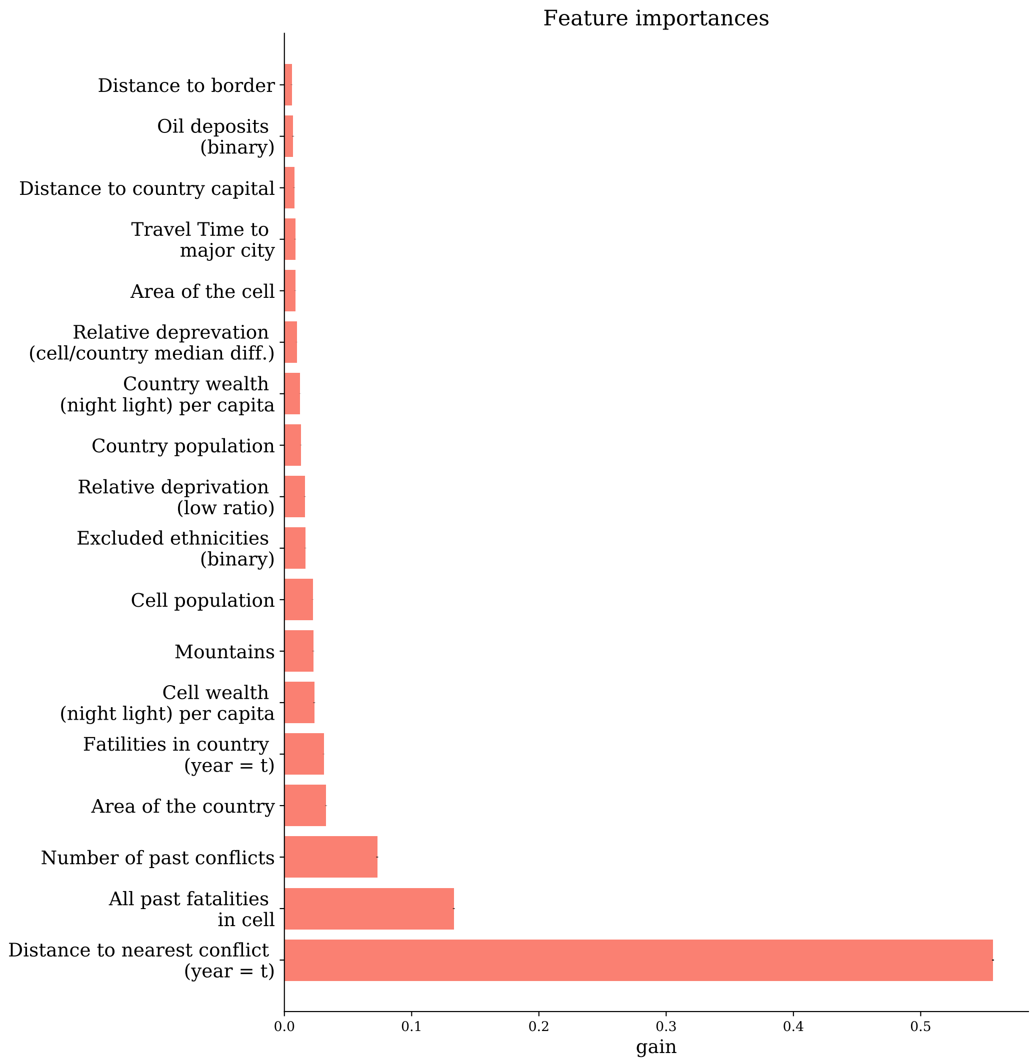

Figure 5: The relative importance of the included features given the use of the xgboost algorithm. Information from all 1000 models included. "Gain" denotes the internal optimization metric which the xgboost algorithm utilizes when choosing where best to split the generated regression tress.

4 Future Challenges and Improvements

Having assessed the predictive capabilities of the framework I find three obvious avenues for improvement: even better handling of the imbalanced data; a better, unified approach to the temporal/spatial inertia and dispersion of conflict; and the inclusion of features capturing more sudden political developments. These three avenues will be discussed in turn in the following subsections, after which I proceed to the summative conclusion.

4.1 Balancing the Future

An initial test showed that the framework did much better if the test data had been balanced in accordance with the train data. Naturally, such a test is only possible because we do have the actual conflict data of the test-set; which of course does not match any practical real world scenario. We could, however, take some precautions. We know from the proceeding section that the best predictors are related to the temporal and spatial proximity of past and present conflicts. Thus, if New Zealand where to experience conflict next year do to some very unforeseen development, the framework would not be able to capture it. With this in mind we could weed out all countries which have never experienced any conflicts in our data sample and which are also relatively far from any ongoing conflicts. This would reduce the number of non-events substantially, thus making the incoming data less imbalanced. Since we would not expect the framework to capture any conflict onset in the disregarded grid cells anyway, the trade-off might be negligible.

An other similar solution would be a "zero-inflated" framework with two or more rounds of estimations. The first round would correspond with the present endeavour. The second round would only include cells with a predicted probability over some appropriately low minimum threshold of conflict risk e.g. 1%. This is more computational burdensome than simply sorting by country, but it is also more systematical and potential more powerful.

4.2 A Unified Temporal-Spatial Approach

Concerning the temporal dimension, most past efforts have included some sort of fixed, random or conditional effects, but also more cleaver modeling of "conflict traps" and "conflict induced grievances" (Collier and Hoeffler, 2004; Hegre and Sambanis, 2006; Cederman et al., 2013; Perry, 2013). While these efforts are commendable, they all appear as ad hoc solutions. An example is Perry (2013) who propose "time deteriorating index" where one would include the number of fatalities for each of the last ten years as features. The trick is then to down-weigh the fatalities by dividing with the number of years past since the corresponding events. Thus, a feature pertaining to fatalities two years ago will have, as its values, half of the fatalities observed that year (Perry, 2013, 14). While the heart is at the right place, it is ad hoc and both methodologically and theoretically underdeveloped. There is no reason to cap the effort at ten years and there is no theoretical or practical reason to choose the suggested deterioration rate. Instead of dividing with years past it might be more appropriate to multiply by 0.8 or 0.4; or the deterioration rate might have an altogether different functional form, like an exponential decay function. The point is that we do not know, and therefor we should estimate the function. Also, estimation at least gives us an indication of how bad our guess is.

