Data-based Computational Approaches to Forecasting Political Violence

Philip A. Schrodt, James Yonamine and Benjamin Bagozzi

Abstract [which will be shortened]This chapter provides a general overview of inductive statistical and computational methodologies used in the analysis and forecasting of political violence, and some of the challenges specific to the issue of analyzing terrorism. The chapter is intended for the non-specialist in the field of technical political forecasting, but assumes a general familiarity with data and computational methods. Our purpose is not to exhaustively explore any of these methods—each technique would typically require tens or hundreds of pages—but instead to focus on the similarities and differences between the approaches in terms of their assumptions about what aspects of the problem can be extracted from data, and what types of predictions can be made. We first provide a general overview of some of the types of data commonly used in technical forecasting models, then consider the two broad categories of model: statistical and algorithmic. [next sentences will be updated when manuscript is complete]Within statistical modeling, we assess the strengths and weaknesses of conventional time series approaches, event-history models, vector-autoregression, and zero-inflated models. Within computational modeling, we consider the general issue of computational pattern recognition and data mining, and then look more specifically at the issue of using event sequences for forecasting.

Philip A. Schrodt
Political Science, Pennsylvania State University, University Park, PA 16801, USA. e-mail: schrodt@psu.edu

James Yonamine
Political Science, Pennsylvania State University, University Park, PA 16801, USA. e-mail: jxy190@psu.edu

Benjamin Bagozzi
Political Science, Pennsylvania State University, University Park, PA 16801, USA. e-mail: beb196@psu.edu

1 Introduction and Overview

The challenge of terrorism dates back centuries if not millennia. Until recently, the basic approaches to analyzing terrorism—a combination of historical analogy and monitoring the contemporary words and deeds of potential perpetrators—have changed little: the Roman authorities warily observing the Zealots in 1st-century Jerusalem could have easily changed places with the Roman authorities combatting the Red Brigades in 20th century Italy.

This stasis has changed with the exponential expansion of information processing capability made possible first by the development of the digital computer, followed by the phenomenal growth in the quantity and availability of machine-readable information made possible by the World Wide Web. Information that once circulated furtively on hand-copied sheets of paper (or papyrus) is now instantly available—for good or ill—on web pages which can be accessed for essentially no cost from anywhere in the world. This expansion of the scope and availability of information in all likelihood will change the dynamics of the contest between organizations seeking to engage in terrorism and those seeking to prevent it. It is almost certainly too early to tell which group will benefit more—many of the new media are less than a decade old—but the techniques of processing and evaluating information will most certainly change.

This chapter provides a general overview of inductive statistical and computational methodologies used in the analysis and forecasting of political violence, and some of the challenges specific to the issue of analyzing terrorism. The chapter is intended for the non-specialist, but assumes a general familiarity with data and computational methods. Our purpose is not to exhaustively explore any of these methods—each technique would typically require tens or hundreds of pages—but instead to provide a sufficient introduction to the basic concepts and vocabulary that the reader can explore further on his or her own. This is a map, not a mine. Throughout our discussion, we focus on the similarities and differences between the approaches—a number of which are illustrated or discussed in far greater detail in later chapters in this volume—in terms of their assumptions about what aspects of the problem can be extracted from data, and what types of predictions can be made.

The psychologist and philosopher William James, in his Lowell Institute lectures in 1906, subsequently published under the title *Pragmatism: A New Name for Some Old Ways of Thinking* notes that the fundamental split in philosophy, dating to the very origins of the field, is between "rationalists" who seek to find an intellectual structure that will reveal a proper order in the world, and "empiricists," who take the disorder of the observed world as a given and simply try to make as much sense of it as they can. More than a century later, we find exactly the same split in formal approaches in the social sciences: the rationalist position found in deductive approaches such as game theory, expected utility models, systems dynamics and agent-based models, which seek to explain behavior from a set of *a priori* first-principles and their consequent emergent properties, and the empiricist approach found in inductive statistical and computational data-mining approaches which extracts structured information for large sets of observed data. Both approaches are

well-represented in this volume [we will insert citations when the table of contents has been finalized] but this review will focus only on the empirical approaches.

Throughout the chapter, we will generally be looking at models which focus on forecasting and understanding political violence in general, not just approaches to terrorism per se. This is done for two reasons. First, due to a combination of data limitations and a paucity of interest in the research community prior to the 2000s, the number of studies was quite limited and focused on a relatively small number of approaches. In contrast, the large-scale research efforts were generally focused on various forms of political violence, not just terrorism. Second, most of the methods used to study the general problem of political violence are at least potentially applicable to the study of terrorism—in fact many are generally applicable to almost any behavior for which large amounts of data are available—albeit we will frequently caveat these possibilities with concerns about some of the atypical aspects of terrorism, and we will be focusing on methods appropriate to rare-events analysis rather than high-frequency behaviors. Finally, in many instances, there is a close relationship between situations of general political violence such as civil war and state failure, and conditions which encourage the formation and maintenance of terrorist groups, so political instability is of considerable interest on its own.

