

# **Events, Patterns, and Analysis: Forecasting International Conflict in the Twenty-First Century**

**Devika Subramanian**  
Department of Computer Science  
devika@rice.edu

**Richard J. Stoll**  
Department of Political Science  
stoll@rice.edu

**Note: both authors are also affiliated with the Center for the Scientific  
Study of International Conflict and Cooperation  
January, 2003**

**Rice University  
Box 1892  
Houston, TX 77251-1892**

**The research discussed in this paper is funded by the National Science  
Foundation, and INTEL.**

## **Introduction**

It seems like a lifetime ago, but if one looks back to the heady days after the fall of the Berlin Wall, and the subsequent demise of the Soviet Union, there was certainly a great deal of speculation that the world in general, and the United States in particular, would experience a significantly lower amount of conflict than was the case during the Cold War. In many ways, the world is a safer place than it was during that era, but a quick glance at the experience of the United States since the end of the Cold War will demonstrate that conflict has been an important part of US foreign policy:

- The United States has engaged in three significant military conflicts: the Gulf War, the NATO campaign against Kosovo, and the conflict against Afghanistan that led to the overthrow of the Taliban.
- The United States has been the target of a number of significant terrorist attacks, culminating in the events of September 9, 2001. This prompted the President to begin a world-wide campaign against terrorist groups, most prominently Al Qaeda (including but not limited to the above mentioned campaign against Afghanistan).
- As of the end of 2002, there was intense speculation that the United States, either as the leader of a multi-nation effort, or by itself, would invade Iraq. Some would argue that the follow-up to the Gulf War began shortly after the end of the war, with the establishment of the two no-fly zones over Iraq; these zones feature frequent combat between US and British aircraft, and Iraqi ground forces.

- At the same time, North Korean actions that facilitated a re-starting of its nuclear weapons program created a situation (the Bush administration denied it was a crisis) that might involve the US in a significant military conflict.

While this list includes the most important conflicts experienced by the United States, it by no means exhausts the significant conflicts throughout the world. Even if we restrict our attention to situations in which the current United States administration has taken a direct interest, at a minimum we would have to add the Israeli-Palestinian and Indian-Pakistani conflicts to the list. In short, despite the end of the Cold War, serious international conflicts – unfortunately - still exist.

A glance at the US conflict involvements shows that there are significant differences. The Gulf War featured both an air campaign and a ground campaign fought according to the doctrines of the US Air Force and the US Army. Kosovo is regarded by many as the first war decided by airpower alone. In the campaign in Afghanistan, a combination of indigenous opposition groups, selective air strikes, and US ground forces prevailed.

While there were major differences in how each of these conflicts was prosecuted, there is one thing they have in common. Despite the unparalleled intelligence capabilities of the US government, to a significant extent, each situation was unanticipated by US decision makers.

If the onset of serious conflicts can surprise the US government, one can only assume that individuals, groups, organizations, and other states throughout the world (with fewer intelligence resources) are even more unlikely to anticipate conflicts. If

conflicts cannot be anticipated, what chance do we have to prevent or stop them?

Consequently, we see many uses for an increased ability to anticipate the outbreak of serious conflict.

This is the goal of our research: To improve the ability to anticipate serious international conflict. We will use a combination of online media sources, analytic techniques, and knowledge derived from research in international conflict to achieve this goal. Since we are at the beginning stages of our research, we are not in a position to present enough information to allow for a final judgment of our approach. However, we can provide enough of a descriptive to allow the reader to develop a good sense of what we propose to do, and to draw a preliminary conclusion about its worth.

## **Conceptual Foundations**

Any research in international conflict is guided by a conceptualization of the process by which it develops. Some researchers make this conceptualization very explicit through the use of formal theory. There are other researchers (usually traditionalists) who claim that they look at “everything,” but this is misleading. They are guided by informal (and often unrecognized) conceptualizations. We believe that it is always better to make these conceptualizations explicit. So before we describe the plan of our research, we need to explain the conceptual foundation of our work. For purposes of our discussion, we will assume that there are only two countries involved in the conflict. This is merely to simplify the verbiage; our work can accommodate a larger number of actors.

