The Research Council of Norway (NFR)

NORSAR

NORSAR Scientific Report No. 1-98/99

Semiannual Technical Summary

1 April – 30 September 1998

Kjeller, October 1998

7.3 Norwegian Experience with IDC Metrics During GSETT-3

Paper presented at the Workshop on Review and Definition of IDC Metrics 7-9 Sep 98

Introduction

The Ad Hoc Group of Scientific Experts (GSE) Third Technical Test, GSETT-3, began full-scale operations on 1 January 1995. In 1997, the responsibility for GSETT-3 was transferred to PrepCom's Working Group on Verification, and the GSETT-3 system is now gradually evolving into the International Monitoring System for the CTBT.

Evaluation has been an essential component of and prerequisite for the success of GSETT-3. Numerous national studies have contributed to these evaluation studies, including a number of papers from Norway. With respect to IDC metrics, the Norwegian contributions have focused on issues such as

- Metrics for event size
- Metrics to define location accuracy
- Metrics for capability estimation
- Metrics for REB completeness
- Metrics for event screening

This presentation gives an overview of some of the main experiences by Norway during GSETT-3, with emphasis on PIDC processing and results. Some more recent studies are also included. The paper focuses on issues and problems that are at the present time still not resolved, and gives suggestions for future improvements.

Metrics for event size

a) Body-wave magnitude m_b

Body-wave magnitude m_b has traditionally been the most common measure of the "size" of a seismic event. While this quantity is in general easy to measure at any given station, it shows a large variability across a seismic network for any given event. For this reason, it has been common practice to calculate the average magnitude measured at the individual stations of a network, and use this network magnitude as a best estimate.

It has long been recognized that this method can create a significant bias at low and intermediate magnitudes, because the stations which do not detect the event (usually those stations with the smallest signals) are selectively excluded from the averaging procedure. Figure 7.3.1 illustrates how this problem affects the magnitude-frequency relationship measured by the ISC network, when compared to NORSAR array magnitudes. It might be worth mentioning that under reasonable assumption, a single station or array produces an unbiased slope in this relationship, since the inherent scatter in single-station magnitudes merely shifts the baseline without affecting the slope (Ringdal, 1975). A similar result is found for IDC magnitudes, where the recurrence relations shows a slope significantly greated than 1.0 (see the IDC Performance Reports).

Looking at the same problem for a different angle, we have compared the IDC magnitudes to NORSAR magnitudes for a sequence of earthquakes from Greece in 1995 (Ringdal, 1995). Figure 7.3.2 illustrates the magnitude-dependent bias in both IDC and PDE magnitudes as compared to the "unbiased" NORSAR m_b.

The maximum-likelihood method (Ringdal, 1976) offers a means to compensate for this bias, but it has not yet been operationally implemented at the IDC. Current efforts aimed at implementing this procedure should be intensified, at the same time as efforts are underway to incorporate new distance corrections to enable the computation of m_b at regional distance ranges. In implementing the maximum likelihood method, the most important consideration is a quality check to ensure that non-operational stations or stations with abnormally low gain are excluded from the calculations.

We believe that the slope of the magnitude-frequency relationship, for various regions and for specified time periods, would be a useful and simple metric to assess the consistency of the IDC magnitude estimates. The actual assessment could be made by comparing this IDC slope to the corresponding slope for the same regions and time intervals as obtained from selected array stations in the IMS network. Such array stations (*e.g.* NOA) must be able to independently provide approximate location estimates in order to ensure that the regions correspond well enough.

b) Surface wave magnitude M_S

The recommendation to introduce the maximum-likelihood approach applies to the computation of network M_S as well as network m_b. In addition, a similar approach should be made to estimate the upper limit of M_S for events for which no surface waves are detected. This provides important information for the M_s:m_b discriminant, in the form of "negative evidence" as has been addressed in many studies in the past.

Recent studies, as *e.g.* documented in Section 7.4 of this report, have shown that the measurement of surface wave magnitudes at regional distances holds significant promise of lowering the limit for applying the M_s:m_b criterion, and would be of particular importance for the event screening currently being implemented at the IDC. Furthermore, regional surface waves have significant energy at shorter periods (down to 5-10 seconds), and this could be exploited in extending the spectral range for useful M_s measurements.