We will first provide a general overview of some of the types of data commonly used in technical forecasting models, then consider the two broad categories of model: statistical and algorithmic. [next sentences will be updated when manuscript is complete]Within statistical modeling, we assess the strengths and weaknesses of conventional time series approaches, event-history models, vector-autoregression, and zero-inflated models. Within computational modeling, we consider the general issue of computational pattern recognition and data mining, and then look more specifically at the issue of using event sequences for forecasting.

1.1 The Development of Technical Political Forecasting

There is both a push and a pull to the current interest in technical political forecasting. As with many aspects of the Information Revolution of the past several decades, some of the interest in driven purely by technology and the fact that have the data, computational power and readily accessible software to do things that we couldn't do before. The impact of intellectual curiosity and a technical itch to scratch is a factor.

But if the technological factors were the only driver, we could satisfying that by clicking on virtual cows and broadcasting our prowess on Facebook (http://www.wired.com/magazine/2011/12/ff_cowclic However, at the level of serious policy analysis—mess up counter-terrorism, and real people die, with no reset—the motivation is different: humans are appallingly bad at predicting human behavior. The pathbreaking studies of Philip Tetlock [59], who assessed the performance of 284 expert political forecasters in the assessments of more than 80,000 predictions over a twenty year period, found that their accuracy was barely better than chance, and for some of the most publicly-visible forecast-

ers, who play to a crowd anxious for dramatic predictions, actually worse than what could be achieved by a dart-throwing chimpanzee. Tetlock's results confirmed—albeit in much great detail—for the political realm similar findings in economics and psychological assessment [41, 34] but were nonetheless a stunning indictment of the "subject matter expert" as prognosticator.

Tetlock's work was particularly timely given the climate for the development of statistical and computational models of political violence. The earliest systematic studies of political violence—largely interstate war—using modern statistical methods and standards for data collection date from the middle of the twentieth century, in the work of British meteorologist Lewis Richardson [49]. This approach gained considerable momentum in the 1960s with the "behavioral revolution" in political science, which used the emerging technologies of social science statistics, digital computing, and machine-readable data to begin the systematic assessment of various theoretical propositions about political violence (and, for the most part, finding the proposed patterns were far less generalizable than their advocates believes, the results on the "democratic peace" [someone have a handy citation or two here?]being the notable exception. This period saw the initial development of many of the large-scale databases relevant to the study of political violence, including the Correlates of War http://www.correlatesofwar.org/ on international and civil war, Polity (http://www.systemicpeace.org/polity/polity4.htm) on state characteristics, and early event data sets [10, 2, 40, 36] on political interactions in general. These data sets collected specifically for the study of political behavior were supplemented by readily-available global data on structural characteristics of states provided by organizations such as the United Nations, International Monetary Fund and World Bank.

The U.S. Department of Defense Advanced Research Projects Agency (ARPA) funded a series of projects for the development of statistical forecasting in the 1970s [1, 13, 30], and parallel efforts, using both statistical and computational ("artificial intelligence") methods continued in the political science community under National Science Foundation funding [12, 31, 58, 24]. However, these early efforts were not particularly successful in the production of robust, timely and accurate forecasts—in retrospect, neither the data nor methods available at the time were sufficient for the task. By the 1980s U.S. government efforts had largely ended, though the basic research in the academic community continued, as did some efforts by international non-governmental organizations [26, 14, 23, 50].

The U.S. government re-engaged with the development of technical political forecasting in the mid-1990s, motivated in part by a recognition of the considerable progress in the development of methods and data since the earlier efforts, and by dramatic intelligence failures by human analysts on such issues as the end of the Cold War [19, 20], state failure in Somalia and the former Yugoslavia, and the Rwandan genocide. The most conspicuous effort was the U.S. multi-agency State Failures Project [17, 16], later renamed the Political Instability Task Force (PITF) [22], a long-term collaboration between government analysts, contractors, and academic researchers. After some initial missteps involving models which were overly complex, PITF developed models that predicted various forms of political instabil-

ity with 70% to 80% accuracy with a two-year time horizon and global coverage. In the mid-2000s, DARPA returned to the issue with the Integrated Conflict Early Warning System (ICEWS) [43] [also cite O'Brien's article in this volume], which achieved similar levels of accuracy.

