Early ViEWS: A prototype for a political Violence Early-Warning System*

Paper presented to the American Political Science Association annual meeting 2016, Philadelphia

Michael Colaresi[1], Håvard Hegre[2,3], and Jonas Nordkvelle[2]

[1]Michigan State University
[2]Peace Research Institute Oslo
[3]Department of Peace and Conflict Research, Uppsala University

August 29, 2016

Abstract

In this paper we present prototype predictions of intra-state violence from the Violence Early Warning System (ViEWS), a new open-source, forecasting project. The goals for the system are to provide the academic and humanitarian communities with useful and accurate forecasts of the timing and location of different forms of political conflict as well as allow public access to the tools and data that created them. We target forecasting violent events at the PRIO-GRID level with lead times of one month. While we are building the project to be global in scope, we begin with a focus on intra-state violence in Africa. Our approach brings together work on the underlying structural risks of violence, which has identified slowly moving indicators such as GDP per capita, with research on quickly emerging triggers observable in news reports. The core of the project is an ensemble of theoretically informed static and dynamic non-linear models, each of which are weighted and averaged into our overall forecast. We evaluate the performance of our sub-models and overall ensemble prediction using proper scoring rules.

*Authors listed in alphabetical order. An earlier version of the paper has been presented to the 2016 convention of the International Studies Association. Thanks to Hanne Fjelde for input to parts of the text, to Margit Bussmann for comments, and to Frederick Hoyles for excellent research assistance. The research has been funded by the Research Council of Norway, project 217995/V10.

1 Introduction

Large-scale political violence kills and maims thousands of people every month. For every person killed, hundreds are forced to relocate within countries and across borders. Armed conflicts have disastrous economic consequences, undermine political systems and public services, prevent developing countries from escaping poverty and hinder international actors in providing humanitarian assistance (Collier et al., 2003; Gates et al., 2012). The challenges of preventing large-scale political violence or mitigating the consequences for local populations are particularly daunting when it escalates in locations and at times where it is not expected. Policy-makers, humanitarian organizations, or UN peacekeepers would benefit greatly from a system that systematically monitors all locations at risk of conflict and continuously assesses the risks of conflict.

This paper presents a prototype for ViEWS – an ambitious, projected political Violence Early-Warning System under establishment at Uppsala University. At this point, we concentrate efforts on the geographical level and to the African continent. We bring together a fairly diverse set of predictors and modeling strategies, and show that Ensemble Bayesian Modeling is well suited to combine the insights from these diverse formulations.

2 Background

Currently, no political violence early-warning system satisfying our criteria exists. The most common 'systems' for early warning are based on assessments by regional experts and journalists (e.g., `www.stratfor/analysis`). These are frequently well-informed, but focus only on specific countries or regions and lack methodologies to rigorously compare risk assessments. Some systems are quite systematic (e.g. CIDCM's Peace and Conflict Instability Ledger) but lack a frequent update schedule or extensive coverage. Most existing systems fall short in terms of transparency, public availability, and replicability. Intelligence and defense communities have invested significant resources into classified early-warning systems, but these are unavailable to the public. Others are non-transparent since they do not fully account for the underlying modelling, such as the Global Peace Index. Some systems may be good, such as the forecasts from Political Risk Services, Michael Ward's CRISP system, or W-ICEWS, but are not freely available so that outsiders can replicate them and assess their quality. Systems such as the Global Conflict Risk Index (`http://conflictrisk.jrc.ec.europa.eu/`) satisfy some criteria, but do not rely systematically on out-of-sample evaluation of their forecasting models.

ViEWS will be designed to help crisis responders to prepare for conflict-induced humanitarian disasters. A detailed description of how the project will proceed to achieve this as well as other aspects of ViEWS are detailed at `http://www.pcr.uu.se/research/views/`.

In order to succeed as a useful political violence early-warning system, ViEWS must achieve an ambitious set of goals. It must have uniform coverage and frequent updates to avoid blind

spots and specifically provide alerts that are as location- and actor-specific as possible. It must be transparent, replicable, and publicly available to anyone in the sense that all risk assessments may be traced back to a fully specified procedure and publicly available data and tools. It must build on the most advanced theoretical and empirical research within the field, and be based on systematic evaluation of predictive performance.

Eventually, ViEWS will provide forecasts of organized political violence up to 36 months into the future, based on publicly available data and specified statistical models. The output will be probabilistic assessments of the risk and likely severity of three types of organized conflict events classified according to the actors they involve (state-based conflicts, and non-state conflicts, and one-sided violence against civilians), at each of three units of analysis: detailed geographical locations, specified actors, and countries. The system will assess the risk that new conflicts will erupt, continue, or spread to new locations, involve new actors, or start targeting civilians, and whether a recent fragile peace will continue or hostilities flare up again.

Developing ViEWS into a system that produces warnings with sufficient quality to be useful is methodologically and theoretically challenging, but feasible. In order to maximize the predictive performance of ViEWS, the project will focus on six core tasks:

(1) carefully survey the theoretical and empirical literature on armed conflict and adapt core models for use in the forecasting system, (2) develop ViEWS at three levels of analysis (detailed geographic locations, specific actors, countries) and ensure that the levels inform each other, (3) exploit information contained in observing various forms of political violence, (4) harness new methodological developments that allow combining insights from the various modeling efforts, (5) use out-of-sample evaluation to refine the system, and (6) build a 'data-processing tool' – a set of data-collection routines as well as a suite of computer programs that organize data and automate core functions. The program will also invest in tools to communicate the results effectively to a broad audience.

The methodological point of departure will be two sets of hierarchical panel models. The first is defined in terms of PRIO-grid units at the lowest level (referred to as 'geographic models' below), with countries at the highest level. The second will have pairs of identified actor organizations at the lowest level ('actor models'), nested within each actor forming the two sides. In this paper, we focus on the geographic model. We intend to eventually formulate models separately for each of three outcomes to forecast (state-based armed conflict, violence against civilians, and non-state conflict), but merge all these into one set of events in the prototype presented here.

Developing ViEWS involves modifying, improving, and integrating these models along six lines of development: (1) model specification, (2) evaluation criteria and procedures, (3) dynamic simulation, (4) Ensemble Bayesian Model Averaging (BMA), (5) missing-data handling, and (6) alternative estimation methods. ViEWS will also consider how these methodological developments can be used to (7) empirically re-evaluate theoretical propositions.

3 Ensemble Bayesian Model Averaging (EBMA)

An important challenge in ViEWS is the need to integrate implications of the disjoint but related disaggregated models across multiple temporal, spatial and actor-specific units. Across a wide range of domains, including conflict forecasting and climatology, combined predictions from several models have improved out of sample performance. Here we rely upon the framework of Ensemble Bayesian Model Averaging to combine a set of constituent models into an overall weighted prediction (Raftery et al., 2005; Montgomery, Hollenbach, and Ward, 2012). EBMA provides guidance on integrating and systematically weighting multiple models and event types together to provide more accurate forecasts of diverse events. EBMA joint forecasts are typically more discriminative and better calibrated than even the most accurate of the component models that go into them (Beger, Dorff, and Ward, 2014; Raftery et al., 2005; Montgomery and Nyhan, 2010; Montgomery, Hollenbach, and Ward, 2012). Below, we show this is the case also for our geographic models.

EBMA involves separating the data into three segments: a training set to compute parameter estimates, a validation set to compute weights for the ensemble, and a test set to validate out-of-sample predictions. The role of the validation set is to estimate a set of parameters and weights that allow for the individual forecasts based on the model sub-components to be combined into one ensemble forecast. This ensemble will then be used with data available at the beginning of the forecasting window to create out-of-sample forecasts for unseen data.

This prototype represents the first step. As new data are collected over the course of project, these forecasts will be evaluated and our models replaced, reformulated, or retrained accordingly. This allows us to gradually improve the forecasting performance of the system by comparing various combinations of risk factors, of statistical model formulations, and projection algorithms.

Ward, Greenhill, and Bakke (2010) show that statistically significant variables are not necessarily good predictors of conflict. Avoiding over-fitting (where a model captures the noise and idiosyncrasies of the data at hand rather than the underlying pattern) is essential to maximize predictive performance (Hastie, Tibshirani, and Friedman, 2009; Blei, 2014). Models with additional significant parameters may not improve forecast accuracy because they are simply accounting for the same variation as other specifications, just with different parameters.

One important element of ensemble models is that they are able to represent generalizable patterns that any one model can not. For example, a linear additive model that maps two features, A and B, to create a vector of forecasts, f_k, can only learn specific types of patterns from training data, those that are linear and additive. A generalized additive model or a high-order polynomial model with enough data, may be able to mirror traditional signals from the data generation process, such as those that are non-linear. A similar process takes place between the mapping from forecasts, which themselves are functions of input features, to outcomes. A single forecast from a model can carry only a finite set of specific patterns, which depends on the relevant input

Unclear!

features and the flexibility of the model representation itself (linear and additive on the log-odds scale, decision trees, etc). A combination of multiple forecasts can potentially represent signals that span across the models. In machine learning, the improved performance of ensemble models is attributed to these models being able to explore an expanded hypothesis space. One can think about forecasts from each single model as constrained ensembles that simply give all their weight to one and only one model. Ensembles can represent patterns with non-zero weights that these constrained only-one-model ensembles cannot, by definition (Domingos, 2012).

Following Montgomery, Hollenbach, and Ward (2012), we split our data into training, calibration and test sets. The parameters for each of our constituent models are computed on the training set. Using those fitted models, forecasts of the probability that an observation equals 1, $f_k^{(raw)}$ are created, where k indexes the constituent model, using the input features, but not the outcomes at time t, for the observations in the calibration set. As suggested in Montgomery, Hollenbach, and Ward (2012), we post-process each forecast vector in the calibration set, $f_k^{(raw)}$ such that the f_k used in the ensemble is equal to

$$\begin{aligned} f_k = &[(1 + logit(f_k^{(raw)}))^{1/3} - 1]\mathbf{I}[f_k^{(raw)} > \frac{1}{2}] \\ &- [(1 + logit(f_k^{(raw)}))^{1/3} - 1]\mathbf{I}[f_k^{(raw)} < \frac{1}{2}] \end{aligned}$$

We use a logistic model to scale and center each of the k forecasts, such that $P(y = 1|f_k) = logit^{-1}(a_{k0} + a_{k1}f_k)$, in the calibration set. The ensemble model for the outcome is then a weighed sum of these post-processed forecasts, f_k. To estimate the weights, we can use the fact that,

$$p(y|f_{1|t}, f_{2|t}, \dots f_{k|t}) = \sum_{k=1}^{K} w_k\{P(y = 1|f_k)\mathbf{I}[y = 1] + P(y = 0|f_k)\mathbf{I}[y = 0]\}$$

with $w_k \in [c, 1]$, where $0 \leq c \ll 1$ represents a lower bound on the weights (potentially equal to 0), and $\sum_{k=1}^{K} w_k = 1$. We estimate the scaling parameters, a_{0k} and a_{1k}, and the weights, w_k, using the EBMAforecast package in R (Montgomery, Hollenbach, and Ward, 2012). Again this step only utilizes the calibration set. The weights are estimated using the EM algorithm. We constrain each weight to be at least .01, representing what Montgomery, Ward, and Hollenbach (2015) note is a "wisdom of the crowd" effect where keeping more models in the ensemble can improve out of sample performance. The overall ensemble forecast is

$$p(y = 1|f_k, w_k, a_{0k}, a_{1k}) = \sum_{k=1}^{K} w_k(a_{0k} + a_{1k}f_k)$$

.

