NRDC Information Product Technical Reference Manual

Version 1.0
August 1, 1992

Prepared by

IRIS's *Joint Seismic Program Center*

Department of Physics
University of Colorado at Boulder
Campus Box 583
Boulder, CO 80309-0583

Distributed by
IRIS DMC

IRIS
DATA
MANAGEMENT
SYSTEM

Incorporated Research Institutions for Seismology
Data Management Center
1408 NE 45th Street
2nd Floor
Seattle, Washington 98105

Table of Contents

1. Introduction

On May 28, 1986 an historic agreement was signed in Moscow by Adrian Dewind, the Chairman of the Natural Resources Defense Council (NRDC), and Academician E. P. Velikhov, Vice-President of the Academy of Sciences of the USSR. This agreement established the first joint seismic observations between the US and USSR aimed at monitoring underground nuclear testing. As part of this agreement, three sites were occupied by American scientists within 300 km of the major Soviet nuclear test sites near Semipalatinsk, Kazakhstan. Data were recorded at these sites during 1986 and 1987 culminating in three large chemical explosions, which were detonated as part of the NRDC-Soviet Academy program, in early September, 1987.

The NRDC program represents the logical predecessor to the Joint Seismic Program and the data collected in Kazakhstan provide us with one of the earliest detailed looks at the seismic characteristics of Central Asia. We at the Joint Seismic Program Center (JSPC) are proud to offer the data collected during Phase-II of the NRDC program as the first JSPC information product. This product combines data, documentation and software into an integrated kit that will make these data immediately useful in seismic research efforts.

The history of the NRDC program in the Soviet Union can be broken down into two parts; the Phase-I deployment and the Phase-II deployment. The Phase-I deployment took place during 1986, starting in July, and mainly involved surface instrumentation for doing noise surveys and for capturing initial seismic events. The Phase-I data were used to establish the detailed site locations for the Phase-II deployment which took place in 1987 and involved digital 3-component surface and borehole instruments.

Those who wish to immediately install the information product and get a look at the data should proceed directly to the *NRDC information product installation/tutorial manual* where the installation is described and a brief tutorial is given. All other details concerning the data, how it is stored and the software for accessing it can be found in this report.

This report summarizes both data and information product characteristics. The people who worked to make the NRDC program a success are listed in the acknowledgements, in section 2. Section 3 gives site and instrumentation characteristics including a map of the sites and figures showing standard instrument responses. A description of the processing done on the data are described in section 4. Associated events are listed in section 5. The data is reviewed in section 6 which also includes a coverage table and describes the quality control measures that were taken during the assimilation of the information product. Section 7 gives a general overview of the NRDC information product. A description of the relational database used for representing the data is given in section 8 and the software included in the product is described in section 9. A comprehensive list of references is listed in section 10.

2. Acknowledgements

The people who worked to bring about the negotiations with the Soviet Academy of Sciences and who made the first Kazakhstan field trips to install seismic instruments pioneered the joint efforts in nuclear verification research that exist today between the U.S. academic seismological community and seismologists from the former Soviet Union. Following is a list of names and institutions that identify the major contributors to the NRDC program and to the data contained in this information product.

name	**institution**
Charles Archambeau	University of Colorado
Jonathon Berger	University of California at San Diego
Paul Bodin	University of California at San Diego
Jim Brune	University of Nevada at Reno
Thomas Cochran	NRDC
Adrian Dewind	NRDC
Jack Evernden	US Geological Survey

name	**institution**
Holly Given	University of California at San Diego
Joan Gomberg	University of California at San Diego
I. L. Nersesov	USSR Institute of the Physics of the Earth
Christopher Paine	US Senate Staff for Ted Kennedy
Keith Priestley	University of Nevada at Reno
M. A. Sardovskiy	USSR Institute of the Physics of the Earth
Aaron Tovish	Parliamentarians for Global Action
E. P. Velikhov	USSR Academy of Sciences
Frank Vernon	University of California at San Diego
Frank von Hipple	Princeton University
William Walter	University of Nevada at Reno

3. Site and Instrumentation Characteristics

During the NRDC Phase-II deployment, three stations were operated at Bayanaul (BAY), Karkaralinsk (KKL) and Karasu (KSU) by a group from UCSD/IGPP. Both surface and borehole instruments were used at these three stations and the nominal instrumentation configuration is described in Berger, et. al. (1988). The data acquired during this deployment starts on 19 March 1987 (day 078), ends on 5 September 1987 (day 248) and consists of 77 days during which one or more stations were operating and reporting events. In addition to the data collected at BAY, KKL and KSU, portable surface instrumentation was deployed at five stations (UNR1, UNR2, UNR3, UNR4 and UNR5) which were operated on 2 September and 3 September by personnel from UNR during three large chemical explosions, which were detonated as part of the NRDC-Soviet Academy program. The station layout is shown in figure 1.

The three UCSD sites were all located on granitic bedrock and nominal 100 m boreholes were drilled at each site. The instrumentation at each site consisted of one borehole 3-component seismometer and two surface 3-component seismometers. The borehole seismometers were Teledyne Geotech 54100 models with nominal seismometer periods of 5 seconds. The surface seismometers were Teledyne Geotech GS-13 models with nominal periods of 1 second and Kinemetrics S-1 models with nominal periods of 5 seconds. The data acquisition units consisted of 15 channels of 16-bit digital data all subjected to six-pole Butterworth anti-aliasing filters with corner frequencies of 80 Hz and sampled at 250 sps. In addition optional low cut filters at 0.1 Hz were used. Since there were 9 physical channels coming from the seismometers at each site, the acquisition units were able to record two of the 3-component instruments at two gain levels and the third 3-component instrument at one gain level. The typical configuration was to record the borehole intermediate period (54100) instrument and the surface short period (GS-13) instrument at high and low gain levels and the surface intermediate period (S-1) instrument at a low gain level. The detailed channel configurations are given in table 1. Although all of the data from BAY, KKL and KSU were recorded at 250 sps, we refer to the GS-13 instruments as being short period and to the 54100 and S-1 instruments as being intermediate period.

Table 1. UCSD channel configurations

Station	**Dates**	**Configuration**
BAY	133-248	high/low gain surface short period, high/low gain borehole intermediate period, low gain surface intermediate period
KKL	078-123	high/low gain surface short period, high/low gain surface intermediate period, high gain borehole intermediate period
	132-248	high/low gain surface short period, high/low gain borehole intermediate period, low gain surface intermediate period
KSU	099-124	high/low gain surface short period, high/low gain surface intermediate period, high gain borehole intermediate period
	230-248	high/low gain surface short period, high/low gain borehole intermediate period, low gain surface intermediate period

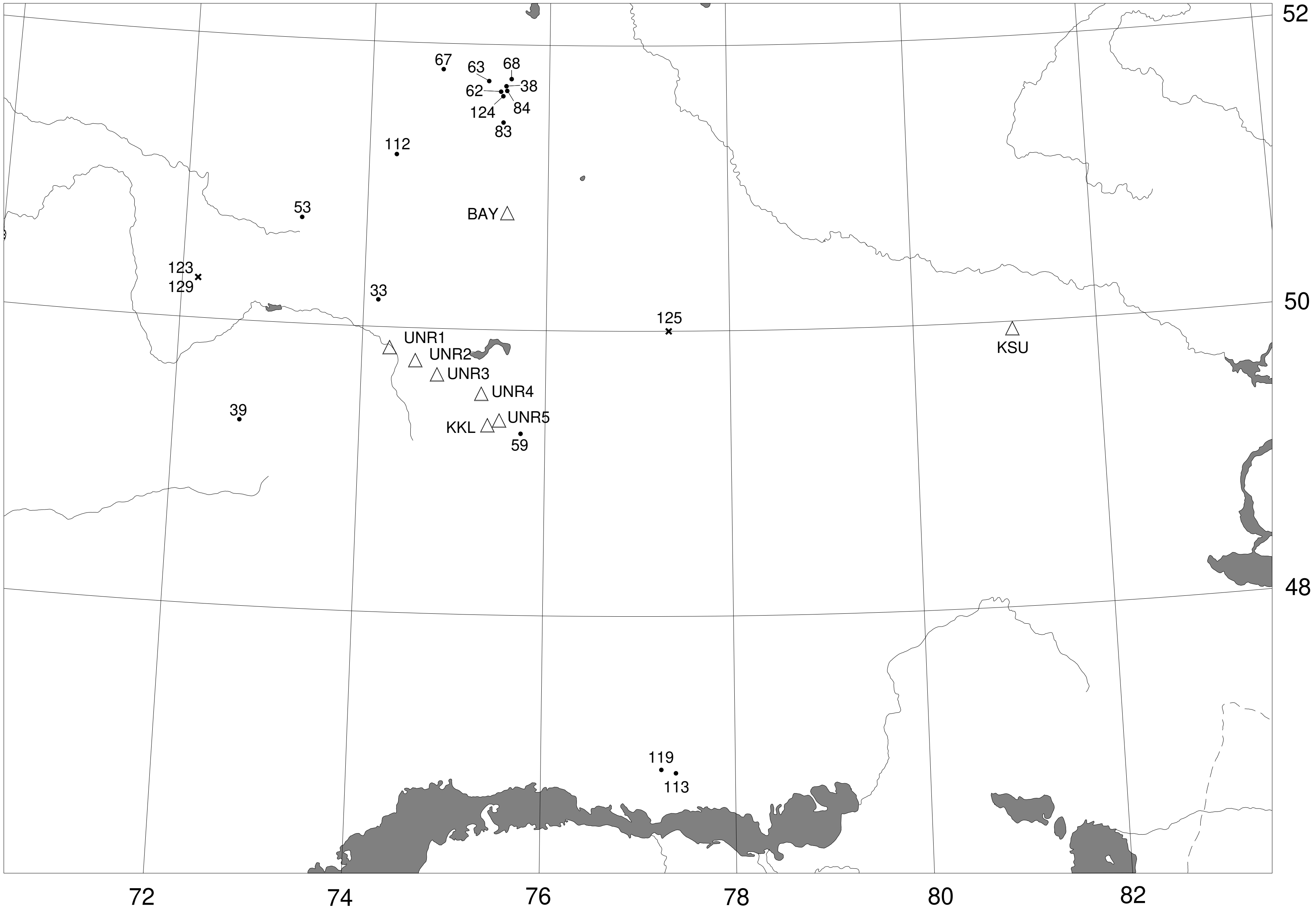


Figure 1. Locations of NRDC stations and local associated events. Triangles denote stations, x’s, the three large chemical explosions, and dots, local mine/quarry blasts. The numbers correspond to **orid** values in table 4.

During the station installation all sensors were carefully aligned to the proper vertical, north and east directions. However, it was later discovered that the local magnetic declinations used in orienting the sensors at BAY, KKL and KSU were not correct and the sensor orientation angles for the raw data in this information product reflect this discrepancy. In addition there were numerous polarity reversals, gain changes and instrument response adjustments.

The portable instruments used at the UNR sites were Teledyne Geotech GS-13 3-component surface instruments and Mark Products L-4C 3-component surface instruments both with nominal 1 second periods. The data acquisition units consisted of 16-bit digitizers with optional anti-aliasing filters with 30 Hz corner frequency sampling at 100 sps.

Nominal instrument response functions in counts per ground velocity for the three instruments at BAY, KKL nd KSU are shown in figures 2-4 and response functions for the two instruments at UNR1 - UNR5 are shown in figures 5 and 6. Note that the actual instrument response files included in this information product give instrument responses in terms of counts per ground displacement.

4. Description of Processing

All of the raw data were obtained from copies of the original field tapes. Header files and corrected station parameter log files were used in conjunction with the field tapes to properly account for time varying data acquisition configurations and instrument characteristics.

We found that the 15 field channel codes used at BAY, KKL and KSU reflected the output of the data acquisition units and did not necessarily correspond to fixed physical channels from the sensors. This caused confusion initially, since a field channel could change from being a surface to a borehole instrument or, in some cases, switched between components for a particular instrument. We decided to establish a set of logical channel codes for the raw data which did retain fixed relationships to physical sensor channels. Since there are three 3-component instruments at two gain levels, the number of logical channels is 18 and we use numerical channel codes to describe the raw data, which are given in table 2. Channel codes for the UNR data were similarly assigned and are given in table 3.

Table 2. Channel naming conventions for BAY, KKL and KSU

Raw	**Processed**	**Description**
01	HSSE	high gain, surface, short period, east (GS-13)
02	HSSN	high gain, surface, short period, north (GS-13)
03	HSSZ	high gain, surface, short period, vertical (GS-13)
04	HBIE	high gain, borehole, intermediate period, east (54100)
05	HBIN	high gain, borehole, intermediate period, north (54100)
06	HBIZ	high gain, borehole, intermediate period, vertical (54100)
07	LSSE	low gain, surface, short period, east (GS-13)
08	LSSN	low gain, surface, short period, north (GS-13)
09	LSSZ	low gain, surface, short period, vertical (GS-13)
10	LSIE	low gain, surface, intermediate period, east (S-1)
11	LSIN	low gain, surface, intermediate period, north (S-1)
12	LSIZ	low gain, surface, intermediate period, vertical (S-1)
13	LBIE	low gain, borehole, intermediate period, east (54100)
14	LBIN	low gain, borehole, intermediate period, north (54100)
15	LBIZ	low gain, borehole, intermediate period, vertical (54100)
16	HSIE	high gain, surface, intermediate period, east (S-1)
17	HSIN	high gain, surface, intermediate period, north (S-1)
18	HSIZ	high gain, surface, intermediate period, vertical (S-1)

