Defense Threat Reduction Agency
8725 John J. Kingman Road, MS 6201
Fort Belvoir, VA 22060-6201

DTRA-TR-03-42

TECHNICAL REPORT

Investigation of Factors Affecting the Transportability of the P/S Amplitude Ratio Discriminant

Approved for public release; distribution is unlimited.

March 2006

DTRA 01-00-C-0043

Douglas Baumgardt, et al.

Prepared by:
ENSCO, Inc.
5400 Port Royal Rd.
Springfield, VA 22151

DESTRUCTION NOTICE

FOR CLASSIFIED documents, follow the procedures in DoD 5550.22-M, National Industrial Security Program Operating Manual, Chapter 5, Section 7 (NISPOM) or DoD 5200.1-R, Information Security Program Regulation, Chapter 1X.

FOR UNCLASSIFIED limited documents, destroyed by any method that will prevent disclosure of contents or reconstruction of the document.

Retention of this document by DoD contractors is authorized in accordance with DoD 5220.22M, Industrial Security manual.

PLEASE NOTIFY THE DEFENSE THREAT REDUCTION AGENCY, ATTN: IMMI, 8725 JOHN J. KINGMAN ROAD, MS-6201, FT. BELVOIR, VA 22060-6201. IF YOUR ADDRESS IS INCORRECT, IF YOU WISH IT DELETED FROM THE DISTRIBUTION LIST, OR IF THE ADDRESSEE IS NO LONGER EMPLOYED BY YOUR ORGANIZATION.

DISTRIBUTION LIST UPDATE

This mailer is provided to enable DTRA to maintain current distribution lists for reports. (We would appreciate you providing the requested information.)

- ❑ Add the individual listed to your distribution list.
- ❑ Delete the cited organization/individual.
- ❑ Change of address.

Note:
Please return the mailing label from the document so that any additions, changes, corrections or deletions can be made easily. For distribution cancellation or more information call DTRA/BDLMI (703) 767-4725.

NAME: _____

ORGANIZATION: _____

OLD ADDRESS	**NEW ADDRESS**
_____	_____
_____	_____
_____	_____

TELEPHONE NUMBER: ( ) _____

DTRA PUBLICATION NUMBER/TITLE	**CHANGES/DELETIONS/ADDITONS, etc.** *(Attach Sheet if more Space is Required)*
_____	_____
_____	_____
_____	_____

DTRA or other GOVERNMENT CONTRACT NUMBER: _____

CERTIFICATION of NEED-TO-KNOW BY GOVERNMENT SPONSOR (if other than DTRA):

SPONSORING ORGANIZATION: _____

CONTRACTING OFFICER or REPRESENTATIVE: _____

SIGNATURE: _____

DEFENSE THREAT REDUCTION AGENCY
ATTN: BDLMI
8725 John J Kingman Road, MS 6201
Fort Belvoir, VA 22060-6201

DEFENSE THREAT REDUCTION AGENCY
ATTN. BDLMI
8725 John J Kingman Road, MS 6201
Fort Belvoir, VA 22060-6201

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
00-03-2006	Technical Report	6/9/2001 to 6/9/2003

4. TITLE AND SUBTITLE

Investigation of Factors Affecting the Transportability of the Regional P/S Amplitude Ratio Discriminant

5a. CONTRACT NUMBER
DTRA 01-00-C-0043

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
1340

6. AUTHOR(S)

Douglas Baumgardt, Zoltan Der, Angelina Freeman, and Meghan Keohane

5d. PROJECT NUMBER
CO

5e. TASK NUMBER
IS

5f. WORK UNIT NUMBER
DH01743

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

ENSCO, Inc.
5400 Port Royal Rd.
Springfield, VA 22151

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Defense Threat Reduction Agency
8725 John J. Kingman Rd.
Fort Belvoir, VA 22060
NTD/ D. Barber

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
DTRA-TR-03-42

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The research under this program has investigated problems associated with transporting regional-phase amplitude ratios, such as Pn/Sn or Pn/Lg ratios. The first study investigated the effect of sensor site effects on the variance of P/S ratios. Using multiple array recordings of groups of events in the same source region, we characterized the factors that contribute bias or the scatter of P/S ratio measurements, after correction for propagation path effects. The variance in the P/S ratio around regional arrays reveals the extent to which site affects cause variations in P/S ratios. The partitioning of the variance between source, path, and receiver effects was examined by analysis of variance (ANOVA). In the second study, we performed a statistical analysis of the transportability of P/S ratio discriminates using separability measures and optimum transformations in order to reduce dimensionality of multiple frequency P/S ratios. These transformations consist of calculating the intra-class and inter-class scatter matrices for P/S ratio discriminants and using the eigenvectors, corresponding to the largest eigenvalues, of the inter-class matrix to compute optimum transformation of discriminants that provide the best separation. We applied this analysis to distance-corrected discriminants in different regions (e.g., China, Eurasia, North America) in order to compare discriminant effectiveness for different regions and to evaluate the transportability of optimum discriminant decision surfaces.

15. SUBJECT TERMS

Discrimination	Regional Arrays	Analysis-of-Variance
Amplitude Ratios	Site Effects	Underwater Explosions

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE	SAR	62	19b. TELEPHONE NUMBER *(include area code)*
Unclassified	Unclassified	Unclassified			

CONVERSION TABLE

Conversion Factors for U.S. Customary to metric (SI) units of measurement.

MULTIPLY → BY → TO GET

TO GET ← BY ← DIVIDE

angstrom	1.000 000 x E -10	meters (m)
atmosphere (normal)	1.013 25 x E +2	kilo pascal (kPa)
bar	1.000 000 x E +2	kilo pascal (kPa)
barn	1.000 000 x E -28	meter^2 (m^2)
British thermal unit (thermochemical)	1.054 350 x E +3	joule (J)
calorie (thermochemical)	4.184 000	joule (J)
cal (thermochemical/cm^2)	4.184 000 x E -2	mega joule/m^2 (MJ/m^2)
curie	3.700 000 x E +1	*giga bacquerel (GBq)
degree (angle)	1.745 329 x E -2	radian (rad)
degree Fahrenheit	$t_k = (t°f + 459.67)/1.8$	degree kelvin (K)
electron volt	1.602 19 x E -19	joule (J)
erg	1.000 000 x E -7	joule (J)
erg/second	1.000 000 x E -7	watt (W)
foot	3.048 000 x E -1	meter (m)
foot-pound-force	1.355 818	joule (J)
gallon (U.S. liquid)	3.785 412 x E -3	meter^3 (m^3)
inch	2.540 000 x E -2	meter (m)
jerk	1.000 000 x E +9	joule (J)
joule/kilogram (J/kg) radiation dose absorbed	1.000 000	Gray (Gy)
kilotons	4.183	terajoules
kip (1000 lbf)	4.448 222 x E +3	newton (N)
kip/inch^2 (ksi)	6.894 757 x E +3	kilo pascal (kPa)
ktap	1.000 000 x E +2	newton-second/m^2 ($N-s/m^2$)
micron	1.000 000 x E -6	meter (m)
mil	2.540 000 x E -5	meter (m)
mile (international)	1.609 344 x E +3	meter (m)
ounce	2.834 952 x E -2	kilogram (kg)
pound-force (lbs avoirdupois)	4.448 222	newton (N)
pound-force inch	1.129 848 x E -1	newton-meter (N-m)
pound-force/inch	1.751 268 x E +2	newton/meter (N/m)
pound-force/foot^2	4.788 026 x E -2	kilo pascal (kPa)
pound-force/inch^2 (psi)	6.894 757	kilo pascal (kPa)
pound-mass (lbm avoirdupois)	4.535 924 x E -1	kilogram (kg)
pound-mass-foot^2 (moment of inertia)	4.214 011 x E -2	kilogram-meter^2 ($kg-m^2$)
pound-mass/foot^3	1.601 846 x E +1	kilogram-meter^3 (kg/m^3)
rad (radiation dose absorbed)	1.000 000 x E -2	**Gray (Gy)
roentgen	2.579 760 x E -4	coulomb/kilogram (C/kg)
shake	1.000 000 x E -8	second (s)
slug	1.459 390 x E +1	kilogram (kg)
torr (mm Hg, 0^0 C)	1.333 22 x E -1	kilo pascal (kPa)

*The bacquerel (Bq) is the SI unit of radioactivity; 1 Bq = 1 event/s.

**The Gray (GY) is the SI unit of absorbed radiation.

TABLE OF CONTENTS

SECTION PAGE NO.

TABLE OF CONTENTS ..... *i*

LIST OF FIGURES ..... *ii*

ABSTRACT ..... *vi*

1.0 INTRODUCTION *1*

2.0 ANALYSIS OF SITE EFFECTS ON P/S AMPLITUDE RATIOS *5*

2.1 Regional Array Recordings of Underwater Explosion Group in the Gulf of Bothnia .5

2.2 Amplitude Ratio Analysis 9

2.3 Analysis of Variance Approach 17

2.4 Array Configurations and Sensor Separations 18

2.5 Single Array Analysis of Variance 19

2.6 Multiple Array Analysis of Variance 22

3.0 DISCRIMINANT SEPARATION MEASURES, OPTIMUM TRANSFORMATIONS, AND TRANSPORTABILITY *25*

3.1 Transportability Issues 25

3.2 Handling of Missing Data - Generalized EM Algorithm 26

3.3 Separability Measures and Optimum Transformations 28

3.4 Discrimination Results 30

3.4.1 China Data Set 31

3.4.2 NTS Explosion vs. Skull Mountain Data Set 35

3.4.3 Steigen-Novaya Zemlya Data Set 40

4.0 CONCLUSIONS *44*

5.0 REFERENCES *46*

DISTRIBUTION LIST *DL-1*

LIST OF FIGURES

FIGURE **PAGE NO.**

Figure 1: Scatter plot of *Pn/Sn* ratios for earthquakes and explosions in Scandinavia compared. The green arrows indicate the *Pn/Sn* ratios of the 31 December 1992 event (From Ryall, et al, 1995). 3

Figure 2: Map showing propagation paths from the event cluster in the Gulf of Bothnia to regional arrays in Scandinavia. 6

Figure 3: Record section of the Gulf of Bothnia underwater explosions of 8 May 1996 recorded at four regional arrays. 6

Figure 4: Record section plot of waveforms for one of the 8 May 1996 Gulf of Bothnia underwater explosions recorded at the center elements of the regional arrays. 7

Figure 5: Spectra recorded at FINES and NORES by one of the 8 May 1996 presumed underwater blasts in the Gulf of Bothnia. 8

Figure 6: Example of cepstral inversions for two of the presumed underwater blasts in the Gulf of Bothnia. 9

Figure 7: Map showing locations of Gulf of Bothnia events and propagation paths to regional arrays that recorded them. 10

Figure 8: Waveforms recorded at the FINES array from one of the 8 May 1996 Gulf of Bothnia underwater explosions. 11

Figure 9: Waveforms recorded at the NORES array from one of the 8 May 1996 Gulf of Bothnia underwater explosions. 12

Figure 10: Bandpass filter analysis of waveforms recorded at the center element of the FINES array (left) and NORES array (right) from one of the 8 May 1996 Gulf of Bothnia underwater explosions showing the focusing of energy in the mid-frequency band (3-6 Hz). 13

Figure 11: Plot of the RMS incoherent beams for the FINES (left) and NORES (right) array recordings of one of the 8 May 1996 Gulf of Bothnia underwater explosions. Regional phases picked on waveforms are indicated at the top. 14

Figure 12: Log (*Pn/Lg*) ratios (left) and log (*Pg/Lg*) ratios (right) measured at each array element of NORES. 15

Figure 13: Log (*Pn/Lg*) ratios measured at each array element of FINES when the numerator and denominator phase signal-to-noise ratio exceeds 3. 15

Figure 14: Scatter plots of *Pn/Lg* ratios versus *Pn* amplitudes (left) and *Lg* amplitudes (right) measured in the 3-6 Hz band for all array elements of NORES that recorded one of the 8 May 1996 underwater explosions..... 16

Figure 15: Scatter plots of *Pn/Lg* ratios versus *Pn* amplitudes (left) and *Lg* amplitudes (right) measured in the 3-6 Hz band for all array elements of NORES that recorded one of the 8 May 1996 underwater explosions..... 17

Figure 16: Array configurations for the NORES (top left), FINES (top right), and Hagfors (bottom)..... 18

Figure 17: Histograms of NORES site terms in three frequency bands for *Pn/Lg* amplitude ratios, *Pn* amplitudes, and *Lg* amplitudes..... 20

Figure 18: Histograms of NORES site terms in three frequency bands for *Pg/Lg* amplitude ratios, *Pg* amplitudes, and *Lg* amplitudes..... 21

Figure 19: Histograms of NORES site terms in three frequency bands for *Pn/Lg* amplitude ratios, *Pn* amplitudes, and *Lg* amplitudes..... 21

Figure 20: Examples of the effects of the arrays and sites in two frequency bands within arrays on the *Pn/Lg* ratio in two frequency bands obtained from ANOVA2. A generalized version of the EM algorithm was used to fill in a small number of missing observations. The site terms for each array are designated by different colors: black-FINESA, red-Hagfors, and green-NORES..... 22

Figure 21: Examples of explosion and earthquake waveforms for events in China. Typically explosions have lower amplitudes in the shear wave related phases *Sn* and *Lg*..... 31

Figure 22: Scatter plot of the *Pn/Lg* spectral ratio data or four frequency bands from China. Red triangles denote explosions and green dots indicate earthquakes..... 32

Figure 23: Eigenvalue spread derived from the analysis of *Pn/Lg* amplitude ratios for the Chinese station WMQ..... 33

Figure 24: Plots of eigenvectors from the analysis of the China data. The first eigenvector has a dominant component for the 3 Hz band where the best separation exists. The

last gives the largest weight to the first frequency band, the least effective discrimination. 33

Figure 25: Discrimination plots for the China data set. Obviously, the best separation between the explosions (red dots) and the earthquakes (green dots) is achieved by the eigenvector #1 associated with the largest eigenvalue. The second eigenvector gives a small shift in means but large overlap, the remaining two overlap completely. 34

Figure 26: ROC curve for the Fisher discriminant acting on the China data set. 35

Figure 27: Examples of waveforms from NTS explosions recorded at LLNL seismic stations. 36

Figure 28: Examples of waveforms from Skull Mountain earthquakes recorded at LLNL stations. 37

Figure 29: Scatterplots of *Pn/Lg* and *Pg/Lg* ratios for Skull Mountain earthquakes and NTS nuclear explosions at the stations Kanab, Utah (KNB) (a) and Mina, Nevada (MNA) (b). 37

Figure 30: Eigenvalue spreads for the KNB and MNA data sets. 38

Figure 31: Eigenvectors for KNB (a) and MNA (b) computed separately from the amplitude ratios of earthquakes and explosions in Figure 29 in the various frequency bands. 38

Figure 32: Resultant transformed amplitude ratios for the KNB and MNA data respectively after applying the weights in Figure 31 to the amplitude ratios in Figure 30. 39

Figure 33: ROC curve for the Fisher discriminant as applied to the Skull Mountain-Nevada data set. 40

Figure 34: Scatter plots of the *Pn/Sn* spectral ratio populations of the Novaya Zemlya explosion (red) and the Steigen earthquakes. 41

Figure 35: Eigenvalue spread for the four frequency bands of the Steigen-Novaya Zemlya data set. 42

Figure 36: Four eigenvectors corresponding to the four eigenvalues (from top to bottom). 42

Figure 37: Discrimination using the coordinate transformations corresponding to the four

eigenvectors of the Steige-Novaya Zemlya data set. 43

ABSTRACT

The research under this program has investigated problems associated with applying regional-phase amplitude ratios, such as *Pn/Sn* or *Pn/Lg* ratios, for discrimination of explosions and earthquakes for monitoring the CTBT. Using multiple array recordings of groups of events in the same source region, we characterized the factors that contribute bias or the scatter of *P/S* ratio measurements, after correction for path effects. These factors include both receiver site effects and source mechanism effects on *P/S* ratios. The study of site effects has focused on arrays where we have seen site variations in *P/S* ratios, including the Scandinavian regional arrays (NORES, FINES), and other new arrays in the International Monitoring System (IMS). The variance in the *P/S* ratio around regional arrays and large aperture arrays reveals the extent to which site affects cause variations in *P/S* ratios around different arrays in different regions. The partitioning of the variance between source, path, and receiver effects is examined by analysis of variance (ANOVA). We have performed an initial study of a group of presumed underwater explosions in the Gulf of Bothnia recorded by regional arrays in Scandinavia. We find that *P/S* and *P/S* amplitude ratios vary by as much as a factor of 3 around the FINES and NORES arrays, with apertures of 3 km, as well as similar variations for the different sources. These variations appear to be driven by variations in *Pn* and *Pg* amplitudes, whereas *Lg* amplitudes appear to be more stable.

We also performed a statistical study of the transportability of *P/S* ratio discriminants using separability measures and optimum transformations in order to reduce dimensionality of multiple frequency *P/S* ratios. These transformations consist of

calculating the intra-class and inter-class scatter matrices for *P/S* ratio discriminants and using the eigenvectors, corresponding to the largest eigenvalues, of the inter-class matrix to compute optimum transformation of discriminants that provide the best separation. We applied this analysis to distance-corrected discriminants in different regions (e.g., China, Eurasia, North America) in order to compare discriminant effectiveness for different regions and to evaluate the transportability of optimum discriminant decision surfaces.

