Using Cluster Analysis to Derive Early Warning Indicators for Political Change in the Middle East, 1979-1996

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

Paper presented at the American Political Science Association
San Francisco, 28 August - 1 September 1996

This paper and the event data set used in the analysis have been posted to the
APSA Political Methodology Section World Wide Web site:
http://wizard.ucr.edu/polmeth/working_papers96/working_papers96_papers.html

The KEDS program, data sets and other information are available at the KEDS web site:
http://raven.ukans.cc.edu/~keds

ABSTRACT

This paper uses event data to develop an early warning model of major political change in the Levant for the period April 1979 to July 1996. Following a general review of statistical early warning research, the analysis focuses on the behavior of eight Middle Eastern actors—Egypt, Israel, Jordan, Lebanon, the Palestinians, Syria, the United States and USSR/Russia—using WEIS-coded event data generated from Reuters news service lead sentences with the KEDS machine-coding system.

The analysis extends earlier work (Schrodt and Gerner 1995) demonstrating that clusters of behavior identified by conventional statistical methods correspond well with changes in political behavior identified *a priori*. We employ a new clustering algorithm that uses the correlation between the dyadic behaviors at two points in time as a measure of distance, and identifies cluster breaks as those time points that are closer to later points than to preceding points. We also demonstrate that these data clusters begin to "stretch" prior to breaking apart; this characteristic can be used as an early-warning indicator. A Monte-Carlo analysis shows that the clustering and early warning measures perform very differently in simulated data sets having the same mean, variance, and autocorrelation as the observed data (but no cross-correlation) which reduces the likelihood that the observed clustering patterns are due to chance.

The initial analysis uses Goldstein's (1992) weighting system to aggregate the WEIS-coded data. In an attempt to improve on the Goldstein scale, we use a genetic algorithm to optimize the weighting of the WEIS event categories for the purpose of clustering. This does not prove very successful and only differentiates clusters in the first half of the data set, a result similar to one we obtained using the cross-sectional K-Means clustering procedure. Correlating the frequency of events in the twenty-two 2-digit WEIS categories, on the other hand, gives clustering and early warning results similar to those produced by the Goldstein scale. The paper concludes with some general remarks on the role of quantitative early warning and directions for further research.

This research was funded by the National Science Foundation through grant SBR-9410023 and the University of Kansas General Research Fund Grant 3500-X0-0038.

Introduction

In recent years, the topic of early warning—moribund for about a decade after substantial research in the late-1970s (see Hoople, Andriole & Freedy 1984; Singer & Wallace 1979; Choucri & Robinson 1979)—has received renewed attention in the international relations literature (see for example Gurr & Harff 1994; Gurr & Harff in press; Rupesinghe & Kuroda 1992). This increased interest is due to at least three factors.

First, following the end of the Cold War the international system appears to have become more vulnerable to sudden outbreaks of serious systematic violence, both international and inter-ethnic. Iraq's invasion of Kuwait, the conflict between Armenia and Azerbijan, the genocidal violence observed in Bosnia and Rwanda, and the violent internal conflicts in Somalia, Chechnya, Haiti, Algeria, and Liberia are all examples of this. The end of the Cold War "removed the lid" from long-simmering regional and ethnic disputes, most conspicuously Armenia-Azerbijan and the former Yugoslavia. The disappearance of communism as an ideological principle for organizing conflict appears to have stimulated lethal disputes organized along ethnic and religious lines, frequently augmented by economically-motivated gangsterism.

With the end of the perceived threat of Communist exploitation of ethnic divisions, the liberal-democratic military powers—the United States, Britain and France—are less inclined to intervene in local or regional disputes. The international community instead has increasingly relied on multilateral responses, including the recycling of Cold War organizations (NATO in the former Yugoslavia, and the United Nations generally), *ad hoc* initiatives (Iraq-Kuwait, Rwanda, Bosnia), and the use of existing non-military organizations in a peace-keeping role (ECOWAS in Liberia). This reliance on multilateral responses—which cannot depend on the threat or deployment of prompt, overwhelming force used by the Cold War powers—in turn enhances the attractiveness of early warning in two ways. First, there is general agreement (Cahill 1996; Crocker & Hampson 1996; Lund 1996; Schmeidl 1997) that smaller amounts of force— and ideally no use of military force, relying only on diplomacy backed by the threat of force or other international sanctions—are required to contain a conflict in its early stages.[1] Second, multilateral responses require substantially longer to orchestrate than the rapid responses of a superpower or Cold War alliance. This has lead to significant interest by international organizations in early warning (e.g. Boutros-Ghali 1992; Dedring 1994; Alker, Gurr and Rupesinghe 1995, Mizuno 1995)

Finally, changes in communications and computer technology have dramatically changed the amount and timeliness of the information available for use in early warning. Information relevant to political early warning is available from the commercial efforts of Reuters, Agence France Press (AFP) and other news agencies, and from the equally vast, if more specialized, networks of IGO and NGO fieldworkers. The Internet and data providers such as NEXIS provide a quantity of real-time information far exceeding that available to the CIA and KGB during most of the Cold War period.[2] Similarly, inexpensive desk-top computers now surpass in capacity most of the computers available to national intelligence agencies until the middle of the last decade. The Internet and the related text-based electronic communications of news organizations, IGOs and NGOs can be processed directly by these computers. Whether this massive quantity of information can be *effectively* processed is another issue—this is the crux of the early warning challenge—but a researcher working with public domain sources in the late 1990s has access to orders of magnitude more real-time information and data processing capability than he or she would have had available even a decade earlier.

---

[1] Nor is the use of military force in the later stages of a conflict necessarily a guarantee of success: the military intervention in Somalia arguably worsened, rather than improved, that situation; the same could be said for the US activities in Lebanon in 1983-84.

[2] This is particularly true when one focuses on strategic political intelligence to the exclusion of the tactical military intelligence provided by satellite imagery and the monitoring of electronic communication.

The purpose of this paper is to explore some approaches to early warning that take advantage of this new situation. The paper starts with a review of the statistical early warning problem and discusses several approaches that could be used. We then apply a quantitative approach—cluster analysis—to an event data set that we have generated by machine-coding the lead sentences of stories taken from the Reuters newswire. The regional focus of the study is the Levant—Egypt, Israel, Jordan, Lebanon, the Palestinians and Syria, plus the United States and USSR/Russia—for the period April 1979 to July 1996.

Statistical Approaches to Early Warning: A Review

This section will review past approaches to *statistical* early warning in order to motivate the clustering approach that we employ in our analysis. It will not consider the large literature on non-statistical (qualitative) approaches to forecasting.[3]

For the purposes of justifying our methodology, 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 a civil or international war) as a dependent variable 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 Ted Gurr and his associates, most recently in the "State Failure Project" [SFP] (Gurr 1995; Esty et al. 1995); Gurr & Lichbach (1986) and Gurr & Harff (in press) provide surveys of these methods more generally. In the field of international instability, the structural approach is illustrated by the work of Bruce Bueno de Mesquita and his associates, and more generally by the Correlates of War project; Wayman & Diehl (1994), Gochman & Sabrosky (1990) and Midlarsky (1993) provide general surveys. These approaches have tended to use standard multivariate linear regression models, although recently 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, the Early Warning and Monitoring System (EWAMS), developed with funding from the U.S. Defense Advanced Research Projects Agency (DARPA; see Hoople 1984; Laurance 1990), 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 a model that looks for behaviors measured with COPDAB event data that fall outside a range of "normal" interactions for the dyad.

Scholars justify the dynamic approach—which is at odds with most statistical modelling in political science in its use of 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 in an early warning indicator; in fact many are essentially fixed (e.g. ethnic and linguistic heterogeneity; historical frequency of conflict; natural resource base). Data on variables that are changing—for example 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.

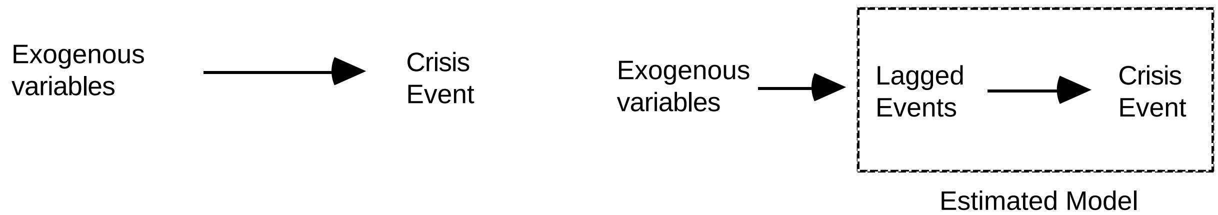
The second justification for the dynamic approach is that it reduces, and focuses, the information required by the model. The data collection effort of the SFP, for example, measures more than 50 independent variables (Gurr 1995:5-7), which requires a great deal of information

---

[3] Contemporary surveys of qualitative approaches can be found in Rupesinghe & Kuroda (1992), Gurr & Harff (1994), and Adelman & Schmeidl (1995); additional comments will be made in our conclusion about the relationship between quantitative and qualitative forecasting. The paper 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) provide surveys of that literature.

from a very wide variety of sources.[4] Event data collections, in contrast, focus on reported political interactions that can be systematically collected in real-time.

This convenience alone, however, cannot legitimately dictate the choice of dependent variable—such a decision would put one in the position of the proverbial drunk who searches for his lost keys beneath a streetlamp because that is where the light is best. 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. The dynamic approach effectively uses the lagged values of the events as a substitute for the structural variables. In other words,



Structural Modeling Approach **Dynamic Modeling Approach**

To take a concrete illustration, Gurr (1995: 7) notes "We think, for example, 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 of 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 different propensity for conflict than those with a low level; democracies are likely to be different than autocracies, and so forth. The difference 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 that to the extent that the variables are relevant for early warning problems, they can be measured indirectly through the patterns of events they generate.[5]

---

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

[5] 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, prices of raw materials, and macroeconomic trends. Technical analysis, in contrast, assumes that these factors will be reflected in the patterns of the movements of the price of a stock (or set of stocks) and therefore analysis of those prices alone will provide sufficient information for forecasting. Fundamental analysis corresponds to the structural modeling approach; 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

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 increase in democracy 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.

Because of the labor-intensive character of human event coding, the primitive statistical methods available at the time, and institutional factors (Daly & Andriole 1980; Andriole & Hoople 1984; Laurance 1990), the event-based early warning research was largely discontinued during the 1980s. Nonetheless, a small set of dynamic modeling efforts continued. These employed increasingly-advanced econometric time-series methods that modeled an interval-level measure of events as an autoregressive time series with disturbances. Goldstein & Freeman (1990) provide a book-length example of this approach; Ward (1982), Dixon (1986), Ward & Rajmaira (1992), Lebovic (1994) and Goldstein and Pevehouse (1996) illustrate the continued development of dynamic models of events. These studies generally used event data to explore political interactions rather than for forecasting but the techniques can be used retrospectively to determine the date at which a change occurred in the past.

