UCRL-95242, Rev. 2
PREPRINT

ELEMENTS OF A SYSTEM FOR VERIFYING A COMPREHENSIVE TEST BAN

W. J. Hannon

UCRL--95242-Rev.2

DE87 007094

For the Proceedings of the Workshop on Global Seismic Monitoring, Canberra, Australia, September 23-25, 1986.

March 6, 1987

Lawrence Livermore National Laboratory

Treaty Verification Program

ELEMENTS OF A SYSTEM FOR VERIFYING A COMPREHENSIVE TEST BAN*

Paper Presented at the Opening of the Australian Seismic Data Centre
September 23-25, 1986

W. J. Hannon

Introduction

Today I will talk about discrimination as part of a seismic monitoring system for a Comprehensive Test Ban (CTB). As a disclaimer, I should point out, at the start, that the views that I express are my own. They do not necessarily represent those of any organization.

Discrimination is only part of the monitoring system for a CTB or a low yield threshold test ban. To place it in perspective, this presentation, I will talk about the goals of the monitoring system, its functions, the challenges to verification, discrimination techniques, and some recent developments. But let me give you the bottom line, at least from from my perspective: Some of the long term benefits of a CTB, especially its contribution to confidence building and stability, are at risk because of the potentially large number of unidentified or misclassified events that will occur. These are disturbances which are detected but cannot be unequivocally identified as a nuclear explosion, a chemical explosion, or an earthquake.

*This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract No. W-7405-Eng-48.

Goals of Verification System

Treaty verification has several goals: (Vugraph 1) 1) timely, assured recognition of significant violations -i.e., catching the other side when it cheats; 2) confidence building -i.e., avoiding false accusations which lessen the confidence of the monitor and the incentive for the potential evader to comply; and 3) deterring militarily significant violations--i.e., increasing the probable cost of non-compliance to a level at which the potential evader will not carry out tests which will provide a significant advantage.

The first two goals are related to the standard statistical type-one and type two-errors. The third goal involves a cost/benefit analysis and has very large uncertainties associated with it. The national monitoring systems and, I would argue by analogy, the international systems chosen for deployment depend upon the weights that are assigned to these various goals. Depending upon the weights assigned, one can design different networks and have different capabilities. The cost implications and the operational implications of the different monitoring choices are significant.

When we discuss the extent to which these goals are met, we need to look at the quantitative performance measures and the integrated military and political decisions. In the international arena, this is difficult to do because of the mix of interest in horizontal and vertical proliferation. In this discussion, I am not going to treat horizontal proliferation. Several attempts have been made to define a militarily significant program from a vertical proliferation viewpoint. (Vugraph 2) In the U. S. literature, one can find estimates of militarily significant yields that range from a .25 kt to 10 kt. Numbers like 1, 7, 10 have been cited as the number of tests necessary to execute a militarily significant program. The significance of the number seven is that, if the probability of detecting a single explosion is 30%, the probability of detecting at least 1 explosion out of a 7 explosion test series is 90%. This illustrates the necessity of considering the degree of confidence desired. We are not going to get 100% confidence from any of the known monitoring systems. Two confidence levels which have been discussed fairly often are 90%, which is more or less related to stressing the assured recognition of violations, and 30%, which is usually related to a deterrence argument.

Finally, we have to consider the number and type of unidentified events per year that we are willing to accept. Such unidentified events will occur and we have to to come to grips with this fact ahead of time. Failure to do so, will result in on-going compliance debates which degrade the political value of the treaty.

I encourage you to read the available literature about the deliberations that have occurred concerning compliance with the threshold test ban. It will give you some very interesting insights into how nations involved in monitoring test ban treaties deal with verification problems: the decisions which must be made; the uncertainties which come up; and, the spectrum of views which has to be considered. All of these views, whether you agree with them or not, are part of the system. Most of them must be addressed if the confidence building benefits of the treaty are to be achieved.

Functions of Monitoring Systems

The monitoring system involves many functions. (Vugraph 3) The well known ones are: data acquisition, seismic wave detection, event detection, association of seismic phases with a given event, location, characterization, and discrimination. Here I have defined characterization as the estimation of various parameters of the source without making any decision about what kind of source it is. For example, we could define the components of a moment tensor representation of the source. Given these components, we could, in a separate step, discriminate among source types using physical and statistical principles. I believe it is useful to view these as separate steps because they involve different disciplines. As we will see, the discrimination function involves value judgments directly related to the first two considerations described under goals.

Less well known and certainly less well discussed elements of the system are functions like site selection, start-up procedures, calibration procedures, the integrated use of national technical means and/or their international equivalents, and ranking events from most explosion like to least explosion like once they have been detected. Although the use of "national technical means" may not seem to apply on the international scene, I suggest that the question needs to be examined. For example, the French SPOT satellite system is a very capable system and it could provide information at the international level which can be integrated with the seismic information.

