

PREDICTING INTRA-STATE CONFLICT ONSET: AN EVENT DATA APPROACH USING EUCLIDEAN AND LEVENSHTEIN DISTANCE MEASURES

VITO D'ORAZIO, JAMES E. YONAMINE, AND PHILIP A. SCHRODT

1. INTRODUCTION

The goal of much of the large- N quantitative literature focusing on conflict onset has been to find generalizable empirical patterns across time and space. To do so, most studies have utilized a time-series cross sectional approach to analyze the effects of “structural” variables aggregated at the state or dyad-year level. These studies have found hundreds if not thousands of statistically significant relationships between state-year variables at a period T and the likelihood of a conflict onset at a period $T+N$ that are generalizable across a broad spatial and temporal domain. However much these types studies enhance our understanding of conflict, they are not conducive to forecasting conflicts. Ward et al. [2010] show that mere statistical significance provides no guarantee that models can predict and most existing models and with the exception of those explicitly designed to predict (e.g. Goldstone et al. [2010], O'Brien [2010]), are poor at prediction. But, as argued forcefully by philosophers of science such as Quine [1951] and Hempel [2001], predictive accuracy is a requirement for any genuinely scientific theory. The difficult of prediction is especially pressing for structural models of intrastate conflict, as key factors that account for the actual sub-annual level events that drive conflict—for example the level and type of interactions between government and rebel forces—are rarely accounted for in structural models.

To this end, a smaller but substantial body of literature often classified as “early warning models” exist, which analyze temporally nuanced data—generally in the form of political event data—and focus on the effects of specific types of events on conflict processes. Event data approaches are able to account for the constant flux of domestic conditions at the sub-annual level. Unlike structural

Date: Version 0.8 : March 24, 2011.

Paper prepared for delivery at the Annual Meeting of the Midwest Political Science Association, Chicago, 31 March - 3 April 2011. This project was funded by a contract from the Defense Advanced Research Projects Agency under the Integrated Crisis Early Warning System (ICEWS) program (Prime Contract #FA8650-07-C-7749: Lockheed-Martin Advanced Technology Laboratories). The views, opinions, and/or findings contained in this paper are those of the authors and should not be interpreted as representing the official views or policies, either expressed or implied, of the Defense Advanced Research Projects Agency or the Department of Defense, nor those of Lockheed-Martin.

models, which predominantly utilize annual, time-series cross-section data, the majority of leading event data models are state or region (such as the Levant or former states of Yugoslavia) specific. When event data studies do analyze conflicts in multiple countries, they tend to either pool the data for a limited number of states or conduct an independent empirical study for each individual country or dyad (e.g. Israel-Palestine or Israel-Lebanon) and then compare findings. Although these approaches have uncovered many important findings about conflict processes, the available data and established methodological approaches have largely inhibited the ability of scholars to test for generalizable patterns across time and space, which serves as the cornerstone for much of the structural model literature.

This lack of event data studies analyzing trends across a large number of states has implications for both our theoretical understanding of the way domestic conflict unfolds and for our ability to generate accurate predictions. If sub-annual conflictual processes—and, more specific to our analysis, escalations towards conflict—exhibit generalizable trends across time and space, then uniform forecasting models could be successfully applied on a broad scale. On the other hand, if conflicts follow unique escalatory processes, then models derived from trends found in large-N studies may not be appropriate approach for forecasting conflict onset.

In this paper, we utilize both a new event data dataset and a novel empirical approach to test for the extent of similarities between event sequences preceding conflict onset. The event data were compiled as part of the Defense Advanced Research Project Agency (DARPA) funded Integrated Conflict Early Warning Systems program (ICEWS; O'Brien [2010]) and consist of machine-coded events between politically relevant actors in 29 Asian countries from 1998-2010, derived from electronic articles reported in 27 regional and international electronic news sources. Our methodological innovation is the use of Levenshtein and Euclidean distance measures to assess the level of similarity between two 12-week sequences of event counts aggregated at the state-week and state-month level. Distances are calculated between all possible pairs of the 53 sequences within our dataset that precede a conflict onset and approximately 100 sequences that precede peaceful period with no domestic conflict. We find that sequences preceding peace exhibit highly similar event count structures while those preceding a conflict onset vary considerably from each other and, importantly, from periods preceding peace. Based on this, we calculate a Baseline Peace Archetype (BPA) that reflects that “average” event count structure of a sequence preceding peace. We then calculate the

distances measures between all sequences preceding conflict and a random sample of those preceding peace against the BPA sequence. These distances allow us to generate out-of-sample forecasts one and two months into the future with greater than 75 percent accuracy.

The paper proceeds by introducing the ICEWS event dataset and discussing our data aggregation techniques and research design. We then present our forecasting results and the BPA models, then conclude with suggestions for further research.

2. OVERVIEW OF EXISTING EVENT DATASETS AND INTRODUCTION OF ICEWS

Political event data have had a long presence in the quantitative study of international politics, dating back to the early efforts of Edward Azar’s COPDAB [Azar, 1980] and Charles McClelland’s WEIS [McClelland, 1967] as well as a variety of more specialized efforts such as Leng’s BCOW [Leng, 1987]. By the late 1980s, the NSF-funded *Data Development in International Relations* project Merritt et al. [1993] had identified event data as the second most common form of data—behind the various Correlates of War data sets—used in quantitative studies including Merritt et al. [1993] and McGowan et al. [1988]. The 1990s saw the development of two practical automated event data coding systems, the NSF-funded KEDS [Gerner et al., 1994, Schrodts and Gerner, 1994] and the proprietary VRA-Reader (<http://vranet.com>; King and Lowe [2004]) and in the 2000s, the development of two new political event coding ontologies—CAMEO [Schrodts et al., 2009] and IDEA [Bond et al., 2003]—designed for implementation in automated coding systems.

While these efforts had built a substantial foundation for event data—for example by the mid-2000s, virtually all refereed articles in political science journal used machine-coded, rather than human-coded, event data—the overall investment in the technology remained relatively small. The KEDS¹ project focused on coding a small number of regions, mostly in the eastern Mediterranean [Schrodts and Gerner, 2000, Shellman, 2000, Schrodts, 2006, 2007, Pevehouse and Goldstein, 1999]. VRA did global coding—and made a 1990-2004 global data set of around 10-million events available through Harvard University (<http://gking.harvard.edu/events/>). Bond et al. [1997] use the KEDS/PANDA dataset to compare conflict escalations in China, Poland, South Korea, and the Balkans, while Jenkins and Bond [2001] analyze a selection of seven countries.

¹Kansas Event Data System, a term used to refer to both the overall project which existed at the University of Kansas from 1990 to 2009, and the original coding program, which was written in Pascal for the Macintosh operating system. That project has now moved to Penn State but, despite months of effort, we have yet to come up with a clever name for it that does not produce an acronym that sounds like an obscure skin disease. Suggestions are welcome. Meanwhile we are just calling it the “Penn State Event Data Project.”

This situation changed dramatically with the DARPA-funded Integrated Conflict Early Warning System (ICEWS; O'Brien [2010], Schrodt [2010]) which invested very substantial resources in event data development, including both TABARI—the open-source successor to KEDS —and VRA (the latter used as one of the sources to generate the ICEWS “events of interest”; [O'Brien, 2010, pg. 91]). The Lockheed-Martin ICEWS effort have made extensive use of TABARI—both in the original form and a direct translation (with some bug corrections) of TABARI into Java called JABARI — and the data produced by that system will be the focus of this paper. The ICEWS dataset uses a 15,000-item actor dictionary to code for a broad range of domestic actors, including but not limited to military, police, rebel groups, and civilians, and this allows for detailed analyses of domestic events.