Intriguingly, estimation would also allow different regions, countries or cells to have different rates of conflict deterioration. I shall not go into the technical details here, but the machine learning technique of Gaussian Processes have been used to similar purpose with good results before (Gelman et al., 2013; McElreath, 2018). Furthermore Gaussian Processes also facilitates forecasting more readily, without the need to construct specific lags or leads in various features. In such a framework one could decide, post-modeling, whether one wanted estimates for next year, the year after or ten years ahead. Even more, the procedure also estimates the uncertainty which inherently grows as we look into the future.

Turning to the spatial aspect, I conveniently find that the same tools also have shown very good results in regards to spatial dispersion, where both regional and country[15] specific variation again is readily incorporated (Gelfand et al., 2003; Gelfand, 2012; Gelfand and Schliep, 2016). The fact that both the temporal and spatial patterns have potential of being modelled according to a unified approach is both practical and encouraging.

To clarify, I do *not* propose to swap the xgboost algorithm, used above, with a new machine learning algorithm; I am suggesting that one could use the proposed techniques to create better features pertaining to the temporal and spatial patterns of conflict. Features which then could be fed some framework utilizing xgboost.

A less technical, but just as important point, is that modelling conflicts as a spatial/temporal function of conflict itself only requires the UCDP[16] conflict data. As motioned in the beginning of this endeavour the UCPD data is substantially more up-to-date compared to the PRIO grid data. And indeed structural data such as that presented in the PRIO database are rarely available for the most resent years whatever the source. Thus, having learned that our model gets most of its predictive power from the UCPD itself, it seems only natural that any future effort to explore the potential in predicting conflict, considers using this database as the primary feature-source.

However, while I do believe such approaches holds great promise for future endeavours, they would not amend the challenge exemplified by Nepal 2007. The next subsection will present a potential solution to this shortcoming.

4.3 Large Scale Automated Text Analysis

The most simple and straight forward way to know whether some ceasefire is underway, is to know whether or not the warring parties are discussing a ceasefire. And furthermore whether or not the public and political sentiment appears to favour a potential ceasefire. Tools and approaches for large scale automated text analysis gets evermore accessible and powerful and specifically

[15] Or indeed any other geographical unit

[16] Or any other relevant data source on conflicts

automated text analysis utilized on online news sources have already shown promise in tasks of conflict prediction (Chadefaux, 2014; Mueller and Rauh, 2016).

Thus, one could enhance the feature-space with information pertaining to the political or public sentiment in various countries or regions by scraping large amounts of text data from social media, news sources or political outlets. Not only would such features potentially amend the problem exemplified by Nepal 2007; more importantly it could help in predicting onsets in otherwise peacefully areas (Mueller and Rauh, 2016). As such, a framework incorporating analysis of text would bring valuable new information to the prediction effort.

The tread-of is naturally that such an effort is vastly more resource demanding than the two other remedies proposed. Never-the-less, Mueller and Rauh (2016) show one way to do it, and with the pace of the present technological development it is surely not an impossible task.

This concludes the discussion regarding future improvements of the framework. The following section will summarize the endeavors main points and conclude on the project as a whole.

5 Conclusion

The present endeavour was structured around three research questions, which have all been answered in turn. To reiterate:

Q_1: Using a collection of modern machine learning tools I have explored to what extent is it possible to predict the geographic location of future intra-state conflicts. This was a preliminary endeavour, yet the results are quite encouraging. Indeed the framework does seem to successfully identify a larger proportion of high risk areas correctly. However, this is not without some imprecision. Thus, if we where to classify all predictions with a risk of conflict above 10% as potential conflicts, we would correctly identify almost 80% of all conflicts. However, given the imprecision of the framework, we would also produce two false positives for each true positive. However, What is entirely more problematic is that the framework proves unable to capture sudden political events such as peace agreements or conflicts onsets far removed in time or space from any other conflicts. This leads us to the conclusion pertaining the second research question.

Q_2: The framework gets most of its predictive capabilities through features pertaining to the temporal and spatial patterns of past and presents conflicts. These features proved remarkable reliable in the prediction effort. But naturally a framework mainly relying on these dimensions will not be able to incorporate the effects of sudden political developments as described above.

This leads us to the question of future improvements.

Q_3**:** Three proposals where given regarding future improvements. The first involves a better handling of the imbalanced nature of the data by disregarding countries which are deemed to have a very low probability of conflicts pre-estimation. Initial tests indicated that this could lower the trade-of between false negatives and false positives substantially.

The second proposed improvement involves creating a more unified framework for creating features pertaining to the temporal and spatial patterns of conflicts utilizing only the conflict data. Since these where the most informative dimensions and also since this data was more topical, resources should be prioritized developing this avenue.

While these two improvements would no doubt enhance the framework, they would not effectively handle the problem posed by sudden political developments. This, however, could be handled by scrapping of country or region specific news sources or political outlets. This data could be analyzed using large scaled automated text analysis. While this would be a powerful addition to the framework, it is also vastly more resource demanding then the other proposed improvements.

With these recommendations regarding future endeavours, and a confident - if cautious - optimism regarding the prospect of a reliable early-warning-system the present project is concluded.

6 Bibliography

References

Blattman, C. and Miguel, E. (2010). Civil war. *Journal of Economic Literature*, 48(1):3–57.