In particular, measurement of such shorter period surface waves at regional distances could contribute to reducing the influence of coda from surface waves of large teleseismic earthquakes, which often mask ordinary surface waves from small events for hours. The reason is that these strong surface waves generally have a dominant period of 20 seconds or more, with far less energy in the shorter period bands. This is illustrated in Figs. 7.3.3. and 7.3.4, which show NORSAR LP beam recordings for two Novaya Zemlya nuclear explosions (m_b=5.8 and 4.5). In the latter case, the surface waves in the "standard" frequency band are masked by an interfering teleseismic earthquake, but by applying a filter around 10 sec, these surface waves can be clearly seen.

As a new metric, we propose regional surface wave magnitudes at a suite of signal periods, *e.g.* 5sec, 10sec, 15 sec, 20 sec, 25 sec. This would in effect amount to providing a "spectrum" for recorded regional surface waves. It would be important to include an indicator of whether the measured level corresponds to noise or signal ("noise" includes possible interfering energy from other seismic events).

Metrics to define location accuracy

The traditional metric for location accuracy is the 95% confidence ellipse around the estimated epicenter. This metric should certainly be retained, but its implementation in the current IDC needs to take into account more realistic uncertainties in the parameters used for location estimation. Studies for many countries (including Fennoscandia) have shown that the location ellipse too often does not encompass the true epicenter. Significant progress in this regard is, however, taking place at the present time.

Looking at the available methods for estimating location, it is widely recognized that regional calibration is a requirement for achieving significantly better accuracy than today. Again, efforts are underway to develop such calibrated procedures at the IDC. There are, however, some factors that are much more difficult to quantify, and that also play a large role in producing mislocations. The most obvious is inaccurate reading of onset time, most often due to emergent signals with low SNR, but in some cases also caused by questionable analyst picks. The IDC experience in comparing picks by two or more independent analysts illustrates this problem well enough. It would be difficult to define appropriate metrics for this type of erroneous reading, but it is necessary to take this possibility into account when defining the error ellipse for small events.

An interesting result obtained by NORSAR in analyzing a sequence of Kola mining explosions with known locations, is that the most accurate locations (in this case) were obtained by including only three stations at close distances and correspondingly high SNR (Kværna and Ringdal, 1994, Ringdal, Kværna and Hokland, 1993). Even though, in principle, the locations should be improved by adding more stations, this did not happen in practice. The obvious reason is the lack of calibration (which is more serious at larger distances) combined with difficulties in reading onset time accurately at remote stations with low SNR.

This result could be important in future evaluation and estimation procedures. For example, if stations at regional distances from a given seismic event have been well calibrated through *e.g.* small chemical explosions or refraction surveys, it may be possible to estimate quite accurate locations using these regional stations only. It is far from obvious that the location accuracy would improve by adding a large number of teleseismic stations, for which the calibration information might be less developed. This question needs to be investigated in the future.

Metrics for capability estimation

The traditional method of estimating network capability is based upon an average statistical assessment of the noise level, the required SNR for detection and the number and types of phases needed to define an event. Recent developments in Threshold Monitoring, documented *e.g.* by Kværna and Ringdal (1998), promise to significantly expand and improve metrics for estimating capabilities, both on a network and station level.

Basically, the two types of network capability estimation can be summarized as follows:

Detection capability:

- The *smallest* hypothetical event that could be *detected* (*e.g*. by three stations)

Threshold capability:

- The *largest* hypothetical event that could have *occurred*

The Threshold capability always gives lower magnitude levels than the Detection capability, with a typical difference of 0.5-1 unit. Among the advantages of the Threshold Monitoring approach is that it can provide estimates of *both* the detection capability and the threshold capability

- continuously
- in near real time
- using the actually observed seismic field

In addition, the Global Threshold Monitoring system, as currently implemented at the PIDC, provides regular (hourly) statistics on individual station performance of the primary network. These performance statistics can be used to monitor the seismic noise level, seismometer gain, data quality (*e.g*. statistics on spikes) and instrument outage.