The key difference between the ICEWS event data coding efforts and those of earlier NSF-funded efforts was the scale. As O'Brien—the ICEWS project director—notes,

... the ICEWS performers used input data from a variety of sources. Notably, they collected 6.5 million news stories about countries in the Pacific Command (PACOM) AOR [area of responsibility] for the period 1998-2006. This resulted in a dataset about two orders of magnitude greater than any other with which we are aware. These stories comprise 253 million lines of text and came from over 75 international sources (AP, UPI, and BBC Monitor) as well as regional sources (*India Today*, *Jakarta Post*, *Pakistan Newswire*, and *Saigon Times*).

To the best of our knowledge, the ICEWS event database, which has been updated to 2010 and contains over 2,000,000 events for 29 Asian countries is the most accurate event data set currently available.² In addition, systematic tests have shown that the current version of JABARI increases coding accuracy substantially over traditional TABARI software and the use of multiple regional news sources provides more comprehensive coverage of countries that tend to receive little media attention from western outlets (Fiji and the Comoros, for example).

3. AGGREGATION TECHNIQUES

For all politically relevant and codeable events, JABARI determines the initiator of the action, the action itself, and the target of the action, classifying all events into 20 parent categories based

²It is interesting to note that despite being produced by completely independent technologies, the ICEWS and VRA data sets have very similar densities: 5300 events per country-year ICEWS and about 4500 events per country-year for VRA on the assumption that most events occur in the 150 or so countries with populations greater than 500,000.

on the Conflict and Mediation Event Observations (CAMEO) coding system. In order to convert raw event data to a format suitable for empirical testing, aggregation accounting for actor, actions, and temporal domains of interest must occur. The high volume and level of detail of event data allow provide researchers with a tremendous level of flexibility regarding aggregation techniques in three primary areas: 1) actors; 2) actions; 3) unit of time. In this section, we address the methods we adopt for aggregating across these three areas while providing a brief overview of approaches used in relevant literature.

3.1. Actor Aggregation.

All existing event datasets attempt to categorize the actors who carry out the events of interest. Given this, researchers must determine the actors of interest between whom an event must occur in order for that event to be included in their study's empirical models. At a minimum this means events that occur between at least one actor affiliated with a country of interest. The justification for this minimal level of actor aggregation is clear; a study focusing on Israeli-Palestinian conflicts would not want to include events between Aceh rebels and the Indonesian army, as these are not relevant to the conflict of interest. Although excluding Indonesian rebel activity would be an obvious, more difficult decisions exist, such as whether or not to include events between members of the Lebanese and Syrian armies, or the governments of the United States and Iran in a study of conflict between Israel-Palestine.

Even more challenging are questions regarding which domestic actors are politically relevant to studies of conflict. For example, should a bar fight between two Israeli citizens over the outcome of a soccer match be treated identically as a fight occurring between an Israeli and a Palestinian government official? Unfortunately, the majority of studies utilizing event data provide incomplete explanations regarding the specific actors between who events must occur to enter the dataset while others provide a conceptual discussion of actors of interest but do not include how these concepts are operationalized Bond et al. [1997] discuss "mass" and "state" actors; Harff and Gurr [2001] discuss "governing elites", "mass followship", "disadvantaged groups", etc; ? address "kindred groups", "communal groups", etc; Shellman [2000] discusses "government" and "dissidents"). Given the lack of formal precedence regarding actor aggregation approaches, we attempt to explain our approach as clearly as possible. Below, we provide a thorough description of our actor-aggregation approach in a level of detail that we believe should exist in studies using event data.

Every coded event in the ICEWS dataset contains two actors in a who-did what to-whom format. ICEWS provides the national identity of both actors (the who and the whom) as well as one of dozens of sub-national level descriptions when applicable. We divide these sub-national level descriptions and collapse them into one of three main classes of actors:

- *Government*, which includes actors identified by ICEWS as:
 - Armed Forces
 - Police
 - Bureaucrats
 - Politicians
- *Rebels*, which includes actors identified by ICEWS as:
 - Insurgents
 - Separatists
 - Rebels
- *Other*, which includes by ICEWS as:
 - Civilians unaffiliated with another group
 - Actors who are unidentifiable by JABARI

Theoretically, we believe that events occurring between governments and rebel groups (be they labeled as insurgents, separatists, terrorists, etc) and between rebel groups and other actor non-governmental sectors in a country drive escalatory processes. Consequently, we aggregate the subcategories into the three broader classes, meaning that both police officers and members of the army are treated uniformly as “Government”. Next, we focus exclusively on events that occur between a “Government” actor and a “Rebel” (called GOV-REB events) and between “Rebels” and “Other” (called REB-OTHER), establishing two classes of variables. This method of actor aggregation is not perfect, but we believe it does provide useful detail.

3.2. Action Aggregation. The majority of extant event data literature either scales all events, assigning them a score on a conflict-cooperation continuum or generates event counts reflecting the number of events that occur within conceptually unique categories. The Goldstein Scale (Goldstein [1992]), which is the most commonly used scaling technique within the event data literature (see Goldstein [1992], Schrodtt [2007], Schrodtt and Gerner [1994], Pevehouse and Goldstein [1999], Hämmerli et al. [2006], for sample uses) assigns a value to all events coded under the World Event

Interaction Survey (WEIS) scheme on a -10 to 10 scale conflict/cooperation scale, with -10 reflecting the most conflictual events and 10 indicating the most cooperative is the most commonly used. Despite its dominance within the event data literature, the Goldstein scale requires additional levels of aggregation beyond the initial scaling, which leads to a number of operational difficulties. For example, consider a day on which an armed killing (which receives a -10 score) and a peace-treaty signing (which receives a +10 score) occur on the same day between the same actors. Summing Goldstein scores would result in a net score of 0 in the previous example, which is the same score that days with no activity receive. Additionally, taking the mean score would also result in a daily score of 0. Theoretically, it is apparent that the nature of events occurring on a day comprised of purely neutral events that receives a score of 0 and that of a day with a -10 event and a +10 event are fundamentally different. While this example of two events exactly canceling is hypothetical, the problem of violent events masking the concurrent presence of cooperative actions—notably negotiations occurring during periods of on-going violence—is very real, and occurs frequently during such periods when the KEDS Levant and Balkans data are aggregated using Goldstein scores.³

Despite the preponderance of the Goldstein scale, a number of other studies (Schrodt et al. [2001], Schrodt and Gerner [2004], Shearer [2006]) utilize count measures. Duval and Thompson [1980] put forth the first event data count model by placing all events into one of the four conceptually unique, mutually exclusive categories below, which we have translated from the WEIS system used in the original article to CAMEO using the categories listed:

- *Verbal Cooperation*: The occurrence of dialogue-based meetings (i.e. negotiations, peace talks), statements that express a desire to cooperate or appeal for assistance (other than material aid) from other actors. CAMEO categories 01 to 05.
- *Material Cooperation*: Physical acts of collaboration or assistance, including receiving or sending aid, reducing bans and sentencing, etc. CAMEO categories 06 to 09.
- *Verbal Conflict*: A spoken criticism, threat, or accusation, often related to past or future potential acts of material conflict. CAMEO categories 10 to 14.
- *Material Conflict*: Physical acts of a conflictual nature, including armed attacks, destruction of property, assassination, etc. CAMEO categories 15 to 20.

