Chapter Four
Clustering Methods

This chapter will focus on the use on the use of cluster analysis to characterize international events, with an emphasis on early warning applications. As we discuss below, the DARPA-sponsored event data research made extensive use of clustering as a metaphor, but the computing power and software available in the early 1970s placed severe constraints on the actual application of statistical clustering methods, so these remained quite simple. In contrast, clustering methods can now be readily implemented on personal computers with multi-megabyte memories, even when large data sets are being used. That increased power has, in turn, also led to a number of developments in the broader clustering literature that can be applied to the early warning problem.

In this chapter, we first use factor analysis to determine whether it is possible to reduce the dimensionality of the vector describing the system by using the correlations between interactions of the various dyads. We then examine the clustering problem both deductively and inductively. In the deductive analysis, we begin by positing a set of behavioral phases in the Middle East political landscape during the 1979-1995 period, based on our contextual knowledge of the region. Using discriminant analysis, we ascertain the extent to which the dyadic behavior reflected by event data can predict these human-identified phases. Examination of the discriminant space also provides insights into the types of behaviors that are most important in determining the phase.

In the inductive part of the research, we attempt to discern system phases directly from the observed data by using cluster analysis (Aldenderfer and Blashfield 1984, Bailey 1994), rather than setting the phases *a priori*. In other words, we look for clustering in the data themselves rather than externally imposing any order upon them. Once we determine these clusters, we look at whether the system tends to remain in a cluster for a period of time, as postulated by the phase model, then try to interpret the political meaning of the clusters based on our contextual knowledge of the situation. We then use patterns of clustering over time to derive and test some early warning indicators that are sensitive to whether the system is ready to shift from one cluster to another.

The data used in this study include all interactions among Egypt, Israel, Jordan, Lebanon, the Palestinians, Syria, the United States and the Soviet Union/Russia (except for USA>USR and USR>USA) reported by Reuters News Service lead sentences downloaded from the NEXIS data service and coded with KEDS according to the WEIS coding scheme. This gives us a total of 54 directed dyads with 192 monthly totals in each dyad.[1]

4.1. Crisis and Clustering

Most of the quantitative DARPA studies worked with continuous indicators of crisis. For example, the DARPA-funded Early Warning and Monitoring System (EWAMS) system

> provides a comprehensive profile for the particular [dyadic] relationship in terms of conflict and other probabilities (for monthly data) and indicator readings (for total activity and other standard system indicators in their raw and transformed or Z-scores versions, for daily, weekly, monthly, quarterly or yearly time intervals selected by the user). (Hopple 1984, 52)

In contrast, a number of contemporary studies of crises assume that political behaviors go through a series of phases that are *qualitatively* delineated by an emphasis on different sets of behavior. In the statistical literature, crisis "phase" has been coded explicitly in the Butterworth international dispute resolution data set (Butterworth 1976), CASCON (Bloomfield and Moulton 1989, 1997) and SHERFACS (Sherman and Neack 1993). Describing the early CASCON work, Sherman and Neack explain that

---

[1] 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. Inclusion of these dyads in the data set makes relatively little difference in the analysis. For instance, the same five factors are found in the factor analysis and the variance explained by each factor changes by less than ±0.2%: Both of the dyads have their maximum loading on the "Camp David" factor with a loading of 0.336 for USA>USR and 0.505 for USR>USA. The accuracy of the discriminant analysis increases from 90.1% to 91.1% when the two superpower dyads are included.

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

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

In the policy literature, crisis phase has emerged as a key aspect of the "preventive diplomacy" concept, because of the assumption that diplomacy can be more effective in the early stages of a crisis (e.g., before the outbreak of military hostility) than in later periods (Rupesinghe and Kuroda 1992; Lund 1996; Bloomfield and Moulton 1997). To illustrate this argument, Lund (1996, 38-39) outlines a series of crisis phases ranging from "durable peace" to "war" and emphasizes the importance of preventive diplomacy during the "unstable peace" phase. In situations where preventive diplomacy is not an option, crisis phase may still be of utility in providing early warning of, for instance, large-scale refugee movements. Depending on the crisis phase, a localized outbreak of military action may be contained without generating large numbers of refugees, or it might rapidly spread, requiring the need for an international response. Finally, much of the literature on ethnic conflict assumes that militarized ethnic disputes such as those found in the former Yugoslavia, the former Soviet Union, Rwanda, Sri Lanka, and elsewhere evolve through a series of relatively predictable phases (Alker, Gurr, and Rupesinghe 1995; Leatherman and Väyrynen 1995).

This approach to early warning 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.

In the techniques developed in this chapter, we will 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. We also use a standardized metric based on correlation, whereas Azar used a Euclidean metric and established distinct critical ranges for the behavior of each dyad. 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.

One of the most substantial difficulties in using event data to analyze behavior in an international subsystem is the sheer quantity of information available. An eight-actor system such as the one we are analyzing has 56 directed dyads (the actions of X toward Y, indicated as X>Y), each with the potential to contribute to the overall interaction patterns of the system. However, much of this activity is inter-correlated because it is generated by a small set of political issues. In order to characterize the behavior of the system, therefore, it is necessary to ascertain the underlying issues that are generating the observed actions, then identify the contributions of various dyads to determining those behaviors. If the concept of crisis phase is correct, we would expect that the observed behaviors would fall into distinct patterns over time and that we could determine the event behaviors that are characteristic of any phase. In addition, if the behaviors determining a phase are distinct and are reflected in event data, it should be possible to identify those phases inductively by looking at the data themselves, without prior knowledge of the phases.

The approach we use to analyze the system is to examine the behavior 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 XY_t is the Goldstein score of events directed from X to Y, aggregated over a month. In other words, we converted individual X>Y events to a numerical score on the Goldstein scale, then totaled these numerical scores by month to obtain a net cooperation score for each directed dyad over time, producing time series data similar to that illustrated in Figure 2.x to 2.y.

The high dimensionality of this space makes it difficult to visualize. This is not a new problem in event data analysis: The response of most of the earlier event data studies was to ignore N-actor systems and instead either focus on a small number of dyads (e.g., Ward 1982; Dixon 1986; Goldstein and Freeman 1990) or examine the interactions of one actor with a number of states (e.g., Howell and Barnes 1993). In situations where there are a small number of clearly dominant dyads—for example, the USA-USSR-PRC triad of the Cold War—this is effective. Our research suggests, however, that in a complex system such as the Middle East, determining *a priori* the relative importance of various dyads is difficult and may distort the analysis.

We define the "behavior" of the system as the path this vector traces over time in an N-dimensional space. A crisis "phase" is indicated by a region in the vector space where points cluster over time. Empirically, a phase typology characterizes the behavior of the system provided that most of the time these behavioral points are found inside a few distinct clusters, with brief transitions between these clusters.

Figure 4.1 is an idealized illustration of this process for the World War II period, using just two hypothetical dimensions: "talking versus fighting" and "local versus global involvement." The years prior to 1936 involved little violent 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 conflict spread first to the USSR, then to the Pacific; 1942-1944 was characterized by a global war. In 1945, this war ended, first in Europe and then in the Pacific, but post-war politics remained global rather than returning to the unilateralism/isolationism of the pre-war period. The 1946-47 cluster continued to characterize the system for most of the Cold War period, with occasional departures from this cluster to take in the Korean War, the Suez War, the Cuban missile crisis, and other militarized events.

