The Research Council of Norway (NFR)

NORSAR

NORSAR Scientific Report No. 2-97/98

Semiannual Technical Summary

1 October 1997 – 31 Mars 1998

Kjeller, May 1998

7.2 Threshold Monitoring: Summary of pipeline processing

Introduction

The Threshold Monitoring software and operations manual have been completed and are in use at the International Data Center (IDC) in Arlington, Virginia. This report is a summary of the pipeline processing discussed in the manual, which describes the TM system.

The Threshold Monitoring (TM) system is intended for continuous assessment of the detection capability of the International Monitoring System's Primary Seismic Network, in support of Comprehensive Nuclear Test Ban Treaty. It accomodates temporary problems which traditional methods based on statistical models may not account for. This includes background noise fluctuations, data quality variations, processing deficiencies, and unrelated seismic events (*e.g.*, large earthquakes).

Software and data files

The TM software resides in a directory structure which also contains the static data files required for processing, such as target lists, beamforming recipe files, *etc.* Scripts for defining TM environment variables are included. The directory structure and the files therein are thoroughly explained in the Operations Manual.

Although some of the software consists of Bourne shell scripts, most of the programs are written in C, with C and FORTRAN subroutines. Arguments for the C programs can be stored in a parameter file rather than entered on the command line. This system makes it easier to run the software.

The input and output files used by TM are binary files which are organized with respect to a reference time T_r. Data observed at time T will begin at the file position corresponding to the remainder of $(T - T_r)/L$, where L is the file size in seconds. Since data will wrap around from the end of the file back to the beginning, these files are referred to as *disk loops*. Raw data are stored in disk loops which are large enough to hold seven days' worth of data. These disk loops are updated continuously.

TM processing

TM produces statistics for every hour of data. The sequence of events is shown below, with program names in bold face.

- **CreateTMSession**: creates the working directory and initializes files.
- **DFX** (Detection and Feature Extraction, Wahl 1996a,b): generates STA (short term average) traces from the raw data.
- **TMthreshold**: calculates thresholds for predetermined targets from the STA data..
- **TMmap**: generates a single disk loop containing merged, resampled threshold data for plotting.
- **TMprod**: generates hourly plots showing station availability, STA data, and worldwide thresholds.
- **TMbulletin**: reads Reviewed Event Bulletin.

- **replotuptime**: adds seismic event information to the station availability plot.

Before TM processing commences, a working directory structure with initialized output files must be created with CreateTMSession. This is done once. Processing is performed continuously in the so-called Alpha and Delta pipelines (see the flowchart in Fig. 7.2.1).

Quality control, beamforming, bandpass filtering, and short-term-average calculations are performed by DFX for each station. These "STA" data are written to new disk loops. DFX processes each ten minute segment of raw data as soon as it becomes available. It runs in the Alpha pipeline, as described in the International Data Center Operations Manual (CTBT/PC/V/WGB/TL/44/Rev.2, 1998). The remaining steps are performed in the Delta pipeline, currently ten hours behind real time.

Network detection thresholds for each of 2562 targets distributed around the globe are calculated by TMthreshold and written to a third set of disk loops. These threshold data are interpolated and resampled by TMmap, which writes the results to a final disk loop. This disk loop is organized with respect to time and is used to generate the threshold maps described below.

The following statistics are generated by TMprod for each hour for which data are available:

- Map showing the location and percent availability of each station (see Fig. 7.2.2).
- Plots of STA traces for each station in the primary network, allowing the user to see the background noise levels, observed signals, and processing gaps (see Fig. 7.2.3).
- Maps showing the average and worst-case detection thresholds for the world (see Fig. 7.2.4).

Finally, TMbulletin reads the Reviewed Event Bulletin, and this information on interfering seismic events can then be included on the station availability map by the script replotuptime.

L. Taylor

References

CTBT/PC/V/WGB/TL/44/Rev.2 (1998): Initial Draft of the Operational Manual for the International Data Centre. Preparatory Commision for the Comprehensive Nuclear-Test-Ban Organization, Vienna, 19-30 January 1998.

Wahl, D. (1996a): User's Manual for the Detection and Feature Extraction program (DFX). SAIC-96/1098.

Wahl, D. (1996b): Programmer's Guide for the Detection and Feature Extraction program (DFX). SAIC-96/1069.

