Cluster-Based Early Warning Indicators for Political Change in the Contemporary Middle East

Philip A. Schrodt and Deborah J. Gerner

Department of Political Science

University of Kansas

Lawrence, KS 66045 USA

phone: 913-864-3523 fax: 913-864-5700

p-schrodt@ukans.edu d-gerner@ukans.edu

Authors' Note: An earlier version of this paper was presented at the 1996 annual meeting of the American Political Science Association, San Francisco. We received valuable suggestions on revisions from the "Beer and Politics Seminar" at the University of Kansas. This research was funded in part by the National Science Foundation through grant SBR-9410023 and the University of Kansas General Research Fund grant 3500-X0-0038. Address correspondence to Philip A. Schrodt, Department of Political Science, University of Kansas, 504 Blake Hall, Lawrence, KS 66045, USA; email: p-schrodt@ukans.edu.

ABSTRACT

This article uses cluster analysis to develop an early warning model of political change in the Levant between April 1979 and July 1996 as reflected in WEIS-coded event data generated from Reuters. We employ a new clustering algorithm that uses the correlation between dyadic behaviors at two points in time to identify clusters of political activity. The transition to a new cluster occurs when the time points are closer in distance to later time points than to preceding points. These data clusters begin to "stretch" prior to breaking apart; this characteristic can be used as an early-warning indicator. The clusters identified by this technique correspond well with phases of political behavior identified *a priori*. A Monte-Carlo analysis shows that the clustering and early warning measures are not random; they perform very differently in simulated data sets having the same mean, variance, and autocorrelation as the observed data.

Introduction

The problem of early warning of political change has long been one of the central foci of event data research. Much of the early impetus—and quite critically, early funding—for the approach came from the United States Defense Advanced Research Projects Agency (DARPA), for example with the Early Warning and Monitoring System (EWAMS; see Hopple 1984; Laurance 1990). These efforts were contemporaneous with the development of a variety of other systematic forecasting techniques (see Choucri and Robinson 1979; Hopple, Andriole, and Freedy 1984; Phillips and Rimkunas 1983; Singer and Wallace 1979) as well as the creation of the World Event Interaction Survey (WEIS) event coding scheme.

These initial efforts failed to take hold in the policy community for a variety of institutional and pragmatic reasons (Andriole and Hopple 1984; Daly and Andriole 1980; Laurance 1990). By the mid-1980s, interest in, and funding of, these statistical early warning projects had essentially ceased. The policy community continued to spend billions of dollars on political forecasts in the form of analysis by intelligence agencies, but virtually all *political* forecasting—as distinct from economic and demographic forecasting—was done using traditional techniques. Because human-coded event data were (and are) expensive to generate, the development of publicly-available data collections ceased and the academic community was constrained to re-analyzing legacy collections of event data that ended around 1977.[1]

Ironically, the emphasis on event-based early warning ended at the same time that two technological changes made it more feasible. First, the revolution in electronic communications made available a vastly greater amount of information about political affairs than that which had been tapped by the earlier efforts, which usually relied on a small number of elite Western

---

[1] An assortment of small, geographically-specific event data sets were collected in the academic community after 1977—for instance, Ashley (1980) and Van Wyk and Radloff (1993). More important, the WEIS data set—the centerpiece of the DARPA efforts—continued to be maintained through a variety of public and private efforts through the 1980s and 1990s (Tomlinson 1993), but it was not widely available or used.

newspapers. By the early 1990s, much of this information was available in machine-readable form, first through commercial services such as NEXIS and Dialog, and now through the World Wide Web. Second, the exponential increase in the computational power available to researchers allowed for the use of a number of statistical and coding techniques that were impossible during the earlier period of DARPA research.

In addition to these technological changes, the increased complexity of the forecasting problem in the post-Cold War world—coupled with some sobering assessments of the failure of traditional forecasting methods to anticipate the end of the Cold War itself—have renewed interest in this problem in the international relations literature (Gurr and Harff 1994; Gurr and Harff in press; Rupesinghe and Kuroda 1992) and by the policy community (Alker, Gurr, and Rupesinghe 1995; Boutros-Ghali 1992; Dedring 1994; Mizuno 1995).

The purpose of this article is to explore some approaches to early warning that utilize the technologies and conceptual approaches of the 1990s. The paper starts with a review of the statistical early warning problem and discusses several possible techniques. We then apply cluster analysis to an event data set to show that these events tend to form temporally-delineated clusters and that the movement of points in those clusters can be used as an early warning indicator. The existence of these clusters is consistent with the theories of "crisis phase" that underlie several earlier approaches to early warning; furthermore, the clusters we find in actual event data differ significantly from those that occur in a set of simulated data having similar statistical characteristics. The regional and temporal focus of the study is the Levant—Egypt, Israel, Jordan, Lebanon, the Palestinians, and Syria, plus the United States and USSR/Russia—between April 1979 and July 1996.

Statistical Approaches to Early Warning: A Review

This section will review past approaches to statistical early warning in order to justify the clustering approach that we employ in our analysis. It will not, however, consider the large literature on non-statistical (qualitative) approaches to forecasting: Contemporary surveys of qualitative approaches can be found in Rupesinghe and Kuroda (1992), Gurr and Harff (1994),

and Adelman and Schmeidl (1995). We also will not deal with the topic of long-range forecasting using formal methods, which is primarily done using simulation; Ward (1985) and Hughes (1993) summarize that literature.

Statistical approaches to early warning can be classified into two broad categories: structural and dynamic. The *structural* category consists of studies that use events (or more typically, a specific category of event such as civil or international war) as dependent variables and explain these using a large number of exogenous independent variables. In the domain of domestic instability, this approach is exemplified by the work of Gurr and his associates, most recently in the "State Failure Project" [SFP] (Esty et al. 1995; Gurr 1995); Gurr and Harff (1996) and Gurr and Lichbach (1986) describe a number of such research projects. In the field of international instability, the structural approach is illustrated by the work of Bueno de Mesquita and his associates, and more generally by the Correlates of War project; Gochman and Sabrosky (1990), Midlarsky (1993), and Wayman and Diehl (1994) provide general surveys. Structural approaches have tended to use multivariate linear regression models. Recently, however, the research has branched out to other techniques; for example, the SFP uses logistic regression, neural networks, and time series methods.

In contrast to the structural approach, in *dynamic* early warning models event data measures are used as both the independent and dependent variables. Most of the event data projects of the late 1970s classified dyads with respect to the likelihood of a crisis based on a set of event-based empirical indicators. For instance, DARPA's EWAMS evaluated three WEIS-based indicators (conflict, tension, and uncertainty) to determine an alert status for any dyad. Azar et al. (1977) use a similar approach based on whether behaviors measured with the COPDAB event scale fall outside a range of "normal" interactions for the dyad. More recent efforts employ increasingly-advanced econometric time-series methods that model an interval-level measure of events as an autoregressive time series with disturbances. Goldstein and Freeman (1990) provide a book-length example of this approach; Dixon (1986), Goldstein and Pevehouse (forthcoming), Lebovic

(1994), Ward (1982), and Ward and Rajmaira (1992) illustrate the continued development of dynamic models of events.

Scholars justify the dynamic approach—which is at odds with most political science statistical modeling in using only lagged endogenous variables—in three ways. The first rationale is that many of the structural variables that are theoretically important for determining the likelihood of conflict do not change at a rate sufficient for use as early warning indicators; in fact, many are essentially fixed (e.g., ethnic and linguistic heterogeneity, historical frequency of conflict, natural resource base). Data on variables that do change—for instance, unemployment rates, economic and population growth rates—are often reported only on an annual basis and the quality of these reports tends to be low in areas under political stress.

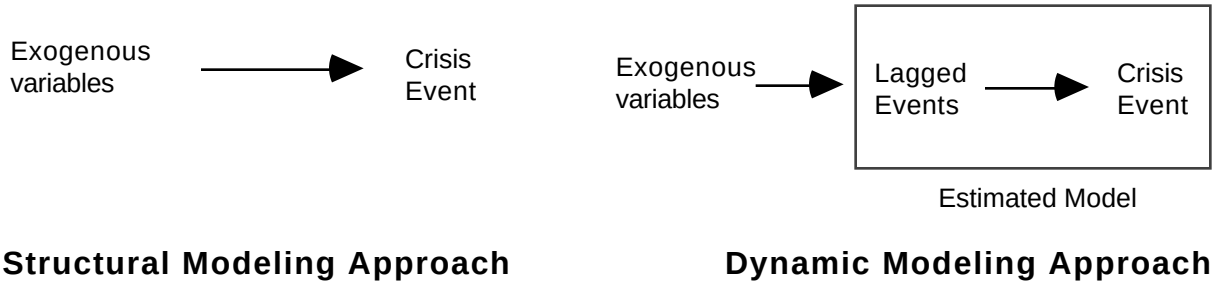
The second justification for the dynamic approach is that it reduces the information required by the model. The data collection effort of the SFP, for example, measures more than 50 independent variables (Gurr 1995:5-7); this requires a large amount of information from a vast number of sources.[2] In contrast, the event data collections used in dynamic models focus on reported political interactions that can be collected systematically in real-time, which increases the predictive utility of the model.

The final justification for dynamic modeling involves the nature of political events themselves: The approach assumes that the effects of exogenous variables used in the structural models will be reflected in the pattern of events prior to a major change in the political system. As illustrated in Figure 1, the dynamic approach effectively uses the lagged values of events as a substitute for structural variables.

---

[2] The final models developed in the project—which unfortunately are still classified—apparently involve only a half-dozen or so variables out of this much larger collection (Gurr, personal communication, August 1996).

Figure 1. Comparison of the structural and dynamic approaches to early warning



To take a concrete illustration, Gurr (1995: 7) notes that "ethnic heterogeneity probably is most significant for state failure when it coincides with lack of democracy and low regime durability." Consequently, the SFP includes measures for those three variables: ethnolinguistic diversity, regime democracy, and regime durability.

A dynamic approach, in contrast, would not measure these aspects of a political system directly, but would instead assume that each would be reflected in the types of events picked up by the international media. The presence of democracy, for instance, would be reflected not only in periodic elections but in a large number of reports discussing disagreements between the government and the elected opposition. A low level of regime durability would be reflected in coups and attempted coups. To the extent that ethnicity was an important political factor, it would be reflected in ethnically-oriented political rallies, outbreaks of violent ethnic conflict, and similar events. A suitably-designed event coding scheme should detect the presence or absence of these events and make the appropriate forecast without directly measuring the underlying variables.