Key Words: discrimination, amplitude ratios, regional arrays, site effects, analysis-of-variance, underwater explosions

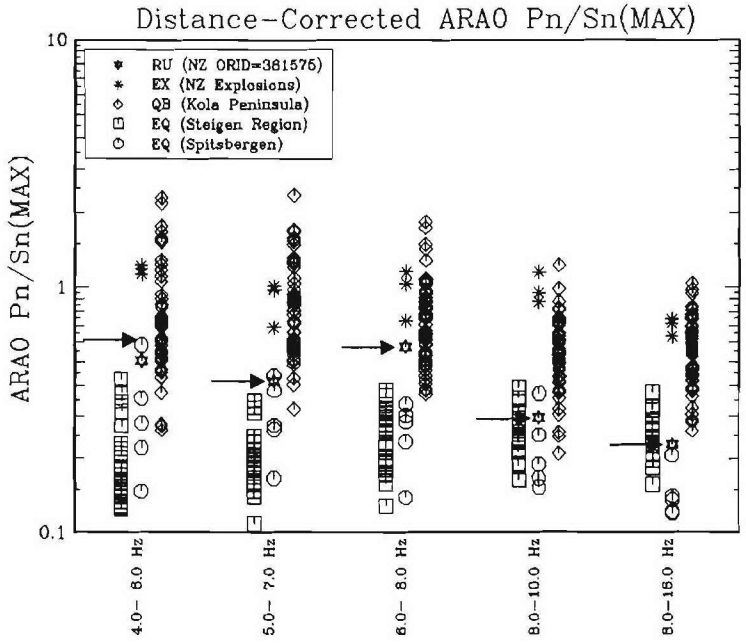
1.0 INTRODUCTION

The International Monitoring System (IMS) for nuclear test monitoring faces the serious challenge of being able to accurately and reliably identify seismic events in any region of the world. This requirement extends to a very low magnitude threshold, mb=2.5, which is in the range of the sizes of local and regional seismic activity, both natural and artificial. Much research has been performed in recent years on developing discrimination techniques that classify seismic events into broad categories of source types, such as nuclear explosion, earthquake, and mine blast.

The seismic waveform discriminant which has been commonly investigated is the regional *P(Pn, Pg)/S(Sn, Lg)* amplitude ratio. Seismic source physics suggests that earthquakes, being dislocation sources, should be intrinsic sources of shear waves whereas explosions, being pure compressional sources, should only generate *P* waves. Therefore, explosions should have higher *P/S* amplitude ratios than earthquakes. This has been generally observed to be true, although the separation of explosions and earthquakes amplitude ratio is larger at high frequency (> 5 Hz) than at lower frequencies.

Observationally, nuclear explosions and earthquakes appear to be well separated by this discriminant (e.g., Baumgardt, 1993; Baumgardt and Der, 1995; Hartse et al, 1997). For example, Russian nuclear explosions observed at a Chinese station WMQ records no shear wave energy at frequencies above 6 Hz whereas Chinese earthquakes produce significant shear wave energy above 6 Hz. However, studies of mine blasts in Scandinavia and Germany (Baumgardt, 1993) indicate that many of the mine blasts seem to be intrinsic sources of shear waves, perhaps because they induce shear in fracturing and spallation in mines. Thus, low *P/S* ratios may be an indication of earthquakes, but many mine blasts may also have low values. However, we generally observe that most nuclear explosions will have high *P/S* ratios at high frequency compared to earthquakes, and mine blasts can also have high *P/S* ratios at high frequency.

Although *P/S* ratios appear to give promising discrimination between earthquakes and explosions, they have also been shown to have a high degree of scatter which reduces the confidence of identification using such discrimination techniques as the outlier method (Fisk et al, 1996). An example of this scatter is shown in Figure 1. This example shows discrimination analysis of an event that occurred on 31 December 1992 near the former Soviet Union test site at Novaya Zemlya. The figure shows *Pn/Sn* amplitude ratios for populations of nuclear explosions on Novaya Zemlya, quarry blasts on the Kola Peninsula, and earthquakes in Scandinavia, all recorded by the ARCES array in northern Norway, that clearly shows that nuclear explosions tend to have higher ratios than earthquakes and most regional recordings of mine blasts. The 31 December 1992 event, indicated by the arrow, falls in the lower part of the explosion category at low frequency and in the earthquake category at high frequency. However, the scatter in the points is quite large, and there is considerable overlap in the ratios for nuclear explosions and earthquakes. Although the 31 December 1992 event seems to be outside the nuclear-explosion population, the large scatter in data points makes it equivocal to identify the event as an earthquake or mine blast.



Arrows Indicate Data Points for 31 December 1992 Event

Figure 1: Scatter plot of *Pn/Sn* ratios for earthquakes and explosions in Scandinavia compared. The green arrows indicate the *Pn/Sn* ratios of the 31 December 1992 event (From Ryall, et al, 1995).

Likely causes of this scatter includes the following:

1. Propagation path effects – These include differential attenuation of *P* and *S* and unmodeled propagation path effects, such as variations in elevation, crustal depth, and depth to basement (sediment thickness). Empirical studies of variations of *P/S* ratios with distance (e.g., Fisk et al, 1996) have resulted in distance corrections. Correlation studies for *P/S* amplitude ratios with crustal parameters (e.g., Zhang et al, 1994; Fan et al, 2001) have demonstrated that these correlations can reduce the variance of *P/S* ratios caused by unmodeled path effects.

2. Source effects – These may include ripple fire patterns in mine blasts, which can usually be identified by spectral techniques, magnitude differences (Xie and Patton, 1999; Ringdal et al, 2000), and possible differential radiation pattern effects on *P* and *S* amplitudes. The latter has usually been assumed to be small for

high frequency regional waves. As mentioned above, mine blasts may also intrinsically excite shear waves to different degrees, depending on the local tectonic environment and the blasting practice, which can contribute high variance in *P/S* ratios. Finally, variations in depth of focus of earthquakes may also produce significant variations in *P/S* ratios.

3. Site effects – These include variations in *P* and *S* amplitudes caused by variations in the geology immediately below the site itself. These effects are usually local and not included in propagation-path corrections. Baumgardt and Der (1995) showed examples of site variance effects around the Iranian Long Period Array (ILPA) where both earthquake- and explosion-like *Pn/Lg* amplitude ratios were observed for sensors separated by several kilometers.

This report addresses possible causes of scatter and bias in the use of regional *P/S* amplitude ratio discriminants that have not been much studied in previous research, although their importance has been noted. These effects may contribute to the residual variance in distributions of *P/S* ratios for earthquake populations even after the application of propagation-path corrections. These include station site effects, perhaps due to variation in amplification of *P* and *S* waves by variable site geology, on the *P/S* ratios, as evidenced by the variation in *P/S* amplitude ratios around regional arrays. Also, earlier research (Baumgardt, 1996) has revealed that source radiation patterns from earthquakes in the Zagros Mountains of western Iran, recorded at ILPA, may cause significant variation in *Pn/Lg* ratios. This study follows up on that observation and investigate the effect in more detail to determine if the effect is due to source or propagation-path effects. Overall, this study provides a method for estimating the likely maximum a priori variance that may be caused by these site effects and which may be useful in discrimination studies that must rely on single site measurements of regional *P/S* ratios.

2.0 ANALYSIS OF SITE EFFECTS ON P/S AMPLITUDE RATIOS

In this section, we describe a method for estimating site variances using regional arrays. Using multiple array recordings of groups of events in the same source region, we characterize the factors that contribute bias or the scatter of *P/S* ratio measurements. These factors include both receiver site effects and source effects on *P/S* ratios. The study of site effects focused on arrays where we have seen site variations in *P/S* ratios, including the Scandinavian regional arrays (NORES, FINES), and other new arrays in the International Monitoring System. The variance in the *P/S* ratio around regional arrays and large aperture arrays reveals the extent to which site affects cause variations in *P/S* ratios around different arrays in different regions.

2.1 Regional Array Recordings of Underwater Explosion Group in the Gulf of Bothnia

We have chosen, as an initial study, to analyze a group of events located in the Gulf of Bothnia, shown in a previous study (Baumgardt, 1999) to be several underwater explosions that occurred there in a single day on 8 May 1996. These events were discovered by searching the Reviewed Event Bulletin (REB) of the Prototype International Data Center (PIDC) for events located in offshore areas. These events constitute an event cluster, with apparently nearly the same magnitudes (Ml between 3.4 and 3.6), that appear to be underwater explosions.

The locations and the propagation paths to Scandinavian regional seismic arrays are shown in Figure 2. A record section of the events recorded at the center elements of the regional arrays pictured in Figure 2 are shown in Figure 3. The phase identifications made on each of the events are shown. The plot shows that on this day eight events occurred, all apparently in the same location.

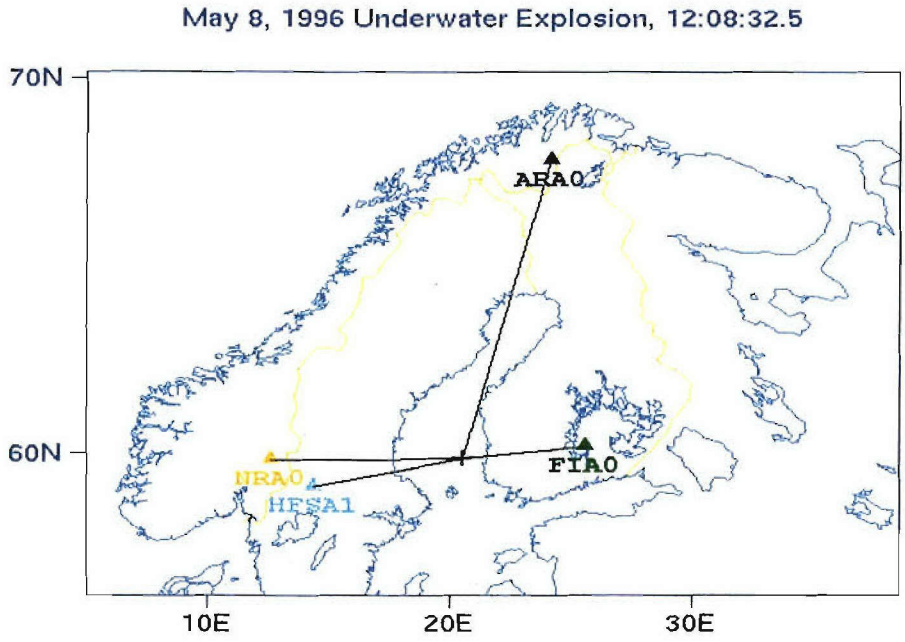


Figure 2: Map showing propagation paths from the event cluster in the Gulf of Bothnia to regional arrays in Scandinavia.

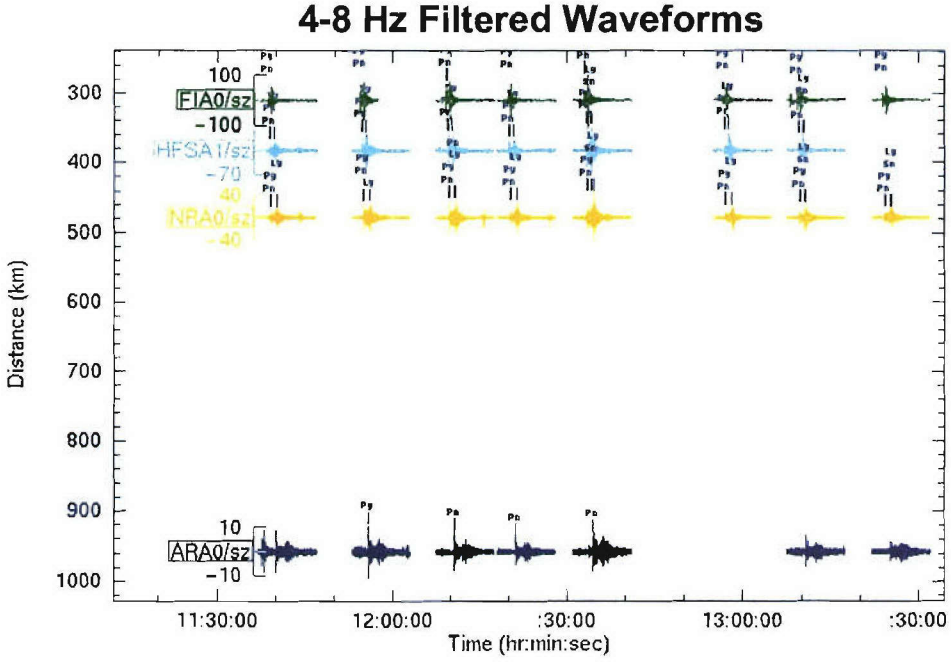


Figure 3: Record section of the Gulf of Bothnia underwater explosions of 8 May 1996 recorded at four regional arrays.

Figure 4 shows a record section plot of one the events in the Gulf of Bothnia. All waveforms were passed through a 4-8 Hz bandpass filter. The regional phases are clearly observed at all the stations out to a distance of 950 km. The large amplitudes suggest that the events are in fact underwater explosions.

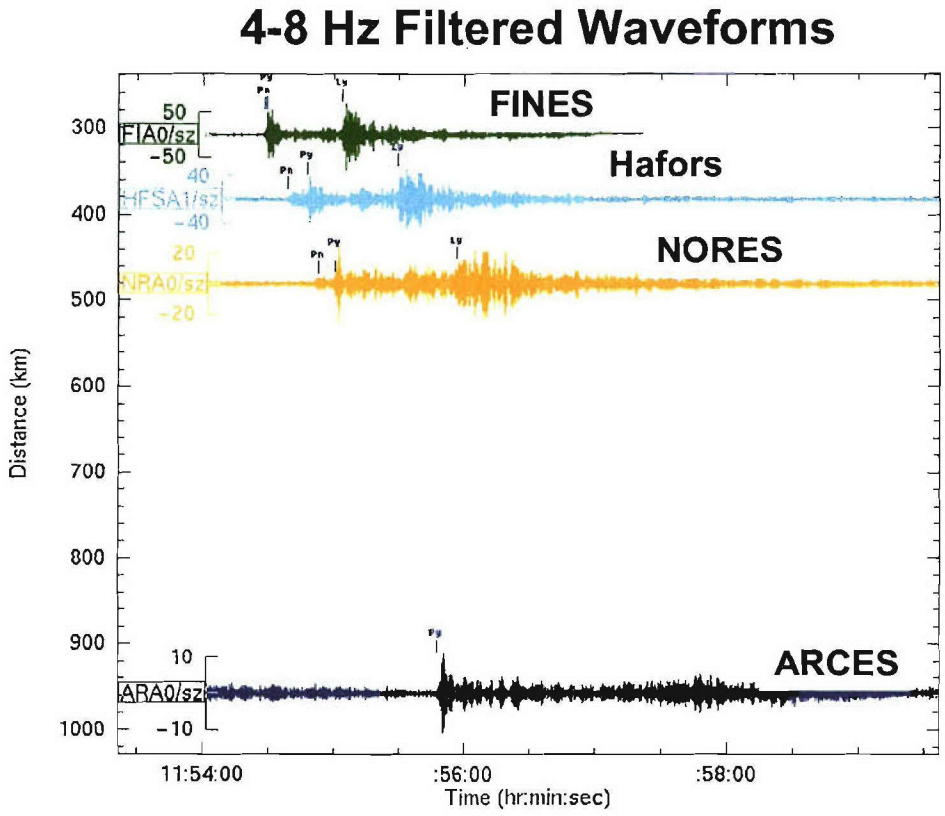


Figure 4: Record section plot of waveforms for one of the 8 May 1996 Gulf of Bothnia underwater explosions recorded at the center elements of the regional arrays.

Baumgardt and Der (1998) previously discovered other events in the Gulf of Bothnia and showed that their spectral and cepstral characteristics were consistent with those expected from underwater explosions. Baumgardt (1999) described a cepstral modeling and inversion approach for inferring the depth and yield of underwater explosions by modeling and inverting cepstra for underwater explosions. For the 8 May 1996 group of presumed underwater explosions, we found that the cepstra were very similar, and the resulting inversions gave very similar results in terms of explosion yield and depth in the water column.

Examples of spectra for waveforms recorded by one of the events at FINES and NORES are shown in Figure 5. The FINES spectra show very strong spectral scalloping, also observed at NORES, although less apparent. This feature results in interference of acoustic waves reflecting in the water column and bubble pulses, which convert to seismic phases at the water-bottom rock interface.

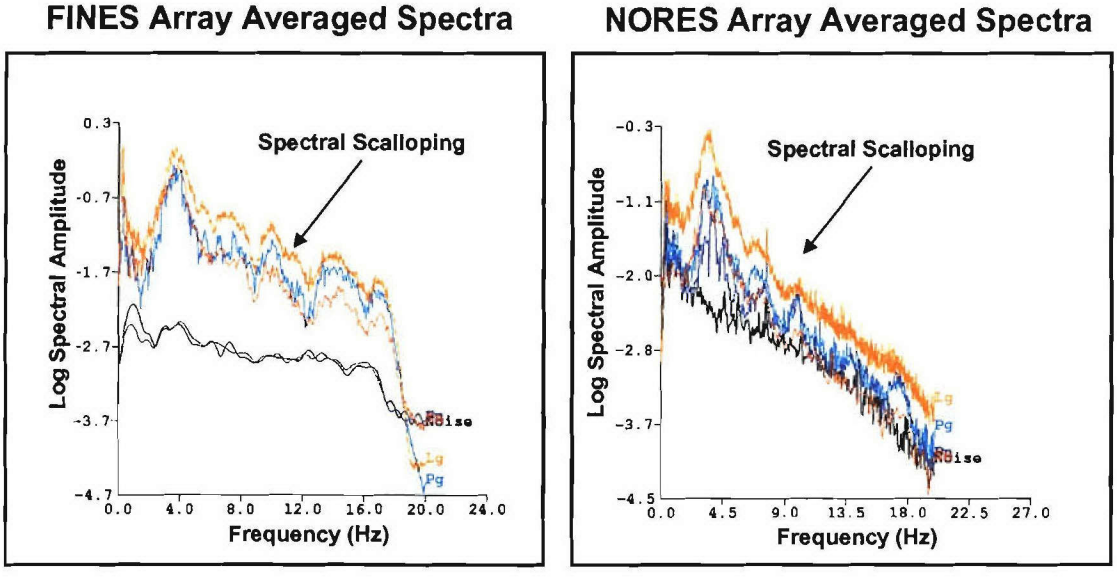


Figure 5: Spectra recorded at FINES and NORES by one of the 8 May 1996 presumed underwater blasts in the Gulf of Bothnia.