We need to consider, too, the whole process by which the science of monitoring interacts with the decision makers and the policy makers. That interface is very poor, at least in my experience, and the fault lies with both the scientists and the decision makers. Neither understands the strengths or the weaknesses of the other's knowledge base. Finally, we must consider how we are going to handle on-site-inspections. Location accuracy is a very key issue for on-site-inspection--if we are not at the right location, the value of on-site-inspections for assumed detection of violations is very limited. The limitations of on-site inspection suggest that its primary value may be one of deterrence rather than assures detection of violations or confidence building.

Uncertainties in Capability Estimates

Any of the technical inputs in the assessments of the capability of monitoring system contain significant systematic and random uncertainties. (Vugraph 4) There are few direct, controlled experiments. We try to estimate the monitoring capability in other countries from our own experience or from the information that is available to us. But this extrapolation process has errors in it. Systematic errors occur, e.g., effects like regional biases. Often we have to estimate them by expert opinion in the absence of any direct measurements. There are formal processes for doing this, and they need to be incorporated in the system. Then, as our knowledge evolves, we can go back and reconstruct how the previous estimate was made and determine the significance of the update.

These are also random uncertainties which arise because of the heterogeneity of the Earth and our inability to identify or measure important parameters which affect the seismic signals.

In addition, we should note that the evasion populations are unlike the natural reference populations. For example, the Group of Scientific Experts of the Conference on Disarmament has generated maps of detection capability. These are based upon the properties of earthquake populations. They are not based on distributions of signals and noise appropriate for evasion scenarios. In particular, they do not account for the fact that a potential evader will choose propagation paths, times, and source conditions which minimize the signal and which maximize the noise at key recording sites. The problem is compounded by the fact that we will continually encounter situations in which we are working in the tails of poorly defined populations. We often approximate these populations by normal distributions. In fact, we often do not know the form of the distribution. Finally, there is the problem posed by the existence of classified and unclassified databases. How will the various nations integrate the national and international databases, procedures and results? Clearly the problem varies from one nation to the next. The problem is compounded by the fact that both capability and the targets of interest vary among nations.

Evasion Scenarios

Evasion scenarios have been constructed for almost all environments. (Vugraph 5) In the underground environment cavity decoupling, hide-in-earthquake and chemical explosions are three that have been discussed extensively. In my opinion, cavity decoupling is the one which ends up driving the monitoring system because it forces us to deal with very small magnitudes. Not only does this increase the demands on the data acquisition system, it also causes problems in processing and in false alarms because the number of earthquakes increases as the magnitude decreases. The chemical explosions also pose a number of problems. The problem of distinguishing between a 1 kt decoupled nuclear explosion and a 5-20 ton well-coupled chemical explosion is one for which I currently know of no solution.

We also have to worry about monitoring the oceans, the atmosphere and space. It is not sufficient to look back at the Limited Test Ban Treaty or other treaties and say that the monitoring problems in these environments were solved. If we have a comprehensive test ban, all the measures which were considered sufficient in the past for monitoring those environments have to be re-examined in terms of the importance of evasion at a couple of kilotons. In these environments, an extremely important issue is the question of attribution. That is, in the oceans and in the atmosphere there are many nations with the capability to test. For example, a ship could go by in the ocean and drop a test package over the side, and sail on. A month later, the package could be detonated. Who did it? We simply would not know. We might know that an event occurred, but we would not know who did it. I think that this factor should to be considered when international system tries to decide how much effort should be put into monitoring the oceans, especially the Southern Hemisphere. If we cannot use the information from the network once we have it, we may initially want to put our priorities elsewhere, e.g., into monitoring the land masses.

Monitoring Assessments

When monitoring land masses, attribution could be a problem, but not in most cases. Usually, successful evasion on land requires that the signals not be detected, or if detected, not identified as coming from a nuclear explosion. The next Vugraph (6) illustrates the low magnitudes that must be monitored, the fact that the estimates are uncertain (the width of the bands and the dashed lines) and the types of networks necessary to meet these levels in a nation the size of the Soviet Union. In addition, there are uncertainties in the magnitudes associated with regional biases and different magnitudes and in the magnitude-yield relationships at low magnitudes. I will first talk about detection thresholds at these low levels.

Although I know that Australia has a long history of experience with arrays, it may still be useful to point out some of their benefits for verification. Arrays do more than just improve signal-to-noise ratios. They allow us to use regional seismic waves which we could not use before because we could not identify them with the signals recorded at a single location. For example, if the high amplitude regional wave Lg was the only wave seen from a small event, we might not be able to use it because we could not identify it and associate it with a specific source. With the array, we can identify it as an Lg wave because we can determine the velocity associated with it. That would make a significant difference in the location and association capabilities. If we talk about detecting a 1 kt decoupled explosion something like 25-30 arrays are required. With this number of stations deployed in a country the size of the Soviet Union, we would end up with about a 1000 km spacing between the stations. The assessments illustrated in Figure 6 involve many assumptions, and a 0.2-0.3 magnitude unit variation is well within their uncertainty.

