The Research Council of Norway (NFR)

NORSAR

NORSAR Scientific Report No. 1-97/98

Semiannual Technical Summary

1 April – 30 September 1997

Kjeller, November 1997

7.7 Recommendations for improvements in the PIDC processing of Matsushiro (MJAR) array data

Introduction

The seismic arrays participating in the International Monitoring System (IMS) to monitor compliance with the Comprehensive Test Ban Treaty (CTBT) play an important role not only in detecting but also in defining and locating seismic events. In particular the measurements of azimuth (i.e. back azimuth, from the station to the epicenter) and ray parameter (or apparent velocity) of observed signals are used to name the onsets and to associate them to form events. Therefore all such measurements should be made as exact as possible to make them usable for all automatic data processing at the Prototype International Data Center (PIDC).

The estimates of the slowness vector of an observed seismic wave (i.e. azimuth and ray parameter) are often associated with errors. First of all the aperture and geometry of the seismic arrays limit the resolution and precision in measuring these parameters; generally spoken: the larger the aperture and the more array elements, the better the resolution capability of an array. These limits defined by the configuration of an array are additionally influenced by the actual background noise, which can distort the results, especially in the case of signals with low signal-to-noise ratios (SNR). One way to avoid (or better: to reduce) this problem is to concentrate the signal analysis in the frequency range with the best SNR. In addition, prefiltering of the data before fk-analysis (Schweitzer, 1994a) has been shown to be helpful to reduce the leakage of especially longer period noise into the frequency band analyzed.

Since the beginning of array seismology it has been well known that all arrays show more or less pronounced deviations of their measured azimuths and ray parameters from the theoretically estimated ones. Such deviations are commonly attributed to lateral heterogeneities of different size and structure beneath the arrays. Many studies have been undertaken to derive such deviations and to have a set of correction values available to obtain better estimates of the event location (e.g. for the arrays analyzed at the NORSAR data center the following studies can be mentioned: NORSAR (Berteussen, 1974; Fyen et al., 1995), ARCES, FINES, GERES, and NORES (Schweitzer, 1994b; Schweitzer, 1995; Schweitzer & Kværna, 1995), and SPITS (Schweitzer, 1994b)).

The results from array measurements of onsets observed for the Matsushiro array (MJAR) in Japan show a large scatter and are thus difficult to handle by the data processing at the PIDC. Therefore the contributions of MJAR to the Reviewed Event Bulletins (REBs) were investigated in more detail to obtain as reported in this contribution recommendations for the data processing at the PIDC.

Evaluation of PIDC phase measurements; comparison of MJAR with other arrays

The results provided in the PIDC Reviewed Event Bulletins (REBs) from the GSETT-3 experiment form the basis for the investigation of mislocation vectors at different arrays. In the time period January 1, 1995 to May 27, 1997, the REBs were searched for first P-arrivals corresponding to arrivals associated by the NEIC to events in its bulletins (monthly or weekly). The azimuth and slowness estimates given in the REBs were then compared to the theoretical val-

ues for P-phases from the NEIC event locations using the IASP91 model (Kennett & Engdahl, 1991) and the tau-spline software (Buland & Chapman, 1983). In order to reduce the influence of less accurate hypocenter estimates only events were considered which were located by the NEIC with observations from at least 20 stations.

To minimize the influence of erroneous phase associations and other obvious measurement problems, the following set of acceptance criteria was introduced:

- The travel-time residual of the first onset must be less than 4 seconds.
- The reported signal period used for magnitude estimation must be larger than 0.25 seconds for teleseismic onsets (phases with theoretical ray parameter less than 10 s/°).
- The length of the mislocation vector (slowness-error vector) must be less than 5 s/°.

If for a given array, there is a large percentage of phases falling outside the limits given above, this gives us a hint that there may be a problem with the data processing of this array at the PIDC.

To obtain an overview of the quality of the estimated azimuths and ray parameters, the mean azimuth errors, the mean ray-parameter errors and the corresponding standard deviations were calculated. Some of the arrays were operational only for a short time period during the time interval for which the REBs were searched, and thus provided few observations, which often showed a large scatter. For further analysis it was therefore required that at least 1000 observations were available. The results are listed in Table 7.7.1 and discussed in the following.

- Table 7.7.1 clearly shows that at some arrays specific problems exist. E.g., the percentage of P-arrivals with travel-time residuals exceeding 4 seconds is high (exceeding 4.9%) for CMAR, ESDC, GERES, KSAR, MJAR and TXAR. This can be caused both by structural heterogeneities beneath the arrays or by data processing problems at the PIDC.
- The relatively high percentage of high-frequency teleseismic onsets at ARCES, NORES and SPITS is clearly due to the data processing at the PIDC. It should be noticed that the frequency analyzed in this study is **not** the frequency providing the highest SNR, but the frequency associated with the amplitude used for estimation of m_b. The standard attenuation curves for calculating event magnitudes are not calibrated for amplitudes measured at such high frequencies.
- The arrays having a large number of arrivals with large mislocation vectors are generally those with a small aperture, as evidenced by the results for FINES, HFS, and SPITS in particular. The MJAR array, however, stands out as an exception to this pattern. This array has an aperture of about 10 km, but still has a large number (22%) of arrivals with large mislocation vectors. This is an indication that there are problems with the MJAR data processing at the PIDC.
- For the number of onsets remaining after outlier rejection (#U in Table 7.7.1), the mean onset-time residual, the mean azimuth residual, the mean ray-parameter residual and the mean slowness error were calculated for each station. The positive mean travel-time residuals for all stations in Table 7.7.1 are due to the use of NEIC source times (calculated with the Jeffreys-Bullen model (1940)) and prediction of the arrivals using the IASP91 model. In

general there exists a well-known base-line shift of about 2 seconds between the IASP91 and the Jeffreys-Bullen model, so from Table 7.7.1 it cannot be concluded that any of the arrays show anomalous travel-time behavior.