Examples of cepstral inversions for two of the events are shown in Figure 6, which give water depths of 60 and 82 m, and explosive yields of 156 and 177 kg. This result is typical of all the events in the group, which gave yields ranging between 141 to 183 kg and depths from 69 to 72 m. These depths were consistent with the known bathymetric depths in the Gulf of Bothnia. There may have been some correlation between local magnitude and yield, since the cepstra for events with local magnitudes of 3.6 gave the higher yields of between 179 to 183 kg whereas the 3.4 to 3.5 events gave yields between 141 and 165 kg.

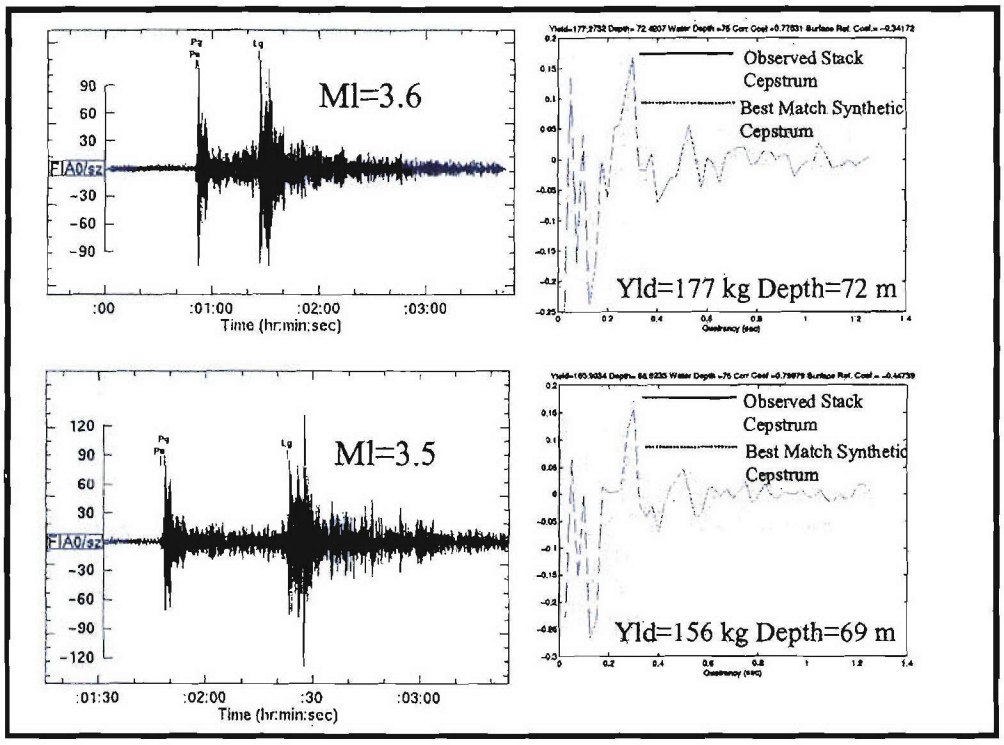


Figure 6: Example of cepstral inversions for two of the presumed underwater blasts in the Gulf of Bothnia.

2.2 Amplitude Ratio Analysis

For the amplitude ratio analysis, we focused on the regional arrays that recorded most of the presumed underwater explosions in the Gulf of Bothnia. Figure 7 shows an expanded map of the region showing the propagation paths to each of the center elements of the regional arrays. NORES and Hagfors are nearly at the same azimuth. In this report, we will focus primarily on the NORES and FINES data.

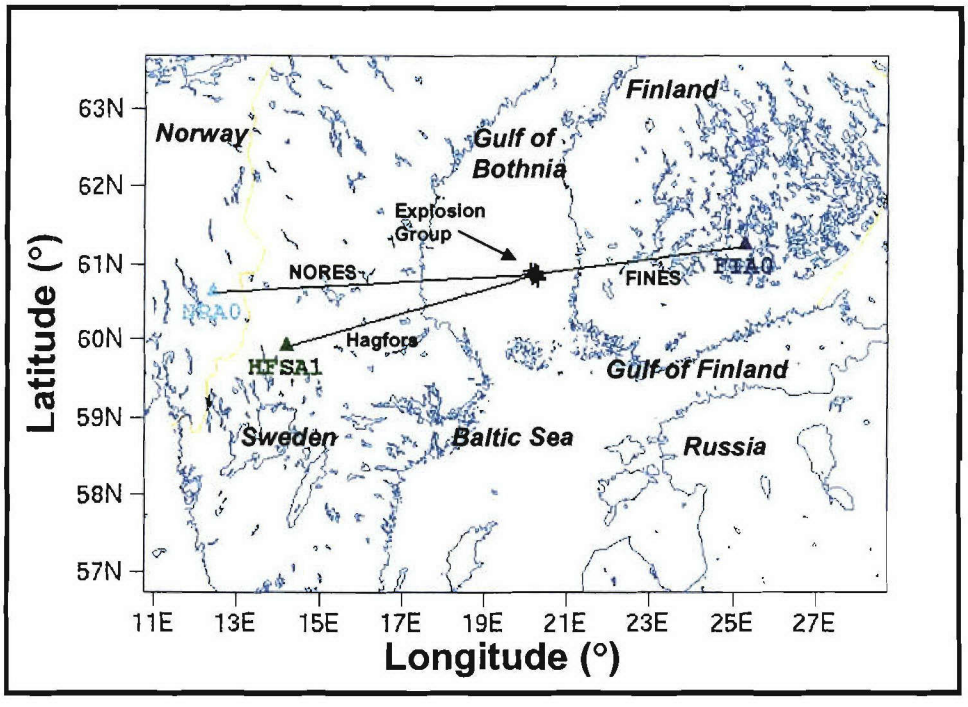


Figure 7: Map showing locations of Gulf of Bothnia events and propagation paths to regional arrays that recorded them.

Figure 8 and Figure 9 show plots of the waveforms at the different array elements at FINES and NORES, respectively, from one of the 8 May 1996 events. The FINES array elements are between 304 and 306 km from the events and *Pn* and *Pg* are difficult to separate there. NORES, on the other hand, at distances between 481 and 484 km, had clearly observable *Pn* and *Pg* phases. Both arrays recorded strong *Lg* waves from all the events.

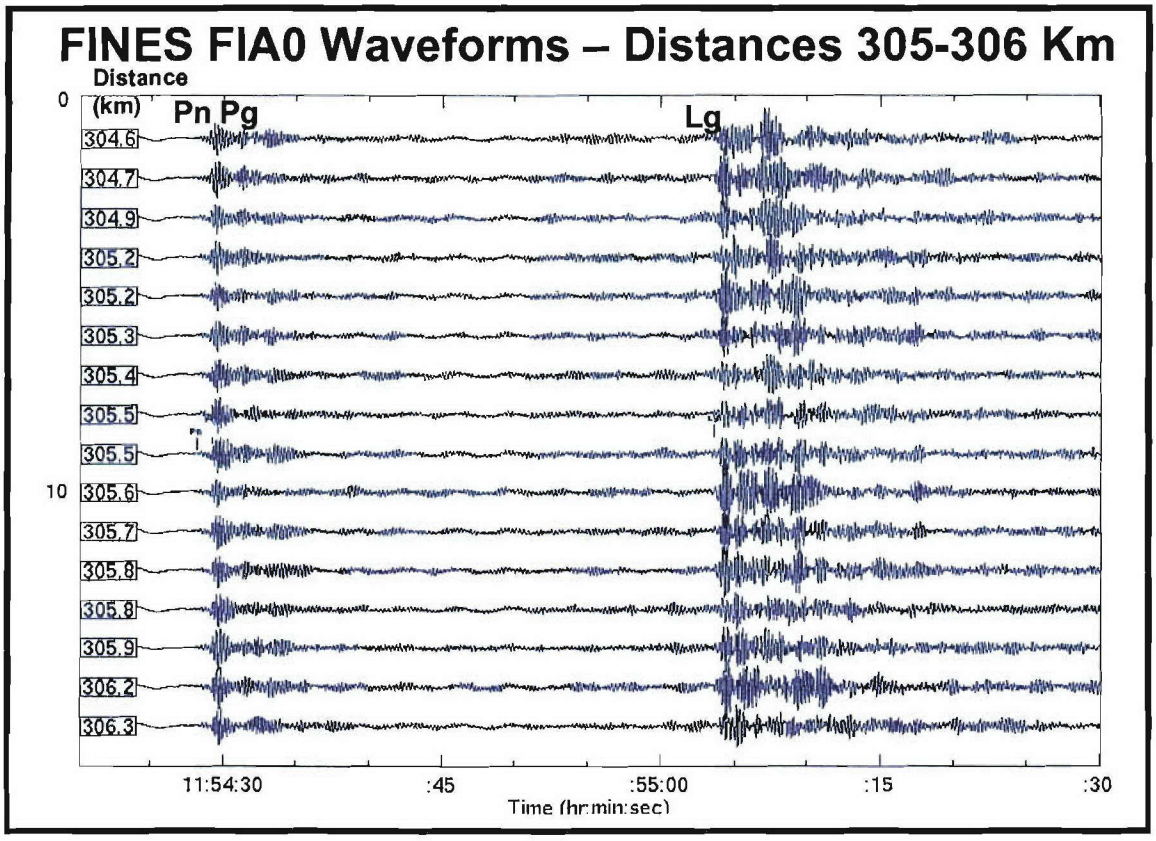


Figure 8: Waveforms recorded at the FINES array from one of the 8 May 1996 Gulf of Bothnia underwater explosions.

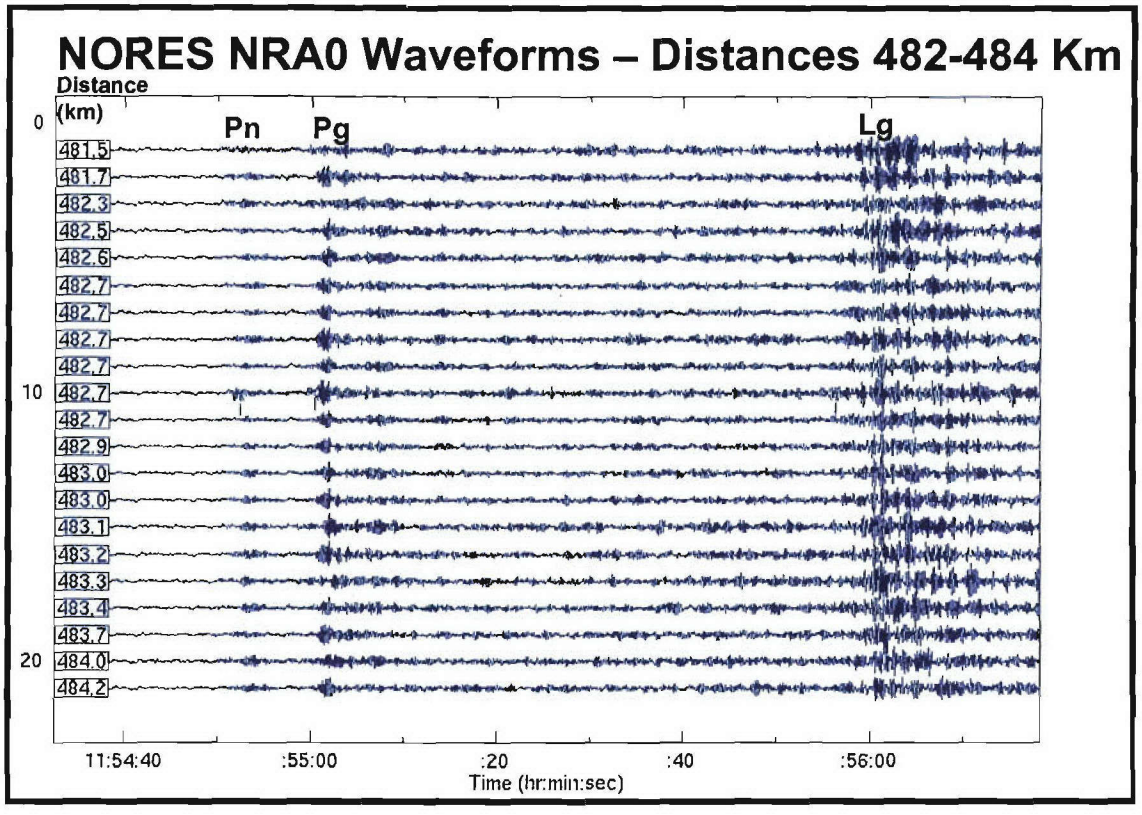


Figure 9: Waveforms recorded at the NORES array from one of the 8 May 1996 Gulf of Bothnia underwater explosions.

Figure 10 shows bandpass filtered waveforms recorded at the center elements of each array. These plots show that the range of frequencies for the highest signal-to-noise ratios is between 1.5 and 18 Hz for FINES and 1.5 and about 12 Hz for NORES. The peaking of the signal-to-noise ratios in the 2 to 4 Hz band at both arrays is due to the spectral modulations produced by the bubble pulse and surface reflections, discussed above.

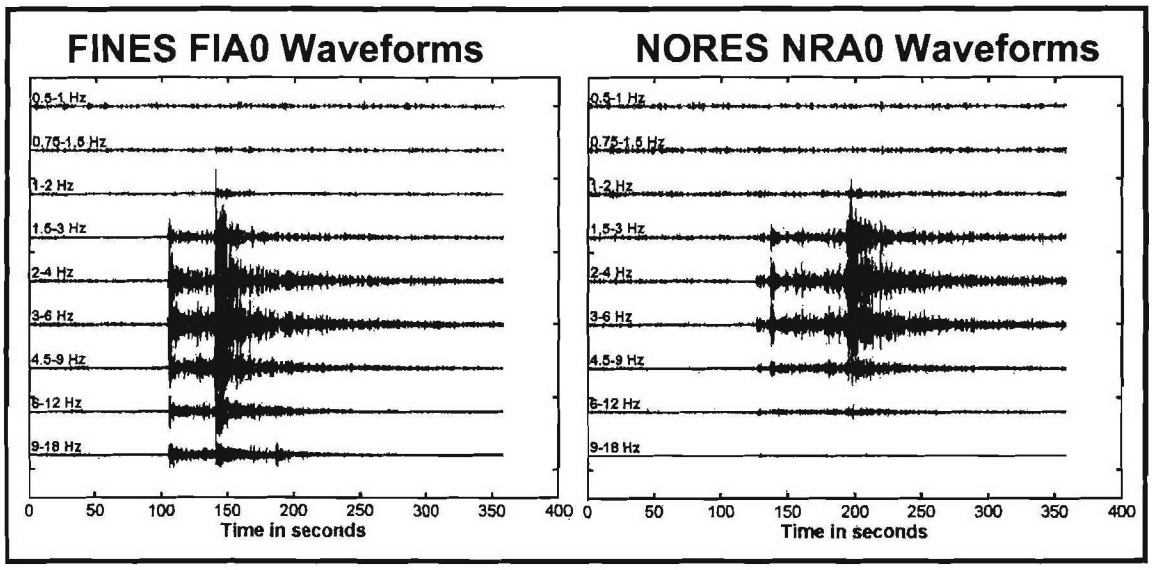


Figure 10: Bandpass filter analysis of waveforms recorded at the center element of the FINES array (left) and NORES array (right) from one of the 8 May 1996 Gulf of Bothnia underwater explosions showing the focusing of energy in the mid-frequency band (3-6 Hz).

We measure the *Pn/Lg* and *Pg/Lg* ratios of filtered waveform envelopes computed by calculating the RMS amplitudes in 1-second windows shifted down the traces. Figure 11 shows the array-stacked envelopes, called incoherent beams, with the phase picks shown. These same envelopes have been computed for each array element at NORES and FINES and the maximum phase amplitudes were measured in a 5-second window following the *Pn*, *Pg*, and *Lg* phase picks. Figure 11 shows that only *Pn* and *Lg* were picked at FINES since the array was too close to the source to observe separation of *Pn* and *Pg*. Also, as shown on the NORES plot, *Sn* phases can be observed on the incoherent beams although they were difficult to pick on the actual waveforms. However, in this study we did not include the *Sn* phase at NORES.