The global TM maps also give immediate indications of any degradation in global detection performance caused *e.g*. by coda of large earthquakes, abnormal noise levels for certain regions or stations or outages of key stations in the IMS primary network.

While the TM data provides a vast amount of potentially useful information, it will be a challenge to develop appropriate "simple" metrics to extract and make use of the most essential parts of this information.

Metrics for REB completeness

This topic is closely tied to the metrics for detection capability discussed above, but addresses some important additional considerations. In particular, the completeness of the bulletin must be seen in relation to the estimates of "expected" capabilities. Thus, even if the system "theoretically" has a certain capability, given a number of assumptions, an obious question to be considered is whether the actual detection performance, as observed in the REB, matches these theoretical estimates.

The PIDC Performance Reports already address this question by comparing the REB to the PDE or NEIC bulletins, and highlights events that are close to the 90% detection threshold of the IMS network but are not reported in the REB. This procedure should be expanded, taking also into account national earthquake bulletins. However, it is mandatory to accompany such comparisons by a realistic assessment of the reference magnitudes used at these non-IMS agencies. Again, this is a considerable challenge for future evaluation work.

An entirely different aspect of this problem is whether the IDC event definition criteria are appropriate for the purposes of the global system. As discussed earlier, there is a significant "gap" between even the theoretical detection capability of the network and the actual "threshold" at which we can monitor the upper limit of the magnitudes of possible occurring events. The current event definition criteria for the REB calls for P-detection at 3 or more primary stations. Obviously, many events could be (and are being) detected and located that do not satisfy this criterion, and consequently are not listed in the REB.

A particularly interesting example, in terms of CTBT monitoring, is the seismic event near Novaya Zemlya on 13 January 1996. This event was well detected (with P and S phases and azimuth estimates) by both the primary array ARCES and the auxiliary array SPITS (see Fig. 7.3.5). In fact ARCES was by itself able to detect and locate this event with reasonable accuracy, and the event thus fulfils the requirement that it should be "detected and located by the primary network". With the inclusion of SPITS, the location estimate could be further refined, as demonstrated by Ringdal (1997).

It will be an important task to develop metrics to assess the completeness of the REB, and to provide improved event definition criteria to enhance the completeness of this bulletin. Such new event definition criteria must carefully consider the tradeoff between achieving improved detectability and the desire to avoid overloading the REB with numerous small local events seen only at one or two IMS stations.

Metrics for event screening

The current event screening procedure employed at the PIDC focus on two criteria: event focal depth and M_s:m_b. These are considered to be by far the most robust criteria currently available, but have the disadvantage that they are difficult to apply to small events or events recorded only by few stations. Section 7.4 of this report describes some recent advances in studying regional recordings of surface waves, and the preliminary results indicate that it would be possible to apply the M_s:m_b discriminant to low magnitude events, perhaps approaching m_b=3.0-3.5 using regional data.

Other criteria, such as the high-frequency P/S ratio, hold the promise of being applicable at much lower event magnitudes. We have carried out extensive studies of this criterion for the Barents/Kara Sea region, and have concluded that at present, the P/S ratio is not sufficiently well understood to be routinely applied in event screening at the IDC (see the study described in Section 7.1 of this report).

In order to further develop the metrics for screening, it is necessary to study extensive historical recordings of nuclear explosions in various tectonic regions. Fortunately, many of the IMS stations have retained such recordings, but nevertheless the majority of IMS stations were not established at the time when the majority of nuclear explosions were conducted. The screening criteria must therefore be developed based to a large extent on non-IMS data. An excellent example is the historical data base of regional (analog) LP recordings retained in Apatity, Kola Peninsula (see Section 7.4).

Furthermore, since event magnitudes are important in most of the envisaged criteria, the problem of computing magnitudes of pre-GSETT-3 events in a way compatible with the current magnitude calculations must be addressed. This question is now being studied by many scientists, but again, we emphasize the need to avoid excessive reliance on past PDE, ISC or NEIC magnitude estimates, because of the potential magnitude-dependent bias discussed earlier in this paper.

