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

Philip A. Schrodt and Deborah J. Gerner
Department of Political Science
University of Kansas
Lawrence, KS 66045-2157 USA
phone: 785-864-3523 fax: 785-864-5700
schrodt@ku.edu gerner@ku.edu

May 2000

Forthcoming, *American Political Science Review* 94,4: 803-818 (December 2000)

Authors' Note:

Philip A. Schrodt (schrodt@ku.edu) is Professor of Political Science, University of Kansas, Lawrence, KS 66045. Deborah J. Gerner (gerner@ku.edu) is Associate Professor of Political Science, University of Kansas, Lawrence, KS 66045-2157.

We received valuable suggestions from Jon Pevehouse, from the "Beer and Politics Seminar" at the University of Kansas, and from the anonymous referees; Rajaa Abu-Jabr provided editorial assistance. This research was funded in part by the National Science Foundation through grant SBR-9410023 and by the University of Kansas General Research Fund grant 3500-X0-0038.

Abstract

This article uses cluster analysis to develop an early warning model of political change in the Levant as reflected in WEIS-coded event data generated from Reuters between April 1979 and December 1998. We employ a new statistical algorithm that uses the correlation between dyadic behaviors at two time points in time to identify clusters of political activity. The transition to a new cluster occurs when a point is closer in distance to subsequent points than to preceding points. These data clusters begin to "stretch" prior to breaking apart; this characteristic can be used as an early-warning indicator. The clusters identified by this technique correspond well with phases of political behavior identified *a priori*. A Monte-Carlo analysis shows that the clustering and early warning measures are not random; they perform very differently in simulated data sets with similar statistical characteristics. These results demonstrate that the statistical analysis of newswire reports can be used to provide systematic early warning indicators and provides empirical support for the theoretical concept of distinct behavioral "phases" in political activity.

Introduction

In recent years, early warning has emerged as a major concern of national and international agencies that respond to political crises (Boutros-Ghali 1992; Davies and Gurr 1998; Dedring 1994; Esty et al. 1995, 1998; Mizuno 1995; Schmeidl and Adelman 1998), as well as a topic of renewed interest in the international relations literature (Gurr and Harff 1994, 1996; Rupesinghe and Kuroda 1992; Schmeidl, 1996). The policy community has been inspired by the success of famine forecasting models that have enabled international agencies to reduce famine-related fatalities by several orders of magnitude over the past twenty years (Rashid 1998, Whelan 1998). Agencies such as the UN High Commissioner for Refugees and the UN Office for the Coordination of Humanitarian Affairs would like to have comparable models for predicting other catastrophic situations—notably refugee flows—that require emergency humanitarian response. Academics have contributed to these efforts, as evidenced by the collaboration between operational agencies and university-based researchers in the State Failures Project (Esty et al. 1995, 1998) and the Forum for Early Warning and Emergency Response (http://www.fewer.org); academic scholarship has also been motivated by sobering assessments of the failure of traditional forecasting methods to anticipate the end of the Cold War (Gaddis 1992).

One of the central foci of event data research has long been the development of tools for the early warning of political change. Much of the initial impetus—and, quite critically, early funding—for event data research came from the United States Defense Advanced Research Projects Agency (DARPA). These efforts, including the Early Warning and Monitoring System (EWAMS), were contemporaneous with the creation of the World Event Interaction Survey (WEIS) and Conflict and Peace Data Bank (COPDAB) event coding frameworks as well as the development of a variety of systematic forecasting techniques (see Andriole and Hopple 1984; Choucri and Robinson 1979; Daly and Andriole 1980; Hopple, Andriole, and Freedy 1984; Hopple 1984; Laurance 1990; Phillips and Rimkunas 1983; Singer and Wallace 1979).

For both institutional and pragmatic reasons, these initial efforts failed to take hold in the policy community (Laurance 1990). By the mid-1980s, interest in, and funding of, statistical early warning projects had virtually ceased. The policy community continued to spend billions of dollars on analytical forecasts by intelligence agencies, but virtually all *political* prediction—as distinct from projecting economic and demographic trends—used traditional, rather than statistical, techniques.

Because human-coded event data were (and are) expensive to generate, the end of policy-oriented event data research meant that the development of publicly-available event data collections declined dramatically and the academic community was limited to re-analyzing legacy collections whose coverage terminated around 1977. The WEIS data set—the centerpiece of the DARPA efforts—continued to be maintained through a variety of public and private efforts through the 1980s and 1990s (Tomlinson 1993), but it was not widely available or used. In addition, several small, geographically-specific event data sets were created by scholars (for instance, Ashley 1980 and Van Wyk and Radloff 1993).

Ironically, the emphasis on event-based early warning research ended at the same time that two technological changes rendered the approach more feasible. First, the revolution in electronic communications made available a vastly increased amount of news about political affairs than was utilized by previous analyses, which usually relied on a small number of elite Western newspapers. By the early 1990s, much of this information was available in machine-readable form, initially through commercial services such as NEXIS and Reuters Business Briefing, and now through the World Wide Web. Second, the exponential increase in the computational power available to scholars allowed for the use of statistical and coding techniques that were impossible during the earlier period of DARPA research.

The purpose of this article is to describe several new approaches to early warning that utilize contemporary technologies and concepts. The paper starts with a review of the methodological challenges involved in statistical early warning research. We then apply two clustering methods—one based on the distances between points, the other on the changes in the density of the cluster—to an event data set covering international behavior in the Levant—Egypt, Israel, Jordan, Lebanon, the Palestinians, and Syria, plus the United States and USSR/Russia—between

April 1979 and December 1998. The Levant has long been regarded as an important example of an international subsystem (Binder 1958, Gause 1999); the relatively large number of actors and varieties of behavior in the region provide for a robust test of any early warning techniques. Our analyses indicate that events tend to form temporally-delineated clusters and that the movement of points in those clusters can be used as an early warning indicator; this finding is consistent with the theories of "crisis phase" that underlie many qualitative approaches to early warning. The clusters identified in our statistical analyses generally correspond to demarcations in the times series that we assigned *a priori*, based on a qualitative assessment of political changes in the region, and differ significantly from clusters found in a set of simulated data having similar statistical characteristics.

Statistical Approaches to Early Warning: A Review

In this section, we discuss some key differences between structural and dynamic models as applied to the early warning problem and describe the dynamic time series approach used in our research.

Structural and Dynamic Models

Statistical approaches to early warning can be classified into two broad categories: structural and dynamic.[1] The *structural* category includes studies that use events (or more typically, a specific category of events such as civil or international war) as dependent variables and explain these using a large number of exogenous independent variables. In the domain of domestic instability, this approach is exemplified by the work of Gurr and his associates, most recently in the "State Failure Project" [SFP] (Esty et al. 1995, 1998). Gurr and Harff (1996) and Gurr and Lichbach (1986) describe a number of such research projects. In the field of international instability, the structural approach is illustrated by the work of Bueno de Mesquita, as well as by the Correlates of War project; Gochman and Sabrosky (1990), Midlarsky (1993), and Wayman and Diehl (1994) provide general surveys. Structural approaches have tended to employ multivariate linear regression models. Recently, however, the research has branched out to other techniques; for example, the SFP uses logistic regression, neural networks, and some elementary time series methods.

In contrast to the structural approach, *dynamic* early warning models use event data measures as both the independent and dependent variables. Most of the event data projects of the late 1970s employed event-based indicators in dynamic models to predict whether or not a pair of political actors would become involved in a crisis. For instance, DARPA's EWAMS research evaluated three WEIS-based measures (conflict, tension, and uncertainty) to determine the “alert status” of any dyad. Azar et al. (1977) took a similar approach based on whether behaviors measured with the COPDAB event scale fall outside a range of "normal" interactions for the dyad. More recent efforts have utilized increasingly advanced econometric time-series methods that model interval-level indicators of events as an autoregressive time series with disturbances. Goldstein and Freeman (1990) provide a book-length example of this approach; Dixon (1986), Goldstein and Pevehouse (1997), Goldstein et al. (2000), Lebovic (1994), Ward (1982), and Ward and Rajmaira (1992) illustrate the continued development of this type of dynamic model.

