PREDICTING FUTURE LEVELS OF VIOLENCE IN AFGHANISTAN DISTRICTS USING GDELT

JAMES E. YONAMINE

1. Introduction

For centuries, key pillars of the philosophy of science like Francis Bacon and David Hume, have stressed that scientific progress occurs through the development of consistently accurate, replicable, and falsifiable predictive models. Building on these argument, numerous scholars of political conflict, including Choucri [1974], Singer and Wallace [1979], Beck et al. [2000], Bueno de Mesquita [2002], and Ward et al. [2010], have similarly stressed the importance of predictive models for two main reasons. First, as Beck et al. [2000], Weidmann and Ward [2010], and others convincingly argue, predictions are vital for the development of theories about the causes of violence, since the most rigorous way to test whether an empirical model is actually reflecting a real-world data generating process, or simply fitting "noise", is to measure its forecast accuracy.[1] Second, accurate conflict forecasts can be tremendously useful in the real world – they can help peacekeepers allocate scarce resources, inform Non-governmental Organizations (NGOs) on potential hot-spots to avoid, and even provide speculative investment opportunities. Although the majority of empirical studies of conflict continue to focus on "explanation" – primarily in the form of interpreting coefficients and standard errors established through in-sample testing – a smaller though considerable number papers and projects exist with the explicit goal of building dynamic forecasts of future levels of violence. Likewise, the goal of this article is to build a forecasting model, though not for theory-building or hypothesis-testing, but rather to create a proof of concept tool for real-time, policy relevant decision making.

Extant empirical forecasting studies focusing on domestic conflict range tremendously in terms of data, methods, and scope. The most coarse studies build forecasts at that state-year level using primarily structural variables like GDP per capita, ethnic diversity, and infant mortality (see Gurr and Harff [1996], King and Zeng [2001], Fearon and Laitin [2003], and Goldstone et al. [2010]), which

Date: Version 0.8 : April 12, 2013.

[1] I use "prediction" and "forecast" interchangeably throughout this article.

are useful in some contexts but unable to build predictions beyond the state-year unit of analysis. The majority of studies attempting to build empirical forecasts of violence use more fine grained, event data coded at the daily and sometimes local level, as these data allow scholars to capture more dynamic patterns of violence and ultimately build more detailed forecasts than those using state-year, structural data. Historically, scholars building empirical forecasting models of violence have used either machine-coded (like the Kansas Event Dataset (KEDS) (see Schrodt [1990], Integrated Conflict Early Warning System (ICEWS) (see O'Brien [2010], 10 Million International Dyadic Events Dataset (see King and Lowe [2004])) or human-coded event data datasets (like ACLED (see Raleigh et al. [2010]) built form open source text, with the majority of scholars utilizing the machine-coded option. Recently, however, WikiLeaks has provided an alternative data set of conflict events that previously required security clearance from the United States Government to access, but have subsequently been illegally obtained and distributed to the public. The logical question, then, is which of these sources of data is more appropriate for this study? Given the goal of this article, an ideal dataset would contain the following five key attributes:

(1) Broad spatial coverage: Global coverage is preferable to one with country or region specific coverage as it would enable a forecasting model to be built for any global location.
(2) Density: Predictive algorithms tend to perform better with more data, meaning that many fine-grained events is preferable to fewer larger scale events.
(3) Geo-coding: Sub-state, geo-spatial predictions require sub-state, geo-coded events.
(4) Accuracy: The data should accurately reflect the events as they occur in reality in order to build relevant predictions.
(5) Future availability in real-time: If the data are not accessible in the future in real or near real-time, then it becomes highly difficult to build actionable predictions.

Each of the five attributes above is a necessary condition to building nuanced predictions on a global scale, and none of the major existing datasets listed above meet all the conditions. For example, WikiLeaks is highly accurate since it is based on first hand accounts, but limited in spatial coverage and not likely readily available in the future. ICEWS, KEDS, and the 10 Million International Dyadic Events datasets are capable of real-time updates, but lack the ability to geo-code, and relying on human-coded datasets makes it difficult, if not impossible to update in (near) real time or maintain on a global scale because human coding is slow.

Recently, however, a breakthrough dataset called the Global Dataset of Events, Location, and Tone (GDELT) was released, contained over 200 million that are machine-coded in near real-time (e.g. daily) based on open source data. Most importantly to this study, GDELT is the first machine-coded dataset capable of performing sub-state geo-coding, providing specific latitude and longitude coordinates for each observation. Thus, GDELT is the first dataset to meet all five criteria above, and is currently the only suitable source of data for building temporally and geo-spatially nuanced forecasts of violence on a global scale in real time.

This is the first study to ever use open-source, machine-coded event data to build forecasts of political violence at a sub-state level of geospatial aggregation. Since the process of aggregating conflict events into sub-state units based on latitude and longitude is currently time and computationally intensive, doing so on a global scale exceeds the scope of a article. Thus, I focus on forecasting conflict in sub-state geospatial units in a single country: Afghanistan. I choose Afghanistan for two reasons. First, there is dense political violence across a long time-frame (2001-2012) with considerable variation at local levels. Second, Mangion-Zammit et al. [2012] have demonstrated the ability to build forecasts with the WikiLeaks data, meaning that to the extent it is possible at all to build temporally and geo-spatially nuanced forecasts of political violence using open source, machine-coded event data, it should be feasible in Afghanistan.

Although I focus primarily on building predictions one-month in advance at the district-month unit of analysis (Afghanistan's smallest administrative unit, N=317), I also build forecasts at the province-month (N=32) and the country-month (N=1) level, which provides a rudimentary test of the effects of geo-spatial aggregation on forecast accuracy. Empirically, I use an autoregressive fractionally integrated moving average (ARFIMA) model, which builds forecasts of levels of material conflict one-month-in-advance that consistently outperforms a naive model assuming that the level of violent in a location during a month will be the same as it was in the same location in the previous month. The ARFIMA model performance decrease relative to the naive model at each additional level of geo-spatial aggregation, suggesting further justification for the use of fine-grained geo-spatial analyses. Additionally, I implement two logical extensions to the univariate ARFIMA model, first by building and modeling additional features, and second by incorporating exogenous drug price data to ARFIMA model, though neither enhance predictive accuracy. The remainder of this article provides a review of relevant literature, details my research design and ARFIMA forecasting model, discusses two logical extensions, and lastly concludes.

2. Literature Review

To facilitate this review of relevant literature, I organize studies that forecast domestic political violence into the three general types of data that they use: machine-coded, human-coded, and WikiLeaks.

2.1. Machine-coded data.

Although a large number of studies utilize machine-coded event data (see Appendix A), a much smaller subset of these studies build forecasts: Schrodt and Gerner [1997] use discriminant analysis to predict conflict phases in the Levant, Schrodt [1999] uses HMMs to forecast conflict in southern Lebanon, Pevehouse and Goldstein [1999] use time-series to predict events in the Serbia-Kosovo conflict, Schrodt and Gerner [2000] forecast unique clusters of conflict in the Levant from 1979 to 1997, Schrodt [2000] uses HMM's to forecast conflict dynamics in the Levant form 1979 to 1997, Bond et al. [2004] forecast conflict in Indonesia, Shellman [2004b] forecasts conflicts between government and dissident actors in Chile and Venezuela, Brandt and Freeman [2005] use Bayesian time-series to forecast dynamics between the United States, Israel, and Palestine, Schrodt [2006] forecasts conflict in the Balkans using HMMs, Shearer [2006] uses HMMs to forecast conflict between Israel and Palestine, Bagozzi [2011] uses zero-inflated count models and D'Orazio et al. [2011] use sequence analysis to forecast domestic conflict in 29 Asian countries, and Brandt et al. [2011] employ Markov Switching Bayesian Vector Autoregression (MS-BVAR) for forecast domestic and inter-state conflict in the Levant in 2010. Although these and other scholars demonstrate the ability to generate accurate forecasts of when and between whom conflict will occur in the future using open-source, machine-coded event data, they have been unable to predict *where* this conflict will occur at a sub-state level since none of the relevant machine-coded event data datasets provided geo-location information prior to GDELT.

2.2. Human-coded data.

A number of geo-located, human-coded event data datasets exist that could allow researchers to build forecasts of violence at specific sub-state geographic units. For example, the Armed Conflict Location and Event Dataset (ACLED), which provides over 75,000 geo-coded violent events with (both atomic and composite) for approximately 60 countries, including all of Africa, and other, conflict-prone countries throughout the world (see Raleigh et al. [2010]), Daly [2012] provides a dataset with 7,729 geo-coded acts of violence in Colombia from 1964-1984, Schneider et al. [2012] presents the Konstanz One-Sided Violence Event Dataset (KOSVED) with 21,458 attacks against civilians in Bosnia, Urdal and Hoelscher [2012] introduces a dataset of 4,003 events occurring in 55 major cities in Asia and sub-Saharan Africa from 1960 to 2008, and Salehyan et al. [2012] introduce the Social Conflict in Africa Database (SCAD), which contains 7,200 events of political unrest occurring in 47 African countries from 1990-2010.