As a consequence of these various efforts, technical forecasting of political violence is now quite well developed and the subjects of numerous efforts in both North America and Europe [9, 5, 60, 61, 8, 27, 28, 44, 29]. While most of these are looking at violence and irregular political activity such as coups, a quantitative literature has also emerged on the topic of terrorism [15, 33, 38, 48, 46, 47] [need to add some Sandler here...do you folks have access to a site where you can get these in BibTeX form? Or is there a good review article somewhere?]

[Need another sentence or two to finish this off, but probably it is about done]

2 Data Sources

In this discussion, we use the following terminology to distinguish between the types of information being coded. An "event" is a discrete incident that can be located at a single time (usually precise to a day) and set of actors, usually a dyad of a source and target. This is distinct from "structural data" such as GDP or Polity scores (`http://www.systemicpeace.org/polity/polity4.htm`). "Episodic" data are those which code the characteristics of an extended set of events such as a war or a crisis: the Correlates of War project (COW; `http://www.correlatesofwar.org/`) is the archetype; International Crisis Behavior (*http://www.cidcm.umd.edu/icb/*) would be a more recent example. "Composite" events are those which occur in a relatively short period of time and limited space—for example a terrorist attack—and multiple characteristics of the incident are coded. Finally, "atomic" events are basic units of political interaction—date, source, target, event—found in classic event data sets such as WEIS and COPDAB, and in contemporary coding schemes such as IDEA, CAMEO and SPEED. As with any typology, not all of the data sets fit clearly into a single category, but most will.

2.1 Structural Data

2.2 Event Data

While the projects coding composite events still use human coding, for real-time coding of atomic events, there is simply no alternative to automated methods. Sustained human coding projects, once one takes in the issues of training, retraining, replacement, cross-coding, re-coding due to effects of coding drift and/or slacker-coders and so forth, usually ends up coding about six events per hour. Individual

coders, particularly working for short periods of time, and definitely if they are a principal investigator trying to assess coding speed, can reliably code much faster than this. But for the *overall* labor requirements—that is, the total time invested in the enterprise divided by the resulting useable events—the 6 events per hour is a pretty good rule of thumb and—like the labor requirements of a string quartet—has changed little over time.

The challenge will be extending the success of automated coding to projects which are coding composite event. This may be possible by further development of specific "data field extraction" methods, for example locating reports of the number of individuals killed or the amount of aid promised. A very sizeable NLP literature exists on this [CiteTBA] and several such methods are used in SPEED. One would then define composite events such as "civil war" by using patterns of the atomic events. This would also make the differences between definitions used by various project unambiguous (or at least comparable) and allow the composite events to be easily constructed out of existing data sets rather than starting every new project from the beginning. MID, in moving from the original episodic definitions to coding composite incidents as well, would be an example of this approach, albeit with human coding.

2.3 Text, Social Media and Other Unstructured Data Sources

2.4 The Challenges of Data Aggregation

3 Statistical Approaches

Much of the work with event data has focused on forecasting political conflict. Within the early warning literature, three primary methodological approaches exist: time series [45, 57, 56, 26], vector auto regression (VAR) [21, 18], and hidden Markov models (HMM) [7, 55, 52, 54]. This paper—like that of the HMM work—will look at event data as patterns since patterns are one of the most common modes of political analysis found in qualitative studies. In particular, various forms of qualitative "case-based reasoning"—see for example [39, 42, 35]—essentially match patterns 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.[1] Instead of analyzing the effects of specific events in a vacuum (like [25] and her focus on specific "triggers" and "accelerators") a pattern-recognition 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 [31, 51, 4, 32] and the HMM studies cited earlier.

---

[1] See [53, chapter 6] for a much more extended discussion of this approach

3.1 Cross-sectional Regression and Logit

Since this is so widely used in PITF and much of the other published literature
  Ordinary least squares regression uses equations of the form

$$y_i = \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i = x_i' \beta + \varepsilon_i, \qquad i = 1, \ldots, n, \tag{1}$$

  Logistic regression, in contrast, is used to predict a values between zero and one—typically interpreted as a probability—and does this by using an equation of the form

$$f(z) = \frac{e^z}{e^z + 1} = \frac{1}{1 + e^{-z}} \tag{2}$$

  where the variable "z" is usually defined as

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_k x_k, \tag{3}$$

3.2 Classical Time Series

Box-Jenkins, etc. Fairly short but introduce the issue of autoregression vs autocorrelated error

3.3 Event History and Hazard Models

3.4 Vector Autoregression Models

This is probably worth a mention due to Sandler, Brandt. We can also mention the Bayesian variant

3.5 Zero-Inflated Models

and more generally discuss some of the rare events models. Hmmm, should we also do Heckman models here?