Blimes, R. J. (2006). The indirect effect of ethnic heterogeneity on the likelihood of civil war onset. *Journal of Conflict Resolution*, 50(4):536–547.

Blyth, S., Groombridge, B., Lysenko, I., Miles, L., and Newton, A. (2002). Mountain watch: environmental change & sustainable development in mountains. *Cambridge, UK: UNEP World Conservation Monitoring Centre*.

Buhaug, H. (2010). Dude, where's my conflict?: Lsg, relative strength, and the location of civil war. *Conflict Management and Peace Science*, 27(2):107–128.

Buhaug, H., Gates, S., and Lujala, P. (2009). Geography, rebel capability, and the duration of civil conflict. *Journal of Conflict Resolution*, 53(4):544–569.

Cederman, L.-E., Buhaug, H., and Rød, J. K. (2009). Ethno-nationalist dyads and civil war: A gis-based analysis. *Journal of Conflict Resolution*, 53(4):496–525.

Cederman, L.-E. and Gleditsch, K. S. (2009). Introduction to special issue on "disaggregating civil war". *Journal of Conflict Resolution*, 53(4):487–495.

Cederman, L.-E., Gleditsch, K. S., and Buhaug, H. (2013). *Inequality, grievances, and civil war*. Cambridge University Press.

Cederman, L.-E., Weidmann, N. B., and Gleditsch, K. S. (2011). Horizontal inequalities and ethnonationalist civil war: A global comparison. *American Political Science Review*, 105(3):478–495.

Chadefaux, T. (2014). Early warning signals for war in the news. *Journal of Peace Research*, 51(1):5–18.

Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794. ACM.

Chen, X. and Nordhaus, W. D. (2011). Using luminosity data as a proxy for economic statistics. *Proceedings of the National Academy of Sciences*, 108(21):8589–8594.

Collier, P. and Hoeffler, A. (1998). On economic causes of civil war. *Oxford Economic Papers*, 50(4):563–573.

Collier, P. and Hoeffler, A. (2004). Greed and grievance in civil war. *Oxford Economic Papers*, 56(4):563–595.

Croicu, M. and Sundberg, R. (2017). Ucdp ged codebook version 18.1. *Department of Peace and Conflict Research, Uppsala University.*

Diamond, J. M. (1998). *Guns, germs and steel: a short history of everybody for the last 13,000 years.* Random House.

Elvidge, C. D., Hsu, F.-C., Baugh, K. E., and Ghosh, T. (2014). National trends in satellite-observed lighting. *Global urban monitoring and assessment through earth observation*, 23:97–118.

Elvidge, C. D., Sutton, P. C., Ghosh, T., Tuttle, B. T., Baugh, K. E., Bhaduri, B., and Bright, E. (2009). A global poverty map derived from satellite data. *Computers & Geosciences*, 35(8):1652 – 1660.

Fearon, J. D. (2004). Why do some civil wars last so much longer than others? *Journal of Peace Research*, 41(3):275–301.

Fearon, J. D. (2005). Primary commodity exports and civil war. *Journal of Conflict Resolution*, 49(4):483–507.

Fearon, J. D. and Laitin, D. D. (2003). Ethnicity, insurgency, and civil war. *American Political Science Review*, 97(1):75–90.

Friedman, J., Hastie, T., and Tibshirani, R. (2001). *The elements of statistical learning.* Springer series in statistics New York, NY, USA:.

Gartzke, E. (1999). War is in the error term. *International Organization*, 53(3):567–587.

Gelfand, A. E. (2012). Hierarchical modeling for spatial data problems. *Spatial statistics*, 1:30–39.

Gelfand, A. E., Kim, H.-J., Sirmans, C., and Banerjee, S. (2003). Spatial modeling with spatially varying coefficient processes. *Journal of the American Statistical Association*, 98(462):387–396.

Gelfand, A. E. and Schliep, E. M. (2016). Spatial statistics and gaussian processes: A beautiful marriage. *Spatial Statistics*, 18:86–104.

Gelman, A., Stern, H. S., Carlin, J. B., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian data analysis - third edition.* Chapman and Hall/CRC.

Girardin, L., Hunziker, P., Cederman, L.-E., Bormann, N.-C., and Vogt, M. (2015). Growup-geographical research on war, unified platform. *ETH Zurich. http://growup. ethz. ch.*

Goldstone, J. A., Bates, R. H., Epstein, D. L., Gurr, T. R., Lustik, M. B., Marshall, M. G., Ulfelder, J., and Woodward, M. (2010). A global model for forecasting political instability. *American Journal of Political Science*, 54(1):190–208.

Gurr, T. R. (1970). *Why men rebel.* Routledge.

Gurr, T. R. (1995). Minorities at risk- a global view of ethnopolitical conflicts. *UNITED STATES INSTITUTE OF PEACE PRESS, ARLINGTON, VA 22210(USA). 1995.*

He, H. and Garcia, E. A. (2008). Learning from imbalanced data. *IEEE Transactions on Knowledge & Data Engineering*, (9):1263–1284.

Hegre, H. and Sambanis, N. (2006). Sensitivity analysis of empirical results on civil war onset. *Journal of Conflict Resolution*, 50(4):508–535.

Hegre, H., Østby, G., and Raleigh, C. (2009). Poverty and civil war events: A disaggregated study of liberia. *Journal of Conflict Resolution*, 53(4):598–623.

Hobbes, T. (1991[1651]). *Leviathan. Cambridge texts in the history of political thought.* New York: Cambridge University Press. edited by Richard Tuck.

Kalyvas, S. N. (2007). Civil wars. In Boix, C. and Stokes, S. C., editors, *The Oxford handbook of comparative politics*, volume 4, chapter 18, pages 417–434. Oxford Handbooks of Political.

King, G. and Zeng, L. (2001a). Explaining rare events in international relations. *International Organization*, 55(3):693–715.