We begin by introducing the concept of a *conflict of interest*. This is a situation in which each of two countries<sup>1</sup> seeks to influence the other to change or modify its position on at least one issue. Three conditions are necessary for these countries to have a conflict of interest:

1. There must be at least one issue over which the two parties have different preferences.
2. The two countries must both be aware of these different preferences.
3. Each party must decide that it is worthwhile to try to influence the other.

These conditions seem simple, even obvious, but they are important. If two countries have exactly the same preferences on all issues, there is no need for either party to try to influence the other, since they are in total agreement. Obviously, such complete and total agreement between two countries (or two individuals for that matter) is highly implausible. It is likely that for any pair of countries, there are a number of issues over which they have different preferences. But unless both parties are aware of these differences in preferences, they cannot consider doing anything about it.<sup>2</sup> Finally, even if two countries are aware of differences in preferences, they may not consider that it is worth an effort to influence one another. One can imagine that the governments of Great Britain and the United States have different preferences about which country's soccer team should win the next World Cup. But it seems highly unlikely that either government would try to influence the other to change that outcome.

Once an influence situation develops, we expect that both countries will take actions. Actions can be words, deeds, or a combination of the two. States have a variety

---

<sup>1</sup> Note: we use the words "country," "state," "party," and "actor" interchangeably in our discussion.

<sup>2</sup> Think of two roommates meeting for the first time. At that point they do not know each other well enough to be aware of differences in preferences. That will come later.

of instruments at their disposal; diplomatic, economic, etc. Our research is restricted to influence situations in which military force is seriously contemplated, or actually used. That is, we are not trying to deal with the entire gamut of activities that comprise foreign policy.

Once the use of military force is considered, the influence situation has become a militarized interstate dispute (Jones, Bremer, and Singer, 1996). Disputes can be “resolved” in a variety of fashions. For example, some sort of written settlement that resolves the discrepancy in preferences between the parties might occur. On the other hand, both countries may simply cease to use military instruments without either side changing its preference ordering. Finally, the dispute may escalate to a war. A war is continuous combat in which each state tries to inflict a military defeat on the other.

We see conflict as a series of steps or phases (Bremer, 1996). There is an implied sequence as countries move through a conflict of interest, a dispute, and a war. But not every conflict of interest represents a significant danger to the parties involved. Often the closest of countries go through periods of time that are characterized as “rocky” or “stormy” but despite this, the use of force may never be considered a viable instrument of policy by both sides. These situations are of little interest to us. But it does illustrate one of our challenges; to successfully separate those conflicts of interest with no potential for violence, from those that will spill over into increasing levels of violence.

This conceptual framework has a number of implications. First, conflicts of interest are *about something*. Countries do not try to influence one another, even to the point of using military instruments, unless there are one or more important issues at stake. This seems like a simple point, but it is important. We often have a tendency to think of

conflict as “caused” by things such as the capability balance between two sides, but we need to remember before governments consider the chances of winning a military confrontation, they have to have a *substantive reason* for that confrontation.

Second, we do not expect the countries to be omniscient, all-knowing, and comprehensively rational actors. We do expect them to behave in a limitedly-rational, goal seeking manner. Their decisions will not be perverse or random. They try their best, given the constraints of the situation, problems with obtaining and correctly evaluating information, and limitations of time, to achieve their goals. But we do not expect them to make optimal decisions.

Third, in deciding on future actions, a state takes into account a variety (but a limited one) of considerations. It considers the recent actions of the other party (not individual actions, but rather an aggregation of recent events), its own recent actions, and a series of factors that play a significant role in evaluating which state is likely to prevail if the contest turns on the relative military capabilities of the two sides. The state forms a summary evaluation that combines these factors together in a fairly simple fashion.

Finally, the confluence of these assumptions will produce a flow of events. These events, when aggregated at the appropriate level, will provide a good trace of the ebb and flow of the interactions between countries. An analysis of these (aggregated) events, along with appropriate other factors will allow us to predict the onset of serious conflict.

### **Forecasting International Conflict: the Building Blocks**

The primary building block of our research is the event:

...an interaction, associated with a specific point in time, that can be described in a natural language sentence that has as its subject and object an element of a set of *actors* and as its verb an element of a set of *actions*, the contents of which are transitive verbs.