- Concerning the estimates of mean azimuth and ray-parameter residuals after rejecting the outliers, there is a good correlation with the array aperture, as the two largest arrays, WRA and YKA, show the smallest deviations from the theoretical values. The scatter in the azimuth and ray-parameter measurements is mostly due to systematic mislocations or noisy data. However, some arrays like PDAR, SPITS and TXAR show very large azimuth errors and/or scatter, which is believed to be caused by the influence of dipping structures beneath the arrays. After removal of the outliers, the azimuth measurements at MJAR show no specific trend. Concerning the mean ray-parameter residuals, only PDAR and SPITS stand out as clearly having a larger scatter than the other arrays. When comparing the mean length of the mislocation vectors (rightmost column of Table 7.7.1), MJAR is one of the arrays having the largest value.

In conclusion, MJAR shows an unexpectedly large amount of "bad" azimuth and slowness estimates which cannot be explained by systematic effects caused by heterogeneities beneath the array site only.

The MJAR data

To investigate the statistics reported above for the MJAR array in more detail, a data base of MJAR recordings was created. The search criteria used were the same as those described above. The original MJAR data and associated onset parameters were retrieved from the PIDC for 294 onsets between May 1 and July 1, 1997. Fig. 7.7.1 shows the error vectors, given as the difference in this ray parameter versus azimuth space, between the PIDC slowness values and the theoretically estimated ones. The symbol (small circle) used in this figure represents the theoretical value, and the end of the line represents the PIDC estimate. As expected from the former paragraph, the amount of erroneous estimates is large quite large, but on the other hand many slowness estimates are quite good. It is also seen that for some source regions systematic shifts can be observed, which is an indication for lateral heterogeneities beneath the MJAR site. The problematic slowness estimates are not those associated with such systematic effects, but the large amount of erroneous estimates with error-vector lengths of more than 5 s/°. These errors are not associated with a specific source region, epicentral distance, nor azimuth; since all slowness values are affected similarly.

The processing parameters used to analyze the data for MJAR at the PIDC were also retrieved and used in the NORSAR analysis program EP to reproduce the results of the PIDC. This did not work perfectly, although the amount of 'bad' estimates, as well as the mean features of the reported values were in general confirmed. The differences found between the PIDC and the EP results can be explained partly by different realizations of filters, fk-analysis and onset-time handling, but it also became clear that the results were strongly dependent on parameter settings such as frequency range used, analysis time window, and slowness range for the fk-analysis.

Improving the MJAR analysis

The MJAR array consists of six sites approximately situated on a circle plus one central site. The aperture of the array is about 10 km and the mean distance between neighbouring sites is about 5 km (see Fig. 7.7.2). This geometry defines the characteristic parameters of this array, which influence the results of the slowness measurements. For a monochromatic wavefront the minimum distance Dmin between neighbouring array sites defines the maximum wavenumber Kmax, which can be resolved by this array without distortions due to aliasing effects:

$$D_{min} = 1/(2 \cdot K_{max})$$

The wavenumber Kmax is related to the minimum apparent velocity Vmin, for which a monochromatic wave with the frequency ν can be resolved:

$$K_{max} = \nu/V_{min}$$

Combining the two equations with the mean distance between neighbouring sites at MJAR of about 5 km, one gets:

$$V_{min} = 10 \cdot \nu$$

This means, e.g., that for a signal with a frequency of 1 Hz the slowness can only be correctly estimated, if the apparent velocity is higher than 10 km/s (or the ray parameter lower than 11.12 s/°). Results of the fk-analysis for local or regional onsets with lower apparent velocities can be distorted by aliasing, and one may obtain a slowness solution, which actually is on a side lobe of the array transfer function. The fk-analysis as implemented at the PIDC and at NORSAR is applied to a frequency band and not to a single frequency. Results of the fk-analysis are usually presented in the form of observed energy vs. slowness. The observed energy should have its maximum at the same slowness value for all frequencies represented in the signal. But the position of the side lobes in this form is different for each frequency. Superimposing the results of all frequencies will amplify the correct slowness and reduce the side lobes. However, if the frequency pass band is relatively narrow, a side lobe may have the largest energy, and erroneous results will occur. Therefore all fk-analysis should be done with a frequency band that is chosen to be as wide as possible.