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

The second argument for the dynamic approach is that it allows for greater parsimony in the specification of the forecasting models than a structural approach. The first phase of SFP, for example, collected data on 75 independent variables (Esty et al. 1995, 9). However, the final models developed in the project found that most of the forecasting power could be accounted for with only three variables: infant mortality, trade openness, and democracy (Esty et al. 1998, viii).[2] In contrast, the event data collections used in dynamic models allow the researcher to monitor political interactions as reported in public media sources. The focus on *events* leads to models that contain a relatively small number of independent variables and the reliance on news sources allows the data to be collected systematically in real-time.

Finally, a dynamic modeling approach assumes that the exogenous variables incorporated in the structural models do not need to be explicitly included because their effects will be reflected in the pattern of events prior to a major change in the political system. In other words, the dynamic approach uses the lagged values of actual events as a substitute for structural variables, as illustrated in Figure 1.

FIGURE 1 ABOUT HERE

To take a concrete illustration, Gurr (1995, 7) notes that "ethnic heterogeneity probably is most significant for state failure when it coincides with lack of democracy and low regime durability." Consequently, the SFP includes measures for those three variables: ethnolinguistic diversity, regime democracy, and regime durability. A dynamic approach, in contrast, would not measure these aspects of a political system directly, but would instead assume that each would be reflected in the types of events picked up by the international news media. The presence of democracy, for instance, would appear in the data not only in periodic reports of elections but in a large number of reports discussing disagreements between the government and the elected opposition. A low level of regime durability would result in coups and attempted coups. To the extent that ethnicity was an important political factor, one would find 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, therefore, dynamic modelers accept the importance of exogenous structural indicators: *Ceteris paribus,* countries with a high level of ethnic heterogeneity will have a greater propensity for conflict than those with a low level, democracies are likely to be different than autocracies, and so forth. The distinction between the two early warning strategies is one of *measurement*: structural modeling seeks to identify and measure the indicators directly, whereas the dynamic approach assumes we can indirectly incorporate the effects of structural attributes through the patterns of events they generate.

This is an optimistic, but not wholly implausible, assumption. For instance, in the Reuters-based domestic 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. Similarly, the existence of ethnoreligious conflict in Lebanon is one of the most conspicuous features of the data set and is in sharp contrast to the relative lack of salience of such conflict in Jordan. Our argument is that an increase in democratization in Syria or a decline in ethnoreligious tensions in Lebanon would be reflected in the political events reported in Reuters, although we have not analyzed this systematically.

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

Statistical Characteristics of the Early Warning Problem

The fundamental goal of time series is to determine the future values of a variable y given some present and past values of that variable and possibly the present and past values of a set of exogenous variables **X**. In other words, time series analysis seeks to determine a function

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

for some k>0. 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). Standard econometric time series methods, however, have only limited utility in the problem of early warning, for which the challenge is to identify a time T such that some indicator variable y_t has consistently different values (above a predetermined threshold Δ) after time T than it had prior to time T:

$$| y_t - y_{t'} | > \Delta, \quad \text{for all } t > T > t'$$

This is what would occur in aggregated event data following a shift in the type of political behavior in which a dyad was engaged. Figure 2 illustrates this situation for the transition in Israeli-Palestinian interactions that occurred with the outbreak of the Palestinian *intifada* in December 1987.

FIGURE 2 ABOUT HERE

An additional distinction between early warning research and conventional econometric time series methods relates to the amount of autoregression in the data. Econometric time series tend to be highly autoregressive: the value at time t is usually close to (or a simple function of) the value at time t-1. Examples of autoregressive series include unemployment figures, prices of consumer goods, and inflation rates. Even series that are potentially less stable, such as stock prices, usually have an autoregressive component combined with generally random noise. For instance, while the stock market crash of October 1929 was sudden, the high unemployment rates of the Great Depression required two or three years to develop fully.

Dynamic early warning models, on the other hand, focus on the shifts in a time-series that are *not* autoregressive, even though the series taken as a whole might be autoregressive. An autoregressive model of war-and-peace will be very accurate, as illustrated by the presumably apocryphal story about the European political analyst who said "Every day from 1910 to 1970, I predicted that Europe would remain at peace when at peace, and remain at war when at war, and I was only wrong four times." This type of model is not, however, very useful. It succeeds according to a *frequency*-based measure but fails according to an *entropy*-based measure that places higher weight on the prediction of low-probability events (Pierce 1980).

The econometric question most comparable to political early warning is the forecasting of sudden economic shifts such as those observed in massive exchange rate fluctuations (e.g., the collapse of the Mexican peso or the European Exchange Rate Mechanism). These problems are similar to political early warning in that they are primarily psychological and do not reflect a major change in the underlying physical reality: The economic fundamentals of the Mexican or European economies did not change dramatically during the days of the exchange-rate crises, but the *perceptions* of the future values of the relevant currencies did shift.

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 probabilistically. Coefficient estimation problems, in contrast, do not have answers: One can always specify an error structure, prior probability, or alternative model structure that places the estimated emphasis on different variables and there is no empirical method of deciding between these specifications. Second, and closely related to the first issue, forecasting problems are not affected by collinearity, which is the bane of coefficient estimation in the social sciences because every behavior tends to be linked to every other behavior. Coefficient estimates with low standard errors are clearly useful for obtaining a theoretical understanding of a situation, but they are not essential for the pragmatic purposes of forecasting (Wonnacott and Wonnacott 1979, 81). For this reason, it is not surprising that models with very diffuse coefficient structures that do not clearly identify the importance of specific variables—for example, neural networks, hidden Markov models, and vector auto-regression (VAR)—are found increasingly in early warning research.

Crisis Phase and Prediction

The concept of "crisis phases" is found frequently in the early warning and preventive diplomacy literatures; recent discussions of the approach include Alker et al. (n.d.), Carnegie Commission for Preventing Deadly Conflict (1997), Gurr and Harff (1994), and Lund (1996). In the empirical literature, crisis phase has been coded explicitly in collections such as the Butterworth international dispute resolution data set (Butterworth 1976), Bloomfield and Moulton's (1989, 1997) "Computer-Aided System for Analyzing Conflicts" (CASCON; http://web.mit.edu/cascon/) and SHERFACS, developed by Frank Sherman. In a review of these collections, 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 phase structures. (Sherman and Neack 1993, 90)

CASCON and SHERFACS, for instance, both code six phases: "dispute," "conflict," "hostilities," "post-hostilities conflict," post-hostilities dispute," and "settlement." A close analog to this approach is found in the DARPA-sponsored work of Phillips and Rimkunas (1983, 181-213), which analyzes WEIS data in a two-dimensional space of "threat" and "uncertainty" using Thom's (1975) "cusp catastrophe" model. Their study successfully locates eight of eighteen crises identified in the WEIS data and produces no false positives.

If the concept of crisis phase is valid, the international behaviors observed should fall into distinct patterns over time. Figure 3 illustrates this informally for the World War II period, using the two dimensions of "talking versus fighting" and "local versus global involvement." Politics in the years immediately prior to 1936 were predominantly local and involved little violent interstate conflict. The system shifted to a series of militarized crises between 1936 and 1938, then erupted into a full-scale European war in 1939. 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 differentiated 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 1946-48 cluster continued to represent the system for most of the Cold War, with occasional departures from that cluster to take in the Korean War, the Suez Crisis, the Cuban Missile Crisis, and similar events.

FIGURE 3 ABOUT HERE

Figure 3 is obviously idealized. Any analysis using event data will be complicated by the aggregation of dyadic behaviors, the existence of multiple issues determining behaviors, and the fact that real-world political behavior is considerably noisier than the simple summary of international politics in the 1930s and 1940s presented above. Nevertheless, if event data capture the behaviors characterizing a phase typology, it should be possible to determine those phases using statistical clustering. A cluster will occur whenever there is an extended period when the actors in the system are responding to each other in a consistent fashion (that is, repeating approximately the same types of actions—cooperative, conflictual, or absent—month after month). When the behavior of the dyads changes—for example, from peace to war or vice versa—the system shifts to a new cluster. If the transitions between these phases are gradual, or if interactions that precede a phase shift are distinct from those found when the system is locked in a single phase, then those behavior patterns can be used for early warning of the transition.