(Gerner, Schrodt, Francisco, and Weddle, 1994: 95)

The process by which events data are created is discussed below. Our analysis will not rely on individual events, but on *aggregations* of them, for reasons we will identify below. Given an aggregate set of events data, we need tools to analyze them. We will be using wavelet analysis as our primary tool. Finally, as a way to increase our understanding, we will move beyond the forecasting of serious international conflicts, to modeling the factors that play a critical role in the escalation and de-escalation of conflict.

Collecting Events Data. In the social sciences, critical information for research rarely appears in a useful form. Instead, social scientists must engage in the process of *data making* (Singer, 1965). Events data -- nominal or ordinal codes recording the interactions between international actors -- are a prime example of this. Events data are a common type of information used in quantitative international relations research. A single record or case of events data consists of the actor (the entity that initiated the action), the action itself, the target of the action, and the date of the action. Most collections also associate a scale score with each event; this score represents the degree to which the event is cooperative or conflictual.

Until recently, the collection of these data was only possible through the use of teams of trained human coders who read through media sources to extract the appropriate

information. This approach was both slow and expensive. Consequently the most widely used events data collections, the Conflict and Peace Data Bank (COPDAB; see Azar, 1980) and the World Event Interaction Survey (WEIS; see Tomlinson, 1993) both cover only a limited period of time in the post-World War II era.

We propose to automate the extraction process. This process will require a variety of state of the art tools. Some of these are already developed. One important tool is the set of software programs developed by the Kansas Events Data Project (KEDS -- see Gerner et al., 1994), although we will also explore using similar tools that have been developed in the field of artificial intelligence (Craven et al., 1999). The creation of events data is basically a process of content analysis and involves three steps:

1. A source or sources of news about political interactions is identified. This could be an internationally-oriented newspaper such as *The New York Times*, a set of regional newspapers and newsmagazines, a news summary such as *Facts on File* or *Deadline Data on World Affairs*, or a newswire service such as Reuters, the Associated Press, or *Agence France Presse*.
2. A coding system is developed, or a researcher may decide to use an existing coding system such as the World Events Interaction Survey (WEIS), the Conflict and Peace Data Bank (COPDAB), or the Conflict and Mediation Event Observations (CAMEO) systems. The coding system specifies what types of political interactions constitute an "event," identifies the political actors that will be coded (for example, whether non-state actors such as international organizations and guerrilla movements will be included in the data set), specifies the categories of events and their codes, and specifies any information to be coded in addition to the basic event. For example, the

COPDAB data set codes a general "issue area"-- whether an action is primarily military, economic, diplomatic or one of five other types of relationship. WEIS, in contrast, codes for specific "issue arenas" such as the Vietnam War, Arab- Israeli conflict, and SALT negotiations. CAMEO was developed in part to evaluate the role of mediation in international conflict, and its event categories reflect that theoretical perspective.

3. In a machine-coding project, coding rules are implemented in a computer program such as Tabari (the program currently being used at the KEDS project) or WebKB (Craven et al., 1999) by using extensive dictionaries (corpora) that identify actors and events. Corpus-driven techniques are more robust than methods that attempt to parse and understand natural language sentences. Instead, corpus-driven techniques take advantage of implicit regularities in the structure of news sources. This allows us to create coding rules from samples of suitably marked up documents (Craven et al., 1999). Such an approach eliminates the need to manually formulate extraction rules.<sup>3</sup>

KEDS has already been used to produce an extensive set of events data on various regions of the world (see for example, Schrodts and Gerner 1994, 1997), and it will be an important part of our effort to convert contemporary information into data. But we will also need to gather information from sources other than those that have been utilized previously by KEDS. We will extend the scope of KEDS-like data gatherers to more text sources on the internet. In this effort we will take advantage of the growing number of open media sources on the internet. These sources contain not only current stories, but

---

<sup>3</sup> Researchers (Schrodts and Gerner, 1994; King and Lowe, 2002) have shown that machine-based events coding is as accurate as human coding.

also archives of past stories, allowing the possibility of updating the classic events data collections to the present.<sup>4</sup>

There are several issues that have to be addressed before we can proceed with the events data collection. First, which sources should we use? Media sources vary widely in their completeness and quality of reporting. We need to find a systematic way to pick the best set of sources. A second issue is the question of duplicate events. Different sources may report the same event in slightly different ways, so it is critical to find a reasonable way to eliminate duplicates.<sup>5</sup>

A third issue is aggregation of events. For several reasons, we believe it is unwise to analyze the events data as individual events. First, we believe that such a large number of events would contain a large amount of noise. This of course is a judgment on our part and other researchers may disagree.