3.1 Model selection/formulation

Ensembles are most effective when they combine heterogeneous models. At the limit, if two models produce identical forecasts, then the ensemble forecast is invariant to the weights we would associate with each model. Similarly, if the models all yield extremely similar predictions, there is little additional information in the ensemble relative to models themselves. To maximize performance, we seek to specify a number of models that are distinct and able to pick up on a broad variety of early-warning signals. Some of these messages might be about recent conflict history locally, in the neighborhood, or nationally, some about government bases/population centers, resource locations, or other potential targets, and some about institutions in the country as a whole, etc.

Since our goal is both prediction and to theoretical insights from the models (and ensemble), most models are interpretable – they are 'thematic' in the words of Mike Ward [ref]. Each 'theme' is like a unique perspective illuminating part (but only part) of the data-generation process. Perspectives with very little signal can be valuable in the ensemble if they do not overlap much with other perspectives and explain non-zero generalizable patterns in the data generation process. This heterogeneity in perspectives also apply to more technical modeling aspects. The EBMA procedure allows to draw on the divergent strengths of models built with very different assumptions (e.g., linear vs. generalized additive models vs. random forest).

We take a pragmatic approach to overlap between models, trying to define them as thematically consistent in most cases. The results presented below show that the EBMA approach is effective in extracting the unique contribution of each model.

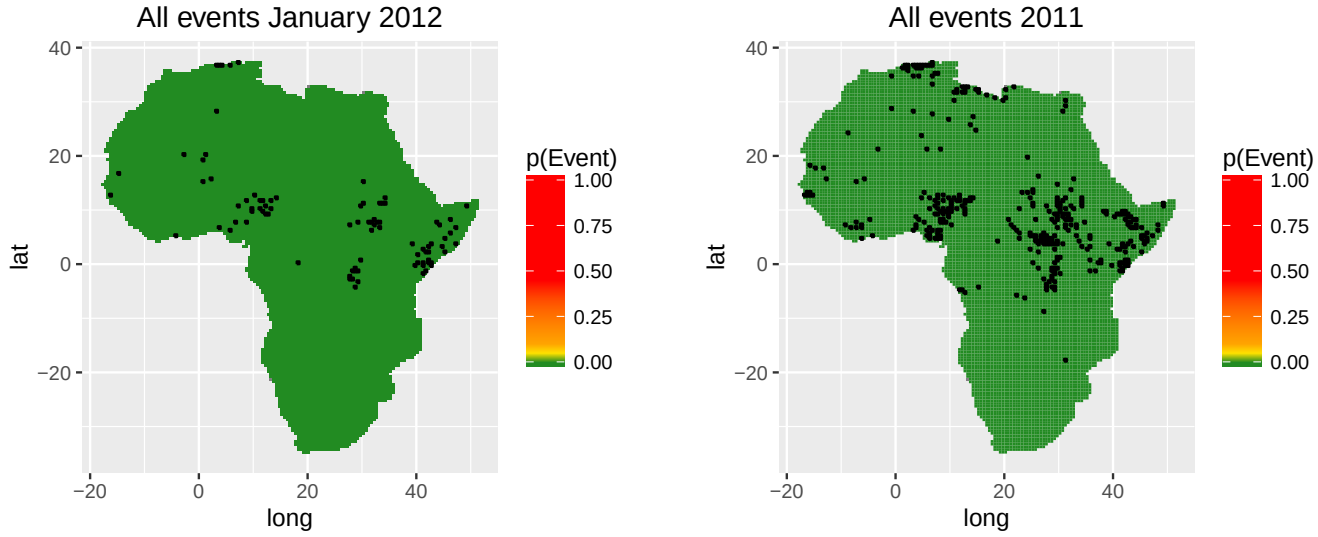
4 A pilot: Predicting the location and timing of conflict events

This paper presents a pilot to consider some methodological aspects of ViEWS. Most importantly, we will explore how useful EBMA is for the problem at hand, what challenges EBMA involves, and look briefly into some model selection issues. More substantively, we will also assess various ways to capture the conflict history at a geographical location, explore some geographic indicators as well as include some country-level characteristics in multi-level models.

We focus here on the geographical aspect of violent conflict. An important part of an early-warning model is being able to pinpoint as exactly as possible time and place of future events. Here we explore a pilot to find out how well we do predictions on PRIO-GRID spatial scale and a monthly temporal scale, abstracting away from who the actors that engage in violence are.[1]

[1] This will be the topic of another ViEWS prototype.

Figure 1: Actual events for January 2012 (left) and January–December 2011 (right)



4.1 Geographical and temporal scope

4.1.1 Dependent variable

We currently run models with a binary outcome (using the logit link), measuring whether or not a UCDP-GED event was recorded in that cell-month or not. A UCDP-GED event is an instance of fatal organized violence: 'The incidence of the use of armed force by an organized actor against another organized actor, or against civilians, resulting in at least 1 direct death in either the best, low or high estimate categories at a specific location and for a specific temporal duration' (Sundberg and Melander, 2013, p. 524).

In this pilot, we have restricted the dataset to all PRIO-GRID cells in mainland Africa (i.e., excluding Madagascar and other islands), at a monthly resolution. To train the models, we use data from January 1991 until December 2005 (we use the first two years (1989–90) of the UCDP-GED dataset to construct a minimum of 24 months of lags for various measures, and leave them out of the analysis). We have separated the test sets into a calibration set (January 2006 until December 2010), and a final test set (January 2011 until December 2014). Although the dataset is large, with XX,000 unique grid-cells sampled for multiple months, the dependent variable is quite sparse. There are 11,799 events in training dataset, and XX,000 and XX,000 in the calibration and test data sets, respectively.

We will look at the predicted and observed outcomes for January 2012 in more detail. The left part of Figure 1 shows the geographical location of the UCDP-GED events of that month. Since conflict history is such an important predictor, we also plot the location of similar events for all of the preceding 12 months to the right in the figure.

Table 1: Overview of constituent models

Model	Theme	Type	Description
A_DistL1	Distance	GAM	Shortest distance to a conflict event occuring at $t-1$
A_MA12Dist	Distance	GAM	The average shortest distance to a conflict event, each month the last 12 months (not including current month)
A_G1500CBD	Distance	GAM	Gravity formulation, only including events within 1500 km. Battle-related deaths set to 1 if 1 or higher, i.e. $\alpha_a \to 0$ if there is at least one fatality in the past event.
A_xycoord	History	GAM	Geographical coordinates as the only predictors; no distinction between the timing of past events
A_HistLagsQ	History	GLM	Lagged conflict in cell at $t-1, t-2, t-3 \sum_{k=4}^{12}(t_k)$, plus similar for lagged conflict in neighboring cells
A_HistLagsGam	History	GAM	Smooth functions of time since past event in same grid cell and time since past event in neighboring cells
A_HistSeverity	History	GAM	Use number of battle deaths in cell and neighboring cells over the past year
B_Accessibility	Geography	GLM	Population density, travel time to capital/urban centers, mountains, night lights, lagged dependent variables
B_Accessibility_gam	Geography	GAM	As B_Accessibility but with smooth functions of population, travel time, mountains
B_Poverty	Geography	GLM	Infant mortality rate, night lights, child mortality rate [CHECK], urbanization, lagged dependent variables
B_Poverty_gam	Geography	GAM	As B_poverty but with smooth functions of population, travel time, mountains
C_Drought	Geography	GLM	Location and timing of droughts, no lagged dependent variables
C_Resources	Geography	GLM	Location of diamonds, gems, gold, oil, no lagged dependent variables
C_Landscape	Geography	GLM	Type of landscape in grid cell, no lagged dependent variables
C_Landscape_gam	Geography	GAM	As C_Landscape, but with smooth functions of all terms
D_CL_history	Country-level history	glmer	Log time since conflict in country up to $t-2$ at country level, population and distance to past events last 12 months at grid-cell level
D_CL_democracy	Country-level democracy	glmer	Democracy score and its square and decay function of time since regime change at country level, population and distance to past events last 12 months at grid-cell level
D_CL_capacity	Country-level state capacity	glmer	Population size and GDP per capita at country level, population and distance to past events last 12 months at grid-cell level
D_CL_acc_poverty	Country-level REs	glmer	Only random intercepts at country level, variables from B_Accessibility and B_Poverty at grid-cell level

4.2 Models in the ensemble

To predict the location and timing of such events, we formulated 20 different models. Table 1 gives an overview of these. A more detailed list of the variables used, sources, etc., is to be found in Table in the Appendix, and the R commands used to estimate the models for the training dataset is found in Appendix A.1.