Despite the geospatial nuance of these datasets, it is somewhat surprising that only Weidmann and Ward [2010] uses one of the aforementioned datasets (ACLED) in order to build predictions, whereas dozens of other articles dimly focus on explanation. Weidmann and Ward [2010] use ACLED's Bosnia dataset in order to build a model that predicts a binary measure of whether a given municipality-month in Bosnia. In total, 4,796 municipality months exists (109 municipalities form March 1992 to October 1995), of which 301 experienced an ACLED conflict event and are treated as a "1". They build a model based on exogenous variables (population, ethnic diversity, borders, and mountains) as well as various endogenous lags of the dependent variable, and utilize a Markov Chain Mote Carlo (MCMC) technique to estimate a logistic regression which is then used to calculate predictions in a rigorous out-of-sample framework, which I discuss in greater detail in Section 4.2.

Despite making major theoretical and empirical contributions to the study of political violence, the fact that the only study to build out-of-sample forecasts using human-coded event data (e.g. Weidmann and Ward [2010]) did so for a conflict that ended five years prior to the release of the study underscores the slow, tedious nature of building human-coded datasets that makes them extremely difficult to update sufficiently close to real time as to build policy-relevant forecast actually for the future.

2.3. **WikiLeaks data.** On July 25, 2010, WikiLeaks publicly released the majority of classified documents comprising both the Afghan War Diary (containing 91,731 documents) and the Iraq War Log (containing 391,832), which contain classified documents that provide a highly detailed account of events occurring in Afghanistan and Iraq from January 2004 through December 2009. Additionally, in 2010, the United States government declassified subsections of the Afghan War Diary and the Iraq war log, called Significant Acts (SIGACT). Although both the WikiLeaks and SIGACT datasets have become difficult to obtain, a number of academic studies have been published that empirically model these data for both Iraq and Afghanistan. Like studies discussed in Section 2.2, the majority of studies using the WikiLeaks and SIGACT data focus on explanation, rather than prediction.

For example, Berman et al. [2011] analyze the effects of sub-state level unemployment data for 297 district-quarters (3 quarters for 99 districts) for Iraq and 2,160 district-months (6 months for between 363 and 365 districts) for Afghanistan on levels of violence using the SIGACT data; Weidmann and Salehyan [2011] use the SIGACT data to analyze the effects of the U.S. surge in Iraq on levels of violence in 85 neighborhoods in Baghdad; O'Loughlin et al. [2010] use hotspot and cluster analysis to compare the Afghan War Diaries data to ACLED's Afghanistan data; Linke et al. [2012] model violence dynamics between the U.S-led coalition forces and insurgent by analyzing 301,374 violent events aggregated at the three-day, 30-by-30 second grid-cell level, and although the authors do assess their model's predictive accuracy, this is done only using in-sample findings as opposed to a proper in-sample/out-of-sample break, meaning that the model is not *actually* building predictions. Among studies drawing on the WikiLeaks or SIGACT datasets, Mangion-Zammit et al. [2012] is the only to actually build out-of-sample forecasts. To do so, Mangion-Zammit et al. [2012] first use the WikiLeaks data to calculate the number of violent events at the province-month level in Afghanistan from 2004 to 2009, which serves as the in-sample training set. Second, they construct and train a point-process model on the 2004-2009 training data. Third, they build future predictions at the province-year level for 2010, based purely on information from 2004-2009. Since WikiLeaks only provides data through 2009, Mangion-Zammit et al. [2012] evaluate their model's predictive accuracy based on data provided by the Afghan NGO Safety Office (ANSO), and find that 62.5% of actual levels of violence fall within 95% confidence intervals of predicted levels.

Although these studies apply innovative methods to address interesting questions, they highlight two major shortcomings to working with WikiLeaks-style of data. First, even when it can be acquired, it does not provide real or near-real time updates. As a result, Mangion-Zammit et al. [2012] needed to use a different data source to obtain data from 2010 since WikiLeaks only covered 2004-2009. Second, all of the studies discussed in Section 2.3 focus on either Iraq or Afghanistan since WikiLeaks only provided dense data for those countries, which clearly means that WikiLeaks data is unsuitable to build predictions for any other states in the world.

The research design I outline in the following sections using the GDELT dataset not only overcome the shortcomings WikiLeaks-style data, but also those of the extant literature relying on human-coded and pre-GDELT machine-coded datasets. In the following section, I outline how I use GDELT to build state- and sub-state levels of political conflict in Afghanistan and discuss my forecasting approach.

3. Research Design

3.1. **Constructing material conflict counts.** As previously mentioned, Afghanistan is spatially divided into 32 provinces and 317 sub-provincial-level districts. Using the GDELT data in conjunction with GIS software, I calculate the number of material conflict events that occur from February 1, 2001 through April 30, 2012 between all actors in each month at three (country, province, and district) geo-spatial levels of analysis. To accomplish this, I first select all material conflict events for which either the source or target actor's primary affiliation (i.e. the first three characters of their actor identification) was with Afghanistan. I use a version of the GDELT data that has duplicate entries eliminated, as my goal in this article is to forecast *actual* the occurrence of events, rather than the perception or intensity of events. This step generates 139,915 material conflict events, each of which contains a specific latitude and longitude coordinate reflecting where the event occured. Next, using shape files and GIS software, I calculate the the number of events that occur within each district and province in each month. I choose to use the month as my level of temporal aggregation because this provides sufficient variation throughout the time-series while reducing the level of noise that is present at daily or weekly levels. Largely for those reasons, the monthly level aggregation is the most commonly used in the relevant literature, employed by Goldstein [1991], Schrodt [1997], Schrodt and Gerner [1997], Schrodt and Gerner [2000], Schrodt and Gerner [2001], Shellman [2004a], Shellman [2004b], Gleditsch and Beardsley [2004], Schrodt [2007], Brandt et al. [2008], Weidmann and Ward [2010], Ward et al. [2010], Shellman et al. [2010], Brandt et al. [2011], D'Orazio et al. [2011],and Mangion-Zammit et al. [2012]. District- and province-months with no material conflict events are assigned a "0". This results in 43,746 district months, 4,352 province months, and 136 country months.[2]

[INSERT FIGURE 1 HERE]

Figure 1 provides a visual overview of the data, illustrating changes in the number of material conflict events from 2001 to 2012 that occur in each district-year.

4. Forecasting Approach

In this section, I outline my forecasting approaches using the univariate data comprised solely of the counts of material conflict events. To facilitate discussion, I detail my forecasting approach

[2]This was done with substantial assistant form John Bieler as well as Josh Steven, who completed all geo-spatial aggregation using GIS.

as applied to the district-month level-of-analysis, though the approach is identical at the province-month and country-month levels-of-analysis as well. Since the structure of the data is time-series cross sectional at highly nuanced unit of analysis – i.e Afghani districts – I am unable to find appropriate exogenous variables to help predict future levels of material conflict.[3] As such, the district-month dataset contains 317 univariate time-series of the count of material conflict events at the district-month level, and I reflect the number of material conflict events occurring in a dingle district month with the notation $District_{it}$.

Since accurate forecasts are so useful across academia, government, and private sectors, there are many different empirical approaches to building forecasts. No one-size-fits all model exists, and it is impossible to know ahead of time which algorithm will generate the greatest degree of predictive accuracy. Due primarily to the large number of observations and amount of information (i.e. location, actors, date, etc.) contained in most event data datasets, including machine-coded, human-coded, and WikiLeaks data, researchers have applied a large number of different forecasting models.

D'Orazio et al. [2011] report that models forecasting domestic conflict largely fall into three general categories: time series (Shellman [2004a], Shellman [2007], Harff and Gurr [2001]), vector auto regression (VAR) (Pevehouse and Goldstein [1999], Goldstein [1992], Freeman [1989], Brandt et al. [2011]), and HMMs (Schrodt [1999], Bond et al. [2004], Shearer [2006], Schrodt [2000], and Schrodt [2006], Petroff et al. [2012]). Additionally, other studies using event data have employed additional methods, such as linear models (Weidmann and Ward [2010], Fearon and Laitin [2003], Gurr and Harff [1996]), clustering algorithms (Schrodt and Gerner [2000] , and point-process modeling (Mangion-Zammit et al. [2012]). Even after choosing a base algorithm, a number of choices must still be made regarding tuning parameters. For example. In addition, a number of techniques, like bagging and boosting can be applied to most of these algorithms (see Schrodt et al. [2012] for a discussion of these techniques in the context of political violence forecasting). As if that did not provide enough choices, a number of approaches combine multiple algorithms into ensemble methods, such as bayesian model averaging (BMA) (Montgomery et al. [2012]).

Despite the nearly infinite number of plausible forecasting approaches, the structure of my data is highly constraining for two main reasons. First, it is a univariate time series, meaning that it does

[3]Exogenous variables on employment and drug prices exist for select districts for select months, but neither variable is available with sufficient coverage to include in an empirical forecasting model at the district-month level. I discuss this further in Section 6.2

not contain exogenous covariates. Most of the methods above specifically designed for datasets with many covariates and are less relevant for my data. Second, my data is temporal. This restricts how I am able to divide my training and test set, since that training set must exclusively contain observations that preceded the test set. This greatly inhibits re-sampling techniques like boosting as a way of enhancing predictive accuracy. In the following section, I outline a forecasting model that achieves highly accurate predictions using a univariate time-series, discuss my out-of-sample forecasting framework, and detail how I build a benchmark to assist with evaluating forecast accuracy.