But we believe there are other problems with using individual events in analysis. There is really no good way to order events within a day. Often news stories do not contain enough information to order events through time. Of course some stories will contain the time that the action took place. But it is less likely that the story will contain information as to when the *target* knew the act took place, and without that information, it is hard to see how one can make use of the sequencing of individual events.

Even if the times are available, we still have to consider just how quickly one can expect governments to react to actions. Consider the Cuban Missile Crisis. The time difference between Washington and Moscow was eight hours. But a reading of accounts

---

<sup>4</sup> We should note one minor impediment to using a wide variety of media sources from the internet. Virtually every source uses a different format to present stories. Consequently the software necessary to extract stories from online archives must be modified (at least a bit) for each different source.

<sup>5</sup> A fairly common technique to achieve this is to code only one event of each type per dyad per day.

of that crisis suggests that the time difference meant that each government responded not to individual actions of the other, but to the *accumulation* of events that took place in the previous day.

All of these problems have led us to conclude that, like generations of events data researchers, it is necessary to aggregate events data.<sup>6</sup> In a number of our preliminary analyses, we have used a two week aggregation period. In the analysis reported below, we used a one week aggregation period. Obviously the issue of the optimal aggregation period is one that needs additional study.

### **Plan for Analysis**

We will pursue two paths of analysis. The first will involve attempting to predict the onset of serious international conflict strictly from patterns in previous events. The second path involves building models that will increase our understanding of the process by which conflict escalates.

To pursue the first path, we will treat the stream of aggregated events as a time series. In order to be successful, we must be able to demonstrate two things. First, serious interstate conflicts have to register as significant changes from the event stream prior to the outbreak of the serious conflict. That is, our measure of events should reflect the onset of serious conflicts as dramatic changes in the level of conflict in the event stream. Second, there must be enough information in the event stream *prior* to the onset of serious conflict to predict it. That is, the pattern of events prior to the onset should

---

<sup>6</sup> A good discussion of issues involved in aggregation is Thomas, 2000.

allow for the prediction of serious conflict without an excessive number of false positives or false negatives.

While the first path can be viewed as an exercise (albeit a sophisticated one – see below) in signal processing, in pursuing the second path we seek to increase our understanding of the process by which conflict breaks out. In some ways this is the easier task because international relations researchers have accumulated a number of strong findings at the dyadic level concerning the outbreak of conflict. But a good model is more than an accumulation of variables that have been significant in other studies. So this is likely to prove to be the more difficult task. On the other hand, it is also the more important task.

If we are successful in the first task, we will have created the foundation for an early warning system. By itself this would be an important step forward. But we would accomplish more if this is accompanied by advancing our knowledge about the process by which conflicts occur. We break out our research into a series of questions.

Step 1: Are serious conflicts visible through the analysis of events data? Our initial examination of conflict data from KEDS datasets leads us to conclude that these events series have significant levels of non-stationarity (singularities). This is both good and bad news. The good news is that it appears that significant conflicts are likely to be reflected in these non-stationarities. The bad news is that many statistical (and machine learning) tools do not handle these data well. Consequently we need to use an analysis technique that will be able to handle singularities. That is, it must be able to cope with and identify these singularities. Wavelet analysis is just such a technique. It is very

useful because, unlike Fourier analysis (the standard tool of signal analysis for many years), it simultaneously decomposes a time series into time/frequency space.

Step 2: Is there an unambiguous signal of the onset of serious conflict? Given that serious conflicts register as singularities in events data, the next question is whether these singularities are preceded by clear signals of the onset of conflict. Our initial efforts will also involve using wavelet analysis to detect these precursor signals. As noted above, to be successful, we have to find precursors that simultaneously do not produce too many false positives and too many false negatives.