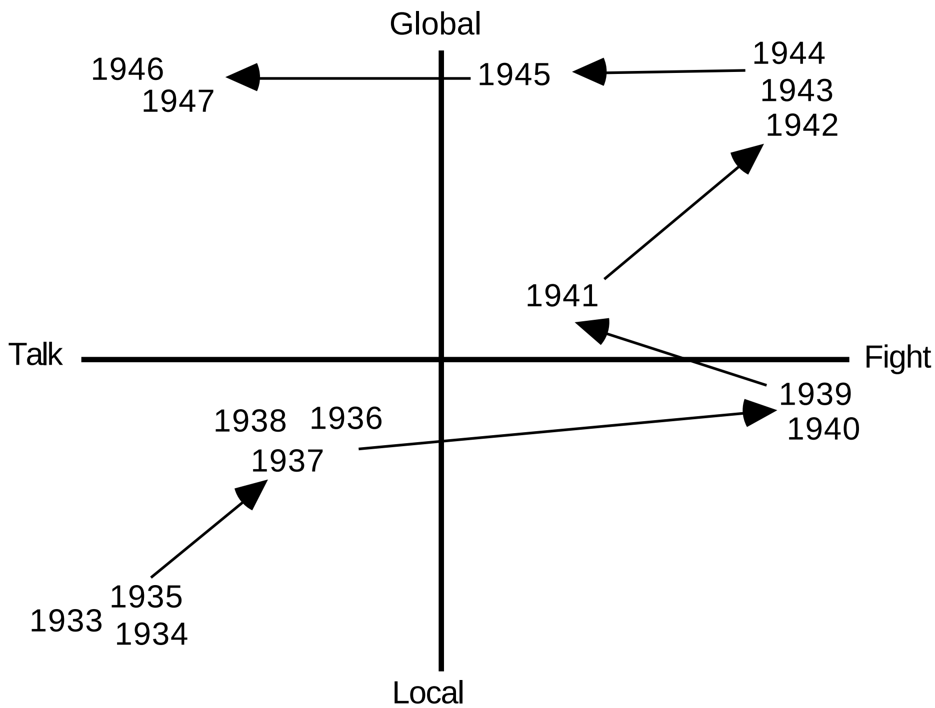


Figure 4.1: Hypothetical Representation of Phases during the WWII Period

Obviously an analysis using event data will be more complicated than this due to the problem of aggregating dyadic behaviors, the existence of multiple issues determining those behaviors, and the fact that actual political behavior is considerably noisier than the short-answer-exam summary of international politics in the 1930s and 1940s presented above. Nevertheless, if interactions in a system can be characterized by a phase typology, and if the behaviors defining those phases are captured by event data, it should be possible to determine behavioral phases using statistical clustering techniques.

The crisis phases identified in the Butterworth, CASCON, and SHERFACS data sets have all been assigned retrospectively by human coders. While this type of coding is obviously necessary in the early stages of a new concept's development, it presents two problems. First, when the coding of a crisis phase is dependent on human judgment, the *de facto* definition of the phase is likely to drift over time. This can happen as a single coder becomes more familiar with the data, or could occur during attempts to transfer the definition of a crisis phase across projects. Consequently, the crisis phases coded in two different data sets may appear to have conflicting implications because the coders were, in fact, working with disparate definitions. In contrast, the statistical identification of phases—combined with the machine-coding of event data—should make it possible to code crisis phase consistently and efficiently within a variety of contexts and from an assortment of different news sources.

Second, the tendency of human analysts to impose order on political events means that in some instances human coders may identify phases that do not correspond to information reported in the data set. If the human coder correctly identifies the phase, but unconsciously makes that assessment based on exogenous knowledge rather than the variables in the data set, any model that attempts to predict the phase using those data will be subject to specification error. Conversely, if the human coder has incorrectly identified the phase, any statistical estimates made with the data will be biased. We suspect that human-coded phase identification contains both types of error.

For the purposes of this analysis, we assigned the months in the data set to the phases identified in Table 4.1. These phases reflect the dominant issues and activities affecting the region and are not intended to parallel exactly the crisis phases of the CASCON and SHERFACS data sets. They do, however, mirror many of the characteristics of the CASCON and SHERFACS phase structures, notably the movement between periods of violent conflict and periods of dispute resolution.

Table 4.1. A Priori Determination of Behavioral Phases in the Levant Region, 1979-1995

	Label	Dates	Months	Defining Characteristic
Camp David	Jun.79-May.82	35		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 to south of the Litani until *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-Jun.95	22		Oslo peace process

4.2. Factor Analysis

The first question in our analysis is whether it is possible to decrease substantially the dimensions of the N-actor system's behavior without significant loss of information. Reducing dimensionality is possible when there are consistent correlations between some of the dyadic behaviors in the system. There were at least three substantive reasons we anticipated this might occur.

First, there is considerable policy coordination between some of the states. In the extreme case, since the late 1980s Lebanon's foreign policy has been greatly influenced by Syria. Lebanon's reported behaviors generally mirror those of Syria, so knowledge of Syria's position toward another actor may by itself provide sufficient information to predict Lebanon's policy toward that actor. A less extreme example is conventional policy coordination: For instance, during the Reagan years US and Israeli policies closely paralleled each other on most issues affecting the region, as did those of Syria and the USSR.

Second, all of the states in the system are reacting to at least some of the same events: Israel's invasion of Lebanon, Syria's eventual establishment of military hegemony in Lebanon, the Palestinian *intifada*, and the Madrid and Oslo negotiations, among others. To the extent that states share similar policy positions, they may react to these external stimuli in comparable ways.

Finally, it is likely that some of the actors in our set of states have very little influence on overall activity within the system. Two candidates are Egypt—which was diplomatically isolated during much of the period we are analyzing due to its signing of the Camp David agreements with Israel—and Jordan, which is comparatively small and neither initiates nor receives many events. It is possible that a simplified model of the system's behavior can ignore these actors without a significant decrease in predictive value.

While there are a variety of methods that can be used to reduce dimensionality through correlation, we focused on the oldest and most well-understood: factor analysis (Kim and Mueller 1978). Factor analysis creates clusters of variables (in this case, the vectors of monthly net cooperation scores for each directed dyad) based on their mutual correlations. Consequently, it allows us to evaluate both the extent of the correlation between dyadic behaviors and the extent to which the dimensionality of the system can be reduced. If foreign policies are primarily determined by exogenous events—whether the foreign policy of another actor or the interactions of other dyads (e.g., the Israeli-Palestinian conflict)—then those policies should show up as a distinct factor. Furthermore, the political content of a factor cluster should be apparent from the dyads with which it is most strongly correlated. Second, if a state is not actively involved in the system (or is uninvolved with some subset of issues), this will be evident from an absence of correlation between its behavior and the factors that dominate the system.

We factor analyzed the monthly aggregated scores for the 54 directed dyads using principal components, then rotated the original factors using the varimax criterion to minimize the number of variables that have high loadings on any given factor. In other words, varimax rotation tries to associate each variable with one and only one factor. All analyses were done using SPSS 6.1 for the Macintosh Power PC.

The results of the factor analysis are presented in Table 4.2, which shows the factors that have the five highest eigenvalues and the directed dyads with which those factors are most strongly correlated. As anticipated, the two factors that explain the highest amount of variance are those associated with (a) Israeli and Syrian involvement in Lebanon and (b) Israeli-Palestinian dispute. In the unrotated solution, the first two factors also emphasize the Lebanon and Israeli-Palestinian disputes but with less separation of the variables. The third factor appears to emphasize the dyads involved in the Camp David peace process, the fourth reflects the US involvement in the Lebanon dispute, and the fifth involves Jordan's interactions with Syria and Lebanon. This last factor is contrary to our expectation that Jordan is relatively unimportant in regional behaviors.