Unfortunately, standard econometric time series methods have only limited utility in the problem of early warning.[6] The statistical problem of early warning is a subset of the general problem of time series analysis. In general, time series seeks to determine a function

$$y_{t+k} = f(y_t, y_{t-1}..., \mathbf{X}_t, \mathbf{X}_{t-1}...) \quad \text{for some } k>0$$

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 vector 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

$$y_t - y_s > \xi \ \forall \ t > T > s$$

for some indicator variable y. In English, this means that the variable y has substantially higher 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.[7]

An additional distinction is that econometric time series generally are 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 fully develop. Furthermore, most

---

"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 might have a similar effect on event data analysis.

6 "Unfortunately" because we would like to be able to use the extensive set of sophisticated theoretical models—to say nothing of the software—that econometricians have developed over the past half-century to analyze time series, rather than developing techniques *de novo*.

7 An early warning problem could also work on a $y_t - y_s < \xi$ situation

econometric time series are measured continuously rather than episodically, so missing data is less of an issue.

In contrast, the early warning problem 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 joke about the European political analyst who said "Every day from 1920 to 1970 I predicted that Europe would remain at peace when at peace, and at war when in war, and I was only wrong four times." This type of model is not, however, very useful.[8] The econometric problem most comparable to political early warning is forecasting sudden economic shifts such as those observed in exchange rate fluctuations (e.g., the collapse of the Mexican peso or the European Exchange Rate Mechanism).[9] These problems are similar to political early warning in the sense 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 & 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.

A Cluster-based Approach to Early Warning

In Schrodt & Gerner (1995) and Schrodt, Huxtable & Gerner (1996), 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 explicitly coded in data sets such as the Butterworth international dispute resolution dataset (Butterworth 1976), CASCON (Bloomfield & Moulton 1989) and SHERFACS (Sherman & Neack 1993).[10] 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 & Neack 1993:90)

---

8 More technically, such a measure succeeds according to a frequency-based measure but fails according to an *entropy*-based measure, which places higher weight on the prediction of low-probability events.

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

10 Sherman & Neack (1993) provide a review of the evolution of these data sets.

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".

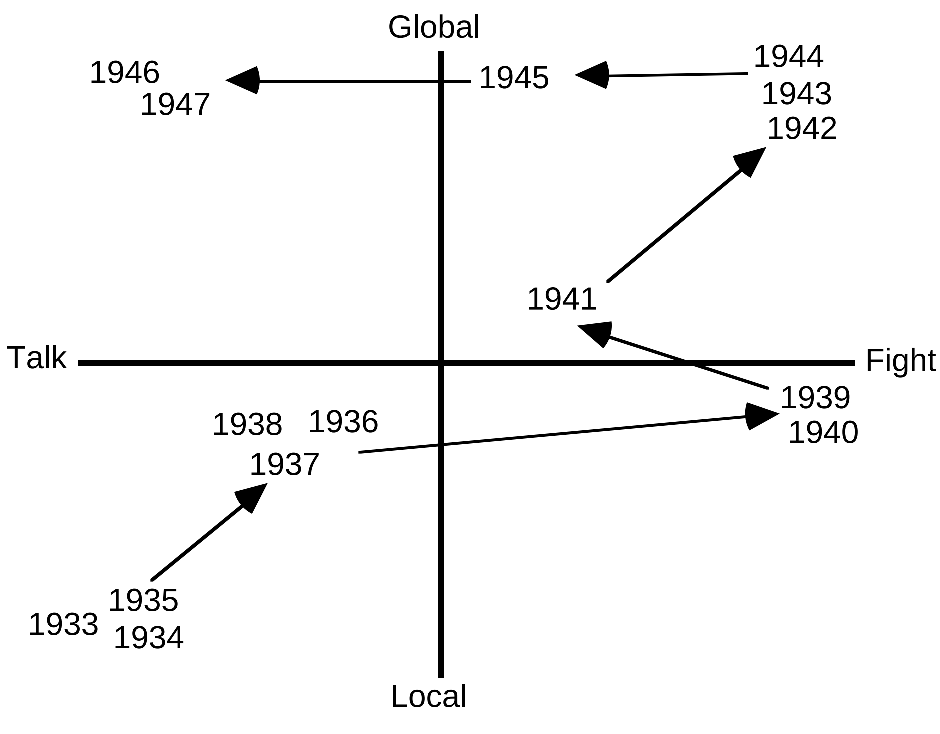
If the concept of crisis phase is valid, the behaviors observed in an international subsystem should fall into distinct patterns over time. If the transitions between these phases are gradual, or if behaviors that precede a phase transition are distinct from those found when the system is locked in a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

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

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.[11] 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 1
Schematic Representation of Phases during the WWII Period



Source: Schrodt & Gerner (1995)

---

11 In other words, we converted each X→Y event to its numerical score on the Goldstein scale, then totaled these numerical scores by month. Schrodt & Gerner (1994) gives a number of time series plots of the data for the 1982-1993 period. We have excluded the USA->USR and USR->USA dyads from our analysis since most of their interactions did not deal with the Middle East.

Figure 1, from Schrodt & Gerner (1995), illustrates this process informally for the World War II period, using the two dimensions of "talking versus fighting" and "local versus global 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 continues 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 so forth. Figure 1 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 short-answer-exam summary of international politics in the 1930s and 1940s presented above.

Nevertheless, if the behaviors characterizing a phase typology are captured by event data, it should be possible to determine those phases using clustering.[12] 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.

This assumption is relatively uncontroversial from the standpoint of actual political behavior: foreign policies are generally stable over periods of months and at times stable over periods of decades. The more difficult question is whether event data will pick up this consistency, since they are based on the *reports* of behavior rather than on the behaviors themselves. Because the international media report novel behaviors more often than than routine behaviors—in that time-worn phrase of journalists, "Dog bites man" is not news, but "Man bites dog" is news—routine behaviors that would, in principle, lead to a cluster may not be found in the news reports.

In the research we have done to date, however, this does not seem to be an insurmountable problem. In Schrodt, Huxtable & Gerner (1996), we looked at two different regional sub-systems: the Levant and West Africa. As expected, missing data was more of a problem the latter than in the former; nonetheless, we could detect most of the conspicuous changes in political behavior even in West Africa. A discriminant analysis of event data scores aggregated using the Goldstein (1992) scale classified monthly points into behavioral phrases that had been identified *a priori* with about 90% accuracy in the Levant, and about 75% accuracy in West Africa. This indicates that sufficient information is present in the event data to determine behavioral phases in a retrospective analysis.

The effectiveness of event-space clustering in early warning, in contrast, depends on whether some measurable characteristic of the behavior of the system changes prior to the phase transition. In some cases no precursors to a phase transition will be present, either because of deliberate 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,[13] journalists who are present in the region, understand the local politics, and can get their stories past editors and onto the newswires, are likely to report the behaviors they perceive to be pre-cursors to any political phase change. If the international media are not present, this information may be available from IGO and NGO fieldworkers. To summarize, the existence of pre-cursors in an

---

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

[13] Analysts within *organizations*—Cassandra of Troy, US foreign service officers in China in the 1940s, CIA analysts in Vietnam in the early 1960s—are not so fortunate...

event data set is dependent on the openness and inertia of the political process, and dependent on that process being reported in the sources used to generate event data, but usually the environment favors the detection of precursors.

The approach we are using to develop an early warning indicator is similar to the "normal relations range" concept proposed by Edward Azar (1972)

> 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.[14] Instead of exceeding a single critical threshold, we will assume that the system is moving away from normal behavior when it nears (or passes) the edge of the cluster. In addition, we look at the *density* of clusters—defined as the average distance between the points in a cluster—over time. Behavior in the NRR should result in dense clusters, whereas when a system moves away from one phase/cluster/NRR and into another, it will usually experience a period where the points do not cluster densely.

Our approach is in the spirit of the early event data projects—EWAMS and Azar's initial work—rather than the structural approach or the econometric time series approach. Like EWAMS and Azar, we are looking for a very general set of indicators, based on a single data source, that can be employed for dynamic early warning. This generally will come at the expense of relevance to specific theories of political behavior—the structural approach clearly excels in that domain—but has the pragmatic advantage of providing a method that does not require multiple data sources, careful tuning for specific dyads, or sophisticated estimation in order to be used.

Data

The data used in this study were machine-coded from Reuters lead sentences downloaded from the REUNA file of the NEXIS data service for the period April 1979 through July 1996; this generates about 80,000 events. We coded these data using the Kansas Event Data System (KEDS), a Macintosh program that generates event data from machine-readable reports; the program is described in Gerner et al. (1994), and Schrodt, Davis & Weddle (1995).[15] KEDS does

---

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

[15] 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! OR KUWAIT! OR IRAQ!)
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 & Gerner 1994; Schrodt & Gerner 1995), this data set was generated under the control of a "complexity filter" that did not code sentences if

- the sentence contained six or more verbs or
- no actor was found prior to the verb.

Sentences that met these criteria had a greater-than-average likelihood of being incorrectly coded by KEDS, thus by using the filter we should have a somewhat less noisy data.

In spot-checking some of the more densely reported dyads (e.g. ISR->PAL and ISR->LEB), we found that this new data set generally results in Goldstein scores that are smaller in magnitude. The bivariate regressions for these two dyads are 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. Schrodt & Gerner (1994), Huxtable & Pevehouse (1996) and Bond et al. (1996) 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 of 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.

Following the approach that we used in Schrodt & Gerner (1995), we assigned the following *a priori* phase identifications to various periods in the time series based on the dominant political interactions during the period. Our discussion of the results of the clustering and the early warning indicator will use these *a priori* clusters as a reference point.

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

Detection of Phase using Clustering over Time

In Schrodt & Gerner (1995), we analyzed a data set for phases using the SPSS K-Means agglomerative clustering algorithm and the Euclidean metric

$$\sum_{i=1}^{54}(x_i - y_i)^2$$

as the measure of the distance between points. The K-Means algorithm starts by finding K cases that are widely separated in the vector space; these are used as the initial cluster centers. It then assigns each of the remaining N-K cases to the cluster whose center is closest to the case. This technique was successful in identifying some of the phases that we had assigned *a priori* in the

---

ISR->PAL	G96 = 0.73 G95 - 2.75	r = 0.93 N = 192
ISR->LEB	G96 = 0.71 G95 - 0.66	r = 0.88 N = 192

where G96 are the Goldstein scores for the data set used in this paper and G95 are the scores for the data set used in Schrodt & Gerner (1995). The overall patterns in the series are generally very similar between the two data sets.

Both data sets, as well as the KEDS program (version 0.9B6.2) and the dictionaries used for this coding session, are available on disk from the authors or from the KEDS web site http://raven.cc.ukans.edu/~keds.

first half of the period but was less successful in the second half. That analysis also seemed to suggest that there was instability in the cluster assignment prior to a change in phase, but 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, in the conclusion of our earlier paper, we suggested several ways that clustering techniques might be modified to work with event data. Most importantly, K-Means does not use 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. While this could represent a return to an equilibrium, the cluster assignments jumped around a lot, which is inconsistent with equilibrium behavior. Furthermore, 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. We also suggested that metrics other than the Euclidean might be useful, and that examining the movement of points in the cluster space to see whether quantitative changes in the distance of a point from its cluster (rather than the change in the assignment of a point to a cluster) could be used for purposes of early warning. In this paper we implement all three of these suggestions.