A Cluster-Based Approach to Early Warning

To test whether behavioral phase shifts can be used to develop early warning indicators, we analyzed the dyadic behavior of the international actors in the Levant between April 1979 and December 1998 using data created from Reuters lead sentences obtained from the NEXIS and

Reuters Business Briefing data services. We coded these leads into WEIS categories using the Kansas Event Data System (KEDS) machine-coding program (Gerner et al. 1994; Schrodt, Davis, and Weddle 1994). We then transformed the individual WEIS-coded events into a monthly *net cooperation score* for each directed dyad—the actions from a specific source to a specific target—by converting every event to an intensity score using the numerical scale in Goldstein (1992) and totaling these numerical values for each of the directed dyads for each month. We examined all the dyads involving interactions among Egypt, Israel, Jordan, Lebanon, the Palestinians, Syria, the United States, and the Soviet Union/Russia; this gives a total of 56 directed dyads with 237 monthly totals in each dyad.[4] Appendix A contains additional details on coding and measurement.

TABLE 1 ABOUT HERE

The behavior of this system can be seen by examining the position of the vector

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

where A, B, ..., H are the actors in the system and each pair of letters represents the total Goldstein-scaled net cooperation for a directed dyad aggregated over one month. The behavior of the system is simply the path that this vector traces over time in a high-dimensional space. In vector terminology, a "phase" is characterized by a region in the vector space where points cluster over time. If political behavior follows a phase typology, this would be evident by the system spending most of its time inside distinct clusters of behaviors that characterize the phases, with brief transitions between the clusters.

The benchmark we use in our empirical tests is a set of phases established *a priori* based on the dominant political interactions during each period. These phases, shown in Table 1, correspond to the interpretations of "area experts" and are consistent with our own fieldwork in the region. Our discussion of the results of the clustering and the early warning indicator will use these *a priori* clusters as a reference point.

The effectiveness of event-space clustering in early warning depends on whether some measurable characteristic of the behavior of the system changes *prior to* the phase transition. In some instances there are few if any precursors to a phase transition, either because of deliberate concealment by political actors or due to the lack of media interest, as in the examples of Chechnya and Somalia.[5] Our conjecture is that most political situations, however, go through a gradual deterioration (or improvement) of affairs over a period of a few weeks or months prior to a phase transition, rather than experiencing a sharp jump. Furthermore, because news-gathering organizations are usually rewarded for correctly anticipating political events, journalists who are present in the region, understand the local politics, and can get their stories past editors and onto the news wires are likely to report the behaviors they perceive to be pre-cursors to any political phase change.

This approach to early warning is similar to the "normal relations range" concept proposed by Edward Azar nearly 30 years ago:

> 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 interactions will cluster, with the diameter of the cluster a function of the “upper critical threshold.” Consistent with Azar’s treatment, we expect that when the system is near the edge of a behavioral cluster, it is in a crisis situation. That crisis will either result in a shift to a new cluster—a phase transition—or be resolved without a transition, in which case the system's interactions will return to the core of the cluster. Unlike Azar, however, we assume the system is moving away from normal behavior when it consistently nears (or passes) the edge of a behavioral cluster, rather than when it exceeds a single critical threshold. We have generalized 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), and we use a standardized metric based on correlation, whereas Azar used a Euclidean metric and established distinct critical ranges for each dyad. Finally, we also look at the *density* of clusters—defined as the average distance between the points in a cluster—over time. Behavior within the NRR should result in dense clusters, whereas when a system moves away from one phase/cluster/NRR and into another there will generally be a period when the points do not cluster densely.

Detection of Phase using Clustering over Time

Most widely-used clustering algorithms are cross-sectional and do not incorporate the time-series element of event data (see Aldenderfer and Blashfield 1984; Bailey 1994; Everitt 1980). A cross-sectional method will not necessarily produce credible time-series clusters: Because the Levantine subsystem as we have defined it does not include all relevant interactions—for example, the end of the Cold War—points that are distant in time may still bear a superficial resemblance to each other.

Including time as the dominant dimension actually simplifies the delineation of clusters. The new clustering algorithm we employ in this study is uncomplicated: a new cluster is established at a point x_t when the distance between x_t and the points before it is greater than the distance between x_t and the points after it. This can be expressed in the equation:

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

"LML" is the "Lagged distance Minus Leading distance;" x_t is the Goldstein-scaled net cooperation score for month t, $\|x_t - x_{t\pm k}\|$ is the distance between x_t and $x_{t\pm k}$ according to some metric, and Δ is a threshold parameter that prevents new clusters from being formed because of random fluctuations in the event data that are unrelated to phase transitions.[6]. (A large negative value of LML_t means that the point is still firmly in the current cluster: It is much closer to paints in the past than to points in the future. When the LML_t value is strongly positive, it indicates the point is far away from past points and close to future points.) From the perspective of cluster analysis, this approach is similar to the "minimum spanning tree" approach (see Backer 1995, Chapter 1) in dividing the clusters at places where a large distance is found between adjacent points; it differs in using the dimension of time rather than a graph to determine which points are adjacent.

In our analysis, "k" is measured in months, although in principle it could represent some other interval of time: weeks or even days. The threshold Δ is a free parameter whose value is dependent on the level of differentiation the analyst wants between the clusters (and the characteristics of the data being used to measure the political activity). For example, if Δ were set to a very high value, such as 0.8, we would see only two clusters in the data: pre-Oslo and post-Oslo. If it were set to a low value such as 0.15, we would find several additional sub-clusters within the *a priori* clusters specified in Table 1.

Figure 4 shows the results of analyzing our Levant data set using this algorithm for k=4 and the correlation metric

$$\|x_t - y_t\| = 1 - r_{x,y}$$

where r is the Pearson product moment. (In other words, this metric is based on the correlation between the Goldstein-scaled scores of the 56 dyads at two points in time.) The vertical lines on the graph correspond to time points where the *a priori* cluster divisions from Table 1 are located.

FIGURE 4 ABOUT HERE

A cluster division occurs when $LML > \Delta$. With the threshold level set at $\Delta = 0.30$, the LML measure, by increasing sharply at or very close to the demarcation lines, correctly identifies four of the seven phases determined *a priori*: Lebanon, Taba, *intifada*, and Oslo. In addition, several other plausible transitions are suggested by peaks that do not correspond to the original phases:

- a pre-Lebanon change, probably reflecting increased tension between Israel and the PLO in 1981;
- two pre-Taba changes that may correspond to shifts in Israeli, Syrian, and U.S. policy in Lebanon (including extensive diplomatic activity related to the May 17, 1983 agreement and its subsequent abrogation);
- a transition in January 1993 that may reflect the change in U.S. policy towards the Middle East that occurred with the shift from the Bush to Clinton administrations.

Finally, the analysis identifies March 1997 as the end of the Oslo period. This is reasonable. While the Netanyahu government came to power in the June-July 1996 period, it was nominally committed to continuing the Oslo peace process initially and the observed nine-month delay before the end of that process, while unanticipated, is in retrospect unremarkable.

Our LML measure misses two previously chosen transitions: the Kuwait war and the transition to the Madrid phase. Including the Kuwait transition as an *a priori* phase was probably an error. Although Iraq's invasion of the emirate and the subsequent war had profound implications for the Levant, the initial interactions characterizing this period came from actors outside the group we are studying, so it is not surprising that the data do not reflect it. It is less obvious why we fail to identify the subsequent Madrid transition. This might also be a side-effect of the Kuwait invasion. Because much of the behavior during late 1990 and 1991 was due to circumstances outside of the immediate region, the very real shift in politics that accompanied the Madrid process may have been less apparent in the data than were the other transitions.