Step 3: Can we develop compelling models of the onset of serious conflict? The final step in our research is to model the outbreak of conflict. In this effort we will build on the work of quantitative international relations scholars. There have been many studies through the years of what conditions are most frequently associated with the escalation (and de-escalation) of conflict. It would be too much to say that a single dominate model has emerged. But in recent years, there has been a convergence of findings, particularly those that account for the escalation of disputes to war; good summaries of the state of the discipline are in Bremer and Cusack, 1996, and Vasquez, 2000. The factors that are most often associated with the escalation or de-escalation of conflict - usually in situations involving a pair of countries - are (Vasquez, 2000: 367):

- The presence of a territorial issue at stake between the two parties. Situations in which territorial issues are present are more likely to escalate; contiguous states are more likely to have territorial issues than states that do not share a common border.

- Some alliances promote war, while others promote peace. Alliances that settle territorial questions or which do not pose a great threat to a third party promote peace. Alliances that are formed by states that have fought a war recently and are dissatisfied with the status quo are likely to promote war (although the war does not break out immediately).
- Disputes involving states which are in an ongoing arms competition are more likely to escalate to war as long as the arms competition does not involve nuclear weapons.
- Disputes between parties that have had repeated conflicts are more likely to escalate to war.
- A dispute in which the parties bargain aggressively is likely to end in war; across a series of disputes, if the parties take on more and more aggressive bargaining strategies, a war is likely.
- Strong states are more war-prone than weak states. If the power balance between a pair of states rapidly moves towards parity, there is an increased chance of war.
- When major states establish norms or common expectations that reduce their freedom to take unilateral actions, the chances of war are reduced.
- Democratic states are very unlikely to go to war with one another.

These are all important findings that have turned up repeatedly in quantitative studies, and to this mix we will add the level of recent conflict within the dyad of states. Together, this is a set of empirical building blocks that we can use to increase our understanding. But we still need to formulate coherent models that combine sets of

factors in a meaningful (i.e., theoretical) fashion. Otherwise, we have a set of interesting empirical relationships, but no real understanding of why conflict happens.

### **An Exploration: the Ebb and Flow of the Cold War, 1966-1978**

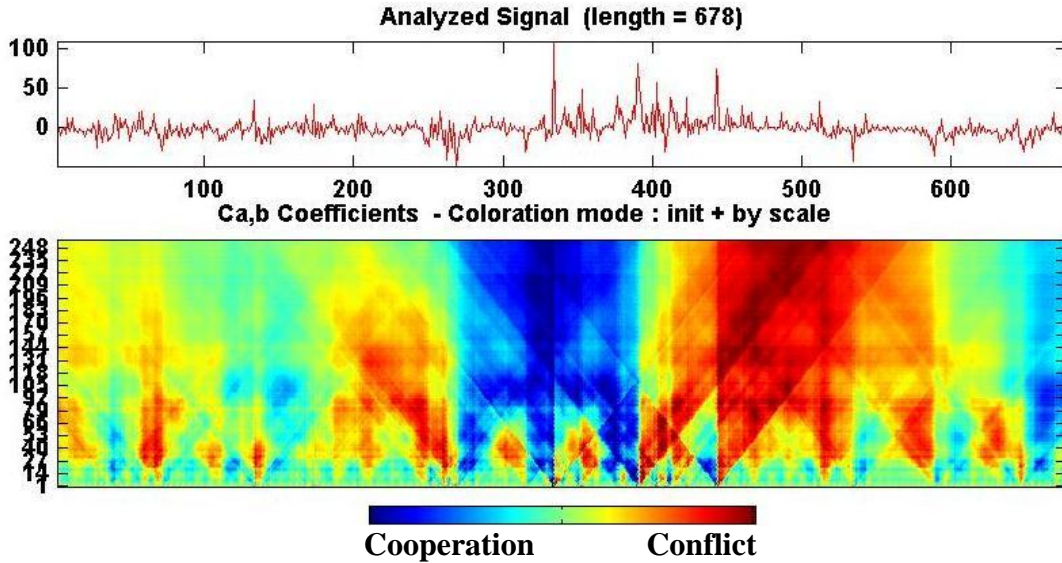
Our research over the next few years will determine whether our approach helps to predict and to understand the onset of serious conflict. For now what we can offer is an exploration of a particular interaction: the United States and the Soviet Union from 1966 through 1978. This analysis uses the original (human coded) WEIS data; this determines the time frame of our study. We begin by selecting just those events from the WEIS data that involve both the United States and the Soviet Union. We then aggregate events to the weekly level, summing the scale scores for all the events during the week. This gives us a dataset with 678 observations. We analyze these data with a Haar wavelet (the simplest possible wavelet).

