STEPPING INTO THE FUTURE: THE NEXT GENERATION OF CRISIS FORECASTING MODELS

MICHAEL D. WARD, NILS W. METTERNICH, CHRISTOPHER CARRINGTON, CASSY DORFF, MAX GALLOP, FLORIAN M. HOLLENBACH, ANNA SCHULTZ, AND SIMON WESCHLE

Abstract. Developing political forecasting models is not only relevant for scientific advancement, but also increases the ability of political scientists to inform public policy decisions. Taking this perspective seriously, the International Crisis Early Warning System (ICEWS) was developed under a DARPA initiative to provide predictions of international crisis, domestic crisis, rebellion, insurgency, and ethnic violence (Events of Interest/EOIs) in about two-dozen countries in the US PACOM Area of Responsibility. As part of a larger project coordinated by Lockheed Martin Advanced Technology Labs, a team at Duke University created a series of geographically informed statistical models for these EOIs. The generated predictions have been highly accurate, with few false negative and positive categorizations. Predictions are made at the monthly level for three months periods into the future. The major variables to generate the predictions include 1) event data culled from FACTIVA reports, 2) structural political characteristics of states, 3) economic characteristics of states, and 4) contextual features of each country. These later characteristics take into account the social-spatial context of each individual country, thereby allowing the models to escape the limitations of treating each country as independent from the influence of events and forces in nearby countries. For each EOI we present a separate prediction model, which captures the unique dynamics of each outcome. Each of these models has a high degree of accuracy in reproducing historical data measured monthly over the past 10 years, and is approximately equally accurate in making three-month forecasts out-of-sample.

1. Introduction

The Arab Spring and its aftermath, continuing violence in Afghanistan, and sudden agreements between Georgia and Russia always raise the same question: could we have predicted these events to allow policy actions to prepare for or even influence their emergence. While our profession is very good fabricating *ex post* explanations for almost any past crisis, we struggle to predict what seems obvious after the fact. Having observed the Arab Spring, the regional contagion effects only seem logically. After ten years of military operations in Afghanistan, there are hundreds of accounts of why the fighting continues. However, we have to be aware that these detailed "obvious" explanations and interpretations are rationalized *ex post*. Would the Arab Spring have turned into a success of the dictators, instead of a victory of the people, the same evidence that is now used to explain the success of protesters (e.g., social media) would be cited as their demise (e.g., too much reliance on social media instead of actual military force). However, relying on *ex post* explanations hazes and biases our ability to find common and underlying explanations for domestic and international crises. *Ex post* explanations have a tendency to highlight the unusual and particular of a crisis. For example, parts of the Arab Spring were organized by social media, hence many commentators focused on this novel tool as an explanation for the success of the revolution. However, we do not know what would have happened without social media. In fact, there were a host of other factors that made revolutions in Arab world highly likely. High unemployment, low growth rates, aging dictators, or religious divisions are long standing explanations for revolutions that were all present in this context.

This article is a reminder that despite the fact that every crisis has their own unique features, we as a discipline should strife for the main prize, that is the identification of general mechanisms that will allow us to make predictions about future events. In fact, the ability to predict future crises can be understood as the gold standard to scientifically advance the study of conflict, peace, and crises.

In political science, prediction is typically conceptualized as a conditional exercise, in which values on a dependent variable are calculated based on some estimated, or conditional, statistical model, and then compared with the actual observed values (Hildebrand, Laing and Rosenthal, 1976). But there is also a recent tradition of attempting to make political predictions about things that have not yet occurred, in the sense that the *Old Farmer's Almanac*, published continuously since the late 18th Century, predicts the weather for the coming year (as well as fashion trends). An early proponent of using statistical models for making such predictions in the realm of international relations was Stephen Andriole, a research director at ARPA in the late 1970s (Andriole and Young, 1977). In 1978, a volume edited by Nazli Choucri and Thomas Robinson provided an overview of the then current work in forecasting in international relations, much of which was done in the context of policy oriented research for the U.S. government during the Vietnam War.[1] There were a variety of efforts to forcast or evaluate forecasting efforts, including Freeman and Job (1979), Singer and Wallace (1979), & Vincent (1980), and a few efforts began to forecast internal conflict (Gurr and Lichbach, 1986), but the median empirical article in political science (as well as sociology and economics) used predictions only in the sense of in-sample observational studies.[2] Doran (1999) and others provided some criticism but most scholars avoided making predictions, perhaps because their models had enough difficulty in describing accurately what *had* happened.

Still there were a few scholars that continued to make predictions (yes, about the future), including Gurr, Harff and Harff (1996), Krause (1997), Davies and Gurr (1998), Pevehouse and Goldstein (1999), Schrodt and Gerner (2000), King and Zeng (2001), O'Brien (2002), de Mesquita (2002), Fearon and Laitin (2003), de Marchi, Gelpi and Grynaviski (2004), Enders and Sandler (2005), Leblang and Satyanath (2006), Ward, Siverson and Cao (2007), Brandt, Colaresi and Freeman (2008), Bennett and Stam (2009), and Gleditsch and Ward (2010), among a few others.[3] However, just in the last years the field of conflict forecasting has tremendously expanded. The surge of prediction research in conflict and peace studies can be attributed to the new availability of spatio-temporal disaggregated data and the application of new estimation strategies. Both developments are a result of increasing computational power that allow access to large data sources and the implementation of complex statistical tools.

2. Integrated Crisis Early Warning System

Predicting crises has been a research priority of the US intelligence and warning community for decades.

For the past several years under DARPA funding, a large, multidisciplinary team of computer and social scientists from universities and small businesses developed the Integrated Crisis Early Warning System (ICEWS). ICEWS provides Combatant Command staffs (COCOMs) with highly accurate and timely forecasting of instability events of interest (EOIs) using an innovative combination of computational social science models O'Brien (2010). ICEWS exploits dynamic, high-volume, heterogeneous data sources to drive these models and provide operators with situational awareness of past and current events in countries of interest. Many components of this system are being transitioned into the Integrated Strategic Planning and Analysis Network (ISPAN) program of record at the United States Strategic Command by Lockheed Martin. ICEWS has also been deployed for user testing and evaluation at the Pacific and Southern Combatant commands in 2010 and 2011.

The basic task of the ICEWS project is to produce predictions for five dependent variables, for 29 countries, for every month from 1997 through the present plus three months into the future. The variables in question are rebellion, insurgency, ethnic violence, domestic crisis, and international crisis. The twenty-nine countries are Australia, Bangladesh, Bhutan, Cambodia, China, Comoros, Fiji, India, Indonesia, Japan, Laos, Madagascar, Malaysia, Mauritius, Mongolia, Myanmar, Nepal, New Zealand, North Korea, Papua New Guinea, Philippines, Russia, Singapore, Solomon Islands, South Korea, Sri Lanka, Taiwan, Thailand, & Vietnam. This set not a random sample, but rather constitutes the countries of population greater than $500,000$ that are in the Area of Responsibility of the US Pacific Command (PACOM). The countries in PACOM include about 50% of the total world population, along with five of the largest military powers (China, Russia, India, North & South Korea). The countries range from democratic to authoritarian, from tiny to large, from landlocked to archipelago, and vary widely on almost any social or economic indicator.