King, G. and Zeng, L. (2001b). Improving forecasts of state failure. *World Politics*, 53(4):623–658.

Lipset, S. M. (1959). Some social requisites of democracy: Economic development and political legitimacy. *American political science review*, 53(1):69–105.

McElreath, R. (2018). *Statistical Rethinking: A Bayesian Course with Examples in R and Stan.* CRC Press.

Mueller, H. F. and Rauh, C. (2016). Reading between the lines: Prediction of political violence using newspaper text. *American Political Science Review*, 2(112):358–375.

Nordhaus, W. D. (2006). Geography and macroeconomics: New data and new findings. *Proceedings of the National Academy of Sciences*, 103(10):3510–3517.

OHCHR, U. (2012). Nepal conflict report.

Perry, C. (2013). Machine learning and conflict prediction: a use case. *Stability: International Journal of Security & Development*, 56(2(3)).

Ross, M. L. (2004). What do we know about natural resources and civil war? *Journal of Peace Research*, 41(3):337–356.

Schrodt, P. A. (2014). Seven deadly sins of contemporary quantitative political analysis. *Journal of Peace Research*, 51(2):287–300.

Skocpol, T. (1979). *States and social revolutions: A comparative analysis of France, Russia and China*. Cambridge University Press.

Sundberg, R. and Melander, E. (2013). Introducing the ucdp georeferenced event dataset. *Journal of Peace Research*, 50(4):523–532.

Tollefsen, A. F., Bahgat, K., Nordkvelle, J., and Buhaug, H. (2016). Prio-grid codebook v.2.0. *Oslo: PRIO*.

Tollefsen, A. F., Strand, H., and Buhaug, H. (2012). Prio-grid: A unified spatial data structure. *Journal of Peace Research*, 49(2):363–374.

Tollefsen, Andreas Forø, K. B. J. N. and Buhaug, H. (2015). Prio-grid v.2.0 codebook. *Oslo*.

UCDP (2017). Ucdp georeferenced event dataset (ged) global version 18.1. http://www.ucdp.uu.se/downloads/. Accessed: 2018-11-13.

Vogt, M., Bormann, N.-C., Rüegger, S., Cederman, L.-E., Hunziker, P., and Girardin, L. (2015). Integrating data on ethnicity, geography, and conflict: The ethnic power relations data set family. *Journal of Conflict Resolution*, 59(7):1327–1342.

Ward, M. D., Greenhill, B. D., and Bakke, K. M. (2010). The perils of policy by p-value: Predicting civil conflicts. *Journal of Peace Research*, 47(4):363–375.

7 Appendices

All python scripts used for the present endeavour are attach as HTML files or can be found on: https://github.com/Polichinel/Conflict_Prediction.

The scripts are embedded in jupyter notebooks, thus they are accompanied by plot and comments. The scripts are: interpolation.html, Feature_eng.html, data_exploration.html, div.html, C_shapes_lookup.html, boosting.html and evaluation.html.

7.1 The Data Sources

The project at hand utilize two different data source. The he Uppsala Conflict Data Program (UCDP) (Sundberg and Melander, 2013; Croicu and Sundberg, 2017) and the PRIO grid(Tollefsen et al., 2012). The first subsection below presents the UCDP while the follow presents the PRIO grid database.

7.1.1 UCDP

Most central to the endeavour at hand lies - of course - the data regarding intra-state conflict itself. This data is obtained through the Uppsala Conflict Data Program (UCDP) (Sundberg and Melander, 2013; Croicu and Sundberg, 2017). Specifically I utilize the UCDP Georeferenced Event Dataset (GED) Global version 18.1 (UCDP, 2017). The dataset contains records of conflict fatalities and the corresponding coordinates. As mentions I utilized data from 1990 through 2010 but data from 1989 through 2017 is available in the database[17]. Conflict fatalities are here defined as:

> *"An incident where armed force was used by an organized actor against another organized actor, or against civilians, resulting in at least 1 direct death at a specific location and a specific date."*(Croicu and Sundberg, 2017, 38).

Further definitions regarding armed force, organized actor ect. can be found in (Croicu and Sundberg, 2017, 10-11).

The project at hand limits itself to intra-state conflict, thus I only included incidents which do *not* include two different nations as the organized actors[18]. This data source provides the

[17]For a smaller less detailed feature space data is available going back through 1975

[18]Naturally the distinction between a intra-state conflict and a proxy war can be very hard to uphold in practise. Never-the-less, though proxy wars are effectively between two states, the phenomenon is arguable more similar

prediction target; the presence of conflict deaths in some geographic location in some future point in time. The data from UCDP is very detailed in regards to both temporal and spatial location, however for the endeavour at hand I aggregate the time unites to 'years' and the spatial resolution to the 0.5×0.5 decimal degrees grid cells provided by the PRIO grid. The target is further 'leaded' such that I am predicting conflicts one year into the future. Thus the prediction target is a binary feature denoting whether or not a given PRIO grid cell will host one or more conflict incidence the following year.

As I will show later on, many of the most important predictive features are also derive from this data source; e.g. the distance from a specific cell to the nearest conflict and the number of previous conflicts in a given cell. The fact that conflict patterns do them selves hold a lot of prediction power regarding future conflict zones might not surprise the reader, but I find the implications of this insight rather consequential in regards to future improvement of the framework, which I shall return to in the discussion.