So the first attempt at solving the problem of the large amount of erroneous fk-results from analysis of MJAR data was to limit the analysis to the maximum slowness values that can be resolved for the reported dominant frequency of the onset. The parameters used in the fk-analysis at the PIDC and also the parameters chosen in this study are given in Table 7.7.2. The statistical results for this first trial - here referred to as method 'AD-HOC I' - are listed in Table 7.7.3, and Fig. 7.7.3 shows a plot of the new error vectors. Table 7.7.3 also contains the statistical values for the original PIDC results: The improvements are obvious.

Fig. 7.7.2 also gives the relative elevation of the single MJAR sites. The elevation difference between the single stations is up to 825 m. The travel-time difference for a plane wave due to this difference is up to 0.18 s for vertical incidence and an assumed mean velocity beneath the array of 4.5 km/s. This is an effect which cannot be neglected, as usually done for other arrays. To investigate the influence of the site elevation on the quality of the fk-results for MJAR data, a test called 'AD-HOC II' was performed, taking the elevation differences also into account

during the fk-analysis. A plot of the error vectors from this test can be seen in Fig. 7.7.4 and the mean errors are also given in Table 7.7.3. Again, the improvement is significant, and up to 50% (with respect to the PIDC solution) of the erroneous fk-results now disappear. But also the non-erroneous results became clearly more stable; note in particular the more consistent results for events from the same source regions and the clear decrease of the median of the slowness errors from 2.25 to 1.94 s/°. The reason for the relatively large mean slowness error of MJAR (see rightmost column in Table 7.7.1) in the PIDC processing could just be the fact that the PIDC does not take the elevation differences between the single sites into account in its fk-analysis.

From the formula given above for Vmin, it is clear that low apparent velocities are better resolved the lower the analysis frequency. Additionally, the coherency for signals above 3 Hz is relatively low, because of the large distance between the single sites of the MJAR array. Therefore a scheme to obtain the 'best' frequency range was developed and applied for the MJAR data set. The principles of this procedure can be seen in the flow chart in Fig. 7.7.5. Firstly, the detection beam is recalculated and the frequency range with the highest SNR is searched for. This frequency range is shifted as far as possible to lower frequencies because the best resolution and best signal coherence can be expected for low frequencies. Especially smaller amplitudes are often disturbed by local noise. Therefore the fk-analysis is not done for the first part of the onset, but for the part of the signal for which the SNR has its highest value. This alone contributes positively to obtaining more stable fk-results, but the best fk-parameters are found, when in an iterative process firstly the lower frequency limit and secondly the higher frequency limit are systematically modified by small steps of 0.15 Hz around the original values. For each modification of the frequency band the fk-analysis is redone. Then the final and 'best' estimate of the slowness is chosen as the one, for which the fk-quality parameter attained its highest value. This procedure adapts more precisely the frequency band to the characteristics of the actual onset. In all cases the site elevations were taken into account in the fk-analysis. Fig. 7.7.6 shows the results for the 'BEST' solution, for which the statistical values can be found in Table 7.7.3. A reduction of erroneous onsets by about 66% and a decrease of the mean length of the slowness-error vector by about 48% (compare values in the first and the last rows of Table 7.7.3) clearly demonstrate the advantages of this procedure.

Conclusions

The anomalous amount of erroneous slowness estimates for onsets at MJAR is the result of several factors. First of all, the mean minimum distances between the single sites (about 5 km) is too large for resolving the slowness of higher frequency signals. This is due to the array transfer function and the lack of signal coherency. The small number of sites makes the array additionally very sensitive to noise and other complications at any one site, since the array has practically no redundancy in its data. Applying some simple plausible changes to the parameters to estimate the slowness, a reduction of the erroneous estimates by about 30% can be achieved. A specific problem with the PIDC processing is that it does not take into account the different elevations of array sites in the fk-analysis. With a special search for the best frequency range to use in the fk-analysis for each onset, the erroneous onsets can be reduced to about 30% and all other slowness estimates are very stable. However, because of the inherent problem of the array configuration the erroneous estimates cannot be removed totally (see Fig. 7.7.6).

Recommendations for the PIDC processing of MJAR data

To stabilize the MJAR data analysis at the PIDC in the short term, the above mentioned changes in the data processing are strongly recommended:

1) The parameters to be used in the fk-analysis of MJAR data should be calculated in accordance to the actual frequency content of each signal, as shown in Table 7.7.2 (case AD-HOC I).

2) The fk-analysis routine should be modified so that elevation differences between the array sites can also be taken into account (Table 7.7.2, AD-HOC II).

3) A further improvement of the results can be obtained by an iterative search for the best frequency band, by choosing the analysis window around the maximum SNR value, and calculating the other parameters as shown in Table 7.7.2 (case BEST).

In a longer perspective, a modification of the MJAR array configuration (i.e. minimum distance between sites, number of sites) would be the better solution, especially in order to improve the capability for resolving also larger slowness values in a higher frequency range. The definition of additional and well analyzed S onsets would contribute to improving the location of seismic events in the whole region surrounding MJAR.

J. Schweitzer

References

Berteussen, K.-A., 1974. NORSAR location calibration and time delay corrections. In: NORSAR Semiannual Tech. Summ. 1 Oct 73 - 31 Mar 74, NORSAR Sci. Rep. **2-73/74**, Kjeller, Norway.

Buland, R. & Chapman C.H., 1983. The computation of seismic travel times. Bull. Seism. Soc. Am. **73**, 1271-1302.