Table 4.2: Factor Analysis of the Monthly Aggregated Goldstein Scores

FACTOR 1
Lebanon Conflict: Regional

ISR>SYR	.77	-.04	.11	-.02	.08
ISR>LEB	.68	-.19	.27	-.11	-.10
LEB>ISR	.62	-.25	.25	.06	-.07
SYR>ISR	.61	.07	-.12	.07	.01
SYR>LEB	.61	.11	.04	-.05	.02
LEB>PAL	.56	-.01	-.08	.22	-.10
LEB>SYR	.47	.12	.05	.07	.31
PAL>LEB	.44	.20	-.10	.22	-.20
USR>ISR	.42	.02	-.08	.09	-.02
SYR>PAL	.22	.12	-.01	.03	-.21
USR>UAR	-.21	-.07	.04	.11	-.04

FACTOR 2
Israeli-Palestinian Conflict

PAL>USA	-.01	.67	.17	.06	.02
USA>PAL	-.07	.67	.40	.11	-.01
PAL>ISR	.17	.62	.33	-.21	-.02
JOR>ISR	-.03	.54	.00	.00	-.04
ISR>JOR	.07	.47	-.23	-.00	.08
USA>ISR	.08	.46	.24	.10	-.05
USA>JOR	.05	.45	.08	.03	-.00
ISR>PAL	.15	.43	.44	-.20	-.18
PAL>JOR	.09	.29	-.14	-.04	.02
JOR>USA	-.12	.28	-.17	.03	-.18
ISR>USA	.01	.27	.31	.01	-.05

FACTOR 3
Camp David

USA>UAR	.03	.00	.66	-.07	-.03
UAR>ISR	-.00	.06	.52	-.09	-.01
UAR>USA	-.23	-.04	.51	.02	-.18
USR>LEB	.13	.13	.50	.08	.09
ISR>PAL	.15	.43	.44	-.20	-.18
LEB>USR	-.08	.05	.41	.41	.08
ISR>UAR	-.00	.12	.37	.06	.08
ISR>USA	.01	.27	.31	.01	-.05
JOR>PAL	-.02	.15	-.22	-.07	.08

FACTOR 4
Lebanon: USA

USA>LEB	.09	-.18	.13	.76	.11
LEB>USA	.05	-.19	.19	.65	.08
USA>SYR	.10	.18	-.13	.64	.03
SYR>USA	.24	.21	-.27	.56	-.01
PAL>SYR	.04	.21	-.02	.41	-.02
LEB>USR	-.08	.05	.41	.42	.08
UAR>JOR	.07	.10	-.18	-.39	.07
USR>SYR	.07	.04	.10	-.26	.22

FACTOR 5
Lebanon: Jordan

LEB>JOR	-.00	-.04	.08	.07	.72
SYR>JOR	-.04	-.04	.03	-.05	.64
JOR>LEB	-.03	-.22	-.01	.12	.63
JOR>SYR	-.03	.06	-.14	-.07	.60
UAR>LEB	.03	-.07	.19	.01	-.45

Variance Explained by Factors:

F1:	7.7%
F2:	6.3%
F3:	4.9%
F4:	4.6%
F5:	4.3%

No loading above 0.20 on first five factors

USR>PAL	-.02	-.08	.10	-.01	.03
ISR>USR	.17	.08	-.04	.13	.12
JOR>UAR	.01	.09	-.15	-.09	-.00
JOR>USR	-.19	-.10	-.02	.14	.07
LEB>UAR	-.03	-.01	-.05	-.04	-.09
PAL>UAR	.04	.20	-.04	.06	.06
PAL>USR	.14	-.11	.08	.03	.13
SYR>UAR	.12	-.02	-.01	.07	-.01
SYR>USR	-.05	-.00	.11	-.04	.00
UAR>PAL	-.01	.15	.20	-.12	.03
UAR>SYR	.02	-.15	-.07	-.04	-.05
UAR>USR	-.15	-.07	.05	.13	-.15
USR>JOR	-.01	-.14	.07	.09	.02

SOURCE: Computation by the authors. Factor analysis of all directed dyads; varimax rotation. Variables are sorted by maximum loading; they are included in two factors if the difference between the two highest loadings is <0.05. Variables are included in a factor if the loading is >0.20 in absolute value.

In the varimax rotation, most of the dyads correlate strongly with only a single factor; the exceptions are ISR>USA and ISR>PAL (correlating with factors 2 and 3) and LEB>USR (correlating with factors 3 and 4). Thirteen dyads have no correlations of 0.20 or higher with the first five factors. All of these dyads involve either the USSR/Russia, which during the period studied was a relatively minor actor in the region, or—as we had anticipated—Egypt.

The first five factors explain only 27.8% of the total variance in the original data; by the usual standards of factor analysis, this is quite small. There are 21 factors with eigenvalues greater than 1.0 (the conventional rule-of-thumb for significant factors) and the factors beyond the five shown in Table 4.2 exhibit a long and very gradual decline in the variance explained. Most of those smaller factors appear to be picking up the idiosyncratic behavior of one or two dyads.

Based on this analysis—and contrary to our expectations—we conclude it is not possible to reduce the behavior of this system to a small number of factors. The high intrinsic dimensionality we find in the data may be due in part to the wide variety of behaviors shown in this long time series and the political diversity of the dyads. In some earlier experiments using only the dyads involving Israel and the Palestinians plus LEB>SYR, LEB>PAL, SYR>ISR and SYR>LEB as sources of behavior between 1982 and 1993, we found that four factors explained about 70% of the variance (Schrodt and Gerner 1995). Factor analysis may therefore be useful as a data reduction technique in systems less diverse than the one we've considered here.

4.3. Discriminant Analysis

Discriminant analysis is used to classify cases in a data set into a group of previously known, nominal categories based on linear combinations of the values of the interval-level independent variables describing each case (Klecka 1980). In the discriminant analysis in this study we use the behaviors of each dyad to classify the monthly data into the phases we assigned in Table 4.1. In other words, the "case" is a month of political activity; the "category" is the *a priori* assignment in Table 4.1 of each month to a phase, and the variables describing each month are the aggregated Goldstein scores reported for the 54 directed dyads. A month that shows activity typical a phase's general behavior will be located close to the center of the cluster containing the months of that phase—near the "centroid" of the phase in the discriminant space—while a month with atypical behavior (such as an outbreak of violence in an otherwise peaceful phase) will be located far from the centroid and will probably be misclassified into another phase.

Discriminant analysis is one method of determining whether the information required to describe the various phases of political activity is actually present in the event data set. If the analysis using the event data available for the 54 dyads cannot assign most months to the correct phase, then we must conclude that those human-identified phases were actually determined by information beyond what is available in the monthly event data, such as the hindsight of the analyst or trends aggregated over more than a month. (Alternatively, accurate phase classification based on event data might be possible through a nonlinear method such as an expert system, neural network, or genetic algorithm, but not through discriminant analysis.)

On the other hand, if discriminant analysis can classify most cases successfully, the characteristics of the discriminant space and the locations of the clusters within that space provide additional information about the data set as a whole. For example, the location of the centroids in the discriminant space usually provides some insight about the attributes that distinguish each phase from the others.

We first ran the discriminant analysis with all of the dyads; we then did a stepwise analysis to determine whether the phases could be identified by looking at only a small number of dyads. The results of these experiments are presented in Tables 4.3a and 4.3b; Figures 2a and 2b show the first three dimensions of the discriminant space.

TABLE 4.3a.
Discriminant Analysis using Monthly Aggregated Goldstein Scores for All Dyads to Predict A Priori Phase

Actual	Predicted: Camp David	Lebanon	Taba	Intifada	Kuwait	Madrid	Oslo	N
Camp David	**100.0%**	0%	0%	0%	.0%	0%	0%	35
Lebanon	2.8%	**86.1%**	5.6%	2.8%	0%	2.8%	0%	36
Taba	0%	0%	**86.7%**	0%	3.3%	10.0%	0%	30
Intifada	3.1%	0%	3.1%	**90.6%**	0%	3.1%	0%	32
Kuwait	0%	0%	0%	0%	**100.0%**	0%	0%	15
Madrid	4.5%	0%	4.5%	0%	.0%	**77.3%**	13.6%	22
Oslo	0%	0%	0%	0%	0%	9.1%	**90.9%**	22

Percent of cases correctly classified: 90.10%

Discriminant Function	Var Explained	Cumulative Pct	Wilks' Lambda	Signif
1	28.74	28.74	.024	<.001
2	24.44	53.18	.072	<.001
3	20.62	73.80	.196	<.001
4	11.12	84.92	.378	<.001
5	8.61	93.53	.650	.030

TABLE 4.3b.
Stepwise Discriminant Analysis using Monthly Aggregated Goldstein Scores to Predict A Priori Phase