Each month we receive a drop of two sets of data. The first of these comprises the five dependent variables in this study. In addition, we receive data for each event that transpires within or involving each of the 29 countries in the sample. These event data are gleaned from natural language processing of a continuously updated harvest of news stories, primarily taken from Factiva™, an open source, proprietary repository of news stories from over 200 sources around the world. The baseline event coder is called JABARI, a java variant of TABARI (Text Analysis By Augmented Replacement Instructions) which has been developed by Philip Schrodt and colleagues (see `http://eventdata.psu.edu/`). It combines a "shallow parsing" technology of prior coders with a richer exploitation of syntactic structure. This has increased accuracy (precision) from 50% to over 70%, as demonstrated in a series of ongoing (informal) evaluations of its output by human graders (peak human coding performance is around 80% (King and Lowe, 2003)).

These data are augmented with a variety of other attribute and network data. In particular we use attributes, coded on a monthly or yearly basis from the Polity, Minorities and Risk (MAR), and World Bank data set. We also include information about the election cycles (if any) in each of the countries. In addition, we use information about relations among the 29 countries, including geography, the length of shared borders, the amount of trade, the movement of people across borders, the number of refugees, as well as the number and types of events between each pair of the 29 countries.

The next section of this paper provides a brief review of the theoretical motivations for each of our EOI models and an overview of the model specifications for each dependent variable. A full explanation of mixed effects models can be found in the section thereafter, followed by a discussion of our main results. Finally, we focus on the predictive power of our models and explore the use of cumulative probabilities to predict the occurrence of any crisis.

3. Modeling events of interest

ICEWS focuses on five EOIs: rebellion, insurgency, ethnic violence, domestic crisis, and international crisis. In modeling each EOI, we draw from relevant literature in political science to suggest a set of conditions under which a specific event is likely to occur. For example, our model of rebellion relies on theoretical insights suggesting that societal conflicts between the government and ethnic/opposition groups are more likely to manifest themselves in situations where democratic institutions are absent or threatened. Rebellions have also been shown to increase the probability of rebellion in surrounding countries. Similarly, the insurgency model is derived from theories highlighting arguments that political exclusion make dissident, ethnic, and religious groups more likely to overcome collective action problems.

3.1. **Rebellion.** Our model for predicting rebellion uses proxies for the level of latent conflict between the government and the opposition, and then models the circumstances under which this latent conflict will lead to rebellions. The proxies are directional measures of the number of conflictual words ("demand", "disapprove", "reject", "threaten") stated from the government towards opposition groups and vice versa. We suggest that the effect of conflict on the probability of rebellion depends on the number of ethnically relevant groups that are excluded from power. When there are no excluded ethnic groups, rebellion should be very unlikely, as disagreements can be solved in the political arena. However, if a large number of excluded groups exist, coordination problems arise, which also mitigate rebellions. Hence, rebellion becomes most likely when few excluded groups exist.

We also include proximity to elections, which can bring about an increase in violence. A recent example is the case of Kenya, where following the victory of incumbent President Mwai Kibaki, the opposition denounced the results and widespread protests led to violence. As Snyder argues, while elections and democracy are often seen as important mechanisms in the peace building process, they can actually increase the likelihood of violence (Snyder, 2000). Additional predictive factors are detailed in Table 1.

3.2. **Insurgency.** Access to power is a key variable to understanding the causes of insurgencies. Insurgencies involve groups attempting to wrest political power from the sitting government, and so groups without access to political power are especially of interest. The larger this excluded population, the more likely violence will be used to change the political landscape. Furthermore, evidence has shown that violence designed to undermine the government is faced with a collective action problem (Kalyvas and Kocher, 2009). However, if anti-government groups observe attacks against the government, they may change their calculus. Thus, we include a measure of dissident groups actions against the government because such actions can be used as a rallying force and recruiting tool, increasing the probability of insurgency. Similarly, it follows that insurgencies in nearby countries may update individuals beliefs about who else will act against the government of their own country. For this reason we include a measure of insurgencies in nearby countries, lagged by three months. We also suggest that nearby insurgencies could potentially disrupt effective government repression, liberating sources of weapons, money, and information for would-be insurgents in the target country.

3.3. **Ethnic Violence.** While most quantitative studies focus on the effect ethnicity has on conflicts between rebels and the government, we are interested in inter-ethnic and inter-religious violence. Thus, our concept of ethnic violence matches ideas of non-state war (Sarkees and Wayman, 2010), non-state conflicts (Eck, Kreutz and Sundberg, 2010), or subnational wars (Chojnacki, 2006), where the primary involved actors are non-state actors. In line with recent work on ethnic conflicts (Cederman, Wimmer and Min, 2010),), we argue that government policies play an important role in explaining these dynamics. Thus, in our models, we include the number of politically excluded ethnic groups in a country and the overall proportion of the excluded ethnic population. The existing literature also points to a polarization effect of political exclusion, which suggests including the squared term of the proportion excluded. In addition, we argue that periods of political transition increase incentives to lock in political power in future institutions. Hence, we include Polity and its squared term to model political transition periods (Hegre et al., 2001).

Finally, we are interested in the spatial component of ethnic conflict. An increasing number of scholars not only highlight the transnational dimensions of civil conflict (Gleditsch, 2007; Buhaug and Gleditsch, 2008; Salehyan, 2007, 2009), but also its ethnic component (Cederman, Buhaug and Rød, 2009; Cederman, Girardin and Gleditsch, 2009). Thus, our model takes into account possible spillover effects from neighboring countries.

3.4. **Domestic Crises.** Domestic violence and protests are frequently triggered by elections that were perceived to be unfair. We include proximity of elections in our model, with different effects depending on the level of executive constraints. We propose this approach because in countries with moderate levels of executive constraints, elections have meaningful implications regarding who holds office, but governments have the latitude to manipulate the elections and therefore domestic crises are more likely to center around elections. A second major factor that we believe affects the propensity of domestic crises onsets is a countrys ethnic composition. When ethnic groups are excluded from political processes grievances are likely to arise. In authoritarian systems this effect is likely to differ from democracies, so in our model, the effect of the number of excluded groups varies by executive constraints. Hence, the likelihood of domestic conflict is conditional on different levels of executive constraints, with the coefficients for the proximity to elections and the number of excluded groups also varying by executive constraints. In addition to the random effects for proximity to election and number of excluded ethnic groups, we control for GDP per capita, population size, and a spatial lag of domestic crises.