As discussed in Section 3.1, we sought to formulate models that bring a variation of perspectives to the ensemble. The models fall in four broad groups. Group A focuses on the recent conflict history in the PRIO-GRID cells and their neighborhoods. The first four of these relate the risk of conflict events as smooth functions of variables defined in terms of the geographical distance to recent conflict events. A_DistL1 and A_MA12Dist are functions of the distance between each grid cell and conflict events in the month or the year preceding the month of observation. A_MA12Dist is based on a measure of the shortest distance in kilometers to an event in each month for the previous 12 months. The distance is then averaged across the 12 values for each cell. A generalized additive model is used to learn how the average distance from the nearest conflict event over the last 12 months maps to the presence or absence of conflict in the next period in that cell. A_DistL1 is similar, except that we use the shortest distance to a conflict event only for the previous month, not all previous 12 months. The A_G1500CBD model is based on a 'gravity model' measure of

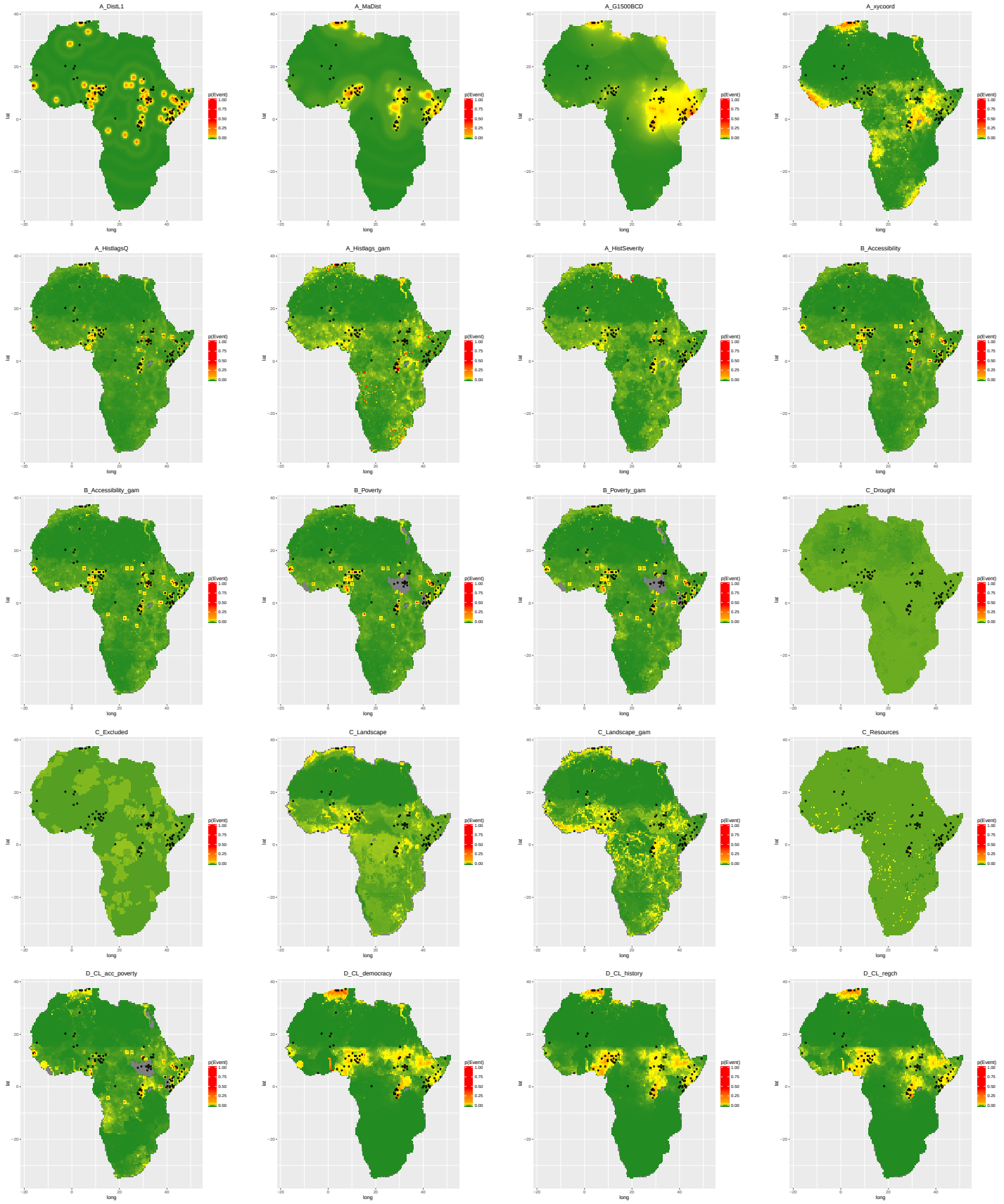
influence from previous events. It accounts for all previous conflict events within 1500 km of each cell's centroid, and uses the number of battle-related deaths as an indicator of the intensity of these evens. The measure is a weighted sum of distances to all previous events (from 1989 onwards). Time is measured in days and again a generalized additive model is used. The weights are constructed so that more recent events are assumed to have more influence than events many years earlier. A detailed description of the measure is found in Appendix A.2. The A_xycoord model only uses the geographical coordinates of locations as predictors.

Figure 2 shows the predicted probabilities for a conflict event within a grid cell in January 2012 (an arbitrary chosen out-of-sample month) for the 20 models we have estimated. The observed locations of violent events are represented by black boxes (these are the same as in Figure 1). The predicted probabilities from each of the models are given as heat maps. Predictions for January 2012 from the A_DistL1 model appear as concentric rings with the events of December 2011 at the centres. Predictions from the A_MA12Dist are in effect the averages of similar patterns for each month of 2011. Since multiple events in the same location contribute individually, the predictions for this model is more concentrated in regions with intense fighting in 2011 – the fighting in Libya in 2011 shows up more clearly here than in the first model. The predictions from the A_G1500CBD model also take more distant events into account, both in time and space. Accordingly, Rwanda and Burundi appear as distinct hotspots, as does the location of Mogadishu. Even North East Egypt appears as a hotspot due to conflicts in the Middle East. The A_xycoord model is a smooth function of the geographical coordinates. In effect, it is a smooth of the average number of conflict events over the *training data* 1991–2005, with no distinction between recent and early events.

The three remaining models in the A group are primarily defined in terms of time since events in the same location or its immediate neighbors. The A_HistlagsQ.pdf is based on a standard logit model with dummy variables for whether there were any conflict events in the grid cell or its neighbors in each of the 12 months before the month of observation. Predictions resemble that of the two first models, but are much more localized. A_Histlags_gam is a GAM model of the same information, but also taking into account events more than 12 months earlier. The prediction map shows that this model portrays many of the locations with conflict many years earlier as high-risk places, such as many locations in Mozambique, Angola, and South Africa. The final model in group A takes the severity of events (in terms of number of people killed) over the past 12 months into account in a GAM model. For January 2012, this model gives particular attention to locations in Nigeria, Libya, Somalia, and South Sudan.

The next two groups of models are based on characteristics of the geographical locations drawn from the PRIO-GRID collection of data (Tollefsen, Strand, and Buhaug, 2012). The B group models all include lagged dependent variables – whether there was a conflict in the grid cell or its immediate neighbor in the month before the month of observation. B_Accessibility is a GLM model that includes variables such as population density and travel time to urban centres. B_Accessibility_gam contains the same predictors but is estimated using a GAM model. The model

Figure 2: Predicted probabilities and actual events for January 2012



picks both that conflict events are likely where population concentrations are high, such as near Kinshasa or densely populated Rwanda, Ethiopia, and Nigeria, but also in inaccessible regions such as Chad or the interior of the Democratic Republic of Congo, and close to interstate borders. The B_Poverty and B_Poverty_gam models are based on data on infant and child mortality rates and degree of urbanization.[2]

The C group models select another set of location characteristics from PRIO-GRID, but do not include lagged dependent variables. The C_Drought model includes variables quantifying local drought – one representing the proportion of the year with drought, and one with the proportion of months in the growing season with drought. These variables area also entered as lagged by one year. The estimates in the drought model are partly in the opposite direction of what one might expect. Hence, Sahara has a low predicted probability of violent events in this model, as seen in Figure 2.

The C_Resources model contains variables denoting the locations of diamonds, gems, petroleum deposits, and surface, placer, or vein gold deposits . The petroleum variable has a negative estimate (explaining the scattered dark green shades in North Africa in the map in Figure 2, all locations with petroleum deposits). The diamonds and surface gold deposits variables have strong positive estimates, as reflected in the yellow shade in Sierra Leone, for instance.

The C_Landscape and C_Landscape_gam models are GLM and GAM versions of models with variables denoting various landscape types. Pasture, savannas, and barren landscapes are estimated as having a low risk of violent events, whereas agricultural landscapes have a high risk. Forest landscapes are more neutral. To some extent the model predictions reflect population density at various locations. This is seen in the map in Figure 2, where predicted risk is high in densily populated areas in the Great Lakes region, in Ethiopia, along the Mediterranean coast and the Gulf of Guinea.

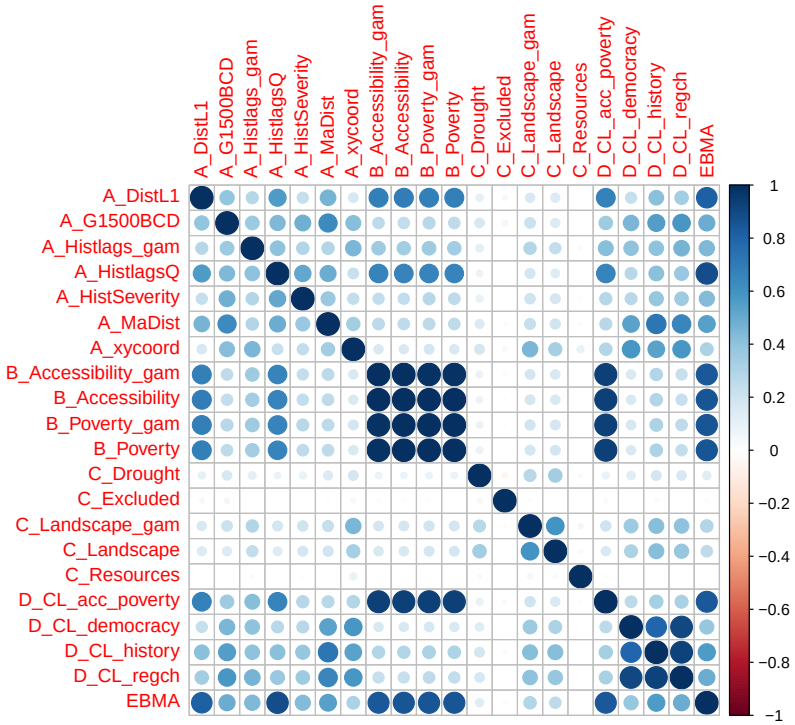
The final set of models are a set of hierarchical models with terms at both the country and grid-cell level. The three first of these are random-intercept/random-slope models and include log population and the MaDist variable at the grid-cell level. As noted above, the MaDist variable is the shortest distance to violent events averaged over the past 12 months.

In addition to this, the D_CL_history model has random effects at the country level as well as the natural logarithm of the number of years without state-based armed conflict up to year $y-2$. For January 2012, for instance, this variable is based on the number of years in peace up to and including 2010, but excluding 2011. As seen in the map in Figure 2, this model qualifies the A_MaDist model considerably. The violent events in Nigeria has a much greater impact in Nigeria or Rwanda than in Benin or Tanzania.