	Predicted							
Actual	Camp David	Lebanon	Taba	Intifada	Kuwait	Madrid	Oslo	N
Camp David	**85.7%**	2.9%	8.6%	0%	2.9%	0%	0%	35
Lebanon	2.8%	**63.9%**	16.7%	2.8%	5.6%	8.3%	0%	36
Taba	10.0%	6.7%	**56.7%**	10.0%	3.3%	13.3%	0%	30
Intifada	0%	0%	9.4%	**68.8%**	12.5%	6.3%	3.1%	32
Kuwait	6.7%	0%	6.7%	0%	**73.3%**	13.3%	0%	15
Madrid	9.1%	4.5%	4.5%	9.1%	0%	**68.2%**	4.5%	22
Oslo	0%	0%	4.5%	0%	4.5%	13.6%	**77.3%**	22

Percent of "grouped" cases correctly classified: 70.31%

Discriminant Function	Var Explained	Cumulative Pct	Wilks' Lambda	Signif
1	32.95	32.95	.156	<.001
2	27.45	60.41	.311	<.001
3	19.97	80.38	.535	<.001
4	10.89	91.27	.746	<.001
5	5.54	96.80	.896	.006

As Table 4.3a indicates, when we include all of the dyads, the discriminant analysis differentiates the phases with a high degree of accuracy: 90% of the months are classified correctly and most of the errors are plausible (e.g., *Oslo* for *Madrid*; *Madrid* for *Taba*). This means we can successfully determine phase assignments from the patterns of monthly dyadic behaviors without additional information. Unlike the factor analysis results, the discriminant analysis concentrates most of the explanatory power in the first three dimensions of the discriminant space: these explain about 74% of the variance in the classification.

The stepwise discriminant analysis reported in Table 4.3b chose the following 12 dyads (in order of selection; the number in parentheses is Wilks' lambda[2]):

[2] The Wilks' lambda reported for each dyad is the proportion of the discrimination between the phases that is *not* explained by the dyads up to and including that dyad. For example, the ISR>PAL, LEB>ISR, PAL>UAR and ISR>JOR dyads together account for 75% of the discrimination between phases that is done by the model.

ISR>PAL (.636)	LEB>ISR (.427)	PAL>UAR (.324)
ISR>JOR (.250)	UAR>USA (.203)	PAL>LEB (.173)
JOR>ISR (.146)	UAR>ISR (.127)	SYR>LEB (.110)
UAR>JOR (.094)	ISR>LEB(.081)	JOR>USA (.071)

With 70% of the cases classified correctly, the stepwise discriminant is considerably less accurate than the system using all of the dyads, although 45.4% of the incorrect classifications occur into phases immediately before or after the correct phase. The stepwise selection appears to pick a single dyad to represent each type of dominant behavior in the system—for example, ISR>PAL is included but not PAL>ISR—and eliminates dyads that are correlated with those selections through reciprocity or coordinated policies. Unsurprisingly, nine of the dyads included in the stepwise discriminant analysis involve either Israel, Lebanon, or the Palestinians.

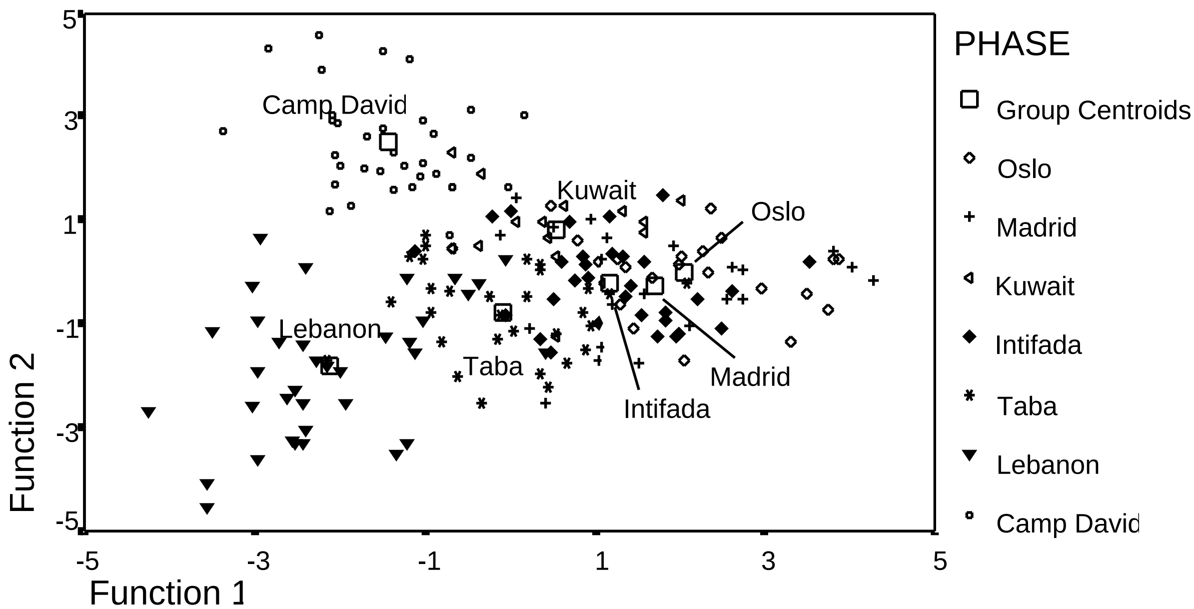


Figure 4.2: Discriminant Space—Functions 1 and 2

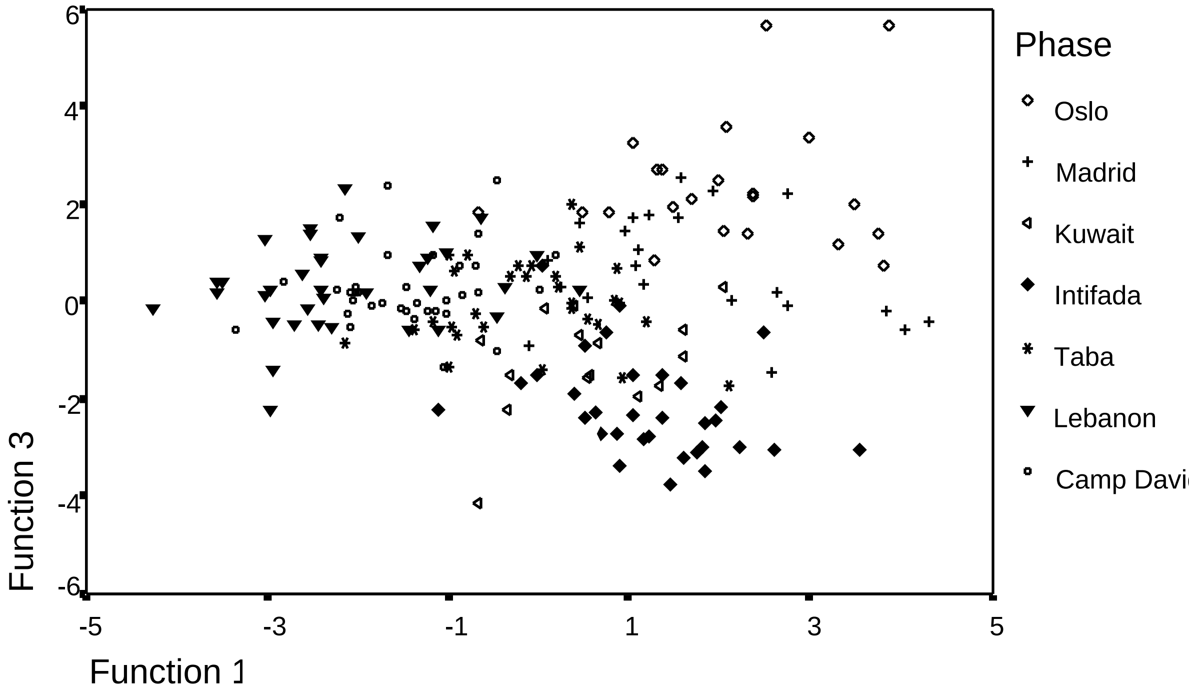


Figure 4.3: Discriminant Space—Functions 1 and 3

Figures 4.2 and 4.3 show the first three dimensions of the discriminant space; in both cases, the diagrams are based on the analysis of all 54 dyads rather than the stepwise analysis. In Figure 4.2, the first (horizontal) dimension discriminates the phases *chronologically*: the centroids are in temporal order from left to right, except for the swapping of *Camp David-Lebanon* and *intifada-Kuwait*. This is particularly interesting given that none of the variables contain explicit chronological information; the discriminant analysis is instead picking up the changes in behavior over time. The interpretation of the second (vertical) dimension in Figure 4.2 is less clear; it may serve primarily to differentiate the *Camp David* and *Lebanon* phases, which strongly overlap in the first dimension.