3.5. **International Crises.** Our model of international crises tries to capture those situations when a leader is unable or unwilling to make the necessary concessions to avoid a crisis. A leaders incentives to avoid international crises will be conditional on domestic political institutions. Leaders in more democratic regimes may be less able to make concessions internationally due to threat of domestic audience costs. The costs of a crisis might be lower for leaders of more autocratic regimes since their constituency will not bear the brunt of any potential fighting (Bueno de Mesquita et al., 2003; Schultz, 2001). To account for systematic differences between the prevalence of crises under different regime types, the model includes a random intercept based on a countrys democracy polity score. In addition, homogeneous populations impose few constraints on the bargaining of leaders in democracy. So, the model also includes a random effect for the number of politically relevant ethnic groups conditional on level of democracy. We also control for population size, international crises in politically similar states and include measures for both domestic political pressure and domestic conflict.

4. Methods: Predicting Events of Interest Using Mixed Effects Models

Hierarchical models with random slopes and random intercepts, called mixed effects models, have the ability to provide a general framework for understanding a phenomenon, without at the same time requiring that the coefficients be exactly the same for each and very case. We could in theory model crises in each country being investigated, and then make the coefficients of the models we estimated for each case, the study of further investigation. In a sense there are two (or more) levels of modeling, in this case one at the level of the country (e.g., Japan over time), and a second at the regional level (e.g., the 29 countries in the US Pacific Command or the groups of democracies and autocracies). Hierarchical models also keep track of the variation across the levels (i.e., between the groupings). This approach allows us a) to learn about processes that may vary slightly from one place or time to another, b) use all the data while compromising between within grouped estimates that are highly uncertain because they are based on averages and the more precise individual estimates that plausibly ignore influences that occur at the level of a group, and c) keep trace of the uncertainty and co-variation across the different levels. As an example, it may well be that accumulated inequalities tend to be associated with rebellious onsets in a fairly predictable way, but that this relationship is perhaps slightly different for monarchies than it is in dictatorships. A simple way to model this is with an interaction term, but that ignores the cross-level variation that may occur between the level of the grouping (dictators may get a lot of foreign aid) and the individual effects within each country. See Gelman and Hill (2007) for a more complete statement of the benefits of this approach; Pinhiero and Bates (2002) and the draft materials on the forthcoming volume by Doug Bates at `http://lme4.r-forge.r-project.org/book/Ch1.pdf`.

We model rebellion, domestic conflict, and international conflict using hierarchical models in which both the intercept and slope vary. Simply stated, this means that we group the data along an indicator, such as level of executive constraints, creating a different intercept for each group. Thus, the varying intercepts correspond to group indicators and the varying slopes represent an interaction between predictor variables x and the group indicators:

$$\Pr(y_{it} = 1) = \text{logit}^{-1}(\alpha_{j[it]} + \beta^G_{j[it]} x^G_{it} + \beta^O_{j[it]} x^O_{it} + Z_{it}\gamma) \tag{1}$$

$$\begin{pmatrix} \alpha_j \\ \beta^G_j \\ \beta^O_j \end{pmatrix} \sim N \begin{pmatrix} \mu_\alpha \\ \mu^G_\beta \\ \mu^O_\beta \end{pmatrix}, \Sigma \Bigg) \tag{2}$$

where i denotes the countries, t the month and j the grouping variable, α_j are the grouping variable's random intercepts. x^G_{it} and x^O_{it} are predictor variables; β^G_j and β^O_j are the associated random coefficients; γ is a vector of fixed effects associated with Z_{it}. Table 1 provides an overview over all model specifications.

For an illustrative example, this equation accurately presents our model of rebellion where i denotes the countries, t the month and j the grouping in which a country falls with respect to the number of excluded ethnic groups. α_j are the grouping-specific random intercepts. x_{it}^G is the number of conflictual words from the government against the opposition in country i at time t and x_{it}^O the number of conflictual words from the opposition against the government; β_j^G and β_j^O are the associated random coefficients. γ is a vector of fixed effects associated with Z_{it}, which is a matrix of other covariates that are commonly found in the literature. All models except ethnic violence take this form, which does not include grouping variables.

5. Results

5.1. **Rebellion.** The results for the random effects are displayed in Table 2. The left column provides the estimates for the group specific intercepts. The second column reports the group specific effects of the number of conflictual words by the government towards the opposition. Finally, the last column in Table 2 shows the group specific effects for the number of conflictual words by the opposition towards the government. The first insight from the random effects is that conflictual interactions between the government and the opposition have a very low probability of escalating to rebellious behavior. We can also see that disagreements between the government and the opposition have the highest probability of turning into rebellions when a country has one or two ethnically relevant groups excluded from power. When there are 3 or more such groups, conflictual words by the government do translate into a higher probability of rebellion, although less so than with one or two groups. However, conflictual words expressed by the opposition do not lead to a higher likelihood of rebellion when there are three or more politically excluded ethnic groups. Finally, the effects of conflictual words for country-months in which there are no ethnic groups excluded from power are close to zero. Overall, these findings provide support for the argument that conflict or disagreement between the government and opposition groups only turns into violence when groups are excluded from power and especially if only few ethnic groups are excluded.

The numerical results for the individual-level effects are shown in Table 3. The proximity to elections only has a small effect on elections and also the z-value is relatively small. Whether there are elections at all, however, does have a strong effect on the probability of rebellion. Rebellions are more likely if executive recruitment is organized by elections rather than

Table 2. *Random Effects Estimates for the Rebellion model*

Grouped by:	Intercept	Gov words	Opp words
Grouped by:	Intercept	→Opp	→ Gov
No excluded groups	-0.33	-0.01	-0.00
1-2 excluded groups	3.22	0.17	0.11
3+ excluded groups	4.95	0.12	-0.01

hereditary succession or designation. One interpretation of this finding is that groups losing elections are likely to use violence to regain political power. There is a lower probability of rebellion when the executive is less constrained, which likely reflects higher oppression levels by the government towards the opposition. The effect of GDP per capita is negative, indicating richer countries are less likely to have rebellions. Finally, countries are less likely to have rebellions if neighboring countries experience them. While this could be a region specific finding, there is some evidence that relate a negative spatial effect to outside support (Dudley and Miller, 1998). Because outside support is a finite resource it may only be available to one rebellion at a time. Furthermore, governments might react to close-by rebellions by increasing their level of oppression, making rebellions less likely.

5.2. **Insurgency.** The main results for the insurgency model are shown in Table 3. As we theoretically expect, the percentage of politically excluded ethnic has a strong effect on the probability of insurgency. However, Figure 1 highlights that this effect is not independent of the number of ethnic groups excluded from political power. When few groups are excluded from power, but these groups represent a large percentage of the overall population the probability of an insurgency is the highest. There is one exception to this finding. When many groups are excluded from political power and they represent a small percentage of the population the probability of insurgency is also relatively high. This could suggest that when divisions of the population into different groups still persist and the excluded population is small, the small population lowers collective action problems within groups, and competition between groups incentivizes actions against the government. In addition, the results show that dissident actions against the government, and insurgencies in nearby countries both have a positive effect on the probability of insurgency.