The discrimination threshold corresponding to a detection threshold of M_b 2.2-2.4 would be m_b 2.8-3.1 or so. (The values are quite uncertain due to the lack of experience with regional discriminants and the fact that the evader can determine the conditions under which the monitor must operate.) There are very few descriptions in the literature about discrimination as part of a national or international verification system. What has been published is based primarily upon U.S. experience. There is little direct information from systems deployed in the Soviet Union. This fact reflects the significant asymmetry in the information that is available to different nations.

As we operate closer to the detection threshold, the signal-to-noise ratio becomes lower and more events will become unidentified. We will require very high performance from our discrimination process. Something like a (0.1%) false alarm rate will be required in order to get the number of unidentified events down to a range which is comparable to the number of inspections that one might reasonably expect. If we end up with several hundred unidentified events, the question is "What are you going to do with them?" This question is equally valid on both the national and international level.

Discriminants

Another area of concern is the definition of discrimination techniques that can be used to identify earthquakes, chemical explosions and nuclear explosions. The location is a fundamental discriminant. The argument is that if a seismic event is more than 25 km offshore, then it is an earthquake. Depth is another fundamental discriminant. An event below 10 or 15 km is considered an earthquake because drilling limitations prohibit testing at that depth. However, one has to take into account the uncertainty in the depth estimates which can easily be 10 - 15 km.

After using these discriminants, one begins to look at differences in waveforms. However, many of the waveform discriminants do not work well for small events near the detection threshold because they depend on high signal-to-noise ratios. For example, a downward P-wave first motion has been considered an indication of an earthquake. However, when operating close to the detection threshold of the network, the signal-to-noise is very low, and the first negative swing maybe lost and only the first positive seen. In this case, an earthquake would be identified as an explosion.

There is a very good article in in the Bulletin of the Seismological Society, (Vol. 62 No. 6B, 1982) by Pomeroy, McEvilly, and Best in which they discuss a wide variety of regional discriminants. Some of these can be grouped under the category of surface waves and/or guided waves versus other kinds of waves. An example of this category is the surface wave magnitude vs. the body wave magnitude discriminant. This is an important, widely-used discriminant. Others of this type are: the long-period energy in a given window vs. the local magnitude, higher mode surface waves, and the ratio of the long-period Love wave energy to the long-period Rayleigh wave energy. Many of these are good discriminants for large events, but do not work well for small events because the surface wave information is lost. Thus a technique which looks like a powerful discriminant at magnitudes of 4.0 to 5.0 will often fail when applied to magnitude 2.5 events.

Other discriminants are based on the relative excitation of various regional phases. Many of the discrimination processes for low magnitude events are highly dependent on the region. Discriminants that work in one source region or one receiver region will not work for others.

Discriminants involving Lg, Rg, and Sn have all been suggested, and we have tried some of them. We have also examined spectral ratios for Pn, Pg and Lg. Some very promising discriminants that compare the energy in various frequency bands have been suggested by Archambeau and others. For example, Pn or Pg vs. Lg are good of discriminants in some regions. In other regions we get either the reverse of the discriminant that we had before, or the discriminant fails. These may prove to be very useful. However the problem of operating at low signal-to-noise ratios will also be difficult for these methods.

There have been other discriminants that have looked very promising. I was personally very much interested in one which used estimations of the seismic moment tensor and its invariants to build discriminants. However, when operating in realistic monitoring environments, there may not be enough stations and enough azimuthal coverage to make good estimates of the moment tensor. Therefore, the method does not hold as great a promise for low magnitude events. It still may be a good discriminant when you are up say 0.6 or 0.7 magnitude units above the threshold. Pearce, from the UK, had an algorithm of this type which worked rather well for larger events. The method mapped out the focal mechanism of the source, by examining all possible solutions and identifying the ones which were most likely. Teledyne-Geotech researchers have shown some of the practical limitations of this method for low signal-to-noise events.

My colleagues, Denny, Taylor, Vergino and Patton are engaged in a discrimination study which uses data from Lawrence Livermore National Laboratory stations which are 200 - 400 km from the Nevada Test Site (NTS).
(Vugraph 7) They used 330 explosions from the test site and 130 earthquakes from the Western U. S. The stations are shown with the small triangles and the earthquakes with asterisks. The test site is, of course, located in southern Nevada. All the explosions are at the test site and the

earthquakes are spread throughout the western U.S. This raises a concern common to many studies. We may be discriminating among regions and propagation paths and not among source types.

Subject to this caveat, we found that the long-period discriminants were better in general than the short-period ones. (Vugraph 8) For Love-wave energy vs. M_L (upper panel) we are able to draw a line through the data for which all of the explosions are below the line and all of the earthquakes are on or above the line. We were not able to achieve as clean a separation for the short period discriminants, involving the ratio of the two regional waves Lg and Pg (lower panel).