Because the *intifada* is one of the most conspicuous features of the data set, a puzzling aspect of Figure 4.2 is the location of the *intifada* cluster, which is thoroughly intermixed with the *Madrid* and *Oslo* clusters. As Figure 4.3 shows, this is an effect of the projection: the third dimension clearly separates the *intifada*, *Madrid*, and *Oslo* phases and generally seems to reflect a violent conflict versus conflict resolution dimension.

4.4. Cluster Analysis

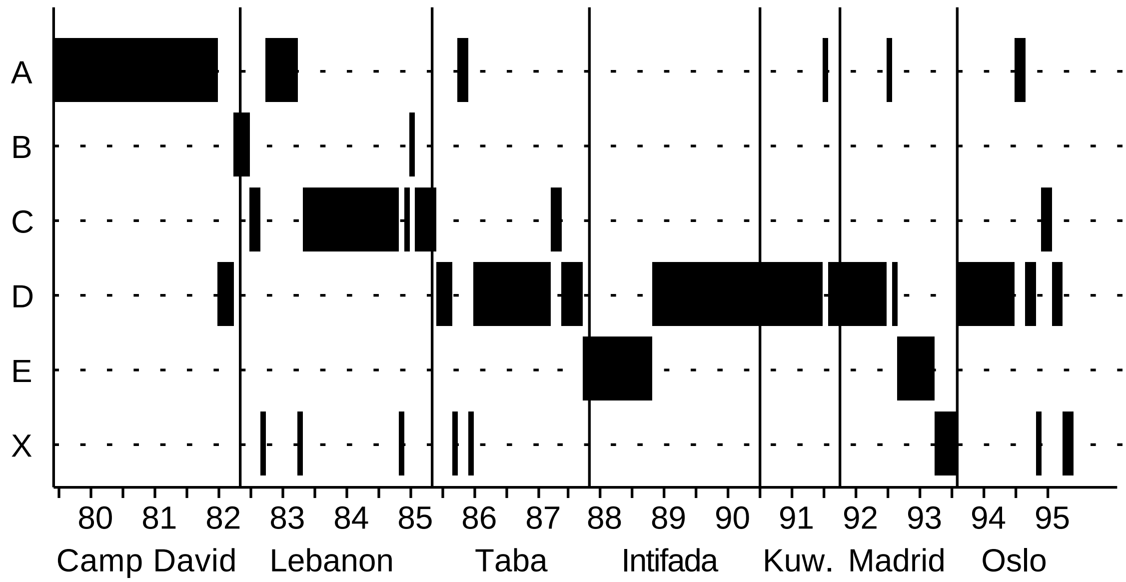
The discriminant analysis demonstrates that we can differentiate previously identified political phases with a high degree of accuracy using information about the dyadic behaviors in the system. This means that the phases are statistically distinct and their identification does not require additional information, such as hindsight bias, that is available only to the human coder. However, the discriminant results do not necessarily mean that the phases we have identified *a priori* are the same as those that would arise naturally from clusters of data points in the 54-dimensional space.

In order to determine whether clusters are actually present in the data, we used the SPSS K-Means clustering algorithm (Norusis 1994) with the Euclidean metric

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

as 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 seeds for generating the clusters. The algorithm then assigns each of the remaining cases in the data set to the cluster whose center is closest to the case. When this is completed, the center of each of the K clusters is recomputed and all of the cases are again assigned according to their proximity to the cluster centers. Because those centers are now computed from the assigned clusters rather than from the location of the seed case initially used to establish each cluster, this will cause some changes in the cluster assignments. This process is repeated until the cluster membership no longer changes; the algorithm usually converges after a small number of iterations.

Because K-Means starts with cluster centers that are widely separated, some of the initial centers will be outliers and their clusters will contain only one or two points after the iterations have been completed. Consequently, we used a two-stage process: In the initial iteration we used a relatively high number of centers (16), then identified the centers of the large clusters. These large clusters were quite distinct: the large clusters contained 20 to 50 points, whereas the remaining clusters contained fewer than five points. We then used only the centers of the large clusters as the starting points for a new clustering that assigned the outlying points to the large clusters. (We also allowed this assignment process to iterate but this had little effect on the final cluster centers.)



Note: Months in the "X" cluster are those for which a mode could not be computed.

Figure 4.4: K-Means Clusters for Monthly Aggregated Goldstein Scores for All Dyads, with Modal Smoothing

The results of the clustering using the 54 directed dyads is shown Figure 4.4. The five clusters identified by the K-Means procedure are labeled "A" through "E"; the presence of a bar indicates that a monthly data point was assigned to this cluster. The vertical lines show the phase divisions that we defined in Table 4.1. The graph has been smoothed using a four-month moving mode; the "X" cluster contains those points where the assignment is changing so quickly that the mode is undefined.[3] Note that because of the modal smoothing, an abrupt change of the form XXXYYYY (e.g., the *Camp David-Lebanon* transition and the *Taba-intifada* transition) shows up in the smoothed data one month before the actual transition.

Three of the clusters in the 54-dyad analysis correspond closely to the phases we had identified in Table 4.1. Cluster A corresponds to the *Camp David* period. It is quite uniform, although it shifts into a new phase several months before Israel's 1982 invasion of Lebanon, probably reflecting the political instability that preceded Israel's attack. This is followed by a short phase—Cluster B—that corresponds to Israel's initial invasion of south Lebanon and the siege of Beirut. The B cluster is found at only one other point, near the end of Israel's withdrawal to south of the Litani River in 1985. The system goes back into the Cluster A/*Camp David* pattern for most of the period when the US-led multinational force was in Beirut, then shifts into a new Cluster C pattern that is maintained until the end of the *Lebanon* period. The Taba period is dominated by a single cluster assignment, D, with a couple of jumps into clusters A and C. The *Taba* phase ends abruptly at the expected transition point corresponding to the outbreak of the *intifada* and a new cluster assignment, E, develops.