Hence, ethnic dimensions of insurgency only matter in the context of access to power. Specifically,

TABLE 1. Fixed Effects and Random Effects per Model

	Grouping Variables	Controlled Effects
Rebellion	Excluded groups grouped by: Conflictual words opposition → government Conflictual words gov → opposition	Proximity to election Competitiveness of executive recruitment Executive constraints GDP per capita (log) Rebellions in surrounding countries
Insurgency	Country	Proportion of population excluded Number of excluded ethnic groups Exclusion interaction: population $*$ number of groups High intensity actions: dissidents → government Insurgencies in nearby countries
Ethnic Violence	None	Number of excluded groups Number included groups Proportion of population excluded Squared proportion of population excluded High intensity actions: ethnic groups → government Polity score Squared polity score Violence in neighboring countries
Domestic Conflict	Level of executive constraints grouped by: Number of excluded groups Proximity to election	Population (log) GDP per capita (log) Crises in neighboring countries
International Conflict	Number of ethnic groups grouped by: Level of democracy	Population (log) Domestic EOIs Conflictual words: any domestic group → government International crises in politically similar countries

All effects are lagged three months, except proximity to election.

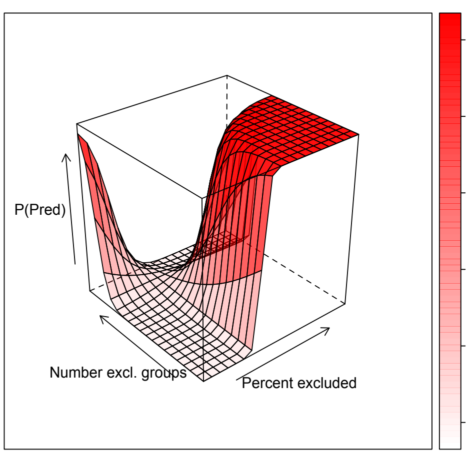


FIGURE 1. Predicted insurgency, by number and divisions in excluded population

as the number of ethnic relevant groups excluded from power increases, the probability of insurgency decreases. When the population of those excluded from power increases, so does the probability of insurgency. However, insurgency is less likely to occur when the excluded population is divided into a large number of groups.

5.3. **Ethnic Violence.** Estimates from the ethnic violence model can be found in Table 3. The model provides support for the theoretical framework that ethnic exclusion and political transition periods are important explanations for ethnic violence. Figure 2 shows that the probability of ethnic violence increases with the number of politically excluded ethnic groups. In addition, Figure 2 highlights the curvilinear relationship between the proportion of politically excluded ethnic groups and the probability of ethnic violence. The probability of ethnic violence is highest in the medium range of observed values for the

Table 3. *Statistical estimates for the Rebellion model, using the monthly test data from 2001-2008.*

Fixed effects:	$\hat{\beta}$	z value	$\hat{\beta}$	z value	$\hat{\beta}$	z value	$\hat{\beta}$	z value	$\hat{\beta}$	z value
Proximity to elections	-0.00	-0.32								
Competitiveness executive recruitment, t-3	0.67	4.37								
Executive constraints, t-3	0.56	5.93								
Log GDP per capita, t-3	-1.11	-8.98								
Rebellion in neighboring countries, t-3	-20.68	-6.51								
Intercept			-22.81	-2.42						
Excluded population			189.60	1.99						
Excluded ethnic groups			3.25	1.83						
Excluded groups * excluded population			-30.94	-1.85						
Dissident high intensity actions$\rightarrow$ Gov.			0.04	2.84						
Insurgencies in nearby countries			12.66	2.81						
Intercept					-4.11	-10.42				
Number of politically excluded groups, t-3					.11	4.14				
Number of politically included groups, t-3					.42	11.19				
Proportion population excluded, t-3					39.35	9.60				
Squared proportion population excluded, t-3					-98.09	-9.40				
High intensity actions of ethnic groups against government, t-3					.14	5.23				
Polity score, t-3					.11	4.74				
Squared polity score, t-3					-.02	-4.66				
Ethnic violence in neighboring countries, t-3					-8.63	-7.83				
Intercept							-9.22	-3.12		
LN Population							0.85	4.9		
LN GDP per Capita							-1.28	-6.38		
Crises in Neighbour Countries							-9.26	-6.02		
Intercept									-56.25	-7.53
LN Population									2.74	6.85
Civil Events of Interest									-2.37	-2.37
Anti-Government Words									0.02	3.64
Spatial Lag of Crises									0.52	3.98

proportion of excluded population. This alludes to polarization effects that drive ethnic violence.

As suggested by our theoretical framework, political transition periods are associated with an increase of ethnic violence. We find a curvilinear relationship between ethnic violence and polity scores that is largely in line with findings for civil war. The political component of ethnic violence is further emphasized by the finding that conflictual interactions between ethnic groups and the government are good predictors of violence between ethnic groups.

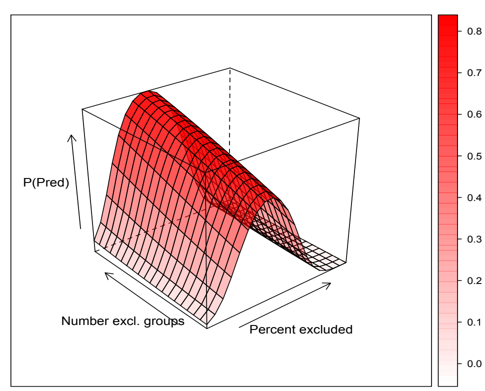


Figure 2. Predicted ethnic violence, by number and percentage of excluded ethnic groups

5.4. **Domestic Crises.** The domestic crises model random effects are provided in Table 4. The random effects vary by nine levels of executive constraint in the countries. The estimates of the fixed effects are shown in Table 3. As expected, larger country populations increase the probability of a domestic crises occurring. Higher GDP per capita, on the other hand, lowers the likelihood of domestic crises. Domestic crises in neighboring countries lower the probability of such an event happening in your own state. The random intercept for executive constraints shows that the likelihood of domestic crisis is lowest at medium levels. Similarly, the number of excluded groups increases the probability of conflict most at low or medium levels of executive constraints. Surprisingly, the closer an election, the less likely a domestic crisis is to occur, although this effect disappears at higher levels of executive constraints. In

TABLE 4. *Group-level Estimates for the Domestic Crises*

Grouped by:	Intercept	Excluded groups	Proximity to election
Level 1	0.03	14.50	-0.012
Level 2	5.88	-0.03	-0.006
Level 3	7.20	-0.80	-0.009
Level 4	-0.69	2.03	-0.006
Level 5	-9.89	4.38	0.008
Level 6	2.19	-0.83	-0.001
Level 7	1.98	-0.23	0.001

TABLE 5. *Random Effects Estimates for International Crises*

Grouped by:	Intercept	Number of ethnic relevant groups
Level 0	11.12	-1.30
Level 1	-0.04	0.00
Level 2	-0.17	0.02
Level 3	-0.06	0.00
Level 4	-0.47	0.01
Level 5	-0.48	0.05
Level 6	11.54	-1.26
Level 7	1.62	-0.62
Level 8	12.22	-2.78
Level 9	-0.73	0.02
Level 10	4.37	-0.77

summary, larger countries experience more domestic crises, as do poorer countries. Executive constraints in interaction with proximity to election or excluded groups have more ambiguous effects.