Change in Cluster Density as an Early Warning Indicator

The “Lagged distance Minus Leading distance” measure cannot be used for early warning: It requires information from both before and after a phase transition has occurred and therefore can only be used to delineate clusters of behavior retrospectively. Consequently, in order to develop an early warning indicator, some alternative measure is required. An examination of Figure 4 shows that the leads-minus-lags (LML) measure often begins a rapid increase several months before a phase transition occurs. This is consistent with the underlying theory of phase transitions: The changing interactions in the system cause the points to pull away from the cluster center before a final break occurs, rather like what happens when pulling on a piece of taffy. This pattern suggests that a decrease in the *density* of the cluster—the extent to which the points within the cluster are close to each other—might serve as an early warning indicator. Since a change in cluster density can be determined solely on the basis of information available up to and including time t, it can be identified prospectively.

Figure 5 shows the behavior of such a "cluster density change" (CDC) measure. This measure is calculated by first computing the total distance between the points in a cluster of four consecutive months

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

and then calculating the difference between CD_t at points that are eight months apart:[7]

$$CDC_t = CD_t - CD_{t-8}$$

The cluster density change measure generally corresponds well with both the *a priori* phases and LML-identified transitions, despite the fact that the LML clusters were based on *post-hoc* information. At each point where the CDC exceeds a threshold set at one standard deviation (0.23), there is a corresponding LML cluster transition within a few months of the point (see Table 2). Unlike the LML measure, the CDC analysis does identify the Madrid transition—it exceeds the 0.23 level in November 1991 ($x_t = 0.24$) and December 1991 ($x_t = 0.27$)—but still fails to show the Kuwait transition, which arguably occurred due to factors exogenous to the system. A peak in the CDC measure in October 1989 ($x_t = 0.24$) probably corresponds to the decline of reports of activity in the Palestinian *intifada* (Gerner and Schrodt 1998); the peaks in early 1995 and 1996 may reflect changes associated with the problems encountered in the Oslo peace process prior to the cluster break identified in March 1997.

FIGURE 5 ABOUT HERE

TABLE 2 ABOUT HERE

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

This is most conspicuously the case for Lebanon in 1981-82. According to the CDC measure, the system shifted into the "Lebanon" phase about a year before the Israeli invasion in June 1982. When the invasion actually occurs, the CDC measure is at one of the lowest points seen in the entire time series. On the one hand, the Israeli policies that culminated in the invasion of Lebanon were put into effect well before the invasion, so placing the true phase shift in mid-1981 is politically plausible. It is widely accepted that Israel planned the invasion of Lebanon as much as a year in advance, then engaged in provocative military maneuvers throughout the first half of 1982 in an effort to goad the Palestinians into a response that would justify Israeli military intervention (Gerner 1994; Hiro 1992; Jansen 1982; Petran 1987; Randel 1983; Reilly 1983). On the other hand, the situation on the ground looked very different in July 1982 than in April 1982, a period during which the CDC measure plummeted. CDC is clearly not a "barometric" early warning indicator that allows a political analyst to say to his boss, "The CDC indicators are real low this month, ma'am: nothing to worry about ..." This may be because CDC is based on a correlation distance and is sensitive to changes in the configurations of policies—who is coordinating policy with whom—rather than to the direction of change. Using CDC as an early warning indicator in combination with a Euclidean measure sensitive to the direction of change might provide both types of information.

Our research generally supports the crisis phase model, but the measures are somewhat *ad hoc.* How do we know our findings are nothing more than a statistical artifact, reflecting a combination of chance and self-deception? In Appendix B we compare our results to those obtained with a randomly-generated simulated data set that has the same mean, variance, and autocorrelation as the observed data. The Monte-Carlo analysis of the LML and CDC measures found that these two indicators behave quite differently in the randomly-generated simulated data sets than in the observed data. This suggests our results are due to actual political characteristics in the region rather than simply an artifact of the clustering methods.

Summary of Findings

Table 2 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; when they do not, the differences are plausible. The LML cluster analysis identifies two phases that were not part of the *a priori* set. The first is an increase in tension between Israel and the PLO prior to the Lebanon invasion; the second is a post-invasion, pre-Taba period corresponding to continuing instability and the Israeli withdrawal from the area around Beirut. The end of the Oslo phase occurs about nine months into the Netanyahu administration; while we did not anticipate this particular date, it is quite plausible.[8] The CDC measure—although not the LML cluster analysis—indicates significant changes following the Oslo peace process phases. Based on CDC, we might also have designated a post-*intifada*, pre-Madrid cluster beginning in late 1989. All of our analyses missed the Kuwait transition.

The CDC measure usually provides two to six months of early warning; however, it gives no signal of the Oslo transition and no distinct alert before the June 1982 invasion of Lebanon. The

CDC measure also has some false-positives where it peaks just below the critical level. This is to be expected: Any measure that does not contain false positives is probably insufficiently sensitive to political events. For example, the several CDC peaks associated with the Oslo and Netanyahu periods are consistent with the on-again-off-again character of that negotiation. More generally, we have not included interactions between domestic actors—for example the various political parties and factions within Israel and within Lebanon—in this model, and sometimes those interactions may have been important. We are not dealing with a deterministic system and at times a false positive may reflect pre-cursors to transitions that failed to occur because of a reaction in the international system or in domestic subsystems that prevented the phase change.

The pre-Lebanon peak in LML may an example of this. In 1981, allies of Israel may have persuaded Prime Minister Menachem Begin 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 the complete destruction of Begin's political career. Only after another year had passed did the contrary and disastrous design of Defense Minister Ariel Sharon prevail.

Conclusion

With the end of the DARPA-sponsored early warning research, the development of quantitative early warning models went into eclipse. This is understandable. When evaluated against what is possible today, the DARPA efforts were necessarily primitive in their dependence on time-consuming and unreliable human coding and on mainframe computers that has only a tiny fraction of the speed and memory now available in a contemporary desktop machine. The event-based quantitative forecasting efforts of the late 1970s were unsuccessful, but then 1970s video games were not very impressive either.

The contemporary situation is quite different. The quantity, consistency and timeliness of event data have improved substantially due to the advent of machine-coded event data sets based on news wire sources. With increased computer power, and the greater sophistication of statistical research in international relations, a variety of new early warning techniques can now be applied to that data.

We draw two general conclusions from our analysis using time-delimited clusters. First, our empirical results support both the theoretical concept of crisis phases and the strategy of analyzing those phases by looking at the movement of a point defined by the vector of dyadic interactions in an international system. The pattern of variation in LML_t seen in Figure 4 is exactly what we expected the phase transition model to generate: brief periods of large movement followed by long periods of little movement.

These time-delineated clusters are *much* cleaner and consistent than clusters determined by the cross-sectional techniques such as the K-Means analysis found in Schrodt and Gerner (1997). The $LML_t > \Delta$ method used to delineate the clusters is conceptually straightforward and computationally efficient; in fact, the algorithm is sufficiently simple that it may be possible to determine analytically some of its statistical properties. Nonetheless, it locates most of the clusters we expected to find.

The CDC measure also appears promising as the basis of an early-warning indicator. The measure provides a two to six month warning for most of the changes in our data set and its behavior is consistent with the theoretical predictions of the crisis phase approach. Both the LML_t and the CDC measures act quite differently in data generated from real-world events than in simulated data, which suggests that they are reflecting underlying political behavior rather than statistical artifacts.

Time-delimited clusters are a dynamic rather than a structural early warning approach, but their effectiveness should not be regarded as evidence against the structural approach; we regard these two research strategies 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 about *why* a specific system is likely to experience problems. This might provide insights into the types of actions that could be taken to ameliorate an impending crisis. Structural models are unlikely to excel at predicting the exact timing of breakdowns, however, because the variables that they have identified as theoretically important change much too slowly. This is where dynamic models come into play.