3.6 Geo-spatial Models

A brief introduction....

3.7 What else am I missing?

4 Algorithmic Approaches

4.1 Supervised Cross-sectional Classification Methods

Neural networks, SVM, linear discriminant analysis, the usual suspects

4.2 Unsupervised Clustering Methods

Again, the usual suspects, and also note that some of these overlap with the statistical approaches. Put in a word for some of the topic-clustering methods Burt is using?

Latent Dirichlet allocation (LDA) models were introduced by [6] and briefly described in the abstract of that article as:

> LDA is a three-level hierarchical Bayesian model, in which each item of a collection is modeled as a finite mixture over an underlying set of topics. Each topic is, in turn, modeled as an infinite mixture over an underlying set of topic probabilities.

In the typical LDA application to document classification, each document is assumed to be a mixture of multiple, overlapping *latent topics*, each with a characteristic set of words. Classification is done by associating words in a document with the topics most likely to have generated the observed distribution of words in the document. The purpose of LDA is to determine those latent topics from patterns in the data.

The latent topics are useful for two purposes. First, to the extent that the words associated with a topic suggest a plausible category, they are intrinsically interesting in determining the issues found in the set of documents. For example, one of the sample data sets in the *R* lda package [11] determines the set of issues discussed in a series of political blogs. Second, the topics can be used with other classification algorithms such as logistic regression, support vector machines or discriminant analysis to classify new documents. The full mathematical details of LDA estimation can be obtained from that paper or the other usual suspects on the web and will not be repeated here, as I am simply using this off-the-shelf (or off-the-CRAN, as the case may be.)

Despite the surface differences between the domains, the application of this technique to the problem of political forecasting is straightforward: It is reasonable to assume that the stream of events observed between a set of actors is a mixture of a variety political strategies and standard operating procedures (for example escalation of repressive measures against a minority group while simultaneously making efforts to co-opt the elites of that group). This is essentially identical to the process by which a collection of words in a document is a composite of the various themes and topics, the problem LDA is designed to solve. As before, the objective of LDA

will be to find those latent strategies that are mixed to produce the observed event stream. These latent factors can then be used to convert full event stream to a much simpler set of measures.

The importance of latent dimensions in event data—rather than specifying the dimensions *a priori* based on some theory—is due to issues of measurement. As I noted in Schrodt (1994), if one is using event data in forecasting models—the objective of ICEWS—coding error is only one potential source of error that lies between "events on the ground" and the predictions of the forecasting model. These include

- News reports are only a tiny, tiny fraction of all of the events that occur daily, and are non-randomly selected by reporters and editors;
- Event ontologies such as WEIS, CAMEO and IDEA are very generic and bin together events that may not always belong together in all contexts;
- Forecasting models always contain specification error and cannot consider everything; for example few if any political forecasting models contain a full economic forecasting component;
- Political systems have a degree of intrinsic randomness due to their inherent complexity, chaotic factors even in the deterministic components of those systems, the impact of effectively random natural phenomena such as earthquakes and weather, and finally the effects of free will, so the error intrinsic to a forecasting model will never reduce to zero.

Because of these sources of error, the ability to determine latent dimension in event data is important in the overall scientific exercise of improving instrumentation for conflict forecasting. The latent dimensions of event data will never be not self-evident (or purely derivable from theory) because of the measurement factors noted above. We do not have a "god's-eye view" of political interactions—we have the highly (and non-randomly) selected view provided by the international media. Consequently determining methods that will allow these to be more effectively used to move the field forward more generally.

The LDA approach is similar in many ways to the hidden Markov approach. In both models, the observed event stream is produced by a set of events randomly drawn from a mixture of distributions. In an HMM, however, these distributions are determined by the state of a Markov chain, whose transition probabilities must be estimated but which consequently also explicitly provides a formal sequence. An LDA, in contrast, allows any combination of mixtures, without explicit sequencing except to the extent—as in this paper—that sequencing information is provided by the events in the model. The HMMs uses in political forecasting also tend to have a relatively small (typically about 5) set of states, and hence distributions, whereas LDA's typically use a larger number.

4.3 Social network analysis model

again, really overlaps with statistical at times.

4.4 Case-based reasoning

Hmmm, or is this really a null set at the moment, beyond essentially toy problems like CASCON

4.5 Sequence Comparison

Snag the D'Orozio-Yonamine lit review?