7.1.2 PRIO

Given the geo-referenced nature of the UCDP data, the number of interesting data sources one could enhance it with are virtually endless. However, the aggregation of various geo-referenced and geo-spatial data accompanied with the appropriate grid construction and feature engineering can be a time consuming endeavour. While it is certainly a interesting and most likely fruitfully undertaking, it is one which will be saved for less preliminary work then the project at hand. Conveniently the Peace Research Institute Oslo (PRIO) has created what they call a "unified spatial data structure" (Tollefsen et al., 2012, 1). More specifically they have divided the world - excluding Greenland and Antarctica - into grid cells of 0.5×0.5 decimal degrees. For each cell PRIO has gather a large selection of features, including economic, geographic and demographic features (Tollefsen et al., 2012). Naturally the PRIO grid also extent itself across time, and the most current data includes cell-years from 1946 to 2015 - with some divergence in the data coverage across the years (Tollefsen et al., 2016). On one side more complete data is available for more recent years; especially after 1990. On the other side no or little data is avialbel for the most recent years. Thus I utilize data from 1990 through 2010. This data includes relatively few random missing observations which are mainly handled through forward interpolation if possible, else backward interpolation or nearest observation. Some variables are only account for in 5-years intervals. This is handled through cell-specific linear interpolation as illustrated and described in the script *interpolation.html*.

The PRIO grid is constructed as geo-spatial data and primed for collaboration with the UCDP data. As such merging and handling these two data source is a trivial task. For now I refrain from

to intra-state conflicts and civil wars then to all out open warfare between to nations the temptation to included all variables available through the PRIO grid data base, and instead chose handpick features which are most often and constantly associated with internal conflict in the corresponding litterateur. There are number of reasons for this choice, but mainly it is to keep the endeavour concise, facilitate convertibility and due to the fact that initial tests indicated that few theoretically sound features outperformed the inclusion of many crude features. The following section will present the included features and also discuss some notable absentees.

7.2 The Included Features - Expanded

Below I give a spend some more ink on the theoretical backdrop support many of the included features. I also included some mathematical definitions where needed.

7.2.1 Wealth and State Capacities

Easily one of the most robust findings in country level studies of civil wars is GDP per capita[19] has a negative effect on the probability of civil war onset (Collier and Hoeffler, 1998; Fearon and Laitin, 2003; Collier and Hoeffler, 2004; Hegre and Sambanis, 2006; Blattman and Miguel, 2010) and also to some extent the conflict duration(Fearon, 2004; Hegre et al., 2009).

A number of mechanisms have been proposed linking GDP to conflict, Two have been especial prolific. The first is championed by Collier and Hoeffler (1998, 2004) sees GDP per capita as a proxy for opportunity-cost. that is what i given citizen have to loss by engaging in conflict. The second story draws on some insight from Skocpol (1979) and also echoes the gospel of modernization theory (Lipset, 1959). Regarding the context at hand it has most prominently been presented by Fearon and Laitin (2003). Here GDP per capita is seen as as proxy for state capacities. Simple put; weak or fragile states have low GDP per capita and these states are more conflict prone(Fearon and Laitin, 2003, 88).

A measure for GCP (Gross cell product) per capita (ppp) is included in the PRIO GRID from the Gecon dataset (Nordhaus, 2006) and this measure could be aggregated creating a feature for GDP per capita (ppp) (Tollefsen and Buhaug, 2015). However the data available from the directly from the PRIO GRID only extent to 2010. Fortunately Elvidge et al. (2009), Chen and Nordhaus (2011) have shown that Night light emission can serve as an proxy for economic activities - especially for countries of areas with low-quality statistical systems and few or no recent population or economic censuses. (Chen and Nordhaus, 2011) - an approach explicitly proposed by (Cederman et al., 2013, p. 101) in regards to conflict studies. A ad hoc illustration of the high correlation between the two features can also be found in the script *data_exploration.html.*

[19]Often logged and adjusted for purchasing power parity (ppp)

The measure of Nigth Light Emmision in the PRIO GRID is borrowed from Elvidge et al. (2014)it extent all the way up to 2015 and calibrated as to better accommodate time series Tollefsen and Buhaug (2015).

Just as the case with GDP ir GCP, it makes intuativly sens to correct the measure *per capita*.

- $\frac{Cell-yearspecificNightLightEmission(mean)}{Cellpopulation}$
- $\frac{Country-yearspecificNightLightEmission(aggregatedmean)}{Countrypopulation}$

Logged, density and measures not corrected for population size was also tired, but I found these measures less effective.

Naturrally one could argue that wealth is a relative concept, which leads us to the next section.

7.2.2 Inequality and Deprivation

If we - for the time being - leave the Strong State proposition behind and focus on the satisfaction of the individual citizens it is only natural to argues this satisfaction should be considered a function of *what we have* and *what we believe we rightfully should have*. This, indeed, is the crux of Robert Gurr's (1970) Relative Deprivation Theory. Perhaps one of the most seminal[20] takes on inequality and conflict, Gurr (1970) defines relative deprivation:

> *"[...] as actor's perception of discrepancy between their value expectation and their value capabilities. Values expectation are the goods and conditions of life to which people believe they are rightfully entitled. Value capabilities are the goods and conditions they think they are capable of getting and keeping."* (Gurr, 1970, 24).

While intuitively appalling, Gurr's theory was award little credit doing the haydays of comparative corss country conflict studies. Supporting statistical results failed to materialize, and the axplanation eccoed the of (Skocpol, 1979, 11); injustice and misery is simply to widespread to account for the rarity of major conflicts (Collier and Hoeffler, 1998, p. 22, Collier and Hoeffler, 2004, p. 22 Fearon and Laitin, 2003, p. 44(FLERE?)). (så lige styr på de 22 side tal der?)

However, Cederman and Gleditsch (2009); Cederman et al. (2013) have noted that the aggregated country level features conventional used as indicators for inequality might lead to misspecifications; that is, they do not properly measure the theoretical concept of relative deprivation or the correct mechanisms through which inequality affects conflict-propensity (Cederman et al., 2013, XX). Acknowledging this critique I utilized the operationalization put forth in (Cederman

[20]At least after Marx

et al., 2013, p. 104-105)[21]:

$$y_g = \text{country year mean} \quad , \quad y_c = \text{cell year value}$$

$$\text{low\_ratio} = y_c/y_g \quad \text{if} \quad y_c < y_g, \quad 1 \quad \text{otherwise}$$

Thus cells which are relatively well of compared to the mean of country takes the value 1. Cells worse of the the country mean takes a value above 1, with the magnitude of this value indicating *how* severe the deprivatione is. Cederman et al. (2013) Uses GCP per capita (ppp) but as mentioned also suggests using night light emission, Thus I here produce ratio feature solely on Night Light emission (cell-year mean).