Figure 1 displays the results of the wavelet analysis.<sup>7</sup> For wavelet analysis to be a useful tool, the results must make sense. That is, there must be a correspondence between what we see in the wavelet analysis, and the ebb and flow of US-Soviet relations. To be sure, this is not a rigorous test of the validity of the analysis technique; we are in essence looking backwards from the wavelet results to the Cold War. On the other hand, if there is no correspondence between events and the wavelet analysis, then great doubt has been cast on the utility of this approach to identify conflicts.

#### **Figure 1. Wavelet Analysis of US-Soviet Interactions, 1966-1978**

---

<sup>7</sup> The wavelet analysis was done using MATLAB version 6.5, release 13, and version 2.2 of the Wavelet toolbox. A haar wavelet was used with a scale setting maximum of 256. Results were displayed with the jet colormap and the init + by scale coloration mode.



Note: a color version of this figure is available at <http://www.ruf.rice.edu/~stoll>

The top part of Figure 1 displays the “raw signal;” it is a plot of the summed scale scores, aggregated to the weekly level as noted above. The lower part of Figure 1 is from the wavelet analysis. The main areas of interest in are colored red and blue. Red areas should correspond to periods of time in which the US and the Soviet Union relations are going through a conflictual period. Blue areas should correspond to periods of time when the US and the Soviet Union are going through a cooperative period. At the present time, we are basically restricted to “eyeballing” the figure in order to identify the specific periods of time. Table 1 displays the identification value of the weeks bounding each area of red and blue, the year and month that corresponds to the weeks, and a list of important events that occurred during that time.

**Table 1. Key Events During Extreme Red and Blue Periods<sup>8</sup>**

Week(s)	Date(s)	Comments
56 – 83	67/1 – 67/8	[Red] Six Day War

<sup>8</sup> Sources used to identify key events were: Department of History, University of San Diego (2002), Goldstein and Freeman (1990), International Institute for Strategic Studies (1978), International Institute for Strategic Studies (1979), and Studenic (2002).

132	68/7	[Red] Czech Crisis (138; 68/8), but NPT signed
187-265	69/8 – 71/1	[Red] Cienfuegos (Cuban sub base)
271-289	71/3 – 71/7	[Blue] Soviet ships leave Cienfuegos.
295-312	71/8 – 71/12	[Red] Soviets criticize Nixon trip to China, Moscow summit
316-334	72/1 – 72/5	[Blue] Moscow summit, SALT, ABM Treaties
341-363	72/7 – 72/12	[Red] US “Christmas bombing” of North Vietnam
364-390	72/12 – 73/6	[Blue] US-Soviet Summit (Brezhnev to US)
391-585	73/6 – 77/3	[Red] October War Crisis; Angola; Carter criticizes Soviets on human rights
620-649	77/11 – 78/6	[Red] Soviet UN employees arrested. Carter criticizes dissident trials. Soviet stage large maneuvers in Central Europe. Ginzburg & Shchransky sentenced.

An informal examination is not a definitive test of the utility of this approach to analysis. In fact, there are actually two parts to this informal examination. In the first part, the wavelet analysis figure is examined to delineate the red and blue zones. The second part involves examining Cold War chronologies, and determining what events took place during the zones. Aside from the informality of the process, it also involves working backwards from the identified zones to the events. But nevertheless, we believe this is a reasonable first step in our investigation.

### **Some Issues in Analyzing Events**

We are generally pleased with the ability of wavelet analysis to identify significant periods of conflict and cooperation between the United States and the Soviet Union.<sup>9</sup> Obviously, we need to formalize the procedures by which we identify zones of conflict and cooperation.<sup>10</sup> But there are other issues that we need to address as well.

---

<sup>9</sup> In previous work (Subramanian and Stoll, 2002) we have used wavelet analysis in a similar fashion to identify periods of conflict and cooperation among eight states in the Persian Gulf region from 1979 through 1999 and reached a similar conclusion.

<sup>10</sup> We are currently implementing a more systematic and precise way to identify these periods.