4.6 Sequence Development: hidden Markov models

[this will be considerably shortened to adjust for the content of the Bond HMM paper]

Hidden Markov models (HMM) are a recently developed technique that is now widely used in the classification of noisy sequences into a set of discrete categories (or, equivalently, computing the probability that a given sequence was generated by a known model), most commonly in speech recognition and comparing protein sequences.

An HMM is a variation on the well-known Markov chain model, one of the most widely studied stochastic models of discrete events (Bartholomew 1975). Like a conventional Markov chain, a HMM consists of a set of discrete states and a matrix $A = a_{ij}$ of transition probabilities for going between those states. In addition, however, every state has a vector of observed symbol probabilities, $B = b_j(k)$ that corresponds to the probability that the system will produce a symbol of type k when it is in state j. The states of the HMM cannot be directly observed and can only be inferred from the observed symbols, hence the adjective "hidden."

In empirical applications, the transition matrix and symbol probabilities of an HMM are estimated using an iterative maximum likelihood technique called the Baum-Welch algorithm. This procedure takes a set of observed sequences (for example the word "seven" as pronounced by twenty different speakers) and finds values for the matrices A and B that locally maximize the probability of observing those sequences. The Baum-Welch algorithm is a nonlinear numerical technique and Rabiner (1989:265) notes "the algorithm leads to a local maxima only and, in most problems of interest, the optimization surface is very complex and has many local maxima."

Once a set of models has been estimated, it can be used to classify an unknown sequence by computing the maximum probability that each of the models generated the observed sequence. This is done using an algorithm that requires on the order of N^{2T} calculations, where N is the number of states in the model and T is the length of the sequence. Once the probability of the sequence matching each of the models is known, the model with the highest probability is chosen as that which best represents the sequence. Matching a sequence of symbols such as those found in daily data on a six-month crisis coded with using the 22-category World Events Interaction Survey scheme (WEIS; [40]), generates probabilities on the order of 10^{T+1}—which is extremely small, even if the sequence was in fact generated by one of the models—but the only important comparison is the relative fit of the various models.

The application of the HMM to the problem of generalizing the characteristics of international event sequences is straightforward. The symbol set consists of the event codes taken from an event data set such as WEIS or CAMEO. The states of the model are unobserved, but have a close theoretical analog in the concept of crisis "phase". In the HMM, different phases would be distinguished by different distributions of observed CAMEO events. A "stable peace" would have a preponderance of cooperative events in the CAMEO 01-10 range; the escalation phase of the crisis would be characterized by events in the 11-15 range (accusations, protests, denials, and threats), and a phase of active hostilities would show events in the 18-22 range. The length of time that a crisis spends in a particular phase would be proportional to the magnitude of the recurrence probability a_{ii}.

The HMM has several advantages over alternative models for sequence comparison. First, if $N << M$, the structure of the model is relatively simple. For example a left-right model with N states and M symbols has $2(N-1) + N*M$ parameters compared to the $M(M+2)$ parameters of a Levenshtein metric. HMMs can be estimated very quickly, in contrast to neural networks and genetic algorithms. While the resulting matrices are only a local solution—there is no guarantee that a matrix computed from a different random starting point might be quite different—local maximization is also true of most other techniques for analyzing sequences, and the computational efficiency of the Baum-Welch algorithm allows estimates to be made from a number of different starting points to increase the likelihood of finding a global maximum. The HMM model, being stochastic rather than deterministic, is specifically designed to deal with noisy output and with indeterminate time (see Allan 1980); both of these are present in international event sequences.

An important advantage of the HMM, particularly in terms of its possible acceptability in the policy community, is that it can be trained by example: a model that characterizes a set of sequences can be constructed without reference to the underlying rules used to code those sequences. This contrasts with the interval-level aggregative methods using event data scales such as those proposed by Azar and Sloan ([3]) or Goldstein ([21]). These scales, while of considerable utility, assign weights to individual events in isolation and make no distinction, for example, between an accusation that follows a violent event and an accusation during a meeting.5 The HMM, in contrast, dispenses with the aggregation and scaling altogether—using

only the original, disaggregated events—and models the relationship between events by using different symbol observation probabilities in different states.

The HMM requires no temporal aggregation. This is particularly important for early warning problems, where critical periods in the development of a crisis may occur over a week or even a day. Finally, indeterminate time means that the HMM is relatively insensitive to the delineation of the start of a sequence: It is simple to prefix an HMM with a "background" state that simply gives the distribution of events generated by a particular source (e.g. Reuters/CAMEO) when no crisis is occurring and this occurs in the models estimated below. A model can simply cycle in this state until something important happens and the chain moves into later states characteristic of crisis behavior.

4.7 What else am I missing?

Also does this organization make sense? – this one is going to be more difficult. Another possibility, I suppose, would put the SNA and geo-spatial models in a separate section?