4.1. **The ARFIMA model.** To build forecasts with the univariate time-series, I implement an Autoregressive Fractionally Integrated Moving Average (ARFIMA) model, which models all univariate time-series (317 at the district-level, 32 at the province level, and 1 at the country-level) independently of each other. Though this is the first time an ARFIMA model has been used to forecast political conflict, a number of studies have demonstrated its ability to generate more accurate and consistent forecasts than other time-series models across various substantive fields. For example, Siew et al. [2008] demonstrate that an ARFIMA model consistently outperforms a traditional ARIMA model in forecasting air pollution rates, Chu [2009] generates more accurate forecasts of tourism levels in Asia with an ARFIMA model than with seasonal ARIMA (SARIMA) models, Barkoulas and Baum [2006] illustrates how ARFIMA models outperform other autoregressive models in forecasting U.S. monetary indices, and Bhardwaj and Swanson [2006] show that the ARFIMA model outperforms both ARIMA models and GARCH models in forecasting returns in the S&P500.

To introduce the ARFIMA model, first consider an ARIMA (p,d,q) model for a univariate time series X$(x_t, x_{t-2}, x_{t-3}, ..., x_{t-n})$ with d=0, which we can write as:

$$x_t = \omega + \epsilon + \sum_{i=1}^{p} \beta_i x_{t-i} + \sum_{i=1}^{q} \alpha_i \epsilon_{t-1} \tag{1}$$

where ω is a constant, x_{t-i} is the lagged dependent variable, ϵ_{t-i} is the lagged error, ϵ_t is the current error, and β_i and α_i are estimated parameters. When a time-series is non-stationary, first-differencing or "integrating" the series can help achieve stationarity. This generates a new time series, Δx$_t$, calculated by the following formula:

$$\Delta x_t = x_t - x_{t-1} \tag{2}$$

Thus, we can convert the ARIMA(p,d,q) model with d=0 to an ARIMA(p,d,q) model with d=1 by replacing the x characters with Δx, as done in the following formula:

$$\Delta x_t = \omega + \epsilon + \sum_{i=1}^{p} \beta_i \Delta x_{t-i} + \sum_{i=1}^{q} \alpha_i \epsilon_t + \epsilon_{t-i} \tag{3}$$

Although the ARIMA(p,d,q) model is among the most commonly used time-series models and has been used successfully to forecast with event data (see Shellman [2007]), it is rigid in that d must be an integer. The key innovation of the ARFIMA model is that it allows for d to take on any real number, which need not be an integer (hence the name "fractionally integrated"). Mathematically, Granger and Joyeux [1980] demonstrates that by allowing d<1, the ARFIMA model is able to efficiently account for a long memory process, which occurs when the time-series tends to revert to a historical mean. Importantly, the ARFIMA model is capable of accounting for the long memory process even without increasing the number of p and q lags.

To implement a flexible ARFIMA(p,d,q) model, I utilize the 'arfima' package in r, which automatically establishes values for the p, d, and q parameters of a univariate time series by determining the estimates for these parameters that maximize the likelihood function. This means that the researcher does not need to pre-specify the number of autoregressive components, moving average components, or degree of fractional integration. I treat each cross-section as a unique time-series, meaning that I train and build forecasts with the ARFIMA model one district and one province at a time through a looping function.[4] The 'forecast' function in the 'arfima' package allows the user to build a prediction N units into the future and provides a mean prediction along with 95% confidence intervals. To establish predictions, I use the mean of the one-month-ahead prediction rounded to the nearest integer. Figure 2 demonstrates the use of the 'arfima' package to build a prediction of the number of material conflict events in Bughran province in April, 2009 using data from February 2001 through March 2009. The prediction in Figure 2 provide the mean (the circle) as well as 90 and 95% confidence intervals, indicated by the light and darker vertical shading.

[4]Many districts have long periods of consecutive months with "0" material conflict events, which causes the 'arfima' package to crash. To allow the 'arfima' package to properly converge, I generate a random number from a uniform distribution from 0 to .1 for each district-month, and add that value to the count of material conflict events.

[INSERT FIGURE 2 HERE]

4.2. **Out-of-sample framework.** In order to calculate out-of-sample performance accuracy of the ARFIMA model, I utilize the same approach implemented by Weidmann and Ward [2010], which I implement on my data according to the steps outlined below, using the district-level model as an example:

- Train the model on an initial in-sample set containing all data from February 2001 until April 2008.
- Predict (and store) the number of material conflict events for May 2008 (i.e. a one-month-ahead out-of-sample forecast.
- Incorporate May 2008 into the in-sample set.
- Retrain the model on this new in-sample set, which now includes all data from February 2001 to May 2008.
- Predict (and store) the number of material conflict events for June 2008.
- Repeat until a final prediction is made for April 2012 (i.e. the last month in the data set), using a model trained on February 2001 through March 2012.

This results in 48 out-of-sample, one-month-ahead forecasts for each of the 317 municipalities. At the province-month level, this approach yields 48 out-of-sample, one-month-in-advance forecasts for each of the 32 provinces, and at the country-month level, this results in 48 one-month-in-advance forecasts for Afghanistan as a whole.

4.3. **Establishing a benchmark.** Since this is the first paper to build nuanced predictions of political conflict in Afghanistan at the monthly level, no existing appropriate benchmark of predictive accuracy exists. Without an appropriate benchmark, it is difficult to assert whether an alternative predictive model is performing well. The literature provides two plausible approaches to assessing how well a predictive model is performing in the absence of other models attempting to predict the same outcome. First, Gurr and Lichbach [1986] provides a strong theoretical argument called “the conflict persistence model”, which suggests that in the absence of an existing benchmark, it is logical to build a naive model that assumes conflict in the future will be the same in a given location as it is today. Second, Mangion-Zammit et al. [2012] reports the percentage of times that the true number of violent events fall within the 95% and 99% confidence intervals of predicted levels of violence. I choose to follow Gurr and Lichbach [1986]'s approach, and construct a naive model that predicts the number of material conflict events in $District_{it} = District_{it-1}$, for three reasons.

First, Mangion-Zammit et al. [2012]'s approach tells actually tells us little about a model's predictive accuracy because it does not penalize for large confidence intervals. Imagine that the true number of violence events occurring in $District_{it}$ is 75. Now, consider two models. Model 1 generates a prediction for the number of violent events in $District_{it}$ with 95% confidence intervals at 12 and 162, while Model 2's prediction for $District_{it}$has 95% confidence intervals at 68 and 74. Mangion-Zammit et al. [2012]'s approach would report that Model 1 is accurate and Model 2 is inaccurate, when in reality, it is difficult to imagine a scenario in which we would prefer Model 1's prediction to that of Model 2. Second, and directly related to the first point, is that Gurr and Lichbach [1986] approach generates a specific point prediction as a benchmark, which creates greater flexibility in assessing model performance. For example, Gurr and Lichbach [1986]'s approach allows me to calculate Mean Absolute Error (as detailed below), which is impossible using Mangion-Zammit et al. [2012]'s approach. Lastly, in many forecasting contexts (especially predicting civil conflict at the state-year level), the Gurr and Lichbach [1986] approach achieves almost perfect accuracy – countries at peace tend to stay at peace and countries at conflict tend to stay at conflict. This naive approach often works so well that it occasionally outperforms far more sophisticated forecasting models.

For example, Montgomery et al. [2012] introduce Bayesian Model Averaging (BMA) approach, and demonstrate how they are able to leverage the predictions of three separate models in order to build accurate forecasts that outperform all of the three component models. Montgomery et al. [2012] report that their BMA technique outperforms all of the three component models, accurately predicting 13 of 35 conflict onsets ("1's") and all 313 of the 313 non-onsets ("0's") in their dataset. While these may appear strong at first, Gurr and Lichbach [1986]'s naive benchmark approach accurately predicts 33 of the 35 conflict onsets and 310 of the 313 non-onsets, which is a dramatic improvement over the not only the BMA, but also the three component predictive models. Based on this, I assume that any model that consistently outperforms the naive t=t-1 assumption to be accurate.

4.4. **Calculating accuracy.** For each of the 48 months that iteratively serve as the out-of-sample test, I calculate the error rates for the naive model (naive_error) and the ARFIMA model (arfima_error rate), which reflect the MAE across the N cross-sections (N=317 for the district-month model, N=32 for the province-month model, and N=1 for the country-month model) according to the Formula (4) and Formula (5).

$$naive\_error_m = \frac{\sum_{i=1}^{N} \mid naive\_prediction_{i,m} - true\_count_{i,m} \mid}{N} \tag{4}$$

$$arfima\_error_m = \frac{\sum_{i=1}^{N} \mid naive\_prediction_{i,m} - true\_count_{i,m} \mid}{N} \tag{5}$$

These formulas result in a naive_error and arfima_error rate for the district-level, province-level, and country-level models for each of the 48 months that serve as the test-month allowing me to determine the extent to which the ARFIMA model outperforms the naive model across the three levels of geo-spatial aggregation (district, province, and country) in the following section.