As mentioned above, the results that we get are quite dependent on the region. We examined the error distributions for the standard western U.S. magnitude determinations for NTS explosions and concluded that the formulas should be modified. Using the standard formulas, we saw considerable overlap in the earthquake and explosion population for data recorded at the Mina, Nevada station. (Vugraph 9 - top panel). When we went back and used all of the information from the well located sites at NTS, we were able to revise the magnitude curves for these regions. With the new magnitude relationships, we then had a finely tuned regional description and obtained much better discrimina- tion. (Vugraph 9 - lower panel). This implies that the discrimination techniques for small events will require region by region calibration. This may not be possible unless specific calibration measures are negotiated as part of any treaty monitoring agreement.

One question which naturally arises is what we could do if we combined results from different discriminants. Vugraph 10 shows the improvement that can be obtained by such combinations. Note that the improvement is significant even at one station. Further improvement could be obtained by combining the results from multiple stations. Vugraph 11 shows the variability that we found in multivariate results at the four stations. The upper panel shows the possibility of misclassifying an earthquake as a function of magnitude in terms of the ratio of the number of misclassified events to total number of events. The lower panel shows a similar misclassifying earthquake. In general we see good performance for both populations down to about magnitude 4. Below that level performance declines significantly.

Implicit in discrimination algorithms of this type are value judgments about the cost associated with misidentifying an earthquake vs. the cost of misidentifying explosions. In this case, those two errors were considered to be of equal concern and on that basis the line was drawn. This weighting is a parameter in most algorithms. For example, one could weight misclassified explosions as being ten times more costly than misclassified earthquakes. If this is done, the decision line shifts. I point this out to emphasize interplay between the science and the value judgments of the policy makers. This is a critical issue which I think is worthy of intense consideration. If the international network is used for discrimination, it, too, will be faced with this problem.

Looking in detail, we see that at one location we get down to something like a 5% error rate. The significance of a 5% error is that there will be many unidentified events for networks capable of detecting small events. In Vugraph 12, we see the number of unidentified events which might occur if 15-20% of the events between the threshold and 0.5 magnitude units above the threshold remain unidentified. We can divide these numbers by 4 if we want to get down to 5%, or even by 10 or so if we project the results that might be achieved by combining the results from 4 or 5 stations. However, what we see in using these numbers, is that we could get somewhere near 100 unidentified events a year at magnitude 2. These will be a continuing source of friction unless some technical or political means of addressing them is found.

We can eliminate some of these unidentified events by techniques such as overhead photography--e.g., the French SPOT system. If some of those events turn up in the middle of relatively unpopulated areas--no roads, no landing sites, no signs of human activity, we can discard that event. However, even with such all-out processing, it appears that there will still be a large number of unidentified events--more than can be inspected.

All of the discrimination discussion so far has been in the context of earthquakes vs. nuclear explosions. In addition, chemical explosions can mask and can be misidentified as nuclear explosions. (Vugraph 13) Relatively rare chemical explosions in the 1-4 kt range do occur. However, these are sufficiently rare that they could be handled by procedures similar to those defined in the Peaceful Nuclear Explosion Treaty. If the large chemical explosions were pre-announced, one could go to the site before the shot, be there when the shot went off and make appropriate measurements. On the other hand, there are many events which are in the range 20-50 tons. Note that a 20 ton chemical explosion may produce signals equivalent to a 1 kt decoupled nuclear explosion.

Such small events are probably too numerous and widespread to be handled by inspections. Discrimination between these 20 ton chemical explosions and 1 kt decoupled nuclear explosions has not been established. Some discrimination may be possible for some chemical explosions which are used in quarry blasts. In these applications, successive salvos split away rock faces. The explosions have distinctive characteristics due to the sequential firing of the salvos and, possibly, the shallow depths. However, a 100 ton quarry shot may have 5 - 20 ton salvos. Therefore, although the overall event will have a very characteristic signature due to the duration of this sequence, any one of the individual salvos is sufficient to mask or be replaced by a nuclear explosion.

Some Possible Improvements in Monitoring Capability

There is some hope that we may be able to address some of these concerns. The use of high frequencies has received much attention. I think that one of the best ways to comment on the use of high frequencies is to suggest that you read Dr. Charles Archambeau's comments in a publication that LLNL put out in 1986. We had a conference on Cavity Decoupling in which Dr. Archambeau made a presentation. One of the things we asked him to do was to give us an assessment of what he thought the strengths and weaknesses of the high frequency approach are.

Basically, what he said was that he had concerns about some of the assumptions and extrapolations made in his assessments. For example, he projected the detection and discrimination behavior observed at magnitude 4 down to magnitude 2, and projected results from measurements at 5-15 Hz to behavior at 30-50 Hz. Finally, he made some observations and projections about attenuation along the path. At the same time that he identified these concerns, he said that he was optimistic and thought that in many cases the performance will be better than assumed. My own assessment is that although the exact contribution remains to be proven, the high frequency method will work on some events. Whether or not it will work at any given station or be independent of azimuth or path, no one currently knows. Nor do we know how to select sites which have the desired properties. One approach might be to detonate calibration explosions as part of a site selection process. Nuclear explosions are politically unrealistic, but chemical explosions may work.