What events coding scheme should be used? Much of the computer-generated events data has used the WEIS data, and this is therefore a reasonable choice. But the developers of KEDS have recently unveiled a new coding scheme called CAMEO (Conflict and Mediation Event Observations; see Gerner, Schrodt, Yilmaz, and Abu-Jabr, 2002).<sup>11</sup>

Another issue is the particular type of wavelet analysis to be used. There are a large number of wavelets that can be applied to data, and there are additional parameters that must be specified. For the time being, we have been using the haar wavelet. This is a very simple wavelet (which is why we choose to use it). But we need to determine which wavelet (or which wavelets) we should use for our analysis.

An additional issue is the difference in predicting the onset of serious conflict, and postdicting it. In our work to date, we have used an extended time series (for example, the 1966-1978 US-Soviet interactions explored above), and applied wavelet analysis to pick out the areas of high conflict and high cooperation across that entire series. But to build a good early warning system, we will have to predict areas of high conflict *before* they occur. This is a different matter. Consider walking across a field that has peaks and valleys. After walking across the entire field, we can look back and pick out the deepest valleys. But in doing this, we can take advantage of knowing the full extent of each peak and valley, and being able to compare them to pick out the deepest valleys. However, as we walk along and begin to descend that first valley, we don't have the additional information that we acquire in the course of entire journey. So how can we tell as we

---

<sup>11</sup> It should be noted that this scheme is optimized for mediation, which may not serve our purposes as well. In addition, currently there is no scale of conflict and cooperation associated with CAMEO events, although this should be available soon.

descend whether we are going down into a deep valley or a shallow one? This is the challenge of prediction that we face.<sup>12</sup>

Finally, there is the question of at what point we go beyond analyzing just events, to incorporating factors that represent the substantive reasons for the onset of conflict. While it would be intriguing to be able to predict the onset of conflict using only previous events, we think this is unlikely happen. We believe that there is good chance that we will have to include substantive variables in order to generate accurate predictions. Since our initial goal is to create an early warning system, this sort of approach (begin with a few variables, add what is needed to reach a particular level of accuracy) is acceptable. But ultimately we hope to shed additional light on the reasons *why* serious conflicts occur. When we move from prediction to explanation, we will have to build explanatory (as opposed to predictive) models. Predictive ability, while still of interest, will be less of a touchstone than developing explanation.

### **Summary: Computer-Aided Prediction of the Onset of Serious Conflict**

International conflict remains a serious problem in the world of today, despite the end of the Cold War and the demise of the Soviet Union. With an increasing concern about the proliferation of weapons of mass destruction, it is very important that we are able to anticipate the outbreak of serious international conflict. In the future, any number of conflicts could escalate and produce extremely large numbers of casualties,

---

<sup>12</sup> We have made a few informal attempts to investigate this problem. We have taken an events data time series, ended it after the first major valley in a series, then subjected this reduced series to wavelet analysis. The wavelet analysis still appears to detect the major valley. Of course we have only run a small number of informal tests, so the ability of this analysis technique to successfully identify conflicts as they occur remains an open question.

particularly among civilians. The ability to anticipate conflict would be a valuable tool for those individuals, groups, countries, and organizations that would be willing to intervene to prevent or curtail the outbreak of conflict.

Unfortunately, if recent history is any guide, our current abilities are woefully inadequate. Even the United States, with its impressive collection of intelligence assets, has been repeatedly surprised at the outbreak of conflict. One can only assume that other actors have had an equally difficult time of anticipating conflicts.

So how shall we go about creating the tools that would facilitate the accurate anticipation of international conflict? There is no single answer to this question. But we believe that recent advances in information available on the internet, the computer processing of text, and signal processing can be married with existing knowledge about the factors associated with the outbreak of conflict.

The internet provides access to an increasing number of news sources, both to current stories and archives of older stories. The computer processing of text allows us to extract international events from news stories. We believe that the use of signal processing techniques will give us a good chance to both identify and predict the onset of international conflict. Finally, we can build on the foundation of results from international relations scholars to develop a more complete understanding of why conflicts occur.

How successful will we be with this approach? That remains to be seen (check back in a few years!). But we believe we have a good plan and some preliminary results that are promising.