At a *theoretical* level, the dynamic-events approach accepts the importance of exogenous structural variables: *Ceteris paribus*, countries with a high level of ethnic heterogeneity will have a greater propensity for conflict than those with a low level; democracies are likely to be different than autocracies, and so forth. The distinction between the early warning approaches is a matter of *measurement*: the structural modeling approach seeks to measure these variables directly, whereas

the dynamic approach assumes we can indirectly measure the effects of these variables through the patterns of events the variables generate.[3]

This is an optimistic, but not wholly implausible, assumption. For example, in the Reuters-based data with which we have been working, there is a clear contrast between Israel and Syria with respect to the presence of a democratic opposition and between Lebanon and Egypt with respect to the importance of ethnicity: The ethnic conflict in Lebanon is one of the most conspicuous features of the data set. Our impression is that the increased democratization in Jordan and the fluctuations in the Egyptian government's acceptance of a democratic opposition would also be reflected in the activities reported in Reuters, although we have not attempted to analyze this.

Statistical Characteristics of the Early Warning Problem

Standard econometric time series methods have only limited utility in the problem of early warning. In general, time series analysis seeks to determine a function

---

[3] An econometric analogy to this is found in the distinction between "technical" and "fundamental" analysis of stock prices. A fundamental analysis attempts to predict price changes on the basis of underlying factors such as marketing, management, raw material prices, and macroeconomic trends. Technical analysis assumes that these factors will be reflected in the patterns of the movements of the price of a stock (or set of stocks); therefore analysis of those prices alone will provide sufficient information for forecasting. Fundamental analysis corresponds to the structural approach to modeling political events; technical analysis to the dynamic.

Until relatively recently, technical analysis generally had a bad reputation, consisting as it did largely of statistically-dubious patterns based on small samples, wishful thinking, and gurus whose fortunes were based more on the sale of books than on trading stock. With the increase in computing power in the 1980s, the situation changed, and "programmed trading systems" can now process sufficiently large amounts of information to generate profits (and periodically throw the market into chaos) working solely with information endogenous to the market itself. The increased information processing capacity in the 1990s—in contrast to that available in the 1970s—may have a similar effect on event data analysis.

$$y_{t+k} = f(y_t, y_{t-1}\ldots, \mathbf{X}_t, \mathbf{X}_{t-1}\ldots)$$

for some k>0. In English, the fundamental problem of time series is to determine the future values of a variable y given some present and past values of that variable and (possibly) the present and past values of a set of exogenous variables **X**. Due to the importance (and potential financial rewards) of accurate economic forecasts, there is a massive literature on time series estimation in econometrics (see Hamilton 1994).

In contrast, the problem of statistical early warning consists of finding a time T such that

$$| y_t - y_s | > \xi \text{ for all } t > T > s$$

for some indicator variable y. This means that the variable y has consistently different values after time T than it had prior to time T, which would occur in aggregated event data following a qualitative shift in the type of political behavior in which a dyad was engaged.

An additional distinction is that econometric time series tend to be highly autoregressive (e.g., GNP, unemployment, prices of consumer goods, and inflation rates) or at least have an autoregressive component combined with generally random noise (e.g., stock prices, exchange rates). The GNP or unemployment rate of a major industrialized economy has tremendous inertia. For instance, while the stock market crash of October 1929 was sudden, the high unemployment rates of the Great Depression required two or three years to develop fully. Furthermore, most econometric time series are measured continuously rather than episodically, so missing data are less of an issue.

The early warning problem, on the other hand, focuses on shifts in the time-series that are *not* autoregressive, even though the series taken as a whole might be autoregressive. An autoregressive model of war-and-peace will be very accurate, as illustrated by the presumably apocryphal story about the European political analyst who said "Every day from 1910 to 1970, I predicted that Europe would remain at peace when at peace, and remain at war when at war, and I

was only wrong four times." This type of model is not, however, very useful.[4] The econometric problem most comparable to political early warning is forecasting sudden economic shifts such as those observed in massive exchange rate fluctuations (e.g., the collapse of the Mexican peso or the European Exchange Rate Mechanism).[5] These problems are similar to political early warning in that they are primarily psychological and do not reflect a major change in the underlying physical reality: The economic fundamentals of the Mexican or European economies did not change dramatically during the days of the exchange-rate crises, but the perceptions of the future values of the relevant currencies did change.

Despite these complications, it should be noted that in two very important respects prediction is an *easier* problem than the typical econometric estimation problem. First, forecasting models have right-and-wrong answers, or at least their accuracy can be evaluated probablistically. Coefficient estimation problems, in contrast, do not have answers: one can always specify an error structure, prior probability, or alternative model structure that places the estimated emphasis on different variables, and there is no empirical method of deciding between these specifications. Second, and closely related to the first issue, forecasting problems are not affected by collinearity, which is the bane of coefficient estimation in the social sciences because every behavior tends to be linked to every other behavior. Coefficient estimates with low standard errors are clearly useful for obtaining a theoretical understanding of a situation, but they are not essential for the pragmatic purposes of forecasting (Wonnacott and Wonnacott 1979:81). For this reason, it is not surprising that models with very diffuse coefficient structures—for example, neural networks and VAR—are found increasingly in early warning research.

---

[4] More technically, such a measure succeeds according to a frequency-based measure but fails according to an *entropy*-based measure (Pierce 1980), which places higher weight on the prediction of low-probability events.

[5] Hamilton's (1989; 1994, chapter 22) work on modelling a time series that shifts between multiple underlying states—following the Goldfeldt and Quandt switching regression scheme—is an econometric approach to this problem and could use further investigation.

A Cluster-based Approach to Early Warning

In Schrodt and Gerner (1997), Schrodt, Huxtable and Gerner (1996), and Schrodt and Gerner (1996) we explored the possibility of using cluster analysis as an alternative to conventional time series methods. We analyzed behavior in the Middle East under the assumption that crises go through a series of phases that are delineated by distinct sets of behaviors. In the empirical literature, crisis "phase" has been coded explicitly in data sets such as the Butterworth international dispute resolution dataset (Butterworth 1976), CASCON (Bloomfield and Moulton 1989, 1997) and SHERFACS (Sherman and Neack 1993).[6] Describing the early CASCON work, Sherman and Neack explain that:

> conflict is seen "as a sequence of phases." Movement from phase to phase in a conflict occurs as "the factors interact in such a way as to push the conflict ultimately across a series of *thresholds* toward or away from violence" (Bloomfield and Leiss 1969). Characteristics of disputes can be visualized as the timing and sequencing of movement between and among phases. Processes of escalation of violence, resolution or amelioration of the seriousness (threat of violence-hostilities) and settlement are identifiable through the use of phrase structures. (Sherman and Neack 1993:90)

CASCON and SHERFACS, for example, code six phases: "dispute phase," "conflict phase," "hostilities phase," "post-hostilities conflict phase," post-hostilities dispute phase," and "settlement phase."

If the concept of crisis phase is valid, the behaviors observed in an international subsystem should fall into distinct patterns over time. If the transitions between these phases are gradual, or if behaviors that precede a phase transition are distinct from those found when the system is locked in a single phase, then those behaviors can be used for the purpose of early warning.

We have been analyzing behavior by monitoring the position of the vector

---

[6] Sherman and Neack (1993) provide a review of the evolution of these data sets.

$$[AB, AC, AD, \ldots, AH, BA, BC, \ldots, BH, CA, \ldots, HF, HG]_t$$

where A, B, ..., H are the actors in the system and XY_t is the total Goldstein-scaled events directed from X to Y aggregated over a month.[7] The behavior of the system is simply the path that this vector traces over time in a 54-dimensional space. In vector terminology, a "phase" is characterized by a region in the vector space where points cluster over time. Empirically, a phase typology would be evident by the system spending most of its time inside these distinct clusters of behaviors that characterize the phase, with brief transitions between the clusters.

Figure 2: Schematic representation of phases during the WWII period

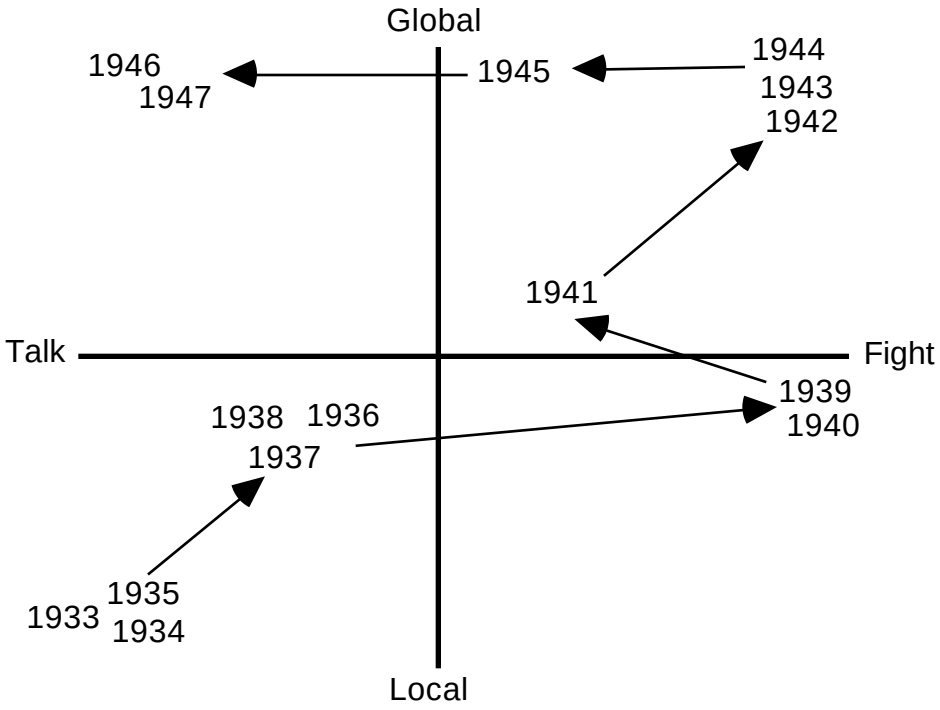


Figure 2, from Schrodt and Gerner (1997), illustrates this process informally for the World War II period, using the two dimensions of "talking versus fighting" and "local versus global

[7] In other words, we converted each X>Y event to its numerical score on the Goldstein (1992) scale, then totaled these numerical scores by month. Schrodt and Gerner (1994) gives a number of time series plots of the data for the 1982-1993 period.

involvement." The years prior to 1936 involved little violent inter-state conflict. The system then shifted to a series of militarized crises during the period 1936-38 and erupted into a full-scale European war in 1939-40. After a lull in the early part of 1941, the war spread, first to the USSR and then to the Pacific; the 1942-1944 period was characterized by a global war. In 1945, this war ended, first in Europe and then in the Pacific, but the post-war politics, rather than returning to the unilateralism/isolationism of the pre-war period, remained global. The 1946-47 cluster continued to characterize the system for most of the Cold War, with occasional departures from that cluster to take in the Korean War, the Suez Crisis, the Cuban Missile Crisis and similar events.

Figure 2 is idealized and any analysis using event data will be complicated by the problem of aggregating dyadic behaviors, the existence of multiple issues determining behaviors, and the fact that real-world political behavior is considerably noisier than the simple summary of international politics in the 1930s and 1940s presented above.[8] Nevertheless, if event data capture the behaviors characterizing a phase typology, it should be possible to determine those phases using clustering.[9] A cluster will occur whenever there is an extended period of time when the countries in the system are reacting to each other in a consistent fashion—in other words, repeating approximately the same types of actions (cooperative, conflictual, or absent) month after month. When the behavior of a dyad or set of dyads changes—for example, from peace to war or vice versa—the system shifts to a new cluster.