The D_CL_democracy model has the square of the country's Polity score as the country-level predictor. Hegre et al. (2001) finds that semi-democracies have a higher risk of armed conflict than autocracies and democracies. This pattern is reflected in this model, too, although the estimate is

[2]For some reason, PRIO-GRID is missing data for South Sudan, Liberia, and some other regions for this variable.

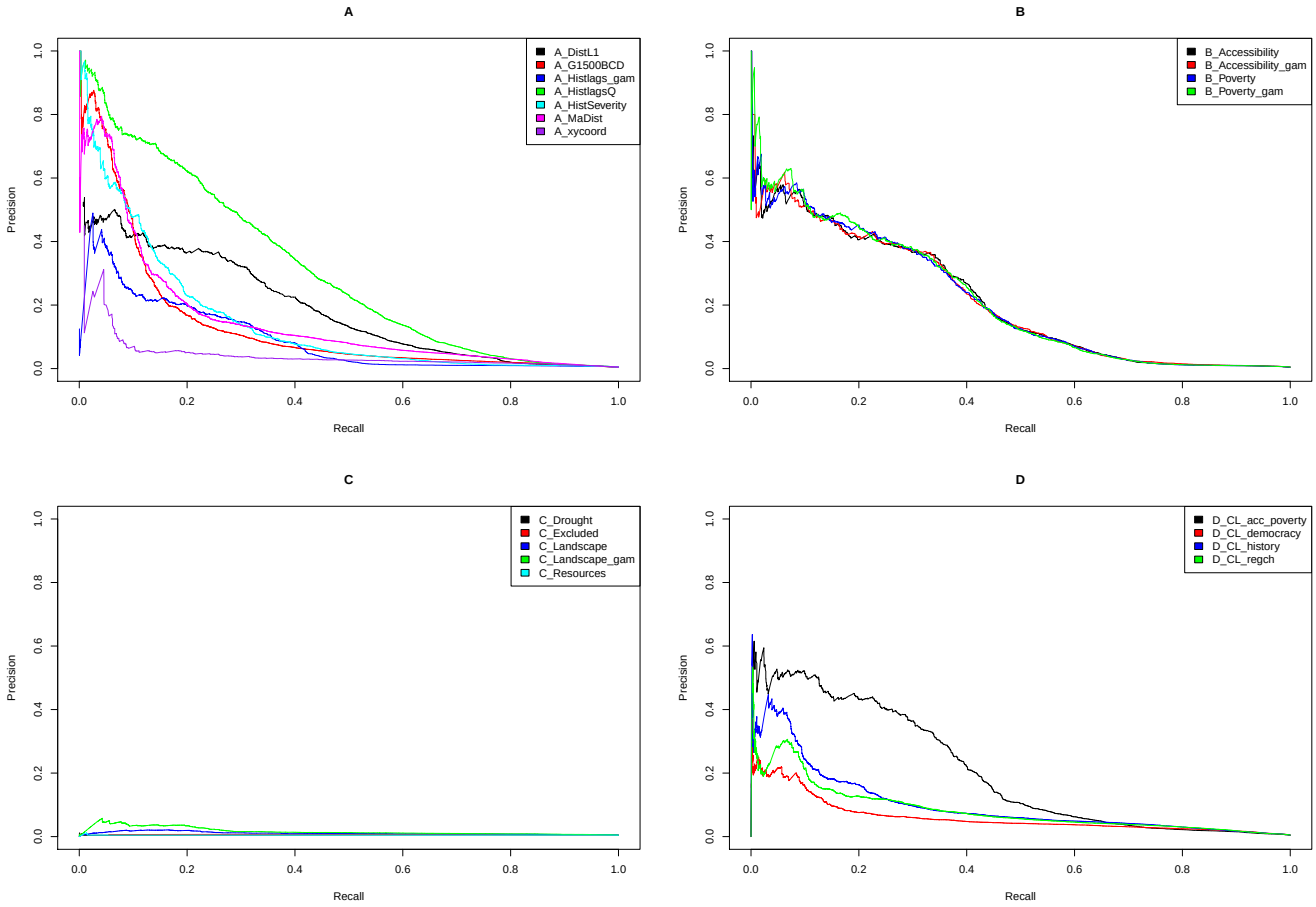
Figure 3: Correlation between out-of-sample predictions



not statistically significant. In Figure 2, this is reflected as high estimated risk of political violence in Togo, Sierra Leone, and Burundi, all countries that score in the middle region between –2 and +6 on the Polity scale. The D_CL_regch captures another finding in Hegre et al. (2001), that countries that have recently gone through regime changes as reflected in changes to the Polity score have a higher risk of armed conflict. Again, the estimate in the current model is not significant but has the same sign as in Hegre et al. (2001). In Figure 2, this is seen as a heightened risk of violence in the same countries as in D_CL_democracy, all of which were fairly new democracies in 2012.

The final multi-level model, D_CL_acc_poverty, has only random intercepts at the country level. At the grid-cell level, it has a combination of the variables from the B_Poverty and B_Accessibility models, with the same form for lagged dependendent variables as in these. A comparison of the maps in Figure 2 for these three models shows how the multi-level model qualifies the two single-level models. Compared to the B_Accessibility model, it predicts a lower risk of political violence in Nigeria which actually has a somewhat low incidence of violence *per capita* than other African countries. It also predicts a much lower risk of violence in countries such as Tunisia, Botswana, Malawi, and Tanzania that had seen very little violence in the 20 years up to 2012.

Figure 4: Precision-recall curves for 2006–2010 calibration data, main models



4.3 Distinctiveness of models

Figure 3 shows the correlation between the predictions for the calibration dataset based on the models estimated on the training dataset. This graph only provides a partial picture of distinctiveness, since all the models that share terms have correlated predictions. This is particularly notable for models that include lagged conflict events in the grid cell and its nearest neighbor (e.g. all the 'B_' models). The 'C_' models are clearly distinct from these models, but this is primarily because they do not include lagged dependent variables, Hence, it is not certatin that they pick up on a much more independent set of early warning signals than the B_ models, for instance, where the lagged dependent variables ensure high correlations between predictions.

4.4 Out-of-sample predictive performance of the 20 constituent models

The models described above were estimated on the training dataset (1991–2005). In Figure 4, we have plotted precision-recall curves comparing the predictions for the calibration dataset (2006–2010) with the observed data for the same period. The upper-left panel shows these curves for the group A models that focus on the political violence history of locations. It shows that A_HistlagsQ

yields better precision for all levels of recall and is clearly the best individual model. The A_xycoord model performs clearly worse than the others.

The upper-right panel compares the B group models. Since the lagged dependent variables have such strong predictive power, the precision-recall curves are quite close to each other. The B_Poverty_gam model performs somewhat better than the others. The lower-left panel shows the C group models. As individual models, they perform quite poorly compared to the A and B group models because they do not include any information on past violent events. Compared to each other, the two landscape models do better than the others, and the GAM version better than the GLM version. Finally, the lower-right panel shows the precision-recall curves for the multi-level models. The democracy and regime change models perform worse than the D_CL_history model. The D_CL_acc_poverty model which has a more effective representation of past events at the grid cell level do considerably better than the other multi-level models, but not discernably better than for instance the single-level B_poverty_gam model.

5 Forecasting with Ensemble Bayesian Model Averaging (EBMA)

We selected five out of the 20 models for the ensemble predictions that we present here.[3] These models were selected based on two criteria. First, we wanted to include high performing models in the ensemble. While EBMA can be effective at pooling weak learners into a stronger predictive system, it is important that the constituent models include signals that generalize out-of-sample. We thus reviewed the AUC and PR curves for each of the models and selected candidate models that had the highest within-category values. Second, we looked at the correlation between the predicted values across the models. Ensembles are effective when they incorporate diverse signals. Including five models that produce predictions that are highly correlated and thus include nearly the same informational content, will not improve an ensemble as much as combining models that include a diversity of the signals, even if some of those signals are only weakly related to the

[3]We also attempted to compute an ensemble that included all of the described models. However, as the objects that hold the model and predictions are between 5 to 10 GB each, and there are 20 total models, the computational resources needed are extensive. We have had an ensemble model running on the large-memory cluster at Michigan State University's HPCC for a wall-time of 10 days, as of this writing. The model uses over 125 GB of memory and the EM algorithm is not surprisingly slow to converge. In the future, we will explore stochastic variational inference as a way to quickly scale EBMA to larger sets of models and big data situations such as ours. In stochastic variational inference, we need to be able to separate our parameter vectors into global and local sets. The local sets are independent conditional on the global parameters, allowing for parallelization. The global parameters are fixed across the local sets, and so the global parameters can updated based on a subset of local variable. In fact, SVI has a connection with the E-M algorithm used in the EBMAforecast package that we currently deploy (Neal and Hinton 1999). The local parameters in EBMA are the posterior probabilities of model k for observation i. The global parameters are the model weights. Another challenge we ran into was that some of our categories of models included missing values in the calibration set. This is a problem when all of the models in an ensemble have missing values for a particular prediction. One task moving forward is to model missing-ness in the relevant data.

		Calibration Set			Test Set		
Model	Weight	Brier	AUC	Accuracy	Brier	AUC	Accuracy
EBMA	–	0.0038	0.941	0.994	0.0059	0.934	0.991
A_HistlagsQ	0.341	0.0038	0.916	0.994	0.0059	0.916	0.991
A_DistL1	0.366	0.0041	0.899	0.985	0.0063	0.899	0.978
B_Poverty_gam	0.095	0.0040	0.873	0.994	0.0061	0.879	0.991
C_Landscape_gam	0.070	0.0047	0.755	0.995	0.0071	0.740	0.992
D_CL_history	0.127	0.0044	0.914	0.991	0.0068	0.899	0.988

Table 2: A comparison of EBMA and constituent model performance measures in the calibration and test sets. Models are sorted by category and Brier score in the calibration set. Lower Brier scores and higher AUC and accuracy values imply better performing models. The ensembles is a weighted average of the constituent predictions, with the weights supplied in the first column. Accuracy is computed with a threshold of .1, suggesting that it is approximately 9 times as costly to miss an observed than it is to miss a month of peace.

outcome of interest. In the latter case, the ensemble is able to blend relevant information from multiple sources, that by definition, generalize out of sample.