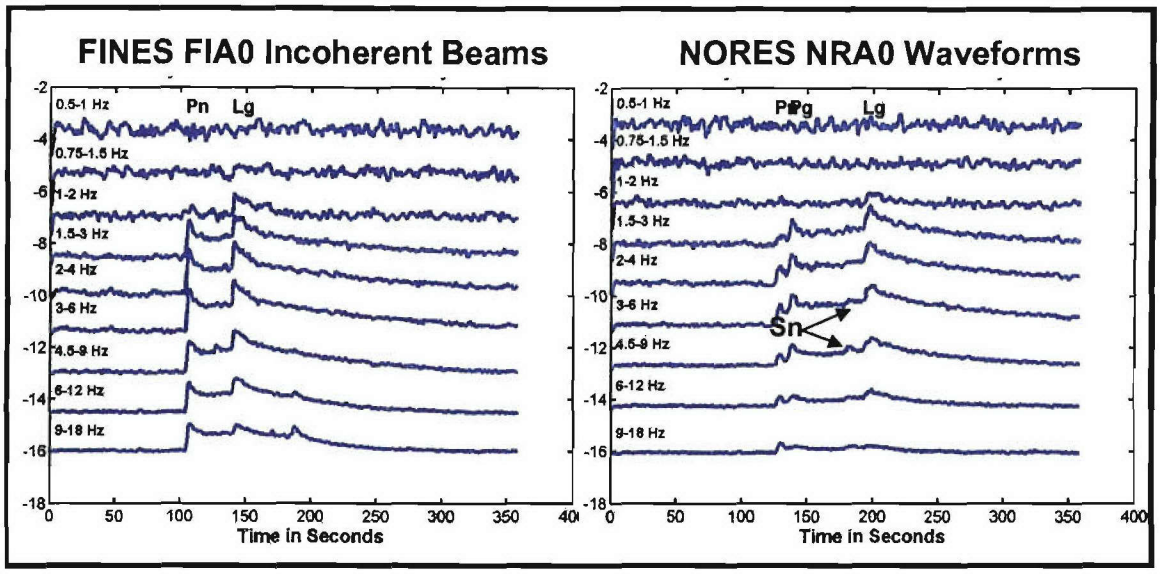


Figure 11: Plot of the RMS incoherent beams for the FINES (left) and NORES (right) array recordings of one of the 8 May 1996 Gulf of Bothnia underwater explosions. Regional phases picked on waveforms are indicated at the top.

Amplitudes of phases are measured off maximum values of RMS envelope plots of each channel and on incoherent beams. *Pn*, *Pg*, and *Lg* amplitudes were determined in all the filter bands shown in Figure 11. *P/S* and *P/S* ratios were computed only when the signal-to-noise ratios exceeded 3. These measurements were made for eight events in the Gulf of Bothnia. The same analysis was also applied to the Hagfors array, although in this study we focus primarily on the FINES and NORES arrays.

Plots of the measurements of *Pn/Lg* and *Pg/Lg* ratios versus frequency for all the NORES array elements are shown in Figure 12. The *Pg/Lg* ratios for the FINES array are shown in Figure 13. In both plots, only ratios where the signal-to-noise ratios of both phases exceeded 3 are plotted, and most of these points fall in the filter frequency bands centered around the 3-to-6 Hz band. For NORES, the ratios increase with frequency as expected. For FINES, the ratios decrease in the mid-frequency band, then increase. This may reflect differences in the site effects at FINES. It should be noted that the "*Pn*" phase at FINES is actually *Pn* and *Pg* in combination, and the different character in the *Pn/Lg* trend may relate to interference effects of the two phases.

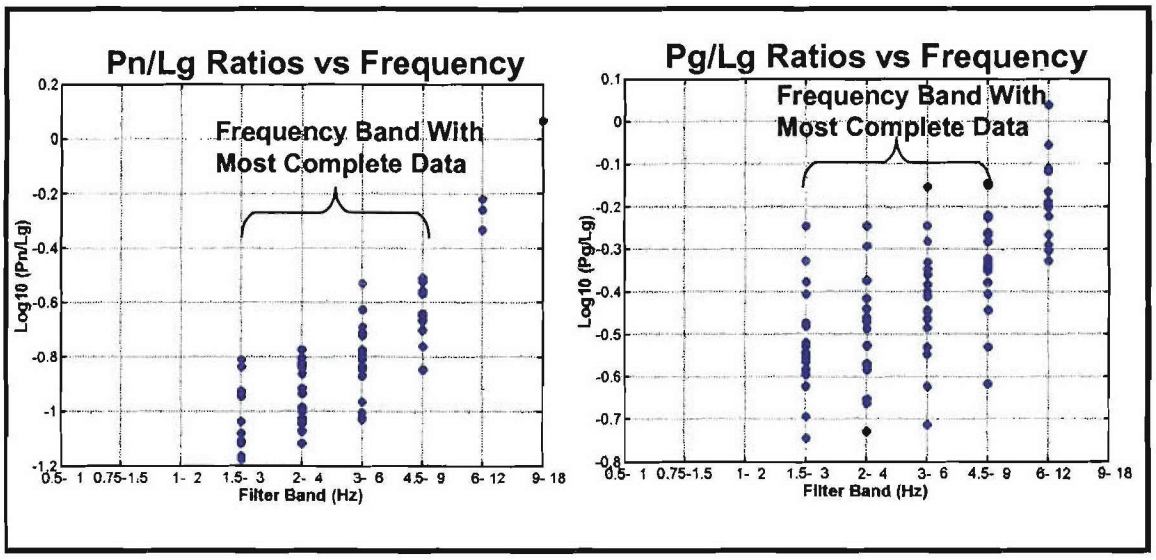


Figure 12: Log (*Pn/Lg*) ratios (left) and log (*Pg/Lg*) ratios (right) measured at each array element of NORES.

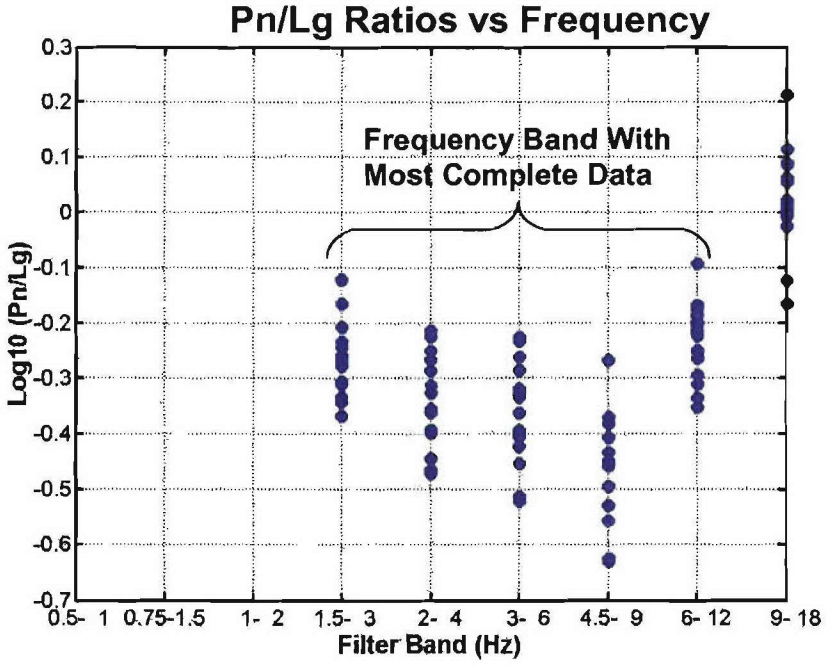


Figure 13: Log (*Pn/Lg*) ratios measured at each array element of FINES when the numerator and denominator phase signal-to-noise ratio exceeds 3.

Figure 12 and Figure 13 show that there is a sizable variance in the ratios. We now investigate how much of this variance is attributable to site effects under the different elements of the arrays.

Figure 14 and Figure 15 show examples of scatter plots of log *Pn/Lg* ratios in the 3 to 6 Hz band at FINES and NORES array elements plotted versus the log *Pn* and *Lg* amplitudes for one of the events. These plots show that *Pn/Lg* amplitude ratios have a range of about .3 to .5 log units, or between factors of 2 to 3. Moreover, the plots of *Pn/Lg* versus *Pn* amplitudes indicate higher correlation than those versus *Lg* amplitudes. This indicates that the *Pn* amplitude variations around both FINES and NORES control the *Pn/Lg* amplitude variations.

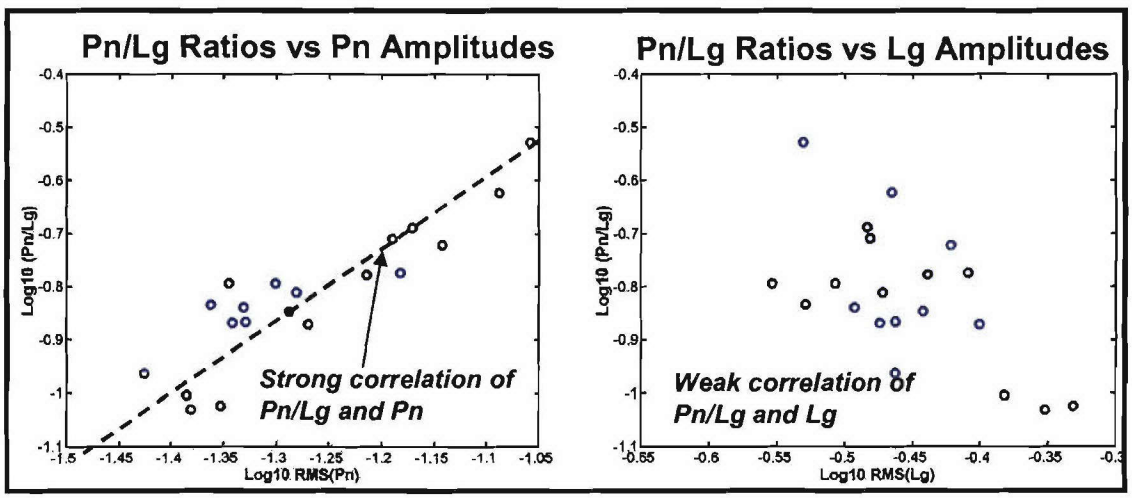


Figure 14: Scatter plots of *Pn/Lg* ratios versus *Pn* amplitudes (left) and *Lg* amplitudes (right) measured in the 3-6 Hz band for all array elements of NORES that recorded one of the 8 May 1996 underwater explosions.

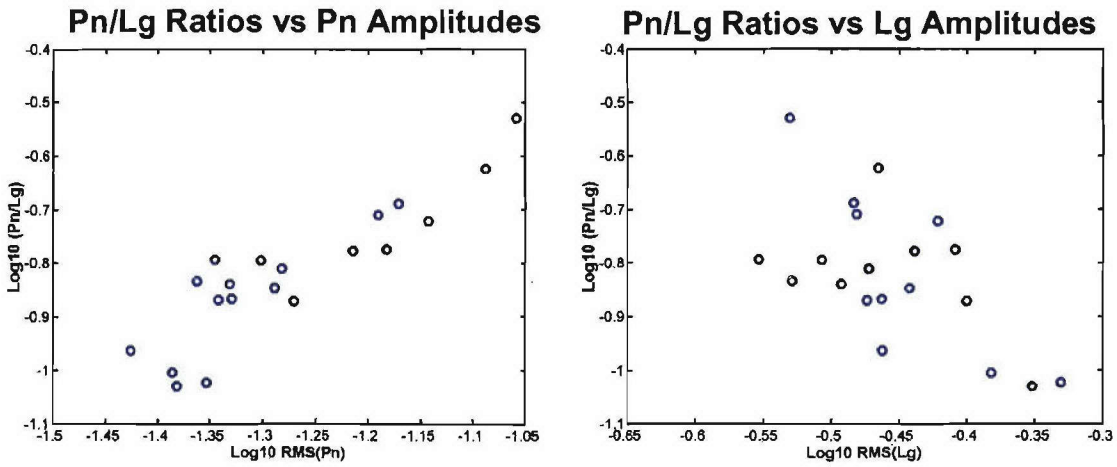


Figure 15: Scatter plots of *Pn/Lg* ratios versus *Pn* amplitudes (left) and *Lg* amplitudes (right) measured in the 3-6 Hz band for all array elements of NORES that recorded one of the 8 May 1996 underwater explosions.

2.3 Analysis of Variance Approach

To explore this more rigorously, we apply a two-way analysis of variance (ANOVA2) test for the variation in the amplitudes of *Pn*, *Pg*, and *Lg* and the amplitude ratios of *Pn/Lg* and *Pg/Lg* for the Gulf of Bothnia events. In ANOVA2, we fit the following model to the amplitudes and amplitude ratios:

$$y_{ijk} = \mu + \alpha_j + \beta_i + \varepsilon_{ijk}$$

where y_{ijk} are the array element amplitudes or amplitude ratios, μ is the mean of all the data, α_j is the source term, β_i, is the site term, and ε_{ijk} is the error term.

This model is fit to the logarithms of amplitudes and amplitude ratios where it is assumed that the amplitudes and ratios are log-normal and dependent on both the additive event and the site effects, both of which are specified to have a zero sum. The errors are assumed to be normally distributed. ANOVA2 in essence tests the hypothesis that the data come from the same population with a common mean (Johnson and Wichern, 1988).

In this study, we focus on the variations in the site terms and ignore for the time being the significance tests on the commonality of the mean. Throughout this study a generalized version of the EM algorithm was used to fill in a small number of missing observations, which consisted of substituting the sum of the event and site means computed from the existing data for the missing values. This is appropriate since the magnitudes are logarithms of amplitudes. Although this does not fit the exact definition of the EM algorithm, such approximate but satisfactory procedures are widely used in practice to avoid undue complexity in calculations. Events with more than three missing values at any array were discarded.

2.4 Array Configurations and Sensor Separations

Figure 16 shows the configurations of the FINES and NORES arrays with color-coding for the different array elements, grouped by rings

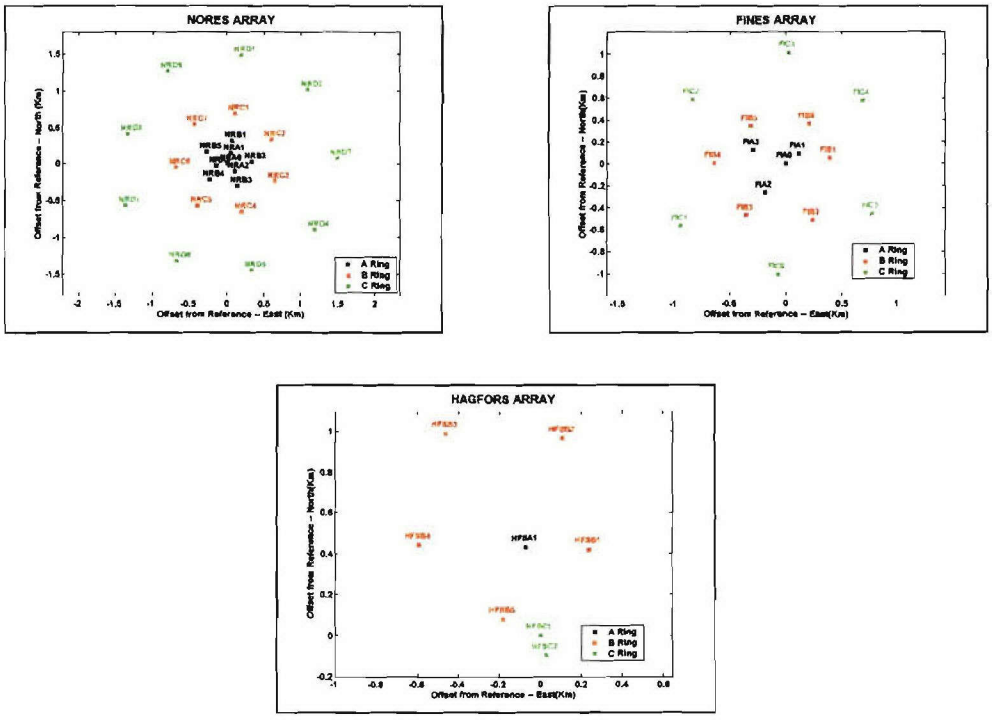


Figure 16: Array configurations for the NORES (top left), FINES (top right), and Hagfors (bottom).

For NORES and FINES, the colors black, red, and green refer to array element rings with increasing separation. For the Hagfors array, where there is actually only one ring, the colors were similarly assigned to represent separation of the sensors. The reason for the color coding is that site terms can be analyzed in terms of their spatial separation. As the colors change from black to red to green for NORES and FINES, the spatial separation of the sensors and the overall aperture of the ring is increasing. For NORES, the spatial dimensions are on the order of ± 0.5 km for black, ± 1 km for red, and ± 1.5 km for green. For FINES, the dimensions are on the order of ± 0.25 km for black, ± 0.75 km, and ± 1 km for green.

The Hagfors array has fewer sensors elements, so it is not actually possible to assign colors to separation. We assigned black to the center element, HFSA1, and red to the single B ring, and green to the HFSC1 and HFSC2 sensors that are outside that ring. Overall, the array is over 1 km across. In the analysis of single array site variations, we focus primarily on the NORES and FINES arrays which have at least three sensor separation groups. We will later include the Hagfors array in a multi-array analysis of site variations.

2.5 Single Array Analysis of Variance

We first consider the application of this analysis to single arrays. In essence, the goal is to determine how the variations in regional phase amplitudes and amplitude ratios depend on the phase ratio used and on site separation. This study was limited to the NORES and FINES arrays.

Figure 17 shows the site terms for NORES in the 2 to 4 Hz, 3 to 6 Hz, and 4.9 to 9 Hz frequency bands obtained by this analysis for *Pn/Lg* amplitude ratios and *Pn* and *Lg* amplitudes.