The effectiveness of event-space clustering in early warning depends on whether some measurable characteristic of the behavior of the system changes *prior to* the phase transition. In some instances no precursors to a phase transition will be present, either because of deliberate

---

[8] A close analog to this approach is found in the DARPA-sponsored work of Phillips and Rimkunas (1983: 181-213), which analyzes WEIS data in a two-dimensional space of "threat" and "uncertainty" using Thom's (1975) "cusp catastrophe" model. Their model successfully locates eight of eighteen crises identified in the WEIS data, and produces no false positives.

[9] For a review of clustering techniques, see Aldenderfer and Blashfield (1984)Bailey (1994), and Everitt (1980).

concealment (Rwanda) or lack of interest by the media (Chechnya, Somalia). Our conjecture, however, is that most political situations go through a gradual deterioration (or improvement) of affairs prior to a phase transition, rather than experiencing a sharp jump. Furthermore, because news-gathering organizations are usually rewarded for correctly anticipating political events, journalists who are present in the region, understand the local politics, and can get their stories past editors and onto the news wires are likely to report the behaviors they perceive to be pre-cursors to any political phase change.

The approach we are using to develop an early warning indicator is similar to the "normal relations range" concept proposed by Edward Azar:

> Over a period of time any two nations establish between them an interaction range which they perceive as "normal." This normal relations range (NRR) is an interaction range ... which tends to incorporate most of the signals exchanged between that pair and is bound[ed] by two critical thresholds—an upper threshold and a lower threshold. The upper critical threshold is that level of hostility above which signals exhibited by either member of the interacting dyad are regarded as unacceptable to the other. Interaction above the present upper critical threshold ... for more than a very short time implies that a crisis situation has set in. (Azar 1972:184)

The NRR model implies that events will cluster and the NNR for each dyad will be the diameter of the cluster in the dimension of that dyad. We generalize Azar's NRR concept by looking at changes in a large number of dyads simultaneously, whereas Azar looked only at one dyad at a time.[10] We assume the system is moving away from normal behavior when it nears (or passes) the edge of the cluster (rather than when it exceeds a single critical threshold). In addition, we look at the *density* of clusters—defined as the average distance between the points in a cluster—over time. Behavior within the NRR should result in dense clusters, whereas when a system moves

---

[10] We also use a standardized metric based on correlation, whereas Azar used a Euclidean metric and established distinct critical ranges for each dyad.

away from one phase/cluster/NRR and into another it will usually experience a period when the points do not cluster densely.

Data

The data used in this study were machine-coded from Reuters lead sentences obtained from the NEXIS data service for the period April 1979 through July 1996. We coded these data using the Kansas Event Data System (KEDS) machine-coding program (Gerner et al. 1994; Schrodt, Davis and Weddle 1994).[11] KEDS does some simple linguistic parsing of the news reports—for instance, it identifies the political actors, recognizes compound nouns and compound verb phrases, and determines the references of pronouns—and then employs a large set of verb patterns to determine the appropriate event code. Bond et al. (1996), Huxtable and Pevehouse (1996), and

---

[11] The NEXIS search command used to locate stories to be coded was

(ISRAEL! OR PLO OR PALEST! OR LEBAN! OR JORDAN! OR SYRIA! OR EGYPT!)

AND NOT (SOCCER! OR SPORT! OR OLYMPIC! OR TENNIS OR BASKETBALL)

We coded only the lead sentences of the stories; this produced a total of 80,519 events. The search command generates a number of events that are outside the 54 directed dyads considered in this study. Those 54 dyads contain 34,707 events.

In contrast to the data that we have used in earlier papers (Schrodt and Gerner 1994; Schrodt and Gerner 1997), these data were generated using a "complexity filter" that did not code any sentence containing six or more verbs or with no actor prior to the verb. Sentences meeting these criteria had a greater likelihood of being incorrectly coded by KEDS; by using the filter we should have a somewhat less noisy data. Schrodt and Gerner (1996:8) compares the two sets statistically.

Both data sets and the source code for the computer programs used in the analysis are available on disk from the authors, from the web site http://www... and upon publication will be deposited in the ICPSR "Publication-Related Archive."

Schrodt and Gerner (1994) discuss extensively the reliability and validity of event data generated using Reuters and KEDS.

We converted the individual WEIS events to a monthly net cooperation score using the numerical scale in Goldstein (1992) and totaling these numerical values for each of the directed dyads for each month. We examined all the dyads involving interactions among Egypt, Israel, Jordan, Lebanon, the Palestinians, Syria, United States, and Soviet Union/Russia except for the USA>USR and USR>USA dyads; this gives a total of 54 directed dyads with 208 monthly totals in each dyad.[12]

Following the approach we used in Schrodt and Gerner (1997), we assigned the *a priori* phase identifications in Table 1 based on the dominant political interactions during each period. Our discussion of the results of the clustering and the early warning indicator will use these *a priori* clusters as a reference point.

Table 1. A Priori Phase Assignments

Label	Dates	Months	Defining Characteristic
Camp David	Apr.79-May.82	38	Before Israel's 1982 invasion of Lebanon
Lebanon	Jun.82-May.85	36	Israeli troops in Lebanon
Taba	Jun.85-Nov.87	30	Israeli withdrawal from most of Lebanon until the *intifada*
Intifada	Dec.87-Jul.90	32	Palestinian *intifada*
Kuwait	Aug.90-Oct.91	15	Iraq's invasion of Kuwait until start of Madrid talks
Madrid	Nov.91-Aug.93	22	Bilateral and multilateral peace talks
Oslo	Sept.93-Jul.96	35	Oslo peace process

[12] In Schrodt and Gerner (1997) we repeated several of our analyses with the USA>USR and USR>USA dyads included. This made only minor changes in the results.

Detection of Phase using Clustering over Time

In Schrodt and Gerner (1997), we analyzed a data set for phases using the SPSS K-Means clustering algorithm and the Euclidean metric as the measure of the distance between points. This technique able to identify the phases that we had assigned *a priori* in the first half of the period but was less successful in the second half. That analysis also seemed to suggest there was instability in the cluster assignment prior to a change in phase; however we did no quantitative analysis of the actual distances between the points and clusters.

K-Means is a very general cross-sectional clustering method and does not incorporate the time-series element of event data. In the final third of the data series, for instance, the K-Means algorithm assigned points to clusters that contained many other points that were quite distant in time. Because the Levantine sub-system does not include all relevant interactions—for example, the end of the Cold War—the resemblance to earlier clusters may be superficial.

Including time as the dominant dimension actually simplifies the delineation of clusters in comparison to K-Means. The clustering algorithm we employ in this study is simple: a new cluster is formed if x_t is closer to the k points following it in time than it is to the k points that precede it in time, plus some threshold that prevents new clusters from being formed because of random fluctuations in the event data that are unrelated to phase transitions.[13] Mathematically, a new cluster is considered to be established at a point x_t when

$$\mathrm{LML}_t = \frac{1}{k}\sum_{i=1}^{k} ||x_t - x_{t-k}|| - \frac{1}{k}\sum_{i=1}^{k} ||x_t - x_{t+k}|| > \Delta$$

"LML" is the lagged distance minus leading distance: ||x-y|| is the distance between x and y according to some metric and Δ is the threshold parameter. From the perspective of cluster analysis, this approach is similar to the "minimum spanning tree" approach (see Backer 1995:

---

[13] Calculations were done with a simple (600-line) Pascal program that produced various tab-delimited files which were read into Excel to produce the figures and tables.

chapter 1) in dividing the clusters at places where a large distance is found between adjacent points; it differs in using the dimension of time rather than a tree to determine which points are adjacent.

Figure 3 shows the results of analyzing our Middle East data set using this algorithm for k=4 and the correlation metric:

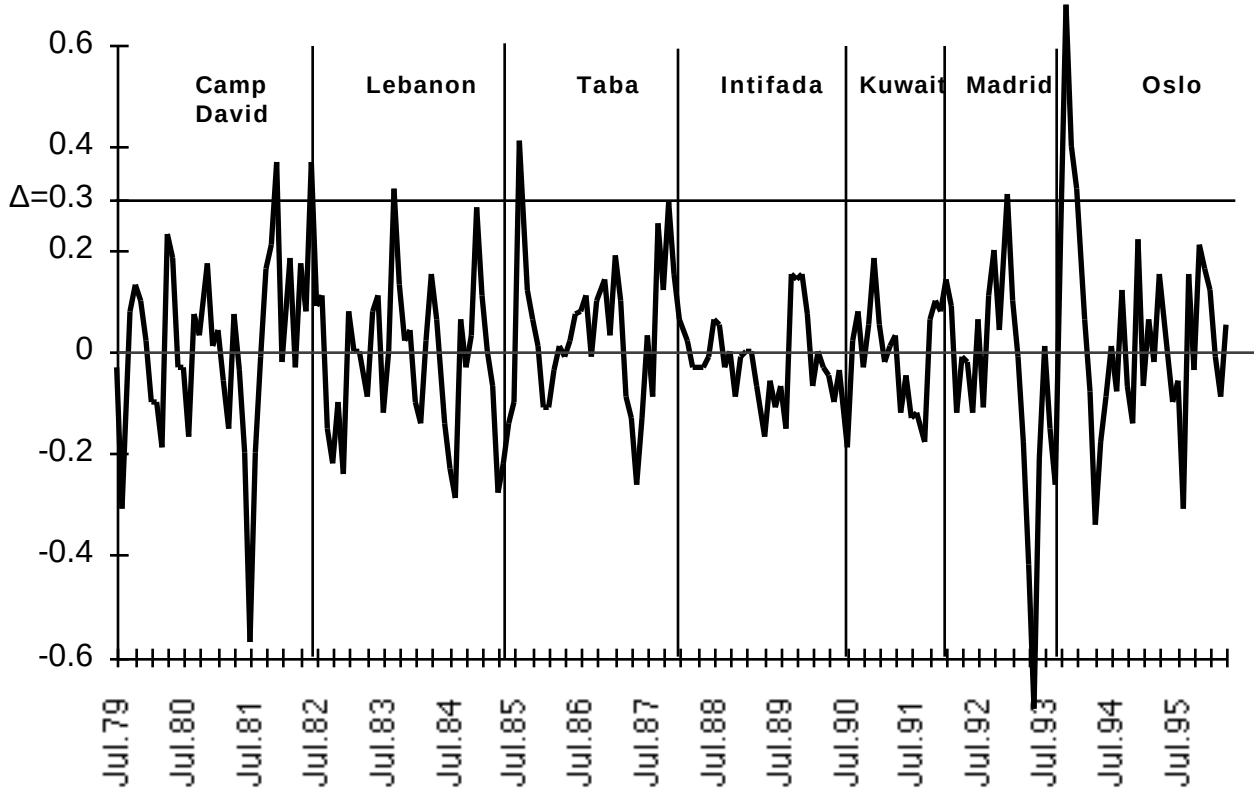
$$||x - y|| = 1 - r_{x,y}$$

where r is the Pearson product moment, The vertical lines on the graph correspond to time points where the *a priori* cluster divisions from Table 1 are located. We also experimented with some larger values of k and the results are much the same as those obtained with k=4. If we set $\Delta = 0.30$ the correlation metric picks four out of the six the *a priori* phase assignments and identifies several other plausible transitions, some of which were also found by the K-Means analysis:

- a pre-Lebanon change, probably reflecting increased tension between Israel and the PLO prior to the actual invasion;
- two pre-Taba changes that may correspond to shifts in Israeli and Syrian policy in Lebanon; K-Means also divided the Lebanon period into at least two phases;
- a peak in January 1993 that may reflect the USA change in policy towards the Middle East that occurred with the change from the Bush to Clinton administrations.