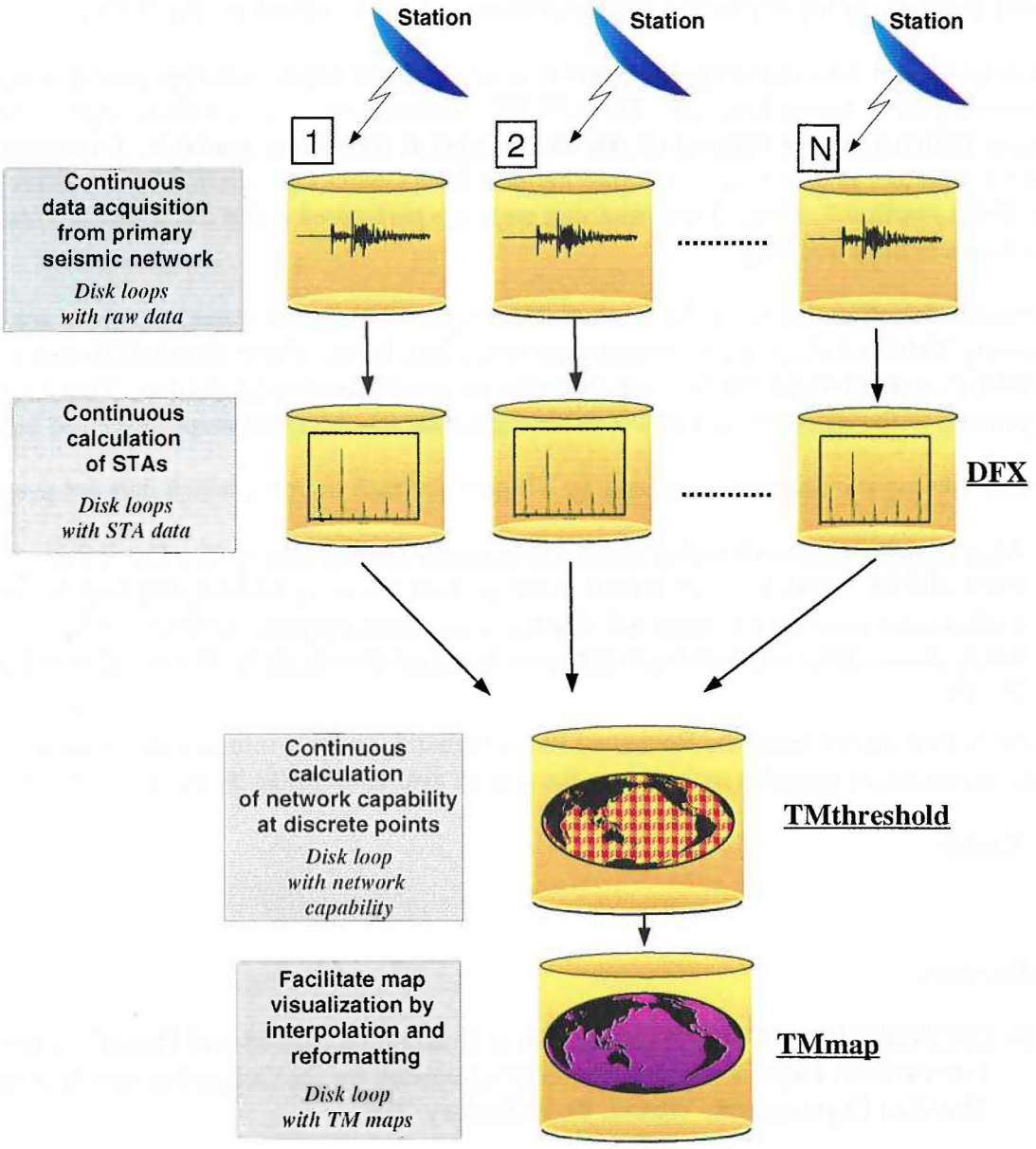