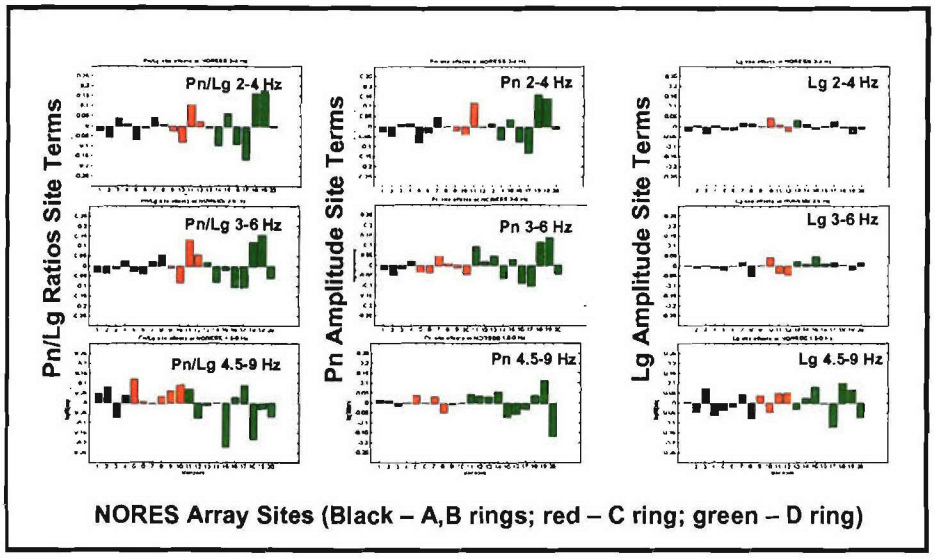


Figure 17: Histograms of NORES site terms in three frequency bands for *Pn/Lg* amplitude ratios, *Pn* amplitudes, and *Lg* amplitudes.

The histograms refer to the site-term values for the different array elements in the rings color-coded as in Figure 16. These plots show that the greatest variability in *Pn/Lg* ratios occurs for the NORES sensors that have the greatest spatial separation or aperture, i.e., rings C and D. The *Pn* phase also has a correspondingly large systematic variation that is largest for the sensors in the rings with the largest aperture. The *Lg* site terms are relatively small, except for the 4.5 to 9 Hz band, and essentially independent of the separation of the sensors. Note also that the site terms for the *Pn/Lg* ratio and *Pn* amplitudes are very similar, which indicates that the *Pn/Lg* ratios are more correlated with the *Pn* amplitudes than with the *Lg* amplitudes. This result is consistent with what was discovered for a single event, shown in Figure 14, and shows that for all the events in the group, variations in *Pn* amplitude drive the variations in the ratio of *Pn* and *Lg* amplitudes.

The same result resulted for the *Pg/Lg* amplitude ratios, and *Pg* and *Lg* amplitudes at NORES, shown in Figure 18. However, site terms in the high frequency band (4.5 to 8 Hz) appear to be more random.

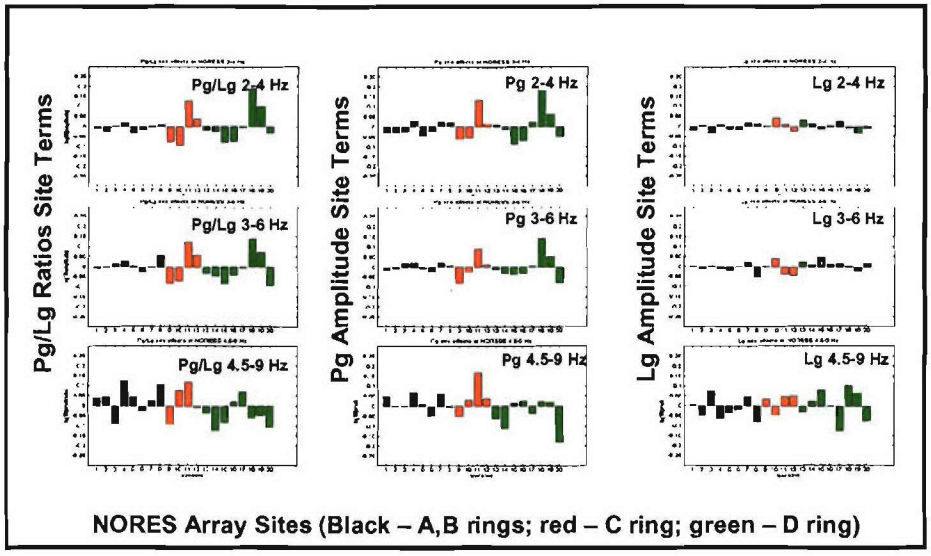


Figure 18: Histograms of NORES site terms in three frequency bands for *Pg/Lg* amplitude ratios, *Pg* amplitudes, and *Lg* amplitudes.

Figure 19 shows the variation in site terms for the *Pn/Lg* ratios at FINES. These again show the same systematic variations.

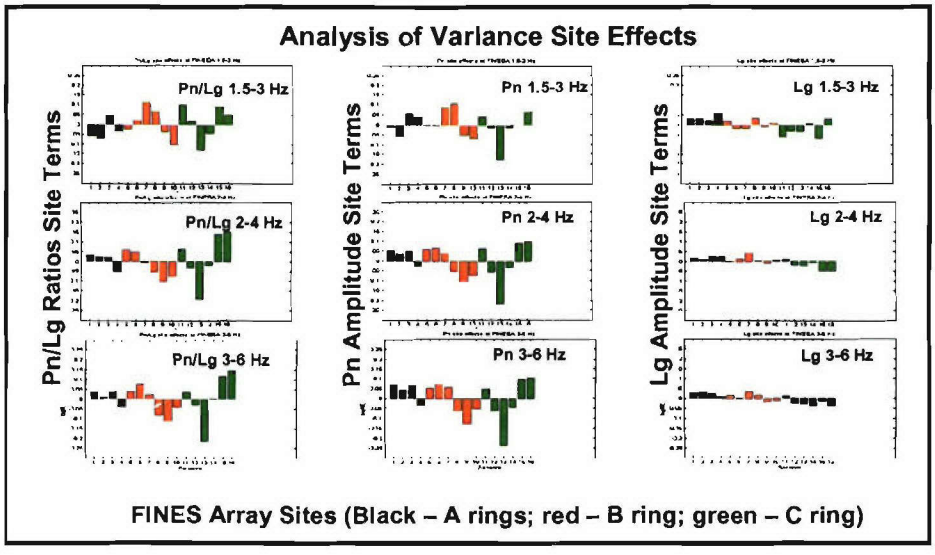


Figure 19: Histograms of NORES site terms in three frequency bands for *Pn/Lg* amplitude ratios, *Pn* amplitudes, and *Lg* amplitudes.

The source terms had a similar variation, which is on the order of 0.3 to 0.5 in the log of the *Pn/Lg* and *Pg/Lg* amplitude ratios. Again, it is apparent that site variations in *Pn/Lg* correlate with site variations in *Pn* more than those, and that these variations increase with

sensor separation. The *Lg* variations around FINES are surprisingly small.

2.6 Multiple Array Analysis of Variance

We also applied the ANOVA algorithm to the *Pn, Lg*, and *Pn/Lg* amplitude ratios measured for three seismic arrays, NORES, FINES, and Hagfors. The results are shown in Figure 20.

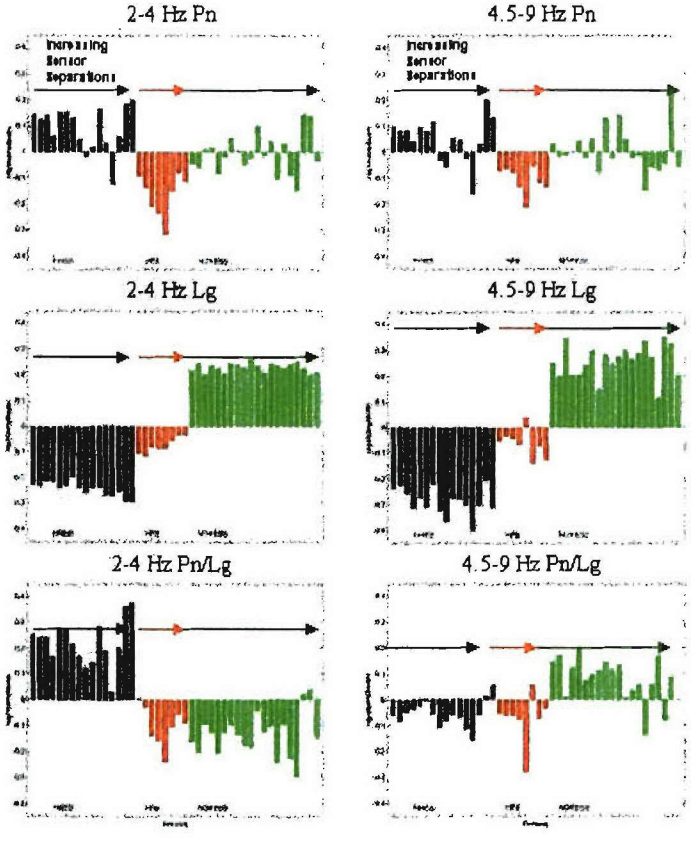


Figure 20: Examples of the effects of the arrays and sites in two frequency bands within arrays on the *Pn/Lg* ratio in two frequency bands obtained from ANOVA2. A generalized version of the EM algorithm was used to fill in a small number of missing observations. The site terms for each array are designated by different colors: black-FINESA, red-Hagfors, and green-NORES.

The results of the analysis in two frequency bands, 2-4 Hz and 4.5-9 Hz in Figure 20 indicate that there are strong biases due to each array and site effects within each array. The ANOVA analysis of single arrays, described by Baumgardt et al (2001), gave *Pn* amplitude variations that controlled the variations in *Pn/Lg* site terms and the *Lg* site terms were more stable across the arrays. This is clearly evident in Figure 20 for the ANOVA applied to three arrays in combination, where the *Pn* site corrections have somewhat random variations of positive and negative values, with the exception of HFS in the 2-4 Hz band. The *Lg* variations, however, all have the same sign for each array. Thus, the random variations in the *Pn* site terms map into the *Pn/Lg* corrections.

The results of this study indicate that the principal site effect contribution to *Pn/Lg* variations is coming from the *Pn* site effects. However, when examining the variations between sites, the *Lg* variations are larger in magnitude than the *Pn* variations. Note that these amplitudes have been corrected for distance, so that these variations appear to be coming primarily from the site effects. Thus, when using the *Pn/Lg* ratio to identify seismic events, this variance needs to be factored into the estimates of the measurement precision for *Pn/Lg* ratios.

The results of this study indicate that the variations in *Pn* amplitudes are greater than those of *Lg* amplitudes, and that these variations increase with sensor separation. One possible explanation for this is that *Pn* amplitude variations around the array may be consistent with a "correlation distance" model encountered in stochastic scattering scattering theory (e.g., Aki, 1973; Toksoz et al, 1991). That is, within a distance of 0.5 km or less, *Pn* amplitude variations caused by scattering from random heterogeneities in the lithosphere beneath the array are small. As sensor separations increase greater than 0.5 km, amplitude variations due to scattering are greater.

Another possible explanation is that *Pn* amplitude variations are caused by focusing and defocusing effects from small-scale heterogeneities in the crust beneath the array, such as those commonly observed for teleseismic *P* waves. However, *Lg* wave amplitudes have

been found to be more stable and have less variability in amplitude around arrays than *P* wave amplitudes. This may be because *Pn* and *Pg* waves have smaller wavelengths than *Lg* waves, which are mainly composed of multiple shear-wave modes, and thus *Pn* and *Pg* phases may "feel" the effects of small scale heterogeneities in the crust more than do *Lg* waves. The stability of high-frequency *Lg* amplitudes have been known for some time and has been the basis for using the *Lg* phase to estimate more stable magnitudes for yield estimation (Hansen et al, 1990).

3.0 DISCRIMINANT SEPARATION MEASURES, OPTIMUM TRANSFORMATIONS, AND TRANSPORTABILITY

3.1 Transportability Issues

Assuming the existence of geophysically well-defined regions, the term *transportability* may mean the following different things:

1) Applicability of the same discriminant with statistics adapted to the specific region using available learning data sets and applying a consistent methodology to classify events (e.g., Fisk et al, 1996).

2) Applicability of a discriminant to various regions without any modifications, i.e., using statistics from other similar regions.

3) Application of the same discriminant with some adaptation of parameters but using assumed statistics for certain types of events (such as nuclear explosions) if learning sets for these are not available in the region.

In this section, we prefer definition (1) while reconciling ourselves with the occasional necessity of (3). Option (2) is not likely to work because of the variability in the definition and physical nature of regional "phases" from region to region. We shall also adapt the usual definitions of regional phases, as they are commonly picked in the seismic bulletins from various regions, even though their excitation and propagation may be different from region to region. Given this regional variability, it seems certain the efficiency of discrimination will vary from region to region and can be described in quantitative terms if sufficiently large learning data sets are available. We have not achieved this as yet, but we are attempting to set up a framework for worldwide transportability studies.

At regional distances, the most common discriminants used are distance-corrected spectral

ratios of various wave groups. These constitute multi-dimensional discriminants because spectral ratio measurements are performed over numerous frequency bands and for multiple combinations of regional arrivals. Inspecting a typical plot of such data for a group of explosions and earthquakes, it is obvious that the distributions of the spectral ratios tend to be similar in various neighboring frequency bands. This provides redundant information because of the strong correlation among the measurements in various bands and phase combinations. Typically, the separation between explosion and earthquake populations is poor at low frequencies (1-2 Hz) and improves with increasing frequency. Classification of a new event is commonly based on the positions of its many data points in such a graph relative to known earthquake and explosion data plotted in the same graph (e.g., Ryall et al, 1995).

Obviously, we do not need all the measurements in all spectral bands because some measurements may be more important than the others. Likewise, we may not need all the spectral ratios either. The question of how to combine such multiple measurements in the most effective way has not been studied extensively in regional seismology. Thus, the dimensionality of the data could be reduced considerably with appropriate data manipulation. Metrics of discrimination effectiveness are also needed to be able to compare various regions with regards to discrimination efficiency and transportability. What follows is a brief description of algorithms used in this study.

3.2 Handling of Missing Data - Generalized EM Algorithm

In our data sets of spectral ratios, ratios in some bands may be missing. This may result from censoring of data because of low signal-to-noise ratios in either the numerator or denominator phases or other data problems. In order to utilize the available data fully, it is desirable to substitute estimated values into the slots of missing data and thus make the rest of the good data available to improve the correlation matrices used in discrimination.

We use the technique of expectation maximization (EM) to interpolate missing data using

the available data values. We apply the following iterative procedure to compute the extrapolated data matrix:

a) Compute the first 'current' correlation coefficient matrix from the events that have complete sets of values in all frequency bands.

b) Extrapolate the missing data points for all events that need them using the "current" correlation coefficient matrix, with the formula below.

c) Compute the "current" correlation coefficient matrix from the extrapolated data set. If the preset iteration number is reached go to (d); otherwise remove the extrapolated values and go to (b).

d) Compute the correlation (not correlation coefficient) matrix for the reconstructed data set. These will be used in the separability analyses. Output the completed data matrix.

The extrapolated missing values a_i are computed as

$$a_i = \frac{\sum_k |C_{ik}|(a_k + \mu_i - \mu_k)}{\sum_k |C_{ik}|},$$

where the C_{ik} are the 'current' correlation coefficients, the summation over k occurs over the existing data values, and i is the index of the missing value. We are essentially correcting for the differences in means (μ) between the existing and missing components and weight these according to the absolute values of the average correlations between them. As the iteration progresses, the values of the means, correlations, and the extrapolated values will be optimized in some sense.

This completes the EM process. Typically a few iterations are quite sufficient. The

approach described above is not the rigorous EM algorithm described in Duda et al (2001) but the one termed *generalized EM* algorithm by the same source which includes a number of *ad hoc* methods for data extrapolation. The exact algorithm involves the iterative estimation of the population parameters from the existing data points without extrapolation of the missing values. It is cumbersome and slow computationally because numerous probabilities, usually assumed to be Gaussian, need to be computed. We feel that such complexity is not justified in this application. We go through this procedure in order to exploit the correlation information contained in the incomplete events and set up extrapolated data matrices that can be used in the handling of dimension reduction and separability issues. Events which had more than one-third of the number of the frequency bands used were discarded in each case described below.

3.3 Separability Measures and Optimum Transformations

Since spectral ratios of multiple regional phases in multiple frequency bands constitute discriminants with high dimensionality, dimensionality reduction is a desirable goal. Moreover, discrimination analysis will define the most efficient way to combine data and eliminate the measurements that contribute little to the solution.

Given the distributions of empirical data containing known event types, one can utilize several measures to examine how well a given category separates from others. The measures listed are based on the comparisons of the within-class scatter matrices $\boldsymbol{S}_W$ to the between-class scatter matrix $\boldsymbol{S}_B$. Different definitions of scatter matrices are given by Fukunaga (1990), Duda et al (2001), and Tou and Gonzalez (1974). The definition given to $\boldsymbol{S}_W$ by Duda et al (2001) for the two-class case is

$$\mathbf{S}_W = \mathbf{S}_1 + \mathbf{S}_2$$

where

$$\mathbf{S}_i = \sum_{\mathbf{x} \in D_i} (\mathbf{x} - \mathbf{m}_i)(\mathbf{x} - \mathbf{m}_i)^t,$$

where $\mathbf{m}_i$ are mean vectors of population i and the sum is over elements of class i. The between-class scatter matrices $\boldsymbol{S}_B$ can be defined as

$$\mathbf{S}_B = (\mathbf{m}_1 - \mathbf{m}_2)(\mathbf{m}_1 - \mathbf{m}_2)^t$$

The solutions to the eigenvalue problem

$$\mathbf{S}_W^{-1}\, \mathbf{S}_B\, \mathbf{w} = \lambda \mathbf{w}$$

from the equation above can be used to determine an optimum transformation for reducing the dimensionality of the problem (Duda et al, 2001). The eigenvector associated with the largest eigenvalue, λ, gives indication of the relative importance of the various orthogonal eigenvector components of the problem. The simplest transformation is the linear combination comprising the Fisher linear discriminant, expressed as,

$$\mathbf{w} = \mathbf{S}_W^{-1} (\mathbf{m}_1 - \mathbf{m}_2).$$

Thus the optimum direction for best linear separation of the two populations is a projection of the vector connecting the means transformed by the matrix $\mathbf{S}_W^{-1}$. If this matrix is close to diagonal, the results should be similar to intuitive expectation based on the scatter diagrams. If not, the results may be hard to explain intuitively. Since we are striving for a simple optimum linear transformation, we shall use this transformation in our study. We have found that a simple linear discriminant derived from the transformation above works well for multi-band regional phase spectral ratios for both a single pair of phases or combinations of several of these. Higher eigenvalues and eigenvectors are often noisy and may not furnish meaningful discrimination capability. We did not consider more complex discriminants because of the paucity of the data and the danger of overfitting. Fukunaga (1990) recommends that several eigenvectors

associated with the largest m eigenvalues be used for discrimination in a lower-dimensional subspace. He used scatter matrices different from those given above.