The measure misses the Kuwait transition, which all of our clustering efforts have failed to pick up, as well as the Madrid transition.

Figure 3. 4-month LML measure



Change in Cluster Density as an Early Warning Indicator

Examination of Figure 3 shows that, in most cases, the LML measure begins a rapid increase several months before a phase transition occurs. This is consistent with the underlying theory of phase transitions because the changing interactions in the system would cause the points to pull away from the cluster before they make the final break, rather like pulling on a piece of taffy. This pattern suggests that the change in the *density* of the cluster might serve as an early warning indicator. The critical difference between this type of analysis and the previous analysis involving LML is that the change in cluster density can be identified solely on the basis of information available up to and including time t—and hence can be done prospectively—whereas computing LML_t requires information after time t and can only be done retrospectively.

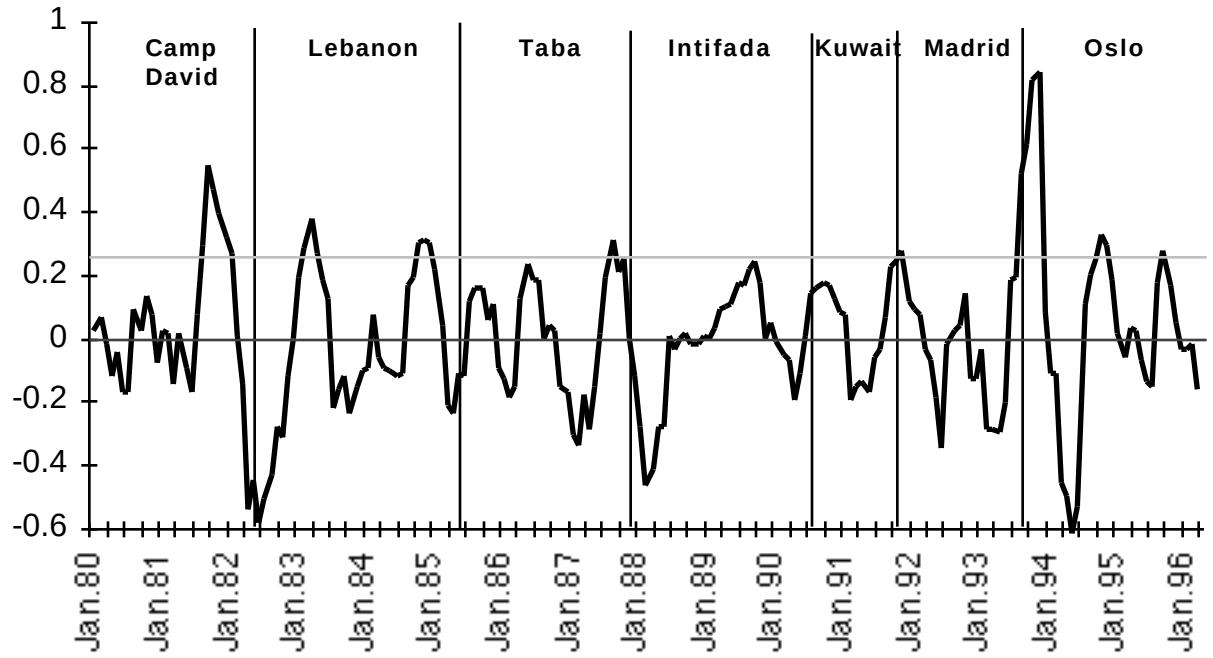
Figure 4 shows such a cluster-density measure, Δ_8CD. This measure is calculated by first computing the total distance between points in a cluster of 4 consecutive months

$$CD_t = \frac{1}{6} \sum_{i=0}^{3} \sum_{j=i+1}^{3} \|x_{t-i} - x_{t-k}\|$$

and then calculating the difference between CD_t at points that are 8 months apart (in other words, $\Delta_8 CD = CD_t - CD_{t-8}$). The Δ_8CD measure generally corresponds well with both the *a priori* and LML transitions, despite the fact that the LML clusters were based on *post-hoc* information. An LML cluster transition occurs in the vicinity of every point where Δ_8CD exceeds one standard deviation (0.23). Unlike LML, the Δ_8CD picks up the Madrid transition, though it still fails to show the Kuwait transition, which arguably occurred due to factors exogenous to the system. A peak in the measure in mid-1989 probably corresponds to the decline of reports of activity in the Palestinian *intifada*;[14] the peaks in early 1995 and 1996 may reflect changes associated with the problems encountered in the Oslo peace process.

[14] This is the "media fatigue" effect that is discussed in Gerner and Schrodt (forthcoming).

Figure 4. 8-month change in 4-month cluster dispersion as an early warning indicator



The Δ_8CD measure is continuous and can be interpreted as being proportional to the probability of a major change occurring, rather than only providing the *yes/no* prediction of change found in many of the event data models developed in the 1970s. The disadvantage of Δ_8CD is that the measure indicates only that some sort of change is going to take place; it tells nothing about *what* that change will be. Furthermore, the phases determined by Δ_8CD do not always correspond to the overt military-political changes that one might wish to forecast with an early-warning system.

This is most conspicuously the case for Lebanon in 1981-82: According to the Δ_8CD measure, the system shifted into the "Lebanon" phase about a year before the actual invasion in June 1982. When the invasion occurs, the Δ_8CD measure is at one of the lowest points seen in the entire time series. On the one hand, the policies that culminated in the invasion of Lebanon were put into effect

well before the invasion and placing the true phase change in mid-1981 is politically plausible.[15] On the other hand, the situation on the ground looked very different in July 1982 than in April 1982, during which period the Δ_8CD measure was plummeting. Δ_8CD is clearly not a "barometric" early warning indicator that allows a political analyst to say to his boss, "The Δ_8CD is real low this month, ma'am: nothing to worry about ..." This may be because Δ_8CD is based on a correlation distance and is sensitive to changes in the configurations of policies—who is coordinating policy with whom—rather than to the direction of change. Δ_8CD as an early warning indicator in combination with a Euclidean measure sensitive to the direction of change might provide both types of information.

Comparison with a Null Model

The results reported above generally support the phase model, but the measures are somewhat *ad hoc* and could easily be due to some combination of chance and ocular self-deception. In this section, therefore, we develop a null model and look at the distribution of various indicators in simulated data generated by that model.

The null model that we use preserves the sample size (192) and number of dyads (54) found in the data set analyzed in Schrodt and Gerner (1997), as well as the mean, variance, and first-order autocorrelation of the data within each dyad.[16] Specifically, we generated simulated data using an AR[1] process

$$y_t = c + \varphi y_{t-1} + \varepsilon$$

---

[15] It is widely accepted that Israel planned the invasion of Lebanon as much as a year in advance, then engaged in provocative military maneuvers throughout the first half of 1982 in an effort to goad the Palestinians into a response that would justify Israeli military intervention (Hiro 1992, Jansen 1982, Petran 1987, Randel 1983).

[16] This analysis—and the analyses of the category weights—were done in April and May 1996, before we generated the new data set. Both analyses are quite time consuming and thus we have not re-done them with the new data set; there is no reason to believe that the results would be any different using the newer data as a basis.

where $c = \mu (1-\rho)$; $\varphi = \rho$; $E(\varepsilon)=0$; $Var(\varepsilon) = s^2 (1-\rho^2)$. As Hamilton (1994:53-54) notes, this will produce a time series with mean μ, variance s^2, and first-order autocorrelation ρ. In order to avoid initial value effects, the simulated data were taken from the interval $[y_{51}, y_{242}]$ with $y_0=\mu$. A sample of 1000 such data sets were generated.[17]

This specification represents a compromise between a null model that is excessively random and one that essentially duplicates the data set. For example, in a null model using white noise (no autocorrelation), points generated by the 54 dyads would jump around in the vector space far more than one would ever expect to see in event data based on actual political behavior and presumably would show only very small clusters. On the other hand, if we also duplicated the cross-correlation between dyads, the simulated data set would have most of the statistical characteristics of the actual data and it would not be surprising if we found similar results. Our choice is an intermediate model, where the simulated time series have generally the same dyadic characteristics but have no relationship to each other.[18]

In comparing the simulated data with the actual data, we looked at the following measures:

- The total number of points where $LML_t>\Delta$, with $\Delta=0.2$.[19]
- The number of $LML_t>\Delta$ points that signal a new cluster: this is defined (somewhat arbitrarily) as an $LML_t>\Delta$ point that had no $LML_t>\Delta$ points in the previous two time periods.[20] These times are called "cluster-defining points."

---

[17] To save computation time, ε were generated by random selection from a table of 5000 normally-distributed random variables produced by Excel 4.0.

[18] Autocorrelation above the first order is significant in only a small number of the dyads in the original data.

[19] $\Delta=0.20$ was the threshold that we found best delineated clusters in the Schrodt and Gerner (1997) data set.

[20] In other words, this definition ignores the strings of consecutive $LML_t>\Delta$ points that are generated by rapid movements away from an existing cluster; these are quite common in the simulated data and are seen in the actual data in the Lebanon and Oslo transitions. This measure should also be less sensitive to the level of Δ.

- The standard deviation of LML_t and the early warning measure Δ_8CD; the means of both measures are zero.
- The number of Δ_8CD measures greater than one standard deviation above Mean(Δ_8CD) at 0, 1 ,2 and 3 "months" prior to a cluster-defining point.
- The number of $LML_t > \Delta$ points within 0, 1 ,2 and 3 months of the six *a priori* cluster transitions we identified in our data set, as a proportion of the total number of $LML_t > \Delta$ points.[21]

Because the Δ_8CD measure can only be computed after twelve months of data are available, and computing the LML_t requires an three additional months, the interval on which these measures were computed contains 192 - 11 - 3 = 178 points.