5 Conclusion

In 1954, the psychologist Paul Meehl [41] published a path-breaking analysis of the relative accuracy of human clinical assessments versus simple statistical models models for a variety of prediction problems, such as future school performance and criminal recidivism. Despite using substantially less information, the statistical models either outperformed, or performed as well as, the human assessments in most situations. Meehl's work has been replicated and extended to numerous other domains in the subsequent six decades and the results are always the same: the statistical models win. Meehl, quoted in Kahneman [34, chpt. 21], reflecting on thirty years of studies, said, "There is no controversy in the social sciences which shows such a large body of qualitatively diverse studies coming out in the same direction as this one."

This has not, of course, removed the human analysts from trying—and most certainly, *claiming*—to provide superior performance, for a wide variety of reasons— dare we call these *pathologies*?—that Kahneman [34, Part 3] discusses in great detail in the section of his book titled, appropriately, "Overconfidence." It is not only that simple statistical models are superior to human punditry, but as Tetlock [59] established, the more confident and well-known the human analyst, the *less* accurate his or her prediction, and in most fields, the past performance of analysts— notoriously, the stock-pickers of mutual funds—provides essentially no guide to their future performance.

At present, the likelihood that human punditry, particularly the opinionated self-assurance so valued in the popular media, will be completely replaced by the unblinking assessments of computer programs, whether on 24-hour news channels or in the brainstorming sessions in windowless conference rooms. Humans are social animals with exquisite skills at persuasion and manipulation; computer programs simply are far more likely to provide the correct answer in an inexpensive, consistent and transparent manner. Boring...

Yet with the vast increase in the availability of data, computational power, and the resulting refinement of techniques, there is some change in the works. While the covers of investment magazines show the latest well-coffed guru who by blind luck managed to have an unusually good year, in fact algorithmic trading now accounts for all but a small fraction of the activity on financial exchanges. Weather forecasts are presented by jocular and attractive individuals with the apparent intelligence of bullfrogs, but the forecasts themselves are the result of numerical models. Even that last bastion of intuitive manliness, the assessment of athletes, can be trumped by statistical models, as documented in the surprisingly popular book and movie *Moneyball* [37].[2]

Similar progress in the adoption of models forecasting political violence, particularly terrorism, is likely to be much slower. As we have stressed repeatedly, one of the major challenges of this research is that political violence is a rare event, and acts of terrorism in otherwise peaceful situations—the "bolt out of the blue" of Oklahoma City, 9/11/2001, Madrid 2004 and Utøya Island, Norway in 2011—are among the rarest. Consequently even if the statistical models are more accurate—and there is every reason to believe that they will be—establishing this will take far longer than is required in a field where new predictions can be assessed by the day, or even by the minute. In addition, a long series of psychological studies have shown that human risk assessment is particularly inaccurate—and hence its validity overestimated—in low probability, high risk situations, precisely the domain of counter-terrorism. Getting the technical assessments in the door, to say nothing of getting them used properly, will not be an easy task, and the initial applications will almost certainly be in domains where the events of interest are more frequent, as we are already seeing with the success of the Political Instability Task Force. As this chapter has illustrated, the challenges are well understood, a plethora of sophisticated techniques await experimentation, and the appropriate data is readily available, all that is needed is the will and skill to apply these.

---

[2] And when *this* book is made into a movie, we want Brad Pitt to play Sean O'Brien...