5. Results

Table 1 provides the arfima_error rate, naive_error rate, and a TRUE/FALSE label indicating whether the ARFIMA forecasts are more accurate on average across all 317 districts for the given month.

[INSERT TABLE 1 HERE]

As Table 1 indicates, the ARFIMA model outperforms the naive model in 47 out of 48 of the out-of-sample months. Additionally, the ARFIMA model reduces the sum of the 48 monthly MAE's by over 16%. Taken together, these are highly impressive finding, especially when considering that naive models (that assume t=t-1) of conflict tend to perform well in forecasting.[5]

[INSERT TABLE 2 HERE]

Table 2 provides the arfima_error rate, naive_error rate, and a TRUE/FALSE label calculated from province-level geo-spatial aggregations, meaning that each of the 48 arfima_error and

[5] A potential critique of these results is that I do not perform any rigorous external validity check, meaning that I may simply be predicting the event-data generating process, rather than *actual* levels of violence. I believe that this is not overly problematic for two main reasons. First, many other forecasting studies likewise rely exclusively on event data and do not perform rigorous external validity checks, which has set a precedent that this is generally accepted practice. Second, the anecdotal story discussed in article 1 serves as an informal external validity check that suggests the GDELT data is accurate.

naive_error rates reflect their respective means across the 32 provinces. At the province-month level, the ARFIMA does not perform as well as at the district-month level, but it still outperforms the naive model in 40 of the 48, or approximately 83% months that serve as the test month. Furthermore, the ARFIMA model reduces the sum of the 48 month MAE by approximately 13%. Even though the ARFIMA performs slightly worse at the province-level than the district-level, it still achieves a respectable level of enhanced accuracy relative the the naive benchmark.

[INSERT TABLE 3 HERE]

Table 3 replicates Table 1 and Table 2, except it reflects the arfima_error rate, naive_error rate, and the TRUE/FALSE label based on a single country-level forecast per month. Table 3 illustrates that at the country-month level, the ARFIMA still outperforms the naive model, but does so at a lower margin than at the district-month or province-month level. Of the 48 months that test sample, the ARFIMA model outperforms the naive model 30 times, or 62.5%. Additionally, the ARFIMA model generates a lower sum of MAE's, but only by approximately 1%, which suggests that the increase in predictive accuracy of the ARFIMA model at the country-month level may be largely meaningless.

Across the district-, province-, and country-month forecasts, the key aspect of the ARFIMA model is that it tends to build forecasts that are between the *naive* model forecast and a longer term moving average. Exactly how much the ARFIMA model shifts forecasts away from the *naive* forecasts and towards the longer term moving average varies based from by month and by cross-section, but in effect, the ARFIMA acts like a smoothing function. Figure 2 visually demonstrates this. The last observed number of material conflict events is approximately 280 in month 99, meaning that the *naive* model would predict 280 events for the month 100. However, we can see that the average number of material conflict events in the previous months is less than 280, so the mean ARFIMA forecast (represented by the black dot) is less than 280. To the extent that the ARFIMA model outperforms the *naive* model, it suggests that levels of future violence tend to exhibit mean reverting characteristics.

6. Future directions

Although the ARFIMA model outlined above largely accomplishes the goal of this paper, in this section I provide preliminary analysis of two logical extensions for the finding in the previous section: first, building features from the univariate time-series to allow for other types of predictive algorithms; second, incorporating exogenous information, such as drug prices.

6.1. **Building features and implementing an ensemble method.** A common approach when building forecasting models is to manipulate existing data in order to build additional features, or covariates, which may uncover meaningful patterns in the data that are hidden in other variables. In many contexts across disciplines, building additional features leads to enhanced predictive accuracy. Note that building features can also decrease predictive accuracy because the additional dimensionality increases the likelihood of over fitting a model. To overcome this, I employ the same out-of-sample predictive framework as previously outline in Section 4.2.

Just like there there is no definitive way to pick the best forecasting algorithm, there are no rules for constructing features. As such, I build 11 new features below, all from the univariate time series, in an attempt to enhance predictive accurate beyond the univariate ARFIMA model outlined in the previous section.

- 2_month_MA = $(\mathrm{count}_t + count_{t-1})/2$
- 3_month_MA = $(\mathrm{count}_t + count_{t-1} + count_{t-2})/3$
- 4_month_MA = $(\mathrm{count}_t + count_{t-1} + count_{t-2} + count_{t-3})/4$
- 5_month_MA = $(\mathrm{count}_t + count_{t-1} + count_{t-2} + count_{t-3} + count_{t-4})/5$
- 6_month_MA = $(\mathrm{count}_t + count_{t-1} + count_{t-2} + count_{t-3} + count_{t-4} + count_{t-5})/6$
- $\Delta\_2\_month\_MA = count_t - 2\_month\_MA$
- $\Delta\_3\_month\_MA = count_t - 3\_month\_MA$
- $\Delta\_4\_month\_MA = count_t - 4\_month\_MA$
- $\Delta\_5\_month\_MA = count_t - 5\_month\_MA$
- $\Delta\_6\_month\_MA = count_t - 6\_month\_MA$
- monthly_sum = the sum of all material conflict events occurring across all spatial units each month

With these additional covariates, I build a number of additional predictive models following the general approach in Section 4.2. Using the 'glm' package in r, I build predictions using linear models comprised of various combinations of the 11 additional covariates above (all lagged one-unit) as well as a one-unit lag of the dependent variable, trying both "gaussian" and "poisson" distributions. I am unable to find a linear combinations of the covariates above (including the lagged dependent variable) capable of outperforming the naive benchmark at the district-month level in more than 35 out of the 48 district-months that serve as the out-of-sample set. Motivated by the enhanced predictive accuracy of the approach in Montgomery et al. [2012], I also implement an ensemble approach. To build an ensemble, I build use two component models, $Model\_1$ and $Model\_2$, which are specified below and estimated using the 'glm' package in r with a gaussian distribution.[6]

(6)

$$Model\_1$$

$$\begin{aligned} Dist\hat{r}ict_{it} = {} & \beta_0 + \beta_1 2\_month\_MA_{i(t-1)} + \beta_2 3\_month\_MA_{i(t-1)} + \beta_3 4\_month\_MA_{i(t-1)} \\ & \beta_4 5\_month\_MA2_{i(t-1)} + \beta_5 6\_month\_MA_{i(t-1)} + \beta_7 monthly\_sum_{i(t-1)} + \beta_8 District_{i(t-1)} \end{aligned}$$

(7)

$$Model\_2$$

$$\begin{aligned} Dist\hat{r}ict_{it} = {} & \beta_0 + \beta_1 \Delta\_2\_month\_MA_{i(t-1)} + \beta_2 \Delta\_3\_month\_MA_{i(t-1)} + \beta_3 \Delta\_4\_month\_MA_{i(t-1)} \\ & + \beta_4 \Delta\_5\_month\_MA_{i(t-1)} + \beta_5 \Delta\_6\_month\_MA_{i(t-1)} + \beta_7 District_{i(t-1)} \end{aligned}$$

Using these two models, I build an ensemble forecasting model according to the six steps below:

(1) Estimate two models on the same in-sample set as in Section 4.2, which contains all data from February 2001 until April 2008, and generate predictions for these in-sample months and store coefficient estimates
(2) Train the *Ensemble* model using the 'glm' function in R on the in-sample predictions from $Model\_1$ and $Model\_2$ according to the formula below, and store coefficient estimates:

$$Ensemble = Dist\hat{r}ict_{it} = \beta_0 + \beta_1 Model\_1_{it} + \beta_2 Model\_2_{it} \tag{8}$$

(3) Build predictions for May 2008 (i.e. one-month ahead out-of-sample forecast) for $Model\_1$ and $Model\_2$ by matrix multiplying the coefficient estimates from Step 1 and the covariates for May 2008, which have been lagged one-month to simulate an actual prediction.

[6] Although the dependent variable is a count, predictions made with the 'glm' package using the gaussian distribution consistently outperforms those build with the 'poisson' distribution.

(4) Calculate and store an *Ensemble* prediction by matrix multiplying the predicted values for $Model\_1$ and $Model\_2$ by their coefficient estimates from the *Ensemble* model trained on the in-sample set in Step 3.
(5) Incorporate May 2008 into the in-sample set.
(6) Repeat Step 1 through Step 4.
(7) Repeat Step 1 through Step 6 until a final prediction is made for April 2012 (i.e. the last month in the data set), using a model trained on February 2001 through March 2012.

This *Ensemble* model outperforms the naive benchmark in 33 out of 48 months. Although this is not a terrible result, it does not approach the accuracy of the more straightforward, univariate ARFIMA model discussed in the previous section. However, given the large number of predictive algorithms and the infinite number of features that can be built from a univariate time-series, scholars in the future may be able to build on my ensemble approach and build a model that eventually outperforms the predictive accuracy of my straightforward ARFIMA model.