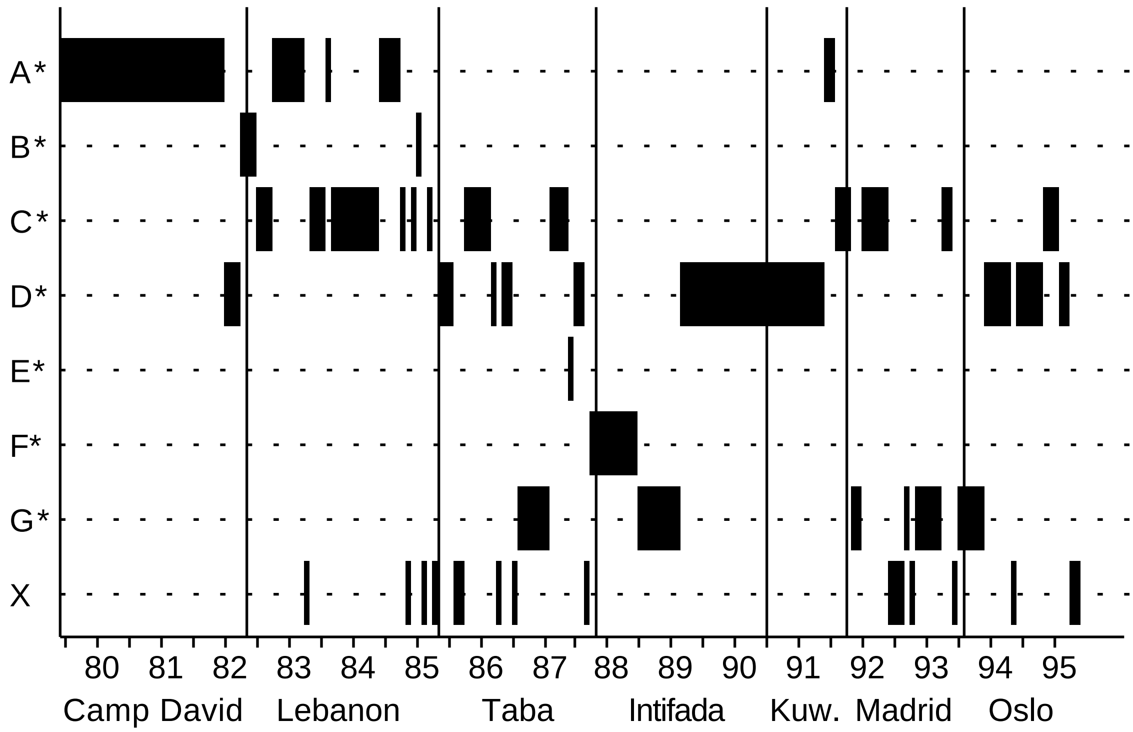
At this point, the phases in Table 4.1 and those identified by the clustering algorithm part company. The *intifada* phase determined by Cluster E is much shorter than we anticipated; after this, the system shifts back to Cluster D, seen earlier during the *Taba* phase. The Cluster D assignment is maintained without any break through the transition between the *intifada* and *Kuwait* phases. In the unsmoothed data, there are short-term changes around the remaining phase transition points we'd identified, but these periods do not form distinct clusters. The system

---

[3] Moving $\text{Mode}_t = \text{Mode}(X_t, X_{t+1}, X_{t+2}, X_{t+3})$ Ties were set to X_t if the tie was generated by the pattern XXYY and set to $\text{Mode}(X_{t-1}, \ldots, X_{t+4})$ for any other tied pattern. Cluster "X" contains those points where the expanded mode also produced a tie, where no mode could be computed because the 4-month interval contained four different clusters, or where the mode occurred in the pattern YZXX, i.e. the actual transition did not occur until t+2. Neither of these assumptions are critical to the smoothing. Graphs containing the unsmoothed cluster assignments can be found at the URL http://www.ukans.edu/~keds/jcr97 or obtained from the authors.

returns to Cluster E near the end of the *Madrid* period, which is consistent with an upsurge of violent incidents between Israelis and Palestinians during this time. The *Oslo* period is characterized by an unusually frequent pattern of cluster transitions, but does not form a distinct cluster.

We were puzzled by the inability of the K-Means algorithm to identify the Madrid and Oslo peace processes as distinct periods. We speculated this might have occurred because we initially defined our phases primarily with respect to the Israeli-Palestinian conflict, whereas the data set includes behavior from a broader set of countries. We therefore did an additional analysis looking only at the dyads involving Israel or the Palestinians as a source or target.



Note: Months in the "X" cluster are those for which a mode could not be computed.

Figure 4.5: K-Means Clusters for Monthly Aggregated Goldstein Scores for Israeli-Palestinian Conflict Dyads Only, with Modal Smoothing

These results are shown in Figure 4.5. The analysis found seven clusters, although neither of the two new clusters correspond to the *Madrid* and *Oslo* phases. Clusters A*, B* and C* delineate the *Camp David* period, the Lebanon invasion, and Israeli pre-1985 occupation of Lebanon in a manner similar to that of the earlier analysis. A single cluster (D*) extends from Spring 1989 until Fall 1992, again covering the *intifada-Kuwait* transition without a break, and Cluster F* distinctly delineates the intense initial months of the *intifada*, much as Cluster E had done in the first analysis.

There are two new clusters in Figure 4.5. Cluster E* is confined entirely to the *Taba* period—in the modal smoothing it is reduced to a single point but in the unsmoothed graph six dispersed months during the Taba period are assigned to this cluster. Cluster G* appears to correspond to intermediate levels of Israeli-Palestinian violence before and after the *intifada* cluster. The *Madrid* and *Oslo* periods still do not receive separate cluster assignments, although the unsmoothed behavior in these two phases jumps around between all of the other clusters rather than settling into a single pattern. In the unsmoothed data this instability is also found during the *Taba* phase; in fact, the cluster we have labeled *Taba* is probably a post-*intifada* phase that also characterizes the *Taba* period.

The only K-Means analysis we have done that successfully finds a "peace process" cluster used the 21 factor scores that had eigenvalues greater than 1.0. In that experiment, the *Camp David* period and the *Madrid-Oslo* period were each in distinct clusters (without a transition between *Madrid* and *Oslo*), but the intervening period did not show plausible patterns—in fact, the factor score clustering was the only analysis not to show any part of the *intifada* as a distinct phase.

The analysis presented here works in a continuous variable space by aggregating events using the Goldstein scale. As we discussed in Chapter 2, the Goldstein-scaled values generated from events coded from Reuters reflect events on the ground fairly well; at the same time it is unlikely that Goldstein scaling—which was developed simply to give a rough estimate of the magnitude of political behavior on a conflict-cooperation continuum—is optimal for deriving factors, clustering, or predicting phase transitions.

An obvious alternative to the Goldstein scale is to use the counts of the nominally-coded events themselves. The reason we have not done this is that it expands the dimensionality of the system's behavior by a factor of 22 if one aggregates to the level of 2-digit WEIS scores and by a factor of 66 if one goes to 3-digit WEIS scores. This would require analyses where the number of variables was considerably larger than the number of data points, not to mention the computational nightmare of dealing with 1188- or 3564-dimensional matrices, most of whose entries are zero.

Table 4.4: Factor Analysis of the Monthly Event Counts in the 2-Digit WEIS Categories

	ISR>LEB			ISR>PAL		
Factor	Eigenvalue	Variance Explained	Corr*	Eigenvalue	Variance Explained	Corr*
1	7.71	35.1%	-.71	5.59	25.4%	-.80
2	1.72	7.9%	.45	2.11	9.6%	.12
3	1.35	6.1%	.13	1.71	7.8%	-.12

* "Corr" is the bivariate correlation of the factor with the Goldstein score.

The Goldstein scale captures the most obvious feature of the event data—the conflict-cooperation dimension. In Table 4.4, we show the results of a factor analysis of the event counts in the 2-digit WEIS categories for two of the densest dyads: ISR>LEB and ISR>PAL.[4] There is a clear distinction between the amount of variance explained by the first factor and that explained by the remaining factors, and the correlation between the first factor and the Goldstein value is quite high. The results in Table 4.4 indicate two characteristics of the relationship between the aggregated Goldstein scores and the disaggregated frequencies of the WEIS event categories. First, over half of the variance in those counts is already reflected in the Goldstein score. Second, there is only a *single* conspicuous dimension to the event counts—one does not find, for instance, distinct and equally important dimensions for conflict and cooperation. These results suggest that an analysis of the event counts would probably not lead to dramatically different results than those found in the analysis of the Goldstein scores, at least for this international subsystem.

[4] Table 4 reports the first three factors; the analysis found 7 factors with eigenvalues greater than 1.0 of the ISR>LEB case and 8 such factors for the ISR>PAL case.

4.5. A Cluster-based Approach to Early Warning

Most of the clustering results in the K-Means analysis appear to be consistent with our assumption that a system will begin to show instability in the behavior space prior to experiencing a phase shift. In the analysis of the 54 dyads, only the *intifada-Kuwait* transition—which was completely exogenous to the Levant—does not show some fluctuation in phase assignments prior to the transitions. This instability is only a necessary, not a sufficient, condition, but as a precursor to the phase transitions it offers the possibility of some form of early warning. The *Madrid-Oslo* period in the Israeli-Palestinian clustering is characterized by almost continuous movement between clusters and it may be possible to use this information to characterize that phase.