³This is further complicated by the fact that comments and meetings have Goldstein scores that are small in magnitude, whereas violent events have a scale score of -10. Consequently a small amount of violence can mask a lot of talking. A similar problem plagued the scaled scores of the COPDAB data set, where the quip was made that “In COPDAB, three riots equals a thermonuclear war.”

The Duval and Thompson [1980] event count structure makes fewer assumptions than the Goldstein Scale about the impact of events. Moreover, because negative counts are impossible, the this approach does not suffer from problems of generating sum or mean scores that affect the Goldstein Scale. In particular, because scaling systems generally assign high magnitude values to uses of force, and low magnitude values to meetings, it is difficult to differentiate conflict periods where the antagonists are also engaging in meetings from those where they are not, and yet the political import of those two situations is quite different. Although this count approach is more simplistic than scaling methods, Shearer [2006] and Schrodts [2006] find strong empirical results with data aggregated into the four counts. Bond et al. [1997] and Jenkins and Bond [2001] utilize a different type of event count structure, which places all events into one of eight boxes which reflect whether an event is violent or non-violent and direct or indirect. Next, Bond et al. [1997], “define carrying capacity as the proportion of contentious or direct action, multiplied by the proportion of the violent (forceful) to all direct action, and subtracted from unity [1] to facilitate interpretation.” This results in a “Conflict Carrying Capacity” score from 0-1 for each country in their study.⁴

Despite the dominance of the Goldstein scale and sophistication of more recent count structure, we employ a parallel count technique to Duval and Thompson [1980] but adjusted for the CAMEO categories as indicated above. We believe that this approach provides most straightforward representation of the number and type of events that occur—especially relative to the “Conflict Carrying Capacity” approach—while making the fewest assumptions about future effects of the events that cannot be known *ex ante*. Given the empirical complexity of the sequence analysis presented in Section 4, this level of simplicity is required in order to generate interpretable findings.

3.3. Temporal Aggregation. Finally, scholars must temporally aggregate data in order to perform empirical analyses at levels appropriate for their theory or empirical models of choice. All of the previously mentioned event data datasets code the exact day on which events occur. As the specific time-of-day that events occurred is not reported, events must at the very minimum be aggregated to the daily level (Pevehouse and Goldstein [1999], Shearer [2006], Schrodts [2006]), though weekly (Brandt and Freeman [2005], Shellman and Stewart [2007]), monthly (Schrodts [2007], Ward et al. [2010]), quarterly (Jenkins and Bond [2001]), and annual level aggregations are present

⁴The VRA CCC approach differs from the Goldstein scale in its use of a ratio of counts, and more generally the “Cambridge” approach of VRA and various Harvard-based studies generally employs ratios and average scaled values rather than the counts and total scaled values used in most of the KEDS project studies. Our sense is that the relative effectiveness of the two approaches is still an open question.

within the literature. Although the majority of studies using event data aggregate according to traditional divisions of time, convenience rather than theory or statistics tends to motivate this approach. Despite this, some level of temporal aggregation is inherently mandatory to any research design using event data. We chose to utilize weekly level aggregations as this allows us to capture event fluctuations that occur across small periods of time and is more robust to noise and than daily level aggregations. Additionally, we assert that the state-week is logical as religious holidays and working schedules tend to occur on a 7-day cycle.⁵ Due to the findings in a number of studies that highlight the way in which different temporal aggregations on the same data can affect empirical results (Alt et al. [2001], Dale [2002], Shellman [2004]), we perform rigorous robustness checks using monthly level temporal aggregations.

To generate weekly and monthly state-counts, we sum the event counts for all relevant actor-source combinations, resulting in 8 total count variables. State weeks and months during which no recorded events occur receive a “0” for all four count variables above. Table 1 provides a four-week example of data output for each state.

TABLE 1. Sample 4-Week Event Count Sequence for State “A”

Count Variable	Week 1, 1999	Week 2, 1999	Week 3, 1999	Week 4, 1999
GOV-REB Verbal Cooperation	5	2	3	4
GOV-REB Material Cooperation	5	7	4	5
GOV-REB Verbal Conflict	4	2	1	4
GOV-REB Material Conflict	0	0	2	3
REB-OTHER Verbal Cooperation	3	3	4	1
REB-OTHER Material Cooperation	0	1	2	5
REB-OTHER Verbal Conflict	4	5	2	2
REB-OTHER Material Conflict	1	5	3	1

⁵Shellman [2000] and others have attempt to move beyond traditional practices by aggregating according to empirically discernible temporal clusters within the data.

These counts provide us with an objective, empirical measure that tracks weekly and monthly level fluctuations of the number and type of interactions between governments and rebel groups at the state-week level. In the follow section, we outline our research design used to analyze this data.

3.4. Dependent Variable.

Our dependent variable is a binary measure of intrastate conflict onset. To determine conflict onset, we use the ICEWS Ground Truth Dataset (GTDS), which provides a monthly, binary measure of whether or not each of the four types of domestic conflict below occur during each state-month⁶.

- | | |
|----------------|-----------------------|
| (1) Rebellion | (3) Domestic Crisis |
| (2) Insurgency | (4) Domestic Violence |

The unit of analysis in the raw GTDS dataset is state-month, which we disaggregate to the weekly level to achieve compatibility with the weekly count variables. If the raw data codes a conflict occurring in June 2008, all four weeks in June 2008 for that state receive a “1” indicating the occurrence of conflict. In order for a conflict to be coded as “onset” rather than “continuation”, it must be preceded by at least 3 months of peace. In total, our dataset contains 53 cases of onset.

4. METHODOLOGICAL APPROACH

Within the early warning literature, three primary methodological approaches exist: time series (Pevehouse and Goldstein [1999], Shellman [2004], Shellman [2000], Harff and Gurr [2001]), vector auto regression (VAR) (Goldstein [1992], Freeman [1989], and hidden Markov models (HMM) (Bond [2004], Shearer [2006], Schrodts [2000], and Schrodts [2006]). Despite the rigor and successful implementation in a number of research designs, these methodological approaches are not well suited to compare event sequences. In this section, we outline a test for the degree of similarity between event structures that do and do not precede an instance of conflict onset in order to determine the extent to which patterns exist in conflict escalations across time and space. To do so, we require a methodological technique that allows us to compare sequences of events occurring in different countries during different periods of time.

⁶DARPA does not make available the strict definitions of each of the four type of conflict or the complete coding procedures used to determine the state-months during which they occurred; for a general description see O'Brien [2010]

We chose to analyze sequences because, in qualitative analysis of political events, these appear to be one of the most common modes of analysis. In particular, various forms of qualitative “case-based reasoning”—see for example May [1973], Neustadt and May [1986], Khong [1992]—essentially match a sequence of events from past cases to the events observed in a current situation (with some substitutions for equivalent events), and then use the best historical fit to predict the likely outcome of the current situation.⁷ To the extent that an event is important, it is likely to generate more events, thereby leading to a spike in event counts. Thus, instead of analyzing the effects of specific events in a vacuum (like Harff [1998] and her focus on specific “triggers” and “accelerators”) the sequence analysis approach allows discrete events or event counts to determine the likelihood of future events. This general concept can be implemented in a variety of different ways—see for example the various “artificial intelligence” approaches in Hudson [1991], Schrodt [1990], Bakeman and Quera [1995], Hudson et al. [2008] and the HMM studies cited earlier.