In this analysis we have not considered an alternative class of dynamic models—those based on event sequences, rules, patterns, and precedents (see Cimbala 1987; Hudson 1991). These models generally provide a greater amount of contextual information than is supplied by the numerical time-series methods. As a consequence, they, too, may be useful in identifying the immediate events leading to a crisis. For instance, while the Kuwait transition is invisible in our cluster analysis, the events preceding Iraq's invasion of Kuwait follow Lebow's (1981) "Justification of Hostility" crisis-type very closely. Recognizing such patterns could be used for very short-term forecasting. An assortment of computationally-intensive non-linear forecasting techniques methods have also been developed in recent years (e.g. Casdagli and Eubank 1992; Richards 2000), although relatively little attention has been paid to these in the quantitative international politics literature. In short, there are still a variety of unexplored techniques 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. The optimal early warning model, for example, might vary depending on the structural precursors found in a specific case. Presumably the internal breakdown in a country such as 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 literature on domestic conflict and state breakdowns (for example Esty et al. 1995, 1998, Gurr and Lichbach 1986) could provide theoretical guidance on this issue.

The reason that such integrated models have not been developed to date is largely one of resources: the political science discipline is still developing accurate structural and dynamic models and at present no researcher has been able to assemble data sets sufficiently large to study structural and dynamic dimensions simultaneously. As the investigation of both types of models identifies more focused sets of variables and techniques, it should be practical to combine the approaches.

Appendix A
Data Sources and Measurement

Our source for the period 15 April 79 to 10 June 97 was the NEXIS "REUNA" file; Reuters Business Briefing (RBB) was used for the period 11 June 97 to 31 December 98. The change of sources was required because Reuters stopped supplying data to NEXIS on 10 June 97. The two data services provide a somewhat different mix of stories but we see no evidence of a significant discontinuity when the stories are coded and aggregated at a monthly level.

The following search command was used to identify relevant stories in NEXIS:

> (ISRAEL! OR PLO OR PALEST! OR LEBAN! OR JORDAN! OR SYRIA! OR EGYPT!) AND NOT (SOCCER! OR SPORT! OR OLYMPIC! OR TENNIS OR BASKETBALL)

To locate stories in RBB, we utilized the RBB search software (version 2.0 for Macintosh) to select "Political" and "General" news stories that dealt with Egypt, Israel, Jordan, Lebanon and Syria; the "Reuters Sports" source was explicitly excluded. (The "Israel" category includes stories on the Palestine National Authority as well as Israel.) Some additional filtering was done on both the NEXIS and RBB downloads to eliminate Reuters "Highlights," historical calendars and other irrelevant material.

The KEDS machine coding program does some simple linguistic parsing of the news reports—for instance, it identifies the political actors, recognizes compound nouns and compound verb phrases, and determines the references of pronouns—and then employs a large set of verb patterns to determine the appropriate event code. Details on the mechanics of the KEDS system are found in Gerner et al. (1994) and Schrodt, Davis, and Weddle (1994). For example, consider the following lead sentence:

> Palestinian President Yasser Arafat accused Israeli Prime Minister Benjamin Netanyahu on Tuesday of intentionally prolonging their peacemaking crisis.

The key components of the sentence are

Subject: Palestinian President Yasser Arafat

Verb: accused

Object: Israeli Prime Minister Benjamin Netanyahu

Based on the dictionaries used by the KEDS program, the proper noun in the subject—"Arafat" (or the adjective "Palestinian")—is assigned the actor code PAL and becomes the source of the event. The proper noun in the direct object—"Netanyahu" (or the adjective "Israeli")—is assigned the actor code ISR and becomes the target. The event is determined by the verb—"accused"—and is assigned the code 121 corresponding to the appropriate code in the WEIS event coding system. Bond et al. (1997), Huxtable and Pevehouse (1996), and Schrodt and Gerner (1994) discuss extensively the reliability and validity of event data generated using Reuters and KEDS.

We coded only the lead sentences of the stories; this produced a total of 92,687 events. The search command generates a number of events that are outside the 56 directed dyads considered in this study. Those 56 dyads contain 43,328 events.

The data and the source code for the computer programs used in the analysis, as well as the KEDS program (version 0.9B6.2) and the dictionaries used for this coding session, are available from the web site http://www.ukans.edu/~keds (15 May 2000).

Appendix B
Comparison with a Null Model

In this appendix, we develop a null model and look at the distribution of various indicators in simulated data generated by that model. We wish to determine whether a Monte Carlo analysis using randomly-generated data having the same means, variances, and autocorrelations as our Levant dyads will show the same patterns of change in distance that we found in the actual data. If this occurs, it would mean our results are simply an artifact of the analytical method, rather a reflection of the underlying politics.

The null model that we use duplicates the sample size (192) and number of dyads (54) found in an earlier version of the data set, as well as the mean, variance, and first-order autocorrelation of the data within each dyad.[9] Specifically, we generated simulated data using a first-order autoregressive (AR[1]) process

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

where $c = \mu\,(1-\rho)$; $\varphi = \rho$; $E(\varepsilon)=0$; $Var(\varepsilon) = s^2\,(1-\rho^2)$. As Hamilton (1994, 53-54) notes, this will produce a time series with mean μ, variance s^2, and first-order autocorrelation ρ. In order to avoid initial value effects, the simulated data were taken from the interval $[y_{51}, y_{242}]$ with $y_0=\mu$. To save computation time, ε were generated by random selection from a table of 5000 normally-distributed random variables produced by Excel 4.0. We created a sample of 1000 such data sets.

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

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

- The total number of points where $LML_t > \Delta$, with $\Delta = 0.2$, the threshold that best delineated clusters in this 1979-1996 data set.
- The number of $LML_t > \Delta$ points that signal a new cluster: this is defined (somewhat arbitrarily) as an $LML_t > \Delta$ point that had no $LML_t > \Delta$ points in the previous two time periods.[10] These times are called "cluster-defining points."
- The standard deviation of LML_t and the early warning measure CDC; the means of both measures are zero.
- The number of CDC measures greater than one standard deviation above zero at 0, 1, 2 and 3 months prior to a cluster-defining point.
- The number of $LML_t > \Delta$ points within 0, 1 ,2 and 3 months of the six *a priori* cluster transitions we identified in our data set, as a proportion of the total number of $LML_t > \Delta$ points.[11]

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

The results of the Monte-Carlo analysis are presented in Table A1, where the "one-tailed proportion" 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 more or less normally distributed; the probabilities are based on the actual distributions of the statistics in the simulated data rather than on a normal approximation.

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

TABLE A1 ABOUT HERE

The relationship between CDC 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 CDC_{t-k} may actually be a better early warning indicator than demonstrated in this data set, although its performance is due to autocorrelation in the data rather than to any more complex political characteristics that involve dyadic interactions.

The large number of $LML_t > \Delta$ points combined with standard deviations of LML and CDC that are higher in the simulated data than in the observed data suggests that the value of Δ—a free parameter that was established arbitrarily—may have been set too low for the simulated data. We re-ran the simulated data sets with Δ=0.35, a level of Δ that gives roughly the same number of cluster-defining points in the simulated data as were found in the observed data with 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 A2.

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

TABLE A2 ABOUT HERE

The results of the Monte-Carlo analysis are affected by the existence of the free parameter Δ, but in no circumstances do the results from the analysis of random data closely resemble those found in the real data. If we assume the $\Delta=0.2$ separation threshold, then the observed data has far fewer clusters than we would expect the null model to generate. By raising the level of Δ, we can match the number of empirically-determined clusters, but the behavior of the CDC statistic and the coincidence of $LML > \Delta$ 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 at which 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 the observed data actually settled into clusters and remained there for a period of time—as predicted by crisis phase theories—rather than jumping around. We suspect that the standard deviation of LML_t is lower in the observed data because of cross-correlation (and in a few dyads, higher-order autocorrelation) of the dyads.

Footnotes

[1] This discussion will not consider the large literature on non-statistical or qualitative approaches to forecasting: Contemporary surveys of these approaches can be found in Davies and Gurr (1998), Rupesinghe and Kuroda (1992), and Schmeidl and Adelman (1998). We also will not deal with long-range forecasting using computer simulation; Ward (1985) and Hughes (1993) summarize that literature.