The results of the Monte-Carlo analysis are presented in Table 2, where the "one-tailed probability" indicates the proportion of the values in the simulated data that are less than (<) or greater than (>) the observed value. The distribution of the values of the statistics are generally smooth, symmetrical, and look more or less Normally distributed;[22] the probabilities are based on the actual distributions of the statistics in the simulated data rather than on a Normal approximation.

---

[21] In the simulated data, these *a priori* transitions are arbitrary—they do not correspond to conspicuous features in the data—but these indicators measure the likelihood of finding $LML_t > \Delta$ points in the vicinity of a set of six transition points spaced at the intervals in the *a priori* set. We look at the proportion because the number of $LML_t > \Delta$ points in the simulated data is substantially higher than in the actual data.

[22] Histograms of these distributions are available from the authors. The exception to the pattern of quasi-normal distributions is the $LML_t > \Delta$ /*a priori* measure at k=0 and k=1: It is bounded at zero and has a small mean and thus is skewed to the left. Schrodt and Gerner (1996:18) shows some examples of the LML and Δ8CD curves that are produced by the simulated data.

Table 2
Statistics Computed from 1000 Simulated Data Sets, Δ=0.2

Statistics for Δ=0.2 (N=1000)		Simulated mean	Simulated standard dev	Observed value	One-tailed probability
Total $LML_t>\Delta$		31.55	5.67	15	0.003 (<)
Cluster-defining $LML_t>\Delta$		15.63	2.61	9	0.006 (<)
Stdev of Δ_8CD		0.30	0.04	0.23	0.026 (<)
StDev of LML		0.25	0.03	0.15	0.001 (<)
CDL at t and Δ_8CD_{t-k}>Stdev	k=0	0.41	0.11	0.56	0.090 (>)
	k=1	0.22	0.10	0.22	0.461 (>)
	k=2	0.21	0.09	0.11	0.893 (>)
	k=3	0.20	0.09	0.11	0.869 (>)
$LML_t>\Delta$ within t±k of *a priori* break	k=0	0.03	0.03	0.07	0.136 (>)
	k=1	0.10	0.06	0.27	0.011 (>)
	k=2	0.17	0.08	0.40	0.006 (>)
	k=3	0.23	0.09	0.47	0.008 (>)

With the exception of one set of statistics—the relationship between Δ_8CD and the cluster-defining points—the values observed in the actual data are substantially different than those found in the simulated data, and vary in the expected direction. The number of $LML_t>\Delta$ points found in the actual data—whether total or cluster-defining—is about half that found in the simulated data. The standard deviations of the LML and Δ_8CD measure are substantially less in the observed data than in the simulated data. Generally, an $LML_t>\Delta$ point is about twice as likely to occur near one of the *a priori* cluster breaks in the actual data than in the simulated data.

The relationship between Δ_8CD and the cluster-defining points is somewhat puzzling. The observed k=0 point is significantly greater (at the 0.1 level) than the simulated values, as we expected. The k=1 value, however, is simply equal to the mean, and the k=2 and k=3 values are actually significantly *less* than the simulated data at the 0.15 level. This suggests that on average Δ_8CD_{t-k} may actually be a better early warning indicator than demonstrated in this data set, although its performance is due to autocorrelation in the data rather than to any complex characteristics involving dyadic interactions.

The large number of $LML_t>\Delta$ points combined with standard deviations of LML and Δ_8CD that are higher in the simulated data than in the observed data suggests that the value of Δ—a free parameter that was established arbitrarily—may have been set too low for the simulated data. We re-ran the simulated data sets with $\Delta=0.35$, a level of Δ that gives roughly the same number of cluster-defining points in the simulated data as were found in the observed data with $\Delta=0.2$. This adjustment of Δ effectively eliminates one additional degree of freedom in the simulated data; the results of this analysis are reported in Table 3.

This modification changes the one-tailed probabilities somewhat, but in general does not alter the conclusions of the analysis. The curious pattern of Δ_8CD and the cluster-defining points is retained—and actually strengthened at k=2 and k=3—except that the k=0 point is no longer significant. The relationship between the $LML_t>\Delta$ measures and the *a priori* breaks is slightly less strong, but the k>0 probabilities are still quite low. We conclude that the behavior of the predictive measures is not solely due to the difference in the number of $LML_t>\Delta$ points.

Table 3
Statistics Computed from 1000 Simulated Data Sets, $\Delta=0.35$

Statistics for $\Delta=0.35$ (N=1000)		Simulated mean	Simulated standard dev	Observed value	One-tailed probability
Total $LML_t>\Delta$		13.56	4.34	15	0.680 (<)
Cluster-defining $LML_t>\Delta$		8.48	2.49	9	0.660 (<)
Stdev of Δ_8CD		0.30	0.04	0.23	0.026 (<)
StDev LML		0.25	0.03	0.15	0.001 (<)
CDL at t and Δ_8CD_{t-k}>Stdev,	k=0	0.54	0.17	0.56	0.462 (>)
	k=1	0.31	0.16	0.22	0.731 (>)
	k=2	0.30	0.16	0.11	0.915 (>)
	k=3	0.28	0.15	0.11	0.903 (>)
$LML_t>\Delta$ within t±k of *a priori* break	k=0	0.03	0.06	0.07	0.247 (>)
	k=1	0.10	0.10	0.27	0.074 (>)
	k=2	0.16	0.13	0.40	0.054 (>)
	k=3	0.23	0.14	0.47	0.060 (>)

The results of the Monte-Carlo analysis are ambiguous due to the existence of the free parameter Δ. If we take as given the $\Delta=0.2$ separation threshold, then the observed data has far

fewer clusters than we would expect the null model to generate. By raising the level of Δ, we can match the number of empirically-determined clusters, although the behavior of the Δ_8CD statistic and the coincidence of LML>Δ points and the *a priori* points are still quite different in the simulated data. Furthermore, the necessity of raising the value of Δ to match the expected number of clusters means that the number of points where a large change occurs in LML_t is greater in the simulated data than in the observed data because the variance of LML is higher in the simulated data. This in turn would be expected if it were the case that the observed data actually settled into clusters and remained there for a period of time, rather than jumping around. We suspect that the standard deviation of LML_t is lower in the observed data because of cross-correlation (and in a few dyads, higher-order autocorrelation) of the dyads.

Optimizing the Weights

The preceding analysis has been done by aggregating the individual events using Goldstein's (1992) numerical weights for the WEIS categories. This aggregation has the advantage of converting the frequencies of the 63 WEIS categories into a single number, which in turn can be analyzed using well-understood interval-level statistical techniques such as correlation. It has the disadvantage that the Goldstein weights—which were determined by averaging "expert" judgments on the general character of the WEIS categories—may not be optimal for clustering and early warning.

We therefore attempted to estimate optimal weights using a genetic algorithm (described in Appendix A) that maximized the following clustering measure:

$$F^c = \frac{\text{average distance between adjacent clusters}}{\text{average distance within clusters}}$$

where "distance" is defined by the correlation metric and the "average distance" is calculated as the average distance between points. F^c is similar to the F-ratio maximized in discriminant analysis except that only the distance between adjacent clusters is considered and the measure uses the total distance between points rather than group variances. The cluster boundaries are dependent on the

value of Δ in the LML criterion: Higher values of Δ consistently produce higher F^c values, but fewer clusters, because of the stricter threshold for establishing a new cluster.

Figure 6 shows the results of several experiments with the genetic algorithm (GA). The thick lines above the X-axis are a histogram of the cluster boundaries identified by 100 GA runs for Δ=0.15; F^c in these runs ranged from 1.70 to 1.51 compared to F^c = 1.08 for the Goldstein weights.[23] The solid squares on or near some of these lines show the number of cluster boundaries found for 25 GA runs with Δ=0.25. The solid triangles just below the X-axis show the location of the cluster boundaries generated by the Goldstein weights for Δ=0.15, and the dotted vertical lines beneath the X-axis show the *a priori* cluster boundaries from Table 1. For comparison, the final four rows of symbols show the divisions identified by four "default" weighting systems.

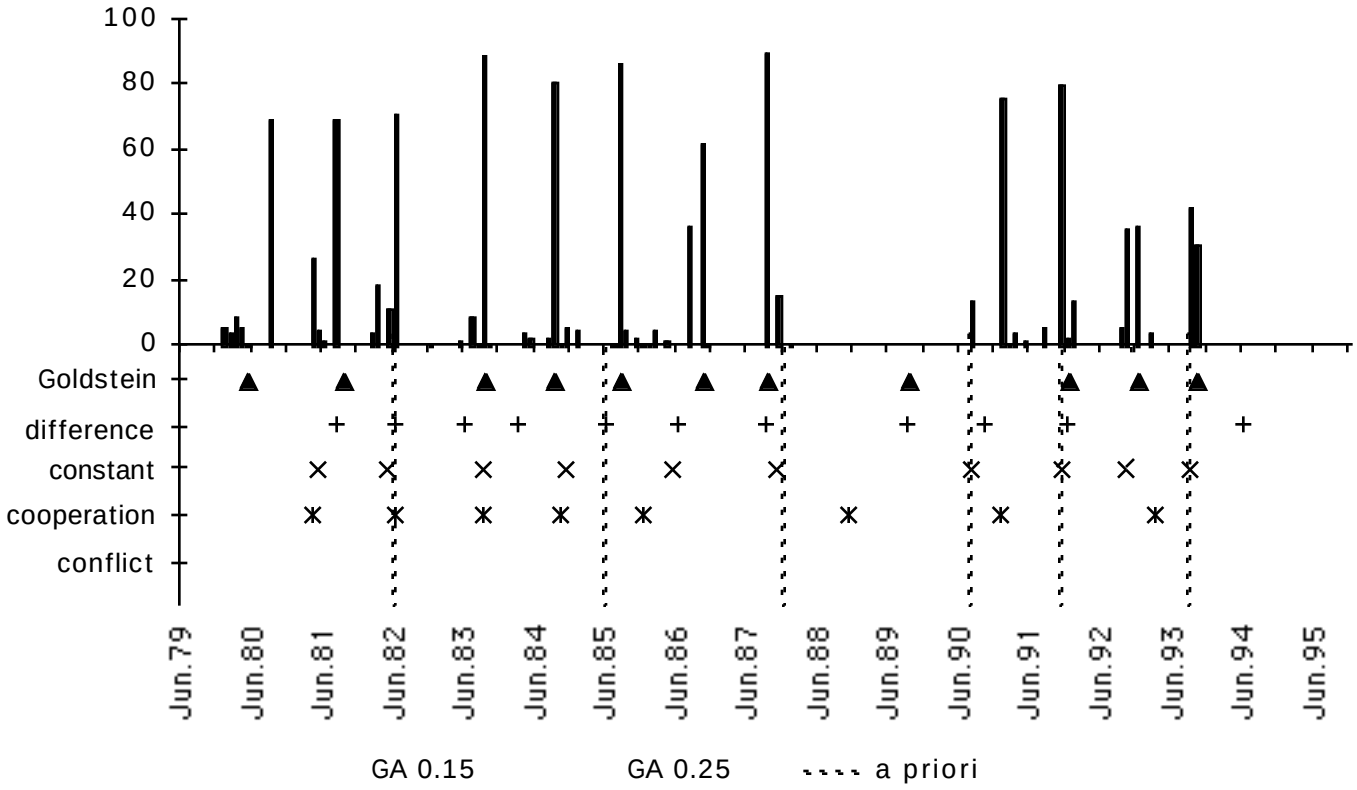
When we set Δ=0.15, the GA finds all of the *a priori* boundaries except the Kuwait transition; this is also true for the periods prior to the *intifada* when Δ=0.25. Most of the remaining boundaries are similar to those found when we used the Goldstein weights (Figure 3) with three exceptions:

- The GA identifies the Lebanon transition in June 1982 rather than earlier in the year;
- The GA finds a January 1991 transition coinciding with the onset of U.S.-led military activity against Iraq;
- The GA does not find an early end to the *intifada*; in fact, none of the 100 experiments found more than one cluster between the beginning of the *intifada* and Iraq's invasion of Kuwait.