## References

- Azar, Edward. 1980. The Conflict and Peace Data Bank (COPDAB) Project. *Journal of Conflict Resolution* 24: 143-252.
- Bremer, Stuart and Thomas Cusack. 1996. *The Process of War: Advancing the Scientific Study of War*. Amsterdam, Netherlands: Gordon and Breach.
- Bremer, Stuart. 1996. Advancing the Scientific Study of War. In Stuart Bremer and Thomas Cusack (eds). *The Process of War: Advancing the Scientific Study of War*. Amsterdam, Netherlands: Gordon and Breach, pp. 1-33.
- Craven, Mark, Dan DiPasquo, Dayne Freitag, Andrew McCallum, Tom Mitchell, Kamal Nigam, and Sen Slattery. 1999. Learning to Construct Knowledge Bases for the World Wide Web. *Artificial Intelligence* 118: 69-113.
- Department of History, University of San Diego. 2002. Cold War Policies 1945-1991. <http://history.acusd.edu/gen/20th/coldwar0.html>. Accessed 1/21/03.
- Gerner, Deborah, Philip Schrodt, Ronald Francisco and Judith Weddle. 1994. The Analysis of Political Events Using Machine Coded Data. *International Studies Quarterly* 38: 91-119.
- Gerner, Deborah, Philip Schrodt, Omur Yilmaz, and Rajaa Abu-Jabr. 2002. The Creation of CAMEO (Conflict and Mediation Event Observations): An Events Data Framework for A Post Cold War World. Paper presented at the Annual Meetings of the American Political Science Association, 29 August – 1 September, 2002.
- Goldstein, Joshua and John Freeman. 1990. *Three-Way Street: Strategic Reciprocity in World Politics*. Chicago: University of Chicago.

- International Institute for Strategic Studies. 1978. *Strategic Survey 1977*. London, UK:  
International Institute for Strategic Studies.
- International Institute for Strategic Studies. 1979. *Strategic Survey 1978*. London, UK:  
International Institute for Strategic Studies.
- Jones, Daniel, Stuart Bremer, and J. David Singer. 1996. Militarized Interstate Disputes,  
1816-1992: Rationale, Coding Rules, and Empirical Patterns. *Conflict  
Management and Peace Science* 15: 163-213.
- King, Gary and Will Lowe. 2002. An Automated Information Extraction Tool for  
International Conflict Data with Performance as Good as Human Coders: A Rare  
Events Evaluation Design. Preprint available at  
<http://gking.harvard.edu/files/infoex.pdf>. Accessed 1/29/2003.
- Putnam, Robert. 1988. Diplomacy and Domestic Politics: The Logic of Two-Level  
Games. *International Organization* 42: 427-460.
- Schrodt, Philip and Deborah Gerner. 1994. Validity Assessment of a Machine-Coded  
Event Data Set for the Middle East, 1982-1992. *American Journal of Political  
Science* 38: 825-854.
- Schrodt, Philip and Deborah Gerner. 1997. Empirical Indicators of Crisis Phase in the  
Middle East, 1982-1995. *Journal of Conflict Resolution* 41: 529-552.
- Singer, J. David. 1965. Data Making in International Relations. *Behavioral Science* 10:  
68-80.
- Studenic, Roman. 2002. The Cold War Guide.  
<http://www.geocities.com/CapitolHill/Lobby/4907/Chron.htm>. Accessed 1/21/03.

- Subramanian, Devika and Richard J Stoll. 2002. Events, Patterns, and Analysis: Forecasting International Conflict in the Twenty-First Century. Position paper presented at the Computer-Aided Methods for International Conflict Resolution Workshop, held at the Austrian Research Institute for Artificial Intelligence, Vienna, Austria, October 25-26, 2002.
- Thomas, G. Dale. 2000. The Machine-Assisted Creation of Historical Events Data Sets: A Practical Guide. Paper presented at the 2000 Annual Meeting of the International Studies Association, Los Angeles, CA, March 14-18, 2000.
- Tomlinson, Rodney. 1993. World Event/Interaction Survey (WEIS) Coding Manual. Manuscript. Department of Political Science. United States Naval Academy, Annapolis, MD.
- Vasquez, John (ed.). 2000. *What Do We Know About War?* Lanham, MD: Rowman and Littlefield.