The performance of the constituent models and ensemble are presented in Table 2.[4] The first column reports the estimated weights for the ensemble. The ensemble prediction is a linear combination of weights (which sum to 1) and the constituent predictions. Interestingly, despite the A_HistlagsQ model having the strongest calibration performance out the constituent models (not including the ensemble), it only has the second highest weight. This fact highlights the need for a diversity of models in an ensemble since the ensemble needs unique information that generalizes to new accurate predictions. Also of note is that both of the historical and spatial (from category A) models have the highest weight and are among the strongest performers. Knowing that a conflict happened nearby or in the recent past in a grid cell provides important information on where conflicts are likely to be. Even knowing that an conflict happened recently at the country level of aggregation is useful, as the category D model (D_CL_history) had the third highest weight. Conversely, the landscape model (from category C) is the weakest performer and has the lowest weight. Both this model and the poverty model suffered from important missing data that contributed to their weak performance overall.

When we turn to the performance measures in the calibration and test sets, we see that the EBMA model is able to tie or exceed the best performing models across each metric, save Accuracy, where it is very close. The ensemble has the highest AUC in both sets. The only model that scores higher in accuracy in the test set is the Landscape model. This occurs because the Landscape model makes very low predictions that nearly all fall below the .10 threshold used to dichotomize the predictions, and conflict is extremely rare. Even so, performance of the ensemble is nearly

[4]These results were produced with the EBMAforecast package in R 3.1.1. We set the minimum weight to .01 so that models do not completely drop out of the ensemble (with zero weights). We initialize the weights with values $(.4, .2, .2, .1, .1)$ reflecting their overall performance in the training set. If these are not useful weights the computed weights will drift away from these values.

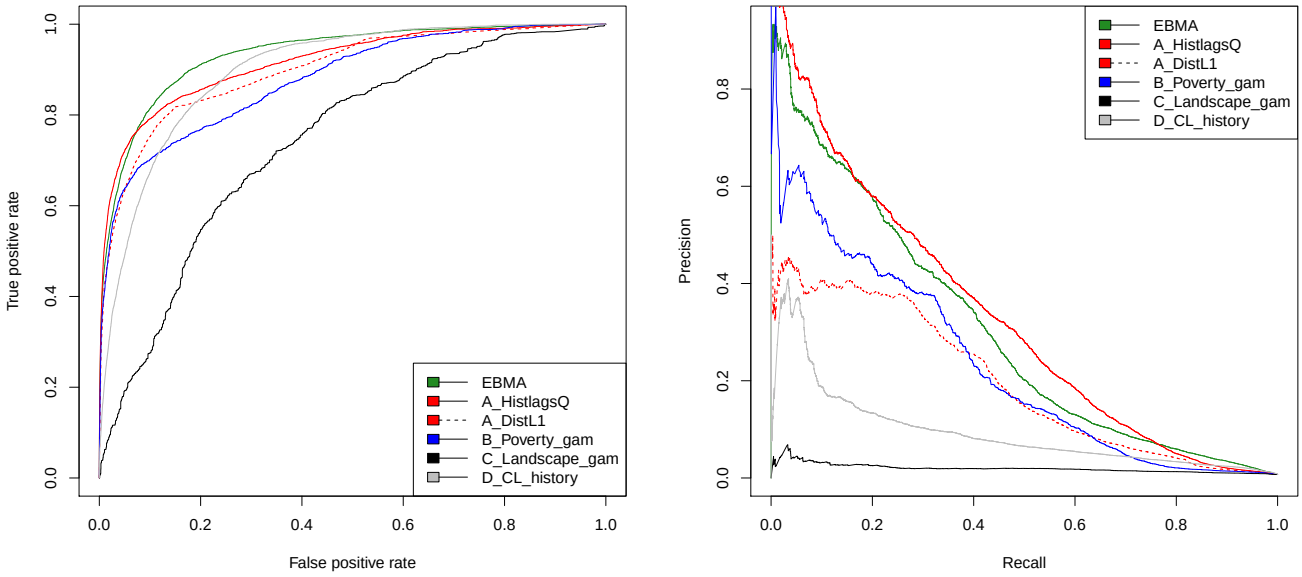


Figure 5: ROC and PR curves for EBMA 2 and constituent models.

comparable.

The value of the ensemble is more apparent in the ROC and precision recall curves. These are presented in Figure 5. The ROC curve plots the true positive rate (sensitivity/recall) on the y-axis and false positive rate (1 minus the true negative rate) on the x-axis. The curve is drawn for each model by varying the threshold used to dichotomize the predictions from that model for each observation. When this threshold is zero, all of the observations are predicted to be conflicts. This results in a true positive rate of 1, as we predict each actual conflict with the correct observed value, but at the price of 1 for the false positive rate, as the true negative rate is 0. As the threshold slides up to just over the lowest predicted value by model, that value is predicted to be a zero, and the rest ones. If that actual observation was a zero, the true positive rate remains 1, but the false positive rate declines. This bends the curve upwards and towards the top-left corner as the calculations are iteratively done for each unique predicted value for the model. A perfect ROC curve is the right angle at the top-left corner of the plot. Predictions that contain only noise and no systematic information about the observations would follow the 45 degree line, trading off equally true positive for false positives and thus not bending upwards and to the left at all. The EBMA model nearly dominates the rest of the models across all thresholds. It is able to have comparable performance to the A_HistlagsQ model at low false positive rates/high thresholds for predicting a conflict and the country level D_CL_history model at higher false positive rates/lower thresholds.

The performance of the ensemble is slightly less impressive on the Precision-Recall plot. Here

recall (the true positive rate) is again plotted, but on the x-axis. he y-axis illustrates precision which is also known as the positive predictive value. It is the probability that given you predicted a conflict, a conflict was observed. Precision-Recall curves are particularly useful when the outcomes are imbalanced, as they are in our data, with only a small fraction of conflict occurrences relative to conflict absence. However, it is important to weigh the relative cost of false positives versus false negatives, where high precision at the cost of missing most conflicts might be detrimental to the policy relevance of the forecast. The curves in Figure 5 are computed by varying the threshold τ from the lowest to the highest (or vice versa) predicted values in each model. If we begin with a threshold of 0, then precision will equal the overall true positive rate for the observed values, as we produced uniformly conflictual predictions. Thus, the curves for all of the models must pass through this point, when recall equals 1.

In PR-space, the best performing constituent model depends on the recall level that is requested. The A_HistlagsQ model has higher precision than the EBMA model across a range of middle recall values. It is only when we have higher values of recall, ensuring we catch more of the observed conflicts, that the EBMA model has higher precision. For example, if we want to catch over 80% of the conflicts with as much precision as possible, we would select the ensemble. On the other hand, if one were interested in only predicting a smaller number of the actual conflicts (lower recall), but with as much precision as possible (certainty that an predicted conflict was an observed conflict), than the historical queens lag model would be preferred.

The difference between the two is revealing. Conflicts cluster in time and space in our sample, but they have to start somewhere. A historically based model has high precision because it catches the recurrences. However, this model misses all or most of the new cases, by design. The other models, including the ensemble, while searching out new cases, lower their precision, but collect higher recall. In the future, building models that better identify new conflicts could improve the ensemble here further. In addition, the landscape model has extremely low relative precision and recall. It predicts very few conflicts, and when it does, they are not observations of violence. It is likely that this model is dragging down, to an extent moderated by its low weight, the PR-curve for the ensemble with its low average predictions for observations of conflict. Its higher Brier score implies that it makes predictions that are farther from the observed values than any other model and the PR curve suggest that this is particularly true for the observations of conflict.

Overall, the ensemble does a superior job of balancing true positive and true negatives, as compared to the constituent models, but at the price of slightly lower precision as compared to the top performing model. The role of the country level model (D_CL_hist) is potentially worth further exploration. Despite making extremely spatially coarse predictions, this model receives the third highest weight overall and improves upon other constituent models in the ROC curve if you are willing to accept a moderate false positive rate. This suggests that multi-level information tying together grid-cell and country-level predictors might be an effective strategy in the next iteration. Another dimension to improve on in the future is the handling of missing data. To calibrate the

ensemble, we need at least one non-missing prediction for each grid-cell month in the relevant set. If we were to fail to make predictions for a given cell, evaluation of the forecasts would be more difficult as different systems may drop more or less difficult to predict observations. Instead, we will work to impute missing data, while filtering forward the uncertainty of these estimates. It is also possible that the fact that specific data is missing might supply us with information that better informs our forecasts. The missing economic indicators appears to cluster near conflict-prone regions.

Figure 6 shows the predicted probabilities for January 2012 for the EBMA model (upper left) and the five constituent models.[5]

6 Conclusions

We have here presented an outline for how to develop an ambitious political Violence Early-Warning System (ViEWS). As an early step, we have also formulated a few very restricted models of how the risk of events in a PRIO-GRID cell month (about 50x50km in size) can be related to previous events, with various formulations for how numerous earlier events at different distances from the cell month in time and space. When estimated separately, the simplest formulations seems to work the best. However, we have also demonstrated the usefulness of Ensemble Bayesian Model Averaging for an early-warning system of complex events. The EBMA model clearly outperforms all the individual models explored at this point.

How well do we predict? The upper left panel in Figure 6 shows that the EBMA model does a good job in predicting the January 2012 conflict events in Algeria, Nigeria, Sudan, South Sudan, and Somalia. As demonstrated in Figure 1, these conflict events occurred in regions that saw a large number of conflict events throughout all of 2011.

Some events were very poorly predicted. One such was an isolated instance of violence against civilians in DR Congo, close to the border with the Republic of Congo. This event occurred in a neighborhood where no conflict events were reported over the 2008–2011 period. A cluster of events in Mali were not among the predicted cases either. These were the first events in the escalation of violence in North Mali in 2012–13.