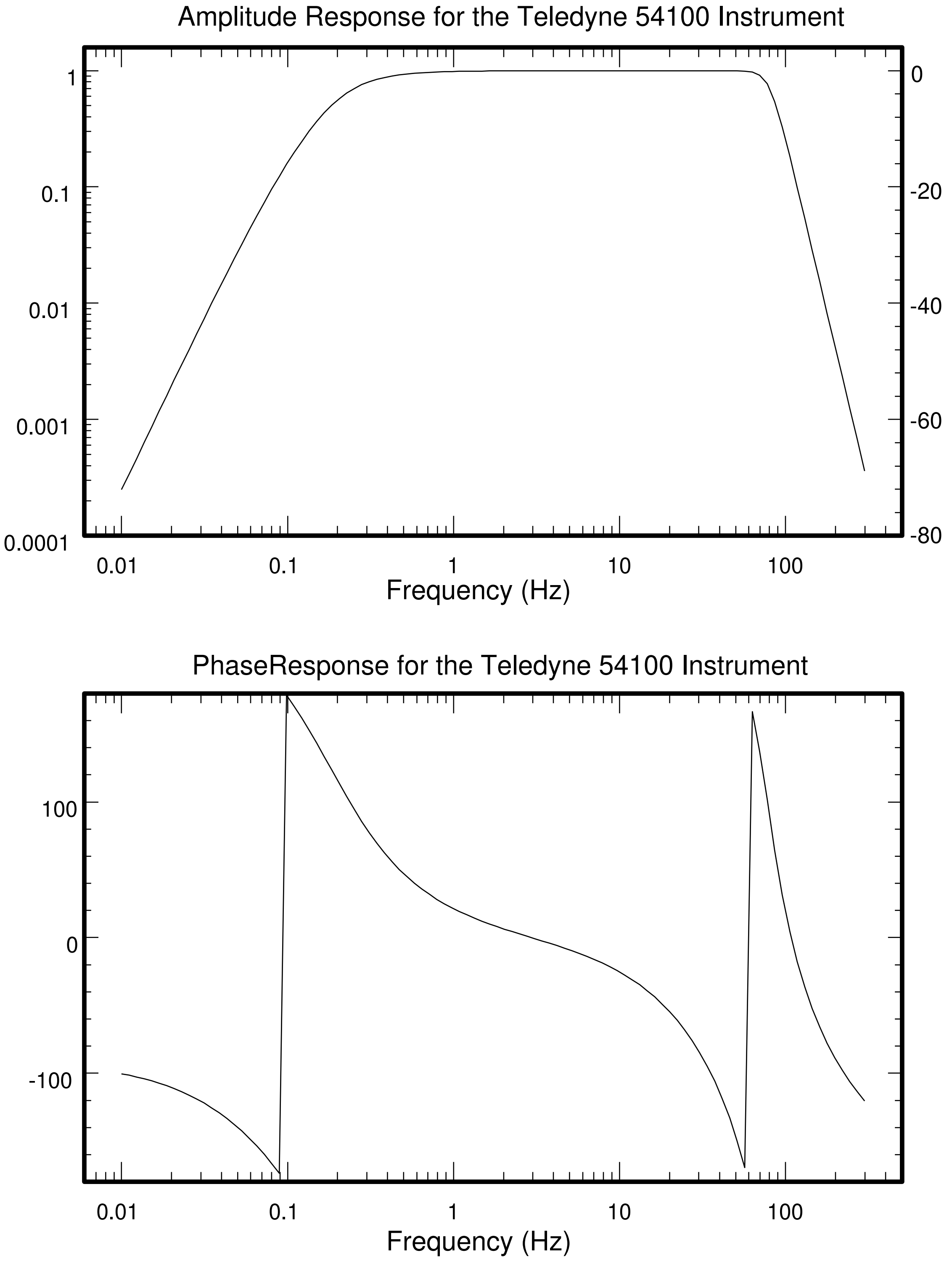


Figure 2. Typical instrument response characteristics for the Teledyne 54100 borehole instruments at BAY, KKL and KSU. These are in terms of instrument output per unit ground velocity and are normalized at 10 Hz.

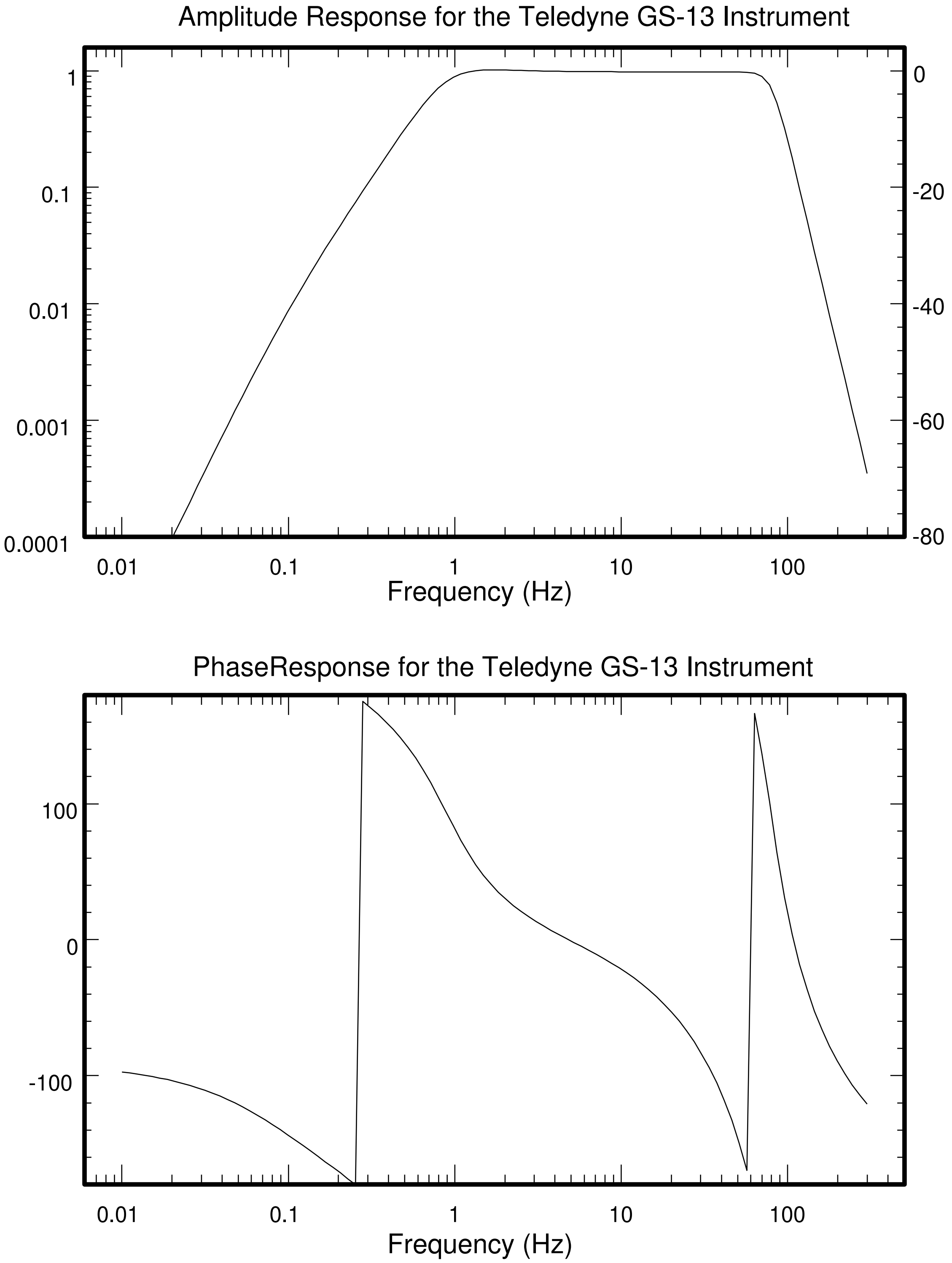
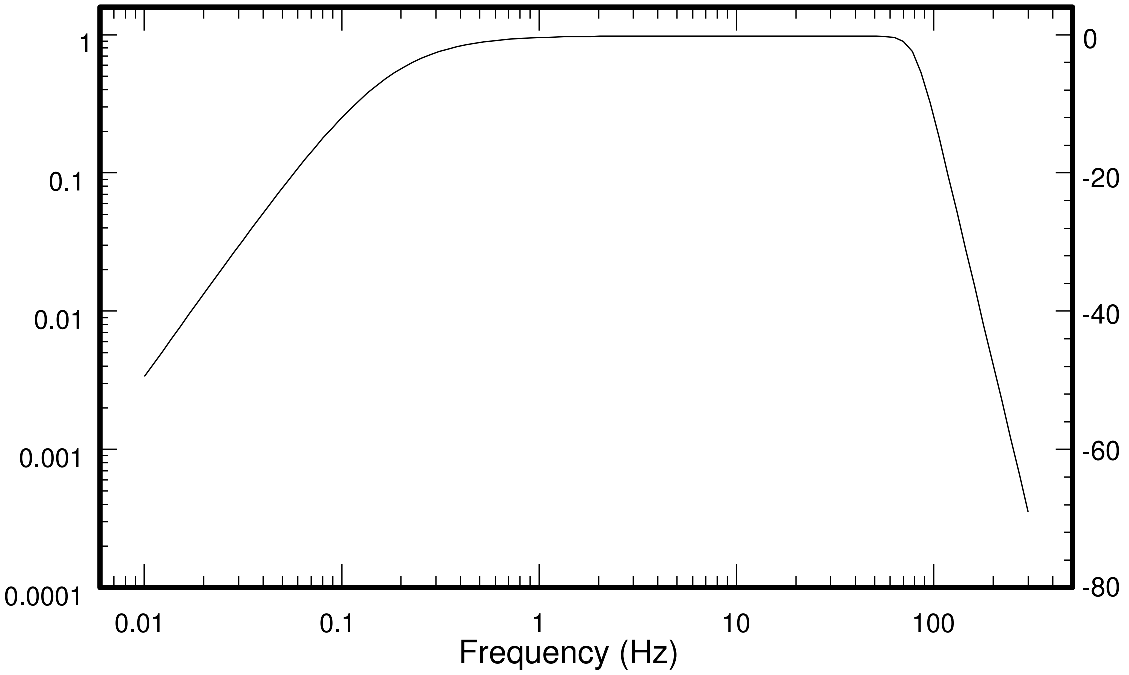
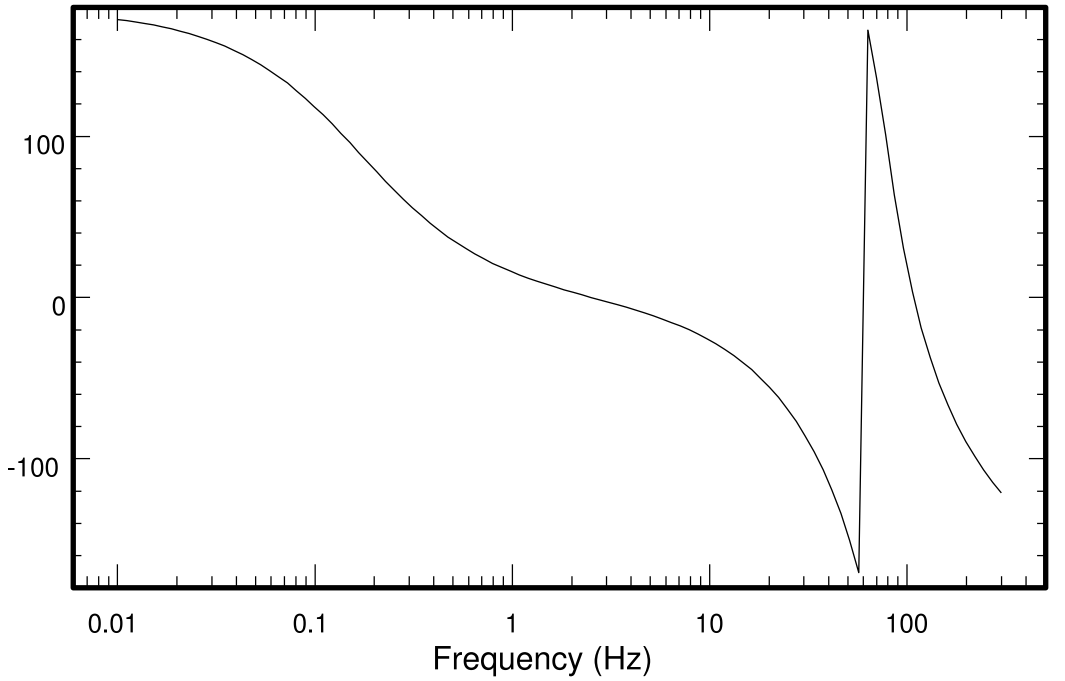


Figure 3. Typical instrument response characteristics for the Teledyne GS-13 surface instruments at BAY, KKL and KSU. These are in terms of instrument output per unit ground velocity and are normalized at 10 Hz.

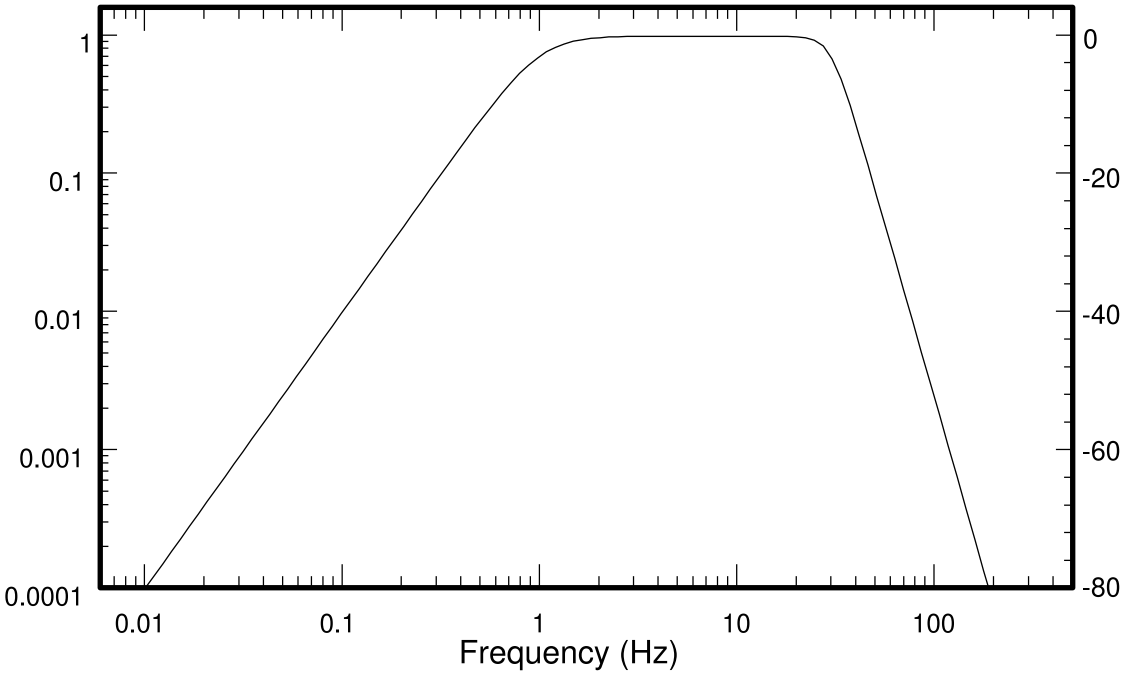


Amplitude Response for the Kinematics S-1 Instrument



PhaseResponse for the Kinematics S-1 Instrument

Figure 4. Typical instrument response characteristics for the Kinematics S-1 surface instruments at BAY, KKL and KSU. These are in terms of instrument output per unit ground velocity and are normalized at 10 Hz.