The LML>Δ Clustering Algorithm for Time-Series

Using time as a dominant dimension actually simplifies the delineation of clusters in comparison to a cross-sectional clustering method such as K-Means. The clustering algorithm we employ 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.[16] Mathematically, a new cluster is considered to be established at a point x_t when

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

where $||x-y||$ is the distance between x and y according to some metric and Δ is the threshold parameter.

Figures 2 and 3 show the results of analyzing our Middle East data set using this algorithm for two metrics and k=4:

Euclidean metric $$||x - y|| = \sqrt{\sum_{i=1}^{54}(x_i - y_i)^2}$$

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 are located. We also experimented with some larger values of k and the results are much the same as those obtained with k=4.

From comparing the two figures, it can be seen if we set Δ = 0.30, the correlation metric picks four out of the six the *a priori* phase assignments and also identifies several other plausible transitions, some of which were also found by the K-Means analysis:

a. A pre-Lebanon change, probably reflecting increased tension between Israel and the PLO prior to the actual invasion;

---

[16] Lagged distance minus leading distance, hence "LML." Given the wide variety of extant clustering algorithms, this technique has undoubtedly been used by someone somewhere in the past, but at this point we've not done the appropriate literature search.

Calculations were done with a simple (600-line) Pascal program; the source code for this is available from the authors. The program produces various tab-delimited files that are read into Excel to produce the figures and tables.

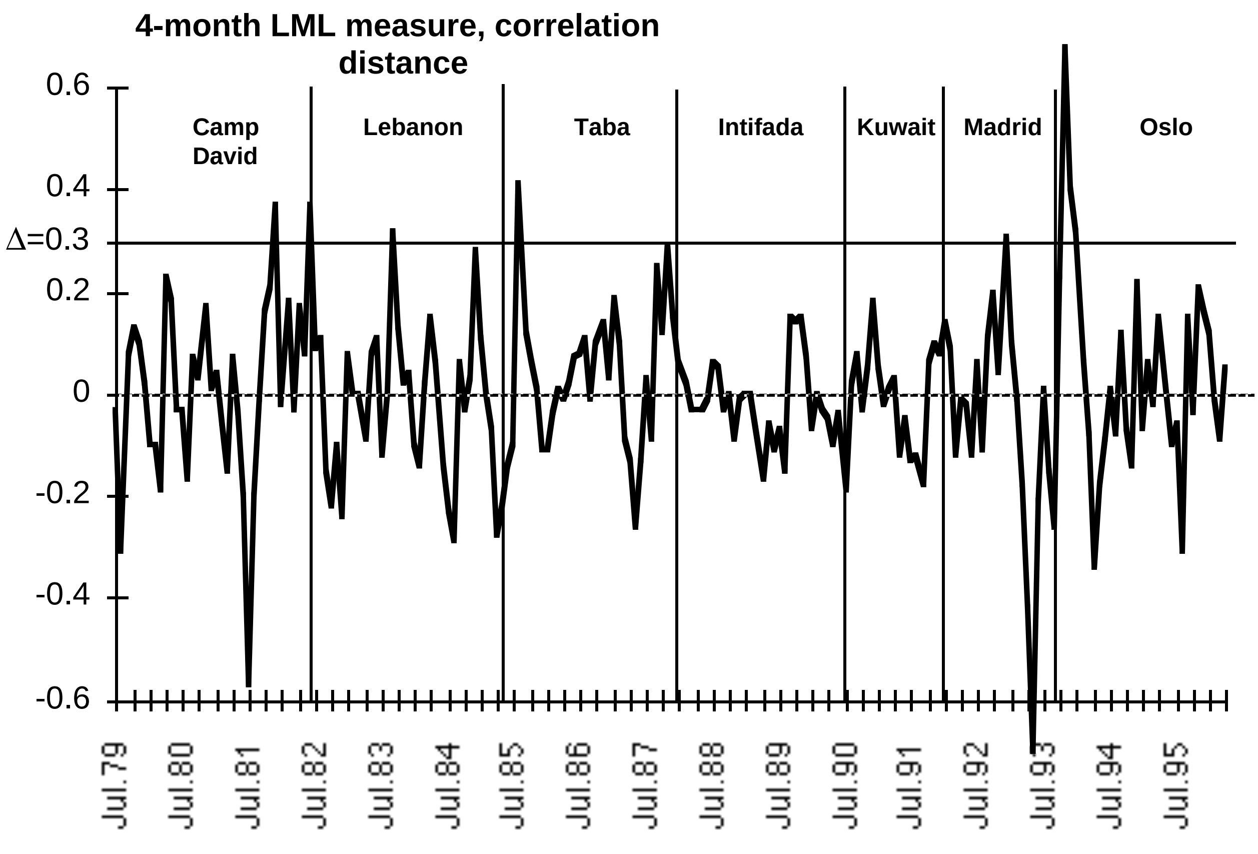
b. Two pre-Taba changes that may correspond to Israeli and Syrian changes of policy in Lebanon; K-Means also divided this period into at least two phases;

c. A peak in January 1993 that may reflect the USA shift in policy towards the Middle East that occurred with the change from the Bush to Clinton administrations.

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

At the Δ=100 level the Euclidean metric, in contrast, picks up only the three big changes in our *a priori* list—Lebanon, *intifada* and Oslo. If one sets Δ=50, a number of additional clusters can be delineated that are generally similar to those found in the correlation analysis, but there is a large difference between lower peaks in the value of the Euclidean LML and those peaks corresponding to the three major changes. The correlation LML, in contrast, produces peaks of roughly the same magnitude for all of the transitions except Oslo. For this reason, the remaining analysis is done with the correlation metric.[17]

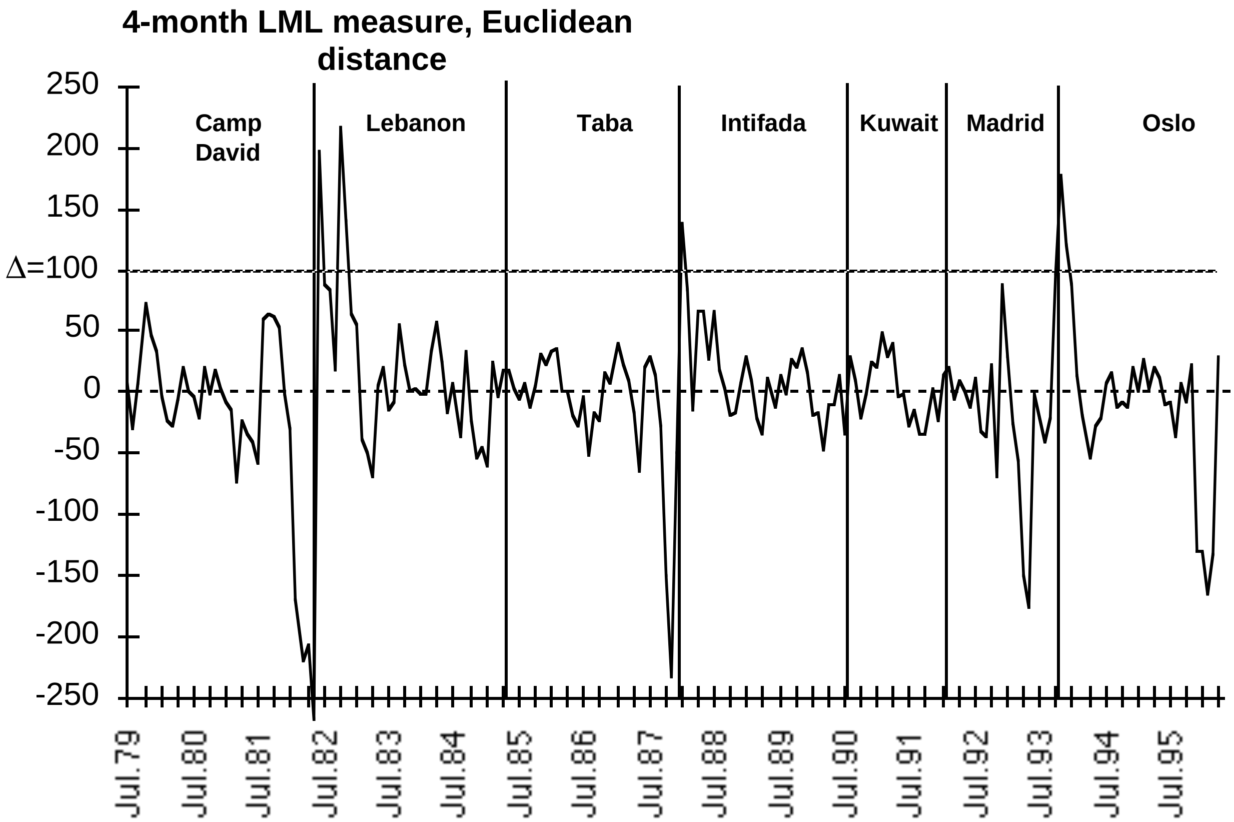
Figure 2.



---

[17] In Schrodt, Huxtable & Gerner (1996) we did this analysis using an earlier data set that was generated without the complexity filter. In that analysis the correlation metric was clearly preferable to the Euclidean. With the new data set—which we believe is less affected by coding errors—the two metrics seem to give similar results except for the greater variance of the Euclidean metric. Our original intention was to abandon the Euclidean metric but that decision may have been unduly influenced by the noise in the earlier data set.

Figure 3.