Concluding remarks

Although the seismological procedures currently implemented at the PIDC are by now considered mature, there is still room for significant improvement, both in the calculation procedures and in the metrics designed to evaluate the IDC products and services. This includes event location, where improvements are needed both in regional calibration and estimation of arrival times at low SNR as well as improvement of metrics to measure location accuracy. Event magnitude is still not measured by maximum likelihood, and upper limits on non-detected surface waves should be included. Threshold monitoring promises to improve significantly the capability estimation, and will also provide metrics for characterizing station performance.

The completeness of the REB needs to be reassessed, with special view to the event definition criteria. In fact, with the current 3-primary station requirement, there are areas where the IMS can detect and locate events an order of magnitude smaller than the current REB threshold. Such a reassessment must, however, be carefully weighted against the undesired effect of including large number of small mining explosions and small aftershocks in the REB.

As detailed in this paper, there are many statistics and results currently forming part of the IDC processing which could give rise to useful metrics for evaluation purposes. An important future challenge will be to compress and synthesize these data to obtain metrics that represent the essence of the performance in a simple and easily understandable way. Furthermore, in the absence of “absolute” criteria against which to evaluate the system, the metrics will need to be assessed in a “relative” sense. Thus it is important to develop metrics which will provide a continuous assessment of the improvements, relative to previous practice, in the IMS and IDC processing as the development progresses in the years to come.

F. Ringdal

References

Kværna, T. and F. Ringdal (1994). Intelligent post-processing of seismic events, *Ann. Geofis.* XXXVII, 309-322.

Kværna, T. and F. Ringdal (1998). Seismic threshold monitoring for continuous assessment of global detection capability. *Semiannual Technical Summary 1 October 1997 - 31 March 1998*, NORSAR Sci. Rep. 2-97/98, Kjeller, Norway.

Ringdal, F. (1975). On the estimation of seismic detection thresholds, *Bull. Seism. Soc. Am.* 65, 1631-1642.

Ringdal, F. (1976). Maximum-likelihood estimation of seismic magnitude, *Bull. Seism. Soc. Am.* 66, 789-802.

Ringdal, F. (1986). Study of magnitudes, seismicity and earthquake detectability using a global network, *Bull. Seism. Soc. Am.* 76, 1641-1659.

Ringdal, F. (1995). Magnitude estimation at the IDC - a case study. *Semiannual Technical Summary 1 April - 30 September 1995*, NORSAR Sci. Rep. 1-95/96, Kjeller, Norway.

Ringdal, F. (1997): Study of Low-Magnitude Seismic Events near the Novaya Zemlya Test Site, *Bull. Seism. Soc. Am. 87 No. 6,* 1563-1575.

Ringdal, F., T. Kværna and B.K. Hokland (1993). Onset time estimation and location of events in the Khibiny Massif, Kola Peninsula, using the Analyst Review Station. *Semiannual Technical Summary 1 April - 30 September 1993*, NORSAR Sci. Rep. 1-93/94, Kjeller, Norway.

Ringdal, F. and E.S. Husebye (1982). Application of arrays in the detection, location and identification of seismic events, *Bull. Seism. Soc. Am.* 72, S201-S224.