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 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, 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 precursors to any political phase change. In short, the existence of precursors in an 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.

K-Means is a very general cross-sectional clustering method and 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.

Because the Levantine sub-system does not include all relevant interactions—for example, the end of the Cold War—the resemblance to earlier clusters may be superficial.

Including time as the dominant dimension actually simplifies the delineation of clusters in comparison to K-Means. The clustering algorithm we employ here 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.[5] 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. In other words, lagged distance minus leading distance, hence "LML." 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 tree to determine which points are adjacent.

Figure 4.6 shows the results of analyzing our Middle East data set (April 1979 through July 1996) using this algorithm for k=4 and the correlation metric:[6]

$\|x - y\| = 1 - r_{x,y}$ where r is the Pearson product moment

The vertical lines on the graph correspond to time points where the *a priori* cluster divisions from Table 1 are located. We also experimented with some larger values of k and the results are

---

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

[6] In Schrodt, Huxtable & Gerner (1996) we did this analysis with both the correlation and Euclidean metrics using an earlier data set that was generated without the complexity filter. In that analysis the correlation metric was clearly preferable to the Euclidean. Experiments reported in Schrodt & Gerner 1996 with our 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.

much the same as those obtained with k=4. If we set $\Delta = 0.30$, the correlation metric picks four out of the six the *a priori* phase assignments and 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;

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

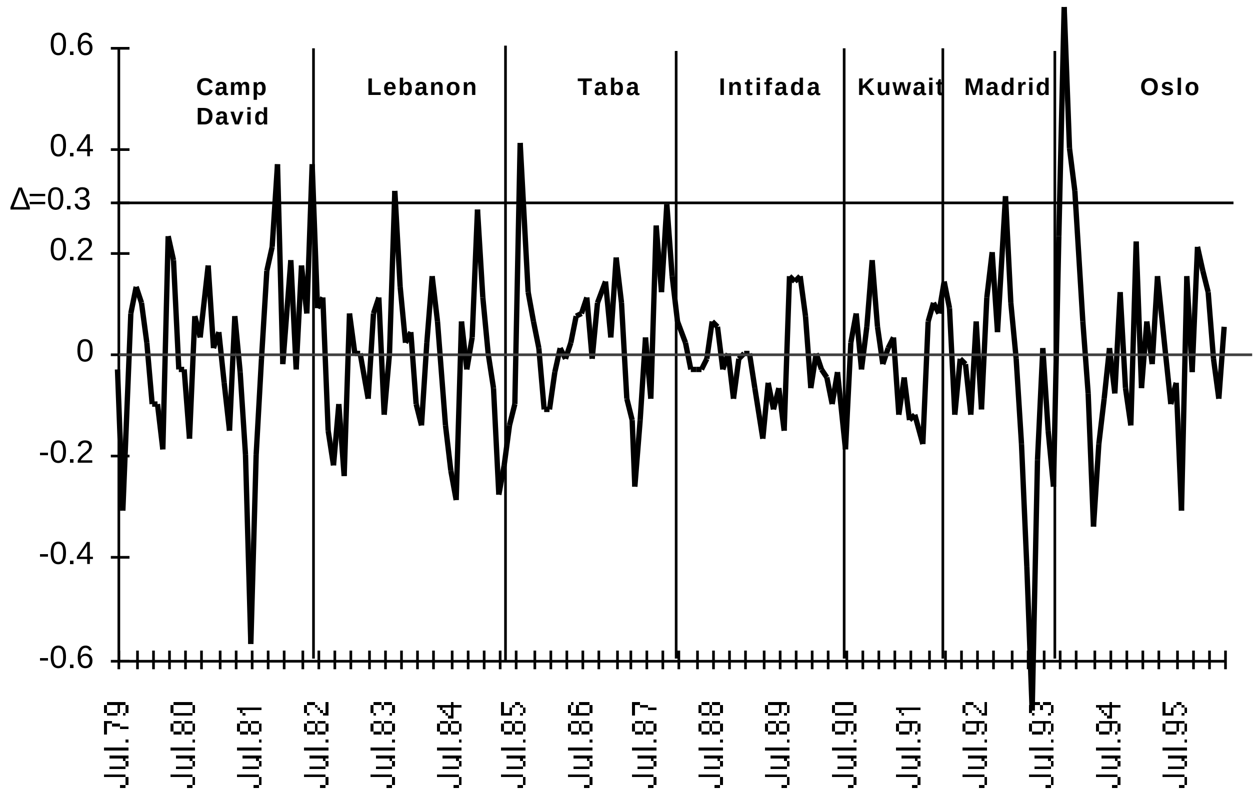


Figure 4.6: 4-month LML measure

4.6. Change in Cluster Density as an Early Warning Indicator

Examination of Figure 4.6 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 4.7 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 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, although 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*; (this is the "media fatigue" effect that is discussed in Gerner and Schrodt 1994). The peaks in early 1995 and 1996 probably correspond to changes associated with the problems encountered in the Oslo peace process.

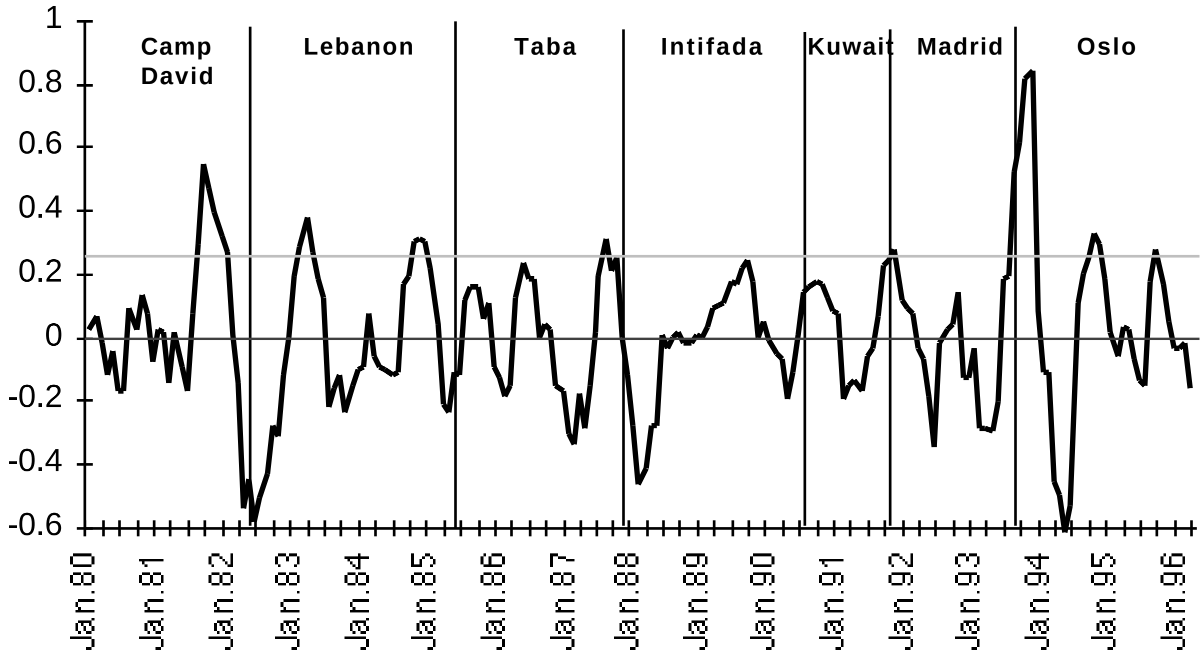


Figure 4.7: 8-month change in 4-month cluster dispersion as an early warning indicator

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.

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 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.

4.7. 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 the data set we analyzed in section 4.3 and 4.4, as well as the mean, variance, and first-order autocorrelation of the data within each dyad. Specifically, we generated simulated data using an 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 generate a time series with mean μ, variance s^2 and first-order autocorrelation ρ. (Autocorrelation above the first order is significant in only a small number of the dyads in the original data.) 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; to save computation time, ε were generated by random selection from a table of 5000 normally-distributed random variables produced by Excel 4.0.