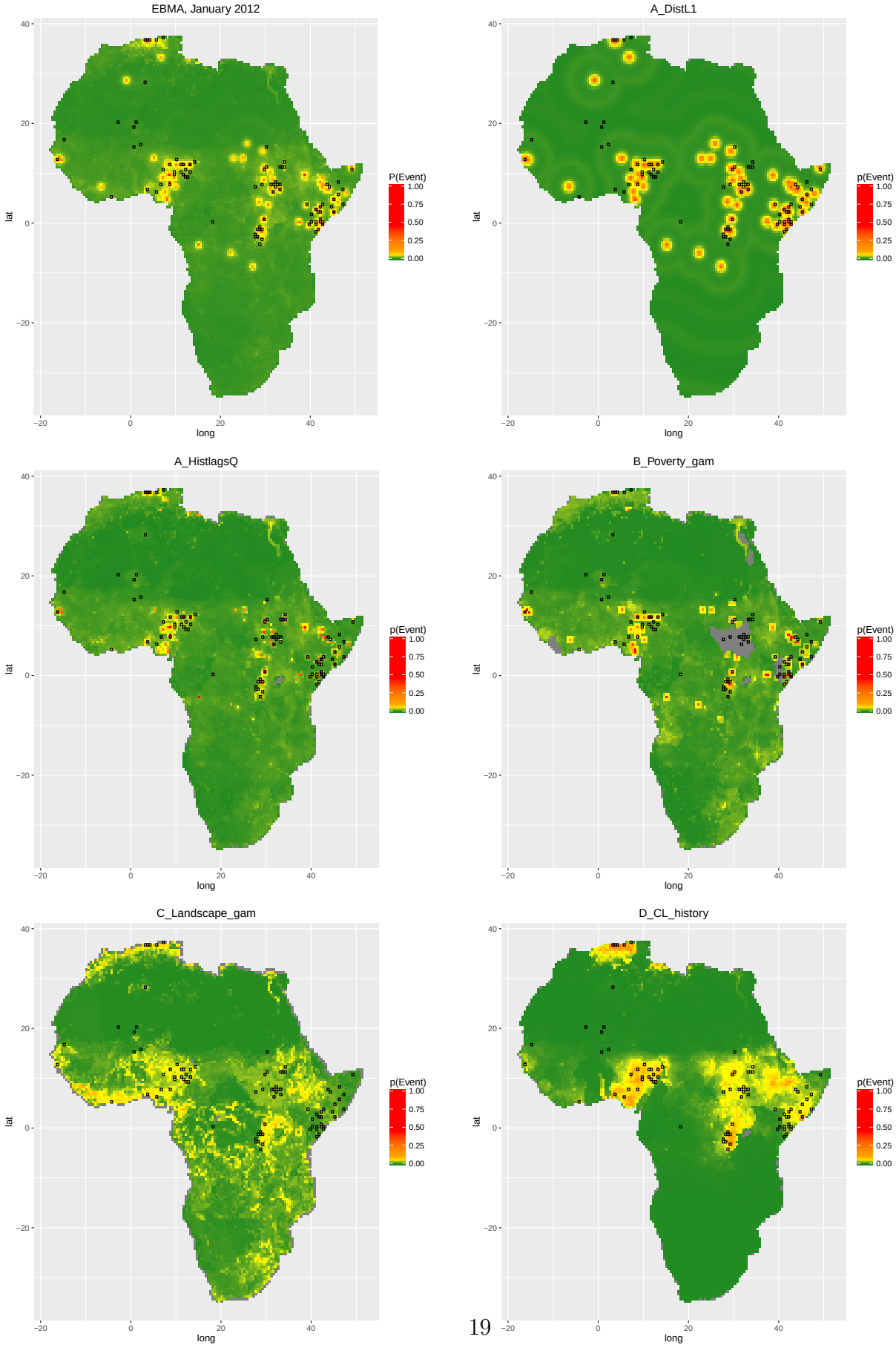
6.1 Next steps

The analysis presented points to a number of areas where further work is required.

First, the size of our dataset, with many thousand grid cells observed monthly, in combination with computationally intensive methods requires research into how to obtain estimates more effectively. Even with excellent access to high-performance computer clusters, the analysis runs into capacity problems. A next step for this project is to explore stochastic variational inference

[5]The five constituent model maps are identical to those in Figure 2.

Figure 6: Predicted probabilities and actual events for January 2012, ensemble and constituent models



to obtain the EBMA weights, and validate the result from such models against the observed data out of sample.

A second step is to develop good routines to fill in missing values, and to evaluate also these routines in terms of how they affect out-of-sample predictive performance.

At the substantive side, the most important task is to specify models that perform better in predicting new conflicts. Some of the models described above but not included in the ensemble would help, as soon as the capacity issues are sorted out. The most efficient approach, however, will be to collect systematic data on signals that violence is on the rise based on news reports on protests, government changes, arms transfers, etc. This is beyond the scope of this pilot, but will be a central activity for ViEWS.

References

.

Balestri, Sara (2012). "Gold and Civil Conflict Intensity: evidence from a spatially disaggregated analysis". In:

Beger, Andreas, Cassy L. Dorff, and Michael D. Ward (2014). "Ensemble Forecasting of Irregular Leadership Change". In: *Research and Politics*. DOI: 10.1177/2053168014557511.

BEGUERíA, SANTIAGO, Sergio M Vicente Serrano, and Marta Angulo-Martínez (2010). "A multiscalar global drought dataset: the SPEIbase: a new gridded product for the analysis of drought variability and impacts". In:

Blei, David M. (2014). "Build, Compute, Critique, Repeat: Data Analysis with Latent Variable Models". In: *Annual Review of Statistics and Its Applications* 1.1, pp. 203–32.

Blyth, Simon (2002). *Mountain watch: environmental change & sustainable developmental in mountains*. 12. UNEP/Earthprint.

Bontemps, Sophie et al. (2011). "GLOBCOVER 2009-Products description and validation report". In:

Collier, Paul et al. (2003). *Breaking the Conflict Trap. Civil War and Development Policy*. Oxford: Oxford University Press.

Croicu, Mihai and Ralph Sundberg (2015). "UCDP Georeferenced Event Dataset Codebook Version 4.0". In: *Journal of Peace Research* 50.4, pp. 523–532. URL: http://www.pcr.uu.se/research/ucdp/datasets/ucdp_ged/.

Domingos, Pedro (2012). "A Few Useful Things to Know About Machine Learning". In: *Communications: ACM* 55.10, pp. 78–87.

Frankel, Jefffrey A. and David Romer (1999). "Does trade affect growth?" In: *American Economic Review* 89.3, pp. 379–399.

Gates, Scott et al. (2012). "Development Consequences of Armed Conflict". In: *World Development* 40.9, pp. 1713–1722.

Gleditsch, Nils Petter et al. (2002). "Armed conflict 1946-2001: A new dataset". In: *Journal of peace research* 39.5, pp. 615–637.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* New York, NY: Springer.

Hegre, Håvard et al. (2001). "Toward a Democratic Civil Peace? Democracy, Political Change, and Civil War, 1816–1992". In: *American Political Science Review* 95.1, pp. 33–48.

Klein Goldewijk, Kees et al. (2011). "The HYDE 3.1 spatially explicit database of human-induced global land-use change over the past 12,000 years". In: *Global Ecology and Biogeography* 20.1, pp. 73–86.

Lights, DMSP-OLS Nighttime (2011). *Image and data processing by NOAA's National Geophysical Data Center. DMSP data collected by the US Air Force Weather Agency.*

Lujala, Paivi (2009). "Deadly combat over natural resources gems, petroleum, drugs, and the severity of armed civil conflict". In: *Journal of Conflict Resolution* 53.1, pp. 50–71.

Lujala, Päivi, Nils Petter Gleditsch, and Elisabeth Gilmore (2005). "A Diamond Curse? Civil War and a Lootable Resource". In: *Journal of Conflict Resolution* 49.4, pp. 538–562.

Lujala, Päivi, Jan Ketil Rod, and Nadja Thieme (2007). "Fighting over oil: Introducing a new dataset". In: *Conflict Management and Peace Science* 24.3, pp. 239–256.

Meiyappan, Prasanth and Atul K Jain (2012). "Three distinct global estimates of historical land-cover change and land-use conversions for over 200 years". In: *Frontiers of Earth Science* 6.2, pp. 122–139.

Montgomery, Jacob M, Florian M Hollenbach, and Michael D Ward (2012). "Improving predictions using ensemble Bayesian model averaging". In: *Political Analysis* 20.3, pp. 271–291.

Montgomery, Jacob M and Brendan Nyhan (2010). "Bayesian model averaging: Theoretical developments and practical applications". In: *Political Analysis* 18.2, pp. 245–270.

Montgomery, Jacob M., Michael D. Ward, and Florian M. Hollenbach (2015). "Calibrating ensemble forecasting models with sparse data in the social sciences". In: *International Journal of Forecasting.*

Polity, IV (2014). "Polity IV: Regime Authority Characteristics and Transitions Datasets". In: *Center for Systemic Peace, Colorado State University.*

Raftery, Adrian E. et al. (2005). "Using Bayesian Model Averaging to Calibrate Forecast Ensembles". In: *American Meteorological Society* 133.1, pp. 1155–1173.

Storeygard, Adam et al. (2008). "The global distribution of infant mortality: a subnational spatial view". In: *Population, space and place* 14.3, pp. 209–229.

Sundberg, Ralph and Erik Melander (2013). "Introducing the UCDP Georeferenced Event Dataset". In: *Journal of Peace Research* 50.4, pp. 523–532. DOI: 10.1177/0022343313484347.

Tollefsen, Andreas Forø (2012). *PRIO-GRID Codebook.* Typescript, PRIO. URL: http://file.prio.no/ReplicationData/PRIO-GRID/PRIO-GRID_codebook_v1_01.pdf.

Tollefsen, Andreas Forø, Håvard Strand, and Halvard Buhaug (2012). "PRIO-GRID: A unified spatial data structure". In: *Journal of Peace Research* 49.2, pp. 363–374. DOI: 10.1177/0022343311431287. eprint: http://jpr.sagepub.com/content/49/2/363.full.pdf+html.

Uchida, Hirotsugu (2009). "Agglomeration index: towards a new measure of urban concentration". In:

Vogt, Manuel et al. (2015). "Integrating Data on Ethnicity, Geography, and Conflict The Ethnic Power Relations Data Set Family". In: *Journal of Conflict Resolution*, p. 0022002715591215.

Ward, Michael D., Brian D. Greenhill, and Kristin M. Bakke (2010). "The perils of policy by p-value: Predicting civil conflicts". In: *Journal of Peace Research* 47.4, pp. 363–375.

Wei, Shang-Jin. *Natural openness and good government.* Working Paper, 7765. Cambridge, MA: National Bureau of Economic Research.

Weidmann, Nils B, Doreen Kuse, and Kristian Skrede Gleditsch (2010). "The geography of the international system: The CShapes dataset". In: *International Interactions* 36.1, pp. 86–106.

World Bank (2015). *World Development Indicators.* URL: http://data.worldbank.org/products/wdi.

Zipf, G. K. (1949). *Human Behavior and the Principle of Least Effort.* Cambridge, MA: Addison-Wesley.