5.5. **International Crises.** The empirical results from calibrating the model on training data are given in Table 3. International crises result in large part from failure for states to reach a peaceful and mutually beneficial settlement. So, low-level conflict between the government and domestic opponents forces a state to choose between mollifying potentially violent domestic groups and rival states: as we expect, the number of conflictual words by domestic groups will limit a government's freedom of action and lead to crisis. However, if a government chooses not to satisfy domestic groups, we will see domestic events of interest–violence, insurgency, rebellion and domestic crises–but a lowered risk of future international conflict. Finally, we expect domestic institutions and the distribution of power (especially along ethnic lines) to mediate a state's ability to avoid domestic crises, and that a state is more likely to enter an international crisis if current international crises involve politically similar regimes.

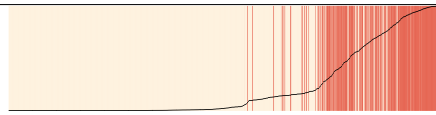
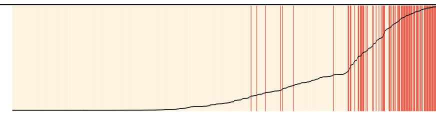
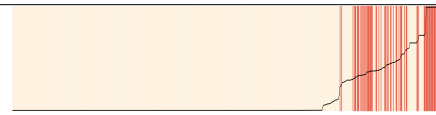
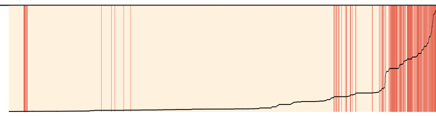
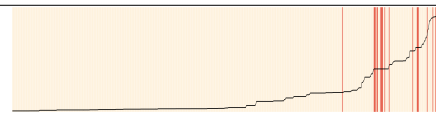
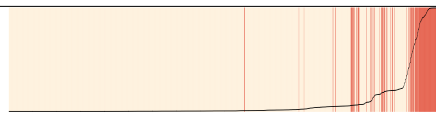
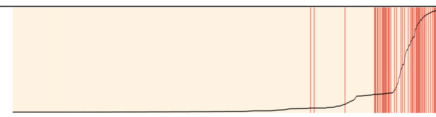
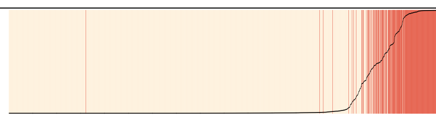
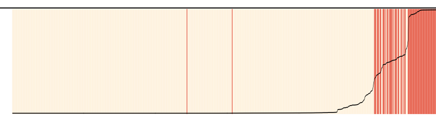
6. Predictive Power

In this section we demonstrate the predictive power of our models and show how to use the individual predictive probabilities to create a cumulative predictive probability of any crisis event occurring. We use separation plots to visualize and assess the predictive power of our models. These plots provide a summary of the fit for each model by demonstrating the range and degree of variation among the predicted probabilities and the degree to which predicted probabilities correspond to actual instances of the event (Greenhill, Ward and Sacks, 2009). Red panels represent events and non-events are left white. The line through the center of the plot represents the expected probability for each model. Thus the probability increases from left to right in the plot and a good fit would be visualized with more red panels (events occurring) stacked at the right end of the plot.

Another way to evaluate predictions that is employed here is the Brier score, defined as the average squared deviation of the predicted probability from the true event (Brier, 1950). It has been shown that the Brier score is one of the few strictly proper scoring rules for predictions with binary outcomes (Gneiting and Raftery, 2007). A Brier scores closer to zero indicate better predictive performance.

In Figure 6 we combine a visual and numeric interpretation of the models to display each of their predictive powers. The separation plots and the provided statistics demonstrate a very good predictive performance of our models. Our model of insurgency serves as a useful illustrative case: in terms of fit, for example, the area under the ROC Curve is 0.98 in-sample and 0.95 out-of-sample. The Brier Score is 0.05 in-sample and 0.06 out-of-sample. The separation plot illustrates the in-sample and out-of-sample fit where probabilities are sorted from low to high in this plot, and country-months with actual occurrences of insurgency are color coded in dark, as opposed to light, colors.

TABLE 6. Predictive Capabilities of All EOI Models

Model	In-sample	Out-of-sample
Rebellion	 AUC=0.96 Brier=0.07	 AUC=0.95 Brier=0.08
Insurgency	 AUC=0.98 Brier=0.05	 AUC=0.95 Brier=0.06
Violence	 AUC=0.90 Brier=0.05	 AUC=0.92 Brier=0.07
Domestic Crisis	 AUC=0.97 Brier=0.03	 AUC=0.96 Brier=0.06
International Crisis	 AUC=0.99 Brier=0.03	 AUC=0.98 Brier=0.03

6.1. **Predicting all Crisis Events - Cumulative Probabilities.** In addition to predicting specific crisis events, we can use the predictive probabilities derived by each of the models to create a cumulative predictive probability of any crisis event occurring. Thus, the goal is to provide a prediction of an onset of any of the five crises predicted by the individual models above. Using simple probability theory the predicted probabilities derived for the individual EOIs can be combined to a cumulative probability.

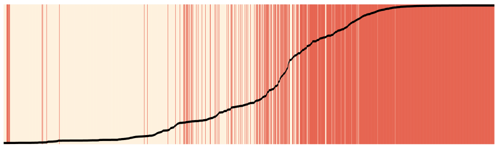
To do so we first create the new crisis variable that indicated the occurrence of any of the five crisis events above. To predict the experience of any crisis event we combine the individual predicted probabilities and calculate a cumulative probability of any crisis occurring. By simple probability theory, the cumulative probability is:

$$P(Y_{i,t}^C = 1) = [1 - P(Y_{i,t}^C = 0)] = 1 - \prod_{k=1}^{5} (1 - P(Y_{i,t}^k = 1))$$

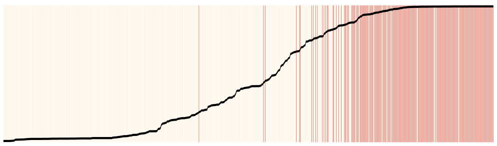
where k ranges from one through five, enumerating the predicted probabilities derived for country i in time t for each of the individual EOIs discussed above.

In words, the probability of a country to experience any crisis event in a given month is one minus the probability of no event occurring. This can be calculated as the product of the individual probability of non-occurrence for each of the individual events. Y stands for the occurrence of a crisis event in country i at time t, where the superscript indicates the type of crisis. We then use the cumulative probability to predict if a country experiences any of the above specified crises in a given month.