[2] Phase II of the SFP involves some limited analysis of dynamic variables, and suggests expanding this approach in future studies.

[3] Until relatively recently, technical stock market analysis generally had a bad reputation, due to its use of statistically-dubious patterns based on small samples, wishful thinking, and gurus whose fortunes were based more on the sale of books than on trading stock. With the increase in computing power in the 1980s, the situation changed, and "programmed trading systems" can now process sufficiently large amounts of information to generate profits working solely with information endogenous to the market itself. The increased information processing capacity that exists today, in contrast to that available in the 1970s, may have a similar effect on the analysis of political behavior using event data.

[4] We repeated some of our analyses without the USA→USSR and USSR→USA dyads; this change generated only minor differences in the results. (The events involving these dyads include only the USA-USSR interactions reported in the lead sentences of Reuters stories dealing with the Levant, not USA-USSR interactions in general.)

[5] On the surface, the Rwanda genocide appears to be an ideal case where we would expect precursors to be absent. Yet in assessing this situation, Ruso (1996) observes that the appropriate facts were available (but incorrectly interpreted) as early the UN Rapporteur's Report in 1993 (Deqni-Ségui 1994). Ruso concludes:

> [T]he authors answer "Yes" to the question, "Did those with the capacity to prevent and mitigate the genocide have the information from which such a conclusion might be drawn?" In fact, they note that specific information about plans and conspiracies towards this end was picked up by the UN system, most significantly in the notorious "Black File" of January 1994. (Ruso 1996:8)

[6] These calculations were done with a simple (600-line) Pascal program that produced various tab-delimited files there were read into Excel to produce the figures.

[7] There are six possible comparisons within a cluster of four months— x_t to x_{t-1}, x_t to x_{t-2}, through x_{t-2} to x_{t-3}. Therefore the divisor six is used to compute the average. The lag of eight months was established empirically. In other regions of the world it might be shorter or longer depending on the time required for political decisions to be made. The choice of eight months is not critical. Because the CDC_t calculation involves data from two periods of four consecutive months, values of CDC_t at lags one or two months shorter or longer than the eight-month lag employed here will be strongly correlated with the eight-month value.

[8] This cluster break occurs near the splice between the NEXIS and Reuters Business Briefing (RBB) sources: the "leading" cluster for March 1997 contains two-and-a-half months of NEXIS data and one-and-a-half months of RBB. While it is possible that the change in data sources affects the delineation of the cluster, this seems unlikely for three reasons. First, this cluster shift is consistent with the changing policies of the Netanyahu administration. Second, the cluster break is anticipated by four months in the CDC measurements that contain no RBB data. Finally, most of the LML scores across the splice are quite low: Many are close to zero and only the March and April 1997 calculations show evidence of a shift in behavior.

[9] This analysis was done in April and May 1996. The simulations are quite time consuming; thus we have not rerun them with the data set that goes through December 1998. There is no reason to believe that the results would be any different using the newer data.

[10] 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 transition. This measure should also be less sensitive to the level of Δ.

[11] In the simulated data, these *a priori* transitions are arbitrary because they were determined by changes in the actual, rather than the simulated, data. Nonetheless, these indicators measure the likelihood of finding $LML_t > \Delta$ points in the vicinity of a set of six transition points spaced at the intervals in the *a priori* set. We look at the proportion because the number of $LML_t > \Delta$ points in the simulated data is substantially higher than in the actual data.

Bibliography

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

Alker, Hayward R., Ted Robert Gurr, and Kumar Rupesinghe, eds. N.d. *Journeys Through Conflict: Narratives and Lessons*. New York: Rowman and Littlefield. Forthcoming.

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* 10 (3/4): 293-309.

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

Azar, Edward E. 1972. "Conflict Escalation and Conflict Reduction in International Crisis: Suez, 1956." *Journal of Conflict Resolution* 16 (June): 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.

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

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

Binder, Leonard. 1958. "The Middle East as a Subordinate International System." *World Politics* 10 (3): 408-429.

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, MA: MIT Center for International Studies.

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

Bond, Doug, J. Craig Jenkins, Charles L. Taylor, and Kurt Schock. 1997. "Mapping Mass Political Conflict and Civil Society: The Automated Development of Event Data." *Journal of Conflict Resolution* 41 (August): 553-579.

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.

Carnegie Commission for Preventing Deadly Conflict. 1997. *Preventing Deadly Conflict.* New York: Carnegie Commission.

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

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

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

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

Davies, John L. and Ted R. Gurr, eds. 1998. *Preventive Measures: Building Risk Assessment and Crisis Early Warning.* Lanham, MD: Rowman and Littlefield.

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

Deqni-Séqui, R. 1994. *Special Rapporteur's Report on the Situation of Human Rights in Rwanda.* UN Commission on Human Rights.

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

Esty, Daniel C., Jack A. Goldstone, Ted R. Gurr, Barbara Harff, Marc Levy, Geoffrey D. Dabelko, Pamela Surko, and Alan N. Unger. 1998. *State Failure Task Force Report: Phase II Findings*. McLean, VA: Science Applications International Corporation.

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

Gaddis, John Lewis. 1992. "International Relations Theory and the End of the Cold War." *International Security* 17 (Winter): 5-58.

Gause, F. Gregory, III. 1999. "Systemic Approaches to Middle East International Relations." *International Studies Review* 1 (May): 11-32.

Gerner, Deborah J. 1994. *One Land, Two Peoples: The Conflict over Palestine*, 2d ed. Boulder, CO: Westview Press.

Gerner, Deborah J., and Philip A. Schrodt. 1998. "The Effects of Media Coverage on Crisis Assessment and Early Warning in the Middle East." In *Early Warning and Early Response,* ed. Susanne Schmeidl and Howard Adelman. New York: Columbia International Affairs Online, Columbia University Press.

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 (March): 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 (June): 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. 1997. "Reciprocity, Bullying and International Cooperation: A Time-Series Analysis of the Bosnia Conflict." *American Political Science Review* 91 (September): 515-530.

Goldstein, Joshua S., Jon C. Pevehouse, Deborah J. Gerner, and Shibley Telhami. 2000. "Dynamics of Middle East Conflict and U.S. Influence, 1979-97." American University. Typescript.

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

Gurr, Ted R., and Mark Irving Lichbach. 1986. "Forecasting Internal Conflict: A Competitive Evaluation of Empirical Theories." *Comparative Political Studies* 19 (April): 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 (July): 3-15.

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

Hamilton, James D. 1989. "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle." *Econometrica* 57 (2): 357-384.

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

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

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

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

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

Hughes, Barry B. 1999. *International Futures.* Boulder: Westview.

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

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

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

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

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

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.

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

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

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

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

Rashid, Abdur. 1998. "The Global Information and Early Warning System on Food and Agriculture." In *Preventive Measures: Building Risk Assessment and Crisis Early Warning*, ed. John L. Davies and Ted R. Gurr. Lanham, MD: Rowman and Littlefield. Pp. 185-193.

Reilly, James A. 1983. “Israel in Lebanon, 1975-82.” *Merip Reports* 108/109: 14-20.

Richards, Diana, ed. 2000. *Non-linear Models and Methods in Political Science.* Ann Arbor: University of Michigan Press.

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

Ruso, Sharon. 1996. "Early Warning: Much Ado About Nothing?" *Refuge* 15 (4): 8-10.

Schmeidl, Susanne, and Howard Adelman, eds. 1998. *Early Warning and Early Response.* New York: Columbia University Press-Columbia International Affairs Online.

Schmeidl, Susanne. 1996. "Selected Efforts/Research in the Area of Early Warning." *Refuge* 15 (4): 29-42.

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 (Fall): 561-588.

Schrodt, Philip A., and Deborah J. Gerner. 1994. "Validity Assessment of a Machine-Coded Event Data Set for the Middle East, 1982-1992." *American Journal of Political Science* 38 (August): 825-854.

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

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

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

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

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

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

Ward, Michael Don. 1982. "Cooperation and Conflict in Foreign Policy Behavior." *International Studies Quarterly* 26 (March): 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 (June): 342-368.