Amplitude Response for the Mark Products L-4C Instrument

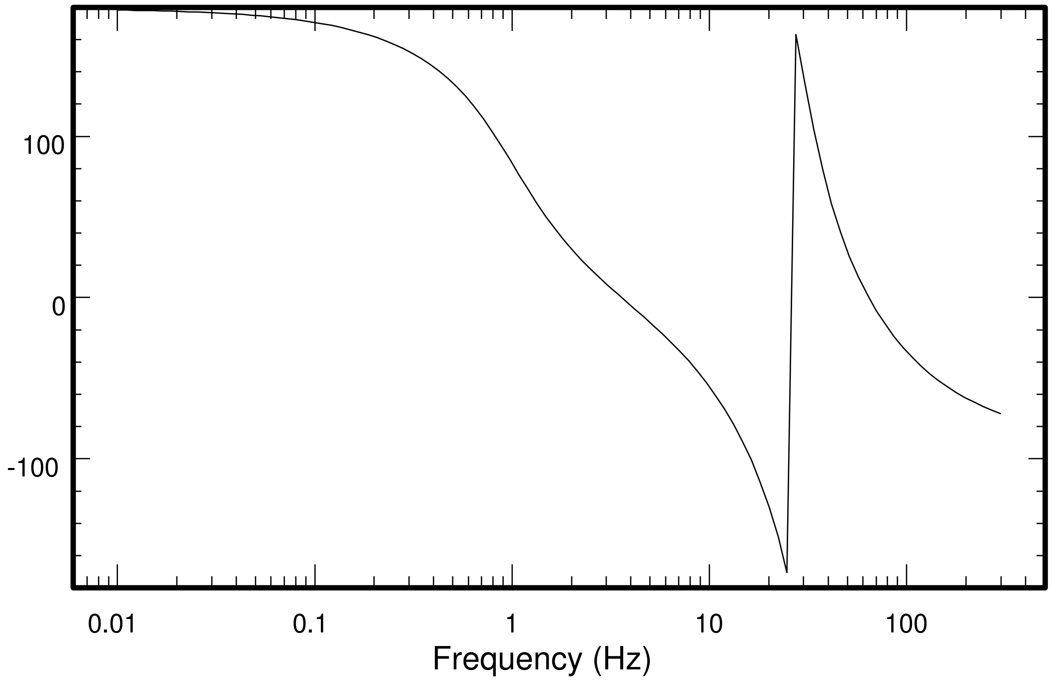
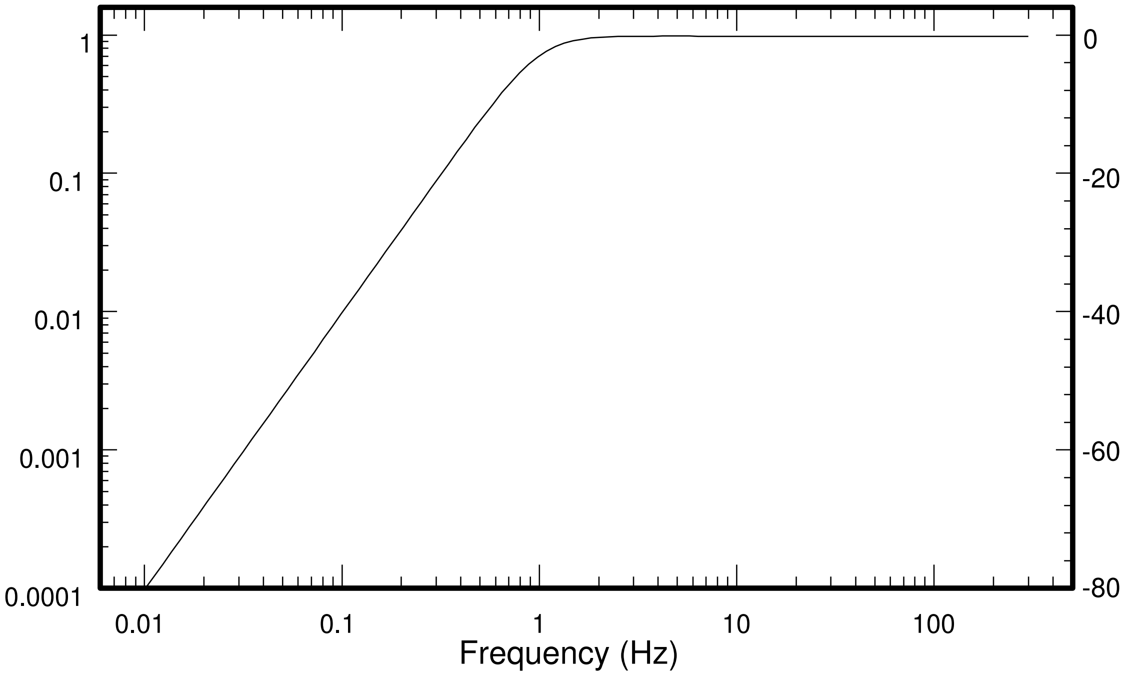

PhaseResponse for the Mark Products L-4C Instrument

Figure 5. Typical instrument response characteristics for the Mark Products L-4C surface instruments at UNR1 - UNR5. These are in terms of instrument output per unit ground velocity and are normalized at 10 Hz.


Amplitude Response for the Teledyne GS-13 Instrument

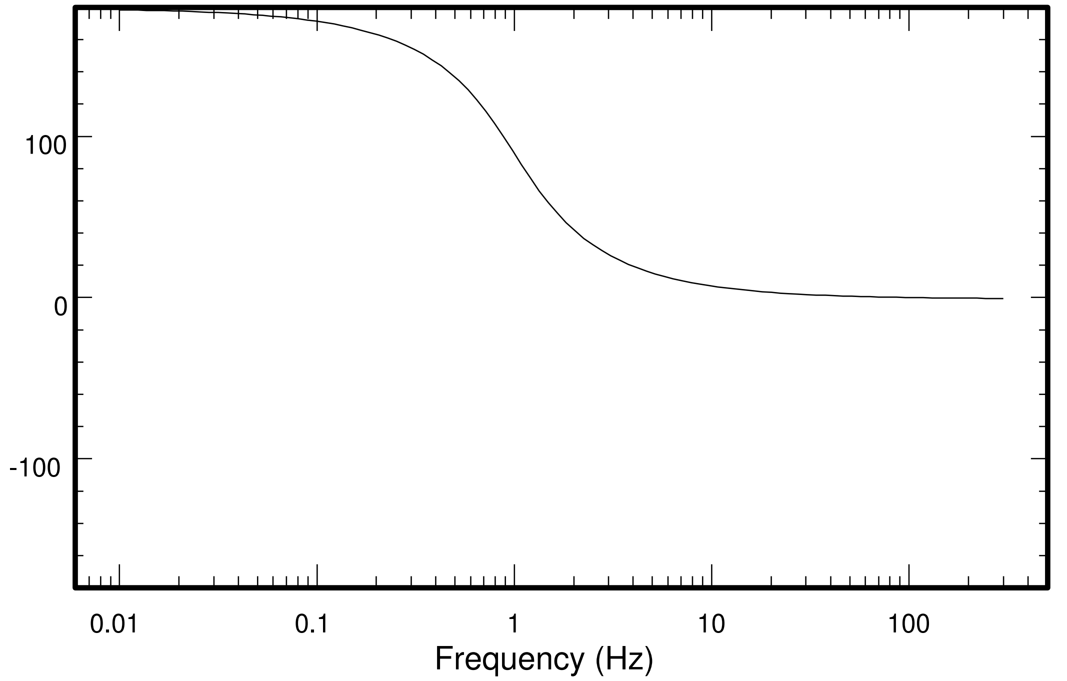

PhaseResponse for the Teledyne GS-13 Instrument

Figure 6. Typical instrument response characteristics for the Teledyne GS-13 surface instruments at UNR1-UNR5. These are in terms of instrument output ment per unit ground velocity and are normalized at 10 Hz.

Table 3. Channel naming conventions for UNR1, UNR2, UNR3, UNR4 and UNR5

Raw	**Processed**	**Description**
01	HSSE	L-4C, east
02	HSSN	L-4C, north
03	HSSZ	L-4C, vertical
04		GS-13, vertical

Initial processing consisted of extracting the 16-bit integer data from field tapes and filling in the relational database with complete and accurate site and instrument characteristics. A detailed description of the database schema is given in section 8. Subsequent processing concentrated on sub-setting the raw data into the "interesting" waveform segments, i.e. those with seismic arrivals. Once the processing was completed, new waveform segments were made and these segments were given the processed channel codes shown in tables 2 and 3. All of the processed waveform segments are in SAC form, in addition to being indexed within the relational database. The detailed processing steps are given below.

1. **Data Review** - Each individual raw waveform was reviewed by an analyst and was given either a pass or fail grade based upon whether or not a seismic arrival was apparent. Clipping of waveforms that passed were also noted.

2. **Data Subsetting** - All of the raw data were batch processed for subsetting using the analyst information from step 1. The rules for subsetting the data were as follows:
   - Time windows were established at each station using the waveform segments that passed step 1. These windows were expanded both in front of and behind the raw segments which passed step 1 to include other nearby detections.
   - All raw waveforms which fell within these time windows were chosen regardless of their grade from step 1. In situations where there was no clipping of the raw data, only the high gain surface short period and borehole channels were used (channels 01, 02, 03, 04, 05 and 06). When clipping occurred on any channel, then all available channels were used.

4. **Deglitching** - All subsetted waveform segments were run through a single point data deglitcher to remove (by interpolation) single point data glitches.

5. **Rotation** - All subsetted waveform segments were rotated into true east, north and up directions. Polarity reversals were removed at this stage.

6. **Conversion to SAC** - All subsetted waveform segments were converted to SAC floating point format. Waveform counts were multiplied by instrument gains resulting in ground velocity units in nm/sec. Site characteristics were written into the SAC headers along with the timing information. Parallel tables within the database were maintained to make these data also available through the relational structure.

7. **Arrival Picking** - First arrival P-wave picks were made by an analyst on the processed waveform data using the following instructions.
   - Only first P-waves were to be picked on a single channel for each station.
   - The high gain surface short period vertical channel (HSSZ) was to be used in most cases.
   - If the arrival was clearly seen at an earlier time on one of the high gain surface short period horizontal channels (HSSE or HSSN), then the pick was made on that channel.
   - If the arrival first break could not be clearly seen on the surface channels, then the high gain borehole channels (HBIZ, HBIE or HBIN) were used instead.

8. **Event Association** - A list of candidate events was composed from the PDE bulletin, the CSS nuclear explosion event database, the CSS NRDC event database, a list of 15 mine and quarry blasts published in Thurber, et. al. (1989) and 3 large chemical explosions with known coordinates and origin times. Associations of picked arrivals were attempted with each candidate event and those with small travel time residuals were retained. The resulting associated events consist of 112 earthquakes, 15 local mine and quarry blasts and the three large chemical explosions. All related information was obtained and entered into the relational database.

5. Associated Events

The list of associated events is shown in table 4. The column labeled **orid** refers to the event index and is used to identify the events. (This is actually the **origin** table index.) The column labeled **nass** refers to the number of picked arrivals (stations) that were associated with the event. The column labeled **src** refers to the source of the event (PDE = PDE, CSS = CSS, THR = Thurber, et. al. list, UCSD = UCSD/IGPP) and the column labeled **type** refers to the event type (eq = earthquake, qb = mine/quarry blast, ex = explosion).

A total of 112 earthquakes, 15 mine and quarry blasts and 3 explosions have been associated. Unfortunately, although we looked carefully, no nuclear test events are contained in these data. Predicted travel times for the associated phases were computed using the IASPEI91 earth model for all of the earthquakes and velocity model A from Thurber, et. al. (1989) for local quarry blasts and explosions (this model is given in file *ndrc*/*data*/*tables*/*nrdc.vmodel*/*kazakh1*).

Table 4. Associated event list

orid	**date/time**	**latitude**	**longitude**	**depth**	**nass**	**Mb**	**Ms**	**src**	**type**
1	078 Mar 19 14:32:15.600	23.6640	64.6880	10.000	1	5.1	4.2	PDE	eq
2	079 Mar 20 13:03:02.300	7.4100	123.7150	607.000	1	4.5	---	PDE	eq
3	080 Mar 21 10:41:35.900	52.0560	-177.5470	93.000	1	6.0	---	PDE	eq
4	083 Mar 24 12:49:47.000	37.4470	137.8650	23.000	1	5.7	5.1	PDE	eq
5	083 Mar 24 21:38:11.200	22.1050	143.8130	122.000	1	4.9	---	PDE	eq
6	099 Apr 09 00:48:55.500	1.2420	128.4320	45.000	1	5.5	6.1	PDE	eq
7	100 Apr 10 10:59:40.500	36.0570	139.7180	70.000	1	5.1	---	PDE	eq
8	101 Apr 11 18:13:28.300	23.9860	122.0270	55.000	2	5.6	---	PDE	eq
9	101 Apr 11 19:02:06.300	1.0240	128.4620	33.000	1	5.2	5.3	PDE	eq
10	102 Apr 12 02:47:20.100	35.5020	23.3700	47.000	1	5.1	---	PDE	eq
11	102 Apr 12 19:30:01.100	5.0940	94.2250	44.000	1	4.8	---	PDE	eq
12	112 Apr 22 20:13:23.100	37.1550	141.5730	30.000	1	6.1	6.6	PDE	eq
13	113 Apr 23 16:39:25.600	5.8760	125.8850	153.000	1	5.3	---	PDE	eq
14	114 Apr 24 04:32:12.300	37.2710	141.2430	57.000	1	4.6	---	PDE	eq
15	115 Apr 25 12:16:48.700	15.8770	120.2260	106.000	1	5.6	---	PDE	eq
16	116 Apr 26 20:02:08.000	31.8760	139.5120	154.000	2	5.2	---	PDE	eq
17	117 Apr 27 04:38:37.300	-3.0560	101.5790	53.000	2	5.3	---	PDE	eq
18	117 Apr 27 05:23:36.100	27.6550	139.8890	478.000	2	4.7	---	PDE	eq
19	117 Apr 27 06:20:46.500	-6.0300	130.5120	149.000	2	5.2	---	PDE	eq
20	117 Apr 27 12:13:34.100	12.9660	123.7480	117.000	1	4.5	---	PDE	eq
21	117 Apr 27 16:59:01.000	37.7880	142.5630	49.000	2	5.1	5.1	PDE	eq
22	118 Apr 28 22:44:33.100	1.9920	99.0400	33.000	1	4.9	---	PDE	eq
23	119 Apr 29 01:45:22.600	27.4370	56.1090	9.000	2	5.9	5.3	PDE	eq
24	120 Apr 30 05:17:37.000	39.7610	74.5730	8.000	2	5.7	5.6	PDE	eq
25	120 Apr 30 13:46:10.800	37.1870	141.3220	61.000	2	4.9	---	PDE	eq
26	121 May 01 06:52:14.400	35.4440	140.6450	74.000	2	4.9	---	PDE	eq
27	122 May 02 19:21:29.600	54.8010	-160.1030	33.000	2	5.1	4.6	PDE	eq
29	123 May 03 17:21:23.400	28.4370	127.6210	217.000	2	5.3	---	PDE	eq
30	132 May 12 20:51:23.500	38.8000	142.1740	56.000	1	4.7	---	PDE	eq
31	133 May 13 23:40:51.539	4.4663	94.9905	65.449	2	3.7	---	CSS	eq
32	134 May 14 00:24:59.200	34.9360	140.6350	56.000	1	4.7	---	PDE	eq
33	134 May 14 09:36:16.400	50.1900	74.1570	-------	2	---	---	THR	qb
35	134 May 14 12:04:03.400	4.8390	122.7660	615.000	2	5.0	---	PDE	eq
36	134 May 14 21:11:35.600	13.6390	120.5970	71.000	2	4.6	---	PDE	eq
37	135 May 15 08:44:04.100	56.3760	-153.1270	33.000	2	4.8	4.5	PDE	eq
38	135 May 15 09:08:35.200	51.7090	75.5140	-------	2	---	---	THR	qb
39	135 May 15 10:35:00.300	49.3040	72.7120	-------	2	---	---	THR	qb
40	136 May 16 03:34:52.100	-7.4510	128.1130	118.000	2	5.4	---	PDE	eq