A Appendix

A.1 R model specifications

```
A_DistL1 <- gam(outcome ~ s(cdistl1), data=train, family=binomial(link="logit"))
A_MaDist <- gam(outcome ~ s(ma12cdist), data=train, family=binomial(link="logit"))
A_G1500BCD <- gam(outcome ~ s(spillover10), data=train, family=binomial(link="logit")
    )
A_xycoord <- gam(outcome ~ te(xcoord,ycoord) + s(ln_pop), data=train, family=binomial
    (link="logit"))
A_HistlagsQ <- glm(outcome ~ outcome_l1 + outcome_l2 + outcome_l3 + outcome_ml4_12 +
    anyneigh_l1 + anyneigh_l2 + anyneigh_l3 + anyneigh_ml4_12 + ln_pop, family=
    binomial(link="logit"), data=train)
A_Histlags_gam <- gam(outcome ~ s(outcome_time_cor) + s(neigh_outcome_time_cor) +
    ln_pop, family=binomial(link="logit"), data=train)
A_HistSeverity <- gam(outcome ~ s(deaths_a_12) + s(deaths_b_12) + s(
    deaths_civilians_12) + s(neigh_deaths_a_12) + s(neigh_deaths_b_12) +s(
    neigh_deaths_civilians_12) + s(ln_pop), family=binomial(link="logit"), data=train
    )

B_Accessibility <- glm(outcome ~ outcome_l1 + neigh_outcome_l1 + ln_pop + ln_capdist
    + ln_bdist1 + nlights_calib_mean + urban_gc + ttime_mean + mountains_mean, family
    =binomial(link="logit"), data=train)
B_Accessibility_gam <- gam(outcome ~ outcome_l1 + neigh_outcome_l1 + te(ln_pop,
    ttime_mean) + s(mountains_mean), family=binomial(link="logit"), data=train)
B_Poverty <- glm(outcome ~ outcome_l1 + neigh_outcome_l1 + ln_pop + imr_mean +
    cmr_mean + nlights_calib_mean + urban_gc, family=binomial(link="logit"), data=
    train)
B_Poverty_gam <- gam(outcome ~ outcome_l1 + neigh_outcome_l1 + te(ln_pop, imr_mean) +
     cmr_mean + nlights_calib_mean + urban_gc, family=binomial(link="logit"), data=
    train)

% B_RF_250tree <- randomForest(outc_f ~ outcome_l1 + neigh_outcome_l1 +
   outcome_time_cor + neigh_outcome_time_cor + ma12cdist + ln_pop + ln_capdist +
   ln_bdist1 + nlights_calib_mean + urban_gc + ttime_mean + mountains_mean + imr_mean
   + cmr_mean + i_pasture_ih + i_agri_ih + i_barren_ih + i_forest_ih + i_savanna_ih +
   excluded, ntree=2000, data=train, na.action=na.omit)
C_Drought <- glm(outcome ~ droughtyr_speibase + droughtcrop_speibase +
    droughtyr_speibase_l1 + droughtcrop_speibase_l1, family=binomial(link="logit"),
    data=train)
```

Figure A-1: Dyadic relations between a cell centroid and UCDP GED events in 2007

```
C_Resources <- glm(outcome ~ diamprim_s + diamsec_s + gem_s + goldplacer_s +
    goldvein_s + goldsurface_s + petroleum_s, family=binomial(link="logit"), data=
    train)
C_Landscape <- glm(outcome ~ i_pasture_ih + i_agri_ih + i_barren_ih + i_forest_ih +
    i_savanna_ih, family=binomial(link="logit"), data=train)
C_Landscape_gam <- gam(outcome ~ s(i_pasture_ih) + s(i_agri_ih) + s(i_barren_ih) + s(
    i_forest_ih) + s(i_savanna_ih), family=binomial(link="logit"), data=train)
C_Excluded <- glm(outcome ~ as.factor(excluded), family=binomial(link="logit"), data=
    train)

D_CL_history <- glmer(outcome ~ ma12cdist + ln_pop + (1|gwno) + (ltsc0|gwno), data=
    train, family=binomial(link="logit"))
D_CL_democracy <- glmer(outcome ~ ma12cdist + ln_pop + (1|gwno) + (polity2sq|gwno) ,
    data=train, family=binomial(link="logit"))
D_CL_regch <- glmer(outcome ~ ma12cdist + ln_pop + (1|gwno) + (proxregch|gwno) , data
    =train, family=binomial(link="logit"))
D_CL_acc_poverty <- glmer(outcome ~ outcome_l1 + neigh_outcome_l1 + ln_pop + imr_mean
     + cmr_mean + nlights_calib_mean + urban_gc + ln_capdist + ln_bdist1 + ttime_mean
     + mountains_mean + (1 | gwno), family=binomial(link="logit"), data=train)
```

A.1.1 Independent variables and models

A.2 A 'conflict gravity field'

A catch-all proxy for many such influences relevant for conflict outcomes is previous conflict history. With monthly grid-cell observations, this has to go beyond a simple one-step temporal lag and

Table A-1: PRIO-Grid variables

Varname	Cite	Description
cmr_mean		Gives the average prevalence of child malnutrition within the grid cell. Based on raster data from the SEDAC Global Poverty Mapping project.
diamprim_s	Lujala, Gleditsch, and Gilmore (2005)	A dummy variable for whether primary (kimberlite) diamond deposits have been found within the given grid cell, based on the Diamond Resources dataset v1a.
diamsec_s	Lujala, Gleditsch, and Gilmore (2005)	A dummy variable for whether secondary (alluvial) diamond deposits have been found within the given grid cell, based on the Diamond Resources dataset v1a.
droughtcrop_speibase	BEGUERíA, Vicente Serrano, and Angulo-Martínez (2010)	Gives the proportion of months in the growing season that are part of the longest streak of consecutive months in that growing season with SPEI-1 values below -1.5. , it uses the Standardized Precipitation and Evapotranspiration Index SPEI-1 from the SPEIbase v.2.3.
droughtyr_speibase	BEGUERíA, Vicente Serrano, and Angulo-Martínez (2010)	Gives the proportion of months out of 12 months that are part of the longest streak of consecutive months ending in the given year with SPEI-1 values below -1.5., as defined by the Standardized Precipitation and Evapotranspiration Index SPEI-1 from the SPEIbase v.2.3.
droughtyr_speibase_l1		droughtyr_speibase lagged by 1 month.
excluded	Vogt et al. (2015)	Counts the number of excluded groups (discriminated or powerless) as defined in the GeoEPR/EPR data on the status and location of politically relevant ethnic groups settled in the grid cell for the given year, derived from the GeoEPR/EPR 2014 update 2 dataset.
gem_s	Lujala (2009)	Is a dummy variable for whether gem deposits have been found within the given grid cell, based on the GEMDATA dataset.
goldplacer_s	Balestri (2012)	Is a dummy variable for whether placer gold deposits have been found within the given grid cell, based on the GOLDATA_L subset of the GOLDDATA v1.2.
goldsurface_s	Balestri (2012)	Is a dummy variable for whether surface gold deposits have been found within the given grid cell, based on the GOLDATA_S subset of the GOLDDATA v1.2.
goldvein_s	Balestri (2012)	Is a dummy variable for whether vein gold deposits have been found within the given grid cell, based on the GOLDATA_NL subset of the GOLDDATA v1.2.
i_agri_ih	Meiyappan and Jain (2012)	Gives the percentage area of the cell covered by agricultural area, based on ISAM-HYDE landuse data. Years between actual observations are linearly interpolated, denoted by the i_ preceeding the variable name.
i_barren_ih	Meiyappan and Jain (2012)	Gives the percentage area of the cell covered by barren area, based on ISAM-HYDE landuse data. Years between actual observations are linearly interpolated, denoted by the i_ preceeding the variable name.
i_forest_ih	Meiyappan and Jain (2012)	Gives the percentage area of the cell covered by forest area, based on ISAM-HYDE landuse data. Years between actual observations are linearly interpolated, denoted by the i_ preceeding the variable name.
i_pasture_ih	Meiyappan and Jain (2012)	Gives the percentage area of the cell covered by pasture area, based on ISAM-HYDE landuse data. In PRIO-GRID, this indicator is available for the years 1950, 1960, 1970, 1980, 1990, 2000, and 2010. Years between actual observations are linearly interpolated, denoted by the i_ preceeding the variable name.
i_savanna_ih	Meiyappan and Jain (2012)	Gives the percentage area of the cell covered by grasslands, based on ISAM-HYDE landuse data. Years between actual observations are linearly interpolated, denoted by the i_ preceeding the variable name.
imr_mean	Storeygard et al. (2008)	Measures mean infant mortality rate, based on raster data from the SEDAC Global Poverty Mapping project.
ln_bdist1	Weidmann, Kuse, and Gleditsch (2010)	Gives the spherical distance in kilometer from the cell centroid to the border of the nearest land-contiguous neighboring country, based on country border data using cShapes v.0.4-2. Included in natural logarithm form, denoted by the ln_ preceeding the variable name.
ln_capdist	Weidmann, Kuse, and Gleditsch (2010)	Gives the spherical distance in kilometers from the cell centroid to the national capital city in the corresponding country, based on coordinate pairs of capital cities derived from the cShapes dataset v.0.4-2. It captures changes over time wherever relevant. Figure 3 visualizes these straight-line distances.
ln_pop	Klein Goldewijk et al. (2011)	Measures the population size for each populated cell in the grid, taken from the History Database of the Global Environment (HYDE) version 3.1. Population estimates are available for 1950, 1960, 1970, 1980, 1990, 2000, and 2005. The original pixel value is number of persons. Included in natural logarithm form, denoted by the ln_ preceeding the variable name.
mountains_mean	Blyth (2002)	Measures the proportion of mountainous terrain within the cell based on elevation, slope and local elevation range, taken from a high-resolution mountain raster developed for UNEP's Mountain Watch Report.
nlights_calib_mean	Lights (2011)	Measures average nighttime light emission from the DMSP-OLS Nighttime Lights Time Series Version 4 (Average Visible, Stable Lights, & Cloud Free Coverages).
petroleum_s	Lujala, Rod, and Thieme (2007)	Is a dummy variable for whether onshore petroleum deposits have been found within the given grid cell, based on the Petroleum Dataset v.1.2.
xcoord	Tollefsen (2012)	Denotes the longitude coordinate (decimal degrees) for the centroid of the grid cell.
ycoord	Tollefsen (2012)	Denotes the latitude coordinate (decimal degrees) for the centroid of the grid cell.
ttime_mean	Uchida (2009)	Is an estimate of the travel time to the nearest major city, derived from a global high-resolution raster map of accessibility developed for the EU.
urban_gc	Bontemps et al. (2011)	Measures the coverage of urban areas in each cell, based on the Globcover 2009 dataset v.2.3.