---

[23] For Δ=0.25 the range of F^c for 25 GA experiments is 2.35 to 1.90; for the Goldstein scale F^c = 1.30.

Figure 6. Cluster boundaries under various weighting systems



Alternative weight vectors:

Goldstein: Goldstein weights averaged within each two-digit WEIS category. $F^C = 1.08$

difference: cooperative events = 1; conflictual events = -1. $F^C = 1.18$

constant: all events = 1. $F^C = 1.46$

cooperation: cooperative event = 1; conflictual events = 0. $F^C = 1.26$

conflict: cooperative event = 0; conflictual events = 1. $F^C = 1.29$

(cooperative events are WEIS 01 through10; conflictual events are WEIS 11 through 22.)

GA Δ=0.25 count has been multiplied by 4

Despite the similarities in the cluster boundaries determined by the various GA experiments, there were no consistent patterns in the values of the weights. The distribution of the 4950 correlations between the weight vectors is generally Normal with a mean of 0.077 and standard deviation of 0.307; the number of significant correlations do not exceed the number expected by chance. The average correlation of the GA weights with the Goldstein weights is -0.008 with a standard deviation of 0.188, and there is virtually no difference in the average correlation of the best-fitting 50% of the vectors (-0.006) and the other 50% (-0.009).

Examination of the intermediate results produced by the GA showed that despite this diversity, the GA is working correctly to generate and select vectors that increase the value of F^c; there are simply a lot of different ways to do this. The lack of convergence of the weight vectors is probably due to an effect comparable to collinearity in a linear model: Because the correlation distance is invariant with respect to a linear transformation of the weights, very different sets of weights can be used to produce essentially the same distances. Consequently, producing an "ideal" set of weights via an estimation procedure, genetic or otherwise, is probably a hopeless task.

The cluster divisions that are produced by the GA generally correspond well to those produced by the Goldstein weights. However, the constant weight vector—where all event types are weighted equally—also produces divisions similar to those found by the other methods; the "difference" vector does as well in matching the *a priori* transitions but (unsurprisingly) corresponds more closely to the Goldstein divisions than does the constant vector. The vector with the least correspondence to the other transitions considers only the cooperative events; this has almost no correspondence with the *a priori* transitions.

We also used a genetic algorithm to maximize the separation of the clusters, given the *a priori* cluster boundaries in Table 1. The purpose of this experiment was to see whether it would be

possible to find a better set of weights than those provided by Goldstein for that set of cluster transitions. Twenty experiments were done, with the GA allowed to run for 128 generations.[24]

The results of this experiment differed substantially from the experiments where both the weights and cluster boundaries are allowed to vary. In particular, there is a significant correlation at the 0.01 level between the weight vectors in about 30% of the cases. None of those weight vectors, however, correlate significantly with the vector of Goldstein weights.

When the LML and ΔBCD curves produced by these new vectors are compared with the curves generated by the default vectors, an even stronger pattern emerges. First, virtually all of the correlations between the curves are significant. The strongest correlations—usually in excess of 0.95 for LML and 0.85 for ΔBCD—were with the curves produced by the constant vector (equal weighting of each category).[25] The second strongest correlations were for the vector that counted

---

[24] We also tried to do this in a linear fashion using discriminant analysis, in which the independent variables were the WEIS event counts (by 2-digit category) totaled across all of the dyads by month. While this technique removes the information on which dyads were interacting (to differentiate the dyads and event categories would involve 1188 variables and we have only 208 data points) we thought it might provide a rough estimate for the event weightings appropriate for the full system.

The results were generally disappointing. The classification accuracy with all variables was only 73% and the first three discriminant functions explained only 75% of the variance. There was no discernible pattern to the weights or functions. With stepwise discriminant the accuracy dropped to 60% but the variables chosen tended to be those with a high density of events: 01 (Yield), 02 (Comment), 03 (Meet), 11 (Reject), 12 (Accuse), 13 (Protest), 21 (Seize) and 22 (Force). The discriminant weights mirrored those of the Goldstein scale to some extent: 01 = 0.65, 02 = 0.70, 03 = 0.23, 11 = -0.48, 12 = -0.09, 13 = 0.22, 21 = -0.32 and 22 = -0.83.

[25] The variance of the weight in many of the optimized vectors is quite small: Half have a standard deviation between 0.5 and 1.0; the remainder have a standard deviation between 5.0 and 7.0 (the Goldstein weights have a standard deviation of 4.34). This bi-modal distribution is entirely a function of whether or not the vector contains both positive and negative weights: the low-variance vectors have only positive weights.

only cooperative events—these were around 0.77 for LML and 0.72 for Δ_8CD. The value of F^c for the optimized vectors is 1.58 while the F^c value for the constant vector is 1.54, so the optimization provides very little improvement.

The high correlation between these optimal vectors and the constant vector is consistent with another experiment we report in Schrodt and Gerner (1996): Computing the distance between points by correlating the frequencies of the 2-digit WEIS events without applying *any* weighting. In general, this measure produced results quite similar to those of the Goldstein measure,[26] particularly in terms of matching the *a priori* cluster boundaries. This analysis again suggests that the clusters are not strongly dependent on the Goldstein weights, and the frequency of coded events alone is the primary factor that the major political features of the data. However, the variance of the Goldstein results is slightly higher than that of the event count results in the LML measure, and noticeably higher in the Δ_8CD measure. This would suggest that the Goldstein weights, despite their somewhat *ad hoc* development and independence of the clustering scheme developed here, are correctly "tuned" to provide sensitive indications of the political changes in which human analysts are likely to be interested.

We draw two general conclusions from this analysis. First, it implies that most of the information being used to differentiate the clusters is found in the event counts themselves, rather than the weighting of events. This could be due to at least two factors. First, about 50% of the dyad-months in the data set have zero values, which are unaffected by any change in the weighting scheme. Second, the existence of any activity in a dyad may be a signal that Reuters reporters or editors think that the activity is important: this is particularly true with respect to verbal activities where Reuters has more of an option of reporting or not reporting activity.[27]

---

[26] The correlation (r) between the Goldstein-weighted and event count LML is 0.63. The correlation between the two Δ_8CD measures is 0.62.

[27] In an experiment, we computed the LML curves for a data set that replaced all event counts that were greater than zero with a value of 1—in other words, the data measured only the presence of events rather than their quantity.

This lack of sensitivity to event weights has an important implication for the use of machine-coded data for forecasting purposes. While machine-coding is more consistent over time than human-coded data, machine-coding is less sensitive to nuances of reported political behavior, and it is possible that those nuances could be very important in a problem such as forecasting. This analysis, however, does not support that conclusion: Because similar results can be obtained from huge differences in the weighting of event categories, there is little evidence that subtle differences in the coding of events would have a major difference on the ability to forecast. Furthermore, machine-coding is very unlikely to make errors in creating an event that is completely unrelated to a dyad.[28]

The forecasting measure Δ_8CD may be shifting primarily due to changes in the importance that various Reuters reporters and editors assign to events. If those reporters anticipate that a political shift is forthcoming in a region, they are likely to devote more coverage to it. In other words, Δ_8CD may actually be an indirect measure of a large number of events that are known by the Reuters organization but not necessarily reflected in the events reported in lead sentences coded by KEDS.

This produces credible cluster breaks (for example it correlates at the 0.78 level with the LML curve from the true data set when the constant vector is used); unsurprisingly, however, the variation of the curve is attenuated compared to that produced by the actual data.

[28] In machine coding, the most common actor-assignment error is confusing the object of an action with an indirect object or a location. Machine coding will not, however, create a actor that is not mentioned in the text. For instance, if a series of events involves Israel, Syria, Lebanon, and the Palestinians, some actions of Israel towards Syria might be incorrectly coded as directed to Lebanon or the Palestinians. However, machine-coding would never create an extraneous Egypt-Jordan event from these texts. Because our forecasting model assumes clusters of activities, it will generally be insensitive to a few incorrect assignments of targets.

Conclusion

With the end of the DARPA early warning research in the early-1980s, the development of quantitative early warning models went into eclipse. This is understandable: When evaluated against what is practical today, the DARPA efforts were necessarily primitive in their dependence on time-consuming and unreliable human coding and computers having only a tiny fraction of the speed and memory available in a contemporary PC. The event-based quantitative forecasting efforts of the late 1970s failed, but then 1970s video games weren't very impressive either.

We draw two general conclusions from our analysis using time-delimited clusters. First, our empirical results continue to support analyzing phases of political behavior by looking at the movement of a point defined by the vector of dyadic interactions. The pattern of variation in LML_t seen in Figure 3 is exactly what we expected the phase transition model to generate: brief periods of large movement followed by long periods of little movement. In addition, the Monte Carlo analysis shows that this pattern is unlikely to occur by chance. Randomly-generated data having the same means, variances, and autocorrelations as our Middle East dyads show a greater amount of variation in the change of distance than we find in the actual data.

Second, the time-delineated clusters are *much* cleaner and consistent than the clusters determined by the cross-sectional K-Means technique, while still preserving most of the *a priori* clusters we expected to find. The $LML_t > \Delta$ method used to delineate the clusters is conceptually simple and computationally efficient; in fact, the algorithm is sufficiently simple that it may be possible to determine analytically some of its statistical properties. The $\Delta_\theta CD$ measure also appears promising as the basis of an early-warning indicator.

Table 5 summarizes the empirically determined clusters in Levantine political behavior for the period that we have studied. For the most part, these divisions correspond to our *a priori* clusters, and the remaining differences are plausible. The LML cluster analysis identifies two phases that we did not: the increase in tension between Israel and the PLO prior to the Lebanon invasion and a pre-Taba period corresponding to the Israeli withdrawal from the area around Beirut that is distinct from the initial period of the invasion. The $\Delta_\theta CD$ measure—although not the LML cluster analysis—

indicates significant changes following the Oslo peace process phases. Based on Δ_8CD, we might also have designated a post-*intifada*, pre-Madrid cluster beginning in late 1989. All of our analyses using the Goldstein weights missed the Kuwait transition, although the weights found by the genetic algorithm and the default vectors usually detected either the invasion or the subsequent war.