6.2. **Incorporating drug prices.** In addition to building features from the univariate time-series as performed in the previous section, another way of potentially improving forecast accuracy is to incorporate exogenous variables. Although a large number of studies have found empirical relationships between many exogenous variables and political conflict, most operate at a state-year level of analysis. Finding relevant exogenous variables at sub-annual and sub-state levels is far more difficult. Even studies that do utilize fine-grained exogenous variables, like Weidmann and Ward [2010] and Berman et al. [2011] face considerable limitations.

For example, Weidmann and Ward [2010] analyze future violence at the municipality-month unit of analysis as a function of past violence as well as a set of exogenous variables comprised of population, ethnic diversity, terrain, and whether the municipality is on an international border. However, these exogenous variables vary cross-sectional (i.e. between municipalities) but not temporally (i.e. from month-to-month for the same municipality), which reduces the extent to which they can improve predictive accuracy. Additionally, Berman et al. [2011] collect unemployment statistics at the province-month level for Afghanistan, Iraq, and the Philippines that do vary at a province-month unit of analysis, but the difficulty in collecting such data limit their temporal domain to just six months in the case of Afghanistan, which also inhibits their effectiveness at enhancing predictive models. Therefore, an ideal set of exogenous variables would vary at a fine grained unit of analysis and span a long temporal range, but these are difficult to collect, especially for conflict-prone countries like Afghanistan.

For Afghanistan, one potential source of an exogenous variables come from the Afghanistan Opium Survey 2012, which is published by the United Nations Office on Drugs and Crime (UNODC).[7] This document provides considerable information at the district-level regarding opium and cannabis prices as well a dataset containing average opium prices at the country-month unit of analysis from September 2004 through March 2012, as illustrated below in Figure 3. Unfortunately, similarly complete time-series data are not publicly provided at the province- or district-month level.

[INSERT FIGURE 3 HERE]

Given the number of empirical studies that either theoretically suggest or empirically demonstrate relationships between drug prices and conflict (see Palmer [1994], Buhaug and Gates [2002], Ross [2003], Ross [2004], and Collier et al. [2004]) it seems reasonable that the addition of opium prices as an exogenous variable may enhance predictive accuracy at the country-month unit of analysis. To test this, I repeat the six steps outlined in Section 4.2 in order to compare the predictive accuracy of the naive model with the original univariate ARFIMA model outlined in Section 4 and Section 6.2 as well as the ARFIMA model that includes the exogenous opium data, which I call the ARFIMA_opium model. Since the opium price data spans a smaller temporal range than my GDELT-derived data on political violence, I set September 2004 through March 2010 as the initial in-sample training set, and use April 2010 through March 2012 as the out-of-sample test months. As Table 3 indicates, the ARFIMA model outperforms the Naive model in 18 of the 24 months that serve as the out-of-sample test months. Interestingly, the ARFIMA_opium model only outperforms the Naive model in 17 out of 24 months. Although this suggests that the inclusion of the drug price data may not actually enhance predictive accuracy, it does not rule out the possibility that more nuanced data on drug prices at the province- or district-level of analysis could lead to more accurate predictions.

7. Conclusion

[7] This document is available at:
www.unodc.org/documents/afghanistan/New%20Country%20Programme/ORAS_report_2012.pdf

This paper is the first to build temporally and geo-spatially nuanced forecasts of future levels of violence relying exclusively on open-source, machine coded event data. The release of the GDELT dataset made this article possible. Before GDELT, the leading open-source, machine-coded datasets did not provide location information, and the hand-coded datasets that did provide location information were too sparse for rigorous empirical forecasting. The Afghan War Diary that was released as past of WikiLeaks provided a notable exception, but this data is not only of questionable legality but also unlikely to be replicable for future conflicts, meaning that forecasting models built from WikiLeaks data may lack real-world applicability moving forward.[8]

Using nothing but GDELT data, I build an ARFIMA model capable of providing forecasts at the district month level that nearly always outperform a naive model that simply assumes that the level of conflict tomorrow will be the same as it is today. My empirical findings suggests three major takeaways: First, it appears that it is feasible to build accurate and nuanced predictions at a sub-state level using only open source, machine-coded event data. Second, the level of forecast accuracy decreased as the degree of geo-spatial aggregation increases: forecasts at the district-month (N=317), province-month (N=32) and country-month (N=1) level outperform their naive benchmarks in 47 out of 48, 40 out of 48, and 30 out of 48 month, respectively. It appears that patterns in violence that are discernible at fine-grained levels of geo-spatial aggregation (i.e. the district-level in Afghanistan) become increasing noisy a higher levels of geo-spatial aggregation. This strong suggests that researcher attempting to build empirical forecasts of violence should use as finely grained geo-spatial aggregations as possible. Third, the fact that the ARFIMA model tends to outperform the naive model suggests that patterns of violence tend to be mean reverting. This means that when we see a major spike in violence during a specific period of time in a specific sub-state location, we should expect violence in the following time period to be more subdued. Conversely, when we see a sudden drop in the level of violence, we should expect a rebound-effect.

Moving forward, a number of logical extensions to this article exist. First, researchers could use the GDELT data to further explore whether the mean-reversion properties present in the levels of violence in Afghanistan hold across other countries. Mean-reversion properties, as first identified by Galton [1886] in his seminal analysis of human heights, is a common and influential property across other substantive fields like biology and economics. Determining whether local levels of violence in

[8]Standard questioning when applying to positions require top-secret clearance is whether you have accessed and used Wikileaks data.

other states also tend to be mean-reverting could be a major theoretical advancement to the study of conflict dynamics.

Second, Section 6.1 provides a basic framework for building additional features from the univariate time series and using these features to construct alternative forecasting algorithms to the ARFIMA model. Although my attempts at enhancing predictive accuracy through this approach were unsuccessful, other scholars find greater success by building additional features and experimenting with other predictive algorithms. Similarly, the inclusion of additional exogenous variables, such as drug prices at finer grained spatial coverage than the country-level data modeling in Section 6.2, terrain, or measures of reflecting potential geo-spatial correlation (i.e. a count of the number of conflictual events occurring in neighboring districts or provinces) may also be helpful.

Third, since GDELT provides event data for all countries in the world (as opposed to WikiLeaks, which only provides detailed data for Afghanistan) researcher could apply a similar forecasting model to that outlined in this article to build geo-spatially and temporally nuanced forecasts of future levels of violence any number of countries with ongoing domestic conflicts, like India or the Democratic Republic of the Congo.

Lastly, since the GDELT data is updated daily, the forecasting approach outlined in this article could be implemented in near real-time. This could provide real-world guidance to a host of potential benefactors, ranging from military leaders hoping to more efficiently allocate resources, to Afghani businessmen trying to identify the safest routes to transport goods. Overall, I hope that this articles seres as a foundation for further forecasting efforts at fine-grained temporal and geo-spatial scales.

References

Benjamin E. Bagozzi. Forecasting civil conflict with zero-inflated count models. Available at: http://eventdata.psu.edu/papers.dir/Bagozzi.ZICM.2011.pdf, December 2011.

John Barkoulas and Christopher F. Baum. Long-memory forecasting of us monetary indices. *Journal of Forecasting*, 25:291–302, 2006.

Nathaniel Beck, Gary King, and Langche Zeng. Improving quantitative studies of international conflict: A conjecture. *American Political Science Review*, 94(1):21–36, 2000.

Eli Berman, Michael Callen, Joseph H. Felter, and Jacob N. Shapiro. Do working men rebel? insurgency and unemployment in afghanistan, iraq, and the philippines. *Journal of Conflict Resolution*, 55(4):496–528, 2011.

Geetesh Bhardwaj and Norman R. Swanson. An empirical investigation of the usefulness of arfima models for predicting macroeconomic and financial time series. *Journal of Econometrics*, 131: 539–578, 2006.

Joe Bond, Vladimir Petroff, Sean O'Brien, and Doug Bond. Forecasting turmoil in indonesia: An application of hidden markov models. Presented at the International Studies Association Meetings, Montreal, 2004.

Patrick T. Brandt and John R. Freeman. Advances in baysian time series modeling and the study of politics: Theory testing, forecasting, and policy analysis. *Political Analysis*, 14:1–36, 2005.

Patrick T. Brandt, Michael P. Colaresi, and John R. Freeman. The dynamics of reciprocity, accountability and credibility. *Journal of Conflict Resolution*, 52(3):343–374, 2008.

Patrick T. Brandt, John R. Freeman, and Philip Schrodt. Real time, time series forecasting of inter- and intra-state political conflict. *Conflict Management and Peace Science*, 28(1):41–64, 2011.

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

Halvard Buhaug and Scott Gates. The geography of civil war. *Journal of Peace Research*, 39(4): 417–433, 2002.

Nazli Choucri. Forecasting in international relations: Problems and prospects. *International Interactions*, 1:63–86, 1974.

Fong-Lin Chu. Forecasting tourism demand with arma-based methods. *Tourism Management*, 30: 740–751, 2009.

Paul Collier, AAnke Hoeffler, and Mans Soderbom. On the duration of civil war. *Journal of Peace Research*, 41(3):253–273, 2004.