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 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 $\Delta = 0.2$.[7]
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.[8] 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 $LML_t > \Delta$ 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

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

[8] 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.

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 4.5, and an example of the statistics generated by one such data set are shown in Figure 4.8. In Table 4.5, 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;[9] the probabilities are based on the actual distributions of the statistics in the simulated data rather than on a Normal approximation.

[9] 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.

Table 4.5

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>Δ		31.55	5.67	15	0.003 (<)
Cluster-defining LML_t>Δ		15.63	2.61	9	0.006 (<)
Stdev of $Δ_8$CD		0.30	0.04	0.23	0.026 (<)
StDev of LML		0.25	0.03	0.15	0.001 (<)
CDL at t & $Δ_8CD_{t-k}$>Stdev k=0		0.41	0.11	0.56	0.090 (>)
	k=1	0.22	0.10	0.22	0.461 (>)
	k=2	0.21	0.09	0.11	0.893 (>)
	k=3	0.20	0.09	0.11	0.869 (>)
LML_t>Δ 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 (>)

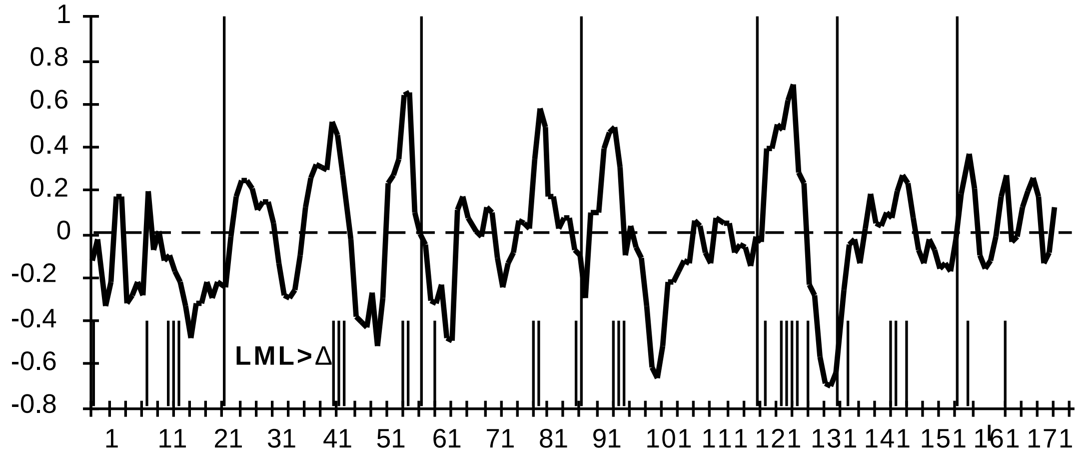


Figure4.8: LML>Δ and Δ_8CD statistics in a set of simulated AR[1] dyads.

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>Δ 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>Δ 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 4.6.

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 4.6

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

4.8. Conclusion

The results of these experiments are generally encouraging for the prospects of studying phase structures systematically using statistical clustering methods. Factor analysis appears to be the least useful approach, at least for this region of the world. The clusters of variables identified by the varimax rotation are politically plausible and correctly reflect what we view to be the two dominant political features of the data set (as well as identifying a politically plausible set of less-important dyads). However, the first five factors explain only about one-quarter of the variance in the data and a large number of additional factors would be required to explain additional variance. Consequently, factor analysis does not appear very promising as a data reduction method; one should instead use the original behavioral variables.

The results of the discriminant analysis, on the other hand, demonstrated that the event data coded from Reuters lead sentences were sufficient to distinguish our *a priori* phase assignments with a high degree of accuracy and, furthermore, the discriminant space involved politically plausible dimensions. The information required to identify these phases is, in fact, present in the event data. Unlike the factor analysis, the discriminant analysis quite substantially reduced the dimensionality of the behavior, focusing about three-quarters of the discriminating power into three dimensions of roughly equal importance. *If used judiciously*, stepwise discriminant might be helpful as a technique for reducing the number of dyads that need to be monitored in the system in order to assess its phase and provide warning of phase transitions. Because any stepwise procedure is highly affected—aye, thrives on—collinearity, we do not suggest simply accepting the stepwise results as given. However, stepwise discriminant analysis might provide a means of simplifying the data requirements of an event monitoring system if it was combined with additional knowledge such as information on the consistency and reliability of the newswire reports used to generate the event data for various dyads.

Despite being a purely inductive method that used only aggregated political behaviors, the K-Means clustering identified several of the same phases we originally assigned to the data, or else identified plausible alternative phase transitions (e.g., an earlier end to the *intifada* phase than we predicted and the "sub-*intifada*" phase in the Israel-Palestinian analysis). The K-Means results also support the proposition that there will be instability in the cluster assignments prior to a shift between phases.

In short, the behaviors recorded in Reuters-based event data—at least in a well-covered area such as the Levant—are sufficiently detailed to discriminate correctly periods of time into the phases that were determined *ex post facto* by human analysts. Given the validity and reliability problems of event data, and the coding ambiguities and hindsight bias that affect human coding of phases, this conclusion was by no means guaranteed. To a more limited extent, the phases determined inductively by cluster analysis also correspond to the phases evident to a human analyst, although in this data set those phase assignments have much greater correspondence in the periods involving military and civil conflict than they do in periods involving negotiation. Because of this, we suspect that studies using crisis phase (whether human- or machine-coded) as an independent variable (for example, in assessing the effects of multilateral intervention on crisis outcome) will not be severely affected by specification error.

The K-Means analysis ignored a critical characteristic of the data: events occur over time. The two time-dependent clustering measures, LML_t and Δ_8CD, explicitly use the dimension of time as an element in determining clusters. The pattern of variation in LML_t seen in Figure 4.6 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. The $LML_t>\Delta$ method used to delineate the clusters is conceptually simple and computationally efficient; in fact the algorithm is sufficiently simple that it may be possible to determine analytically some of its statistical properties. The Δ_8CD measure also appears promising as the basis of an early-warning indicator.

Table 4.7 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. Based on Δ_8CD, we might also have designated a post-*intifada*, pre-Madrid cluster beginning in late 1989. All of our analyses using the Goldstein weights missed the Kuwait transition, although the weights found by the genetic algorithm and the default vectors usually detected either the invasion or the subsequent war.

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 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.

Table 4.7. Clusters Determined by the Analysis

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

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

To the extent that the Middle East is typical, this analysis has at least three implications for the use of event data in early warning research. First, it is clearly possible to analyze the behavior of an N-actor system taken as a whole, rather than looking only at individual dyads or small sets of dyads. This can be done either with readily-available and well-understood statistical techniques such as factor analysis, discriminant analysis and K-Means, or with specialized methods such as the LML_t and Δ_8CD measures. The ability to analyze N-actor systems is particularly important once one begins to consider sub-state actors such as political factions and ethnic groups. Coding sub-state actors is relatively straightforward with automated coding and the techniques presented here should work as well with sub-state actors as they have with international actors. It would be interesting, for instance, to look for phase shifts in the Lebanon conflict using data that differentiates the activities of the various political groups within Lebanon.

The fact that we were generally unsuccessful in reducing the dimensionality of the system also argues for focusing on the analysis of N-actor systems rather than looking only at representative dyads or factor scores. The clusters identified in the analysis of the factor scores are generally less stable and less plausible than those we found using the Goldstein scores. Part of the problem with generating political plausible interpretations of the factor scores may be that the factors are orthogonal (uncorrelated): This is useful for statistical purposes such as regression but makes the factors very difficult to interpret politically because political interactions are generally *not* orthogonal—events in Lebanon continually affected Israeli-Palestinian interactions and vice versa. The Levantine political subsystem is admittedly quite complex but in this respect it may be more typical of the situations that will be encountered in the post-Cold War world than were the bipolar and tripolar systems that have been the primary focus of most event data analyses in the past.