Fyen, J., Ringdal, F. & Paulsen B., 1995. Development of improved NORSAR time delay corrections. In: NORSAR Semiannual Tech. Summ. 1 April - 30 September 94, NORSAR Sci. Rep. **1-95/96**, Kjeller, Norway.

Jeffreys, H. & Bullen, K.E., 1940. Seismological tables, British Association for the Advancement of Science, London.

Kennett, B.L.N. & Engdahl, E.R., 1991. Traveltimes for global earthquake location and phase identification, Geophys. J. Int. **105**.

Schweitzer, J. 1994a. Some improvements of the detector / SigPro - system at NORSAR. In: NORSAR Semiannual Tech. Summ. 1 Oct 93 - 31 Mar 94, NORSAR Sci. Rep. **2-93/94**, Kjeller, Norway.

Schweitzer, J. 1994b. Mislocation vectors for small aperture arrays - a first step towards calibrating GSETT-3 stations. In: NORSAR Semiannual Tech. Summ. 1 April - 30 September 94, NORSAR Sci. Rep. **1-94/95**, Kjeller, Norway.

Schweitzer, J. 1995. An assessment of the estimated mean mislocation vectors for small-aperture arrays. In: NORSAR Semiannual Tech. Summ. 1 April - 30 September 95, NORSAR Sci. Rep. **1-95/96**, Kjeller, Norway, 128-139.

Schweitzer, J. & Kværna, T., 1995. Mapping of azimuth anomalies from array observations. In: NORSAR Semiannual Tech. Summ. 1 Oct 1994 - 31 March 95, NORSAR Sci. Rep. **2-94/95**, Kjeller, Norway.

Table 7.7.1: The table gives the results from analysis of P-phase measurements at the different GSETT-3 arrays. #T is the total number of analyzed phases. NRES gives the number and percentage of phases rejected from analysis due to travel-time residuals exceeding 4 seconds. NPER gives the number and percentage of phases rejected from analysis in accordance with the requirement that the largest onset of a teleseismic phase must have a period larger than 0.25 seconds. NSLOW gives the number and percentage of phases rejected from analysis due to the length of the mislocation vector exceeding 5 s/°. #U is the number of phases remaining after the rejection of outliers, as defined by the three criteria found in the text. DT gives the mean onset-time residual and the associated standard deviation. DPHI gives the mean azimuth residual and the associated standard deviation. DR gives the mean ray-parameter residual and the associated standard deviation. DS gives the mean length of the mislocation vectors.

ARRAY	#T	NRES #	NRES %	NPER #	NPER %	NSLOW #	NSLOW %	#U	DT [s]	DPHI [°]	DR [s/°]	DS [s/°]
ARCES	6367	77	1.2	81	1.3	426	6.7	5783	0.59±1.35	3.51±19.82	0.62±1.32	1.64
ASAR	7238	181	2.5	22	0.3	278	3.8	6757	1.42±1.06	0.70±11.22	-0.14±1.12	1.12
CMAR	5737	471	8.2	4	0.1	124	2.2	5138	2.14±1.18	1.48±18.15	-0.56±1.33	1.85
ESDC	3429	255	7.4	0	0.0	165	4.8	3009	1.48±1.24	-2.02±17.98	0.17±1.04	1.36
FINES	6993	109	1.6	31	0.4	769	11.0	6084	0.63±1.20	3.98±25.85	0.31±1.55	1.99
GERES	6322	446	7.1	21	0.3	365	5.8	5490	1.43±1.19	6.43±26.61	-0.29±1.25	1.55
HFS	4680	105	2.2	20	0.4	804	17.2	3751	0.86±1.71	5.69±27.28	-0.38±1.86	2.35
ILAR	1374	16	1.2	0	0.0	29	2.1	1329	0.30±1.24	2.43±20.67	-0.68±1.18	1.85
KSAR	1024	57	5.6	0	0.0	37	3.6	930	1.84±1.05	1.42±12.93	0.28±0.95	1.20
MJAR	4470	261	5.8	10	0.2	979	21.9	3220	1.29±1.21	3.52±18.21	-0.11±1.24	2.19
NORES	5262	125	2.4	77	1.5	387	7.4	4673	0.63±1.57	3.39±19.02	0.37±1.48	1.76
PDAR	5109	176	3.4	1	0.0	750	14.7	4182	1.63±1.07	-8.35±40.03	-1.18±2.21	2.94
SPITS	2482	43	1.7	84	3.4	1262	50.9	1093	0.86±1.39	19.43±40.09	-0.30±2.33	3.17
TXAR	5723	282	4.9	8	0.1	256	4.5	5177	1.95±1.17	-11.28±38.59	-0.29±1.63	2.32
WRA	5633	121	2.2	13	0.2	203	3.6	5296	1.07±1.58	2.61±8.54	0.24±0.80	0.85
YKA	3403	54	1.6	1	0.0	116	3.4	3232	0.78±1.15	1.60±6.90	-0.07±0.58	0.59

Table 7.7.2: Processing parameters to estimate slowness values at MJAR. FP1 and FP2 are the lower and upper cut-off frequencies of the prefilter, respectively. FREQ is the signal frequency as measured during the detection process. FK1 and FK2 define the frequency range for the broadband-fk analysis. SMAX is the largest slowness for the fk-analysis. T1 is the lead time before the detection time or the lead time before the time of the maximum SNR value. T1 thus defines the start time of the time window for the fk-analysis; T2 is the length of this time window. The column Z indicates whether the station elevations were taken into account in the fk-analysis. For further details see the text.