Table 4. Associated event list (cont)

orid	date/time	latitude	longitude	depth	nass	Mb	Ms	src	type
41	136 May 16 09:25:06.249	27.9415	139.5055	433.923	2	3.4	---	CSS	eq
43	137 May 17 05:12:11.900	-13.5680	167.1540	177.000	1	5.6	---	PDE	eq
44	137 May 17 12:11:09.700	3.0230	97.1190	73.000	1	5.1	---	PDE	eq
45	138 May 18 01:53:51.000	25.2710	94.2020	50.000	2	5.7	5.9	PDE	eq
46	138 May 18 03:07:34.100	49.2820	147.6930	542.000	1	6.1	---	PDE	eq
47	138 May 18 07:18:03.200	44.4080	147.1020	111.000	2	5.3	---	PDE	eq
48	138 May 18 07:23:24.100	-2.2290	100.0330	33.000	2	5.3	6.0	PDE	eq
49	139 May 19 12:56:25.200	-30.2840	-71.4840	36.000	1	5.6	6.0	PDE	eq
50	139 May 19 17:37:46.900	22.0910	144.2490	84.000	2	5.5	---	PDE	eq
51	140 May 20 05:40:19.400	56.5630	-152.7750	33.000	1	5.1	4.8	PDE	eq
52	141 May 21 04:55:37.400	32.0570	141.7050	33.000	1	4.6	---	PDE	eq
53	141 May 21 09:16:43.300	50.7440	73.2790	-------	2	---	---	THR	qb
54	141 May 21 09:20:22.700	-7.4650	128.4160	33.000	1	5.0	---	PDE	eq
55	141 May 21 12:25:27.000	24.1960	125.1830	42.000	1	5.0	5.1	PDE	eq
56	141 May 21 13:39:42.500	24.2150	125.2260	43.000	1	4.5	---	PDE	eq
57	141 May 21 15:36:08.800	24.2140	125.2340	39.000	1	4.9	4.6	PDE	eq
58	143 May 23 07:01:04.000	51.3070	-176.0740	33.000	1	4.7	---	PDE	eq
59	143 May 23 08:49:22.700	49.2750	75.7380	-------	2	---	---	THR	qb
60	143 May 23 17:09:04.400	8.0470	125.4100	32.000	2	5.1	5.2	PDE	eq
61	144 May 24 13:13:56.200	39.4240	14.9550	313.000	1	4.5	---	PDE	eq
62	145 May 25 09:27:43.700	51.6700	75.4540	-------	2	---	---	THR	qb
63	145 May 25 09:56:40.900	51.7430	75.3160	-------	2	---	---	THR	qb
64	145 May 25 11:31:54.300	63.8500	-19.7280	8.000	1	5.8	5.8	PDE	eq
65	146 May 26 01:37:42.200	-5.5190	133.9590	33.000	2	5.3	---	PDE	eq
66	146 May 26 01:49:52.200	29.9260	142.1740	68.000	2	4.7	---	PDE	eq
67	146 May 26 05:31:04.800	51.8190	74.7970	-------	2	---	---	THR	qb
68	146 May 26 08:33:26.500	51.7600	75.5710	-------	2	---	---	THR	qb
69	146 May 26 12:07:52.500	38.2220	142.3180	37.000	2	5.3	5.2	PDE	eq
70	146 May 26 13:44:19.900	42.9230	78.0630	20.000	1	4.6	---	PDE	eq
71	146 May 26 19:23:57.100	31.6180	60.2090	17.000	1	4.8	4.0	PDE	eq
72	149 May 29 03:26:06.400	-5.0660	102.6520	45.000	1	5.1	4.2	PDE	eq
73	149 May 29 06:27:50.700	34.0760	48.2660	41.000	1	4.9	4.6	PDE	eq
74	149 May 29 17:37:37.300	22.4750	143.2650	119.000	1	5.2	---	PDE	eq
75	151 May 31 18:32:17.000	0.7480	121.9410	79.000	1	5.2	---	PDE	eq
76	152 Jun 01 01:19:30.300	-2.3080	138.9810	33.000	1	4.8	3.8	PDE	eq
77	152 Jun 01 21:33:37.900	24.9530	122.3660	161.000	1	4.7	---	PDE	eq
78	153 Jun 02 16:44:51.400	-4.5370	151.9410	145.000	1	5.3	---	PDE	eq
79	160 Jun 09 06:19:25.400	-6.3660	148.8320	54.000	1	5.3	6.2	PDE	eq
80	161 Jun 10 14:50:11.500	37.2300	21.4650	41.000	2	5.2	5.1	PDE	eq
81	161 Jun 10 16:03:55.100	4.1350	94.8100	52.000	1	5.4	5.3	PDE	eq
82	161 Jun 10 17:25:56.700	8.1670	93.6930	90.000	2	4.8	---	PDE	eq
83	162 Jun 11 12:42:04.900	51.4540	75.4880	-------	2	---	---	THR	qb
84	162 Jun 11 12:50:34.300	51.6770	75.5250	-------	2	---	---	THR	qb
85	163 Jun 12 18:01:33.600	-5.3640	150.3110	217.000	1	5.2	---	PDE	eq
86	164 Jun 13 08:58:04.300	-2.4490	138.6700	44.000	1	5.4	4.4	PDE	eq
87	164 Jun 13 14:00:39.300	44.6710	150.3920	42.000	1	5.8	5.8	PDE	eq
88	164 Jun 13 15:41:10.800	44.6250	150.3870	44.000	1	5.6	5.4	PDE	eq
89	165 Jun 14 05:07:26.800	10.4290	126.1470	13.000	1	5.6	5.9	PDE	eq
90	165 Jun 14 07:19:17.900	21.1130	94.7550	125.000	1	4.5	---	PDE	eq

Table 4. Associated event list (cont)

orid	date/time	latitude	longitude	depth	nass	Mb	Ms	src	type
91	166 Jun 15 00:18:50.700	-7.0850	129.4400	140.000	1	5.5	---	PDE	eq
92	175 Jun 24 21:02:54.000	7.9730	126.6440	77.000	1	4.9	---	PDE	eq
93	176 Jun 25 02:49:39.700	47.2970	-27.4630	10.000	1	5.5	5.4	PDE	eq
94	176 Jun 25 03:27:07.400	41.5190	143.6120	38.000	1	5.4	4.7	PDE	eq
95	177 Jun 26 22:04:15.500	-3.4260	139.4460	10.000	1	4.9	4.4	PDE	eq
96	178 Jun 27 00:17:04.600	-2.1640	138.1700	21.000	1	5.7	6.5	PDE	eq
97	183 Jul 02 13:10:41.000	43.4090	148.0590	31.000	1	5.3	5.3	PDE	eq
98	232 Aug 20 21:14:56.300	22.1040	144.1980	116.000	1	4.7	---	PDE	eq
99	233 Aug 21 18:20:10.000	-5.5280	151.8850	46.000	1	5.1	---	PDE	eq
100	233 Aug 21 18:22:37.400	-5.4830	151.7320	36.000	3	5.7	5.7	PDE	eq
101	233 Aug 21 19:44:19.000	-6.5260	154.8770	62.000	3	5.6	---	PDE	eq
102	234 Aug 22 00:21:52.900	43.8110	85.2900	58.000	2	4.4	---	PDE	eq
103	234 Aug 22 05:09:14.400	52.1640	174.0660	33.000	2	5.5	4.8	PDE	eq
104	235 Aug 23 23:08:08.400	46.5300	153.6130	33.000	2	5.0	4.4	PDE	eq
105	235 Aug 23 23:53:59.900	46.4720	153.7330	33.000	1	5.0	---	PDE	eq
106	236 Aug 24 03:57:40.700	21.2070	144.1780	33.000	2	5.0	4.4	PDE	eq
107	236 Aug 24 09:24:40.500	23.0890	94.4020	93.000	3	5.1	---	PDE	eq
108	237 Aug 25 09:47:02.400	43.4250	145.8920	78.000	1	4.9	---	PDE	eq
109	238 Aug 26 01:40:30.400	9.3420	122.4280	24.000	1	5.2	4.6	PDE	eq
110	238 Aug 26 01:44:46.800	9.3570	122.3630	30.000	2	5.2	4.7	PDE	eq
111	238 Aug 26 19:32:49.500	2.1290	126.6170	45.000	2	5.2	4.4	PDE	eq
112	239 Aug 27 08:52:53.000	51.2130	74.3020	-------	2	---	---	THR	qb
113	239 Aug 27 09:38:34.800	46.9000	77.3890	-------	3	---	---	THR	qb
115	241 Aug 29 00:31:49.500	-4.1440	152.3340	162.000	3	5.0	---	PDE	eq
116	241 Aug 29 10:12:57.800	37.4770	142.9510	39.000	1	4.7	4.3	PDE	eq
117	242 Aug 30 05:57:52.800	33.2810	57.0660	33.000	3	4.8	---	PDE	eq
118	244 Sep 01 03:44:38.800	43.8080	85.9480	-------	2	---	---	THR	eq
119	244 Sep 01 09:09:52.000	46.9240	77.2410	-------	2	---	---	THR	qb
120	244 Sep 01 13:36:43.600	32.3790	141.1200	60.000	1	5.0	---	PDE	eq
121	244 Sep 01 21:38:24.100	36.3310	69.1330	33.000	3	4.2	4.1	PDE	eq
122	245 Sep 02 00:37:13.700	24.5100	123.2610	79.000	2	5.0	---	PDE	eq
123	245 Sep 02 07:00:00.320	50.2810	72.1720	0.025	7	---	---	UCSD	ex
124	245 Sep 02 08:02:10.200	51.6390	75.4810	-------	3	---	---	THR	qb
125	245 Sep 02 09:27:04.950	50.0000	77.3360	0.017	4	---	---	UCSD	ex
126	246 Sep 03 00:52:55.300	21.2700	93.8290	53.000	1	4.8	---	PDE	eq
127	246 Sep 03 01:15:33.700	-2.8360	129.5100	20.000	2	5.7	4.8	PDE	eq
128	246 Sep 03 04:10:13.400	9.3600	122.2790	52.000	2	5.3	4.7	PDE	eq
129	246 Sep 03 07:00:00.060	50.2810	72.1720	0.025	7	---	---	UCSD	ex
130	246 Sep 03 12:39:55.000	35.8950	30.6230	45.000	1	5.0	4.4	PDE	eq
131	246 Sep 03 20:35:17.500	9.3670	122.4460	33.000	1	4.9	4.2	PDE	eq
132	247 Sep 04 01:40:29.900	45.6650	26.4720	161.000	1	4.4	---	PDE	eq
133	247 Sep 04 04:27:08.800	49.2930	156.4100	33.000	2	5.9	6.2	PDE	eq
134	247 Sep 04 23:31:17.200	43.3430	146.5820	51.000	1	4.6	---	PDE	eq

Arrival phase identifiers have been modified for those arrivals which have been associated with events by appending the **orid** to the normal phase code. Station-event distances, azimuths, back azimuths and time residuals have been computed for each associated arrival and are given in the **assoc** table. Travel time residual mean and standard deviation in seconds for all of the stations broken down by event type are given in tables 5 and 6. We found that three of the original event origin times in the Thurber publication were apparently exactly one minute in error and we corrected these times in this database accordingly.

Table 5. Travel time residual statistics for BAY, KKL and KSU

	BAY		KKL		KSU	
type	**mean**	**stdv**	**mean**	**stdv**	**mean**	**stdv**
eq	1.52	1.28	2.03	1.02	1.94	1.42
qb	-0.38	0.85	0.05	0.54	0.49	0.11
ex	-0.24	0.45	0.18	0.59	1.05	0.11

Table 6. Travel time residual statistics for UNR1 - UNR5

	UNR1		UNR2		UNR3		UNR4		UNR5	
type	**mean**	**stdv**	**mean**	**stdv**	**mean**	**stdv**	**mean**	**stdv**	**mean**	**stdv**
ex	1.34	0.59	3.43		-3.94	9.73	3.68	2.31	5.89	

The locations of the explosions and quarry blasts are shown in figure 1 where each location is labeled with its **orid** value. The locations of the earthquakes are shown in figure 7 where each event is symbol coded according to magnitude. As an aid for relating station to event azimuths and distances, an orthographic projection of the earth with station BAY at the origin is shown in figure 8. Each circle on this plot is at ten degree increments from BAY.