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 system would be expected to pull away from the cluster before it makes 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, $\Delta_8CD=CD_t - CD_{t-8}$). The 4- and 8-month periods were chosen by eyeballing[18] and probably

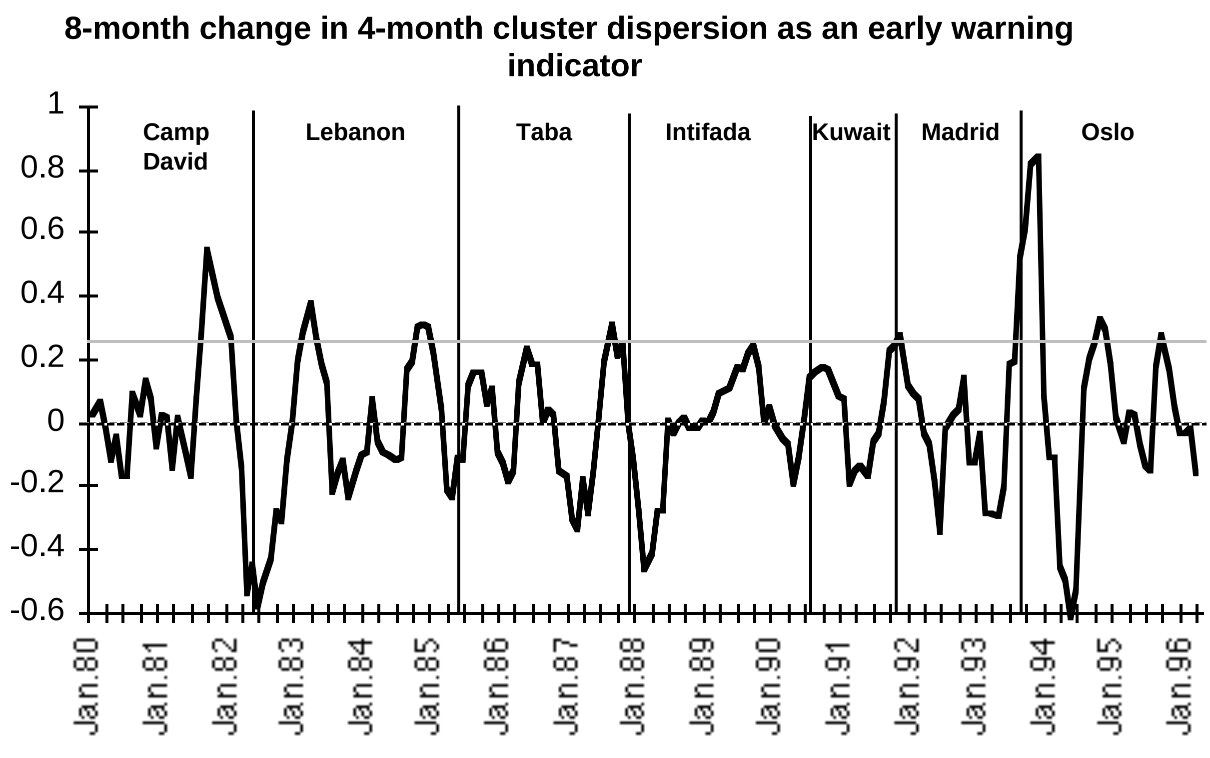
[18] And very little of that: k=4 was used because it was the LML interval already incorporated in the computer program and the 8-month differencing interval was the first we tried. Honest! Evidence that eyeballing does not lead to an exercise in "correlate wheat prices with the phase of the moon and the number of chickens in the barnyard" is given below in the Monte-Carlo analysis.

aren't optimal; the purpose of this measure is simply to tap the increase in cluster dispersion that occurs before many of the transitions.

In Figure 4, 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 occurs 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*;[19] the peaks in early 1995 and 1996 probably correspond to changes associated with the problems encountered in the Oslo peace process.[20]

Δ_8CD measure is a continuous measure 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 only indicates that some sort of change is going to take place and does not indicate *what* that change will be. 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.

Figure 4.



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 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 . On the other hand, the actions on the ground looked very different in July 1982 than in May 1982, during

---

[19] This is the "media fatigue" effect that is discussed in Gerner & Schrodt (1994).

[20] The period now referred to by both Israelis and Palestinians as "the *so-called* Oslo peace process."

which period the Δ_8CD measure was plummeting. Δ_8CD is clearly not a "barometric" early warning indicator where a political analyst can 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 so it 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.

Relationship between Clusters

The K-Means analysis showed a number of points where the system departed from one cluster, then returned to it later. In particular, there was some evidence of an "equilibrium" cluster that the system returned to when nothing extraordinary (e.g. invasion of Lebanon, *intifada*) was occurring.

To see whether the time-delimited clusters also reflected this phenomenon, we plotted the times of the three closest points, as measured by the correlation metric, to each point in the data set.[21] This is shown in Figure 5; the frequencies and column percentages are shown in Table 1. The distribution shows little or no evidence of an equilibrium cluster, although the figure shows a couple of other interesting characteristics. In this plot, an equilibrium cluster would be evident by a clear clustering of points off the main diagonal, as points in one time period were closely associated with other points distant in time (for example the Camp David and Madrid periods). This is not found: most of the points are relatively close to the main diagonal, so the dominant factor in the data seems to be trend.

A couple of features of Figure 5 and Table 1 are nonetheless interesting. First, there are some clear voids in the distribution of points. For instance, points in the Lebanon period are only rarely associated with points elsewhere in the data. Second, the Madrid period is unusually ambiguous, with points in that period being found close to points in almost all other periods (including Lebanon). Finally, some points in the early *intifada* period are located close to a number of other time points following the *intifada*. We suspect that these correspond to sporadic intervals of high levels of conflict between Palestinians and Israelis. These incidents tend to generate a lot of international reaction—particularly from the USA—that is reported by Reuters and therefore generate a disproportionate signal in the event data series.

---

[21] We also attempted to do this using by comparing the distance between the cluster centroids with the average within-cluster distance, but this proved to be completely useless. Virtually all of the centroids are located closer to each other than the value of the average within-cluster distance. We suspect that this is due to the large number of zero values in the dyadic time series causing all of the centroids—which are defined by the mean value of each dyad within the cluster—to be near the origin, so that the inter-centroid distance is quite small. The within-cluster distance, in contrast, includes points that are far from the origin.

Figure 5.

Scattergram of Three Closest Points to each Point in Data Set

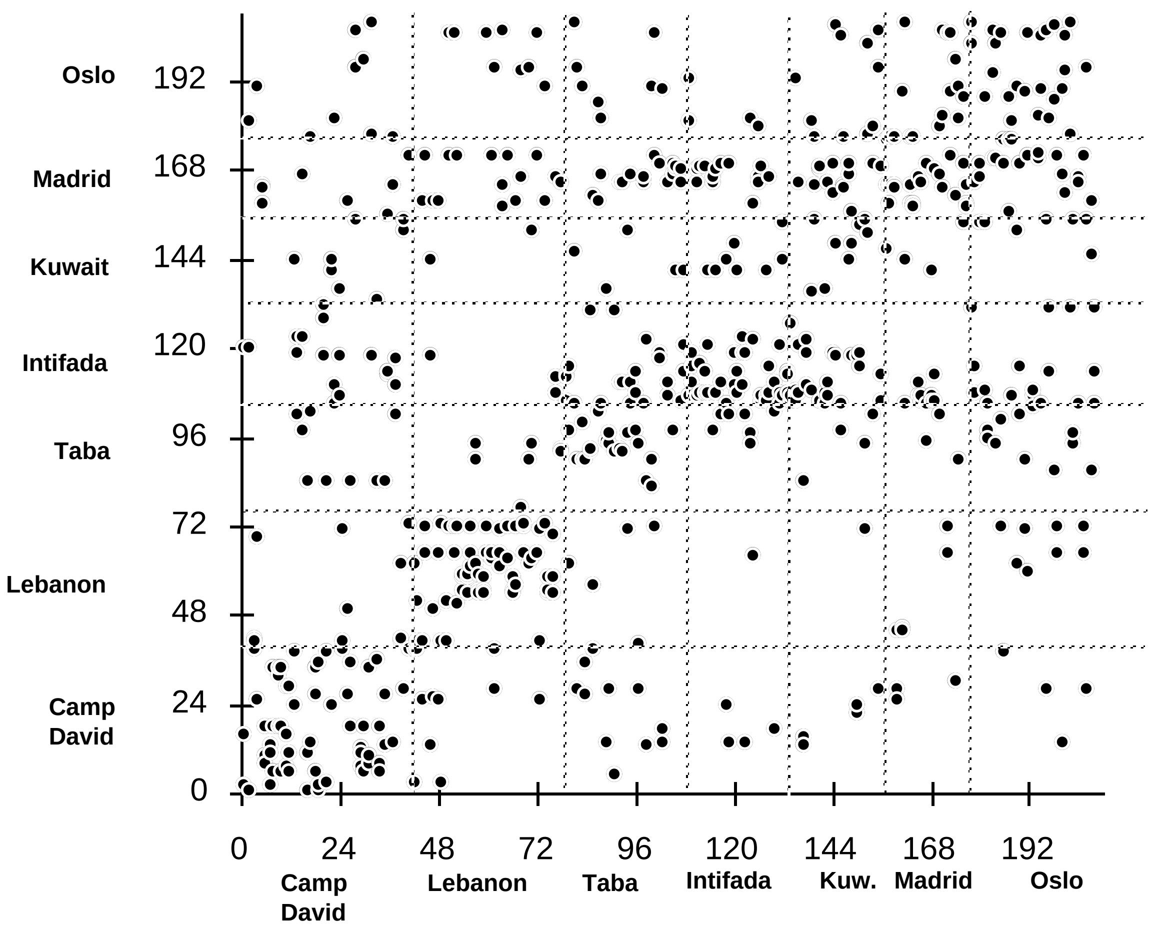


Table 1: Distribution of Three Closest Points by Cluster

N Col %	Camp David	Lebanon	Taba	Intifada	Kuwait	Madrid	Oslo
Oslo	7 6.1%	12 11.1%	4 4.4%	10 10.4%	6 13.3%	15 22.7%	38 36.2%
Madrid	11 9.6%	10 9.3%	13 14.4%	29 30.2%	13 28.9%	19 28.8%	27 25.7%
Kuwait	2 1.8%	1 0.9%	1 1.1%	7 7.3%	5 11.1%	3 4.5%	4 3.8%
Intifada	15 13.2%	2 1.9%	31 34.4%	41 42.7%	16 35.6%	18 27.3%	17 16.2%
Taba	7 6.1%	6 5.6%	23 25.6%	7 7.3%	3 6.7%	4 6.1%	8 7.6%
Lebanon	4 3.5%	69 63.9%	8 8.9%	0 0%	0 0%	5 7.6%	7 6.7%
Camp David	68 59.6%	8 7.4%	10 11.1%	2 2.1%	2 4.4%	2 3.0%	4 3.8%

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 will use preserves the sample size (192) and number of dyads (54) found in the data set analyzed in Schrodt & Gerner (1995), as well as the mean, variance, and first-order autocorrelation of the data within each dyad.[22] Specifically, we generated simulated data using an AR[1] process

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

where $c = \mu\,(1-\rho)$; $\phi = \rho$; $E(\varepsilon)=0$; $Var(\varepsilon) = s^2\,(1-\rho^2)$. As Hamilton (1994:53-54) notes, this will generate 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.[23]

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 clusters that were very small in size. 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[24] but have no relationship to each other.

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

1. Total number of points where $LML_t>\Delta$, where $\Delta=0.2$.[25]
2. The number of $LML_t>\Delta$ points that would signal a new cluster: this was defined (somewhat arbitrarily) as an $LML_t>\Delta$ point that had no $LML_t>\Delta$ points in the previous two time periods.[26] These times are called "cluster-defining points."
3. The standard deviation of LML_t and the early warning measure Δ_8CD; the means of both measures are zero.
4. The number of Δ_8CD measures that were greater than one standard deviation above Mean(Δ_8CD) at 0,1,2 and 3 "months" prior to a cluster-defining point.
5. 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 $LML_t>\Delta$

---

[22]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.

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

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

[25] $\Delta=0.20$ was the threshold that we found best delineated clusters in the Schrodt & Gerner (1995) data set.