While I do appreciate this operationalization I also construct my own relative deprivation feature, which differs in a number of small but relevant ways. First we know that wealth distributions in general are highly skewed, and this is no different when we use Nigt Light Emission as indicator. Thus, given Gurr's original conceptualization I find it more realistic that individuals should compare themselves to the median - not the mean. That is citizens in one cell compare their living standard to the most common living standard in the country over all. In the same vein, instead of a fraction, I simply calculate the difference between the cell-year values and the country-year median. Lastly I let the value start at 0. This I do to insure that interactions create later only "activates" if the cell is actually depraved. In the case of Cederman et al. (2013) when a cell is not depraved, the value of a interaction takes the naked value of the other interacted feature, which is arguably somewhat imprudent. Mathematical the features is formulated as such:

$$y_g = \text{country year median} \quad , \quad y_c = \text{cell year value}$$

$$\text{low\_diff\_median} = y_g - y_c \quad \text{if} \quad y_c < y_g, \quad 0 \quad \text{otherwise}$$

Thus the feature takes the value 0 if the cell is on level or above with the rest of the country. If the cell is deprived the value correspond to the difference between the country-year median and the cell value. Naturally - and as we I shall return to later - this measure and the measure by Cederman et al. (2011) are highly correlated, nut that does not change the fact that this last feature appears somewhat closer to the notion of relative deprivation, and more impotently it

[21] These scholars alos argues the higher conflict propensity might be fund in the other end of the inequality spectrum. That is; one could imagine a mechanism akin to relative deprivation in one end, while simultaneous at the other end, one might see well-off people wanting to secede from or take over a country if they fear to much redistribution. However, they find little statistical backing for this proposition, neither did my initial tests. Thus I only use the measures corresponding to actual Relative Deprivation

handles interactions somewhat more appropriate or at least conventional. As before the feature uses on Night Light emission (cell-year mean) as indicator for wealth.

7.2.3 Ethnicity and Exclusion

Denotes the number of excluded ethnic groups (e.i. discriminated or powerless) in a given cell at a given year. The measures are originally from the GeoEPR/EPR data by Vogt et al. (2015). To better suit the theoretical argumentation laid forth in Cederman et al. (2013)(side) (Tollefsen and Buhaug, 2015). I create a dummy (excluded_binary) which simply denote *if* there are any excluded ethnic groups.

As with inequality the link between ethnic diverse societies and conflict propensity have been ridden with disagreement and controversies. In the quantitative literature results have remained somewhat inconsistent (Blattman and Miguel, 2010, 23-24). A number of studies have fund different - and sometime quite convoluted - relationships between ethnicity or discrimination and conflict (Collier and Hoeffler, 1998; Fearon, 2004; Blimes, 2006; Hegre and Sambanis, 2006; Goldstone et al., 2010), while other studies have fund little or no trace of the connection (Fearon and Laitin, 2003; Collier and Hoeffler, 2004)(more? CH 2004 right?).

As with the problem with inequality the lack of discernible results have often been attributed to poor feature specifications, a framework not capturing the proposed theoretical mechanism and not least country aggregated data (Blimes, 2006; Blattman and Miguel, 2010; Cederman et al., 2013). Mirroring they effort concerning inequality Cederman et al. (2013) have recently made use of new desegregated data (Girardin et al., 2015) to closer model the theoretical mechanism proposed. Without going to much in-depth here the feature Cederman et al. (2013) utilizes aims to capture the effect of *horizontal inequalities* as defined by (Cederman et al., 2013, 31-35). That is the systematic discrimination or political exclusion of and coherent ethnic group. Thou not framed in the theoretical context of horizontal inequalities Goldstone et al. (2010) finds results supporting the argument using the Minority at Risk data from Gurr (1995).

Conveniently, a measure from Girardin et al. (2015) of how many excluded ethnic groups reside in each PRIO GRID cell is readily aviable in the PRIO GRID. To mimic the theoretical proposition lead foth in Cederman et al. (2013) I binarize this variable to simply indicate whether or not a given cell is inhabited by a politicly excluded or discriminate ethnic group[22].

Not surprisingly Cederman et al. (2013) futher finds that the properbilty of conflict gets even larger if political exclusion is followed by group deprivation, but the xgboost algorithm constructs it own interaction so I shall not be creating the interaction variable my self here.

[22]Also initial test did not show much prospect of incorporating the full count

7.2.4 Population size and density

Returning to a variable with a rather flawless record regarding conflict onset we have "country population size" (Collier and Hoeffler, 1998; Fearon and Laitin, 2003; Collier and Hoeffler, 2004; Hegre and Sambanis, 2006)[23]. Furthermore (Fearon, 2004, 287) also finds that country population is correlated with longer civil wars.

Lastly when it comes to disaggregated grid data, the "grid population size" also seems a rather robust predictor (Buhaug, 2010; Cederman et al., 2013)[24]. Thus, whether the features influence conflict duration is more disputed (Collier and Hoeffler, 1998; Fearon, 2004). The most common implementation is to use a log transformed version of the country or grid population count[25] which is also implemented in the project at hand.

One very simple explanation might be the various definitions of conflict and civil war used throughout the litterature. These definitions almost always refer to some minimum fatalities count [eks and refs]. Naturally high population counts makes these threshold relativly less restrictive.

There are however also more theoretical propositions for the relationship. One being that conflict mediation becomes inherently more difficult as social systems and societies grow (Diamond, 1998, p- 271-272).