Table 5.

Clusters Determined by the Analysis

Initial date of cluster	Political characteristics	*a priori* cluster?	LML cluster Δ>.30	nearest Δ_8CD peak
July 1979	Camp David; pre-Lebanon	yes	NA	NA
December 1981	Increase in Israeli activity against PLO in Lebanon prior to the June 1982 invasion	no	yes	Oct-81
June 1982	Israeli invasion of Lebanon	yes	yes	Oct-81
September 1983	Period of Israeli withdrawal from Lebanon; increased Shi'a attacks against Israeli and international forces	no	yes	Apr-83
August 1985	Israel withdraws to south of Litani; Taba negotiations	yes(1)	yes	Apr-85
November 1987	Palestinian *intifada*	yes	yes	Sep-87
August 1990	Kuwait invasion	yes	no	Oct-89(2)
December 1992	Madrid peace process	yes(3)	yes	Dec-91
November 1993	Oslo peace process	yes	yes	Oct-93
January 1995	Post-Oslo period	no	no(4)	Nov-94

Table Notes:

(1) The *a priori* cluster break was two months earlier, in June 1985
(2) This Δ_8CD score probably corresponds to the end of the *intifada* and the Syrian consolidation of power in Lebanon rather than a forecast of the Kuwait invasion
(3) The *a priori* cluster break was almost a year earlier, in November 1991
(4) This cluster is based only on the two peaks post-Oslo peaks in the Δ_8CD score

The Δ_8CD measure usually provides two to six months of early warning; however, it gives no signal of the Oslo transition, and no distinct alert before the June 1982 invasion of Lebanon. The Δ_8CD measure also has some false-positives where it peaks just below the critical level. This is to be expected: Any measure that does not contain false positives is probably insufficiently sensitive to political events. We are not dealing with a deterministic system, and at times a false positive may reflect pre-cursors to transitions that failed to occur because of a reaction in the international system that prevented the phase change.[29]

The focus in the final part of this paper was on optimizing the weights of the individual event categories. This was not particularly successful and, in general, event counts were more important than weights. We suspect that this is because much of the necessary weighting has already been done for us by Reuters. If the reporters and editors of Reuters are good intuitive political analysts—and there is little reason to assume otherwise, particularly for this intensely covered region—then the frequency of reported events in important situations that may be undergoing political change will be higher than reports on situations that are relatively static. From a "god's eye view," this is sloppy and introduces an additional possible source of error. But we aren't gods; we are event data analysts and we can only study what is available in Reuters or comparable sources. This is not to make a virtue of the necessity of relying on Reuters, but simply an observation that Reuters' filtering seems to be doing a reasonable job for purposes of forecasting.

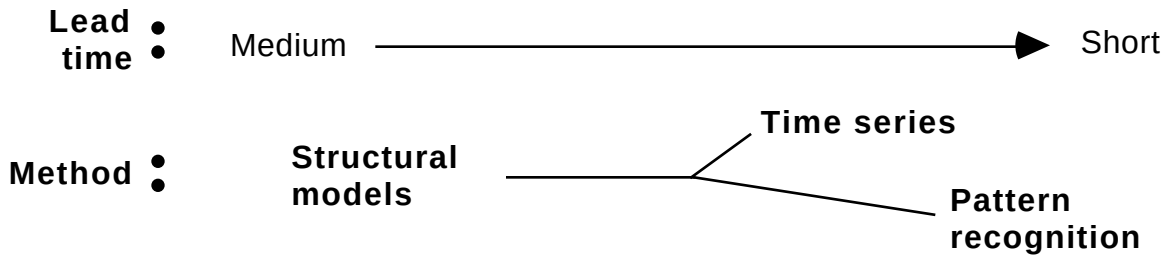
Time-delimited clusters are a dynamic rather than a structural early warning approach, but its effectiveness should not be regarded as evidence against the structural approach: We regard these as complementary rather than competitive. Structural methods are particularly good for mid-level warning: telling analysts where to look for potential trouble. Structural methods are also more

---

[29] The pre-Lebanon peak in LML may be such an example. In 1981, allies of Israel may have persuaded Menachim Began that an Israeli invasion of Lebanon would result in eventual Syrian hegemony in Lebanon, the development of militant Islamic fundamentalist movement on Israel's northern border, and completely destroy Begin's political future. Only after another year did the contrary advice of Ariel Sharon prevail.

likely to provide theoretical guidance about why a system is likely to experience problems, which might provide insights as to the types of actions that could be taken to ameliorate an impending crisis. Structural models are unlikely to excel at predicting the exact timing of breakdowns, however; the variables that they have identified as theoretically important change too slowly. This is where dynamic models come into play. The relationship of the approaches, then, may be something like the situation illustrated in Figure 7.

Figure 7. Time horizon for early warning methods.



In this analysis we have not considered an alternative class of dynamic models—those based on event sequences, rules, patterns, and precedents (see Cimbala 1987; Hudson 1991). These are likely to provide a greater amount of contextual information than supplied by the numerical time-series methods, and as a consequence may be useful in identifying the immediate events leading to a crisis. For instance, while the Kuwait transition is invisible in our cluster analysis, the events preceding Iraq's invasion of Kuwait follow Lebow's (1981) "Justification of Hostility" crisis-type very closely; such patterns could be used for very short-term forecasting. An assortment of computationally-intensive non-linear forecasting techniques methods have also been developed in recent years (e.g. Casdagli and Eubank 1992), though relatively little attention has been paid to these in the quantitative international politics literature. In short, there are still a variety of unexplored methods that could be applied to the early warning problem.

We suspect that the ideal early warning model would combine elements of both the structural and dynamic approaches: It should be possible to refine dynamic early warning models based on different categories of structural precursors. Presumably the internal breakdown in a Lebanon—which is relatively wealthy and highly differentiated by religion—occurs in a different fashion than a breakdown in Rwanda, which is relatively poor and not differentiated by religion. The reason that such integrated models have not been developed to date is largely one of resources: the political science discipline is still in the process of developing accurate structural and dynamic models, and at present no researcher has been able to assemble data sets sufficiently large to study both simultaneously. As the research on both types of models identifies more focused sets of variables and techniques, it should be practical to combine the approaches.

Appendix A: A Genetic Algorithm for Estimating Optimal Event Weights

Genetic algorithms (Goldberg 1989; Grefenstette 1987; Holland 1975) are a general purpose optimization method that is particularly effective in situations with are a large number of local maxima. Because our clustering algorithm determines the cluster boundaries as well as determining the weights, the problem is non-linear and required the use of a numerical optimization method rather than an analytical optimization method such as discriminant analysis.

A cluster break was any point that met two conditions:

1. $LML_t > \Delta$

2. No cluster boundaries in the previous 8 months (i.e., minimum cluster size of 8 months)

We experimented with several values of Δ in the range Δ=0.15 to Δ=0.30. The number of clusters found is inversely proportional to the value of Δ and the $LML_t \approx 0.20$ threshold is comparable to the level found to produce cluster boundaries corresponding to the *a priori* clusters when the Goldstein weights are used. A minimum cluster size is necessary because a sharp change in behavior will produce several consecutive months where LML_t is high.

The genetic algorithm is straightforward: The optimization operates on a vector of weights for the twenty-two WEIS 2-digit categories:

$$\mathbf{w} = [w_1,\ldots,w_{22}]$$

For a given set of weights, an aggregated monthly score is computed for each dyad

$$XY_t = \mathbf{w} \bullet \mathbf{c_t} = \sum_{i=1}^{22} w_i \; c_{it}$$

where c_{it} = number of events in WEIS 2-digit category i directed from X to Y in month t.

Once these scores are calculated, the LML measure is computed, the boundaries between clusters are determined using the $LML_t > \Delta$ threshold and minimum size rules discussed above, and the fitness measure

$$F^c = \frac{\text{average distance between adjacent clusters}}{\text{average distance within clusters}}$$

is computed with "distance" $\| x_i - x_j \|$ is defined by the correlation metric. The "average distance" is calculated as the average distance between points:

$$\text{Between cluster distance} = \frac{1}{N_1 N_2} \sum_{i \in C_1} \sum_{j \in C_2} \| x_i - x_j \|$$

$$\text{Within cluster distance} = \frac{2}{N_1(N_1-1)} \sum_{i \in C_1} \sum_{j>i} \| x_i - x_j \|$$

where N_i = number of points in cluster *i*. We measure the points in adjacent clusters rather than comparing the distance of a cluster to all other clusters is done to allow the possibility of the system returning to an equilibrium behavior, so that clusters that are separated in time might occupy the same space.

The genetic algorithm uses 32 **w** vectors that are initially set randomly to numbers between -10.0 and +10.0, the same range as the Goldstein weights. After the fitness of each vector is computed (a "generation" in the genetic algorithm), the vectors are sorted according to the value of F^c and the 16 vectors with the lowest fitness are replaced with new vectors created by recombination and mutation of the top 16 vectors.[30] The probability of a vector becoming a "parent" is proportional to the relative fitness of the vector (in other words, vectors with higher fitness are more likely to be used to produce new vectors). Mutation involves adding a random number between -1 and +1 to the weight, and mutation is done on 50% of the weights in the new vectors.

---

[30] One new vector was generated by taking the average weight of the top 16 vectors, on the logic that weights that were not important in the distance calculations (notably those for codes that occur infrequently in the data set) would go to zero as the random weights canceled out. These average vectors were tagged so that their survival in future generations could be tracked. These were rarely selected in the first set of experiments where both the weights and cluster breaks were allowed to vary, but were frequently selected in the second set of experiments where the weights were optimized for a given set of cluster divisions.

Most of the results we report are based on runs in which the system ran for 48 generations. This was usually sufficient to find a vector that showed little or no change. We also did a few runs where the system was allowed to run for 128 generations, which produced essentially the same results as the shorter runs.

This system was implemented in a C program; the source code is available from the authors. We originally ran the program on a 50 Mhz 68040 CPU (Macintosh Powerbook 520c) and it was fairly slow: each 48-generation experiment took about three hours to complete. The same source code compiled for a 80 Mhz PowerPC 601 processor (Macintosh 7100/80) was faster by a factor of 50 (!), reducing our computing time from days to hours. It is not entirely clear why the increase in speed is so dramatic, but possibly the RISC architecture of the 601 chip (and its substantially larger memory caches) is particularly well-suited to running a genetic algorithm. Whatever the reason, this speedup simplifies considerably the task of working with GAs.

Bibliography

Adelman, Howard, and Susanne Schmeidl. 1995. "Early Warning Models and Networking." International Studies Association, Chicago.

Aldenderfer, Mark S., and Roger K. Blashfield. 1984. *Cluster Analysis*. Newbury Park: Sage.

Alker, Hayward, Ted Robert Gurr, and Kumar Rupesinghe. 1995. "Conflict Early Warning Systems: An Initial Research Program." International Studies Association, Chicago.