References

1. Andriole, S.J., Hopple, G.W.: "the rise and fall of events data: From basic research to applied use in the u.s. department of defense". International Interactions **10 (3/4)**, 293–309 (1984)
2. Azar, E.E.: The conflict and peace data bank (COPDAB) project. Journal of Conflict Resolution **24**, 143–152 (1980)
3. Azar, E.E., Sloan, T.: Dimensions of Interaction. University Center for International Studies, University of Pittsburgh, Pittsburgh (1975)
4. Bakeman, R., Quera, V.: Analyzing Interaction: Sequential Analysis with SDIS and GSEQ. Cambridge University Press, New York (1995)
5. Besley, T., Persson, T.: Repression or civil war? American Economic Review **99**(2), 292–297 (2009)
6. Blei, D.M., Ng, A., Jordan, M.: Latent dirichlet allocation. Journal of Machine Learning Research **3**, 993–1022 (2003). http://www.cs.princeton.edu/~blei/papers/BleiNgJordan2003.pdf
7. Bond, J., Petroff, V., O'Brien, S., Bond, D.: Forecasting turmoil in Indonesia: An application of hidden Markov models (2004). Presented at the International Studies Association Meetings, Montréal
8. Brandt, P.T., Freeman, J.R., Schrodt, P.A.: Real time, time series forecasting of inter- and intra-state political conflict. Conflict Management and Peace Science **28**(1), 41–64 (2011)
9. Buhaug, H.: Relative capability and rebel objective in civil war. Journal of Peace Research **43**(6), 691–708 (2006)
10. Burgess, P.M., Lawton, R.W.: Indicators of International Behavior: An Assessment of Events Data Research. Sage Publications, Beverly Hills (1972)
11. Chang, J.: Package lda: Collapsed Gibbs sampling methods for topic models. (2010). http://cran.r-project.org/web/packages/lda/; Version date: 24-October-2010
12. Choucri, N., Robinson, T.W. (eds.): Forecasting in International Relations: Theory, Methods, Problems, Prospects. W.H. Freeman, San Francisco (1979)
13. Daly, J.A., Andriole, S.J.: The use of events/interaction research by the intelligence community. Policy Sciences **12**, 215–236 (1980)
14. Davies, J.L., Gurr, T.R. (eds.): Preventive Measures: Building Risk Assessment and Crisis Early Warning. Rowman and Littlefield, Lanham, MD (1998)
15. Drakos, K., Gofas, A.: The devil you know but are afraid to face: Underreporting bias and its distorting effects on the study of terrorism. Journal of Conflict Resolution **50**(5), 714–735 (2006)
16. Esty, D.C., Goldstone, J.A., Gurr, T.R., Harff, B., Levy, M., Dabelko, G.D., Surko, P., Unger, A.N.: State Failure Task Force Report: Phase II Findings. Science Applications International Corporation, McLean, VA (1998)
17. Esty, D.C., Goldstone, J.A., Gurr, T.R., Surko, P., Unger, A.N.: State Failure Task Force Report. Science Applications International Corporation, McLean, VA (1995)
18. Freeman, J.R.: Systematic sampling, temporal aggregation, and the study of political relationships. Political Analysis **1**, 61–98 (1989)
19. Gaddis, J.L.: "international relations theory and the end of the cold war". International Security **17**, 5–58 (1992)
20. Gaddis, J.L.: The United States and the end of the Cold War. Oxford University Press, New York (1992)
21. Goldstein, J.S.: A conflict-cooperation scale for WEIS events data. Journal of Conflict Resolution **36**, 369–385 (1992)
22. Goldstone, J.A., Bates, R., Epstein, D.L., Gurr, T.R., Lustik, M., Marshall, M.G., Ulfelder, J., Woodward, M.: A global model for forecasting political instability. American Journal of Political Science **54**(1), 190–208 (2010)
23. Gurr, T.R., Harff, B.: Conceptual, research and policy issues in early warning research: An overview. Journal of Ethno-Development **4**(1), 3–15 (1994)