Conclusion

In conclusion, technical, military and political efforts are required to establish and verify test ban treaties which will contribute to stability in the long term. It currently appears that there will be a significant number of unidentified events. Therefore, some risks are unavoidable and these must be clearly identified and evaluated. (Vugraph 14).

Given the plurality of views that are present and effective in free societies, approaches must be found that will be acceptable to a large section of the middle of the spectrum of opinion over the long term. These approaches must be robust in the face of changes in political and technical perceptions which will occur in the course of time. A significant contribution can be made by carrying out excellent technical work and presenting the results in a manner which makes the significance of the work and the inherent uncertainties clear to the policy makers. Your new centre should provide the basis for such work. I congratulate you on this fine facility and excellent staff and look forward to cooperating with you in efforts to improve verification methods.

5213H

TREATY VERIFICATION HAS SEVERAL GOALS WHICH MUST BE MET IN THE LONG TERM

- Timely, assured recognition of militarily significant violations
  "Catch them if they cheat"

- Confidence building
  False accusations affect both sides

- Deterrence of militarily significant violations
  Involves evader's cost/benefit analysis

EMPHASIS DETERMINED BY NATIONAL SECURITY CONSIDERATIONS.

MONITORING REQUIREMENTS AFFECTED BY EMPHASIS.

DEFINING QUANTITATIVE PERFORMANCE MEASURES REQUIRES INTERRELATED MILITARY AND POLITICAL DECISIONS

- Describe significant military programs
  - Yield levels (0.25, 1, 5, 10, ? kt)
  - Number of tests (1,7,10, ?)

- Define desired degree of confidence (30%, 90%, ~~100%~~)

- Define acceptable number of unidentified events/year

Monitoring Systems Involve Many Functions

- Well Known
  - Data acquisition
  - Seismic wave detection
  - Association/Event detection
  - Location
  - Characterization
  - Discrimination

- Often Overlooked
  - Site selection
  - Start up procedures - calibration and procedures
  - Ranking
  - NTM integration
  - Reporting to decision makers
  - OSI

Almost all of these involve uncertainty

Technical inputs contain significant systematic and random uncertainties involving both measurement and expert opinion.

- Few "controlled" experiments
- Soviet populations estimated from US experience (bias)
- Systematic errors bounded – expert opinion
- Evasion populations unlike reference populations
- Tails of populations poorly defined
- Classified and unclassified databases