In this study, we develop a new methodology that looks for clusters of sequences, measuring the distances of a sequence to other sequences in our data using Euclidean (in two forms) and Levenshtein distance measures. To the best of our knowledge, this is the first application of this methodology within the social sciences. We further narrow the focus of our analysis by looking only at countries where there is a reasonable risk of conflict. For example, Australia and New Zealand are two countries in the ICEWS dataset that are fairly unlikely to experience an onset of civil conflict. As such, our (or any) model predicting future peace for Australia and New Zealand would almost certainly be correct and increase its overall predictive accuracy, but this provides little insight into the patterns of conflict onset in countries that are likely to experience it. To avoid inflating our models accuracy by using such low hanging fruit, we only analyze data from the 15 states that experienced at least one month of conflict from 1998-2010. We focus on event counts over 12 week periods, meaning that our unit of analysis is a 12 week event count sequence for each state in our data. We have a total of 10,020 state-week observations and 3,984 possible 12-week sequences that are peaceful and precede another a month of continued peace. Additionally, 53 12-week sequences of peace exist that precede a conflict onset.

In order to empirically test for similarities between periods that precede peace and those that result in conflict, we utilize Euclidean and Levenshtein distance measures to assess for the degree of difference between paired 12-week periods. To establish pairs, we select all 12-week periods in

⁷See [Schrodt, 2004, chapter 6] for a much more extended discussion of this approach

our dataset that precede conflict and a random sample of periods that precede peace at a ratio of 2:1 peace preceding:conflict preceding. This results in a random sample of 113 12-week sequences that precede peace and the 53 12-week sequences that precede conflict onset. We then calculate the Euclidean and Levenshtein distances between the sequences of all potential pairs 13,695 pairs.

4.1. Euclidean and Levenshtein Distance Measures.

In order to provide an example of Euclidean and Levenshtein distance calculation, consider the two vectors below, which reflect one row and five columns of the 8X12 matrices used in the actual analysis, each column represents the count of GOV-REB material conflict for one week in a pair of 12-week sequences being compared to assess the level of similarity. In the actual process, we calculate the 12-week Euclidean distance for each of the 8 variables presented in Figure 1.

TABLE 2. Vectors of Material Conflict Counts for State “A” and State “B”

Count Variable	Weekly Counts					
GOV-REB (State “A”) Verbal Cooperation	8	2	1	2	5	7
GOV-REB (State “B”) Verbal Cooperation	2	3	0	1	2	5

Euclidean distance is a conceptually and mathematically straightforward approach used to calculate the “ordinary” distance between two points. At the most basic level, Euclidean distance between two points can be measured with a ruler if those points exist physically. This basic approach is commonly applied to numerical matrices according to the formula below, in which A and B are the counts in each of the two sequences, i is the column (i.e. week), and j is the specific row (i.e. count variable):

FIGURE 1. Euclidean Distance Formula

$$\sqrt{\sum_{i=1}^{12} (A_{ij} - B_{ij})^2}$$

Additionally, we calculate a second form of Euclidean distance, which we call Euclidean Absolute Value (EAV) by summing the absolute value of differences, as demonstrated below:

Likewise, this process results in 8 distances, which serve as our independent variables.

FIGURE 2. Euclidean Absolute Value (EAV) Distance Formula

$$\sum_{i=1}^{12} |C_{1ij} - C_{2ij}|$$

In addition to the two Euclidean distances, we also calculate the Levenshtein distance, which was created by the Russian mathematician Vladimir Levenshtein in 1965 as an algorithm to determine the number of "edits" needed to match two strings and it remains prominent in both applied and theoretical settings. The Levenshtein distance algorithm is commonly used in linguistics, including "spell check" applications of word processing software to provide the nearest match for misspelled words (see Hodge and Austin [2003] and van Noord and Gerdemann [2001]) as well as in studies of the similarities between different languages and dialects (see Freeman et al. [2006] and Gooskens and Heeringa [2004]). Additionally, it is frequently used to match and identify genomic DNA sequences (Troncoso-Pastoriza et al. [2007], Saha et al. [2008]). Mefford [1985] proposed this as a method of comparing international event sequences, and Schrodt [1991] applied it to the classification of war and non-war outcomes in the Behavioral Correlates of War data set. [Schrodt, 2004, Chapter 6] provides an extended discussion of the method and its relevance to political sequence analysis.

The major benefit of Levenshtein distances in the event data context is that it allows sequences to use shifts and substitutes to minimize the distance scores between two matrices. The Levenshtein algorithm assigns a cost of one for all:

- (1) Calculate the substitution/replacement cost at original alignment
- (2) Calculate the substitution/replacement cost after shifting sequence forward one unit, incurring an additional "shift cost" of 1
- (3) Calculate the substitution/replacement cost after shifting sequence backward one unit, incurring an additional "shift cost" of 1
- (4) Assign the lowest of the three costs as the final Levenshtein distance

Theoretically, Levenshtein distances this may be a better assessment of the actual degree of similarity between a sequence of event counts than the more rigid Euclidean measures. Consider the two five column vectors in Table 2. The third, fourth, and fifth column of vector 1 contain counts of 1, 2, and 5 respectively. The fourth, fifth, and sixth column of vector two also contain 1, 2, and 5 respectively. These three week counts are identical, but 1, 2, 5, sequence in the second

vector began one week too late to align with the 1, 2, 5 that occurred in the first vector. The Levenshtein function is able to recognize this similarity by calculating the EAV after shifting the second vector forward one column and backwards one column from its original position. In effect, the Levenshtein distance of the two sequences in Table 2 is the EAV distance after shifting the second sequence backwards one column.

Shifting the second vector backwards one column generates a total Levenshtein distance between the two vectors of 4, whereas the Euclidean distance and Euclidean EAV return distances of 7.21 and 14, respectively for the same two vectors. Overall, therefore, we calculate 24 total distance measures: the Euclidean, EAV, and Levenshtein distances for each of the 8 variables presented in Table 1.

4.2. Empirical Test. Euclidean, EAV, and Levenshtein distances are calculated for all possible sequence pairs within our sample dataset (which includes all 12-week periods that precede conflict and a random sample of 12-week periods that precede peace). These distances are calculated for each actor (GOV-REB and REB-OTHER) and action (verbal cooperation, verbal conflict, material cooperation, material conflict), providing 8 independent variables for each temporal aggregation (weekly and monthly). Out dependent variable is the “pair type”, which reflects whether neither one or both of the sequences preceded a conflict onset:

- Type “0”–Neither of the compared 12-week periods precede conflict onset
- Type “1”–One of the compared 12-week periods precede onset and the other did not
- Type “2”–Both of the 12-week periods precede conflict The histograms below reflect the distribution of distances for each of the

The histograms below reflect the distribution of distances for each of the three “pair types”. Overall, if event sequences of periods preceding conflict are similar across time and space, then Type “2” distances should be small. Conversely, large distances between Type “2” pairs suggests that event sequences of periods preceding conflict onset may not follow broader trends across time and space. The same logic applies to Type “0” pairs; these Type “0” distances will be small If periods proceeding to peace exhibit similar event structures across time and space, and will increase according to the variance between event structures. Lastly, distances among Type “1” pairs reflects the extent to which periods preceding peace and conflict onset vary. Small distances (relative to Type “0” and Type “2” distances) suggests that we may be unable to discern between sequences