[26] 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.

points. In the simulated data, these *a priori* transitions are essentially arbitrary—unlike the actual data, they do not correspond to conspicuous features in the data—but this measure gives some indication of the likelihood of finding $LML_t > \Delta$ points in the vicinity of any arbitrarily-chosen set of six transition points spaced at the intervals we chose. The proportion is used to compensate for the fact that simply by chance, the number of $LML_t > \Delta$ points near an arbitrarily-chosen transition will increase as the number of $LML_t > \Delta$ points increases, and the number of $LML_t > \Delta$ points in the simulated data is substantially higher than in the actual data.

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 192-3-11=178 points. All of the analysis was done using the correlation metric.

The results of the Monte-Carlo analysis are presented in Table 2, and an example of the statistics generated by one such data set are shown in Figure 6. In Table 2, 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;[27] the probabilities are based on the actual distributions of the statistics in the simulated data rather than on a Normal approximation.

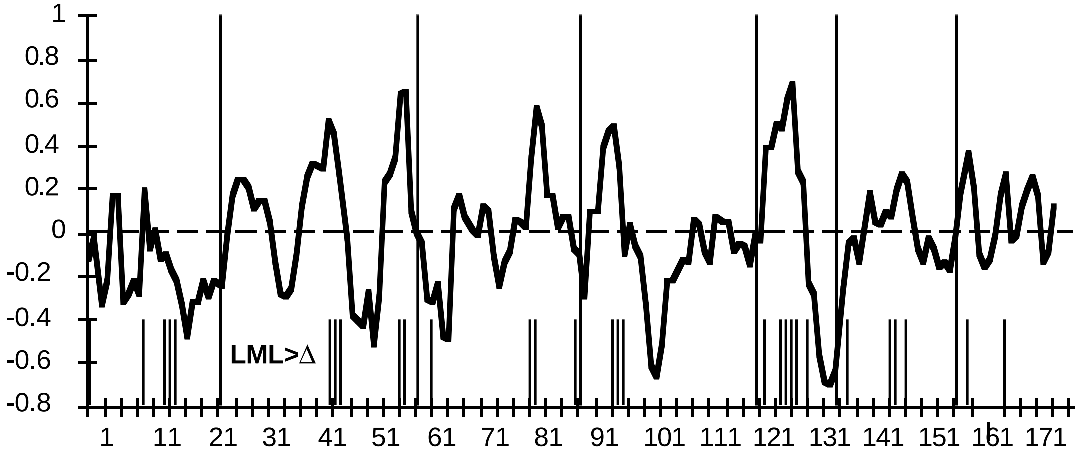
Table 2
Statistics Computed from 1000 Simulated Data Sets, Δ=0.2

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 & $\Delta_8 CD_{t-k}$>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 (>)

[27] 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.

Figure 6.

LML>Δ and Δ8CD statistics in a set of simulated AR[1] dyads



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 differ 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 be found 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, but also that its performance is due simply to autocorrelation in the data rather than any more complex characteristics involving dyadic interactions.

The high number of $LML_t>\Delta$ points combined with the fact that the standard deviation of LML and Δ_8CD are higher in the simulated data than in the observed data suggests that the value of Δ—a free parameter that was set arbitrarily—may have been set too low for the simulated data. We re-ran the simulated data sets with Δ=0.35, a level of Δ which gives roughly the same number of cluster-defining points in the simulated data as were found in the observed data with Δ=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. Consequently the behavior of the predictive measures does not seem to be solely due to the difference in the number of $LML_t>\Delta$ points.

Table 3
Statistics Computed from 1000 Simulated Data Sets, Δ=0.35

Statistics for Δ=0.35 (N=1000)		Simulated mean	Simulated standard dev	Observed value	One-tailed probability
Total LML_t>Δ		13.56	4.34	15	0.680 (<)
Cluster-defining LML_t>Δ		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 & 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>Δ 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 somewhat ambiguous due to the existence of the free parameter Δ. If we take as given the Δ=0.2 separation threshold, then the observed data has far fewer clusters than we would expect to find in a set of data following the null model. By raising the level of Δ, we can match the number of empirically-determined clusters, though 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 deviations of LML_t is lower in the observed data because of cross-correlation (and in a few dyads, higher-order autocorrelation) of the dyads.

Estimating the Weights

All of the earlier 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—are not necessarily optimal for clustering and early warning.

We therefore attempted to estimate optimal weights using a genetic algorithm (Holland 1975; Grefenstette 1987; Goldberg 1989) that maximized the following clustering measure:

$$\text{Fitness} = \frac{\text{average distance between adjacent clusters}}{\text{average distance within clusters}}$$

where "distance" $\| x_i - x_j \|$ is defined by the correlation metric and 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*. The measurement of the points in adjacent clusters (rather than comparing the distance of a cluster to all other clusters, as done in discriminant[28]) is done to allow [again] the possibility of the system returning to an equilibrium behavior, so that clusters that are separated in time might occupy the same space. We provided for this possibility based on evidence for the existence of equilibrium clusters in Schrodt & Gerner (1995) and a clustering study of the Rosecrance typology of European systems in Schrodt (1995a).

The optimization allowed the system to determine the cluster breaks as well as determining the weights: this made the problem non-linear and required the use of a numerical optimization method rather than an analytical optimization method such as discriminant. A cluster break was any point that met the following conditions:

1. $LML_t > 0.20$
2. No cluster breaks in the previous 8 months (i.e., minimum cluster size of 8 months)

Because the weights are estimated, the 0.20 threshold is somewhat arbitrary—in theory the system should be able to adjust the weights to this level of correlation—and the $LML_t > 0.20$ threshold is simply comparable to the level found to produce cluster breaks 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

$$XY_t = \mathbf{w} \bullet \mathbf{c_t} = \sum_{i=1}^{22} w_i \, c_{it} \qquad \text{where}$$

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 breaks between clusters are determined using the $LML_t > 0.20$ threshold and minimum size rules discussed above, and the fitness measure is computed.

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 fitness and

---

[28] Discriminant also measures the distances between the *group means* of each cluster, rather than measuring the distances between individual points.

the 16 vectors with the lowest fitness are replaced with new vectors created by recombination and mutation of the top 16 vectors.[29] 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. This system was implemented in a C program; the source code is available from the authors.

The results produced by this system were consistent, if disappointing. The genetic algorithm works reliably, if slowly, and most of the runs showed no pathological behavior such as genetic drift or multiple local maxima.[30] The best fitness level, when the system was allowed to run for 48 generations and with the system constrained to at least five breaks in the top 16 vectors, ranged from 1.90 to 2.02. The global maximum, however, is rather uninteresting, as shown in Figure 7. Figure 7 aggregates the cluster breaks found in the top 16 vectors in 11 different runs of the GA and shows the number of cluster breaks occurring in each month of the data set. Most of the cluster breaks occur at 5 points. Two of these correspond to the onset of the Israeli invasion of Lebanon (Jun-82) and the *intifada* (Dec-87); the breaks near July-83 probably correspond to the increased attacks against Israeli and international forces in Lebanon; the others do not seem to correspond with obvious political changes in the region.

Even more problematic is the fact that these weights provide no differentiation in the second half of the data set, which contains the Madrid and Oslo peace processes. This result is not entirely surprising, since the data have a large trend component. The second half of the period not only contains the *intifada*, but also the virtual cessation of hostilities in Lebanon after Syria consolidated control in 1989. Politically, the second half of the data set is quite distinct from the first half.

We also looked at the cluster breaks estimated after only four generations of the GA. These results are shown in Figure 8; Figure 9 shows the Goldstein LML measure superimposed on this.[31] The failure to differentiate the post-1987 period is evident even at the fourth generation, with only a small number of vectors identifying breaks, though the breaks that are identified correspond closely to the Kuwait, Madrid and Oslo clusters that we designated *a priori*. Curiously, there is a much lower correspondence in the pre-1987 period between the LML measure and the cluster breaks identified by the fourth generation estimates.

The consistency of the GA results imply that the algorithm is working appropriately to locate true clusters in the data set and then modifying the weights to exploit these: Different runs of the GA give similar results and some of the break points (notably 82-06 and 87-11) correspond to obvious political changes in the region. These presumably correspond to a true and fairly conspicuous global maximum in the search space in terms of the cluster breaks, though not necessarily in terms of the weights themselves. The average standard deviation of the 22 weights in the 176 final vectors generated in our GA experiments is 5.2, with a minimum standard deviation of 3.8 and a maximum of 6.4, with a very uniform distribution within this range. Given that the initial weights were set uniformly in the range [-10, +10], this level of variance seems quite high—a Normal 95% confidence interval around zero covers the entire initial range. This in turn is consistent with a global maximum being found quite quickly: Apparently there are a variety of different sets of weights that produce this same limited set of clusters.

---

[29] 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. They rarely survived more than a couple of generations and appear to have contributed little to the optimization.

[30] Two of the runs started with a fitness around 1.80 and stayed there, with all of the vectors identifying the same four break points: May-82, Nov-84, Oct-86, and Nov-87. This is probably a local maximum, but it is fairly similar to the somewhat higher maximum identified in the other runs.

[31] The scale on Figure 9 has been truncated.

Figure 7.

Cluster Breaks in Final Generation

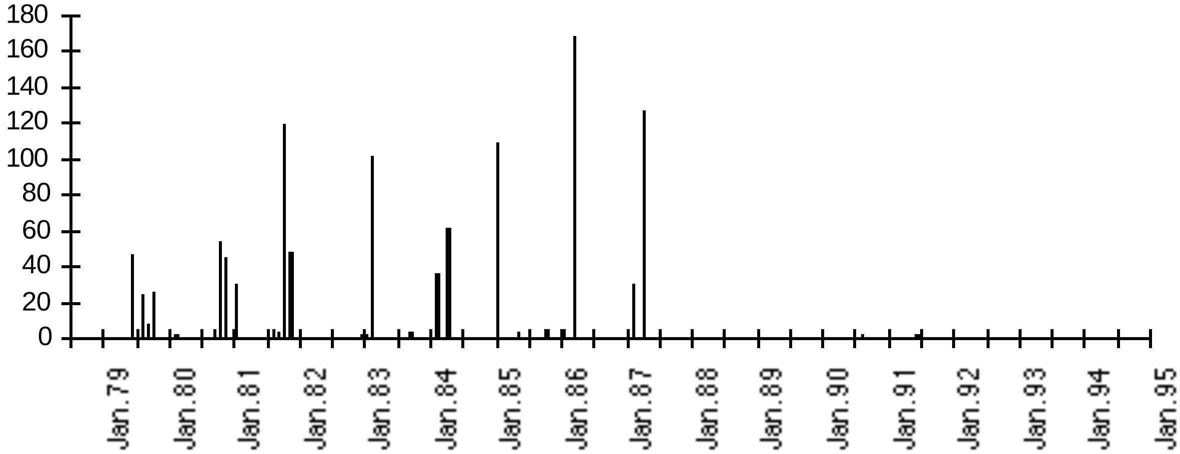
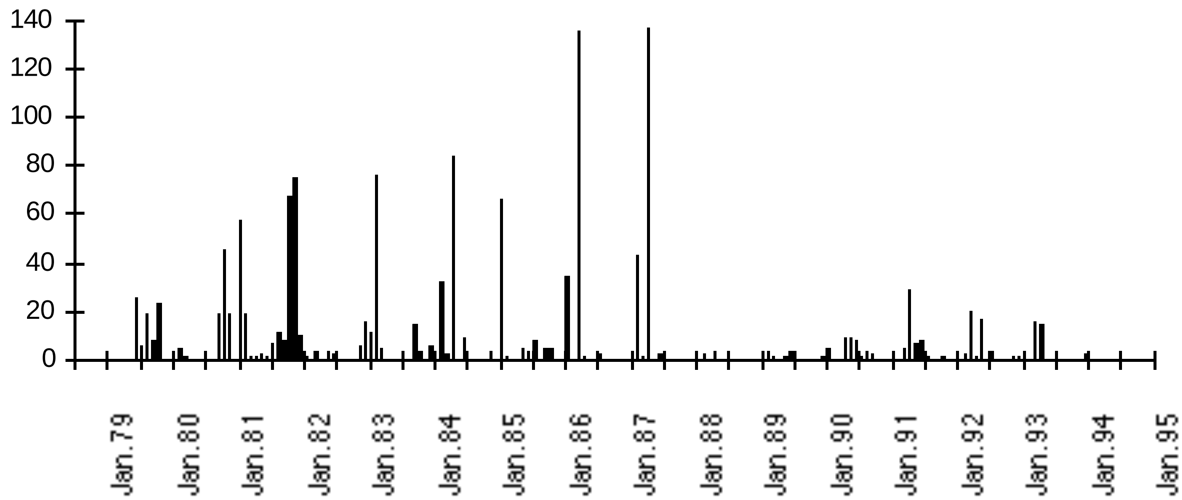


Figure 8.