6. Review of Data - Quality Control

All of the field data acquisition units produced segmented data, in most cases, based upon a real time event detector algorithm and, in some cases (e.g. for the three large chemical explosions), based upon pre-specified time windows. The original raw data set consists of approximately 20,000 such waveform segments and the processed data consists of approximately 6,000 segments. In processing the data, we obtained the reduction of waveform segments mainly through the elimination of low gain channels in cases where there was no clipping. To a lesser extent we reduced the number of processed data samples by eliminating flat traces and apparent false triggers.

For the most part, the data recorded at BAY, KKL, and KSU were of high quality with few data glitches or dropouts. In addition to visually inspecting the data and picking phase arrivals, we performed two independent quality control checks; phase arrival association with known events and station to event azimuth residual calculations based upon P-wave polarization analysis. Considering the reasonable values of the travel time residuals shown in table 5 for stations BAY, KKL and KSU, we have no indications of timing errors at these stations. The results of the polarization analysis are less conclusive and the results are shown in figure 9. We have plotted the azimuth residual as a function of azimuth to the event for BAY, KKL and KSU to check for systematic residual vs. azimuth effects. The E's represent earthquakes and the Q's represent quarry blasts or explosions. The polarization computations were done by diagonalizing the covariance matrix formed from the three components over a fixed 0.5 second time window that started at the picked arrival time for each event. We made no attempt to tune the averaging window and we batch processed all events at the same time. Only residuals whose associated rectilinearity was higher than 0.9 are plotted in figure 9. The overall statistics for these residuals are shown in table 7. Earthquake are type 'eq', quarry blasts and explosions are type 'qb' and the total statistics are type 'tl'.

Table 7. Azimuth residual statistics for BAY, KKL and KSU

	BAY		KKL		KSU	
type	**mean**	**stdv**	**mean**	**stdv**	**mean**	**stdv**
eq	-9.35	12.89	-4.87	11.63	-11.32	18.67
qb	2.60	16.94	4.50	37.81	-11.72	11.30
tl	-6.96	14.47	-3.23	19.04	-11.38	17.56

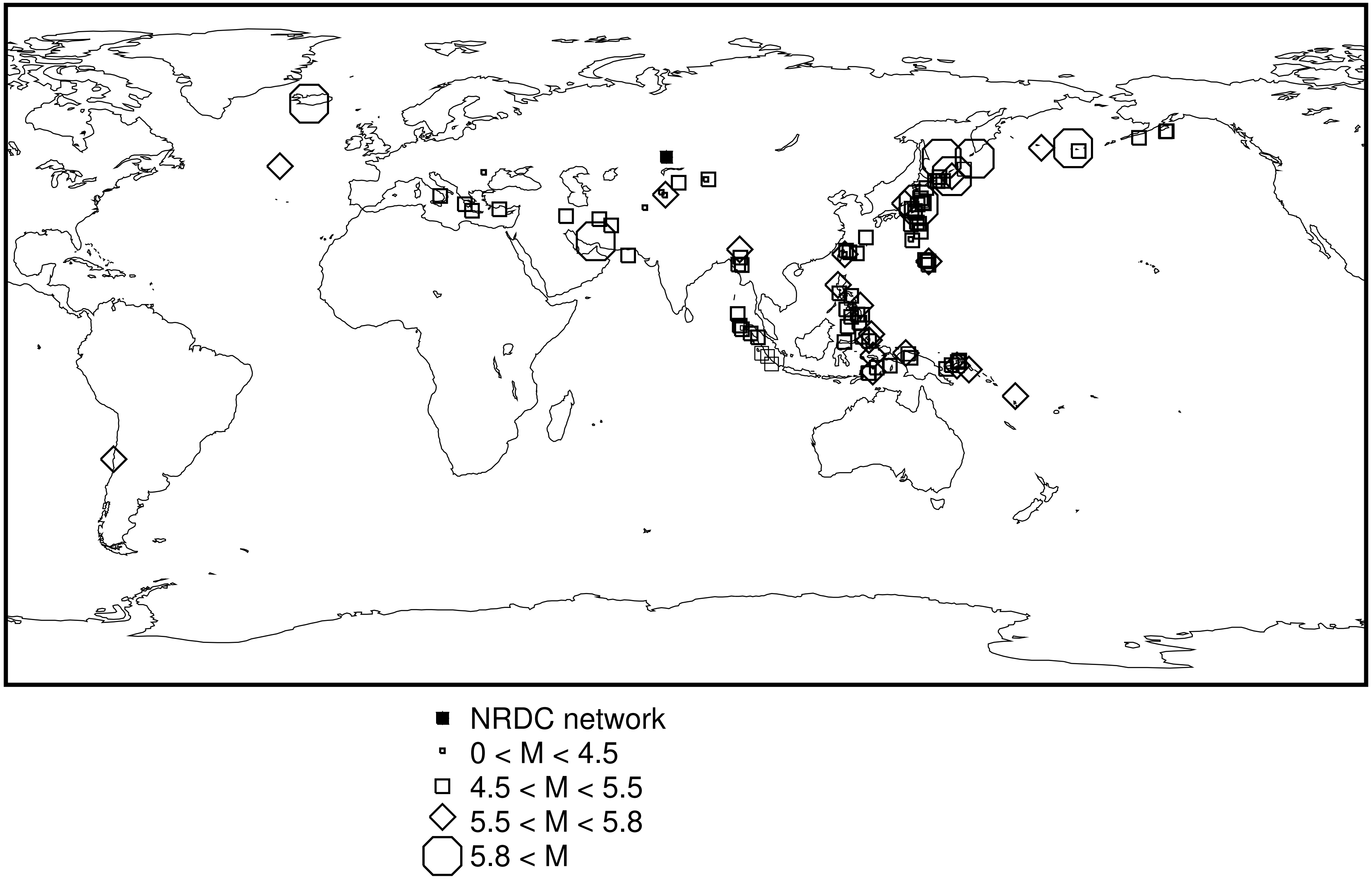


Figure 7. Teleseismic events associated with phase arrivals from the NRDC data.

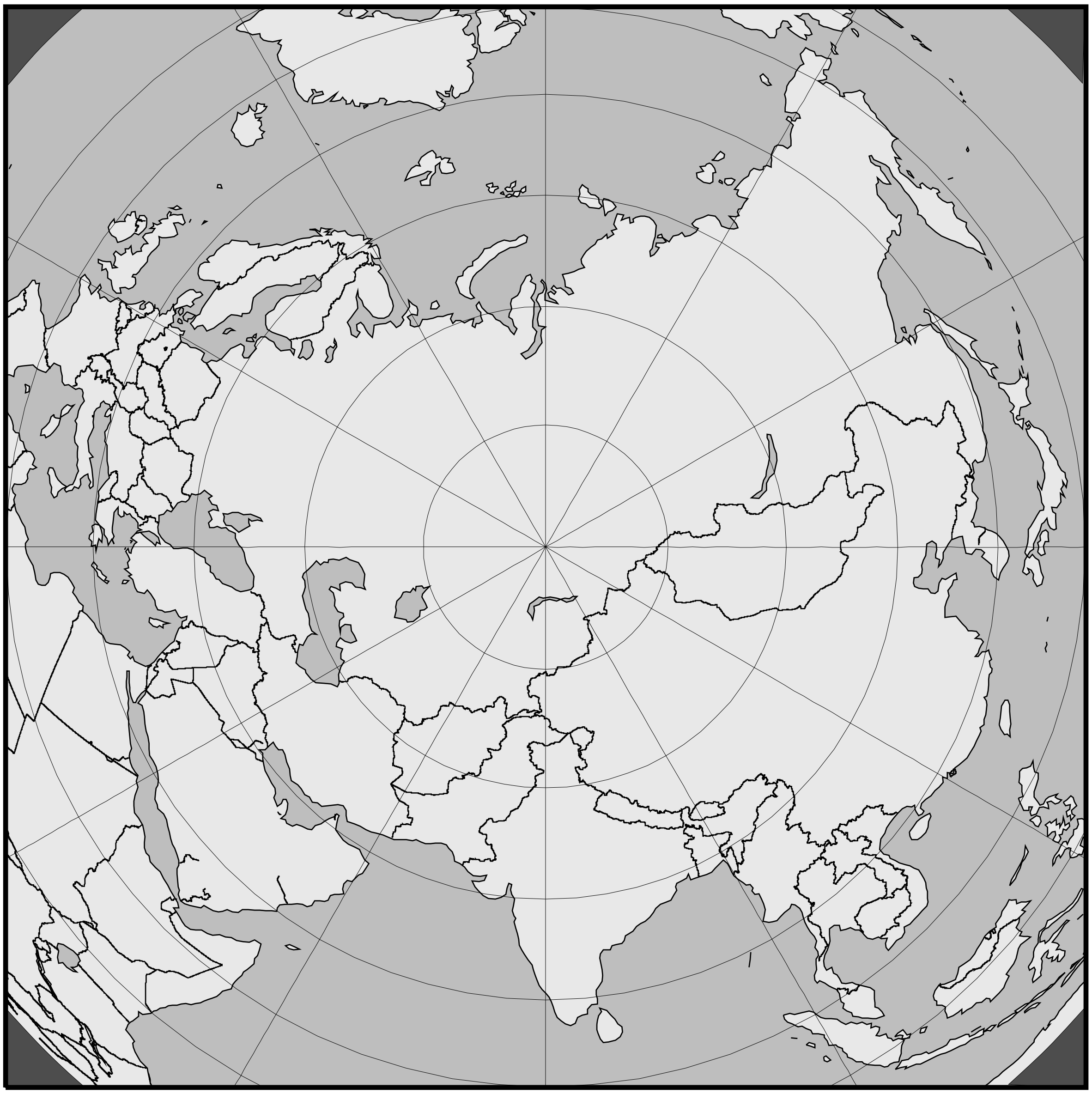
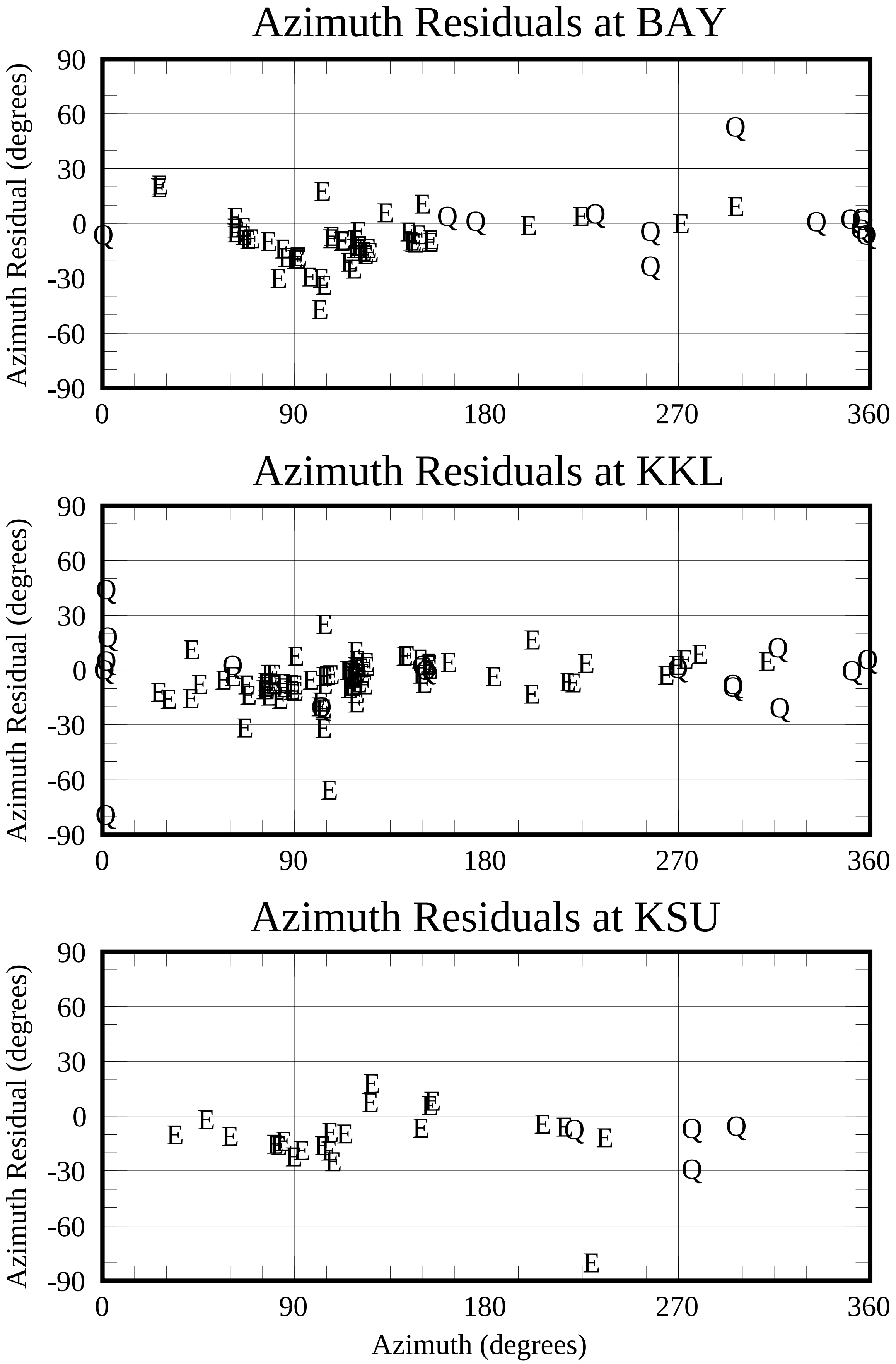


Figure 8. Orthographic projection from station BAY. Circles are at 10 degree distance intervals.



Azimuth Residuals at BAY

Azimuth Residuals at KKL

Azimuth Residuals at KSU

Figure 9. Azimuth residuals based upon P-wave polarization analysis. A 0.5 second time window was used. E’s represent earthquakes and Q’s quarry blasts or explosions.

There seems to be consistent negative azimuth residuals which could indicate that there are still errors in the magnetic declination used to align and subsequently correct the sensor orientation angles. Most of the scatter is due to the relatively crude polarization analysis that we used, however this analysis does confirm polarity and channel identification. In each case where the residual was outside of 30 degrees we looked at the waveforms individually and were able to identify the problem as being either 1) incorrect time window for the polarization analysis, 2) low signal to noise level resulting in extreme noise contamnination or 3) filtering was required to enhance the coherent part of the incoming P-wave. With only one exception, event 112 at station BAY, we were able to confirm that the polarity exhibited by the data was consistent with the know event azimuth. The residuals are rather large especially from events coming from the east. We can see in figure 8 that these are from southern Japan and there may be a local slab effect at the source that is creating a dog-leg path in the ray that would deflect it to the north. At any rate this is a subject for future research.