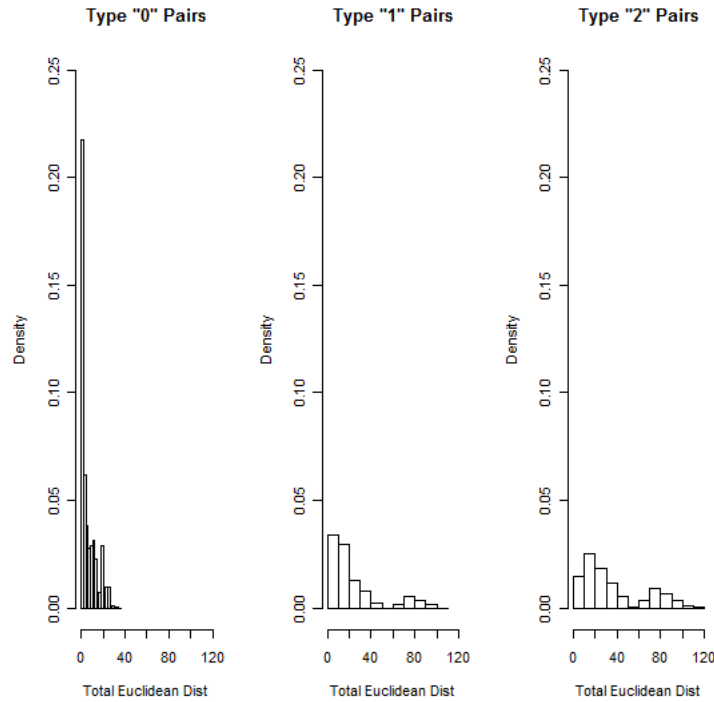


FIGURE 3. Histogram of the Distribution of Total Distances by Pair Type

preceding peace and conflict. The histograms below reflect the distribution of distances by pair type.

It is evident from the plots above that distances between sequences for Type “0” dyads tend to be small, whereas distances for both Type “1” and Type “2” dyads appear considerably larger.

To determine the extent to which a model can classify pair types based purely on the 8 distance measures for each pair, we utilize a bivariate logit model McCullagh and Nedler [1989]. To begin, we have two dichotomous variables: $YA \in \{0,1\}$ and $YB \in \{0,1\}$, corresponding to onset in State A and onset in State B. The resulting four possible combinations of outcomes are $YAB \in \{(0,0);(0,1);(1,0);(1,1)\}$. If we believe that conflicts escalate similarly, then the hypotheses that follow are that (0,0) outcomes and (1,1) outcomes will both exhibit small distances between sequences while (0,1) and (1,0) will exhibit larger distances. Using a bivariate logit model allows us to not only calculate the marginal probabilities of YA and YB (i.e. the probability of onset in State A and the probability of onset in State B separately), but it also estimates the joint probabilities of YAB (i.e. the probability of no conflict in both or the probability of conflict in both), thus providing us with an estimate of the relationship between the two marginal probabilities Imai et al. [2007]. Given the fact that our covariates are describing the distance of sequences between State

A and State B, we expect that the marginal probabilities of YA and YB are strongly correlated, making the bivariate logit a more natural and a better theoretical fit for our data. Furthermore, but summing the predicted probability of YA=1 and YB=1, we have the non-directed predicted probability that will allow us to assess our hypothesis.

We assess our model's performance using a simple decision rule for classifying cases and some straightforward performance measures. First, we estimate our model on training data consisting of a random sample of half our observations. These results are reported in the Appendix. Second, using the estimates derived from the training data, we calculate predicted probabilities for each pair type using the remaining data. Finally, our decision rule is to simply classify each case by the highest predicted probability. For example, if the predicted probabilities of a Type 0, Type 1, and Type 2 pair are 0.7, 0.2, and 0.1 respectively, our decision rule predicts Type 0 to be positive while Type 1 and Type 2 are negative. The performance measures used to assess our models predictions are sensitivity, specificity, positive predictive value and negative predictive value:

- Sensitivity—the percent of all positive observations correctly classified

$$\frac{\#ofTruePositives}{\#ofTruePositives + \#ofFalseNegatives}$$

- Specificity—the percent of all negative observations correctly classified

$$\frac{\#ofTrueNegatives}{\#ofFalsePositives + \#ofTrueNegatives}$$

- Positive Predicted Value—the percent of positive predictions that are accurate

$$\frac{\#ofTruePositives}{\#ofTruePositives + \#ofFalsePositives}$$

- Negative Predicted Value—the percent of negative predictions that are accurate

$$\frac{\#ofTrueNegatives}{\#ofFalseNegatives + \#TrueNegatives}$$

The results above provide a number of important findings. First, the multinomial logit model can effectively predict which distances were generated by Type “0” pairs. The sensitivity scores indicate that across all four model specifications, we correctly predict that a distance was generated by a Type “0” pair between 83% and 87% of all positive observations. For Type “0” pairs, the distances tend to be small, indicating that periods preceding peace are characterized primarily

TABLE 3. Forecast Performance of Multinomial Classification Logit

Performance Measure	Aggregation	Type “0” Pair	Type “1” Pair	Type “2” Pair
Sensitivity	Weekly	87.34	53.45	14.33
	Weekly Lagged	83.37	48.10	14.33
	Monthly	83.16	56.43	17.22
	Monthly Lagged	82.96	53.91	18.82
Specificity	Weekly	46.01	73.99	70.80
	Weekly Lagged	41.68	74.84	68.73
	Monthly	49.21	71.16	69.91
	Monthly Lagged	47.42	71.27	68.56
Pos. Pred. Value	Weekly	69.48	63.89	26.15
	Weekly Lagged	66.86	63.02	25.12
	Monthly	70.79	63.04	25.5
	Monthly Lagged	71.43	61.77	22.48
Neg. Pred. Value	Weekly	86.23	67.84	90.67
	Weekly Lagged	86.40	65.88	90.65
	Monthly	83.67	67.70	91.03
	Monthly Lagged	83.68	66.40	91.04
# of each Pair Type	Weekly	4001	2537	390
	Weekly Lagged	4207	2315	406
	Monthly	3558	2665	458
	Monthly Lagged	3518	2598	565
N-Monthly=13,203		N-Weekly=13,695		

by the absence of events involving domestic rebel groups. Given this, the model classifies most small distances as having been generated by Type “0” pairs. The histograms of the total distances illustrates that although the distributions of Type “1” and Type “2” distances are relatively larger with far fewer total distances equaling 0, a considerable portion of distances are still quite small. This suggests one of two things: first, it is possible that some conflict onsets are not preceded by high levels of rebel activity; second, rebel activities may be occurring but not reported in the news. Given this, one might expect that the high sensitivity score is a product over greatly over predicting Type “0” pairs. However, the large positive predicted values of approximately 70% for Type “0” pairs across all four model specifications of approximately indicates that 7 out of every 10 predictions of a Type “0” pair are accurate.