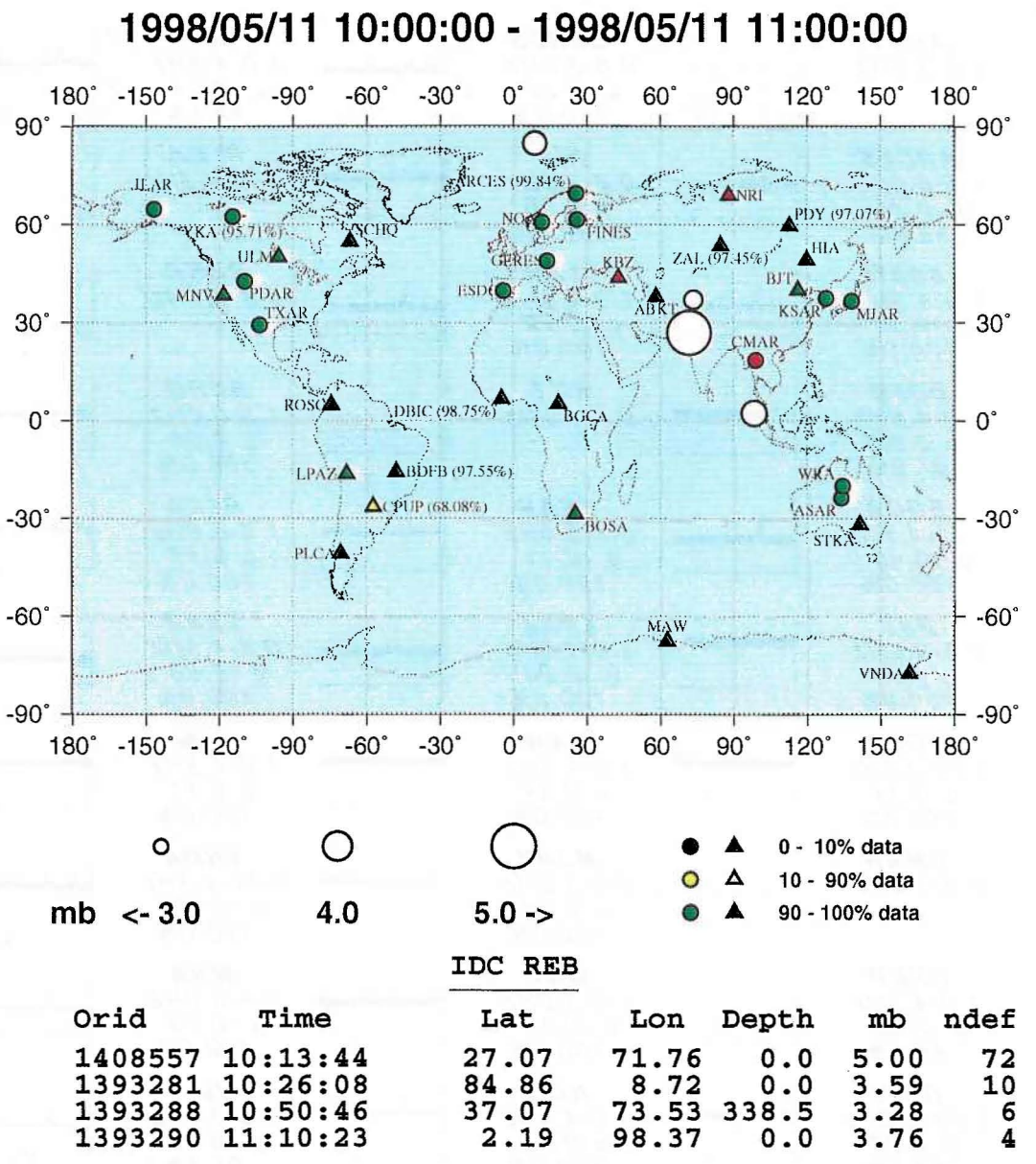