Cluster Breaks in Fourth Generation

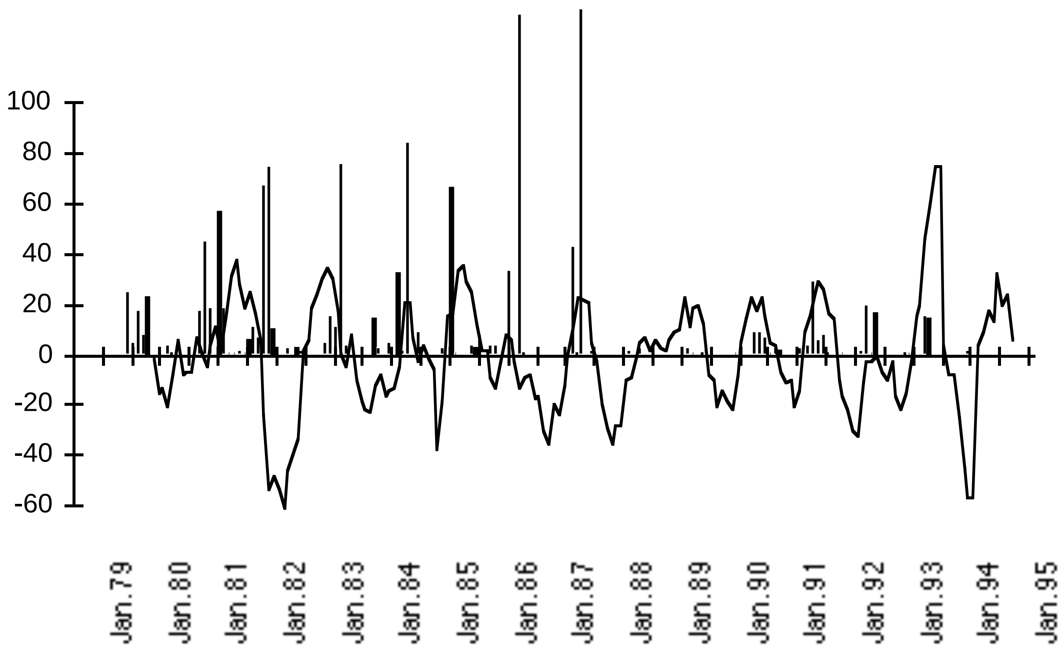


A comparison between the average of the GA weights and the Goldstein weights—shown in Figure 10—finds very little correspondence between the two in either sign or magnitude.[32] The only point where there is some correspondence between the GA and Goldstein weight is in the WEIS categories 11 (Reject: 4.0% of total events), 12 (Protest: 6.6%) and 13 (Deny: 1.2%), with additional correct signs (though not magnitudes) on 17 (Threaten: 1.2%) and 18 (Demonstrate: 1.8%). It is possible that the system is identifying some change in the pattern of accusations and counter-accusations, in addition to picking up the major features of the Lebanon invasion and the intifada.

Our sense is that the weights determined by the GA are picking up some distinctive features of the data, but those distinctive features aren't what we are interested in as political analysts. The 1982 invasion of Lebanon and the *intifada* are the two most conspicuous features of the data, so those are found, but after locating those clusters the GA seems to be going after something else, and the GA has no information as to *what* is important. The upshot of this analysis is that the "Martian approach" of just telling the system to find clusters, any clusters, doesn't work at any level of subtlety.

Figure 9.

Generation 4 Clusters vs Goldstein LML Score

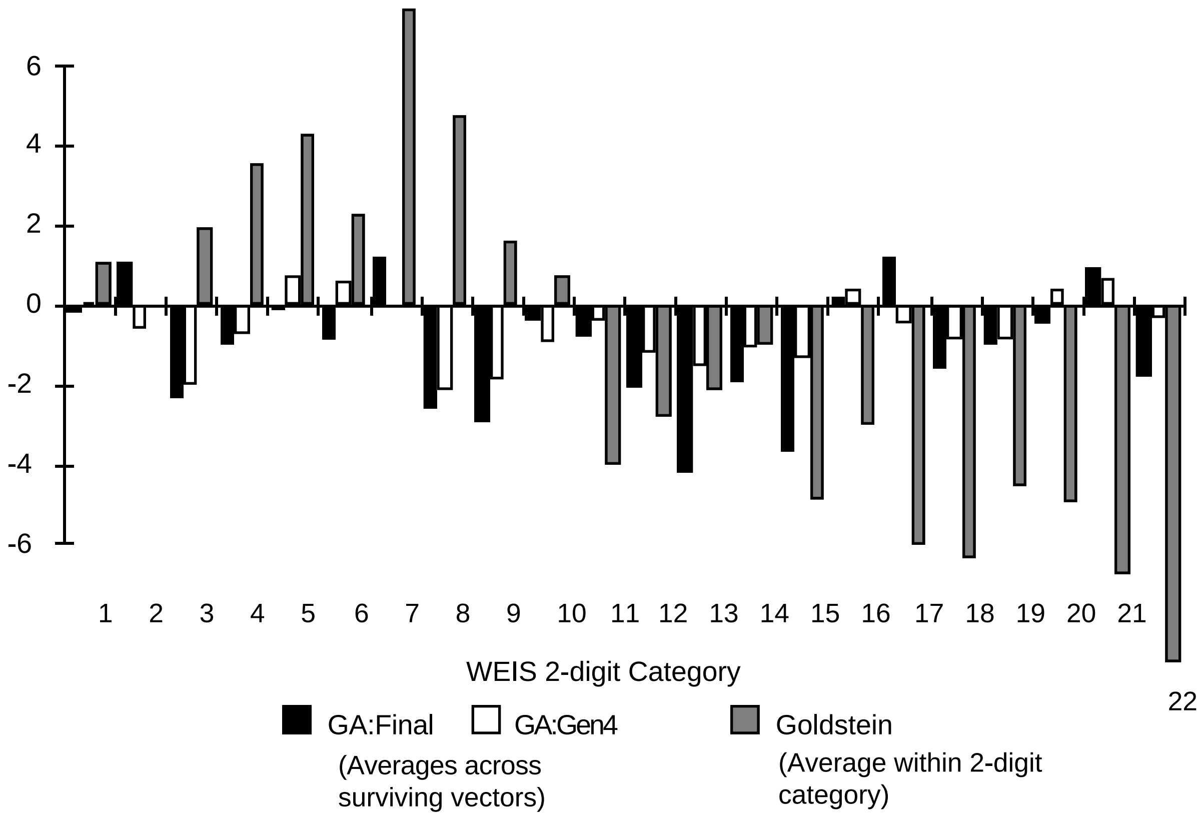


---

[32] It may be possible that the 2-digit WEIS categories do not differentiate behavior as well as the 3-digit categories employed in the Goldstein weighting scheme, but given the plausibility of the 2-digit WEIS scale in terms of differentiating a conflict/cooperation dimension, this seems unlikely to account for the difference between the Goldstein and GA clusters. A more likely complication is something analogous to collinearity, which could explain why some of the signs of the coefficients are the reverse of those in the Goldstein scheme. There also appears to be little relationship—direct or inverse—between the absolute value of the weights and the number of events of each time that are found in the data.

Figure 10.

Comparison of Goldstein weights and weights determined by genetic algorithm



An unweighted correlation metric

With the unsatisfactory results of the estimation using the genetic algorithm, we tried one additional experiment: computer the distance between points by correlating the frequencies of the 2-digit WEIS events without applying any weighting (in principle this method could also be applied to 3-digit categories). The standard LML measure was computed, with the only difference being that the correlation was computed on vectors containing counts of the twenty-two 2-digit WEIS events for each dyad-month, so each correlation used 22 x 56 = 1232 points rather than the 56 points of the vectors containing the Goldstein scores.

Figures 11 through 13 show the results of this analysis for both the LML and Δ_8CD early warning measure; "CodeCorr" refers to the results of computing the correlation distance using the counts of the individual codes. In general, the CodeCorr measure produces results quite similar to those of the Goldstein measure[33]—particularly in terms of matching the *a priori* cluster breaks —which could argue for not using weights at all! At worst, this analysis suggests that the clustering method is not strongly dependent on the Goldstein weights, and the frequency of coded events alone is sufficient to differentiate the major political features of the data.

---

[33] The correlation (r) between the Goldstein and CodeCorr LML is 0.63 (and 0.66 for a 4-month moving average of the two measures). The correlation between the two Δ_8CD measures is 0.62.

Figure 11.