Additionally, our model struggles to classify Type “2” pairs. The sensitivity scores indicates that our model only predicts between 14% and 18% of the 697 Type “2” pairs. To borrow a medical analogy, consider a random sample of patients whose vital signs are monitored weekly. Variation between patients vital signs (i.e. Type “0” pairs) is likely to be low. However, the distance between the vital signs of a patient who continues to maintain health and a patient who is falling ill is likely

large. In the medical world, a multitude of illnesses exist, each of which contain different symptoms. Additionally, the distance between vital sign sequences of two patients approaching an illness (i.e. a Type “2” pair) may likely be high, as different illnesses may alter vital signs in very different ways. Similarly, the large distances and difficulty for our model to distinguish between Type “1” and Type “2” pairs may reflect that domestic tensions escalate to conflict in very different ways.

The goal motivating the preceding methodological approach was to determine the extent to which generalizable trends may be drawn about event sequences preceding conflict onset. Given the large distances of Type “2” pairs and the inability of the multinomial logit model to correctly discern between Type “1” and Type “2” pairs, we can say with confidence that escalatory sequences do not follow consistent patterns across time and space. Despite this, all is not lost regarding our ability to build a forecasting model that is accurate across time and space.

4.3. Using the Baseline Peace Archetype (BPA) to Create a Forecasting Model. The multinomial logit classification matrix and the histogram of the distribution of Type “0” distances both indicate that sequences that precede peace contain highly similar event structures. We introduce a concept of a “Baseline Peace Archetype” (BPA) to reflect the average event event count structure of 12-week sequences that precede peace. To calculate the BPA, we randomly sample 120 12-week sequences that precede peace and calculate the mathematical mean of the event counts for each of the 8 event categories at the weekly level⁸. Substantively, this reflects the average structure of events involving rebel groups during periods that precede peace in 11 countries from 1997 to 2010. Based largely on Figure 3, we expect that the likelihood that a 12-week period precedes a conflict onset should increase as its distance from the BPA sequence increases.

To test this, we calculate the Euclidean, EAV, and Levenshtein distances between all 53 periods that precede conflict onset and a random sample of 128 periods that precede peace against the BPA, resulting in 168 “pairs.” Because the BPA reflects the average sequence of events that precedes peace, calculating distances results in only Type “0” and Type “1” pairs (a Type “2” pair becomes impossible to achieve). We utilize a logit model to generate predicted probabilities that reflect the likelihood that the distance measures were generated by a Type “1” pair. In this framework, predicting a Type “1” pair is synonymous with forecasting a conflict onset. As with the multinomial logit, we must first implement a decision rule that determines in which class each out-of-sample

⁸120 sequences results from randomly sampling from all possible sequences that precede peace at a ratio of 2:1 against sequences that precede conflict onset

observation is to be placed. For assessing the performance of logit or other binary choice models, the decision rule may simply be a cutoff between 0 and 1 where all predicted probabilities below the cutoff are classified as 0 and all above the cutoff are classified as 1. Although there is no generally accepted method for determining such a cutoff, any decision rule should be in-line with the theoretical underpinnings of the model.

Our decision rule can be thought of as a trade-off in sensitivity and PPV. In our setting, perfect sensitivity can be achieved by predicting every case to be a case of conflict onset. However, if a government based its foreign policy decisions on just one model's predictions, and that model optimized sensitivity at the expense of all else, that government will find itself constantly preparing for civil conflicts that only materialize a small fraction of the time. In general, if a false positive is costly, then optimizing sensitivity at the expense of all else is not desirable; rather, we must also incorporate PPV into our decision rule. PPV can be interpreted as the percent of countries that experience onset when an onset is predicted. One way to achieve a balance in predictions is to use the cutoff associated with the highest f-score, which can be thought of as the weighted mean of PPV and sensitivity (Rijsbergen). Other measures of performance are then calculated using the cutoff as our decision rule.

Despite the size and scope of the ICEWS dataset, it only contains 53 conflict "onsets" according to our operationalization (i.e. a positive observation of conflict preceded by at least 12 weeks/3 months of peace). The scarcity of conflict onsets is especially problematic for out-of-sample forecasts, which requires dividing the data into a minimum of two separate sets. To best increase the robustness of our out-of-sample forecasts given our small number of positive observations, we utilize the following bootstrapping technique. First, we randomly select 50 percent of distances to train our model then apply the stored coefficient estimates to the remaining 50 percent to generate out-of-sample predicted probabilities. Second, we calculate the following three measures of forecast accuracy based on the out-of-sample predicted probabilities:

- Sensitivity—the percent of onsets that our model predicts that are actually onsets
- Specificity—the percent of non-onsets that our models predicts that are actually non-onsets
- Area under the curve (AUC)—the area under the line formed by a receiver operator curve (ROC), which plots the relationship between true positives and false negatives across all possible cutoff values

Third, we utilize an algorithm to repeat this process 1,000 times based on a different random sampling to determine the 50% of observations that comprise the in-sample and out-of-sample populations. This results in 1,000 different sensitivity, specificity, and AUC scores, which allows us to determine the mean and confidence intervals of these measure. The graphs below reflect the mean and 95% confidence intervals (indicating that 95 % of observations occur in this range) of sensitivity, specificity and AUC scores for one and two month forecasts of all three distance measures for both weekly and monthly temporal count aggregations. In total, this approach generates the 12 different classes of descriptive statistics below, each of which provide data for all three distance measures.

- | | |
|-------------------------------|---------------------------------|
| (1) Weekly Sensitivity | (7) Monthly Sensitivity |
| (2) Lagged Weekly Sensitivity | (8) Lagged Monthly Sensitivity |
| (3) Weekly Specificity | (9) Monthly Specificity |
| (4) Lagged Weekly Specificity | (10) Lagged Monthly Specificity |
| (5) Weekly AUC | (11) Monthly AUC |
| (6) Lagged Weekly AUC | (12) Lagged Monthly AUC |

The plots above illustrate a number of important results. First, and most importantly, they demonstrate that our model is highly accurate. The mean sensitivity across all 12 model specifications is over 0.6, meaning that over 53% of the onsets that our model predicts are actually conflict onsets. Additionally, the specificity of all 12 models is over 0.8, indicating that over 80% of all predictions of non-conflict are accurate. Lastly, in a binary model, 50% accuracy corresponds to a 0.5 AUC value. The mean AUC for all 12 models exceeds 0.7, which reflects strong forecasting accuracy. These figures are especially impressive given that we restrict our sample to only “difficult” countries that experienced at least one case of conflict onset within our temporal domain from 1997 to 2010. Even among these countries with the greatest “risk profiles”, our model is able to accurately determine which event sequences are likely to precede a conflict onset and which are likely to result in peace.