We have used the event association and polarization analyses to verify timing, channel orientations and polarities for the stations at BAY, KKL and KSU and we presently have no indications of other timing or gain errors for these data. We encourage users of this information product to contact us if any problems are uncovered which we have not found. Table 8 lists the number of waveform segments (**nsegs**), the number of picked arrivals (**narrs**), the number of associated events (**nevts**) and the total number of seconds recorded (**time**) for the processed data broken down by station and day. All of these numbers are on a per channel basis.

The large travel time residuals shown in table 6 indicate that there may be timing errors in the data recorded at UNR1, UNR2, UNR3, UNR4 and UNR5. We discovered, and attempted to fix, a large timing error in UNR4 channel HSSN for the first chemical explosion (event 123). The source of this error was a chunk of missing data samples in the middle of the buffer that caused the event onset to be too early. We also suspect that the system gains may be wrong for UNR1 for the first and second explosions, although they look ok for the third explosion. In a future release of this information product we hope to clear up some of these problems.

Table 8. Processed data status for BAY, KKL and KSU

	BAY				**KKL**				**KSU**			
day	**nsegs**	**narrs**	**nevts**	**time**	**nsegs**	**narrs**	**nevts**	**time**	**nsegs**	**narrs**	**nevts**	**time**
078					1	1	1	47				
079					4	4	1	236				
080					6	3	1	27				
083					3	3	2	113				
099					6	3	1	655	2			273
100					6	4	1	740				
101					2	2	2	136	2	1	1	104
102					2	2	2	108	1			61
103					4	2		127	1	1		57
112					4	2	1	789				
113					9	5	1	383				
114					10	6	1	696				
115					4	3	1	1673				
116					2	1	1	135	3	1	1	125
117					14	10	5	610	7	6	4	355
118					5	4	1	274	1	1		54
119					3	2	1	1345	6	3	1	754
120					5	4	2	1421	7	4	2	1146

Table 8. Processed data status for BAY, KKL and KSU (cont)

	BAY				**KKL**				**KSU**			
day	**nsegs**	**narrs**	**nevts**	**time**	**nsegs**	**narrs**	**nevts**	**time**	**nsegs**	**narrs**	**nevts**	**time**
121					1	1	1	55	1	1	1	28
122					1	1	1	43	1	1	1	89
123					1	1	1	95	3	2	1	391
124									2	1		58
132					1	1	1	45				
133	10	9	1	956	4	2	1	216				
134	10	8	4	920	9	7	3	449				
135	9	8	3	1069	9	5	3	587				
136	6	5	2	633	6	4	2	377				
137	1	1	1	53	2	1	1	121				
138	5	4	4	3854	3	3	3	1670				
139	10	5	2	820	3	2	1	577				
140	7	3	1	437	5	3		387				
141	2	1	1	10	8	7	6	453				
142	7	4		513	4	2		264				
143	2	2	2	217	3	3	3	289				
144					2	2	1	156				
145	5	3	3	588	7	6	2	968				
146	10	8	5	1362	8	8	7	1917				
147					2	2		172				
148	11	11		1565								
149	16	15	3	2979								
150	4	4		389								
151	2	2	1	334								
152	3	3	2	290								
153	13	13	1	1690								
159					1	1		883				
160					4	3	1	442				
161	4	4	2	709	4	3	3	989				
162	12	12	2	1988	2	2	2	196				
163	9	4		1441	2	1	1	156				
164	11	10	3	1816	1	1		132				
165	2	2	2	416								
166	1	1	1	330								
175					1	1	1	103				
176					8	8	2	1305				
177	1	1	1	80								
178	1	1	1	430								
183					2	2	1	318				
230									2	1		189
231					2	1		226	1	1		80
232					6	5	1	756				
233	2	2	2	287	7	5	3	797	4	3	2	624
234	7	5	1	1445	4	3	2	944	5	3	1	826
235	3	1	2	236	3	3	1	281	2	1		312
236	23	21	2	3044	14	10	2	1784	6	5	1	1016

Table 8. Processed data status for BAY, KKL and KSU (cont)

	BAY				**KKL**				**KSU**			
day	**nsegs**	**narrs**	**nevts**	**time**	**nsegs**	**narrs**	**nevts**	**time**	**nsegs**	**narrs**	**nevts**	**time**
237	14	12	1	1457	2	2		174	2	2		355
238	13	10	2	1309	3	3	3	438	1			77
239	17	15	2	2070	14	10	2	1632	8	8	1	1280
240	15	15		2035	9	8		878	6	2		648
241	6	5	1	798	4	4	2	499	7	3	1	719
242	2	1	1	372	8	3	1	863	9	6	1	1039
243	11	11		1310	11	7		1018	9	7		1058
244	15	11	2	2013	9	5	3	1586	8	5	3	1311
245	15	12	4	3177	15	12	4	2516	13	5	3	2765
246	10	13	3	2281	15	12	5	2267	10	8	2	1818
247					7	8	3	1012	4	3	1	663
248					5	5		580				
tot	327	278	71	47821	332	250	105	41386	134	85	28	18286

7. Information Product Overview

The NRDC information product consists of three major parts; the data, including raw and processed waveforms along with all other parametric information, documentation, including a postscript version of this document along with software and database documentation, and software, including complete source code, make files, SUN-sparc executable files and demonstration programs and scripts. In addition to the product itself, a comprehensive installation script is provided which allows the user to bring in sub-sets of the total data volume based upon station, channel and time sifting keys.

The intent of this information product is to provide the scientific community with a tightly integrated package of data, software and documentation which can be immediately useful in research efforts. In addition to the normally provided information associated with raw seismic data, such as detailed instrument and site characteristics, the NRDC information product also provides a base level of "vanilla" processing, such as coordinate rotation into fixed directions and phase arrival picking and event association. We hope that we have eliminated the necessity to wade through large volumes of data looking for the "good stuff" and the waste of valuable time hunting out peripheral, but important, pieces of information associated with the data.

Although we have spent a consider effort to weed out the "good stuff" and to identify all of the peripheral, but important, information, we are not perfect and we encourage the users of this product to let us know if we have left anything out or if there are inaccuracies in the data we have distributed. Our general plan is to release new versions of information products whenever we have significantly increased the information content or when inaccuracies have been identified and fixed and we welcome any help from those who use these products in defining new releases and other future information products.

At the JSPC, we use relational database technology for the assimilation of new information products. We have found that this technology gives us a high degree of flexibility, in terms of layering information as it is obtained and fixing errors that are discovered at a later time. All relational databases are simply a collection of flat tables (or *relations* in database parlance). By flat, we mean a matrix of rows (*tuples*) and columns (*attributes*) with each cell occupied by a single number or character string. The database model that we use is, strictly speaking, a heterogeneous relational database, since we store the actual seismic waveform data outside of the flat tables in their own binary files. The database tables reference the waveform data through file name-byte offset-format description pointers within the tables.

A particular definition of a set of database tables and the columns within each table is known as a *schema*. The database schema is the most important aspect of the design of a database and it usually matures over a period of time as it is being used. We use a schema that was specifically designed for seismic data, especially in the context of nuclear verification and non-proliferation research, by DARPA's Center for Seismic Studies (CSS). The CSS version 3.0 relational database schema has evolved from previous versions that have been in use for over ten years and this schema is at the heart of a currently operational seismic processing system used by DARPA to provide global monitoring for underground nuclear testing.

We provide information product distributions in the original form that we used in assimilating the products. The NRDC information product contains a heterogeneous relational database which provides the major organization for all of the information contained in the product. A problem with using relational database technology is that it is typically tied to proprietary, and often expensive, software products called relational database management systems (RDBMS) which are used to arrange, sort and join information from the various tables. Most RDBMS's store the data tables within opaque structures that make it impossible to access the information without using the RDBMS. In order to avoid this problem, we store all tables as ASCII fixed-format files and we have provided a simple set of open and freely available software that will provide the necessary database management functions. The database schema description and the associated files for this product are described in section 8. The software is described in section 9.

The raw data within this product are all stored, according to their original 16-bit integer forms, as two byte binary integers within separate, headerless files for each waveform segment. The tables within the relational database are used to store all other related information, such as station codes, timing information and instrument characteristics. The IRIS DMC will provide a SEED version of the raw data upon request. The software provided in this product can be used to display the raw data and all other information within the database. Programmatic interfaces are provided for access to the database and the waveform samples along with programs to convert data from the relational database to SAC and AH formats. The processed data waveform files are all in SAC form with appropriate SAC headers. These same files are referenced by the relational database so that the processed waveforms can be accessed in the same manner as the raw data. This information product provides users with multi-tiered access to the data. In the case of the processed data, users can avoid the relational database completely and use the SAC files directly. Alternatively, users can take advantage of the software that is included with the product to access the data directly through the relational database.

The NRDC information product distribution in its entirety consists of approximately 27,000 files organized within a UNIX directory tree structure, which is shown in figure 10. The contents of the various sub-directories are given below.

nrdc/*doc*/

This sub-directory contains the general documentation files including a postscript version of this file (*tech_report.ps*), a postscript version of the software UNIX manual pages (*man.ps*) and a postscript version of the database schema description (*schema.ps*).

nrdc/*data*/

This sub-directory contains the data sub-directories *raw*/, *proc*/ and *tables*/.

nrdc/*data*/*raw*/

This sub-directory contains 77 raw data sub-directories each corresponding to one day's worth of raw data with names *1987078*/ - *1987248*/. These sub-directories in turn contain local databases with only one day's worth of raw data. The waveform reference tables (**wfdisc**) are stored in files with names *1987078.wfdisc*, for example, within these day sub-directories. Also in every day sub-directory is another sub-directory, *wf*/, which contains all of the raw waveform segment files for that day. The raw waveform segment file names are coded by station, channel and waveform index.

nrdc/*data*/*proc*/

This sub-directory contains 76 processed data sub-directories each corresponding to one day's worth of processed data with names *1987078*/ - *1987248*/. These sub-directories in turn contain local databases with only one day's worth of processed data. The waveform reference tables (**wfdisc**) are stored in files with names *1987078.wfdisc*, for example, within these day sub-directories. Also in every day sub-directory is another sub-directory, *sac*/, which contains all of the processed waveform segment SAC files for that day. The processed waveform segment SAC file names are coded in a manner similar to that used by the program **rdseed**, provided by the

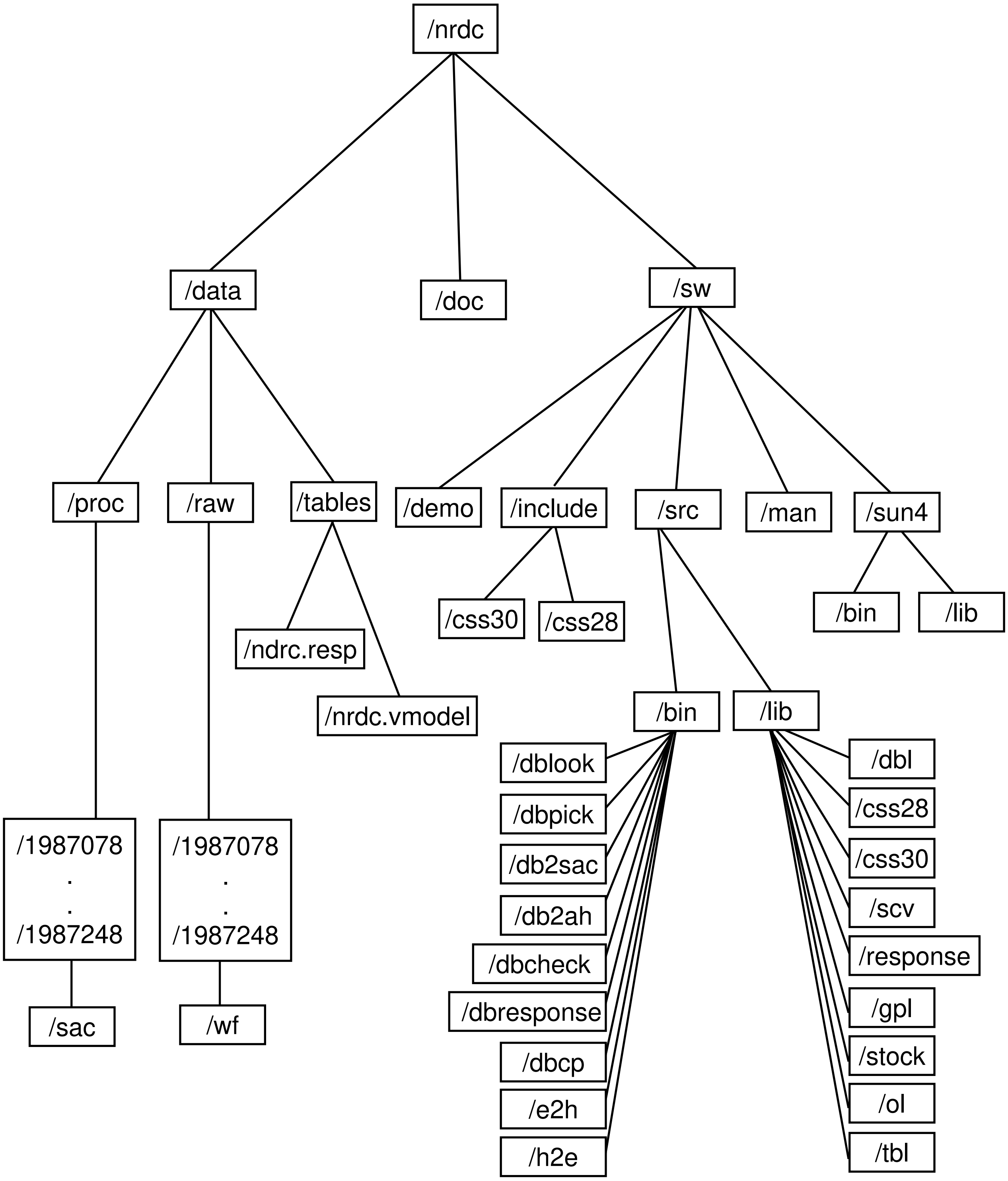


Figure 10. NRDC information product distribution directory tree.