Fig. 7.2.1. Flowchart of processing within the TM system. Text in sans serif typeface describes each process, whereas text in italics describes the results and the type of storage. The names of the programs used at each step are underlined in the figure.



IDC REB

Orid	Time	Lat	Lon	Depth	mb	ndef
1408557	10:13:44	27.07	71.76	0.0	5.00	72
1393281	10:26:08	84.86	8.72	0.0	3.59	10
1393288	10:50:46	37.07	73.53	338.5	3.28	6
1393290	11:10:23	2.19	98.37	0.0	3.76	4

Fig. 7.2.2. Station availability map created by TMprod. The colors of the station symbols indicate the percent availability for detecting events occuring during the time period indicated (11 May 1998, between 10:00 and 11:00 GMT). Three component stations are marked by triangles, the circles represent arrays. When the Reviewed Event Bulletin is complete, the locations of seismic events occuring within or shortly after the given time interval are added to the map. The event with mb=5.0 represents the three simultaneous nuclear tests conducted in India at 10:13:44 GMT.

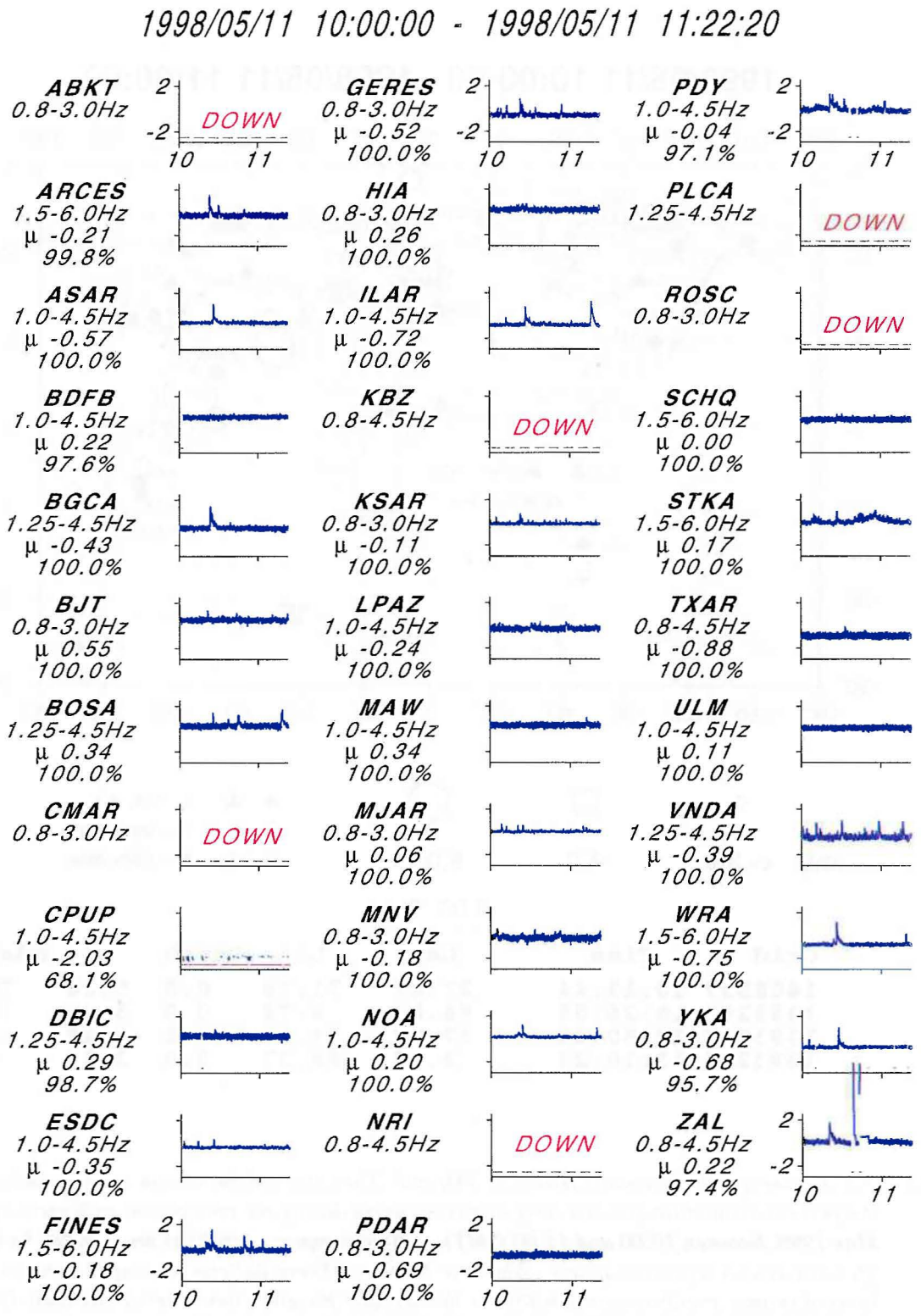


Fig. 7.2.3. Continuous log (A/T) equivalents derived from STA traces are shown in blue for each station in the primary network. Periods of down time are shown in red. The data interval is extended beyond the hour to allow for the travel times of events originating near the end of the hour. The station name, filter cutoffs, average values, and percent availability are shown next to each trace.