Andriole, Stephen J., and Gerald W. Hopple. 1984. "The Rise and Fall of Events Data: From Basic Research to Applied Use in the U.S. Department of Defense." *International Interactions* 11:293-309.

Ashley, Richard K. 1980. *The Political Economy of War and Peace.* London: Frances Pinter.

Azar, Edward E. 1972. "Conflict escalation and conflict reduction in international crisis: Suez, 1956." *Journal of Conflict Resolution* 16,2:183-202.

Azar, Edward, R.D. McLaurin, Thomas Havener, Craig Murphy, Thomas Sloan, and Charles H. Wagner. 1977. "A System for Forecasting Strategic Crises: Findings and Speculations About Conflict in the Middle East." *International Interactions* 3:193-222.

Backer, Eric. 1995. *Computer-Assisted Reasoning in Cluster Analysis.* New York: Prentice-Hall.

Bailey, Kenneth D. 1994. *Typologies and Taxonomies: An Introduction to Classification Techniques.* Thousand Oaks, CA: Sage Publications.

Bloomfield, Lincoln P., and Amelia C. Leiss. 1969. *Controlling Small Wars.* New York: Knopf.

Bloomfield, Lincoln P., and Allen Moulton. 1989. *CASCON III: Computer-Aided System for Analysis of Local Conflicts.* Cambridge Mass.: MIT Center for International Studies.

Bloomfield, Lincoln P., and Allen Moulton. 1997. *Managing International Conflict.* New York: St. Martin's Press.

Bond, Doug, J. Craig Jenkins, Charles L. Taylor, and Kurt Schock. 1996. "Contours of Political Contention: Issues and Prospects for the Automated Development of Event Data." International Studies Association, San Diego.

Boutros-Ghali, Boutros. 1994. "Strengthening of the coordination of emergency humanitarian assistance of the United Nations." Secretary-General's Report to the General Assembly A/49/177, 2 September 1994.

Butterworth, Robert Lyle. 1976. *Managing Interstate Conflict,1945-74: Data with Synopses.* Pittsburgh: University of Pittsburgh University Center for International Studies.

Casdagli, Martin, and Stephen Eubank. 1992. *Nonlinear Modeling and Forecasting*. Reading, MA: Addison-Wesley.

Choucri, Nazli, and Thomas W. Robinson, eds. 1979. *Forecasting in International Relations: Theory, Methods, Problems, Prospects.* San Francisco: W.H. Freeman.

Cimbala, Stephen. 1987. *Artificial Intelligence and National Security.* Lexington, MA: Lexington Books.

Daly, Judith Ayres, and Stephen J. Andriole. 1980. "The Use of Events/Interaction Research by the Intelligence Community." *Policy Sciences* 12:215-236.

Dedring, Jürgen. 1994. "Early Warning and the United Nations." *Journal of Ethno-Development* 4:98-105.

Dixon, William J. 1986. "Reciprocity in United States-Soviet Relations: Multiple Symmetry or Issue Linkage." *American Journal of Political Science* 30:421-45.

Esty, Daniel C., Jack A. Goldstone, Ted R. Gurr, Pamela Surko, and Alan N. Unger. 1995. *State Failure Task Force Report.* McLean, VA: Science Applications International Corporation.

Everitt, Brian. 1980. *Cluster Analysis* (2nd ed.). New York: Wiley/Halsted.

Gerner, Deborah J., and Philip A. Schrodt. 1994. "Foreign Policy Interactions in the Middle East: An Initial Examination of Three Cases of Conflict." International Studies Association, Washington.

Gerner, Deborah J., and Philip A. Schrodt. forthcoming. "The Effects of Media Coverage on Crisis Assessment and Early Warning in the Middle East." In *Early Warning Methodologies*, ed. Susanne Schmeidl and Howard Adelman. York University: Centre for Refugee Studies.

Gerner, Deborah J., Philip A. Schrodt, Ronald A. Francisco, and Judith L. Weddle. 1994. "The Machine Coding of Events from Regional and International Sources." *International Studies Quarterly* 38:91-119.

Gochman, Charles S., and Alan Ned Sabrosky. 1990. *Prisoners of War?* Lexington, MA: Lexington Books.

Goldberg, David E. 1989. *Genetic Algorithms in Search, Optimizations and Machine Learning*. Reading, MA: Addison-Wesley.

Goldstein, Joshua S. 1992. "A Conflict-Cooperation Scale for WEIS Events Data." *Journal of Conflict Resolution* 36: 369-385.

Goldstein, Joshua S., and John R. Freeman. 1990. *Three-Way Street: Strategic Reciprocity in World Politics*. Chicago: University of Chicago Press.

Goldstein, Joshua S., and Jon C. Pevehouse. forthcoming. "Reciprocity, Bullying and International Cooperation: Time-Series Analysis of the Bosnia Conflict." American Political Science Review.

Grefenstette, John J., ed. 1987. *Genetic Algorithms and their Applications.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Gurr, Ted R. 1995. "The State Failure Project: Early Warning Research for International Policy Planning." International Studies Association, Chicago.

Gurr, Ted R., and Mark Irving Lichbach. 1986. "Forecasting Internal Conflict: A Competitive Evaluation of Empirical Theories." *Comparative Political Studies* 19:3-38.

Gurr, Ted R., and Barbara Harff. 1994. "Conceptual, Research and Policy Issues in Early Warning Research: An Overview." *Journal of Ethno-Development* 4:3-15.

Gurr, Ted R., and Barbara Harff. 1996. *Early Warning of Communal Conflict and Humanitarian Crisis.* Tokyo: United Nations University Press, Monograph Series on Governance and Conflict Resolution.

Hamilton, James D. 1989. "A new approach to the economic analysis of nonstationary time series and the business cycle." *Econometrica* 57,2:357-384.

Hamilton, James D. 1994. *Time Series Analysis*. Princeton: Princeton University Press.

Hiro, Dilip. 1992. *Lebanon: Fire and Embers.* New York: St. Martin's Press.

Holland, John H. 1975. *Adaptation in Natural and Artificial Systems*. Ann Arbor: University of Michigan Press.

Hopple, Gerald W., Stephen J. Andriole, and Amos Freedy, eds. 1984. *National Security Crisis Forecasting and Management.* Boulder: Westview.

Hopple, Gerald W. 1984. "Computer-Based Early-Warning: A Staircase Display Option for International Affairs Crisis Projection and Monitoring." pp. 47-84 in Gerald W. Hopple, Stephen J. Andriole, and Amos Freedy, eds. *National Security Crisis Forecasting and Management.* Boulder: Westview.

Hudson, Valerie, ed. 1991. *Artificial Intelligence and International Politics.* Boulder: Westview

Hughes, Barry B. 1993. *International Futures.* Boulder: Westview.

Huxtable, Phillip A., and Jon C. Pevehouse. 1996. "Potential Validity Problems in Events Data Collection." *International Studies Notes* 21,2: 8-19.

Jansen, Michael. 1982. *The Battle of Beirut.* Boston: South End Press.

Laurance, Edward J. 1990. "Events Data and Policy Analysis." *Policy Sciences* 23:111-132.

Lebovic, James H. 1994. "Before the Storm: Momentum and the Onset of the Gulf War." *International Studies Quarterly* 38: 447-474.

Lebow, Richard Ned. 1981. *Between Peace and War*. Baltimore: Johns Hopkins University Press.

Midlarsky, Manus I., ed. 1993. *Handbook of War Studies.* Ann Arbor: University of Michigan Press.

Mizuno, Jiro. 1995. "Humanitarian Early Warning System: Progress and Prospects." United Nations: Department of Humanitarian Affairs.

Petran, Tabitha. 1987. *The Struggle over Lebanon.* New York: Monthly Review Press.

Phillips, Warren R., and Richard Rimkunas. 1983. *Crisis Warning*. New York: Gordon and Breach.

Pierce, John R. 1980. *An Introduction to Information Theory*. New York: Dover.

Randel, Jonathan C. 1983. *Christian Warlords, Israeli Adventurers, and the War in Lebanon.* New York: Vintage Books.

Rupesinghe, Kumar, and Michiko Kuroda, eds. 1992. *Early Warning and Conflict Resolution.* London: MacMillan.

Schrodt, Philip A., Shannon G. Davis, and Judith L. Weddle. 1994. "Political Science: KEDS—A Program for the Machine Coding of Event Data." *Social Science Computer Review* 12,3: 561-588.

Schrodt, Philip A., and Deborah J. Gerner. 1994 . "Validity assessment of a machine-coded event data set for the Middle East, 1982-1992." *American Journal of Political Science* 38: 825-854.

Schrodt, Philip A., and Deborah J. Gerner. 1996. "Cluster-based Early Warning Indicators for Political Change in the Middle East, 1979-1996." American Political Science Association, San Francisco.

Schrodt, Philip A., and Deborah J. Gerner. 1997. "Empirical Indicators of Crisis Phase in the Middle East, 1982-1995." *Journal of Conflict Resolution* 41,4: 529-552.

Schrodt, Philip A., Philip A. Huxtable, and Deborah J. Gerner. 1996. "Events Data and the Analysis of Political Behavior: The Middle East and West Africa, 1979-1995." International Studies Association, San Diego.

Sherman, Frank L., and Laura Neack. 1993. "Imagining the Possibilities: The Prospects of Isolating the Genome of International Conflict from the SHERFACS Dataset." pp. 87-112. In *International Event-Data Developments: DDIR Phase II*, ed. Richard L. Merritt, Robert G. Muncaster, and Dina A. Zinnes. Ann Arbor: University of Michigan Press.

Singer, J. David, and Michael D. Wallace, eds. 1979. *To Augur Well: Early Warning Indicators in World Politics.* Beverly Hills: Sage.

Thom, Rene. 1975. *Structural Stability and Morphogenesis.* Reading, Mass.: Benjamin Publishers.

Tomlinson, Rodney G. 1993. "World Event/Interaction Survey (WEIS) Coding Manual." Manuscript, United States Naval Academy, Annapolis, MD.

van Wyk, Koos and Sarah Radloff. 1993. "Symmetry and Reciprocity in South Africa's Foreign Policy." *Journal of Conflict Resolution* 37:382-396.

Ward, Michael Don. 1982. "Cooperation and Conflict in Foreign Policy Behavior." *International Studies Quarterly* 26:87-126.

Ward, Michael D., ed. 1985. *Theories, Models and Simulations in International Relations.* Boulder: Westview Press.

Ward, Michael, and Sheen Rajmaira. 1992. "Reciprocity and Norms in U.S.-Soviet Foreign Policy." *Journal of Conflict Resolution* 36,2: 342-368.

Wayman, Frank W., and Paul F. Diehl, eds. 1994. *Reconstructing Realpolitik.* Ann Arbor: University of Michigan Press.

Wonnacott, Ronald J., and Wonnacott, Thomas H. 1979. *Econometrics* (2nd ed.). New York: Wiley.