Second, the Levenshtein distance measure outperforms both the Euclidean and EAV distances in 9 of the 12 different classes of descriptive statistics. This provides empirical support for our theoretical expectation that the ability of the Levenshtein formula to allow flexibility of shifts and substitutions between sequences better reflected the actual level of similarity between event

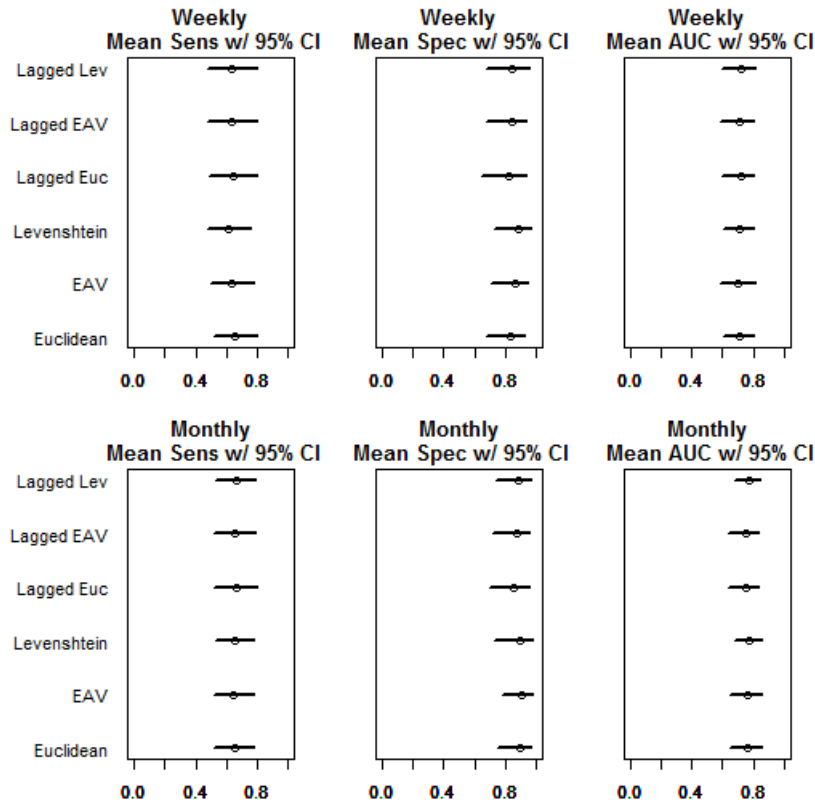


FIGURE 4. Forecast Accuracy Across Distance Measures and Temporal Aggregations

structures. However, as the plots indicate, the means and 95% confidence intervals tend to be similar across all tests indicating that the Levenshtein distance only mildly outperforms the Euclidean and EAV measures on average.

Third, these plots demonstrate that our results are consistent across provide both weekly and monthly level temporal aggregations. This consistency increases our confidence in the accuracy of our model, especially in light of the findings in Shellman [2004] that many inferences drawn from event data tend to vary based on temporal aggregation. Interestingly, the accuracy of our one and two-month (reflect by the “lagged” title) appear indistinguishable. However, the confidence intervals for the lagged forecasts are slightly larger on average, which is expected given basic forecasting theory: variance of forecast accuracy tends to increases as function of how far into the future the forecast predicts.

5. CONCLUSION

The central goal of this paper is to provide accurate, monthly predictions of the onset of intrastate conflict within a diverse set of states for which conflict is a real possibility. Achieving this aim required a more comprehensive event data dataset than had been previously used in extant literature as well as an innovative methodological approach. To overcome shortcomings in currently used event data datasets, we introduced the ICEWS dataset, which provides daily level event data in a who-did what-to whom, when, and where format for 29 Asian countries from 1997-2010. Prior to building a comprehensive forecasting model able to provide accurate predictions for a large number of heterogeneous state, we first had to determine the extent to which periods preceding a conflict onset exhibit similar characteristics across time and space; if generalizable trends exist across states, then it is likely that one model could be fruitfully applied to forecast conflict onsets in a broad range of states. We accomplish this in four steps: first, we convert the ICEWS dataset to reflect the number and type of politically relevant domestic events aggregated at the state-week level; second, we build sequences from all possible consecutive 12 state-week periods, 53 of which precede conflict onset and a sample of the 3,984 total sequences that precede continued peace; third, we calculate the Euclidean and Levenshtein distance between the 53 sequences that precede conflict and a random sample of those that precede peace, which generates 13,695 sequence pairs; fourth, we utilize a bivariate logit model to determine the extent to which an empirical model can determine the pair “type” (i.e. whether one, both, or neither sequence preceded conflict based on the distances between the pairs).

We find distances between pairs that both precede peace are relatively small, but distances between pairs that precede a conflict onset were large, suggesting that intrastate conflicts do not follow generalizable patterns across time and space. This finding alone is an important contribution as it is the first rigorous test of its kind. Although periods preceding different conflicts follow diverse paths, they are similar in that they all differ considerably from periods that precede peace. In fact, we found that periods preceding peace exhibit highly similar event count structure. Therefore, this finding allowed us to achieve the central goal of this paper and build an accurate forecasting model in four main steps: first, we calculate a Baseline Peace Archetype (BPA), reflecting the “average” peace-preceding sequence; second, we calculate Euclidean and Levenshtein distances between the 53 sequence preceding conflict and a new sample of 120 sequences preceding peace against the BPA sequence; third, we build out-of-sample forecasts with bootstrapping to determine the accuracy

with which we can predict conflict onsets one and two months in based on the distance between the 173 sequences in our sample and the BPA sequence. This approach generates high accurate forecasts, consistently achieving over 80% cumulative accuracy. We are highly confident in our results given their consistency across a range of difficult robustness checks, including rigorous out-of-sample methods and varied temporal aggregations.

This study introduced a new dataset and an innovative methodological approach. As such, a number of potential shortcomings and extensions exist. For example, we were able to demonstrate that periods preceding conflict exhibit highly different event structures and although analyzing the causes of this variation and testing for potential clustering of onset types fell outside the scope of this paper, it would serve as a valuable enterprise in the future. Moreover, our model performs best when politically relevant events precede a conflict onset. However, events may not precipitate certain cases of intrastate conflict onsets, which may be sparked by a sudden event. Overall, we hope that this study emphasizes the importance of nuanced data that reflects the constant flux of domestic conditions and encourages future research in the area.

REFERENCES

- James Alt, Gary King, and Curt S. Signorino. Aggregation among binary, count, and duration models: Estimating the same quantities from different levels of data. *Political Analysis*, 9:21–44, 2001.
- Edward E. Azar. The conflict and peace data bank (COPDAB) project. *Journal of Conflict Resolution*, 24:143–152, 1980.
- R. Bakeman and V. Quera. *Analyzing Interaction: Sequential Analysis with SDIS and GSEQ*. Cambridge University Press, New York, 1995.
- Doug Bond, J. Craig Jenkins, Charles L. Taylor Taylor, and Kurt Schock. Mapping mass political conflict and civil society: Issues and prospects for the automated development of event data. *Journal of Conflict Resolution*, 41(4):553–579, 1997.
- Doug Bond, Joe Bond, Churl Oh, J. Craig Jenkins, and Charles L. Taylor. Integrated data for events analysis (IDEA): An event typology for automated events data development. *Journal of Peace Research*, 40(6):733–745, 2003.
- 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, Montréal, 2004.
- Patrick T. Brandt and John R. Freeman. Advances in bayesian time series modeling and the study of politics: Theory testing, forecasting, and policy analysis. *Political Analysis*, 14:1–36, 2005.
- Thomas G.. Dale. Event data analysis and threats from temporal aggregation. Presented at the Florida Political Science Association Meeting, Sarasota, 2002.
- Robert D. Duval and William R. Thompson. Reconsidering the aggregate relationship between size, economic development, and some types of foreign policy behavior. *American Journal of Political Science*, 24(3):511–525, 1980.
- Andrew T. Freeman, Sherri L. Condon, and Christopher M. Ackerman. Cross linguistic name matching in english and arabic: a "one to many mapping" extension of the levenshtein edit distance algorithm. In *Proceedings of the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics*, pages 471–478. Association for Computational Linguistics, 2006.
- John R. Freeman. Systematic sampling, temporal aggregation, and the study of political relationships. *Political Analysis*, 1:61–98, 1989.