LML Computed with Event Code Correlations

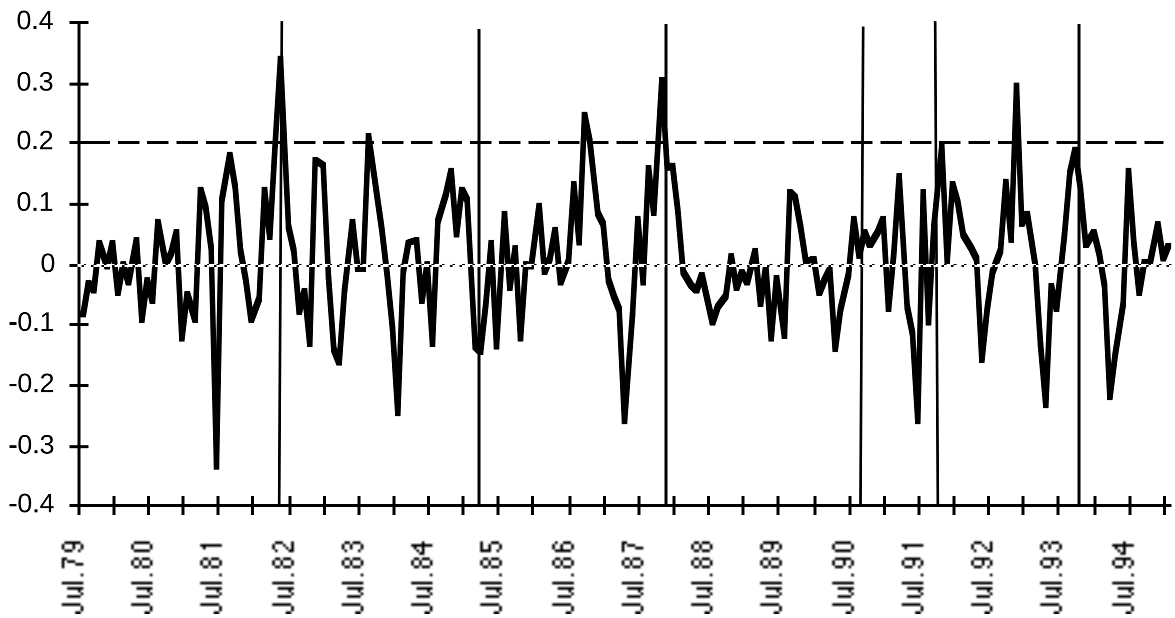


Figure 12.

Comparison of CodeCorr and Goldstein LML

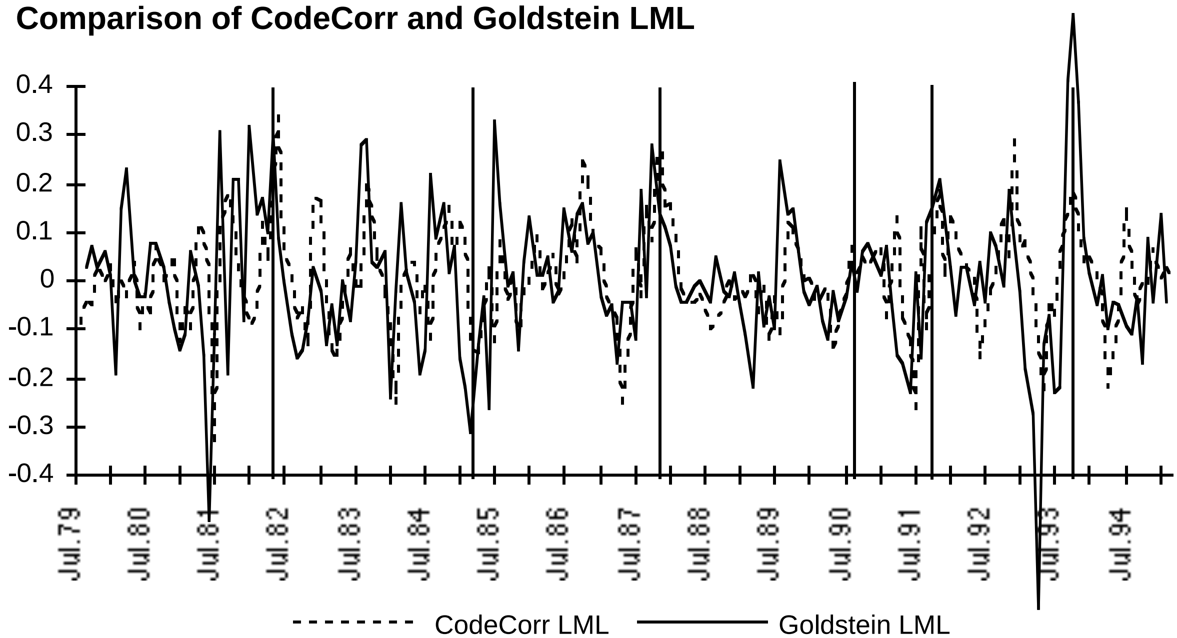
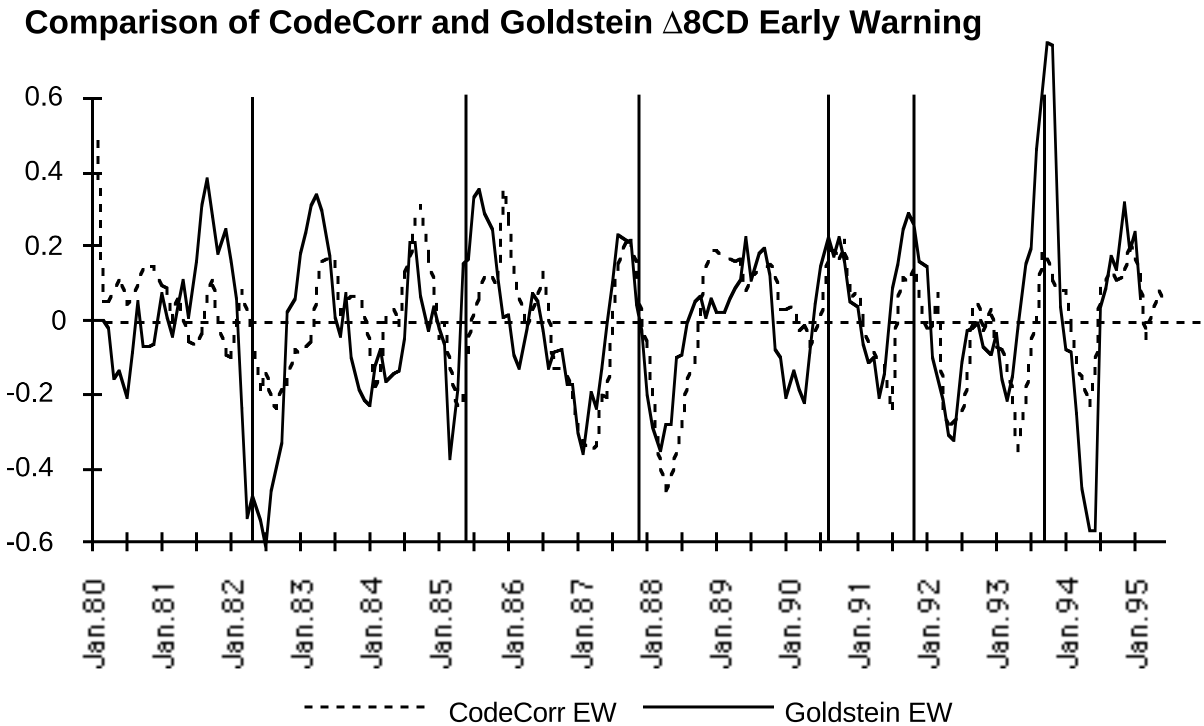


Figure 13.



But before abandoning the Goldstein weights, note in Figures 12 and 13 that the variance of the Goldstein results is slightly higher than that of the CodeCorr results in the LML measure, and noticeably higher in the Δ_8CD measure.[34] Furthermore the Goldstein measure gives a clear early warning on Lebanon, which the CodeCorr measure misses altogether, picking up only the cluster break. The Goldstein measure goes off the chart (literally...) on Oslo, whereas the CodeCorr measure shows only a modest rise. 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.

Conclusion

Following the demise of the DARPA early warning efforts in the mid-1970s, efforts at the development of quantitative early warning models went into eclipse. Yet when measured against what is practical in the today, the DARPA efforts were quite primitive in their reliance on time-consuming and unreliable human coding and their use of 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 much to look at either.

In this concluding section, we will address two sets of issues. First, we will evaluate the utility of the clustering approach based on our analysis of data from the Levant. Second, we will comment more generally on the role of various approaches to quantitative early warning, and on the relationship between quantitative and qualitative early warning.

We draw three general conclusions from our analysis using time-delimited clusters. First, our empirical results continue to support the approach of analyzing phases of political behavior by looking at the movement of a point defined by the vector of dyadic interactions. The pattern of

---

[34] This difference is clearer in the colored charts produced with Excel than in the black-and-white versions presented here.

variation in LML_t seen in Figures 2 and 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.

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. This was not surprising for the Euclidean metric but we did not necessarily expect it to hold for the correlation measure. The $LML_t > \Delta$ method used to delineate the clusters is conceptually simple and computationally efficient;[35] in fact the algorithm is sufficiently simple that it may be possible to determine analytically some of its statistical properties. The Δ_8CD measure also appears promising as the basis of an early-warning indicator.[36]

Table 4 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 Δ_8CD measure—although not the LML cluster analysis—indicates significant changes following the Oslo peace process phases. (Given that we did some of this analysis during a period when the Palestinian population and the authors of this paper were confined to their towns by an Israeli military closure, differentiating a post-Oslo period seems like a pretty good call.) Based on Δ_8CD, we might also have designated a post-*intifada*, pre-Madrid cluster beginning in late 1989.

All of our analyses missed the Kuwait transition, and, in retrospect, we should not have included this as one of our *a priori* clusters. The events leading to this crisis were completely exogenous to the dyads we are studying, and the long-term changes that it may have caused in the Levant only show up several months after the onset of the crisis.

Δ_8CD usually provides two to six months of early warning. It provides no early warning for the Oslo transition, and provides no distinct warning of the June 1982 invasion of Lebanon. The Δ_8CD measure also has some false-positives where the measure peaks just below the critical level. This is to be expected—any measure that does not contain false positives (for example our Euclidean LML measure) is probably insufficiently sensitive to political events. We are not dealing with a deterministic system here, 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. The pre-Lebanon peak in LML may be such an example—for 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.

---

[35] Running on a standard Macintosh Powerbook 520, the Monte-Carlo program did 200 experiments per hour running inside the Think Pascal system. This included generating the simulated data; the only time-consuming aspect of the cluster analysis itself is computing the distances between points.

[36] Or, of course, Δ_8CD may be shifting due to changes in Reuters coverage of the region. Consequently, another possible interpretation of the success of the Δ_8CD measure might be that it reflects, in an aggregate fashion, 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.

Table 4.
Clusters Determined by the Analysis

Initial date of cluster	**Political characteristics**	**a priori cluster?**	**LML cluster $\Delta > .30$**	**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-83
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 focus in the second half of this paper was on determining the weights of the individual categories. This was not particularly successful. Another way to optimize the measure would be to weight the various dyads differentially . For example, in the politics of the Levant, it is clear that Syrian-Lebanese relations are more important than, say, Israeli-Russian or USA-Jordanian. This is the approach of the discriminant analysis in Schrodt & Gerner (1995), where the discriminant function coefficients weight various dyadic behaviors.[37]

We've not extended our analysis to differentially weight the dyads for two reasons. First, it is likely that much of this weighting has already been done for us by Reuters. If we give the reporters and editors of Reuters credit for being 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 dyads will be higher than those of unimportant dyads. 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. This is not to make a virtue of the necessity of relying on Reuters, but simply an observation that Reuters' intrinsic weighting seems to be doing a pretty good job for purposes of forecasting.