A number of separability measures are listed by Fukunaga (1990). A measure often used is the Bhattacharayya distance, expressed as,

$$B=1/8(\mathbf{m}_1-\mathbf{m}_2)^t\left[\frac{\Sigma_1+\Sigma_2}{2}\right]^{-1}(\mathbf{m}_1-\mathbf{m}_2)+\ln\frac{\left|\frac{\Sigma_1+\Sigma_2}{2}\right|}{\sqrt{|\Sigma_1||\Sigma_2|}},$$

where Σ_1 and Σ_2 are the multivariate covariance matrices of the two populations and $\mathbf{m}_1$ and $\mathbf{m}_2$ are the means. Since the Bhattacharayya distance applies strictly to Gaussian populations, it may not be informative if the distributions are much different from the normal distributions (Duda et al, 2001). The larger this measure, the better is the separability. Such measures are being used in this study to evaluate relative effectiveness of dicriminants in various regions.

Using a discriminant involving a single dimension has the advantage that it is possible to characterize the discrimination efficiency using another type of performance measure, the receiver operating characteristics (ROC). ROC are curves of detection probabilities plotted against the false alarm probabilities. ROC is an efficient way to compare the performance of a discriminant in various regions (i.e., transportability). In the following we shall present such curves for three example data sets.

3.4 Discrimination Results

We applied the above algorithms to two data sets. In both cases, the best-known distance corrections were used (Jenkins et al, 1998). The only exception is China where we have modified the *Pg* distance correction to correct for visible mismatches with our data.

3.4.1 China Data Set

The first data set is from China, containing both nuclear explosions and earthquakes. The data set consists of earthquakes around the WMQ region, Kazakh nuclear explosions, and one nuclear explosion at the Chinese test site, Lop Nor (see Figure 21). The scatter plot of *Pn/Lg* ratios in four frequency bands is shown in Figure 22 and some examples of seismograms are shown in Figure 21.

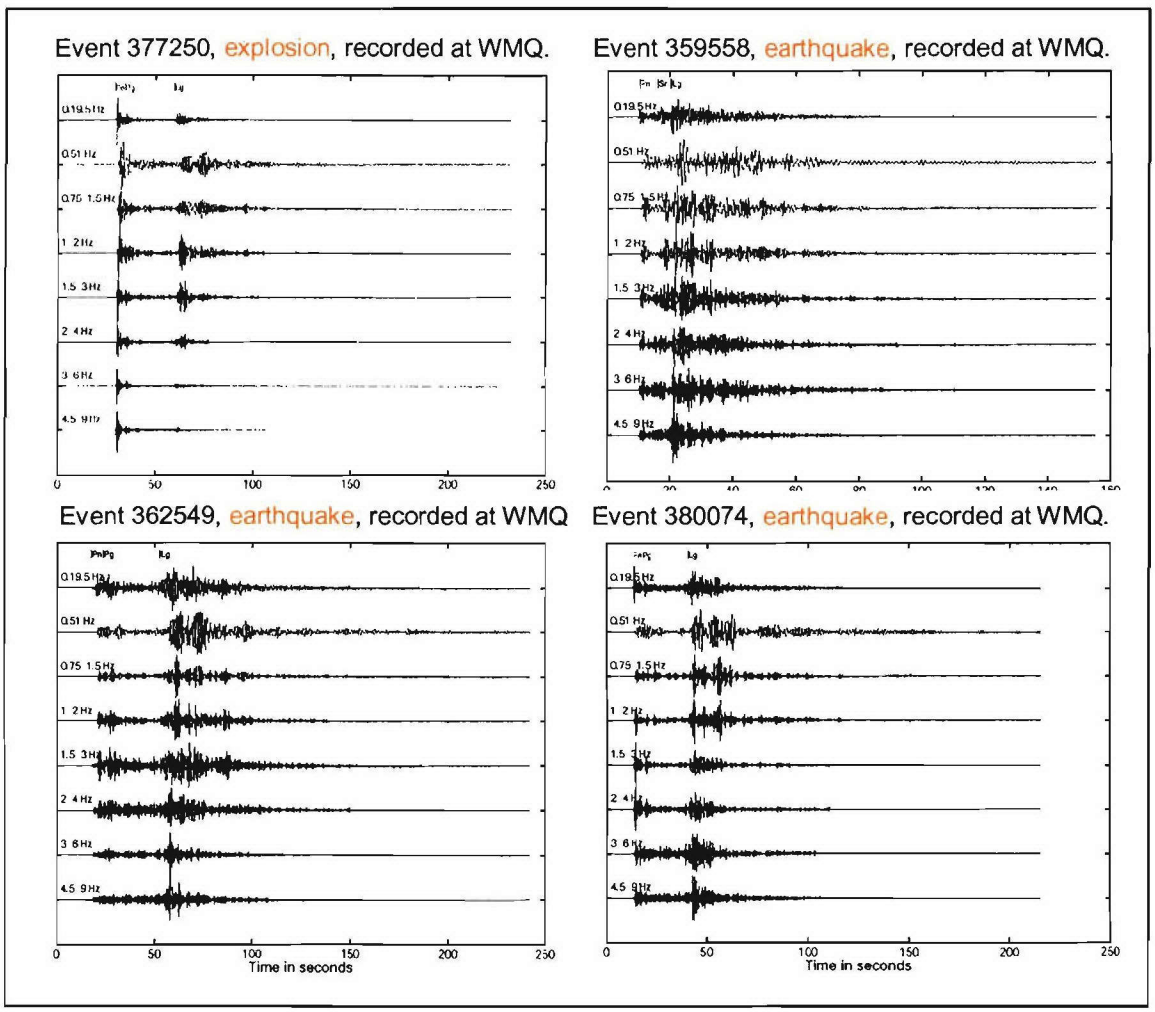


Figure 21: Examples of explosion and earthquake waveforms for events in China. Typically explosions have lower amplitudes in the shear wave related phases *Sn* and *Lg*.

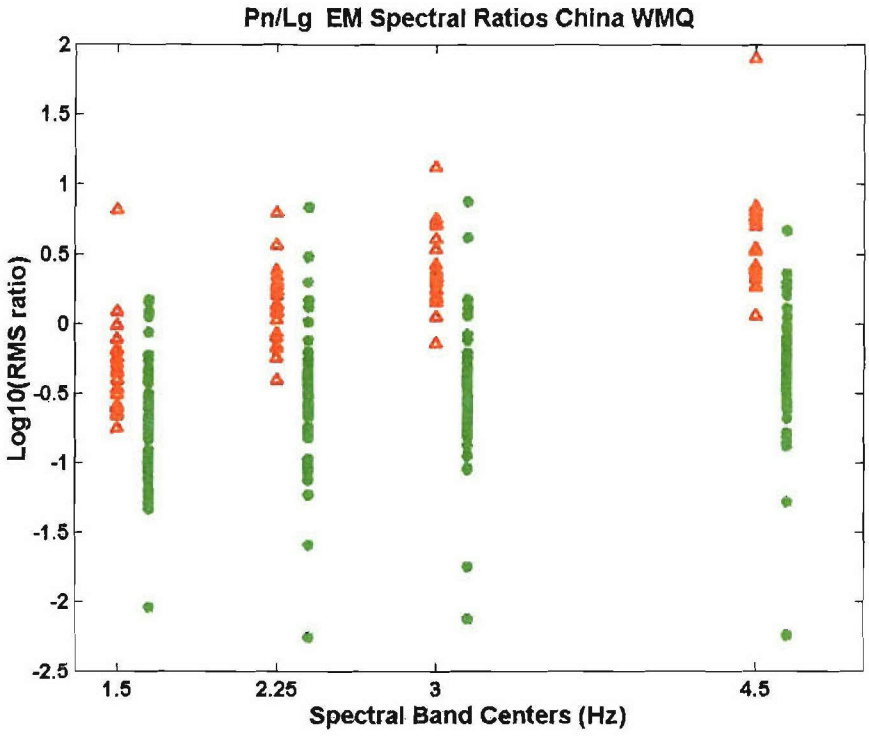


Figure 22: Scatter plot of the *Pn/Lg* spectral ratio data or four frequency bands from China. Red triangles denote explosions and green dots indicate earthquakes.

The data show poor separation generally, but the separation between explosion and earthquake populations improves with increasing filter band center frequency. After computing the various scatter matrices and subjecting the results to the eigenvalue-eigenvector analysis, the eigenvalues (Figure 23) show that we have at most two components, associated with the largest eigenvalues that can be used for effective discrimination.

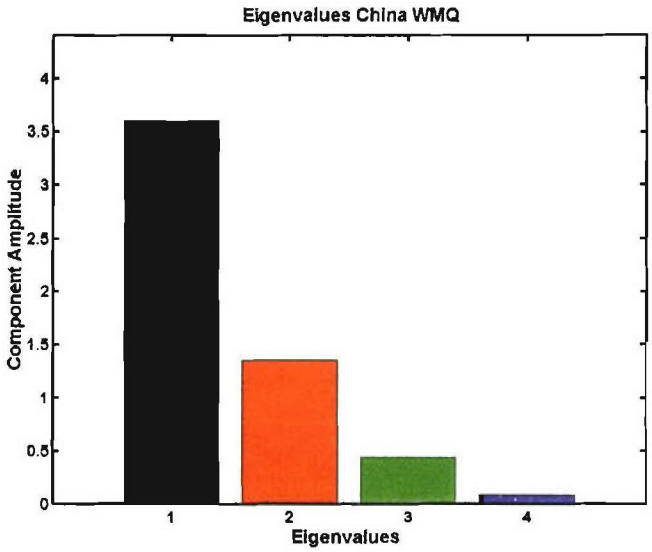


Figure 23: Eigenvalue spread derived from the analysis of *Pn/Lg* amplitude ratios for the Chinese station WMQ.

Plots of the eigenvectors in Figure 24 show how the *Pn/Lg* spectral ratio values in various bands can be combined to provide the discriminants in decreasing order of importance.

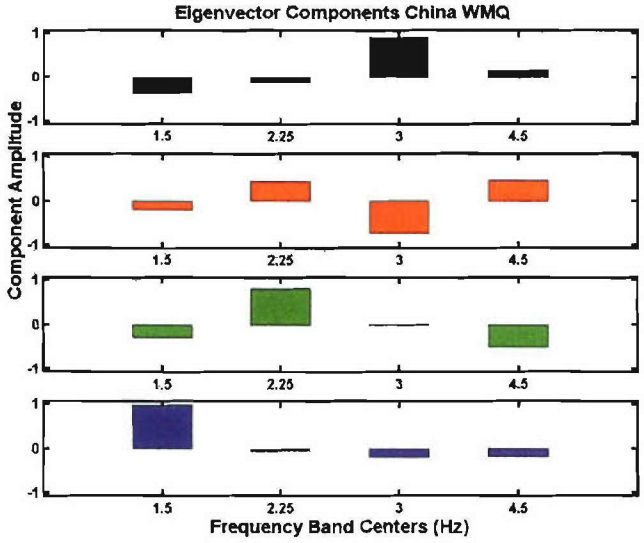


Figure 24: Plots of eigenvectors from the analysis of the China data. The first eigenvector has a dominant component for the 3 Hz band where the best separation exists. The last gives the largest weight to the first frequency band, the least effective discrimination.

Figure 25 shows plots of the ratios transformed with the eigenvectors shown in Figure 24. That is, linear combinations of original *Pn/Lg* ratios were computed using the weights shown in Figure 24 and these combinations are plotted versus eigenvector in Figure 25. The first eigenvector clearly produces the greatest separation. Explaining the details of the rest would be hard. Nevertheless, the first two combinations will give the best separation. There is a definite shift between the means of the projection on the second eigenvector that could also be used to complement the results from the first. The rest of the transformed ratios in Figure 25 clearly overlap completely.

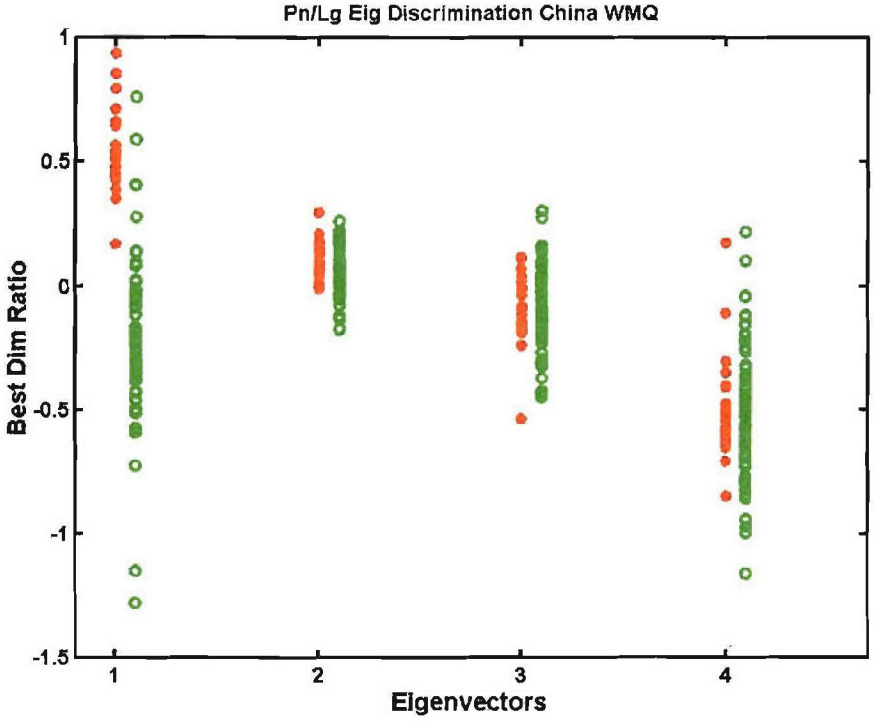


Figure 25: Discrimination plots for the China data set. Obviously, the best separation between the explosions (red dots) and the earthquakes (green dots) is achieved by the eigenvector #1 associated with the largest eigenvalue. The second eigenvector gives a small shift in means but large overlap, the remaining two overlap completely.

The Receiver Operating Curve (ROC) for the discrimination results is shown in Figure 26. This curve gives the relationship between the false alarm rate (probability) and the

detection rate for the Fisher discriminant involving the first eigenvector. The ROC curve shows that in order to have a detection rate as low as 90% one would have a 15% false alarm rate, which is quite high. The probable reason is that the data set covers many paths over Asia with quite variable propagation characteristics thus making reliable discrimination difficult. This indicates that more reliable discrimination would require much finer regionalization of regional discriminants.

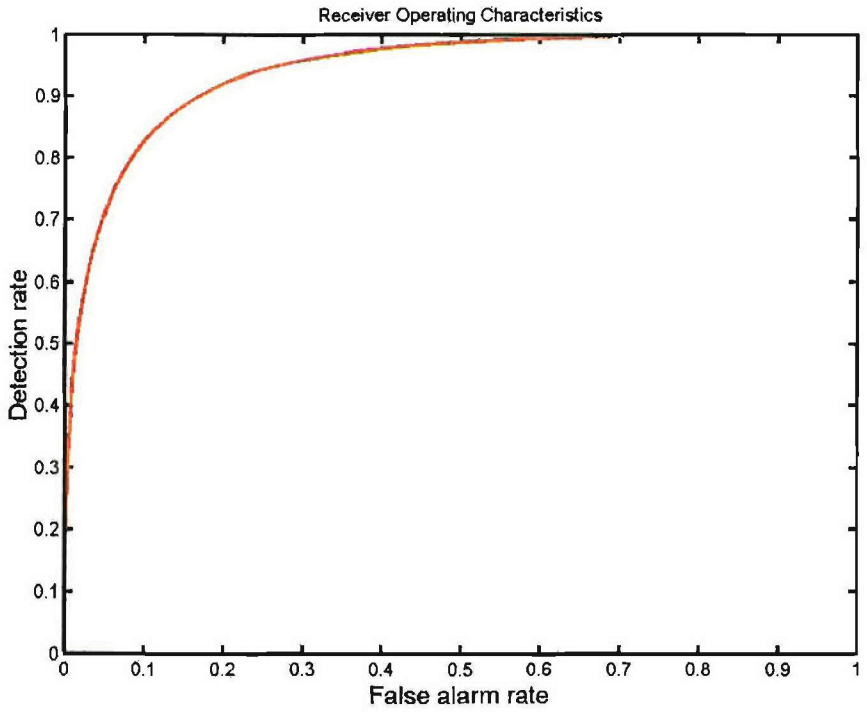


Figure 26: ROC curve for the Fisher discriminant acting on the China data set.