IRIS DMC. A concatenated version of the single day processed databases is contained in *nrdc/data/proc/* and the waveform reference table is named *proc.wfdisc*.

nrdc/data/tables/

This sub-directory contains the 12 global database tables along with sub-directories *nrdc.resp*, which contains 58 instrument response files, and *nrdc.vmodel*, which contains the local Kazakh velocity model and travel time functions in file *kazakh1*. The 12 global database tables are contained in the following files.

nrdc.affiliation - **affiliation** table
nrdc.arrival - **arrival** table
nrdc.assoc - **assoc** table
nrdc.centryd - **centryd** table
nrdc.instrument - **instrument** table
nrdc.moment - **moment** table
nrdc.network - **network** table
nrdc.origin - **origin** table
nrdc.sensor - **sensor** table
nrdc.site - **site** table
nrdc.sitechan - **sitechan** table
nrdc.sregion - **sregion** table

See section 8 for descriptions of these tables.

nrdc/sw/

This sub-directory contains the data sub-directories *src/*, *include/*, *sun4/*, *man/* and *demo/* which contain software source code, man pages, executable programs and libraries, include files and demonstration programs and scripts. All binary programs and libraries are statically linked SUN-Sparc versions.

8. Database Layout

The database schema that we use is derived from the CSS version 3.0 schema as defined in Anderson, et. al. (1990). Our schema is, for the most part, a direct sub-set of the official CSS schema with a few extensions that we describe below. Figure 11 is a schema map which shows the tables and the links between the tables and the *Schema Reference Manual* gives a detailed description of all tables (*relations*) and columns (*attributes*).

8.1 Database Layout - Tables

affiliation	This table provides a means for clustering seismic stations into network or array groups by defining network-station affiliations.
arrival	This table contains information relating to seismic phases as observed at a single station-channel. Each row in the table corresponds to a single seismic phase arrival at a single station-channel and contains arrival time and phase code information along with other associated information, such as first motion, quality, signal to noise ratio, observed azimuth and incidence angles (from polarization analysis) and estimated signal type (e.g. local, teleseismic, data glitch). At present we are only filling in the arrival time and phase code information. We plan on providing the other information in future information product releases. Each row in the **arrival** table is given a unique arrival id (**arid**) which is used in the **assoc** table to reference arrivals.
assoc	This table contains event-arrival association information. Event hypocentral information is stored in the **origin** table and the **assoc** table has information that connects arrivals (i.e. rows in the **arrival** table) to a particular event (i.e. row in the **origin** table). Predicted phase arrivals from an event are associated with observed phase arrivals by creating rows in this table. In addition to the linkage information, the **assoc** table contains station to event distance, station to event azimuth, event to station azimuth, travel time residual and the velocity model used to compute the predicted travel times.

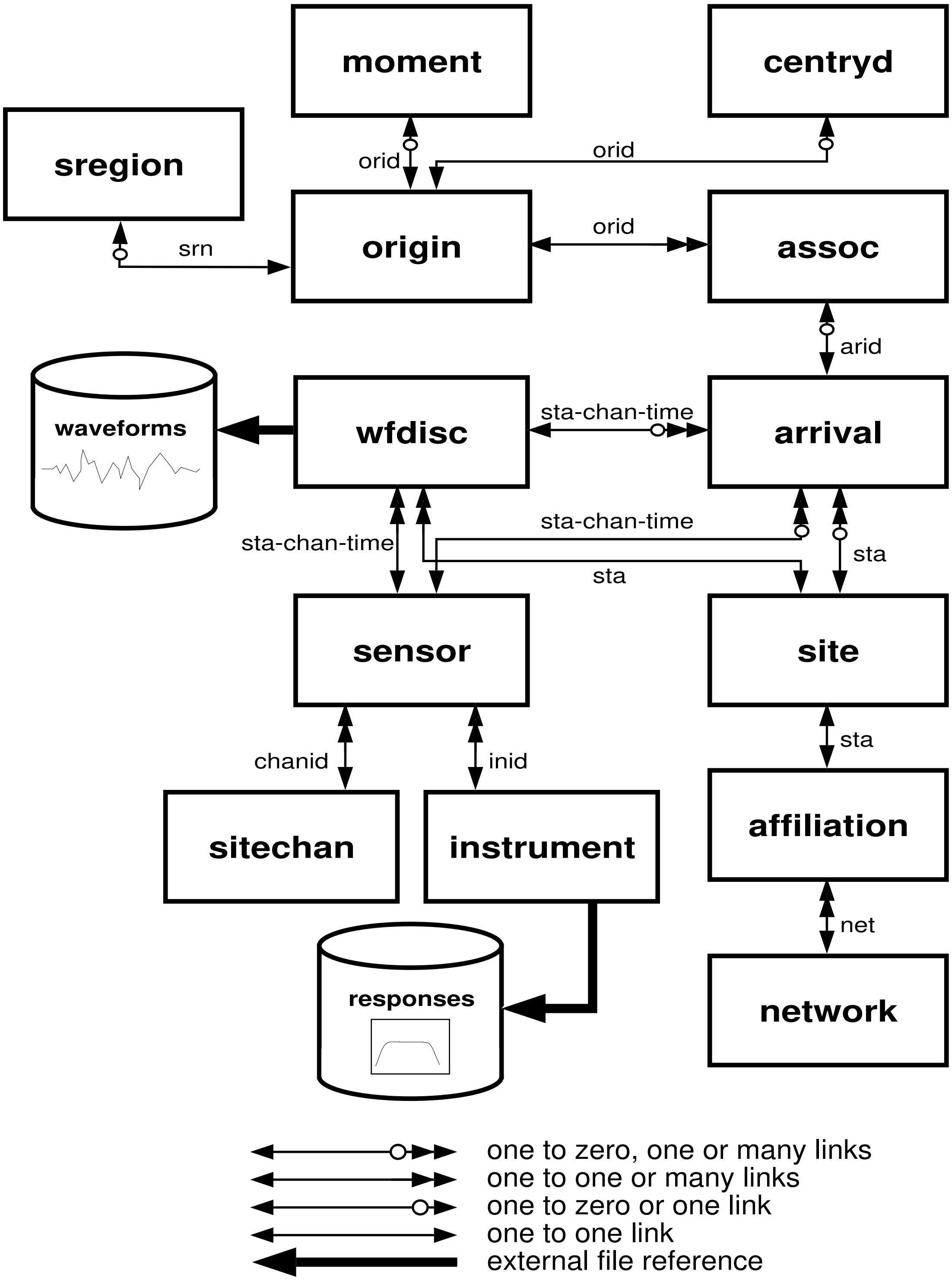


Figure 11. Database schema map for JSPC information products

centryd	This table contains information on the centroid locations of seismic events. This is not part of the normal CSS v. 3.0 schema.
instrument	This table references detailed instrument response functions. Each unique instrument response is stored in an external ASCII file which contains poles and zeroes, FIR coefficients and/or frequency-amplitude-phase tables. Either nominal instrument responses or those determined by direct calibration can be used. Multiple instrument stages are also allowed. The **instrument** table contains pointers to the response files along with instrument names, descriptions, nominal sampling rates and a unique instrument id (**inid**) which is used in the **sensor** table to reference instrument responses.
moment	This table describes the moment tensors for seismic events. This is not part of the normal CSS v. 3.0 schema.
network	This table gives general descriptive information for networks and arrays including names and types (e.g. array, local area, world wide).
origin	This table provides a summary of hypocentral information for seismic events. Either derived origins from local location efforts or reported origins from published bulletins can be used. For this information product we are only using reported origins, however for future products we will also include derived origins. In addition to latitude, longitude, depth, time and body, surface, local magnitudes, the **origin** table also includes Flinn-Engdahl seismic region number, author (e.g. PDE) and algorithm used (e.g. HYPOCENTER1). Every event origin is given a unique origin id (**orid**) which is used in the **assoc**, **centryd** and **moment** tables to reference origins.
sensor	This table provides fine scale time varying instrument response, time and gain calibration adjustment and channel orientation and depth information for each station-channel. The time and gain calibration adjustments are given directly in the **sensor** table and the instrument response and channel orientation and depth information are referenced to the **sitechan** and **instrument** tables through the **chanid** and **inid** ids.
site	This table gives station code, latitude, longitude and elevation. We consider a seismic station to define a single latitude-longitude coordinate on the earth.
sitechan	This table contains time varying geometric station-channel information including sensor orientation angles, sensor emplacement depth and a description of the channel. Every row in the **sitechan** table is given a unique sitechan id (**chanid**) which is used in the **arrival**, **sensor** and **wfdisc** tables to reference the station-channel characteristics.
sregion	This table contains the Flinn-Engdahl seismic region names (see Flinn, et. al. (1974)). Every row in the **sregion** table is given a unique sregion id (**srn**) which is used in the **origin** table to reference the Flinn-Engdahl seismic region names.
wfdisc	This is the most important and central table in the database schema and it acts as a hub for many queries into the database. This table references seismic waveform segments that are stored in external disc files. Station code, channel code, timing and gain information are given in the **wfdisc** table along with external data format description, external file name and byte offset information which provide a complete reference to the external data file containing the waveform samples. Each row in the **wfdisc** table corresponds to a single continuous waveform segment. Waveform segmentation can either occur as a natural result of the data acquisition system, such is the case with this information product, or can be arbitrarily done to continuous data streams for pragmatic reasons. The seismic samples in the external data files are assumed to run continuously for a given waveform segment and the current software supports two byte and four byte binary integer (SUN byte order) and four byte binary IEEE-floating point formats. Since a byte offset can be specified for each waveform segment, fixed header data files, such as SAC and ah, can be used as external waveform files within this schema with no modifications. The schema places no restrictions on how waveform segments are distributed across files, except that the data for a single segment must be continuously written within the file so that file formats that store the data in a fragmented fashion, such as SEED, cannot be used directly.

8.2 Database Layout - Linkages

Each database table is related to one or more other tables through linkage keys contained within the tables, as can be seen in figure 11. The simplest keys are the integer ids, **arid**, **orid**, **chanid**, **inid** and **srn**. These are locally defined integer values which are only used for referencing rows within a particular database, i.e. they generally have no global meaning. There are other linkage keys, such as **sta** (station code) and **net** (network code), which do have global meaning and significance external to the database. All of these linkage keys are simple keys, however there are also complex linkage keys, such as **sta-chan-time** (station code, channel code, absolute time), which consist of multiple columns in the database tables.

There are a number of different linkages that can exist between the rows of two tables as can be seen in figure 11. The simplest linkage is a *one to one* link in which a single row in table A is linked to a single row in table B. Such a link exists between the **site** and **affiliation** tables where each site has a single network affiliation. A variant on the *one to one* link exists when a single row in table A is linked to zero or one row in table B which can be seen in the linkage between the **arrival** and **assoc** tables where a single arrival may be unassociated or associated to a single event. The most common linkage is the *one to many* link in which a a single row in table A is linked to one or more rows in table B. Such a link exists between the **site** table and the **wfdisc** table where many waveform segments are recorded at the same station.

The power of a relational database comes from the logical grouping of information into tables. This allows us, for instance, to list the station latitude and longitude coordinates one time in a single table instead of repeatedly within fixed data headers attached to the waveforms. In addition to avoiding the redundant data storage this makes it easy to maintain the accuracy of the information since we can go to one place to find or change the information. The process of fixing erroneous information is also eased. As we were compiling this information product we initially made a mistake in the latitude at station BAY which was not discovered until we started doing event associations, after making over 26,000 waveform files. We fixed the station latitude for the ˜8,000 raw waveform files by changing one line in the **site** table. However, we were required to remake the ˜2,500 processed waveform data files since we were storing them in SAC form where the station coordinates are stored in the file headers.

The difficulties of using a relational database comes from the logical grouping of information into tables. Since the information has been partitioned from a conceptual "super" matrix into a number of more compact submatrices, which are the database tables (this process is known as *database normalization*), linkages between rows in the tables are required in order to retrieve associated pieces of information. Thus, if we want to know the station latitude and longitude for a particular waveform segment, we must effect a linkage between the **wfdisc** table and the **site** table, which is known as *joining* the two tables. It is the *joins* that are usually required to make effective use of a relational database, which are normally accomplished by the use of a proprietary RDBMS, that have discouraged the widespread use of relational databases within the scientific community.

We have gotten around this major difficulty by providing a light weight, yet effective, set of non-proprietary software tools that can be used to access the database tables directly, to automatically write data into SAC and AH forms, to view waveform data and to interactively pick arrivals, and to implement some simple fixed table *joins* which are commonly needed by scientific users. This software will be discussed in more detail in section 9.

8.3 Database Layout - Internal Conventions

The precise definitions of the table columns (*attributes*) are given in the *Schema Reference Manual* including units, data ranges and null values. For the most part the data values within the tables are self explanatory, however we list below some of the more important conventions used in the database.

units Consistent units are used for data types within the various tables. Distance units are in km, so that columns like **depth** (source depth) and **edepth** (receiver depth) are both expressed in km. All angles are expressed in decimal degrees. Time and amplitude units are described in more detail below.

time Times are expressed as seconds and dates are expressed as Julian dates (e.g. 1987078 = day 78 of year 1987). Absolute times are also in seconds and are called epoch times within the database tables. The epoch time is the *nominal* number of seconds since 1970 Jan 1 00:00:00 GMT. By nominal we mean that leap seconds have not been included, although leap days have. Epoch times are internally represented as double precision floating point numbers which gives sub-millisecond accuracy. Note that an epoch time can be positive or negative, to account for times prior to 1970 Jan 1. We provide a range of software utilities for converting from epoch to "human" time and vice versa. For the most part, however, the epoch time is hidden from the user and most input and output time values are automatically converted to/from epoch time. The epoch to human time conversion is completely reversible and is as accurate as using normal human time as the internal convention.

waveform amplitudes

The units of the actual waveform data samples within the external files can be anything. Waveform amplitude units are standardized through the **calib** and **calper** columns within the **wfdisc** table. By definition, **calib** is the amplitude factor, in nm (nanometers) per external amplitude unit, that is valid at **calper** period for that waveform segment. The external waveform data file can be read and the sample values can be multiplied by **calib** to give ground displacement units in nm. Since most seismometers are velocity instead of displacement sensors, this conversion to nm of ground displacement is only valid at **calper** period. We computed **calib** by first computing the total instrument gain in nm/sec per count and then we divided by the calibration angular frequency ($2\pi/$**calper**), which is set to 10 Hz for all waveform segments. Users can easily convert the ground displacement gain back to ground velocity gain, which will be valid over a much broader frequency range, by multiplying the scaled data values by $2\pi/$**calper**. This was done with all of the processed data stored in the SAC files, so that all of these data have amplitude units of nm/sec.

instrument responses

All instrument responses are in terms of instrument output per ground displacement. The instrument response files in directory *ndrc/tables/ndrc.resp/* are ASCII files containing poles and zeroes along with comment lines that define the instrument type and its characteristics in terms of cutoff frequencies, natural periods and damping ratios. These instrument response files are referenced by the **instrument** table and their contents are automatically included as part of the waveform segments in the software libraries. Instrument responses are automatically normalized at **calper** period so that the total system response function, in units of output amplitude unit per nm, can be computed by dividing the normalized response amplitude by **calib**.