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

Whelan, William. 1998. "The U.S. Agency for International Development's Famine Early Warning System." In *Preventive Measures: Building Risk Assessment and Crisis Early Warning*, ed. John L. Davies and Ted R. Gurr. Lanham, MD: Rowman and Littlefield. Pp. 194-202.

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

Table 1

A Priori Phase Assignments

Label	Dates	Months	Defining Characteristic
Camp David	Apr 79-May 82	38	Before Israel's 1982 invasion of Lebanon
Lebanon	Jun 82-May 85	36	Israeli occupies much of Lebanon
Taba	Jun 85-Nov 87	30	Partial Israeli withdrawal from Lebanon
Intifada	Dec 87-Jul 90	32	Palestinian *intifada*
Kuwait	Aug 90-Oct 91	15	Iraq's invasion of Kuwait and aftermath
Madrid	Nov 91-Aug 93	22	Madrid process: Bilateral, multilateral peace talks
Oslo	Sept 93-Jun 96	34	Oslo peace process
Netanyahu	Jul 96-Dec 98	30	Netanyahu government in Israel

Table 2

Clusters Determined by the Analysis

Initial date of cluster	**Political characteristics**	***a priori* cluster?**	**LML cluster Δ>.30**	**nearest CDC 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	Partial Israeli withdrawal from Lebanon; increased conflict between Lebanese and international forces	no	yes	Apr-83
August 1985	Israel withdraws to south of Litani River; Taba negotiations	yes(1)	yes	Apr-85
December 1987	Palestinian *intifada* begins	yes	yes	Sep-87
August 1990	Invasion of Kuwait	yes	no	Oct-89(2)
December 1992	Madrid peace process	yes(3)	yes	Dec-91
November 1993	Oslo peace process	yes	yes	Oct-93
March 1997	Post-Oslo period	yes(4)	yes	Dec-96

Table Notes:

(1) The *a priori* cluster break was two months earlier, in June 1985
(2) This CDC 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) A break at some point during the Netanyahu government period was anticipated; the exact timing was not.

Table A1

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 proportion**
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 CDC		0.30	0.04	0.23	0.026 (<)
StDev of LML		0.25	0.03	0.15	0.001 (<)
Cluster break at t and CDC_{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 (>)

Table A2

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 proportion**
Total $LML_t > \Delta$		13.56	4.34	15	0.680 (<)
Cluster-defining $LML_t > \Delta$		8.48	2.49	9	0.660 (<)
Stdev of CDC		0.30	0.04	0.23	0.026 (<)
StDev of LML		0.25	0.03	0.15	0.001 (<)
Cluster break at t and CDC_{t-k}>Stdev	k=0	0.54	0.17	0.56	0.462 (>)
	k=1	0.31	0.16	0.22	0.731 (>)
	k=2	0.30	0.16	0.11	0.915 (>)
	k=3	0.28	0.15	0.11	0.903 (>)
$LML_t > \Delta$ within t±k of *a priori* break	k=0	0.03	0.06	0.07	0.247 (>)
	k=1	0.10	0.10	0.27	0.074 (>)
	k=2	0.16	0.13	0.40	0.054 (>)
	k=3	0.23	0.14	0.47	0.060 (>)

Figure 1. Comparison of Structural and Dynamic Approaches to Early Warning

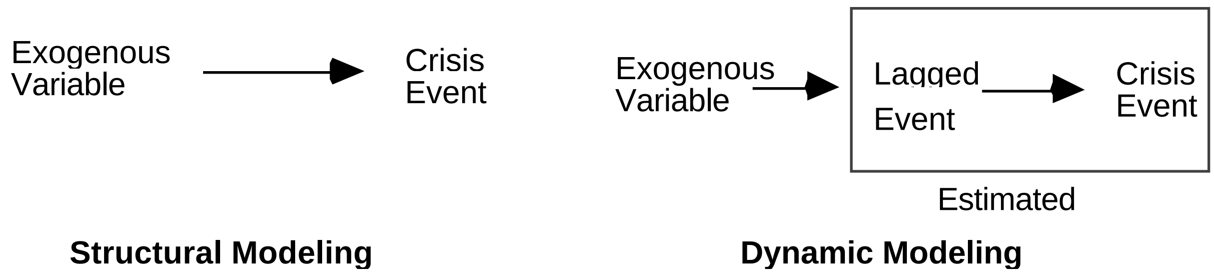
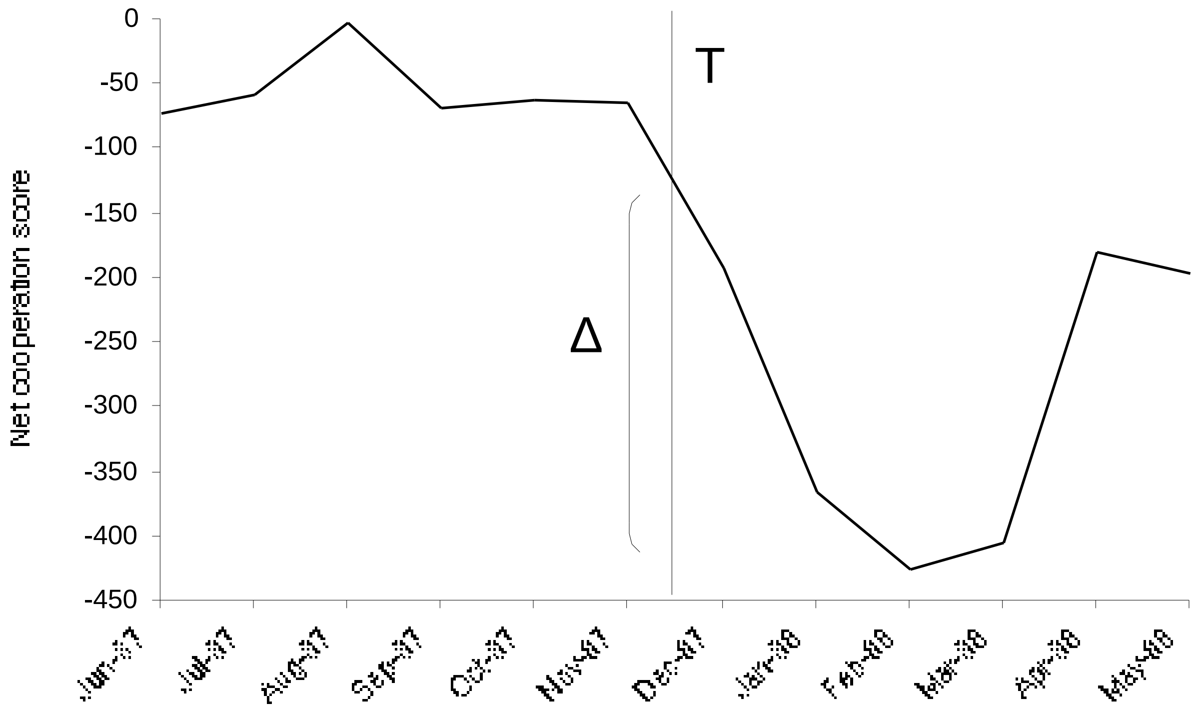
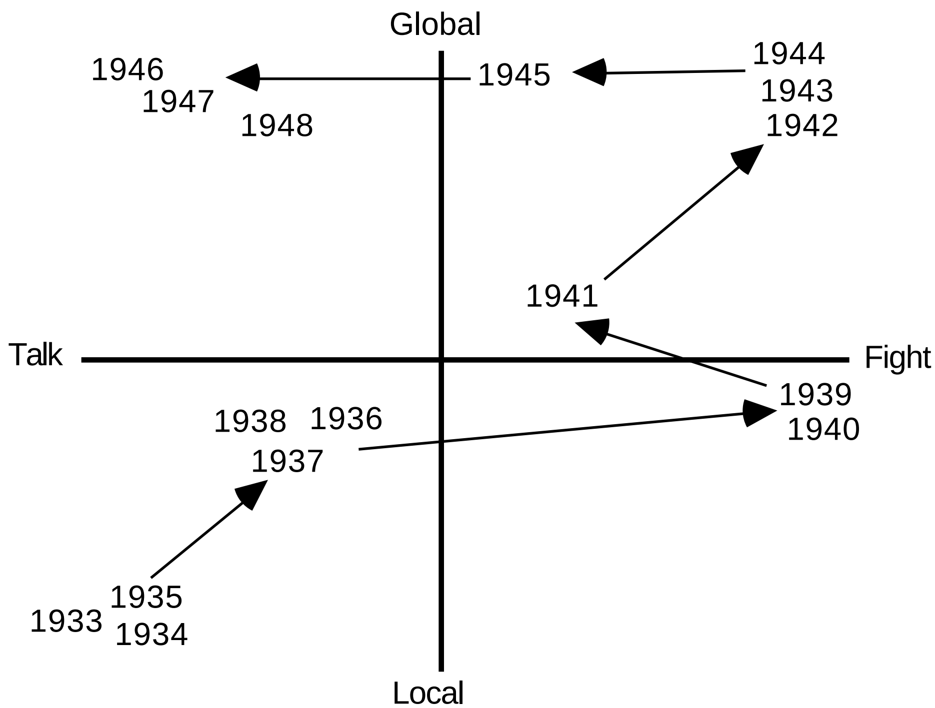