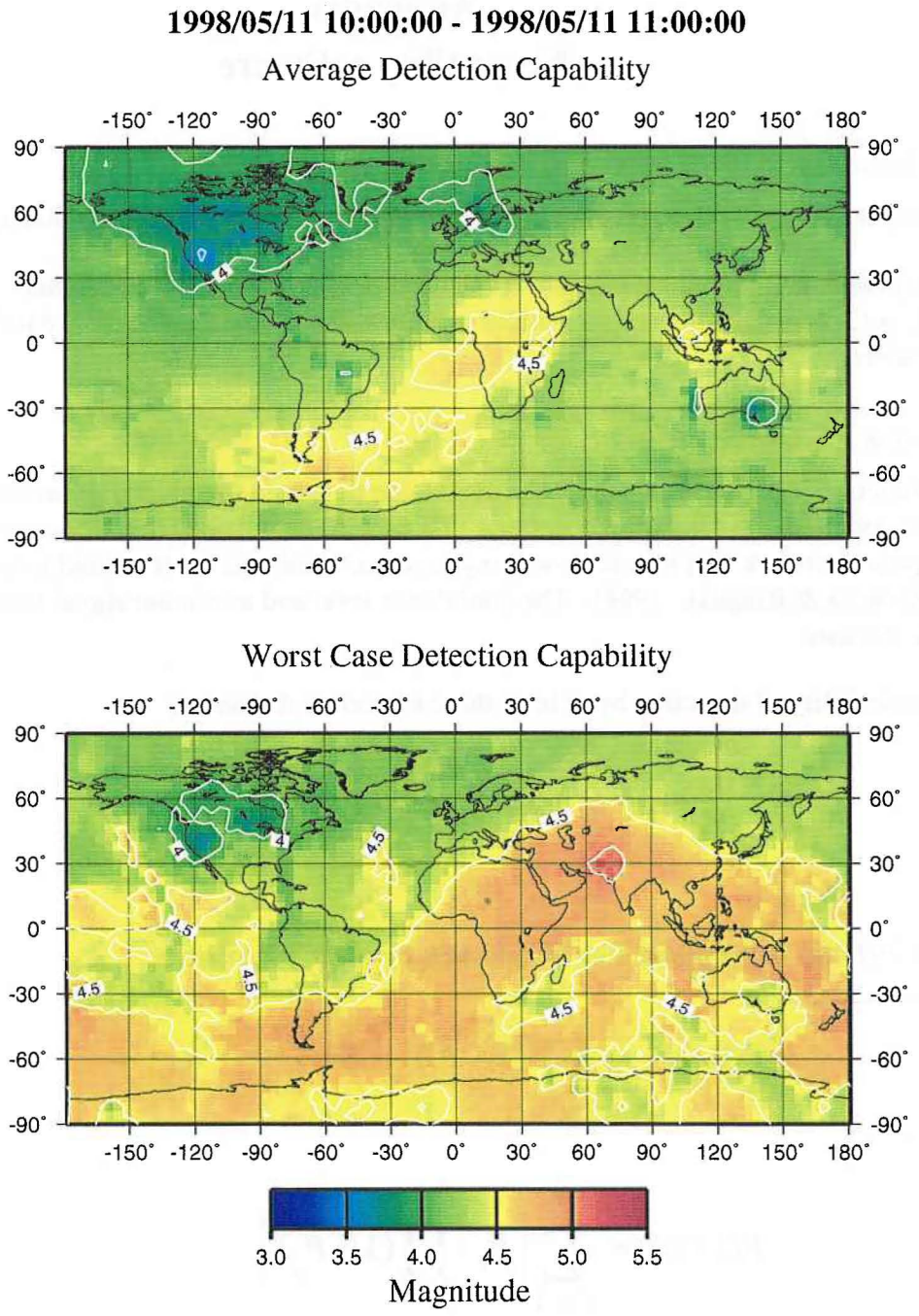


Fig. 7.2.4. Detection capability maps for the given one hour interval. The average capability (top) may vary from hour to hour depending on lengthy station outages, fluctuating background noise levels at different stations, and large long-duration seismic signals. The nuclear explosions in India temporarily lowered the capability over much of the world, as shown in the worst case map.

Appendix
New utility software

Introduction

This report discusses two software modules that have been written during this reporting period:

- **detprob3**, a C function designed to calculate network detection thresholds.
- **wf**, a C program which creates tables of information related to the IDC database waveform headers.