METHOD	PREFILTER LOWER CUT-OFF FP1 [Hz]	PREFILTER UPPER CUT-OFF FP2 [Hz]	FK1	FK2	SMAX [s/km]	LEAD TIME T1 [s]	WINDOW LENGTH T2 [s]	Z
PIDC	0.75	8.0	FREQ*2/3	2*FK1	0.36	1.1	2.4	no
AD-HOC I	as FK1	as FK2	FREQ*2/3	2*FK1	0.1/FREQ (≥ 0.1)	10*SMAX	1 / FREQ + 20*SMAX (≥ 4.0)	no
AD-HOC II	as FK1	as FK2	FREQ*2/3	2*FK1	0.1/FREQ (≥ 0.1)	10*SMAX	1 / FREQ + 20*SMAX (≥ 4.0)	yes
BEST	0.95*FK1	1.05*FK2	best, see text	best, see text	0.1/FP1	1/FK1 before the max SNR	2 / FK1 + 2 / FP1 + 1 (≥ 5.0)	yes

Table 7.7.3: Results of the different slowness estimates of MJAR data. DS is the length of the slowness-error vector, listed both as mean value and as a median of all estimates. NUMBER OF DS > 5 (or 10) gives the number of slowness estimates for which the observed error is larger than the given value. Listed are also the mean values of the lower and upper limit of the frequency range for the broadband-fk analysis (FK1 and FK2, respectively) and the largest slowness (SMAX) for which the fk-analysis is performed.

METHOD	NUMBER OF PHASES	MEAN DS [s/°]	MEDIAN DS [s/°]	NUMBER OF DS > 10 [s/°]	NUMBER OF DS > 5 [s/°]	MEAN FK1 [Hz]	MEAN FK2 [Hz]	MEAN SMAX [s/km]
PIDC	294	5.15	2.37	45	53	0.99	1.98	0.36
Ad-Hoc I	294	3.90	2.25	30	41	0.86	1.86	0.10
Ad-Hoc II	294	3.32	1.94	24	38	0.86	1.86	0.10
BEST	294	2.66	1.84	13	18	0.66	1.45	0.18

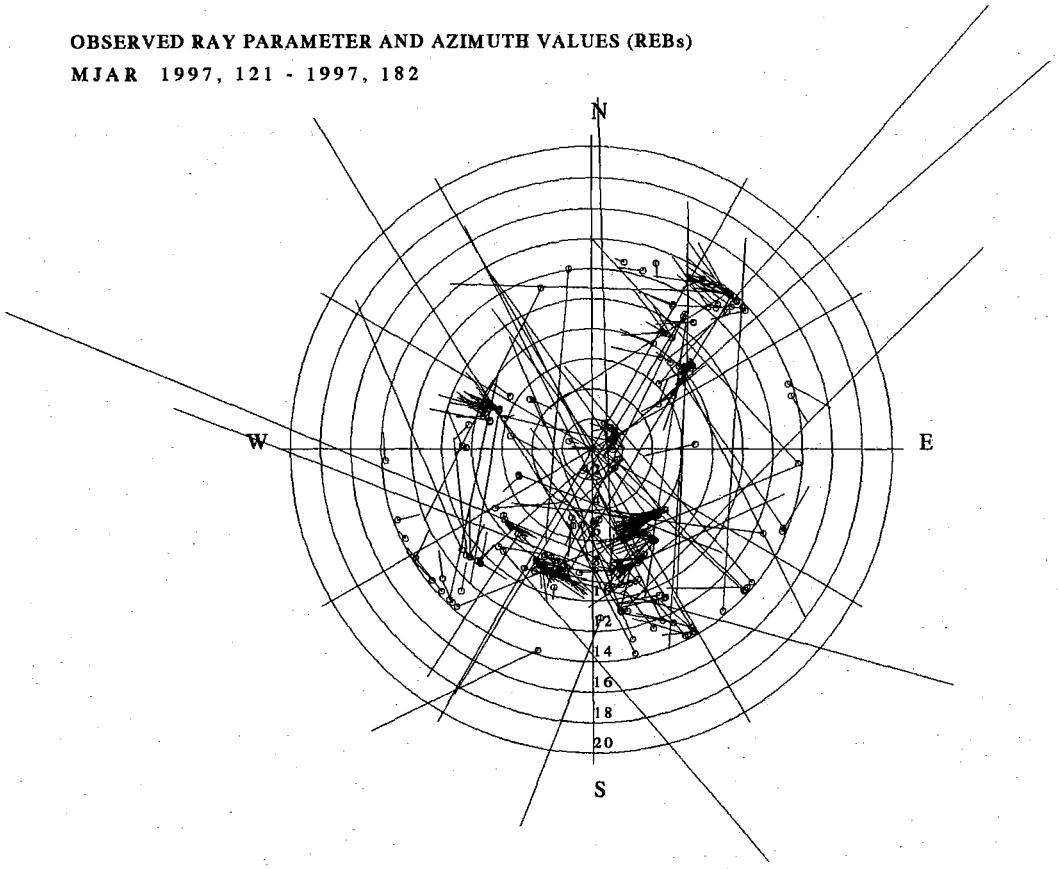

Fig. 7.7.1. Theoretical azimuth and ray-parameter values for the 294 onsets investigated are given by small circles. The corresponding values as reported by the PIDC in its final bulletins (REBs) are given as the end points of the lines.