8.4 Representation of the Database

The relational database contained in this information product is represented as a set of ASCII fixed-format files. Each table is contained in one or more files and table file names are always composed as *databasename.tablename*. Although we think of this information product as containing a single database, with the name **nrdc**, we have split the waveforms into 153 smaller databases each consisting of one day’s worth of raw or processed data. In addition we concatenated all processed data into a single database with the name **proc**. This division was done to improve overall performance and to make it easy for users to install arbitrary sub-sets of the complete data volume. Note that all tables, other than the **wfdisc** tables, are contained in single files with the database name **nrdc** in the distribution directory *nrdc/data/tables/*. It is only the overall **wfdisc** table that has been split into more manageable pieces. A mechanism is provided with the software for logically grouping together tables with different database names.

The ASCII database file formats are precisely defined in the *Schema Reference Manual*. All of the software included in this information product adheres strictly to this standard. In particular, the software in this distribution checks all candidate database files for proper line lengths as a quality control measure and will reject files which do not adhere to the fixed-format definitions. Users who wish to modify existing tables or to create new ones should keep this in mind. It is easy to mis-align fields when tables are hand edited which will result in a file that will not be recognized as a database table. We encourage users to utilize the various programs and utilities that we provide with this distribution for changing and adding to the database tables.

9. Software Tools

The software contained in this information product consists of programs, libraries and examples both in source and executable form. In addition, UNIX style manual pages are provided. The major thrust of this software is to provide a comprehensive and robust interface to the relational database. Most of this software has been in use at the University of Colorado for the last three years and the programs have been fairly well tested. However, there is a fundamental difference between locally developed and used software and software that is distributed and used by the community at large. We have attempted to provide the level of documentation and quality assurance that is normally associated with public distributions, however we cannot insure that there will not be problems with the software or functional omissions. As with the data itself, we encourage users of this information product to let us know about software related difficulties they have encountered and we will do our best to fix problems in future releases.

We have concentrated on providing a strong read interface to the database, which includes database viewing programs, format conversion programs and programmatic interfaces for joining tables and for accessing any of the information in the database. We assume that most users will either convert data to their own favorite form for further processing using familiar seismic processing packages, or modify their existing programs to access the data directly and thus we have not supplied comprehensive data processing or analysis software. We have also put minimal emphasis on providing software for writing to the database.

9.1 Software Tools - Programs

The programs fall within three categories; database viewing programs, database conversion programs and database utility programs. All programs which take a relational database as input begin with **db**. A list of the programs with brief descriptions is given below.

dblook	This program is a read only spreadsheet-like program which provides viewing of any of the tables within a database. It is mouse driven and allows efficient scrolling through tables with large numbers of rows and rearranging of table columns. Each table is represented by a separate window. Clicking the mouse within rows in the **wfdisc** table causes additional windows to appear with the waveform segments displayed as trace plots. This program uses SUN's XView graphics library.
dbpick	This is an interactive waveform viewing and phase arrival picking/editing program. Multiple stations and channels can be displayed within the same window. Subwindows can be spawned with different arrangements of traces and magnify subwindows can be made for arrival picking. Phase arrivals can be added or deleted. The time and phase code of existing arrivals can be edited. Time scrolling/zooming and arrival editing are done with the mouse. On-the-fly filters can be applied to any traces. A variety of amplitude scaling options exist. In addition to a mouse driven user interface, a type-in command driven user interface can be used. Keyboard accelerators can be used to enable/disable automatic data deglitching or to jump scroll to the next or previous waveform segment or phase arrival. Waveform segments can either be read into memory at start time or can be directly accessed through disc files every time they are needed. This program uses a home brewed, Xlib based graphics library.
dbresponse	This program will print out instrument responses for specified station, channel and time values. Either the contents of the ASCII instrument response file are printed or a table of frequency-amplitude-phase values are computed and printed.
db2sac	This program will convert data from a database to SAC form.
db2ah	This program will convert data from a database to ah form.
dbcheck	This program will check the integrity of the tables in a database. Files which do not appear to conform to the fixed-format standards will fail the check and information

	identifying the file and line number are printed.
dbcp	This is a utility program for copying a database along with all of its referenced waveform segments.
e2h	This is a utility program for converting epoch time to human time.
h2e	This is a utility program for converting human time to epoch time.

9.2 Software Tools - Libraries

The software libraries provided with this information product fall into three categories; database management tools, interactive graphics tools, and general utility tools. The graphics and utilities libraries are mainly provided so that the programs can be re-compiled and linked by the user and, although some documentation is included, we have not made special efforts to generalize and augment the documentation, or the software itself, for the general user community. Our emphasis is on the database management libraries which provide both C and FORTRAN programmatic interfaces to the relational databases. We encourage the users of this information product to examine the demonstration programs given in *ndrc/sw/demo/*, which show how these libraries can be used, and we hope that these database management tools will ease the job of accessing database information from locally developed programs. following is a list of the libraries provided with this information product release.

dbl	The **dbl** library provides basic access to database tables (*relations*) and their data columns (*attributes*). Both read and write access are implemented. Table rows can either be kept on disc file or read into memory and accessed from there. Fetching and putting routines are provided on both the row and column level. Both C and FORTRAN interfaces are provided. In addition to the generic database table routines, a number of database utility routines are provided including time conversion routines. This library provides generic support for accessing and manipulating relational databases stored in fixed-format ASCII files. Particular definitions of the schema are not a part of this library. The **dbl** library must be used with a *schema description* library.
css30	This is a schema description library that defines our version of the CSS v. 3.0 relational database schema. Both C and FORTRAN interfaces are provided. This library, along with the **dbl** library, provide complete access to a CSS version 3.0 relational database stored as a set of fixed-formatted ASCII files. The format definitions for these files are given in the *Schema Reference Manual*. Databases are created, opened and read with calls from the **css30** library. Other **css30** calls allow the programmer to read and write waveform samples to/from external files. All other calls are made with the **dbl** library. The **css30** library must be used in conjunction with the **dbl** library.
css28	This is a schema description library that defines our version of the CSS v. 2.8 relational database schema. We provide this to support the programs **dblook** and **dbpick** which can use either schema.
scv	The **scv** library provides a high level schema-format independent interface to seismic data by creating and maintaining data abstractions called *station-channel views*. A single station-channel view consists of all waveform segments and phase arrivals associated with a station-channel over an arbitrary time period along with all station, channel and instrument parameters including station latitude, longitude and elevation, sensor orientation angles, sensor emplacement depth, total system gain and detailed instrument response function. We assume that a *station* corresponds to a single latitude, longitude, elevation coordinate that does not change with time. All other channel and instrument parameters can be time varying within a station-channel view. The **scv** library acts as a connector between the station-channel view abstraction and a particular external data format. Presently this library supports the fixed-format ASCII flat file CSS v. 3.0 and 2.8 relational database formats. We are in the process of developing **scv** connections to raw SEED volumes and to SAC files. Regardless of the particular connections provided, the user interface to the **scv** library is the same, which allows for schema-format independent programming with seismic data. Both C

	and FORTRAN interfaces are provided. At present the **scv** library must be used with the **css30**, **css28**, **response** and **dbl** libraries.
response	This library provides utility procedures for reading the external instrument response files and for computing complex instrument responses as a function of station, channel, time and frequency.
gpl	The **gpl**, or gather plot, library provides the X-window based interactive trace plotting functionality used in programs **dblook** and **dbpick**.
ol	This library is a simple home-brewed X-window widget library that is used by the **gpl** library.
tbl	This library provides a simple XView based table spreadsheet window and is used by program **dblook**.
stock	This library provides general utility functions such as parsing argument lists and dealing with errors.

The **dbl**, **css30** and **scv** libraries provide a fairly comprehensive interface to the relational database contained in this information product. The **dbl** and **css30** libraries can be used to read or write any of the information from the database tables. These libraries do not, however, provide the type of *joining* functionality that is needed to put the various tables together into a custom *view*. The **scv** library provides *joining* functionality by composing a fixed view which is intended to satisfy most user queries that might be made about a single channel of seismic data. The station-channel view abstraction was created primarily from a research seismologist's standpoint and will hopefully be useful for those who wish to access the database directly without mastering the details of the schema. At present, the **scv** library interface to the database is mainly a read interface.

9.3 Software Conventions

Following is a list of conventions, restrictions and miscellaneous information relating to the software.

1. Although we have tested the installation scripts on other UNIX platforms, the programs and libraries have only been made and tested on SUN-Sparc platforms using SUNOS v. 4.1.1. With the exceptions listed below, the software should be portable to other UNIX environments, however we have not tried such ports.

2. All of the graphics software was developed in the X-window environment. We have tested the programs with SUN's OpenWindows v. 2.0/3.0 servers and with the generic MIT X11R5 server. **dblook** uses SUN's XView library and **dbpick** uses the base level Xlib. The binary programs provided with this release have been statically linked so that they should run on any SUN Sparc station regardless of the existence of the XView and Xlib libraries. If these programs are remade, then the XView and Xlib libraries, along with their associated include files, must be made available.

3. When running the program **dbpick**, keyboard accelerators can be used to jump scroll the display to the next or previous arrival or waveform segment. In order for these to work properly when using the OpenLook window manager (**olwm**), it is necessary to include the following line in your *.Xdefaults* file in your home directory.

   ```
   OpenWindows.FocusLenience:      True
   ```

   If this line is not already in the *.Xdefaults* file, then it should be added and the window manager should be restarted before using **dbpick**.

4. There is a single UNIX environment variable, **JSPC_DBPATH**, which is critical when using any of the database software, either the programs or the libraries. This must be set to a MANPAGE style directory search path which specifies where the database table files can be found. This would typically be set to something like *.:/install_dir/nrdc/data/tables/nrdc* where */install_dir* is the absolute path to where the information product was installed. When a database name *dbname*

is specified in a program, the database library procedures automatically search through the directories specified by **JSPC_DBPATH** for files of name *dbname.tablename* or *dbpath.tablename*, where *dbpath* is one of the paths specified in **JSPC_DBPATH**. This software feature allows us to put most of the tables into a single place and to merge them with local tables at run time. Thus, if one were to use the example given above for setting **JSPC_DBPATH**, go into directory /*install_dir*/*nrdc*/*data*/*proc* and run program **dblook** by typing *dblook proc*, not only the *proc.wfdisc* file would be displayed, but also all of the *nrdc.tablename* files in directory /*install_dir*/*nrdc*/*data*/*tables* would be displayed as well.

10. References

Following is a list of references relating to this information product. We have tried to make this list comprehensive, including both the references mentioned in this report and other references describing research that has been done with these data. We have concentrated on listing published reports although we have many more manuscripts that are not included in this list that, to our knowledge, were never published. Please let us know of any published research that is missing from our list and we apologize in advance for unintentional omissions.

Anderson, J., Farrell, W., Garcia, K., Given, J., Swanger, H., Center for seismic studies version 3 database: schema reference manual, *CSS Technical Report C90-01*, 1990.

Antonenko, A., Deep structure of the earth's crust in Kazakhstan, *Academy of Sciences of the Kazakh SSR,* Alma-Ata, 274, 1984.

Berger, J., et al., A new U.S.-U.S.S.R. seismological program, *Eos Trans. AGU, 68*, 110-111, 1987.

Berger, J., Eissler, H., Vernon, F., Nersesov, I., Gokhberg, M., Stolrov, O., Tarassov, N., Studies of high-frequency seismic noise in Eastern Kazakhstan, *Bull. Seismol. Soc. Am., 78*, 1744-1758, 1988.

Eissler, H., Berger, J., Vernon, F., Gomberg, J., Chavez, D., Tarasoz, N., Zhuravlev, Observations of chemical explosions on the steppes of Central Asia, *Eos Trans. AGU, 68*, 1363-1364, 1987.

Flinn, Engdahl, Hill, *Bull. Seismol. Soc. Am., 64*, 791-992,1974.

Given, H., Tarasov, N., Zhuravlev, V., Vernon, F., Berger, J., Nersesov, I., High-frequency seismic observations in Eastern Kazakhstan, USSR, with emphasis on chemical explosion experiments, *J. Geophys. Res., 95*, 295-307, 1990.

Gomberg, J., Bodin, P., Martinov, V., Seismic system calibration using cross-spectral methods from *in situ* measurements: the Kazakh, USSR, Phase I array, *Bull. Seismol. Soc. Am., 78*, 1380-1386, 1988.

Gurrola, H., Minster, J., Given, H., Analysis of high-frequency seismic noise in the western United States and eastern Kazakhstan, *Bull. Seismol. Soc. Am., 80*, 951, 1990.

Hedlin, M., Minster, J., Orcutt, J., The time-frequency characteristics of quarry blasts and calibration explosions recorded in Kazakhstan, U.S.S.R., *Geophys. J. Int., 99*, 109-121, 1989.

Kim, W., Simpson, D., Richards, P., High-frequency spectra of regional phases from earthquakes and chemical explosions, from *Chemical Explosions and the Discrimination Problem, Phillips Laboratory Final Report PL-TR-91-2285*, 43-96, 1991.

Leith, W., Geology of NRDC seismic station sites in Eastern Kazakhstan, USSR, *U.S. Geol. Surv. Open File Rep.*, 87-597, 1987.

Li, Y., Thurber, C., Hypocenter constraint with regional seismic data: A theoretical analysis for the Natural Resources Defense Council network in Kazakhstan, USSR, *J. Geophys. Res., 96*, 10,159-10,176, 1991.

Priestley, K., Zandt, G., Randall, G., Crustal structure in Eastern Kazakhstan, U.S.S.R., *Geophys. Res. Lett., 15*, 613-616, 1988.

Thurber, C., Eissler, H., Berger, J., Zhuravlev, V., Tarasov, N., Location of regional explosions and earthquakes with the NRDC-Soviet Academy of Sciences seismic network in Kazakhstan, *Eos Trans. AGU, 69*, 1332, 1988.

Thurber, C., Given, H., Berger, J., Regional seismic event location with a sparse network: application to Eastern Kazakhstan, USSR, *J. Geophys. Res.*, *94*, 17,767-17,780, 1989.