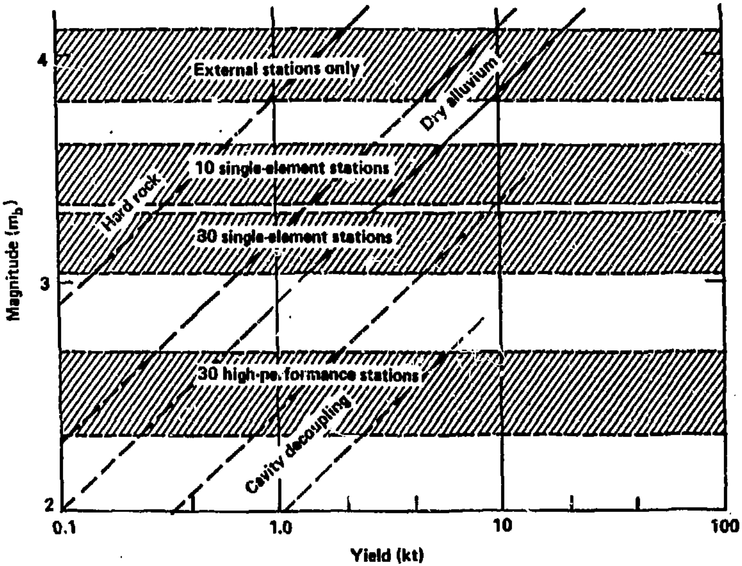
CTBT EVASION SCENARIOS HAVE BEEN CONSTRUCTED FOR ALL ENVIRONMENTS

- UNDERGROUND
  - CAVITY DECOUPLING
  - HIDE-IN-EARTHQUAKE (HIE)
  - CHEMICAL EXPLOSIONS
- OCEAN
- ATMOSPHERE
- SPACE

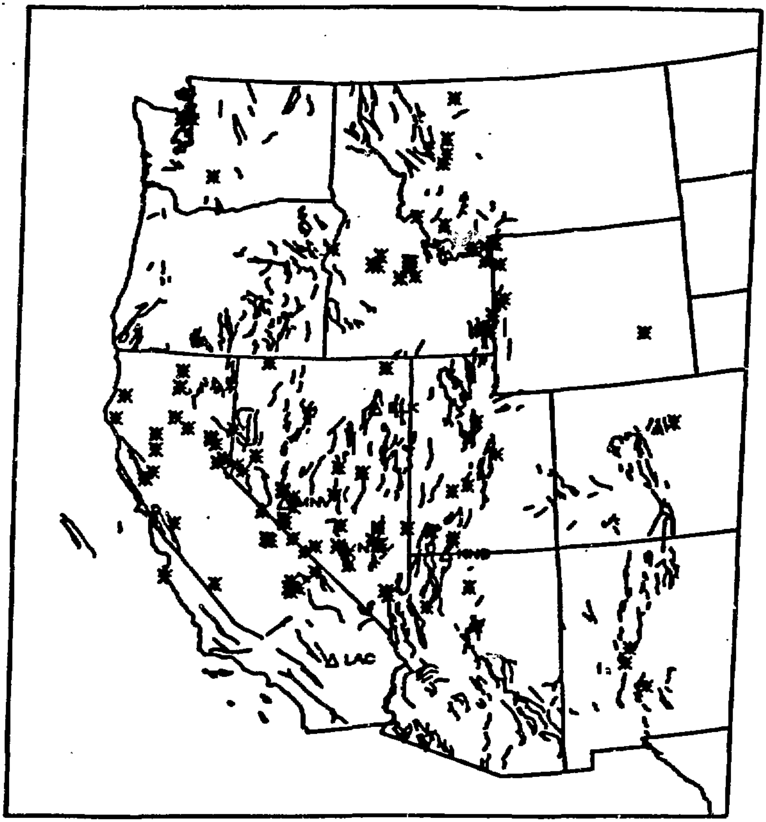
DETECTION

ATTRIBUTION

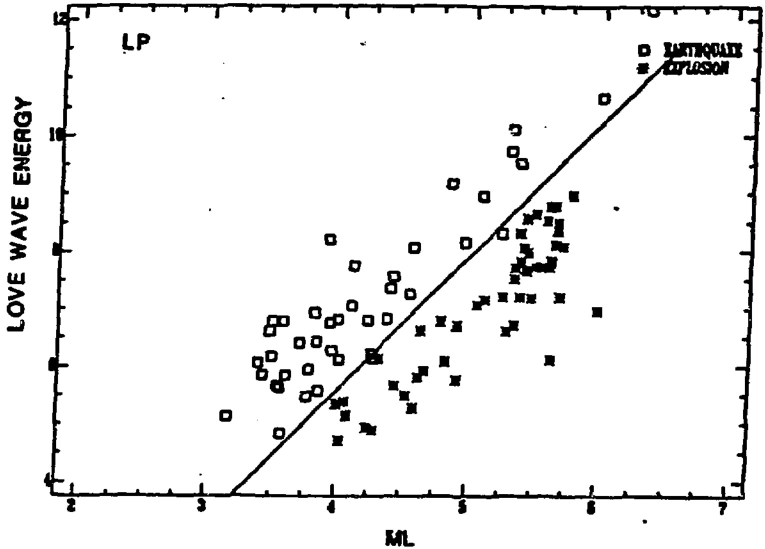
IN-COUNTRY SEISMIC STATIONS WOULD BE REQUIRED TO PRECLUDE CLANDESTINE TESTS > 1kt