MJAR CONFIGURATION

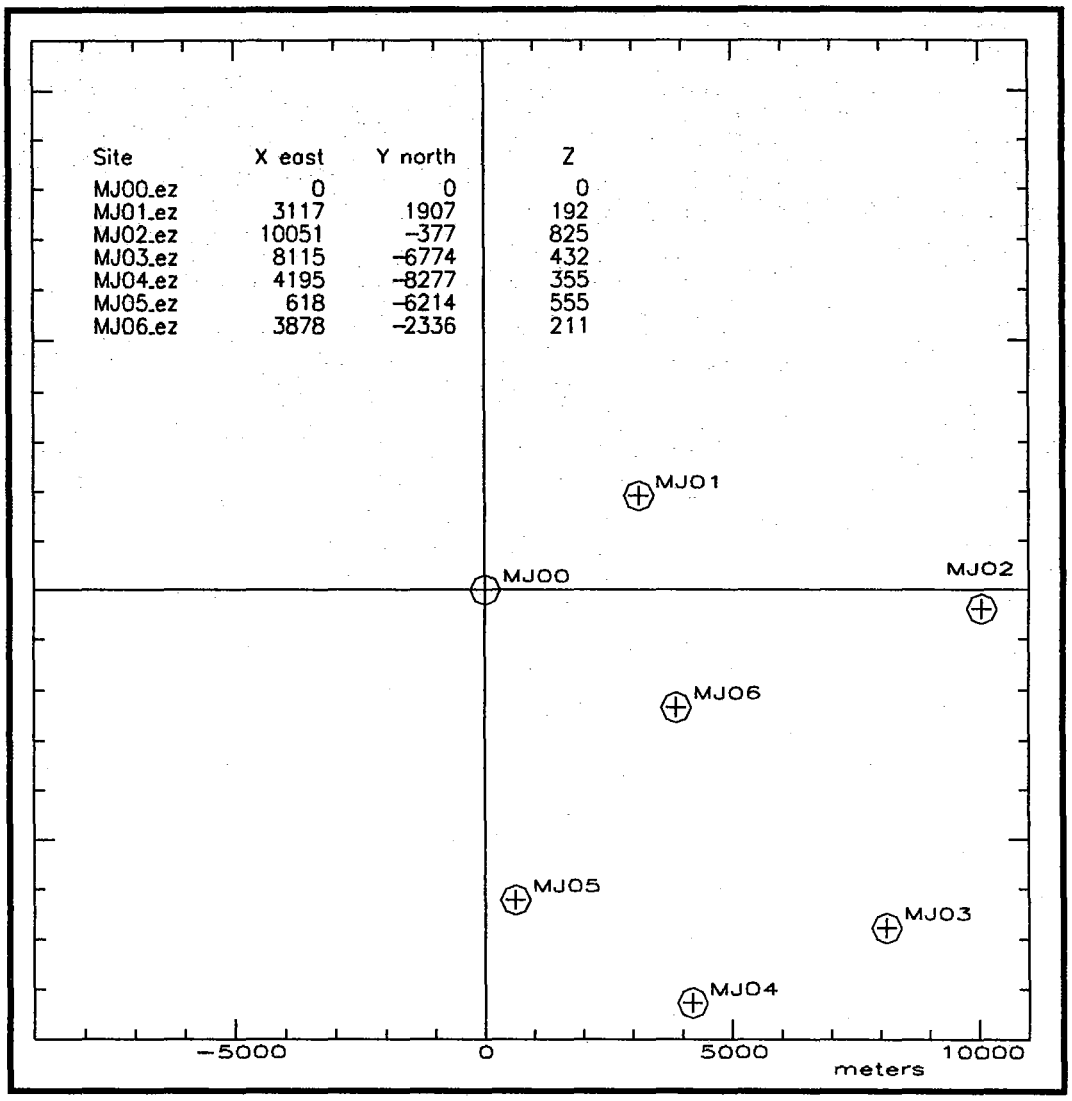

Fig. 7.7.2. The configuration of the MJAR array, with all coordinates given in m. The coordinates are relative to the array reference site MJ00. Note the large differences in the site elevations.