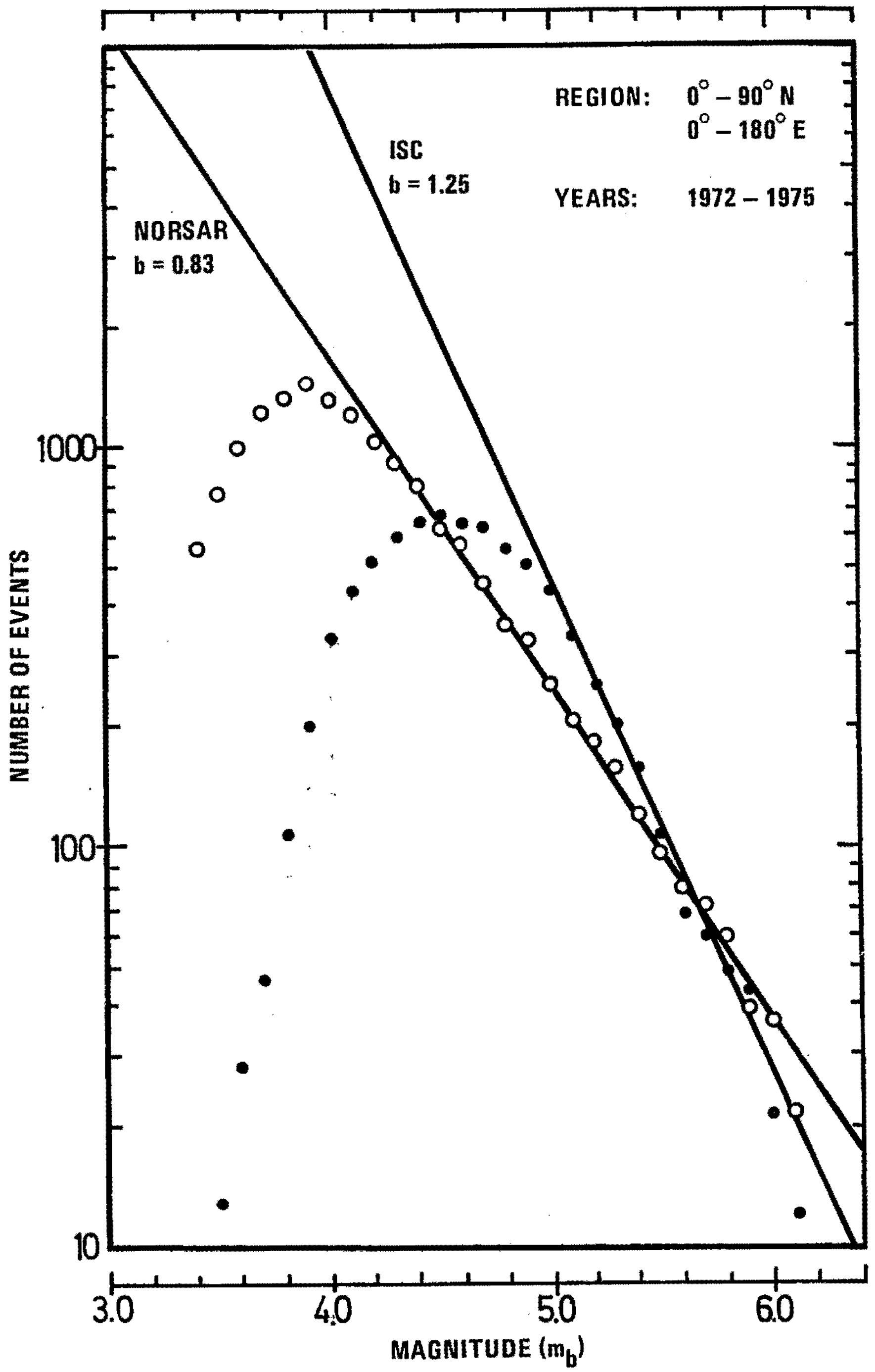

Fig. 7.3.1. ISC and NORSAR magnitude-frequency statistics for seismic events in the region 0-90 degrees North, 0-180 degrees East over the four year period 1972-1975. The filled circles (ISC) and open circles (NORSAR) correspond to incremental number of reported events at m_b intervals of 0.1 unit. Note the significant difference in the apparent slope of the respective recurrence relations. (After Ringdal and Husebye, 1982).