Initial I include features for cell-year population, aggregated country-year population and corresponding population densities:

$$\text{grid\_pop\_dens} = \frac{\text{grid\_pop}}{\text{grid\_area}}$$

$$\text{country\_pop\_dens} = \frac{\text{country\_pop}}{\text{country\_area}}$$

However, I find no gain to be made by including either the density or the logged count. Thus the raw count are utilized in the effort at hand.

[23]though see Goldstone et al. (2010)

[24]Though see Hegre et al. (2009)

[25]Though Collier and Hoeffler (1998) use the plain count in 10.000's

7.2.5 Geography and Accessibility

Accessibility As noted earlier, a strong state have often been presented as the prime inhibitor of internal conflicts. Naturally though, the strength of state must be considered relative to the territory over which it claims sovereignty - both in regards to size and permeability. Fearon and Laitin (2003) pointed to rough terrain and mountains as natural obstacles hindering effective projection of state power. Following this example Hegre and Sambanis (2006) concludes that a feature for rough terrain is found to robustly positively correlated with civil war across a large number of model specifications.(Hegre and Sambanis, 2006, 526-529)[26]. The Prio Grid includes a readily available feature measuring the proportion of mountainous terrain within the cell based on Blyth et al. (2002) wich I utilize.

Another natural hindering for projecting state power is sheer distance (Fearon, 2004; Buhaug et al., 2009; Cederman et al., 2009; Buhaug, 2010).

> *"The projection of power across distance comes at a cost. [...] In particular, large hinterlands and isolated peripheries are favorable to insurgency. In sum, this suggests that large countries are relatively more exposed to intrastate conflict"*(Buhaug, 2010, 113-114)

A number of interesting features could be derived from Buhuag's assertion above (and the paper in general). What I have included is the distance to the nations capital[27], the travel time to the nearest major city, and the total size of the countryTollefsen and Buhaug (2015).

7.2.6 Trans-boarder Influences

A number of different mechanism have been proposed and explored (Blattman and Miguel, 2010, 29-30), the one explored here is rather simple and follows Hegre and Sambanis (2006) I simply incorporate a measure for the distance to the nearest (land) border. I trued with the logged distance, but this yeilded no improvements. This is a very rough measure and is arguably a bit far removed from any specific theoretical concept, but for this preliminary project the measures will have to do.

[26]Though see Goldstone et al. (2010)

[27]The measure utilized by Buhaug (2010)

7.2.7 Prime Commodities and the Recourse course

A lot have been written on this subject, but the empirical results have varied a lot (Collier and Hoeffler, 1998; Fearon and Laitin, 2003; Fearon, 2004; Ross, 2004; Collier and Hoeffler, 2004; Fearon, 2005; Buhaug, 2010; Hegre et al., 2009). In this endeavour, I include a dichotomous feature denoting whether or not a given cell is known to hold oil deposits. Thus, a cell is not denoted as having oil before oil is actually discovered.

7.2.8 Inertia, dispersion, traps and time trends

Conflict traps and inertia have been modelled or proposed by many. Collier and Hoeffler (2004) uses a linear decay since last conflict; Hegre and Sambanis (2006) uses a linear decay since last peace; Cederman et al. (2013) count the number of previous conflicts; and Perry (2013) propose a "deterioration index". Meanwhile, cross-country dispersion has been used by Goldstone et al. (2010). I include a number of features which aims to capture the pattern of conflict as it moves through space and time. Distance from cell center to nearest conflict, yearly fatalities in cell, yearly fatalities in the country as a whole and the total number of years the given cell has been a conflict zone in the past and .

7.2.9 Notable Absentees

Notable missing include infant mortality, political system and intra elite conflict (Goldstone et al., 2010). These should be tested in feature endeavours, but was not included here due to the limited scope of the project. The PRIO dataset did included a measure of infant mortality but it was ridden with missing values. Regarding the two other dimension, these would require the introduction of a third data source.

7.3 Bayesian Prior Correction

A small issue with the undersampling procedure above is that the probabilities predicted is somewhat inflated. This does not as such hurt the prediction effort. This might seem counter intuitively, but since the probability threshold beyond which we denote a observation to be a conflict can be chosen at will, it is the ratio between such threshold and the probabilities that. Not one or the other. To illustrate: normally one would set a threshold at 50%. If the probability of conflict is above 50% we predict a conflict and if it is below we do not predict a conflict. This threshold, while conventional, is entirely arbitrary and for some purposes it might well make more sense to set the threshold at 20% or 80%. Thus knowing that my estimates are inflated do to undersampling

I could just easily set my threshold a bit higher then normal, say at 60 or even 80%. However, I aim to not implement any hard threshold but instead presents the actual probabilities and their certainty for reader to survey. As such, for these probabilities to be intuitively meaningful they need to match what we actually know regarding the propensity of conflict events. This is easily implemented with an Bayesian prior correction[28]. Denoting the the estimated probabilities of an event $Pr(E_{estimated})$, estimated probabilities of a non-event $Pr(NE_{estimated})$ the and the corrected probabilities of events as $Pr(E_{corrected})$ the correction is as follows:

$$Pr(E_{corrected}) = \frac{Pr(E_{estimated}) \times 0.2}{(Pr(E_{estimated}) \times 0.2) + (Pr(NE_{estimated} \times 0.8)} \tag{1}$$

Where 0.2 is the prior, which is here set as the factor between the under sampled ratio of events of non-events and the actual ratio between events and non events. 0.8 is simply $1 - prior$. The correction is a very crude utilization of Bayes theorem (Gelman et al., 2013, 7-8), and could be much improved. Even more it is only semi-Bayesian since I use hard priors - scalars - instead of some appropriate distribution. A fruitful approach might also be to create cell, country or region specific priors possible in some hierarchical setting. This however is a challenge for an other endeavour. For now the crude correction will do - and as sown in Figure 6, Figure 7 and Figure 8 it does rather well. The point is that the non-corrected probabilities depicts a overly dangerous world. This can also be seen be setting an arbitrary threshold at 0.5 for classification. With the non-corrected probabilities this leads to a precision around 0.12 and a recall above 0.9, while the corrected probabilities splits it almost even. This proofs nothing, but it is a nice indication of more realistic estimates.