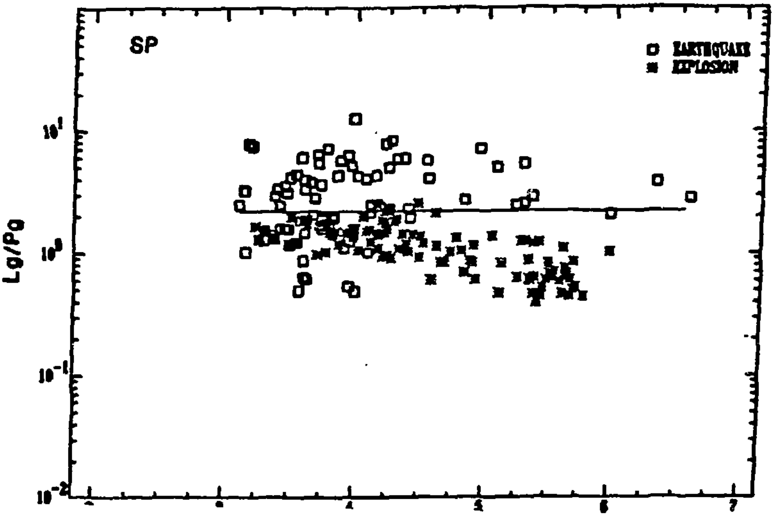


THE DATABASE INCLUDED LLNL RECORDINGS OF 330 NTS EXPLOSIONS AND 130 WUS EARTHQUAKES

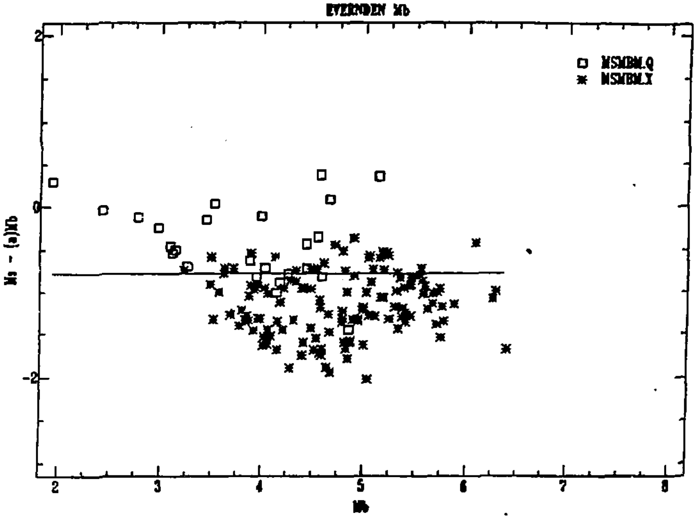


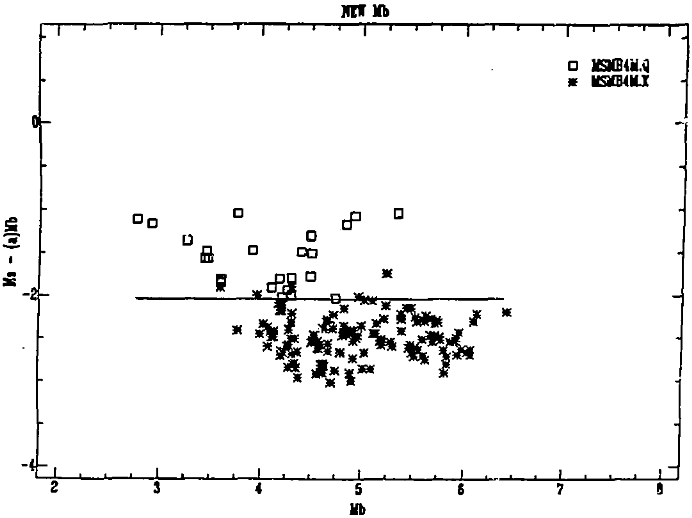
LP DISCRIMINANTS SEPARATE BETTER THAN SP DISCRIMINANTS
BUT LP MEASUREMENTS ARE DIFFICULT