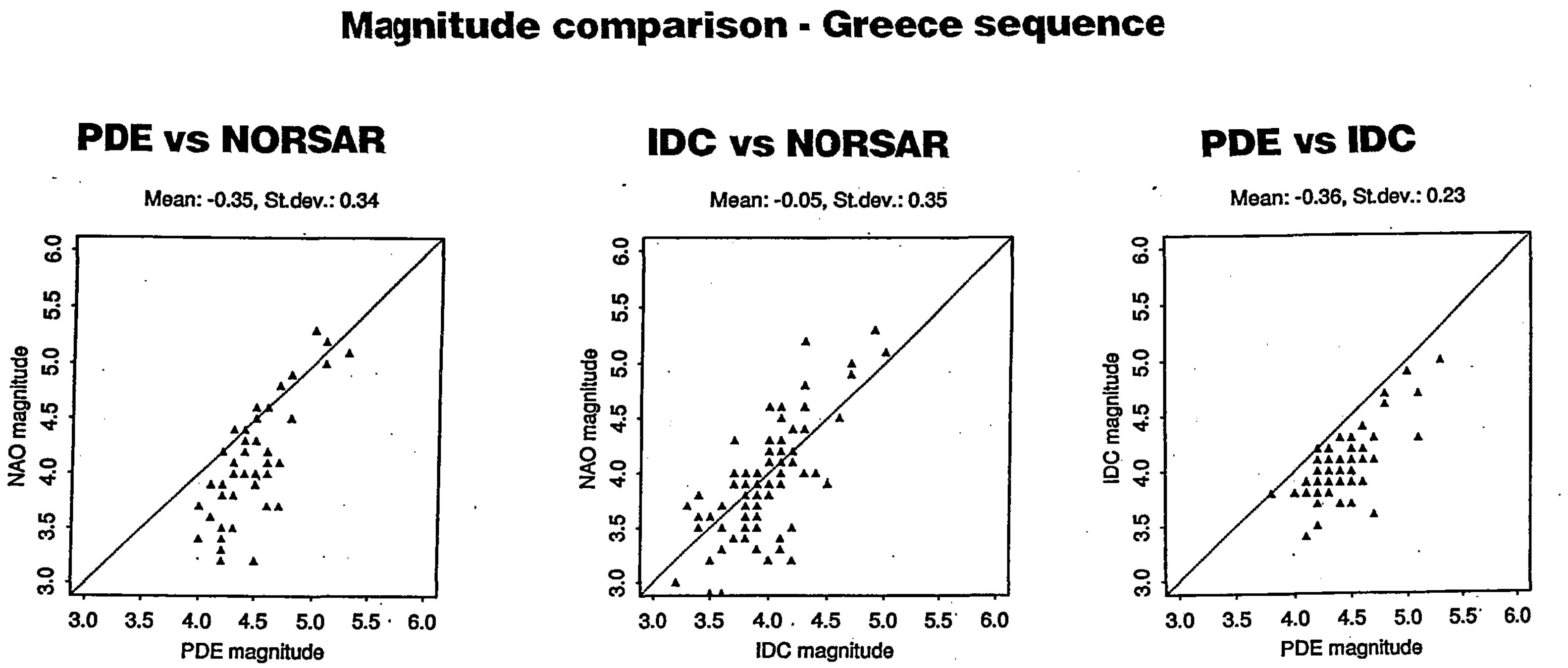

Fig. 7.3.2. Magnitude comparisons for various reporting agencies for an earthquake sequence in Greece during 1995. Note the network magnitude bias, which is particularly pronounced in the comparison of PDE and NORSAR magnitude. Note also the negative bias in IDC magnitudes compared to PDE. (After Ringdal, 1995).