Figure 3 shows the separation plots of predicting any crisis using the cumulative predictive probabilities. There are very few actual crises that are missed, and relatively few false positives. Especially on the out-of-sample data, the cumulative prediction missed very few actual crises. A curious case is displayed on the in-sample separation plot, which shows a number of actual events on the left side with very low predicted probabilities of crisis occurrence. This is the



(a) In-sample



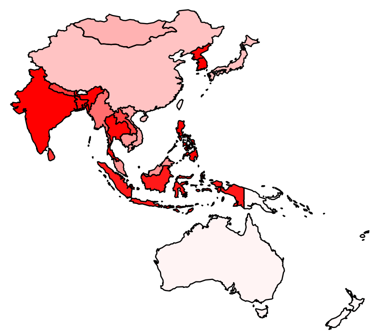
(b) Out-of-sample

FIGURE 3. Separation plot for any crisis, indicating fit of the model, in and out-of-sample.

case of the Solomon Islands in 2003, which experienced ethnic violence in five months during that year and was missed by the individual prediction models. Since the cumulative prediction is dependent on the individual models, the same case is missed in the in-sample prediction of the ethnic violence model. For the in-sample observations, the Brier score for predicting the occurrence of any crisis is 0.07, while the Brier score for the out-of-sample observations is 0.1. Thus, the model predicts quite accurate using the individual predictions to calculate a cumulative probability of any crisis occurring.

While the cumulative probability can predict any crisis is very well, it generally over-predicts the occurrence of crises. This is visible in the separation plots and can easily be explained using an example. If the predicted probability for each event is 0.25, a relatively low probability of an individual crisis event to take place, one can calculate that the cumulative probability of any of these events occurring is actually $1-(1-0.25)^5 = 1-0.75^5 = 0.76$. Thus while the predicted probability for each individual event is relatively low, the overall probability of one or more of these events to happen is quite high. Thus in general the cumulative probability is more likely to over-predict compared to the individual models, than to under-predict. However, combining predictions makes it unlikely that a crisis occurs without being forecasted. Figure 4 maps the average monthly event occurrence in 2010 and 2011 against the average monthly predicted risk for any event of interest. Again, this shows a slight over-prediction of the cumulative probability model but also highlights the otherwise preciseness of our predictions.

(a) Average monthly event occurrence



(b) Average monthly predicted risk

FIGURE 4. Average monthly event occurrence in 2009 and 2010 against the average monthly predicted risk for any event of interest

6.2. **Predicting beyond the training and test sample.** In Phase III of the ICEWS project, we were able to incorporate data until the end of 2010. This implies that even though we separated the data into a training set for insample predictions and a test set for out of sample predictions, we optimized our models in order to making good in and out of sample predictions. Even though this procedure is aimed to minimize over-fitting and maximize predictive power, there is a certain risk of what one might want to call "second-order over-fitting". From our experience this kind of over-fitting is especially likely in models with a small number of cases, a large number of variables, and short time series. Hence, we believe model evaluation by out-of sample prediction is especially powerful if the out-of sample data was not available to the researcher during the modeling phase. To attain this gold-standard, we incorporate new data on insurgencies in 2011 that was made available to us after the modeling phase. This new data is part of the Worldwide-ICEWS (W-ICEWS) project, which extends the ICEWS modeling approach to all countries in the world. This project is funded by the Office of Naval Research and coordinated by Lockheed Martin Advanced Technology Labs. One advancement in this project phase is that we not only have access to whether a project defined event of interest (insurgency, rebellion, ethnic violence, domestic crisis, and international crisis) takes place, but also the underlying count of conflictual interactions that informs the project's main events of interest. For example, in regard to insurgencies, we know whether a country experiences an insurgency but also have the count of insurgent conflictual events. This implies that the W-ICEWS project will be able to generate more fine grained models that will predict onset, occurrence, and the intensity of events. However, in the context of this article, we use the information on intensity only to evaluate the predictive power of our models.

In this subsection, we focus on our model of insurgencies as this is the first event of interest that we model in the the first phase of the W-ICEWS project. We define the existing data 2001-2010 as the training set and make predictions for the first three months of 2011 with data that has not been used in any way to inform the insurgency model. In Table 7, we present the average predicted probability for January, February, and March 2011 and sort the countries from high to low predicted risk. In addition, we provide the average number of insurgency coded months and the underlying average number of conflictual insurgency events. On average, we predict the highest probability of conflict for the Philippines and India, which both are identified as experiencing insurgencies in the W-ICEWS project data. When investigating the events during this time period, we find that in India, a Maoist insurgency actively continued with Naxalite guerillas grounding their bases in the country's southern mountainous regions. While the Naxalites have been known to have bases in the Himalayas, their operations are widespread: they reportedly maintain operations in 60 districts throughout India. The Philippines are also faced with a Maoist insurgency, led primarily by the New People's Army. Unlike the Naxalites, the New People's Army lacks sufficient force to control substantial territory and seize power. However, they were still responsible for 11 conflictual insurgent events in our prediction period. In addition, to being well predicted, these cases highlight central components of our model that focus on ethnic diversity and power-sharing dynamics of the country.

Bangladesh, Laos, and Indonesia are also identified as having a high risk of an insurgency. Even though the W-ICEWS project does not classify these countries as having an ongoing insurgency they nonetheless experience a number of insurgency related events. In some sense Nepal is our only real false-positive case as we predict a relatively high probability of insurgency, but in our prediction period neither an insurgency is coded nor insurgency events recorded. Finally, Cambodia is our only false-negative case with one recorded insurgency related event in February 2011. We believe that this event might be miscoded and relate to the interstate conflict between Cambodia and Thailand which escalated on 4 February when Thai and Cambodian troops exchanged fire in the vicinity of the Preah Vihear temple. The Uppsala Conflict Encyclopedia reports that it was unclear who provoked the clash which ended with a truce later the same day. However, the truce proved to be fragile as the movement of Thai tanks was enough to provoke renewed clashes. For three successive days, Thai and Cambodian troops exchanged fire, despite new attempts to end the violence through a ceasefire agreement on 5 February. Both sides suffered fatalities and three civilians were reportedly killed as stray bullets and rockets hit adjacent settlements. In addition, the skirmishes left thousands of civilians displaced.

7. Conclusion

In this article we have demonstrated the utility of creating forecasting models for predicting political conflicts in a diverse range of country settings. We have shown that this series of geographically informed statistical models is highly accurate, generating few false negative and positive predictions. These models can serve the public policy community as well as shed light on an array of critically important components of the political science literature on conflict dynamics. Moving forward, this project is currently extending these models beyond their current geographical domains to the Worldwide-ICEWS phase.