Table A-2: Variable definitions

Varname	Cite	Description
anyneigh_l1	Croicu and Sundberg (2015)	Any conflict event, either onesided, statebased or nonstate, in the neighboring grid cell. Lagged by 1 month.
anyneigh_l2	Croicu and Sundberg (2015)	Any conflict event, either onesided, statebased or nonstate, in the neighboring grid cell. Lagged by 2 months.
anyneigh_l3	Croicu and Sundberg (2015)	Any conflict event, either onesided, statebased or nonstate, in the neighboring grid cell. Lagged by 3 months.
anyneigh_ml4_12	Croicu and Sundberg (2015)	Sum of any conflict events in the cell, either onesided, statebased or nonstate, in the neighboring grid cell during the previous 4-12 months.
cdistl1	Croicu and Sundberg (2015)	One month lag of distance to closest conflict event.
deaths_a_12	Croicu and Sundberg (2015)	Conflict deaths in neighboring cells for side A as coded by UCDP GED. Summed over the last 12 months.
deaths_b_12	Croicu and Sundberg (2015)	Conflict deaths in neighboring cells for side B as coded by UCDP GED. Summed over the last 12 months.
deaths_civilians_12	Croicu and Sundberg (2015)	Conflict deaths in neighboring cells for civilians as coded by UCDP GED. Summed over the last 12 months.
ma12cdist	Croicu and Sundberg (2015)	12 month moving average of cdistl2,
neigh_deaths_a_12	Croicu and Sundberg (2015)	Conflict deaths in neighboring cells for side A as coded by UCDP GED. Summed over the last 12 months.
neigh_deaths_b_12	Croicu and Sundberg (2015)	Conflict deaths in neighboring cells for side B as coded by UCDP GED. Summed over the last 12 months.
neigh_deaths_civilians_12	Croicu and Sundberg (2015)	Conflict deaths in neighboring cells for civilians as coded by UCDP GED. Summed over the last 12 months.
neigh_outcome_l1	Croicu and Sundberg (2015)	One month lag of conflict event in neighboring cell.
neigh_outcome_time_cor	Croicu and Sundberg (2015)	Months since last conflict event in a neighboring cell.
outcome_l1	Croicu and Sundberg (2015)	One month lag of conflict event in the cell.
outcome_l2	Croicu and Sundberg (2015)	Two month lag of conflict event in the cell.
outcome_l3	Croicu and Sundberg (2015)	Three month lag of conflict event in the cell.
outcome_ml4_12	Croicu and Sundberg (2015)	Sum of conflict events in the cell during the previous 4-12 months.
outcome_time_cor	Croicu and Sundberg (2015)	Months since last conflict event in the cell.
spillover10	Croicu and Sundberg (2015)	Gravity, conflict events within 1500
lngdpcap_c	World Bank (2015)	World Bank GDP per capita (NY.GDP.PCAP.CD) in current US$. Observed at the country level as denoted by the suffix _c. Included in natural logarithm form, denoted by the ln preceeding the variable name.
lnpop_c	World Bank (2015)	World Bank total country population (SP.POP.TOTL). Total population is based on the de facto definition of population, which counts all residents regardless of legal status or citizenship. Included in natural logarithm form, denoted by the ln preceeding the variable name.
polity2	Polity (2014)	Polity score
polity2sq	Polity (2014)	Polity score squared
proxregch	Polity (2014)	Exponential function of negative time since regime change divided by 2. Regime change is defined as a one year change in polity score of at least two.
ltsc0	Gleditsch et al. (2002)	Captures country level conflict history. Defined as one year lag of the natural log of time in status, status being conflict or abscence thereof. Equals log of years since independence in countries with no recorded conflict. Equals 0 if the country saw conflict the previous year. Equals log of the years since the previous conflict in first year of a conflict.

Table A-3: Definitions of variables

VarName	Description
L1	1st order temporal lag within the same cell.
QL1	Queen's case 1st order spatial lag at $t-1$
DistL1	Shortest distance to a conflict event occuring at $t-1$
MA12DistL1	The average shortest distance to a conflict event, each month the last 12 months (not including current month).
G4000	Gravity formulation, only including events within 4000 km (but including all previous events since 1989).
G1500	Gravity formulation, only including events within 1500 km
G500	Gravity formulation, only including events within 500 km
G250	Gravity formulation, only including events within 250 km
G4000GWNO	Gravity formulation, only including events within 4000 km and within same country as cell.
G1500A1B1C2	Gravity formulation, only including events within 1500 km. $\alpha_a = 1$, $\alpha_b = 1$ and $\alpha_c = 2$ – distance in time is discounted more than distance in space.
G1500A1B2C2	Gravity formulation, only including events within 1500 km. $\alpha_a = 1$, $\alpha_b = 2$ and $\alpha_c = 2$ – the intensity of past events is given less weight relative to distance in time and space.
G1500A2B1C1	Gravity formulation, only including events within 1500 km. $\alpha_a = 2$, $\alpha_b = 1$ and $\alpha_c = 1$ – the intensity of past events is given more weight relative to distance in time and space.
G1500CBD	Gravity formulation, only including events within 1500 km. Battle-related deaths set to 1 if 1 or higher, i.e. $\alpha_a \to 0$ if there is at least one fatality in the past event.

beyond the immediate neighbor. Conflict actors might lie dormant for a couple of months before making a new attack, and the next attack might be more than 50-100 km from the last attack.

We think of the following data-generating process underlying what we are measuring within the grid-cell month. The UCDP-GED dataset gives us "sensor" readings at the village-day level (Sundberg and Melander, 2013).[6] We denote each PRIO-GRID cell centroid as i and each day within a month as t (Tollefsen, Strand, and Buhaug, 2012). Each i is mapped to one and only one gridcell I, and each t is mapped to one month T. We can imagine we have three sensors at each point, one for each of the three latent variables y^*_{1it}, y^*_{2it} or y^*_{3it} (the types of violence).

The sensors only fire when the values are greater than three associated thresholds T_1, T_2, T_3, at time t in location i. Thus only when $y^*_{jit} > T_j$ do we get an event, y_{jit} (this is observed). We assume the thresholds are constant and that time is discrete. What we are measuring with gridcells and months are the summation of events over the discrete days and grid points $y_{jIT} = \sum_{i \in I} \sum_{t \in T} (y_{jit}),\ \forall\ j \in (1, 3)$. We can define the latent propensity for counts in the cells as $y^*_{jIT} = \sum_{i \in I} \sum_{t \in T} (y^*_{jit}),\ \forall\ j \in (1, 3)$

Of course, we could just add nth-order lags both in time and space, but to get the same coverage as a single t-1 lag at the country-year level, we would need an unmanageable number of terms. For a country which is 11*11 cells large, for instance, we would need 5*12 = 60 dummy variables. Below, we present one approach to simplify this structure. This also facilitates linking the results to identifiable theoretical accounts of conflict diffusion.

Here, we explore dependence on past conflict along three axes: intensity (measured through battle-related deaths), distance (in km) and time (in days since an event). The basic construct is the dyad between any PRIO-GRID cell centroid in Africa (Tollefsen, Strand, and Buhaug, 2012), and any UCDP-GED 3.0 event that happened at a time before t, where t is any month from 1989–2014 (Sundberg and Melander, 2013).[7] In Figure A-1, we have plotted all UCDP GED events in 2007 and their dyadic relations with a single PRIO-GRID cell located in Northern Nigeria. From such a dyadic list, it is possible to define several alternative measures, such as shortest distance to a conflict event within a specific time period, or shortest time since a conflict within a specific distance. Interestingly, by testing different such measurements for out-of-sample performance, it might also be possible to learn something about general properties of spatio-temporal dependence of previous conflicts. For instance, how far back in time are conflict events relevant for predictive power, how far in space? Even more interesting is that we can include information about the conflict events themselves, such as battle-deaths, or even information about the surface-area between the cell centroid and the conflict event, or dyadic information of which areas that are in close relation (such as whether or not the cell centroid and the conflict event happened in the same country).

[6]This holds in the majority of UCDP-GED events where UCDP has access to adequate information. For a substantial minority of observations, information may be on a more coarse level, but the data-generating process is in principle the same also for these.

[7]A dyadic list only measuring distances (not accounting for time) leads to more than 1 billion entries and take up 60Gb of storage in PostgreSQL.

A model for 'conflict exposure' to PRIO-GRID cell i from the environment based on the gravity model of interaction (Zipf, 1949):[8] The gravity model stipulates that the volume of interaction V_{ij} between locations (grid cells) i, j is proportional to

$$V_{ij} = \frac{P_i^{\alpha_a} P_j^{\alpha_b}}{D_{ij}^{\alpha_c}}$$

where $P_{i,j}$ are the populations of the two cells and D_{ij} the distance between them. D_{ij} is initially just geographical distance but could be made more complicated later (accounting for international borders, travel costs, etc), but then we need to estimate weights between different concepts of distance. In log form,

$$ln(V_{ij}) = \alpha_a ln(P_i) + \alpha_b ln(P_j) - \alpha_c ln(D_{ij})$$

Initially, we may assume that all α terms are 1, as is often the case in applications.

Conflict spillover from cell j to cell i is then proportional to the per-capita intensity f_j of conflict in j and the gravity model of interaction:

$$ln(CS_{ij}) \sim ln(f_j) + ln(P_j) - ln(D_{ij}) = ln(F_j) - ln(D_{ij})$$

where $ln(F_j) = ln(f_j) + ln(P_j)$, the total number of fatalities in cell.

Total exposure CE_i to conflict spillover, then, is

$$CE_i \sim \sum_{j=0}^{J} (CP_{ij}) = \sum_{j=0}^{J} exp((ln(F_j) - ln(D_{ij}))$$

On might generalize D_{ij} also to take time T since events into account; the conflict spillover from cell j at time t is then

$$ln(CS_{ijt}) \sim ln(F_{jt}) - ln(D_{ijt}) - ln(T_{ijt})$$

In the following section, we define the gravity measurement as

$$CE_{it} \sim \sum_{j=0}^{J} (CP_{ijt}) = \sum_{j=0}^{J} exp((\alpha_a ln(F_{jt}) - \alpha_b ln(D_{ijt}) - \alpha_c ln(T_{ijt}))$$

Unless stated otherwise, all α terms are set to 1. In some formulations, we change them to allow that 'distances' in time might be weighted differently than distances in space.

[8] Inspired by the concept of 'natural openness' in Frankel and Romer (1999) and 'remoteness' in Wei (*Natural openness and good government.*).