[28]Related to what is presented in King and Zeng (2001b,a) and Goldstone et al. (2010)

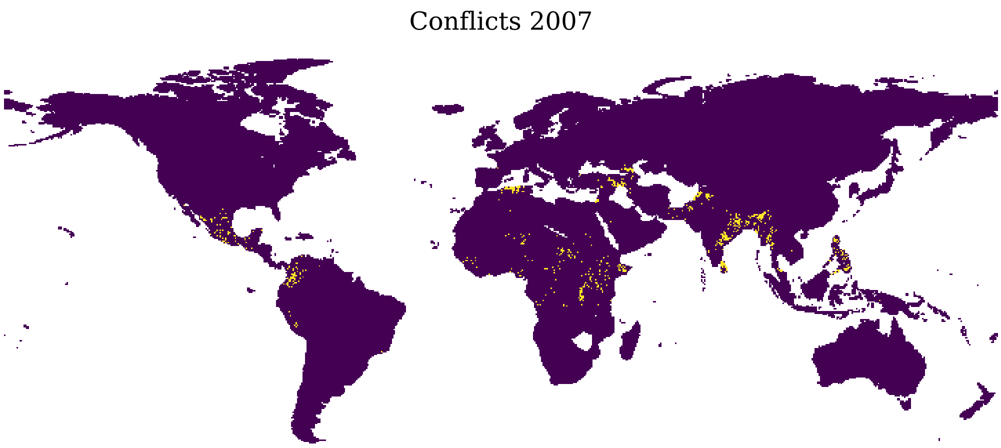

Figure 6: Conflicts observed in 2007 by UCDP aggregated at PRIO grid cell level. The measure is binary, with yellow denoting one or more conflicts in the given cell. Afghanistan, Iraq, Turkmenistan, Georgia and Zimbabwe are missing due to the coding rules of UCDP.

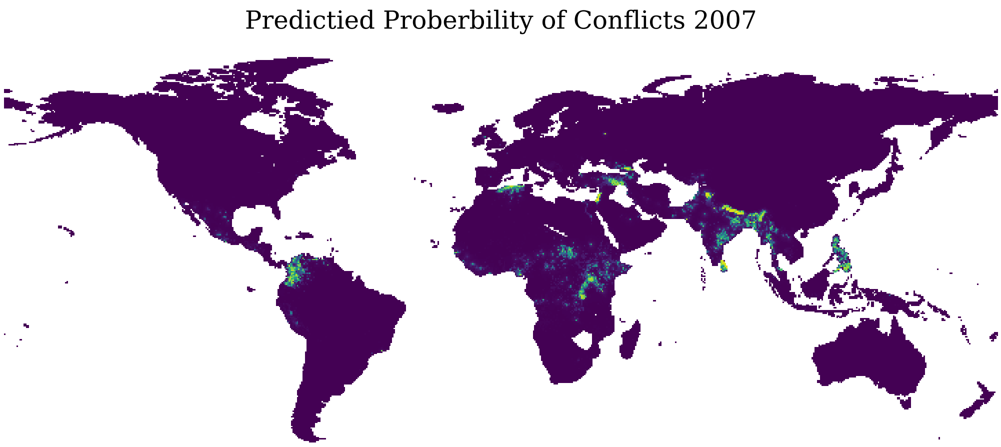

Figure 7: The estimate probability (prior corrected) of conflicts in 2007 using data from 2006 and the model trained on data from 1990 through 2005. The measure is between 0 and 1, where a 1 would denote certain conflict and be colored bright yellow as in Figure 3. Afghanistan, Iraq, Turkmenistan, Georgia and Zimbabwe are missing due to the coding rules of UCDP.

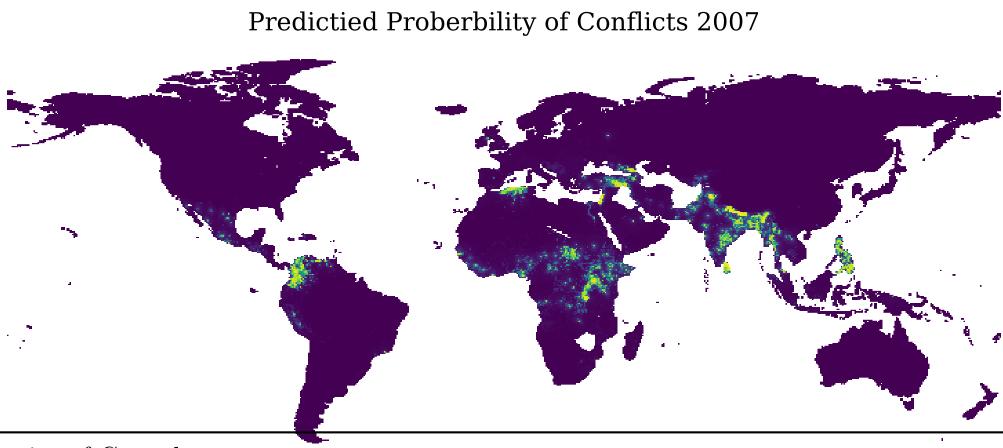

Figure 8: The estimate probability (NOT prior corrected) of conflicts in 2007 using data from 2006 and the model trained on data from 1990 through 2005. The measure is between 0 and 1, where a 1 would denote certain conflict and be colored bright yellow as in Figure 3. Afghanistan, Iraq, Turkmenistan, Georgia and Zimbabwe are missing due to the coding rules of UCDP.