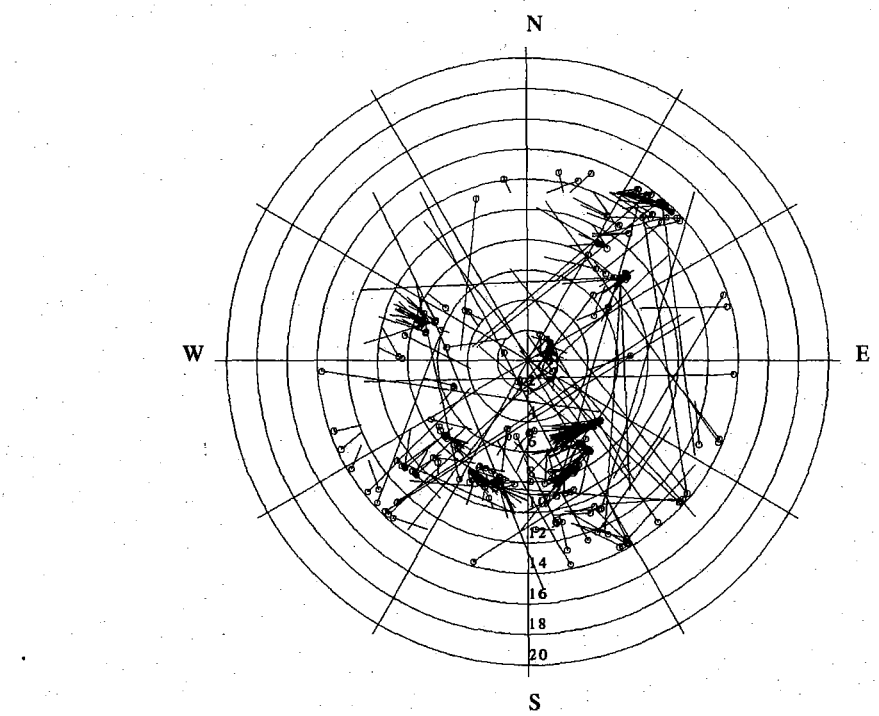

Fig. 7.7.3. Same as Fig.7.7.1 but here corresponding to parameters selected for the 'Ad-Hoc I' test to improve the fk-results, see Table 7.7.3 and the text.

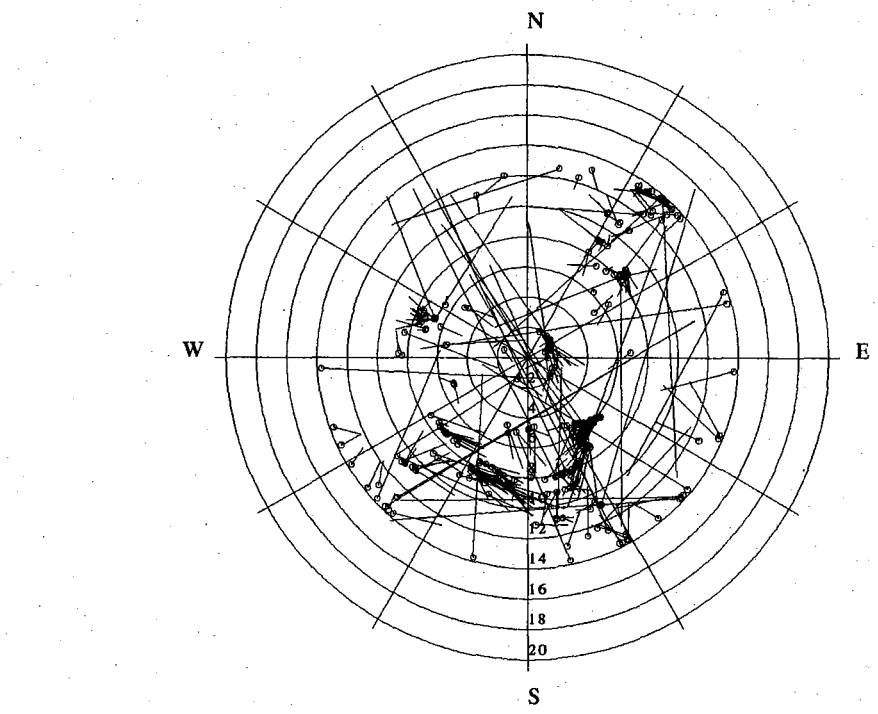

Fig. 7.7.4. Same as Fig. 7.7.3, but here the elevation differences are also taken into account in the fk-analysis.