One frequent, and quickly surfaced, criticism of predictions in the social sciences is that social phenomena such as international crises are simply too complicated to predict by any means. Another, is

Table 7. Average predicted probabilities for insurgencies, January-March 2011

Country	Probability	Insurgency	Insurgency Events
Philippines	0.996	1	3.67
India	0.487	1	3.67
Bangladesh	0.481	0	1.33
Laos	0.473	0	1.00
Indonesia	0.347	0	2.67
Nepal	0.198	0	0.00
Thailand	0.043	0	0.00
Vietnam	0.002	0	0.00
China	0.000	0	0.00
Mongolia	0.000	0	0.00
Taiwan	0.000	0	0.00
North Korea	0.000	0	0.00
South Korea	0.000	0	0.00
Japan	0.000	0	0.00
Bhutan	0.000	0	0.00
Myanmar	0.000	0	0.00
Sri Lanka	0.000	0	0.00
Cambodia	0.000	0	0.33
Malaysia	0.000	0	0.00
Singapore	0.000	0	0.00
Australia	0.000	0	0.00
Papua New Guinea	0.000	0	0.00
New Zealand	0.000	0	0.00
Solomon Is.	0.000	0	0.00
Fiji	0.000	0	0.00

that experts are far better than any models at understanding and even predicting change in the realm of so-called "real world" politics.

Precisely because political conflicts are quite complicated presents a rather compelling reason to expand reason into mechanisms that can support these complications. Indeed, complex systems involve a wide variety of mechanisms and phenomena that are not easily described, let alone understood in isolation. A good example is meteorology, wherein we each receive a variety of forecasts every day. These forecasts are typically generated by combining a large number of forecasts that are based on meteorological models of weather that are based on the physics and chemistry of what is governing the various interacting systems. These systems each use a vast amount of measured data on the stocks and flows of various physical characteristics. These systems permit heterogeneity, so the predictions are not the same everywhere. And, they permit an increasingly accurate scale of prediction.

Indeed, the first attempt at weather prediction comes from Richardson (1922) over a century ago, when he used his mathematical approach to predict (retrospectively) the weather for 20 May 1910 by hand, using data to predict the weather six hours hence. When corrected by modern smoothing techniques, Richardson's predictions were quite accurate, although he did not perceive them to be adequate at the time (Lynch, 2006). To Richardson a global model of weather-forecasts would have taken thousands of human calculators, which from his perspective seemed impossible. However, today with the available computational power and exact global data, we are able implement weather models that are based on Richardson's ideas.

An interesting aspect of this is that Richardson turned away from weather predictions and wrote the probably first book on the statistical analysis of war: Statistics of Deadly Quarrels. After his experiences in the First World War, he thus focused on another complicated phenomenon to predict. Given advancements in theory, data collection, statistics, and computational power, we might be at an important point to push the boundaries of predicting political phenomenon beyond what we believed was possible only a few years ago. To preemptively declare defeat at the forecasting task seems foolish.

It may be that experts are better than expert systems. This could be true for every single task. Yet, we actually don't necessarily have experts for every problem we face. And the experts we do have, are not uniformly expert. Tetlock (2005) demonstrated this in his study of more than 250 experts in a variety of fields. Analyzing more than 28,000 specific predictions, he found that "experts" were slightly more accurate than flipping a coin, but not by much. Further, Tetlock (2005) showed that the most visible and prominent among the expert community, tended to be worse than chance. Even beyond his evidence that experts may have feet of clay, exhibit hindsight, outcome, and confirmation bias, as well as are fooled by sunk-cost, narrative, planning, and conjunction fallacies along with a variety of illusions and neglects. Further, we know that there is a long list of so-called intelligence failures in which big, momentous changes where not predicted by those experts. Indeed, even the death of Kim-Jong-il was initially missed by both South Korean and US intelligence services, which presumably devoted substantial resources to monitoring this particular Mr. Kim. There is reason to believe that a diversity of predictive activities will improve our knowledge about the world.

Finally, while prediction may be especially important in the policy realm, there is little risk that decision makers will be fooled by predictive models, or that they will be persuaded to ignore other sources of information. That said, the real benefit of using prediction may actually be as a heuristic allowing further probing of the empirical validity of specific models. Political science–especially where samples and experiments are not feasible–has an enormous vulnerability to over reliance on the available data. In a statistical sense this is often seen as overfitting. We use all the data to generate models that are dependent on all the data. That no longer seems like a very good research design. Being able to use our models to describe data we haven't seen before should be one gold standard criterion for model evaluation. The fact that in the face of a torrent of new data about the world we could now do this in almost real time, permits the possibility of generating predictions about the future that may be useful, not just toward validating our theories, but more generally.

References

Andriole, Stephen J. and Robert A. Young. 1977. "Toward the Development of an Integrated Crisis Warning System." *International Studies Quarterly* 21(1):107–150.

Bennett, D.S. and A.C. Stam. 2009. "Revisiting Predictions of War Duration." *Conflict Management and Peace Science* 26(3):256.

Brandt, P.T., M. Colaresi and J.R. Freeman. 2008. "The Dynamics of Reciprocity, Accountability, and Credibility." *Journal of Conflict Resolution* 52(3):343.

Brier, Glenn W. 1950. "Verification of Forecasts Expressed in Terms of Probability." *Monthly Weather Review* 78(1):1–3.

Bueno de Mesquita, Bruce, Alastair Smith, Randolph M. Siverson and James D. Morrow. 2003. *The Logic of Political Survival.* Cambridge, MA: M.I.T. Press.

Buhaug, Halvard and Kristian Skrede Gleditsch. 2008. "Contagion or confusion? Why conflicts cluster in space 1." *International Studies Quarterly* 52(2):215–233.

Cederman, Lars-Erik, Andreas Wimmer and Brian Min. 2010. "Why Do Ethnic Groups Rebel? New Data and Analysis." *World Politics* 62(1):87–119.

Cederman, Lars-Erik, Luc Girardin and Kristian Skrede Gleditsch. 2009. "Ethnonationalist Triads: Assessing the Influence of Kin Groups on Civil Wars." *World Politics* 61(3):403–437.

Cederman, Lars, Halvard Buhaug and Jan Ketil Rød. 2009. "Ethno-nationalist dyads and civil war: A GIS-based analysis." *Journal of Conflict Resolution* 53(4):496–525.

Chojnacki, Sven. 2006. "Anything New or More of the Same? Wars and Military Interventions in the International System, 1946-2003." *Global Society* 20(1):25–46.

Choucri, Nazli and Thomas W. Robinson, eds. 1978. *Forecasting in International Relations: Theory, methods, problems, prospects.* W.H. Freeman.

Davies, J.L. and T.R. Gurr. 1998. *Preventive measures: building risk assessment and crisis early warning systems.* Rowman & Littlefield Publishers.

de Marchi, Scott, Christopher Gelpi and Jeffrey D. Grynaviski. 2004. "Untangling Neural Nets." *American Political Science Review* 98(02):371–378.

de Mesquita, Bruce Bueno. 2002. *Predicting Politics.* Columbus, Ohio: Ohio State University Press.

Doran, Charles F. 1999. "Why Forecasts Fail: The Limits and Potential of Forecasting in International Relations and Economics." *International Studies Review* 1(2):11–41.