24. Gurr, T.R., Lichbach, M.I.: Forecasting internal conflict: A competitive evaluation of empirical theories. Comparative Political Studies **19**, 3–38 (1986)
25. Harff, B.: Early warning of humanitarian crises: Sequential models and the role of accelerators. In: J.L. Davies, T.R. Gurr (eds.) Preventive Measures: Building Risk Assessment and Crisis Early Warning, pp. 79–94. Rowman and Littlefield, Lanham, MD (1998)
26. Harff, B., Gurr, T.R.: Systematic early warning of humanitarian emergencies. Journal of Peace Research **35**(5), 359–371 (2001)
27. Hegre, H., Ostby, G., Raleigh, C.: Property and civil war: A disaggregated study of liberia. Journal of Conflict Resolution **53**(4), 598–623 (2009)
28. Hill, D.W., Bagozzi, B.E., Moore, W.H., Mukherjee, B.: Strategic incentives and modeling bias in ordinal data:the zero-inflated ordered probit (ziop) model in political science. http://qssi.psu.edu/files/NF4Hill.pdf (2011). Paper presented at the New Faces in Political Methodology meeting, Penn State, 30 April 2011.
29. Holmes, J.S., Pineres, S.A.G.D., Curtina, K.M.: A subnational study of insurgency: Farc violence in the 1990s. Studies in Conflict & Terrorism **30**(3), 249–265 (2007)
30. Hopple, G.W., Andriole, S.J., Freedy, A. (eds.): National Security Crisis Forecasting and Management. Westview, Boulder (1984)
31. Hudson, V. (ed.): Artificial Intelligence and International Politics. Westview, Boulder (1991)
32. Hudson, V.M., Schrodt, P.A., Whitmer, R.D.: 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)
33. Hultman, L.: Battle losses and rebel violence: Raising the costs for fighting terrorism and political violence. Terrorism and Political Violence **19**(2), 205–222 (2007)
34. Kahneman, D.: Thinking Fast and Slow. Farrar, Straus and Giroux, New York (2011)
35. Khong, Y.F.: Analogies at War. Princeton University Press, Princeton (1992)
36. Leng, R.J.: Behavioral Correlates of War, 1816-1975. (ICPSR 8606). Inter-University Consortium for Political and Social Research, Ann Arbor (1987)
37. Lewis, M.: Moneyball. W.W. Norton, New York (2011)
38. Li, Q.: Does democracy promote or reduce transnational terrorist incidents? Journal of Conflict Resolution **49**(2), 278–297 (2005)
39. May, E.R.: "Lessons" of the Past: The Use and Misuse of History in American Foreign Policy. Oxford University Press, New York (1973)
40. McClelland, C.A.: World Event/Interaction Survey Codebook (ICPSR 5211). Inter-University Consortium for Political and Social Research, Ann Arbor (1976)
41. Meehl, P.: Clinical and Statistical Prediction: A Theoretical Analysis and a Review of the Evidence. University of Minnesota Press, Minneapolis (1954)
42. Neustadt, R.E., May, E.R.: Thinking in Time: The Uses of History for Decision Makers. Free Press, New York (1986)
43. O'Brien, S.: Crisis early warning and decision support: Contemporary approaches and thoughts on future research. International Studies Review **12**(1), 87–104 (2010)
44. O'Brien, S.P.: Crisis early warning and decision support: Contemporary approaches and thoughts on future research. International Studies Review **12**(1), 87–104 (2010)
45. Pevehouse, J.C., Goldstein, J.S.: Serbian compliance or defiance in Kosovo? statistical analysis and real-time predictions. The Journal of Conflict Resolution **43**(4), 538–546 (1999)
46. Piazaa, J.: Incubators of terror: Do failed and failing states promote transnational terrorism? International Studies Quarterly **52**(3) (2008)
47. Piazaa, J.: Poverty, minority economic discrimination and domestic terrorism. Journal of Peace Research **48**(3), 339–353 (2011)
48. Piazaa, J., Walsh, J.I.: Transnational terrorism and human rights. International Studies Quarterly **53**, 125–148 (2009)
49. Richardson, L.F.: Statistics of Deadly Quarrels. Quadrangle, Chicago (1960)
50. Rupesinghe, K., Kuroda, M. (eds.): Early warning and conflict resolution. St. Martin's Press, New York (1992)
51. Schrodt, P.A.: Parallel event sequences in international relations. Political Behavior **12**(2), 97–123 (1990)

52. Schrodt, P.A.: Pattern recognition of international crises using hidden Markov models. In: D. Richards (ed.) Political Complexity: Nonlinear Models of Politics, pp. 296–328. University of Michigan Press, Ann Arbor (2000)
53. Schrodt, P.A.: Detecting united states mediation styles in the middle east, 1979-1998. In: Z. Maoz, A. Mintz, T.C. Morgan, G. Palmer, R.J. Stoll (eds.) Multiple Paths to Knowledge in International Relations, pp. 99–124. Lexington Books, Lexington, MA (2004)
54. Schrodt, P.A.: Forecasting conflict in the Balkans using hidden Markov models. In: R. Trappl (ed.) Programming for Peace: Computer-Aided Methods for International Conflict Resolution and Prevention, pp. 161–184. Kluwer Academic Publishers, Dordrecht, Netherlands (2006)
55. Shearer, R.: Forecasting Israeli-Palestinian conflict with hidden Markov models (2006). Available at `http://eventdata.psu.edu/papers.dir/Shearer.IP.pdf`
56. Shellman, S.: Process matters: Conflict and cooperation in sequential government-dissident interactions. Journal of Conflict Resolution **15**(4), 563–599 (2000)
57. Shellman, S.: Time series intervals and statistical inference: The effects of temporal aggregation on event data analysis. Security Studies **12**(1), 97–104 (2004)
58. Sylvan, D.A., Chan, S.: Foreign Policy Decision Making: Perception, Cognition and Artificial Intelligence. Praeger, New York (1984)
59. Tetlock, P.E.: Expert Political Judgment: How Good Is It? How Can We Know? Princeton University Press, Princeton, NJ (2005)
60. Ward, M.D., Greenhill, B.D., Bakke, K.M.: The perils of policy by p-value: Predicting civil conflicts. Journal of Peace Research **47**(4), 363–375 (2010)
61. Weidmann, N.B., Toft, M.D.: Promises and pitfalls in the spatial prediction of ethnic violence: A comment. Conflict Management and Peace Science **27**(2), 159–176 (2010)