3.4.2 NTS Explosion vs. Skull Mountain Data Set

To compare this analysis with another region, we show the results of the analysis of *Pn/Lg* and *Pg/Lg* ratio data from the NTS and Skull Mountain earthquake dataset, originally studied by Walter et al (1995). This data set consists of Skull Mountain earthquakes and NTS nuclear explosions as recorded at the Lawrence Livermore National Laboratory (LLNL) stations Kanab, Utah (KNB), Elko and Mina, Nevada (MNA). In this analysis,

we combine *Pn/Lg* and *Pg/Lg* measurements into a 14-dimensional scheme. We originally combined data from all three stations, but realized later that the combination yielded worse discrimination than the analyses for each station individually. We have also judged the station Elko to be of lower quality. Therefore, we are presenting results only for KNB and MNV.

Figure 27 shows examples of NTS explosions recorded at LLNL stations. Waveforms for Skull Mountain earthquake waveforms are shown in Figure 28. These seismograms of the explosions and earthquakes display some of the prominent characteristics of regional seismograms in western North America; relatively small Pn, large Pg and Sn often missing. Obviously, Lg is much smaller for explosions.

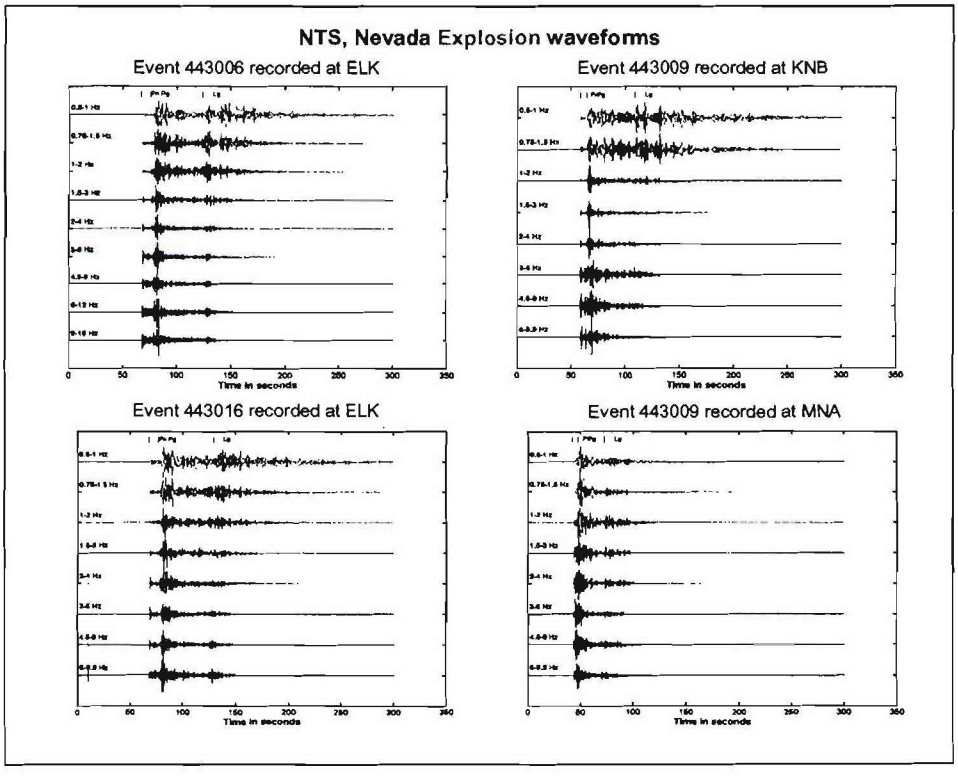


Figure 27: Examples of waveforms from NTS explosions recorded at LLNL seismic stations.

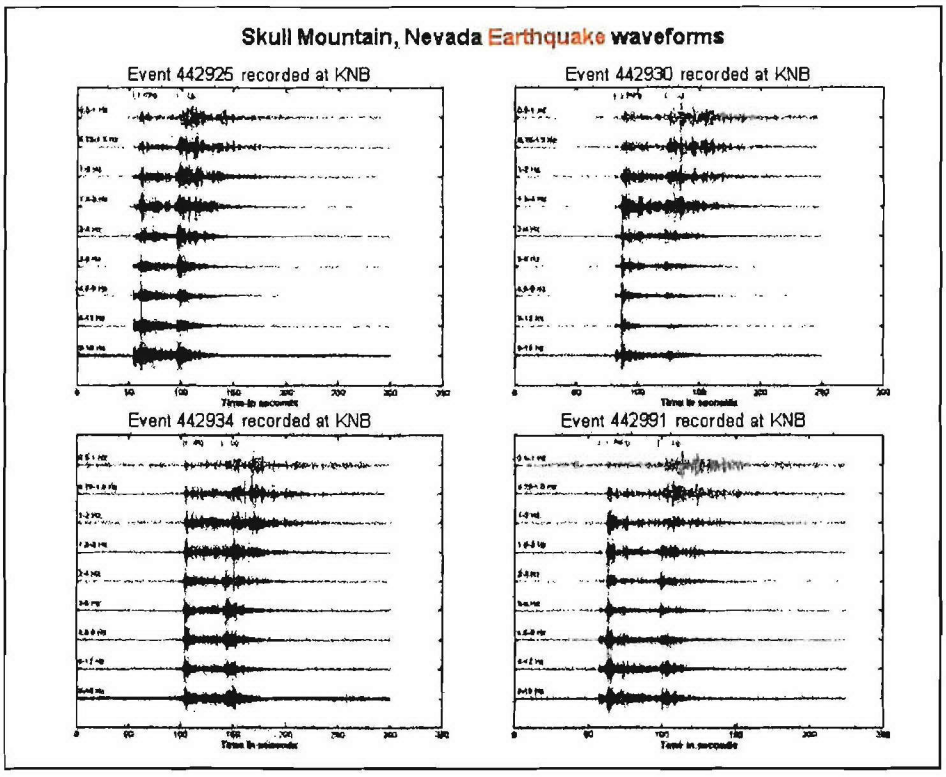
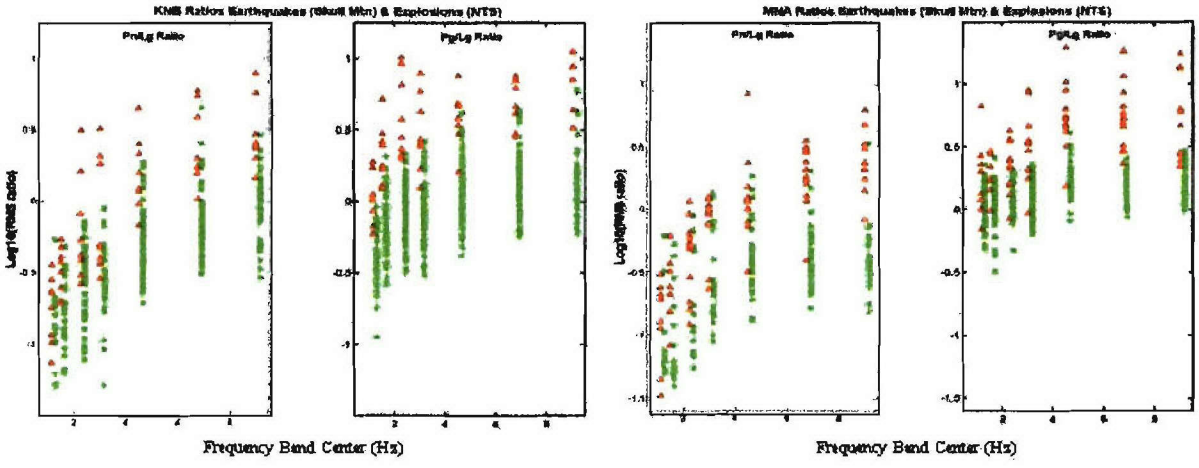


Figure 28: Examples of waveforms from Skull Mountain earthquakes recorded at LLNL stations.

The amplitude ratios, plotted as a function of frequency, are shown in Figure 29.



(a) (b)

Figure 29: Scatterplots of *Pn/Lg* and *Pg/Lg* ratios for Skull Mountain earthquakes and NTS nuclear explosions at the stations Kanab, Utah (KNB) (a) and Mina, Nevada (MNA) (b).

The evidently improved scatter in both ratios at both stations with frequency has been pointed out by Walter et al (1995). It has been a generally accepted tenet of discrimination that the earthquake/explosion separation increases with frequency (e.g., Goldstein, 1995).

The eigenanalysis was applied to the combination of *Pn/Lg* and *Pg/Lg* ratios. Figure 30 shows histogram plots of the eigenvalues and Figure 31 shows the corresponding eigenvector weights.

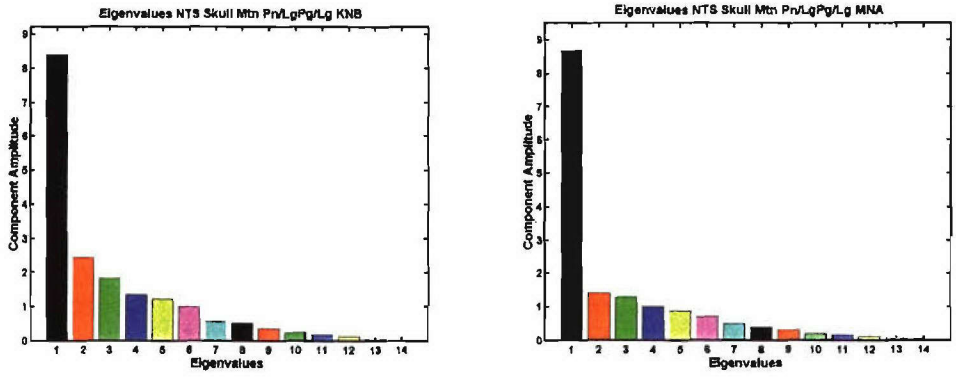
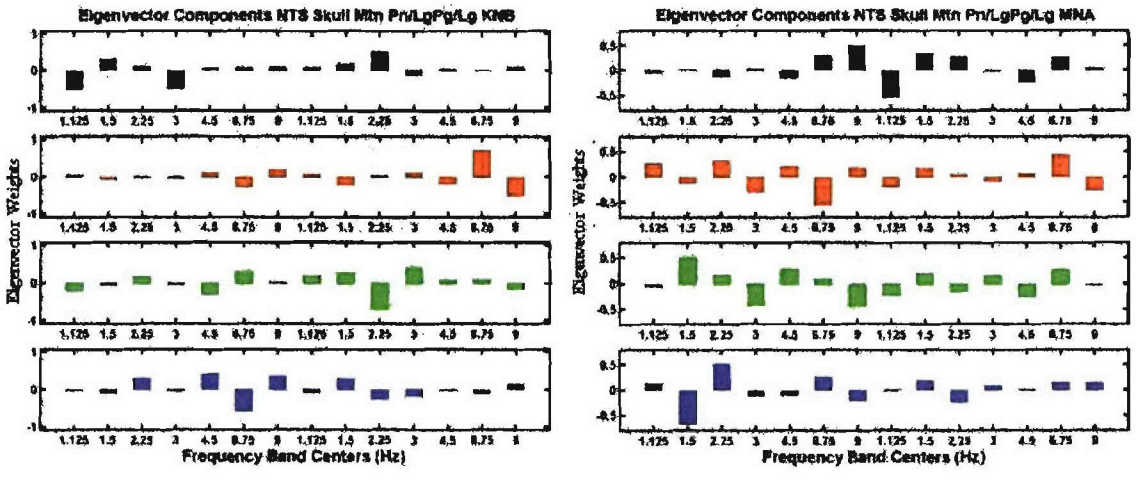


Figure 30: Eigenvalue spreads for the KNB and MNA data sets.

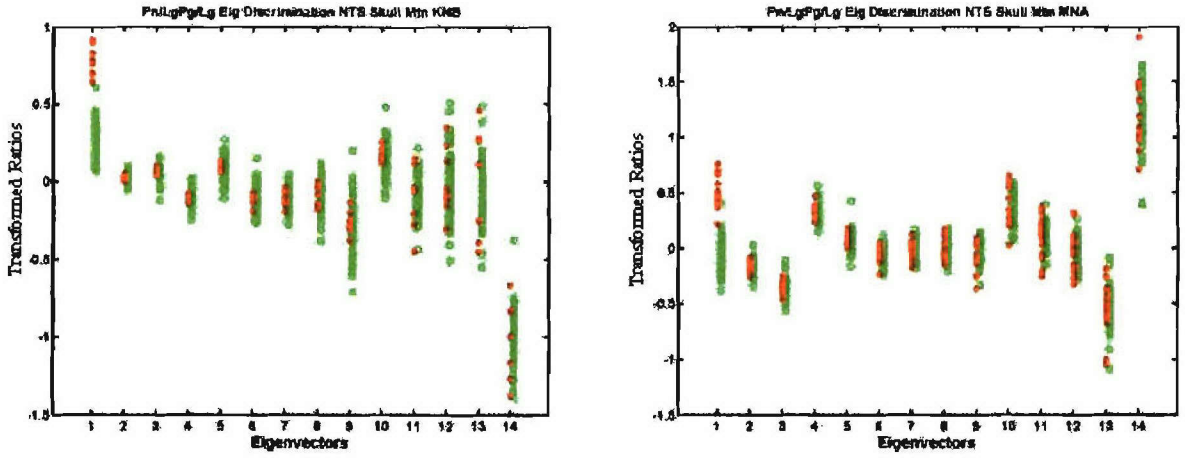


(a) (b)

Figure 31: Eigenvectors for KNB (a) and MNA (b) computed separately from the amplitude ratios of earthquakes and explosions in Figure 29 in the various frequency bands.

The eigenvalue spreads for the separate KNB and MNA data sets show that the first eigenvalues are much larger than the rest. The weights for the principal eigenvalue, which are the first rows in Figure 31, do not seem to be much larger for the high-frequency ratios. In fact, for KNB in Figure 31 (a), the weights seems strongest for lower frequency *Pn/Lg* amplitude ratios.

Finally, Figure 32 shows the resultant transformed discriminants.



(a) (b)

Figure 32: Resultant transformed amplitude ratios for the KNB and MNA data respectively after applying the weights in Figure 31 to the amplitude ratios in Figure 30.

The transformations clearly show the best separation on the first eigenvalue, although the separation is better for KNB than for MNA, while the other eigenvectors do not contribute much. Why the eigenanalysis seems to weight lower frequencies more than higher at KNB is not clear. Perhaps the actual means of the explosion and earthquake populations are separated more at low frequency than at high, even though there appears to be more overlap in the low-frequency bands. One could make the case for some effective separation at some of the other lower eigenvectors as well by noting that the transformed explosion populations have less scatter thus making it possible to identify earthquakes as tending to occupy the details of distributions with similar means.

The ROC curve for the Fisher discriminant involving the first eigenvector defined above is shown in Figure 33. The combined spectral ratio Fisher discriminant in this case performs better than in the case of the China data set described above. A detection rate of 90% now corresponds to about a 5% false alarm rate. This would still require the further examination of a fairly large number of events using criteria other than the Fisher discriminant defined above. Just as in the case of the Chinese data set, this area is also quite variable with respect to propagation characteristics.

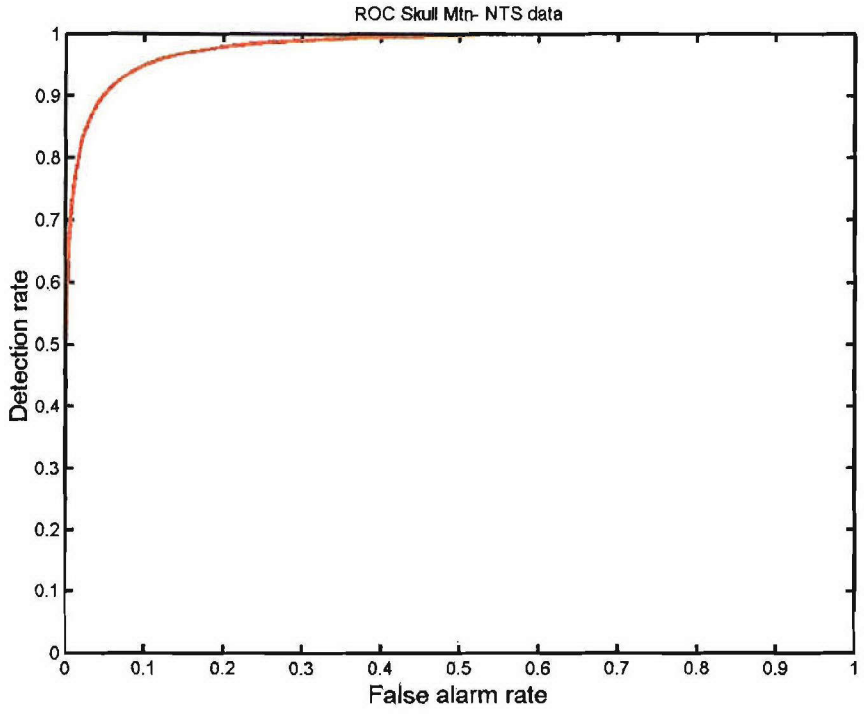


Figure 33: ROC curve for the Fisher discriminant as applied to the Skull Mountain-Nevada data set.