Sarah Zukerman Daly. Organizational legacies of violence: Conditions favoring insurgency onset in colombia, 1964-1984. *Journal of Peace Research*, 49(3):473–491, 2012.

Vito D'Orazio, James E. Yonamine, and Philip A. Schrodt. Predicting intra-state conflict onset: An event data approach using euclidean and levenshtein distance measures. Presented at the annual Midwest Political Science Association meeting, Chicago, 2011.

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

John R. Freeman. Systematic sampling, temporal aggregation, and the study of political relationships. *Political Analysis*, 1:61–98, 1989.

Francis Galton. Regression towards mediocrity in hereditary stature. *Journal of the Anthropological Institute of Great Britain and Ireland*, 15:246–263, 1886.

Kristian Skrede Gleditsch and Kyle Beardsley. Noisy neighbors: Third-party actors in central american conflicts. *Journal of Conflict Resolution*, 48(3):91–119, 2004.

Joshua S. Goldstein. Reciprocity in superpower relations: An empirical analysis. *Journal of Conflict Resolution*, 36:369–385, 1991.

Joshua S. Goldstein. A conflict-cooperation scale for weis events data. *Journal of Conflict Resolution*, 36:369–385, 1992.

Jack A. Goldstone, Robert H. Bates, David L. Epstein, Ted Robert Gurr, Michael B. Lustik, Monty G. Marshall, Jay Ulfelder, and Mark Woodward. A global model for forecasting political instability. *American Journal of Political Science*, 54(1):190–208, 2010.

Clive William Granger and Roselyne Joyeux. An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis*, 1(1):15–29, 1980.

Ted Robert Gurr and Barbara Harff. Early warning of communal conflict and humanitarian crisis. In *Monograph Series on Governance and Conflict Resolution*. United Nations Press, 1996.

Ted Robert Gurr and Mark Irving Lichbach. Forecasting internal conflict: A competitive evaluation of empirical theories. *Comparative Political Studies*, 19(3):1–37, 1986.

Barbara Harff and Ted Robert Gurr. Systematic early warning of humanitarian emergencies. *Journal of Peace Research*, 35(5):359–371, 2001.

Gary King and Will Lowe. An automated information extraction tool for international conflict data with performance as good as human coders: A rare events evaluation design. *International Organization*, 57(3):617–642, 2004.

Gary King and Langche Zeng. Improving forecasts of state failure. *World Politics*, 53:623–658, 2001.

Andrew M. Linke, Frank D Witmer, and John O'Loughlin. Space-time granger analysis of the war in iraq: A study of coalition and insurgent action-reaction. *International Interactions*, 38: 402–425, 2012.

Andrew Mangion-Zammit, Michael Dewar, Visakan Kadirkamanathan, and Guido Sanguinetti. Point process modeling of the afghan war diary. *Proceedings of the National Academy of Science*, 109(31):12414–12419, 2012.

Jacob Montgomery, Florian Hollenbach, and Michael D. Ward. Improving predictions using bayesian model averaging. *Political Analysis*, 20(3):271–291, 2012.

Sean O'Brien. Crisis early warning and decision support: Contemporary approaches and thoughts on future research. *International Studies Review*, 12(1):87–104, 2010.

John O'Loughlin, Frank D.W. Witmer, Andrew M. Linke, and Nancy Thorwardson. Peering into the fog of war: The geography of wikileaks afghanistan war logs, 2004-2009. *Eurasian Geography and Economics*, 51(4):472–495, 2010.

David Scott Palmer. Peru, drugs, and the shining path. In Bruce M. Bagley and Wiliam O. Walker III, editors, *Drug Trafficking in the Americas*, pages 179–191. North-South Center Press, 1994.

Vladimir Petroff, Joe Bond, and Doug Bond. Using hidden markov models to predict terror before it hits (again). In V.S. Subrahmanian, editor, *Handbook on computational approaches to counterterrorism*. Springer, 2012.

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

Clionadh Raleigh, Andrew Linke, Havard Hegre, and Joakim Karlsen. Introducing acled: An armed conflict location event dataset. *The Journal of Peace Research*, 47(5):551–560, 2010.

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

Michael L. Ross. Oil, drugs, and diamonds: The varying role of natural resources in civil war. In Karen Ballentine and Jake Sherman, editors, *The Political Economy of Armed Conflict: Beyond Greed and Grievance*, pages 47–70. Lynne Rienner, 2003.

Idean Salehyan, Cullen S. Hendrix, Jesse Hamner, Christina Case, Christpher Linebarger, Emily Stull, and Jennifer Williams. Social conflict in africa: A new database. *International Interactions*, 38:503–511, 2012.

Gerald Schneider, Margit Bussman, and Constantine Ruhe. The dynamics of mass killings: Testing time-series models of one-sided violence in the bosnian civil war. *Journal of Peace Research*, 49 (3):473–491, 2012.

Philip A. Schrodt. Parallel event sequences in international relations. *Political Behavior*, 12(2): 97–123, 1990.

Philip A Schrodt. Early warning of conflict in southern lebanon using hidden markov models. Presented at the annual meeting of the American Political Science Association, Washington D.C., 1997.

Philip A. Schrodt. Early warning of conflict in southern lebanon using hidden markov models. In Harvey Starr, editor, *TThe Understanding and Management of Global Violence: New Approaches to Theory and Research of Protracted Conflict*, pages 131–162. St. Martin's Press, New York, 1999.

Philip A. Schrodt. Pattern recognition of international crises using hidden markov models. In Diana Richards, editor, *Political Complexity: Nonlinear Models of Politics*, pages 296–328. University of Michigan Press, Ann Arbor, 2000.

Philip A. Schrodt. Forecasting conflict in the balkans using hidden markov models. In Robert Trappl, editor, *Programming for Peace: Computer-Aided Methods for International Conflict Resolution and Prevention*, pages 161–184. Kluwer Academic Publishers, Dordrecht, Netherlands, 2006.

Philip A. Schrodt and Deborah J. Gerner. Empirical indicators of crisis phase in the middle east, 1979-1995. *Journal of Conflict Resolution*, 25(4):803–817, 1997.

Philip A. Schrodt and Deborah J. Gerner. Cluster-based early warning indicators for political change in the contemporary levant. *American Political Science Review*, 94(4):803–817, 2000.

Philip A. Schrodt and Deborah J. Gerner. Analyzing the dynamics of international mediation processes in the middle east and the former yugoslavia. Presented at the annual meeting of the

International Studies Association, Chicago, 2001.

Philip A. Schrodt, James Yonamine, and Benjamin E. Bagozzi. Data-based computational approached to forecasting political violence. In V.S. Subrahmanian, editor, *Handbook on computational approaches to counterterrorism*. Springer, 2012.

Phillip A. Schrodt. Inductive event data scaling using item response theory. Presented at the Summer Meeting of the Society of Political Methodology. Available at `http://eventdata.psu.edu`, 2007.

Robert Shearer. Forecasting israeli-palestinian conflict with hidden markov models. Available at `http://eventdata.psu.edu/papers.dir/Shearer.IP.pdf`, 2006.

Stephen Shellman. Time series intervals and statistical inference: The effects of temporal aggregation on event data analysis. *Political Analysis*, 12(1):97–104, 2004a.

Stephen Shellman. Process matters: Conflict and cooperation in sequential government-dissident interactions. *Security Studies*, 15(4):563–599, 2007.

Stephen Shellman, Clare Hatfield, and Maggie Mills. Dissagregating actors in intrastate conflict. *Journal of Peace Research*, 47(1), 2010.

Stephen M. Shellman. Measuring the intensity of international political interactions event data: Two interval-like scales. *International Interactions*, 30(2):109–141, 2004b.

Lim Ying Siew, Lim Ying Chin, and Pauline Mah Jin Wee. Arima and integrated arfima models for forecasting air pollution index in shah alam, selangor. *The Malaysian Journal of Analytics Sciences*, 12(1):257–263, 2008.

David Singer and Michael David Wallace. *To Auger Well: Early Warning Indicators in World Poliics*. Sage Press, Beverly Hills, CA, 1979.

Henrik Urdal and Kristian Hoelscher. Explaining urban social disorder and violence: An empirical study of event data from asian and subsaharan african cities. *International Interactions*, 38: 512–528, 2012.

Michael D. Ward, Brian D. Greenhill, and Kristin M. Bakke. The perils of policy by p-value: Predicting civil conflicts. *Journal of Peace Research*, 47(5), 2010.

Nils B. Weidmann and Idean Salehyan. Violence and ethnic segregation: A computational model applied to baghdad. Available at http://nils.weidmann.ws/publications/weidmann13baghdad, 2011.

Nils B. Weidmann and Michael D. Ward. Predicting conflict in space and time. *Journal of Conflict Resolution*, 54(6):883–901, 2010.