Figure 2. Example of a Shift in a Time Series at Time T: Values: Israeli Net Cooperation Directed to Palestinians, June 1987 to June 1988



Note to editors: Some of the labels in this figure printed poorly in MS-Word but are okay when the figure is printed directly from MS-Excel. We will produce the camera-ready versions from Excel.

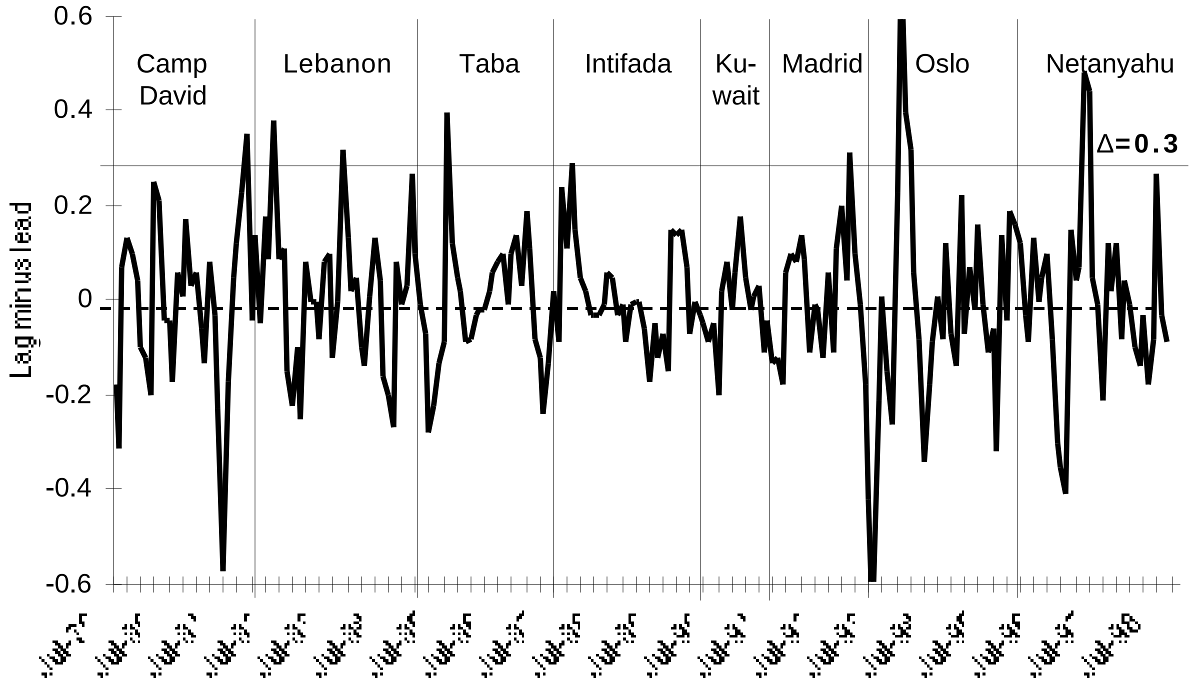
Figure 3. Informal Schematic Representation of Phases during the World War II Period



Note: This is an example of how behavior measured in two dimensions can cluster over time.

The arrows show the transitions between major phases of behavior.

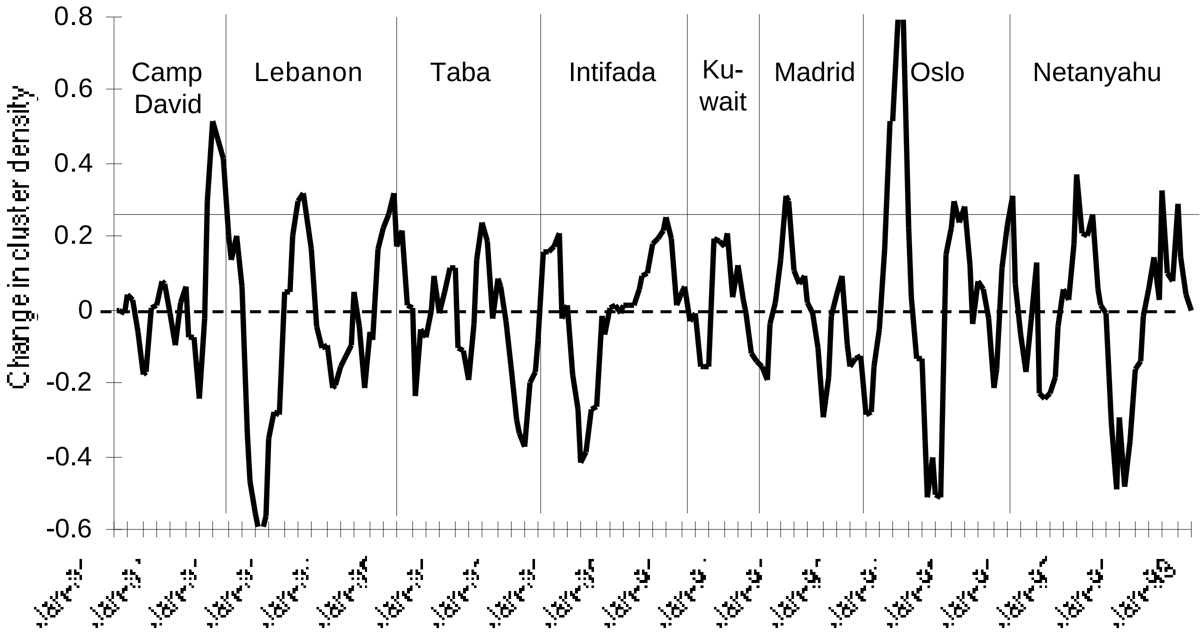
Figure 4. Four Month Lag minus Lead Measure of the Difference in the Distance of a Data Point to Points Preceding and Following It



Note: The Δ = 0.3 line shows the threshold that delineates most of the a priori clusters in the data. The distances are computed using a correlation metric. Vertical lines correspond to the a priori phase transitions identified in the Table 1. X-axis tick marks are 3-month intervals.

Note to editors: Many of the labels in this figure printed poorly in MS-Word but are okay when the figure is printed directly from MS-Excel. We will produce the camera-ready versions from Excel.

Figure 5. Eight Month Change in Four Month Cluster Density as an Early Warning Indicator



Note: The horizontal line shows the threshold that provides an early warning indicator for most of the a priori clusters in the data. Vertical lines correspond to the a priori phase transitions identified in Table 1. X-axis tick marks are 3-month intervals.

Note to editors: Many of the labels in this figure printed poorly in MS-Word but are okay when the figure is printed directly from MS-Excel. We will produce the camera-ready versions from Excel.