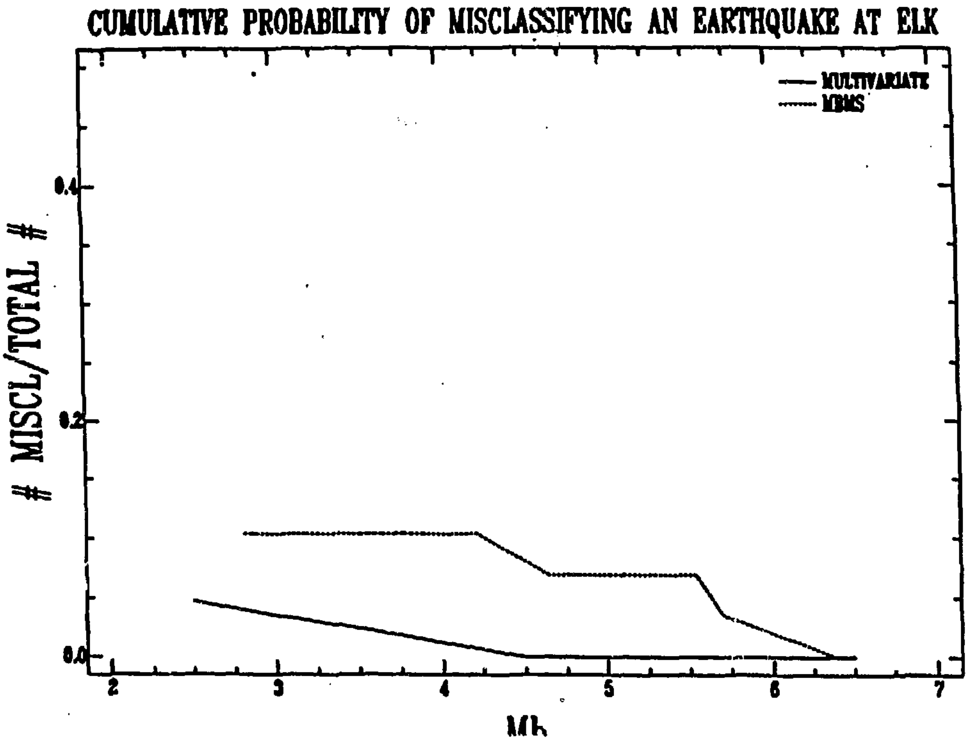


NEW m_b AND M_s FORMULAS HAVE REDUCED THE PROBABILITY OF MISCLASSIFYING AN EXPLOSION BY 3 - 9%



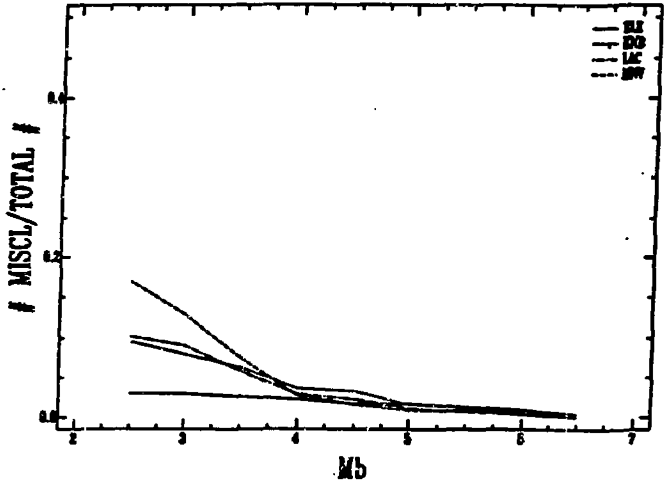


MULTIVARIATE RESULTS REDUCE THE NUMBER OF OF MISCLASSIFICATIONS CONSIDERABLY

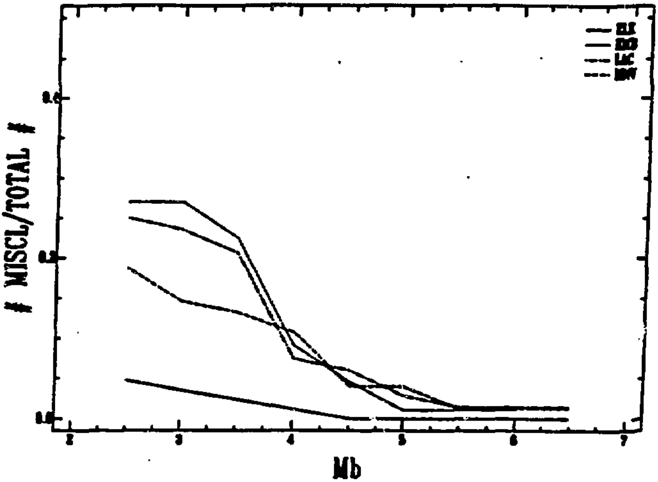


MULTIVARIATE RESULTS DEPEND ON MAGNITUDE, SOURCE TYPE, AND STATION

CUMULATIVE PROBABILITY OF MISCLASSIFYING AN EXPLOSION



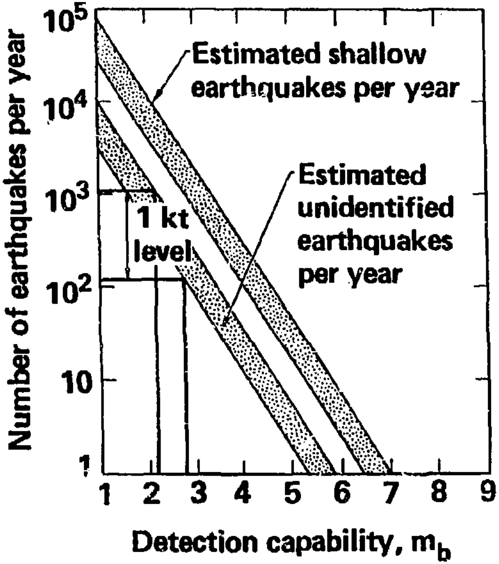
CUMULATIVE PROBABILITY OF MISCLASSIFYING AN EARTHQUAKE



WE EXPECT LARGE NUMBERS OF UNIDENTIFIED EVENTS FOR NETWORKS CAPABLE OF DETECTING SMALL DECOUPLED EXPLOSIONS

Shallow earthquakes and assumed* unidentified earthquakes in the Soviet Union

*Assumes 20% of the earthquakes in the interval $0 \leqslant m_b$ - detection capability $\leqslant 0.5$ will be unidentified



CHEMICAL EXPLOSIONS (CE) CAN MASK AND BE MISIDENTIFIED AS NUCLEAR EXPLOSIONS (NE)

– RARE CE $\cong$ 1 – 4 KT

– MANY CE $\cong$ 20T $\cong$ 1KT DECOUPLED

+ SOME CEs DISTRIBUTED IN TIME

TECHNICAL, MILITARY AND POLITICAL EFFORTS ARE REQUIRED FOR STABILITY. RISKS WILL EXIST.

- Technical
  - Seismic research in discrimination
  - Extensive, capable in-country networks
  - Calibration events & startup period
  - Additional measures (pre-announcement, OSIs)
  - Ability to resolve ambiguous US events

- Military/R&D
  - Definition of militarily significant violations

- Political
  - Procedures to handle uncertainty
  - Resolution of false alarm/assurance balance
  - Proper treatment of OSIs, calibration events
  - Linkage to progress in other areas

TEX