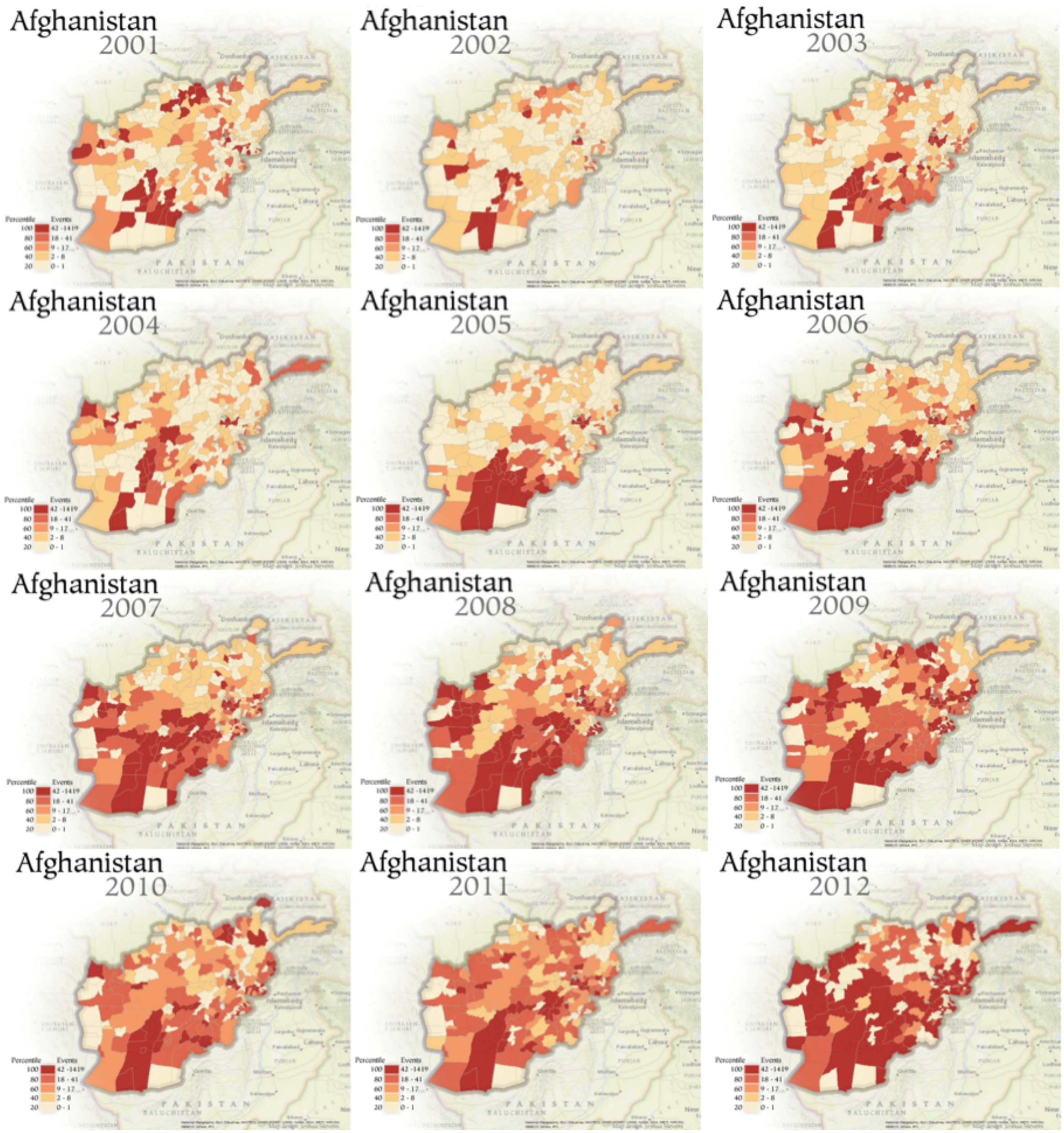

Figure 1. The Number of Material Conflict events per Afghani District from 2001 to 2012

8. Appendix

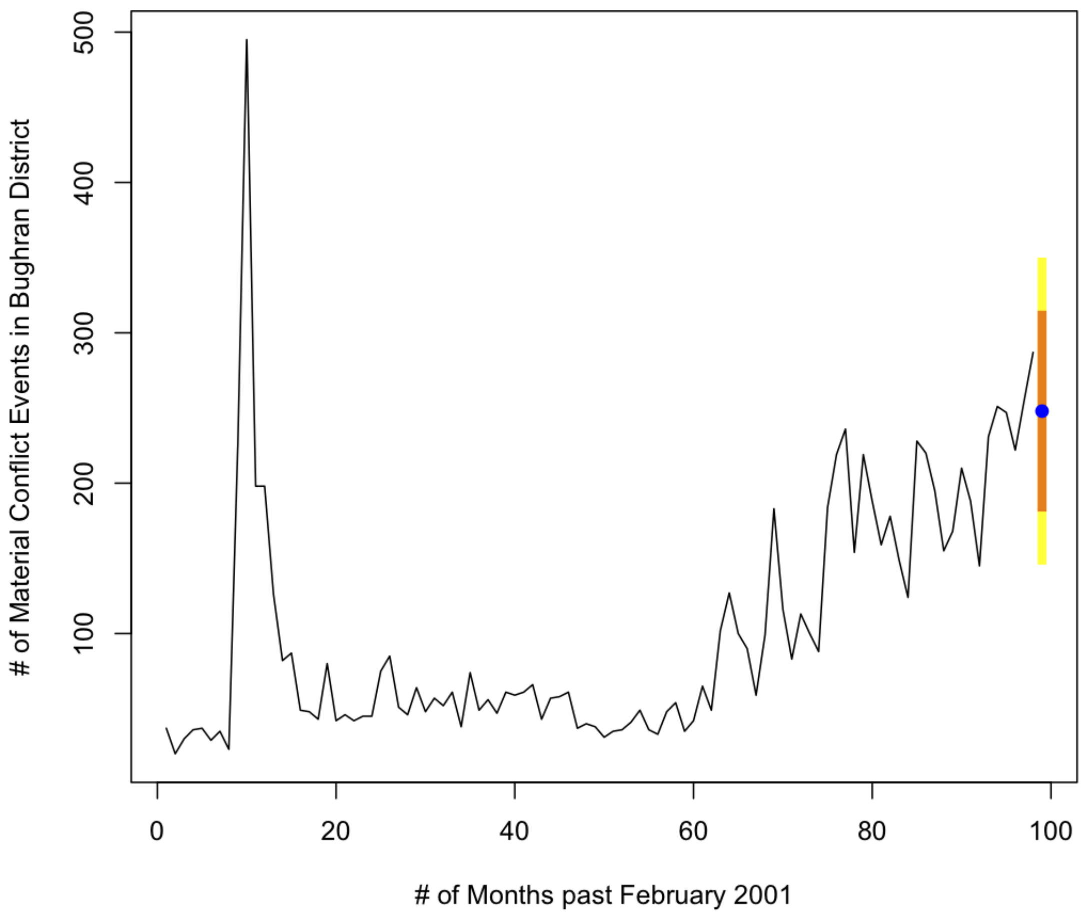

FIGURE 2. One-month Forecast of the of Material Conflict Events in Bughran District using 'arfima' package, with mean, 90%, and 95% confidence intervals.

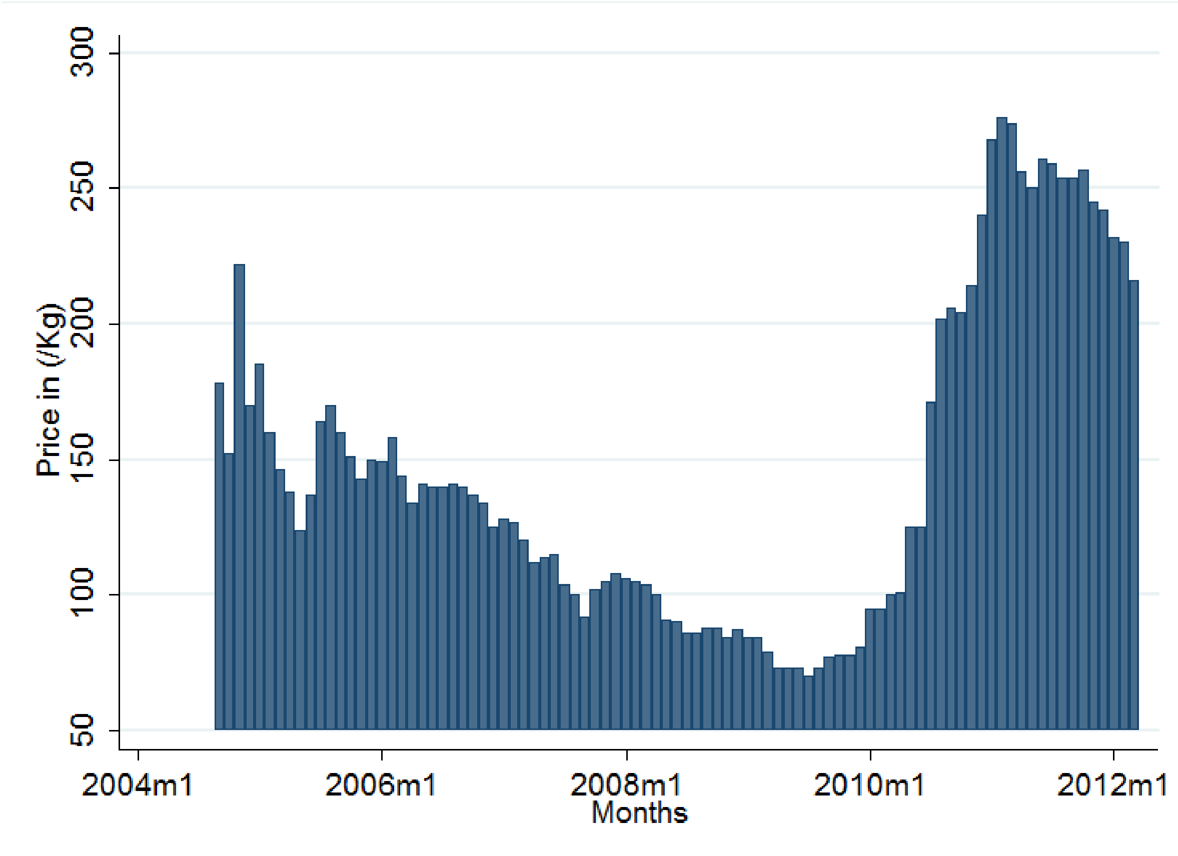

Figure 3. Average Farm-Gate Prices for Dry Opium in Afghanistan, September 2004-March 2012