Differential weighting might be important in some other situations, particularly those having a steady level of reporting such as might be found in reports from IGO or NGO field workers. There is clearly no difficulty in principle (other than the degrees of freedom issue discussed below) in weighting dyads, but it does not seem necessary in the case of Reuters reports on the Levant. Given the limited funds available to IGOs and NGOs, even field reports probably show some weighting: human right monitors, for example, are more likely to be deployed in Rwanda than Iceland.

The other reason we've not weighted dyads is the degrees-of-freedom issue: 22 (or 63) WEIS weights plus 54 dyadic weights introduces a lot of parameters, which become all that more complicated due to the collinearity of the data. The principle of parsimony says that parameters should not be introduced unnecessarily and the model seems to be working fairly well as it is; thus dyadic weighting does not seem to be a priority.

Before looking at the issue of the relationship between quantitative and qualitative early warning, a couple of remarks are in order on the relationship between our clustering approach and other forms of quantitative early warning. First, we regard the structural and the dynamic approaches 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 likely to provide theoretical guidance as to why a system is likely to experience problems,

---

[37] Discriminant analysis could also provide an estimate of the coefficient weights given an *a priori* set of clusters. The analysis in Schrodt & Gerner (1995) constructed discriminant functions of the form

$$df_t = \sum_i d_i\, g_{it} \qquad \text{where } g_{it} = \text{Goldstein scaled score for dyad i at time t}$$

However, since

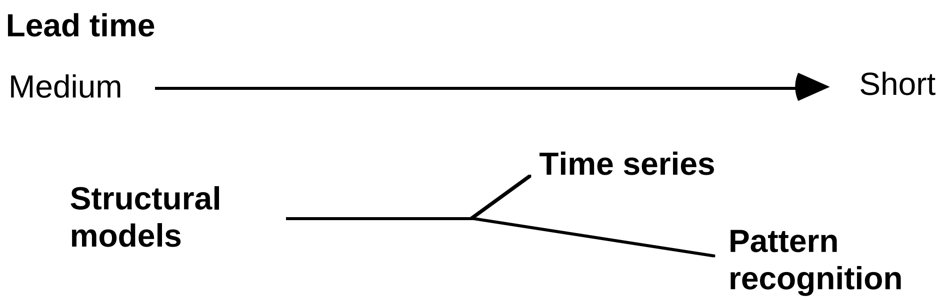
$$g_{it} = \sum_{j=1}^{22} w_j\, c_{jt} \qquad \text{where } c_{jit} = \text{number of events in WEIS 2-digit category j for dyad i at time t}$$

one could rerun the discriminant as

$$df_t = \sum_{j=1}^{22} w_j \sum_i c_{jit}$$

to estimate the optimal weights for the WEIS categories under the assumption of equal weights for the dyads. Alternatively one could use the dyadic weights determined by the first discriminant analysis. Because the choice of dyadic weights can be expected to have a substantial impact on the estimates of the category weights, the assumption of equal weighting is probably preferable if one is trying to obtain a general estimate of scaling weights for the WEIS categories.

which might provide insights as to the types of actions that could be taken to ameliorate an impending crisis. Structural models are unlikely ever to excel at predicting the exact timing of breakdowns—the variables that they have identified as theoretically important change too slowly—and this is where dynamic models come into play. The relationship of the approaches, then, may be something like the following:



In this analysis we have not considered an alternative class of dynamic models—those based on the analysis of event sequences, rules, patterns and precedents (see Cimbala 1987; Hudson 1991; Schrodt 1995a). These are likely to provide a greater amount of contextual information than provided by the numerical time-series methods, and as a consequence may be useful in identifying the immediate events leading to a crisis. For example, while the Kuwait transition is invisible in our cluster analysis, the pattern of events preceding Iraq's invasion of Kuwait follows Lebow's (1981) "Justification of Hostility" crisis very closely, and such patterns could be used for very short-term forecasting. An assortment of computationally-intensive non-linear forecasting techniques methods also have 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 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 models and accurate 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.

Finally, some comments are in order on the relationship between quantitative, statistical methods of forecasting and the traditional qualitative, non-statistical methods. We regard statistical early warning indicators as a supplement to, rather than a replacement for, traditional qualitative early warning methods. Because political behavior is a human activity (in contrast, for example, to weather or earthquakes), human understanding and intuition are likely to be powerful tools in predicting that behavior. Early warning is also an "ill-defined problem" (Moray 1984, 11) within a complex system, where neither the relevant variables nor the relevant processes have been fully, or even adequately, identified. We also face the practical constraint that purely statistically-based warning systems are unlikely to be accepted in the qualitatively-oriented policy community (Laurance 1990).

At the same time, statistically-based forecasting methods fill two gaps that are inherent in human-based qualitative approaches. First, while human intuition is a valuable tool in understanding political behavior, cognitive biases can blind an analyst to a situation that is rapidly changing despite his or her expectations to the contrary. Major United States intelligence lapses such as the failure to anticipate the establishment and stability of the Islamic Republic of Iran and the failure to predict the collapse of communism in Eastern Europe illustrate the extent to which this problem can affect even well-funded and experienced analysts. Second, statistically-based methods are capable of consistently monitoring a much larger amount of information than a human analyst can monitor. A system based on computerized analysis of machine-readable sources can monitor 24-hours-a-day without fatigue, distractions, political pressure or committee meetings.

Based on our field experience in Palestine, the reports of IGO and NGO field workers—appropriately filtered—are an important untapped source of information for early warning purposes.[38] NGOs increasingly provide access to quality information about stressed populations—for example, refugees and minorities—that is not available from other sources. As long as those reports were submitted in a reasonably standard form (e.g. with the subject-verb-object placement in natural language reports corresponding to the source-event-target variables of event data) it should be possible to code them by machine. However, these reports must be carefully filtered for bias and attempts to deliberately manipulate the early warning system,[39] and might well be adjusted to account for the political sophistication, timeliness, reliability, and sensitivity of the source. The machine-coding of field reports for use in statistical early warning systems provides an additional opportunity for merging the qualitative and quantitative approaches.

Finally, statistical systems designed specifically for early warning *may* be able to utilize general models of behavior that can apply in a number of different circumstances, rather than depending on the area-specific knowledge of individual analysts. This does not eliminate the need to employ analysts with area-specific knowledge after a potential problem has been detected. Warning about a problem is not the same as understanding it: When your car's thermometer goes to "HIGH", you are well-advised to stop the engine, but the thermometer reading alone cannot determine whether the problem is a broken fan belt, a leaky radiator or a malfunctioning thermostat. Statistical systems are never going to replace area-specific knowledge, but they may usefully supplement it, particularly if a large number of areas are being monitored.

The qualitative opportunities for receiving information relevant to early warning has increased dramatically in the past five years with the availability of inexpensive machine-readable commercial news sources and the proliferation of reports available from IGOS and NGOs via the internet. During this same period the challenges have also increased, for example in the potential dissolution of some states in the post-Cold War period and the appalling resurgence of genocidal outbreaks such as those witnessed in Cambodia, Rwanda and Bosnia. Consequently we believe that there is an important role for the development of quantitative indicators. To the extent that an area is adequately monitored by electronically-readable sources, real-time quantitative forecasting using machine-coded event data is quite inexpensive and can easily operate in the background as a supplement to qualitative forecasting.

---

[38] That same field experience leads us to doubt the utility of using reports from the visual media such as CNN. Particularly CNN...watching CNN reporters in action is comparable to watching sausage being made, and one of us is vegetarian! As we reported earlier in Schrodt & Gerner (1994), Reuters reports generally seem to correspond well with what we observed on the ground.

[39] In Schrodt (1995b) one of us suggested that a major potential weaknesses in an Internet-based early warning system was its vulnerability to manipulation. Shortly thereafter a concrete example of this occurred in a survey done by the editors of the computer magazine *Byte* (September 1996: 32), who attempted to use the Internet to survey users on the future of various operating systems. The survey failed due to the efforts of users of two relatively obscure operating systems—undoubtedly coordinating their effort through email networks—to "stuff the ballot box," leading the *Byte* editors to end their article with the frustrating observation, "And, to the individual who voted over 80 times in the survey (your IP address is 198.182.4.224): Get a life." While ideologies rarely evoke the level of emotional commitment shown by computer users to their operating systems, the example is instructive with respect to the ease with which Internet sources of information can be manipulated, and the likelihood of this manipulation occurring.

Bibliography

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

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* 3,3:193-222.

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.

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.

Cahill, Kevin M., ed. 1996. *Preventive Diplomacy: The Therapeutics of Mediation.* New York: Basic Books.

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.

Clark, Lance. 1989. *Early Warning of Refugee Flows.* Washington, D. C.: Refugee Policy Group.

Crocker, Chester and Fen Osler Hampson, eds. 1996. *Managing Global Chaos: Sources of and Response to International Conflict.* Washington, D.C.: United States Institute of Peace Press.

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* 4,1:98-105.

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., 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. 1996. "Reciprocity, Bullying and International Cooperation: Time-Series Analysis of the Bosnia Conflict." International Studies Association, San Diego.

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 meetings, Chicago, February 1995.

Gurr, Ted R. and Mark Irving Lichbach. 1986. "Forecasting Internal Conflict: A Competitive Evaluation of Empirical Theories." *Comparative Political Studies* 19,1:3-38.

Gurr, Ted R. and Barbara Harff. 1994. "Conceptual, Research and Policy Issues in Early Warning Research: An Overview." *Journal of Ethno-Development* 4,1:3-15.

Gurr, Ted R. and Barbara Harff. in press. *Early Warning of Communal Conflict and Humanitarian Crisis.* Tokyo: United Nations University Press.

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.

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.

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* 38,3: 447-474.

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

Leatherman, Janie and Raimo Väyrynen. 1995. "Structure, Culture and Territory: Three Sets of Early Warning Indicators." International Studies Association, Chicago.

Lund, Michael S. 1996. *Preventing Violent Conflicts: A Strategy for Preventive Diplomacy.* Washington, D.C.: United States Institute for Peace.

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.

Moray, Neville. 1984. "Humans and their Relations to Ill-Defined Systems." In Oliver Selfridge, Edwina L. Rissland and Michael A. Arbib, eds. *Adaptive Control of Ill-Defined Systems* New York: Plenum.

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

Schmeidl, Susanne. 1997. *From Root Cause to Preventive Diplomacy: Possibilities and Limitations of the Early Warning of Forced Migration.* Westport, CT: Praeger.

Schrodt, Philip A. 1995a. *Patterns, Rules and Learning: Computational Models of International Behavior.* Manuscript posted at http://wizard.ucr.edu/polmeth/ working_papers95 /schro95.html

Schrodt, Philip A. 1995b. "Implementation and Methodology Present More Constraints than Theory: Comments on An Early Warning System for the United Nations." *Mershon International Studies Review* 39,2: 319-320

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. 1995. "Empirical Indicators of Crisis Phase in the Middle East, 1979-1995." Peace Science Society, Columbus, Ohio.

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. 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 Richard L. Merritt, Robert G. Muncaster, and Dina A. Zinnes, eds. *International Event-Data Developments: DDIR Phase II.* 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.

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.