- Deborah J. Gerner, Philip A. Schrodtt, Ronald A. Francisco, and Judith L. Weddle. The machine coding of events from regional and international sources. *International Studies Quarterly*, 38: 91–119, 1994.
- Joshua S. Goldstein. A conflict-cooperation scale for WEIS events data. *Journal of Conflict Resolution*, 36:369–385, 1992.
- Jack A. Goldstone, Robert Bates, David L. Epstein, Ted Robert Gurr, Michael 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.
- Charlotte Gooskens and Wilbert Heeringa. Perceptive evaluation of levenshtein dialect distance measurement using norwegian dialect data. *Language Variation and Change*, 16:189–207, 2004.
- August Hämmerli, Regula Gattiker, and Reto Weyermann. Conflict and cooperation in an actor’s network of chechnya based on event data. *Journal of Conflict Resolution*, 50(159):159–175, 2006.
- Barbara Harff. Early warning of humanitarian crises: Sequential models and the role of accelerators. In John L. Davies and Ted Robert. Gurr, editors, *Preventive Measures: Building Risk Assessment and Crisis Early Warning*, pages 79–94. Rowman and Littlefield, Lanham, MD, 1998.
- Barbara Harff and Ted Robert Gurr. Systematic early warning of humanitarian emergencies. *Journal of Peace Research*, 35(5):359–371, 2001.
- Carl G. Hempel. Explanation and prediction by covering laws. In Carl G. Hempel and James H. Fetzer, editors, *The philosophy of Carl G. Hempel : studies in science, explanation, and rationality*, chapter 5. Oxford University Press, Oxford, 2001.
- Victoria J. Hodge and Jim Austin. A comparison of standard spell checking algorithms and a novel binary neural approach. *IEEE Transaction on Knowledge and Data Engineering*, 15(5): 1073–1081, 2003.
- Valerie Hudson, editor. *Artificial Intelligence and International Politics*. Westview, Boulder, 1991.
- Valerie M. Hudson, Philip A. Schrodtt, and Ray D. Whitmer. Discrete sequence rule models as a social science methodology: An exploratory analysis of foreign policy rule enactment within palestinian-israeli event data. *Foreign Policy Analysis*, 4(2):105–126, 2008.
- Kosuke Imai, Gary King, and Olivia Lau. *bilogit: Bivariate Logistic Regression for Dichotomous Dependent Variables*, 2007.
- Craig J. Jenkins and Doug Bond. Conflict carrying capacity, political crisis, and reconstruction. *Journal of Conflict Resolution*, 45(1):3–31, 2001.

- Yuen Foong Khong. *Analogies at War*. Princeton University Press, Princeton, 1992.
- 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.
- Russell J Leng. *Behavioral Correlates of War, 1816-1975. (ICPSR 8606)*. Inter-University Consortium for Political and Social Research, Ann Arbor, 1987.
- E. R. May. *“Lessons” of the Past: The Use and Misuse of History in American Foreign Policy*. Oxford University Press, New York, 1973.
- Charles A. McClelland. World-event-interaction-survey: A research project on the theory and measurement of international interaction and transaction. University of Southern California, March 1967.
- Peter McCullagh and John A. Nedler. *Generalized Linear Models, Second Edition*. Chapman and Hall, New York, 1989.
- Patrick McGowan, Harvey Starr, Gretchen Hower, Richard L. Merritt, and Dina A. Zinnes. International data as a national resource. *International Interactions*, 14:101–113, 1988.
- Dwain Mefford. Formulating foreign policy on the basis of historical programming. In Urs Lutembacher and Michael Don Ward, editors, *Dynamic Models of International Conflict*. Lynne Rienner Publishing, Boulder, 1985.
- Richard L. Merritt, Robert G. Muncaster, and Dina A. Zinnes, editors. *International Event Data Developments: DDIR Phase II*. University of Michigan Press, Ann Arbor, 1993.
- R. E. Neustadt and E. R. May. *Thinking in Time: The Uses of History for Decision Makers*. Free Press, New York, 1986.
- Sean O’Brien. Crisis early warning and decision support: Contemporary approaches and thoughts on future research. *International Studies Review*, 12(1):87–104, 2010.
- 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.
- Willard Van Orman Quine. Two dogmas of empiricism. *The Philosophical Review*, 60:20–43, 1951.
- Surya Saha, Susan Bridges, Zenaida Magbanua, and Daniel Peterson. Computational approaches and tools used in identification of dispersed repetitive dna sequences. *Tropical Plant Biology*, 1: 85–96, 2008.

- Philip A. Schrodt. Parallel event sequences in international relations. *Political Behavior*, 12(2): 97–123, 1990.
- Philip A. Schrodt. Pattern recognition of international event sequences: A machine learning approach. In Valerie Hudson, editor, *Artificial Intelligence and International Politics*. Westview, Boulder, 1991.
- 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. *Patterns, Rules and Learning: Computational Models of International Behavior*. <http://eventdata.psu.edu/papers.dir/Schrodt.PRL.2.0.pdf>, 2nd edition, 2004.
- 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. Automated production of high-volume, near-real-time political event data. Presented at the American Political Science Association meetings, Washington, 2010.
- Philip A. Schrodt and Deborah J. Gerner. Validity assessment of a machine-coded event data set for the Middle East, 1982-1992. *American Journal of Political Science*, 38:825–854, 1994.
- 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. An event data analysis of third-party mediation. *Journal of Conflict Resolution*, 48(3):310–330, 2004.
- Philip A. Schrodt, Deborah J. Gerner, Rajaa Abu-Jabr, Ömür Yılmaz, and Erin M. Simpson. Analyzing the dynamics of international mediation processes in the middle east and balkans. Presented at the American Political Science Association meetings, San Francisco, 2001.
- Philip A. Schrodt, Deborah J. Gerner, and Ömür Yılmaz. Conflict and mediation event observations (CAMEO): An event data framework for a post Cold War world. In Jacob Bercovitch and Scott Gartner, editors, *International Conflict Mediation: New Approaches and Findings*. Routledge, New York, 2009.
- 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. Process matters: Conflict and cooperation in sequential government-dissident interactions. *Journal of Conflict Resolution*, 15(4):563–599, 2000.

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

Stephen Shellman and Brandon Stewart. Predicting risk factors associated with forced migration: An early warning model of haitian flight. *Civil Wars*, 9(2):174–199, 2007.

Juan Ramón Troncoso-Pastoriza, Stefan Katzenbeisser, and Mehmet Celik. Privacy preserving error resilient dna searching through oblivious automata. In *Proceedings of the 14th ACM conference on Computer and communications security*, pages 519–528, New York, NY, 2007. ACM.

Gertjan van Noord and Dale Gerdemann. An extendible regular expression compiler for finite-state approaches in natural language processing. In Oliver Boldt and Helmut Jorgensen, editors, *Automata Implementation*, volume 2214 of *Lecture Notes in Computer Science*, pages 122–139. Springer Berlin / Heidelberg, 2001. URL http://dx.doi.org/10.1007/3-540-45526-4_12.

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.

ADDRESS FOR ALL AUTHORS:, DEPARTMENT OF POLITICAL SCIENCE, PENNSYLVANIA STATE UNIVERSITY, UNIVERSITY PARK, PA 16802

E-mail address: `vjd125@psu.edu`

E-mail address: `jxy190@psu.edu`

E-mail address: `schrodt@psu.edu`

URL: <http://eventdata.psu.edu>