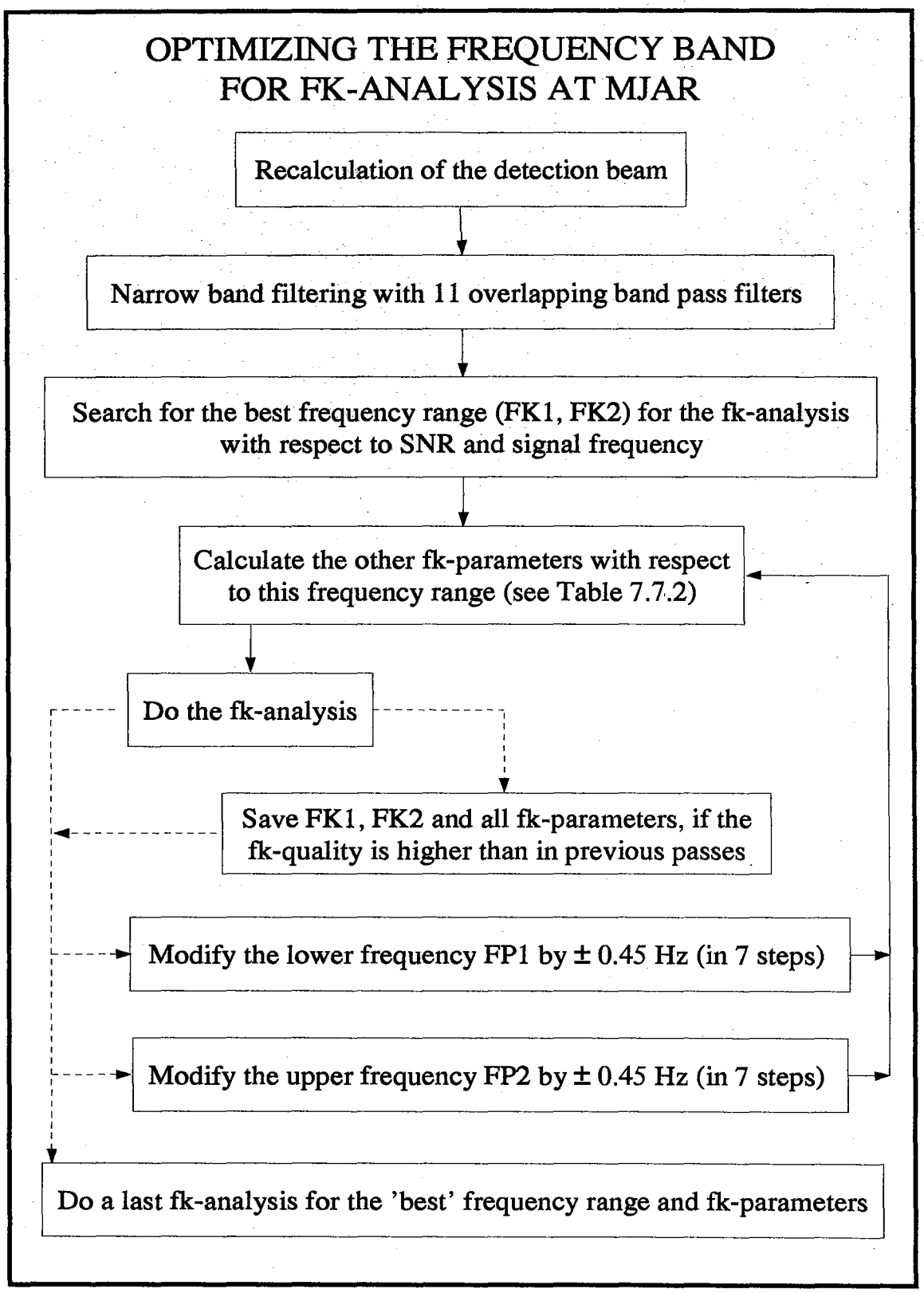

Fig. 7.7.5. This flow chart provides the details for an optimized MJAR fk-processing.

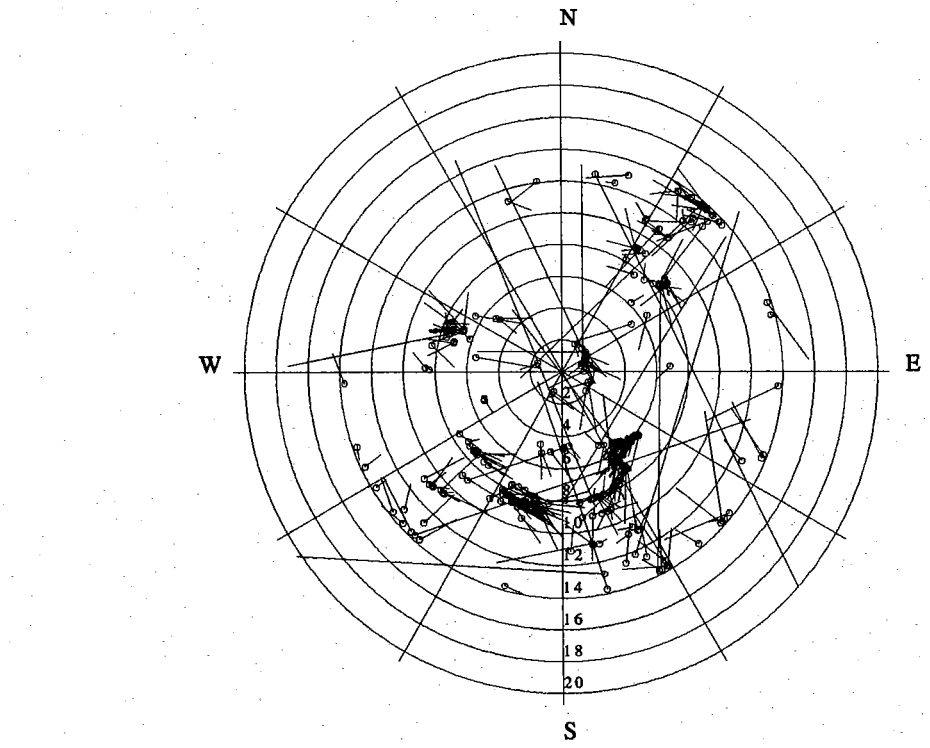

Fig. 7.7.6. Same as Fig.7.7.1, but here corresponding to parameters selected for the 'BEST' test to improve the fk-results, see Table 7.7.3 and the text.