TABLE 1. Assessing Accuracy at the District Level

m	month	arfima_error	naive_error	arfima_error $<$ naive_error
1	May 2008	2.73	3.59	TRUE
2	June 2008	2.75	3.21	TRUE
3	July 2008	3.76	4.41	TRUE
4	August 2008	2.40	3.66	TRUE
5	September 2008	2.58	3.13	TRUE
6	October 2008	3.51	3.60	TRUE
7	November 2008	2.24	2.87	TRUE
8	December 2008	1.71	2.58	TRUE
9	**January 2009**	**2.72**	**2.63**	**FALSE**
10	February 2009	2.11	2.53	TRUE
11	March 2009	2.53	3.11	TRUE
12	April 2009	2.43	2.85	TRUE
13	May 2009	3.58	4.02	TRUE
14	June 2009	2.85	4.02	TRUE
15	July 2009	3.56	4.07	TRUE
16	August 2009	5.50	6.25	TRUE
17	September 2009	4.15	5.00	TRUE
18	October 2009	4.18	4.79	TRUE
19	November 2009	3.38	4.09	TRUE
20	December 2009	3.07	3.09	TRUE
21	January 2010	2.50	3.62	TRUE
22	February 2010	3.82	4.11	TRUE
23	March 2010	3.37	4.15	TRUE
24	April 2010	1.66	1.70	TRUE
25	May 2010	1.99	2.12	TRUE
26	June 2010	2.02	2.25	TRUE
27	July 2010	1.77	2.09	TRUE
28	August 2010	3.63	4.06	TRUE
29	September 2010	3.28	3.28	TRUE
30	October 2010	2.05	2.57	TRUE
31	November 2010	1.77	2.22	TRUE
32	December 2010	2.00	2.31	TRUE
33	January 2011	2.02	2.44	TRUE
34	February 2011	1.91	2.34	TRUE
35	March 2011	1.89	2.18	TRUE
36	April 2011	3.47	4.03	TRUE
37	May 2011	2.91	3.64	TRUE
38	June 2011	2.32	3.07	TRUE
39	July 2011	3.06	3.60	TRUE
40	August 2011	2.64	3.30	TRUE
41	September 2011	3.02	3.47	TRUE
42	October 2011	1.97	2.79	TRUE
43	November 2011	2.32	2.68	TRUE
44	December 2011	1.81	2.10	TRUE
45	January 2012	2.21	2.42	TRUE
46	February 2012	2.09	2.41	TRUE
47	March 2012	2.81	3.08	TRUE
48	April 2012	2.98	3.58	TRUE
Total:	May 2008 - Apr 2012	129.76	155.07	47 TRUE, 1 FALSE

TABLE 2. Assessing Accuracy at the Province Level Level

m	month	arfima_error	naive_error	arfima_error < naive_error
1	May 2008	16.16	23.69	TRUE
2	**June 2008**	**22.34**	**21.53**	**FALSE**
3	July 2008	29.16	33.63	TRUE
4	August 2008	21.13	27.69	TRUE
5	September 2008	15.38	17.69	TRUE
6	**October 2008**	**24.16**	**24.09**	**FALSE**
7	November 2008	12.78	18.59	TRUE
8	December 2008	6.91	14.59	TRUE
9	**January 2009**	**18.03**	**17.06**	**FALSE**
10	February 2009	12.69	15.53	TRUE
11	March 2009	18.84	22.5	TRUE
12	April 2009	11.00	15.53	TRUE
13	May 2009	28.75	30.41	TRUE
14	June 2009	21.69	28.50	TRUE
15	**July 2009**	**31.91**	**31.22**	**FALSE**
16	August 2009	40.75	44.25	TRUE
17	September 2009	21.31	28.22	TRUE
18	**October 2009**	**36.44**	**36.16**	**FALSE**
19	November 2009	21.56	32.49	TRUE
20	**December 2009**	**21.34**	**19.31**	**FALSE**
21	January 2010	18.19	26.91	TRUE
22	February 2010	23.63	27.88	TRUE
23	March 2010	35.38	36.16	TRUE
24	**April 2010**	**19.00**	**10.38**	**FALSE**
25	May 2010	12.94	14.66	TRUE
26	June 2010	14.38	17.13	TRUE
27	July 2010	13.47	14.59	TRUE
28	August 2010	32.03	33.00	TRUE
29	September 2010	8.75	19.66	TRUE
30	October 2010	12.25	12.91	TRUE
31	November 2010	12.81	13.94	TRUE
32	December 2010	11.75	14.75	TRUE
33	January 2011	14.34	16.13	TRUE
34	February 2011	17.13	17.53	TRUE
35	March 2011	12.94	16.25	TRUE
36	April 2011	26.22	31.66	TRUE
37	May 2011	21.19	27.19	TRUE
38	June 2011	12.81	18.56	TRUE
39	July 2011	20.25	25.56	TRUE
40	August 2011	20.29	24.03	TRUE
41	September 2011	24.19	26.69	TRUE
42	October 2011	14.97	19.41	TRUE
43	November 2011	16.72	20.63	TRUE
44	December 2011	12.75	14.97	TRUE
45	January 2012	14.88	16.34	TRUE
46	February 2012	14.31	15.97	TRUE
47	March 2012	21.72	24.24	TRUE
48	**April 2012**	**93.09**	**91.66**	**FALSE**
Total:	May 2008 - Apr 2012	1,004.56	1,151.50	40 TRUE, 8 FALSE

Table 3. Assessing Accuracy at the Country Level Level

m	month	arfima_error	naive_error	arfima_error < naive_error
1	May 2008	52	94	TRUE
2	**June 2008**	**462**	**393**	**FALSE**
3	**July 2008**	**474**	**410**	**FALSE**
4	August 2008	358	426	TRUE
5	September 2008	96	238	TRUE
6	**October 2008**	**383**	**277**	**FALSE**
7	November 2008	5	81	TRUE
8	December 2008	135	231	TRUE
9	**January 2009**	**293**	**204**	**FALSE**
10	**February 2009**	**39**	**25**	**FALSE**
11	**March 2009**	**156**	**78**	**FALSE**
12	**April 2009**	**51**	**23**	**FALSE**
13	**May 2009**	**709**	**629**	**FALSE**
14	June 2009	393	444	TRUE
15	**July 2009**	**455**	**309**	**FALSE**
16	**August 2009**	**845**	**754**	**FALSE**
17	September 2009	135	227	TRUE
18	October 2009	220	311	TRUE
19	November 2009	235	417	TRUE
20	**December 2009**	**484**	**336**	**FALSE**
21	January 2010	633	693	TRUE
22	February 2010	71	20	TRUE
23	March 2010	1,035	1,087	TRUE
24	April 2010	213	244	TRUE
25	May 2010	244	259	TRUE
26	June 2010	137	146	TRUE
27	July 2010	139	143	TRUE
28	**August 2010**	**1,037**	**1,028**	**FALSE**
29	September 2010	515	519	TRUE
30	October 2010	85	85	TRUE
31	November 2010	170	200	TRUE
32	**December 2010**	**185**	**174**	**FALSE**
33	January 2011	160	204	TRUE
34	February 2011	220	223	TRUE
35	March 2011	253	270	TRUE
36	**April 2011**	**539**	**507**	**FALSE**
37	**May 2011**	**144**	**66**	**FALSE**
38	June 2011	135	230	TRUE
39	**July 2011**	**386**	**298**	**FALSE**
40	**August 2011**	**126**	**33**	**FALSE**
41	September 2011	120	202	TRUE
42	October 2011	314	371	TRUE
43	November 2011	214	266	TRUE
44	December 2011	55	57	TRUE
45	January 2012	5	23	TRUE
46	February 2012	112	145	TRUE
47	March 2012	453	475	TRUE
48	**April 2012**	**2,759**	**2,737**	**FALSE**
Total:	May 2008 - Apr 2012	16,439	16,612	30 TRUE, 18 FALSE