Detection probability

This function (detprob3) calculates the magnitude threshold for which it is assumed that events will be detected by at least three stations. Thresholds calculated for the observed noise levels in a global network will be a time-varying function similar to the threshold monitoring output (see Kværna & Ringdal, 1998). The confidence level and minimum signal to noise ratio are set by the user.

The probability of detection by at least three stations is defined as

$$F(\mu) = 1 - P(0/m) - P(1/m) - P(2/m)$$

where $P(k/m)$ is the probability of detection by exactly k stations:

$$P(0/m) = \prod_{i=1}^{N} (1 - P_i)$$

$$P(1/m) = \sum_{i=1}^{N} \left[P_i \cdot \prod_{j \neq i} (1 - P_j) \right]$$

$$P(2/m) = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \left[P_i \cdot P_j \cdot \prod_{k \neq i,\, j} (1 - P_k) \right]$$

The probability of detecting an event of magnitude μ at station i is given by the error function:

$$P_i = \Phi\left(\frac{\mu - m_i - t}{\sigma}\right)$$

where σ^2 is the assumed signal variance, t is the signal to noise level specified by the user, and the m_i are the observed magnitudes. The probability function F is calculated for μ ranging from $m_3 + lowlim$ through $m_3 + uplim$, where m_3 is the third smallest magnitude observed at any station, and the limits are set by the user. $F(\mu)$ is then linearly interpolated to the desired confidence level, and the corresponding magnitude is returned to the user.

Usage

The function requires two arrays containing the observed magnitudes and sigmas, as well as the signal to noise ratio, the confidence level, the lower and upper limits to be added to the test magnitude, and the step size.

A main program (magthresh.c) has been written which will read the magnitudes and sigmas from an ASCII file before calling detprob3. The first line of the file should contain the number of magnitudes; two columns should follow which contain the magnitudes and sigmas, respectively. Other inputs can be entered in a par file (default values are shown:

file=<datafile>

snrthresh=0.4

conflev=0.9

step=0.1

lowlim=-1.0

uplim=1.0

To execute: magthresh file=<filename> par=<parfile> tkfr verbose

There is an option to use $m_3 + t + 1.258 \cdot \sigma_3$ as the test magnitude by including "tkfr" in the calling sequence. If not present, m_3 is the default.

Examples

The default values shown above are used, and the test magnitude (using the "tkfr" option) $m_3 + 0.4 + 1.258 \cdot \sigma_3$ is 4.78 for each case. There are a total of 50 stations.

Case 1: Values for three stations equal 4.0 and the remaining 47 are effectively infinite. We get a magnitude threshold of 4.952.

Case 2: Values for eight stations equal 4.0, two are $-\infty$, and the remainder are ∞. The magnitude threshold is 4.203.

Case 3: Same as Case 2, but with only four stations having values of 4.0. The result is 4.359.

The difference relative to the test magnitude for these extreme cases ranges from -0.58 to +0.17 magnitude units.

Waveform headers

This software (wf) creates new versions of the IDC site, sitechan, sensor, and instrument tables. Information that is pertinent to a set of online waveforms is extracted from the ASCII versions of these tables. Response files listed in the instrument table will also be copied to the output directory.

These four tables contain the following information:

IDC Table	Contents
sensor	Calibration information for specific sensor channels.
site	Station location information.
sitechan	Channel information for each station.
instrument	Calibration information for each instrument.

The wfdisc (waveform header) table contains descriptive information about data which are stored on disk. Entries in the four tables discussed above are copied verbatim to new tables if the station and channel are found in the wfdisc table. The starting and ending times for each entry must overlap with those found in wfdisc.

The instrument table lists the paths for the response files. These are replaced with "." in the output file. Otherwise, the entries in the output files are identical to those in the original tables.

The arguments for wf may be entered in a parameter file:

infile=</path/name>
tabledir=</path/tablename>
outdir=<outputdirectory>

To execute: wf par=<parfile> verbose

/path/name (the wfdisc table) is read, and the program searches for matching entries in /path/tablename.site, /path/tablename.sensor, etc.

L. Taylor
F. Ringdal

References

Carter, J. A. & J. R. Bowman (1997): IDC Database Schema, Tech. Rep. CMR-97/28.

Kværna, T. & F. Ringdal, (1998): Seismic Threshold Monitoring for continuous assessment of global detection capability, *Semiannual Tech. Summ. 1 October 1997 - 31 March 1998,* NORSAR Sci. Rep. 2-97/98 , NORSAR, Kjeller, Norway.