3.4.3 Steigen-Novaya Zemlya Data Set

We now show this same analysis applied to the data sets studied by Ryall, et al (1995) recorded at the ARCES array in northern Norway, shown earlier in Figure 1. This data set was assembled to further illustrate and test the methodology described above and serve as

an example of well separated explosion and earthquake populations. Such a data set would not be encountered in the practice of seismic discrimination since the two populations, the Steigen explosion data set and Novaya Zemlya nuclear explosions, could be easily discrimnated by means other than regional *P/S* ratios, such as the standard *Ms-mb* measures. Moreover, these differ greatly in magnitude and thus scaling laws would have to be considered. However, this example shows how these discriminants behave in Baltic Shield crustal region, which is quite distinct from the more tectonically active areas of China and western US, discussed above.

The scatter plot of *Pn/Sn* ratios for the Steigen earthqakes and Novaya Zemlya explosions are shown in Figure 35.

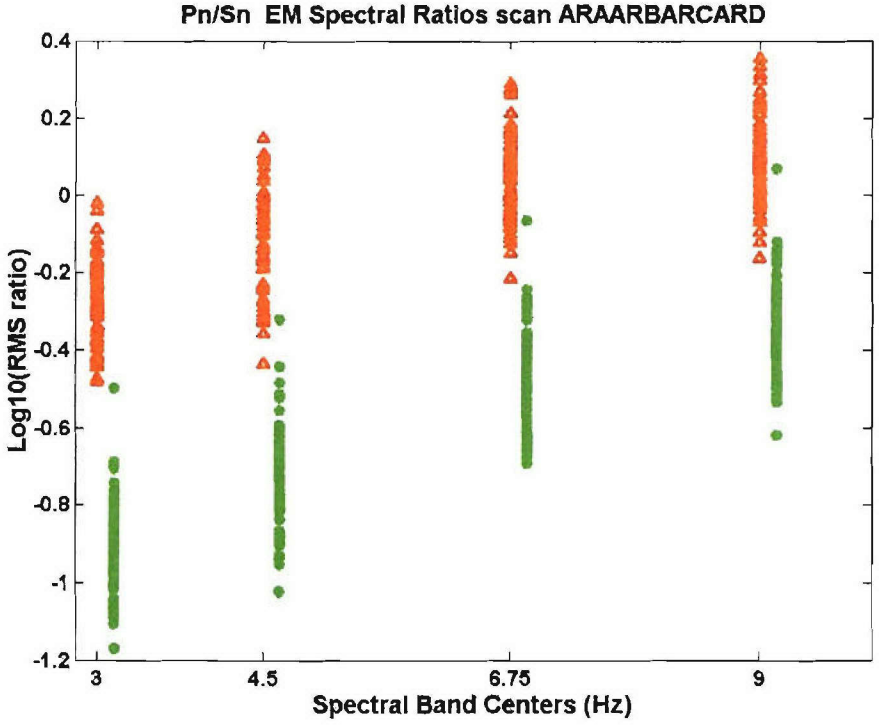


Figure 34: Scatter plots of the *Pn/Sn* spectral ratio populations of the Novaya Zemlya explosion (red) and the Steigen earthquakes.

(Note: *Pn/Lg* ratios were not studied because *Lg* is propagation from Novaya Zemlya explosions to ARCES is blocked in the Barents Sea (Baumgardt, 2001)). As pointed out by Ryall, et al (1995), these ratios show a very good separation between the two types of

events.

Consequently, the eigenvalues in Figure 35 show that a single eigenvector dominates indicating that the Fisher discriminant would be quite effective.

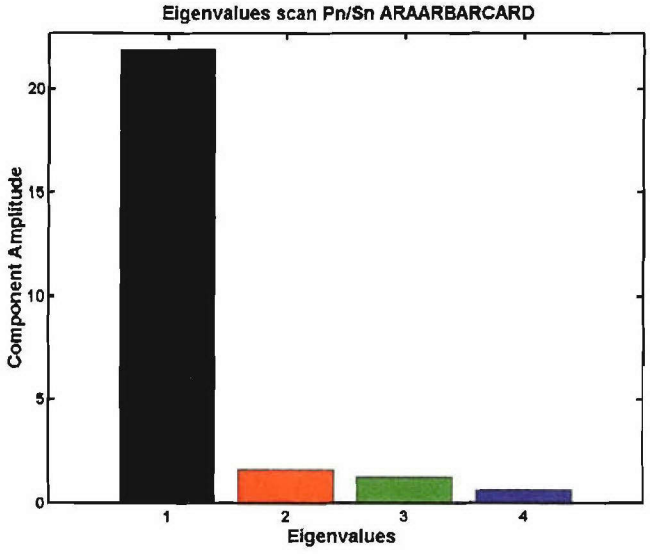


Figure 35: Eigenvalue spread for the four frequency bands of the Steigen-Novaya Zemlya data set.

The eignevalue weights corresonding to these eigenvalues are shown in Figure 36.

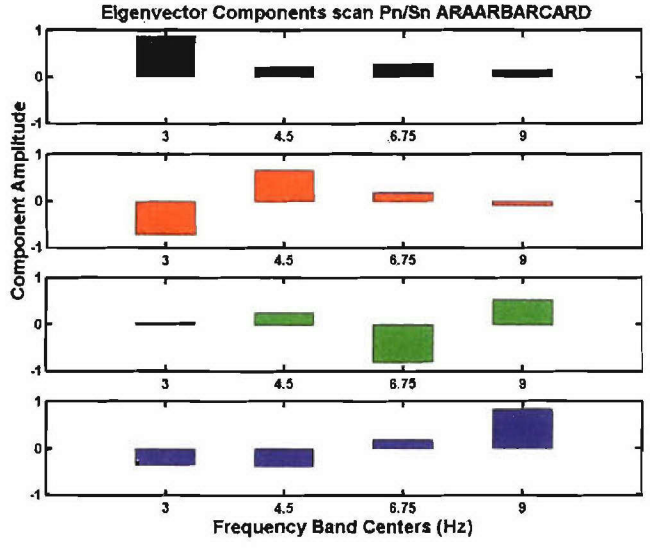


Figure 36: Four eigenvectors corresponding to the four eigenvalues (from top to bottom).

The eigenvector components of the Fisher discriminant (top) indicate that the optimum discrimination emphasizes the lowest frequency band. This is also obvious from the inspection of the scatter diagrams in Figure 37.

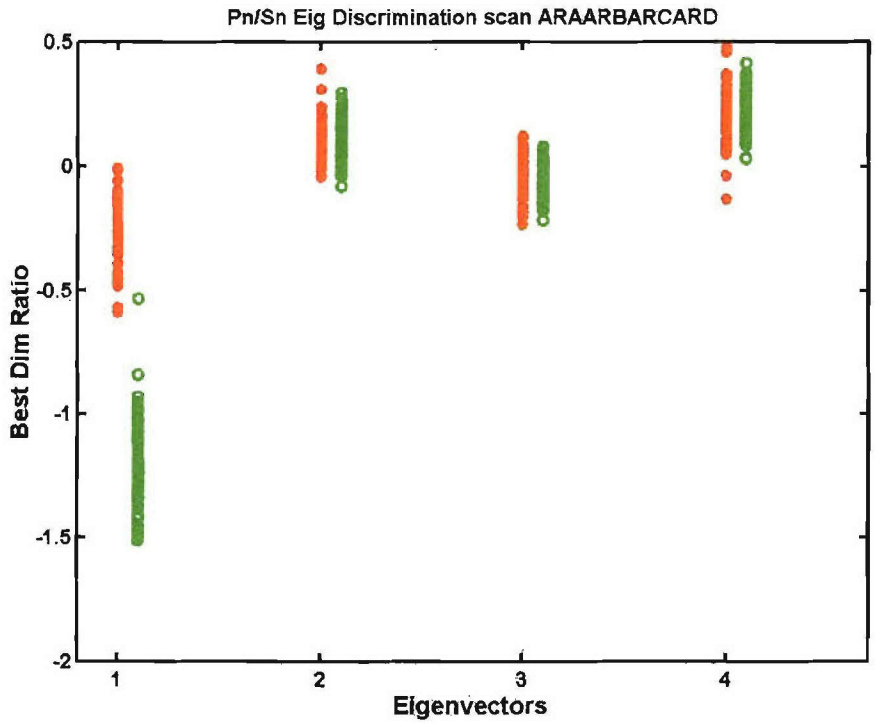


Figure 37: Discrimination using the coordinate transformations corresponding to the four eigenvectors of the Steige-Novaya Zemlya data set.

The least amount of overlap also corresponds to the coordinate transformation according to the first eigenvalue components used as weights.

4.0 CONCLUSIONS

This report has described a method for analyzing site variance in regional *P/S* ratios and a preliminary application to data recorded by Scandinavian regional arrays for a group of underwater explosions in the same location using ANOVA2. We have found that the amplitude ratios *Pn/Lg* and *Pg/Lg* varied by between a factor of 2 to 3 around the elements of both NORES and FINES regional arrays in all the frequency bands with signal-to-noise ratios in excess of 3. This variation, which occurs only over a spatial aperture of 3 km for both arrays, appears to be systematic with the largest variations between sensors with the largest separation, which is up to 3 km, and correlates more strongly with *Pn* and *Pg* amplitudes than with *Lg* amplitudes. *Lg* appears to be more stable than *Pn* and *Pg*, or less sensitive to site effects, except perhaps in the higher frequency bands, where stronger but more random variations in *Lg* site terms are observed.

Overall, our studies have shown that the stability of *Pn/Lg* and *Pg/Lg* amplitude ratio discriminants, in terms of site effects, is driven primarily by the stability of the numerator phase, *Pn* and *Pg*, at least for the arrays we have studied in the Scandinavian shield. The *Lg* phase amplitudes, on the other hand, appear to be relatively stable, exhibiting much smaller variation in amplitude across the arrays. Future research will address whether this same result also holds true for arrays located in other tectonic regions, and to what extent this scatter degrades the effectiveness of the *Pn/Lg* and *Pg/Lg* ratio discriminant.

We have begun a statistical study of the transportability of *P/S* ratio discriminants using separability measures and optimum transformations in order to reduce dimensionality of multiple frequency *P/S* ratios. These transformations consist of calculating the intra-class and inter-class scatter matrices for *P/S* ratio discriminants and using the eigenvectors, corresponding to the largest eigenvalues, of a matrix product constructed of such matrices to compute optimum transformation of discriminants that provide best separation. We have applied this analysis to distance-corrected discriminants in different regions (e.g.,

China, Eurasia, North America) in order to compare discriminant effectiveness for different regions and to evaluate the transportability of optimum discriminant decision surfaces. Analysis of variance showed that most of the variance is associated with the coordinate transformed variable due to the first eigenvector with smaller contributions from the rest of the eigenvectors. These contributions seem to be of dubious value and may represent only computational noise or source scaling unaccounted for by the analysis (Ringdal et al, 2000). Thus, the use of the Fisher discriminant appears to be sufficient in discrimination based on multiple spectral amplitude ratios. Moreover, since the problem can be reduced to a single dimension, performance measures such as the receiver operator characteristics (ROC) can be used for comparing the efficiency of such discriminants in various regions. In our case we have found that the best separation existed for the Steigen-Novaya Zemlya comparison, a data set which is not practically meaningful, however. The discrimination results for the NTS explosion - Skull Mountain earthquake pairing were significantly better than for the China data set. This is not surprising given the path variability in Asia and the fact that the former data set combined spectral two kinds of spectral ratios over larger range of frequencies.

5.0 REFERENCES

Aki, K. (1973). Scattering of *P* waves under the Montana Lasa, *Jour. Geophys. Res.*, **78**, 1334-1346.

Baumgardt, D.R. (1993). Regional characterization of mine blasts, earthquakes, mine tremors, and nuclear explosions using the Intelligent Event Identification System, *Final Report to ARPA, SAS-TR-94-12*, ENSCO, Inc., Springfield, VA.

Baumgardt, D.R., and Z. Der (1995). Investigation of the transportability of the *P/S* ratio discriminant to different tectonic regions, *Scientific Report No. 1, PL-TR-94-2299*, 6 December 1994, ENSCO, Inc.

Baumgardt, D.R. (1996). Investigation of *Lg* blockage and the transportability of regional discriminants in the Middle East, *Scientific Report No. 1, Pl-TR-96-2294*, ENSCO, Inc., Springfield, VA.

Baumgardt, D.R. (1999). Characterization of underwater explosions by spectral/cepstral analysis, modeling, and inversion, Annual Report submitted to DTRA, November, 1999, ENSCO, Inc., Springfield, VA.

Baumgardt, D.R. and Z. Der (1998). Identification of presumed shallow underwater chemical explosions using land-based regional arrays, *Bull. Seism. Soc. Am.*, **88**, 581-595.

Baumgardt, D.R., Z. Der, and A. Freeman (2001). Investigation of the partitioning of source and receiver-site factors on the variance of regional *P/S* amplitude ratio discriminants, *Proceedings of the 23rd Seismic Research Review: Worldwide Monitoring of Nuclear Explosions – October 2-5, 2001*, Jackson Hole, WY.

Duda, R.O., P.E. Hart, and D.G. Stork (2001). *Pattern Classification*, Wiley-Interscience.

Fan, G-W., T. Lay, and S. Bottone (2001). Path corrections for source discriminants: a case study of two International Seismic Monitoring stations, submitted to *Pure and Applied Geophys.*

Fisk, M., H.L. Gray, and G.D. McCartor (1996). Regional event discrimination without transporting thresholds, *Bull. Seism. Soc. Am.*, **86**, 1545-1558.

Fukunaga, K. (1990). *Introduction to Statistical Pattern Recognition*, Academic Press.

Goldstein, P. (1995). Slopes of *P*- to *S*-wave spectral ratios-a broadband regional seismic discriminant and a physical model, *Geophys. Res. Lett.* **22**, 3147-3150.

Hansen, R., F. Ringdal, and P.G. Richards (1990). The stability of RMS *Lg* measurements and their potential for accurate estimgateion of the yields of Soviet underground nuclear explosions, *Bull Seism. Soc. Am.*, **80**, 2106-2126.

Hartse, H.E., S.R. Taylor, W.S. Phillips, and G.E. Randall (1997). Regional event discrimination in central Asia with emphasis on western China, *Bull. Seism. Soc. Am.*, **87**, 551-568.

Husebye, E.S., Y.V. Fedorenko,and Y.S. Karadzhov (2002). Parametrized seismic source discrimination in aseismic regions, *Seism. Res. Lett.*, **73**, 356-368.

Isaaks, B.L., and C. Stephens (1975). Conversion of *Sn* to *Lg* at a continental margin, *Bull. Seismol. Soc. Am.*, **65**, 224-235.

Jenkins, R.D., T.J. Sereno, and D.A. Brumbaugh, (1998). Regional attenuation at PIDC stations and the transportability of the *S/P* discriminant. Final Report, *SAIC-98/3012/AFRL-VS-HA-TR-98-0046*, Science Application International, Inc., La Jolla, CA.

Johnson, A, and D.W. Wichern (1988). *Applied Multivariate Statistical Analysis*, Prentice Hall, Englewood Cliffs, NJ.

Nuttli, O.W. (1973). Seismic wave attenuation and magnitude relations for eastern North America, *J. Geophys. Res.*, **78**, 876-885.

Patton, H. (2001). Regional magnitude scaling, transportability, and Ms:mb discrimination at small magnitudes, *Pure Appl. Geophys*, **158**.

Ringdal, F., E. Kremenetskaya and V. Asming (2000). Observed characteristics of regional seismic phases and implications for *P/S* discrimination in the European Arctic, submitted to *PAGEOPH Special Issue on CTBT Monitoring*.

Ryall, Jr., A.S, D.R. Baumgardt, M.D. Fisk, and F. Riviere-Barbier. (1995). Resolving regional discrimination problems: some case histories, in *Monitoring a Comprehensive Test Ban Treaty*, E.S. Husebye, and A. M. Dainty, eds., 721-742.

Toksoz, M.N., A.M. Dainty, and E.E. Charrette (1991). Coherency of ground motions at regional distances and scattering, *Phys. Earth. Planet. Int.*, **67**, 162-179.

Tou, J.T., and R.C. Gonzalez, (1974). *Pattern Recognition Principles*, Addison-Wesley.

Xie, J., and H. Patton (1999). Regional phase excitation and propagation in the Lop Nor region of Central Asia and implications for the physical basis of *P/Lg* discriminants, *J. Geophys. Res.*, **104**, 941-954.

Zhang, T., S.Y. Schwartz, and T. Lay (1994). Multivariate analysis of waveguide effects on short-period regional wave propagation in Eurasia and its application in seismic discrimination, *J. Geophys. Res.*, **99**, 21929-21945.

Walter, W.R., K. Mayeda, and H.J. Patton (1995). Phase and spectra ratio discrimination between NTS earthquakes and explosions Part 1: Empirical observations, *Bull. Seism. Soc. Am.*, **85**, 1050-1067.

DEPARTMENT OF DEFENSE

DEFENSE TECHNICAL
INFORMATION CENTER
8725 JOHN J. KINGMAN ROAD,
SUITE 0944
FT. BELVOIR, VA 22060-6201
2 CYS ATTN: DTIC/OCA

DEFENSE THREAT REDUCTION
AGENCY
8725 JOHN J. KINGMAN ROAD,
STOP 6201
FT. BELVOIR, VA 22060-6201
2 CYS ATTN: NTD/D. BARBER

DEPARTMENT OF DEFENSE CONTRACTORS

ITT INDUSTRIES
ITT SYSTEMS CORPORATION
1680 TEXAS STREET, SE
KIRTLAND AFB, NM 87117-5669
2 CYS ATTN: DTRIAC
ATTN: DARE

ENSCO, INC.
5400 PORT ROYAL RD.
SPRINGFIELD, VA 22151
ATTN: D. BAUMGARDT