Dudley, R. and R.A. Miller. 1998. "Group Rebellion in the 1980s." *Journal of Conflict Resolution* 42(1):77–96.

Eck, Kristine, Joakim Kreutz and Ralph Sundberg. 2010. *Introducing the UCDP Non-State Conflict Dataset*. Uppsala University: Unpublished manuscript.

Enders, Walter and Todd Sandler. 2005. "After 9/11: Is it All Different Now?" *The Journal of Conflict Resolution* 49(2):259–277.

Fearon, James D. and David D. Laitin. 2003. "Ethnicity, Insurgency and Civil War." *American Political Science Review* 97(1):75–90.

Freeman, John R. and Brian L. Job. 1979. "Scientific Forecasts in International Relations: Problems of Definition and Epistemology." *International Studies Quarterly* 23(1):113–143.

Gleditsch, Kristian Skrede. 2007. "Transnational dimensions of civil war." *Journal of Peace Research* 44(3):293–309.

Gleditsch, Kristian Skrede and Michael D. Ward. 2010. "Contentious Issues and Forecasting Interstate Disputes." Presented to the 2010 Annual Meeting of the International Studies Association.

Gneiting, Tilmann and Adrian E. Raftery. 2007. "Strictly Proper Scoring Rules, Prediction, and Estimation." *Journal of the American Statistical Association* 102(477):359–378.

Greenhill, Brian D., Michael D. Ward and Audrey Sacks. 2009. "A New Visual Method for Evaluating the Predictive Power of Binary Models.".

Gurr, Ted Robert and Mark I. Lichbach. 1986. "Forecasting Internal Conflict: A Competitive Evaluation of Empirical Theories." *Comparative Political Studies* 19(1):3–38.

Gurr, T.R., B. Harff and B. Harff. 1996. "Early warning of communal conflict and humanitarian crisis." *Tokyo: United Nations University Press, Monograph Series on Governance and Conflict Resolution* .

Hegre, Håvard, Tanja Ellingsen, Scott Gates and Nils Petter Gleditsch. 2001. "Toward a Democratic Civil Peace? Democracy, Political Change, and Civil War, 1816-1992." *The American Political Science Review* 95(1):33–48.

Hildebrand, David K., James D. Laing and Howard Rosenthal. 1976. "Prediction Analysis in Political Research." *The American Political Science Review* 70(2):509–535.

Kalyvas, S.N. and M.A. Kocher. 2009. "The Dynamics of Violence in Vietnam: An Analysis of the Hamlet Evaluation System (HES)." *Journal of Peace Research* 46(3):335.

King, Gary and Langche Zeng. 2001. "Improving Forecasts of State Failure." *World Politics* 53(4):623–658.

King, Gary and Will Lowe. 2003. "An Automated Information Extraction Tool for International Conflict Data with Performance as Good as Human Coders: A Rare Events Evaluation Design." *International Organization* 57:617–642.

Krause, George A. 1997. "Voters, Information Heterogeneity, and the Dynamics of Aggregate Economic Expectations." *American Journal of Political Science* 41(4):1170–1200.

Leblang, David and Shanker Satyanath. 2006. "Institutions, Expectations, and Currency Crises." *International Organization* 60(01):245–262.

Lynch, Peter. 2006. *The Emergence of Numerical Weather Prediction: Richardson's Dream*. Cambridge: Cambridge University Press.

O'Brien, Sean P. 2002. "Anticipating the Good, the Bad, and the Ugly: An Early Warning Approach to Conflict and Instability Analysis." *The Journal of Conflict Resolution* 46(6):791–811.

O'Brien, Sean P. 2010. "Crisis Early Warning and Decision Support: Contemporary Approaches and Thoughts on Future Research." *International Studies Review* 12(1):87–104.

Pevehouse, J.C. and J.S. Goldstein. 1999. "Serbian compliance or defiance in Kosovo? Statistical analysis and real-time predictions." *The Journal of Conflict Resolution* 43(4):538–546.

Richardson, Lewis F. 1960. *Arms and Insecurity*. Chicago and Pittsburgh, PA: Quadrangle/Boxwood.

Richardson, Lewis Frye. 1922. *Weather Prediction by Numerical Process*. Cambridge: Cambridge University Press.

Salehyan, Idean. 2007. "Transnational Rebels: Neighboring States as Sanctuary for Rebel Groups." *World Politics* 59(2):217–242.

Salehyan, Idean. 2009. *Rebels without borders: transnational insurgencies in world politics*. Ithaca: Cornell University Press.

Sarkees, Meredith Reid and Frank Wayman. 2010. *Resort to War: 1816 - 2007*. Washington, DC: CQ Press.

Schrodt, Philip A. and Deborah J. Gerner. 2000. "Using Cluster Analysis to Derive Early Warning Indicators for Political Change in the Middle East, 1979-1996." *American Political Science Review* 94(4):803–818.

Schultz, Kenneth A. 2001. *Democracy and Coercive Diplomacy*. New York: Cambridge University Press.

Singer, J.D. and M.D. Wallace. 1979. *To augur well: early warning indicators in world politics*. Sage Pubns.

Snyder, Jack. 2000. *From Voting to Violence: Democratization and Nationalist Conflict*. New York: Norton.

Tetlock, Philip. 2005. *Expert Political Judgment: How Good is it? How Can we Know?* Princeton, NJ: Princeton University Press.

Vincent, Jack E. 1980. "Scientific Prediction versus Crystal Ball Gazing: Can the Unknown be Known?" *International Studies Quarterly* 24(3):450–454.

Ward, Michael D., Randolph M. Siverson and Xun Cao. 2007. "Disputes, Democracies, and Dependencies: A Reexamination of the Kantian Peace." *American Journal of Political Science* 51(3):583–601.

Michael D. Ward: Department of Political Science
Current address: Duke University, Durham, NC, 27708, USA
E-mail address: michael.d.ward@duke.edu

Nils W. Metternich: Department of Political Science
Current address: Duke University, Durham, NC, 27708, USA
E-mail address: nm86@duke.edu

Chris Carrington: Department of Political Science
Current address: Duke University, Durham, NC, 27708, USA
E-mail address: chris.carrington@duke.edu

Cassy Dorff: Department of Political Science
Current address: Duke University, Durham, NC, 27708, USA
E-mail address: cassy.dorff@duke.edu

Max Gallop: Department of Political Science
Current address: Duke University, Durham, NC, 27708, USA
E-mail address: max.gallop@duke.edu

Florian M. Hollenbach: Department of Political Science
Current address: Duke University, Durham, NC, 27708, USA
E-mail address: florian.hollenbach@duke.edu

Anna Schultz Department of Political Science
Current address: Duke University, Durham, NC, 27708, USA
E-mail address: anna.schultz@duke.edu

Simon Weschle: Department of Political Science
Current address: Duke University, Durham, NC, 27708, USA
E-mail address: simon.weschle@duke.edu