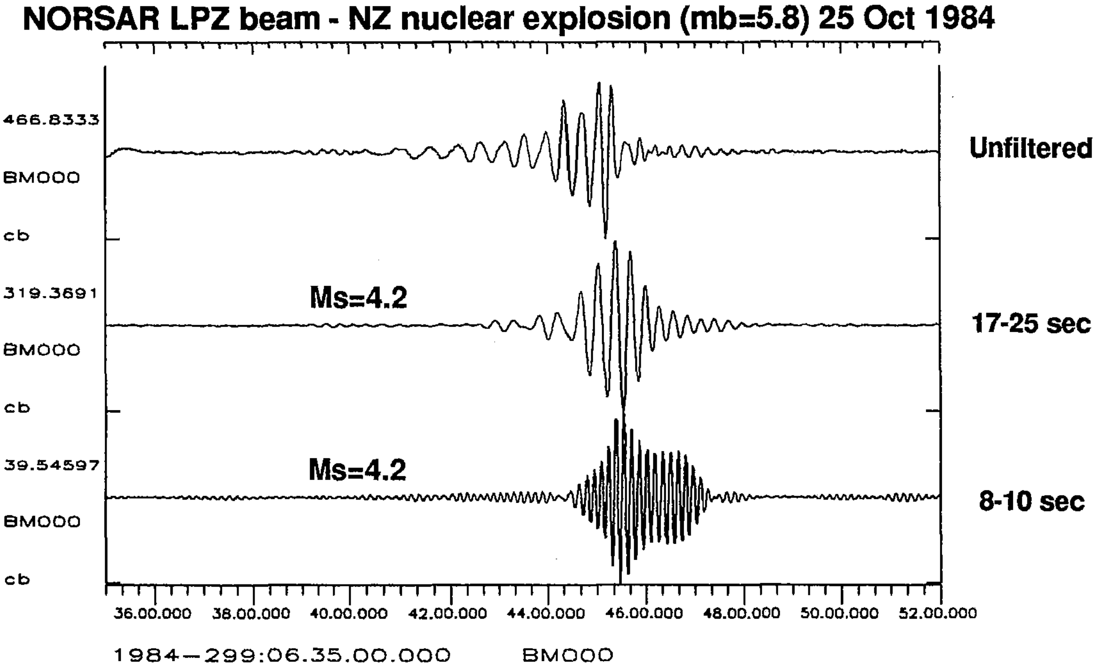

Fig. 7.3.3. NORSAR LPZ array beam recordings of a nuclear explosion (m_b=5.8) at Novaya Zemlya on 25 October 1984. An unfiltered beam is shown together with the beam filtered in the "standard" 17-25 seconds band and a "high-frequency" 8-10 seconds band. Note the high SNR of this regional recording (distance =20 degrees) even at the higher frequencies.

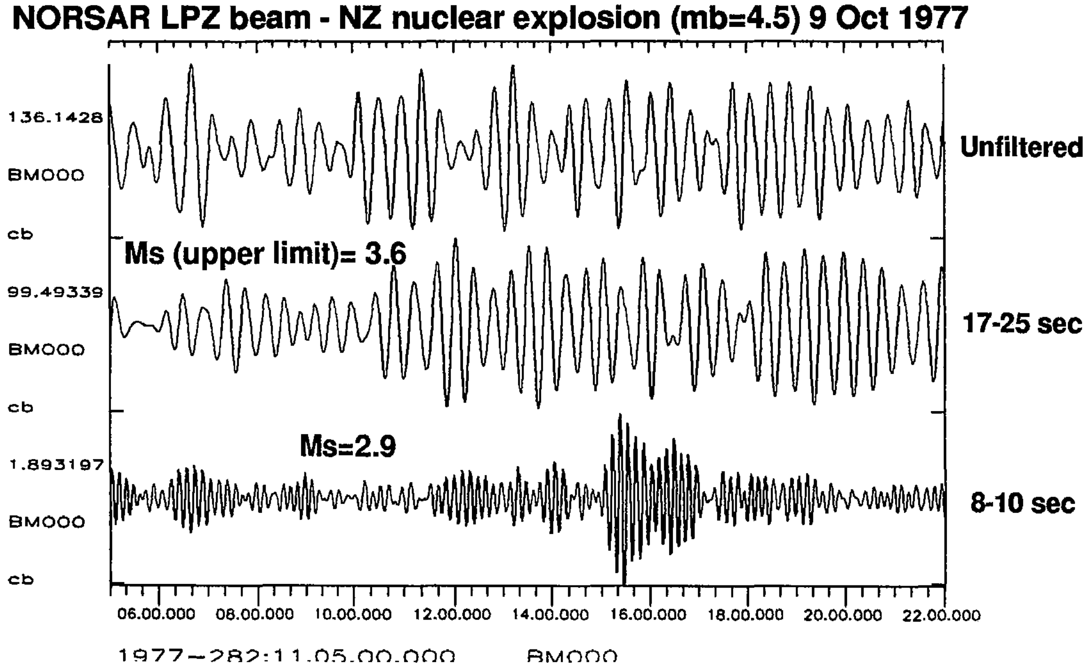

Fig. 7.3.4. NORSAR LPZ array beam recordings of a nuclear explosion (m_b=4.5) at Novaya Zemlya on 9 October 1977. An unfiltered beam is shown together with the beam filtered in the "standard" 17-25 seconds band and a "high-frequency" 8-10 seconds band. Note that an interfering event masks the explosion surface waves in the 17-25 seconds band, whereas the explosion signal is clearly seen in the 8-10 seconds band.

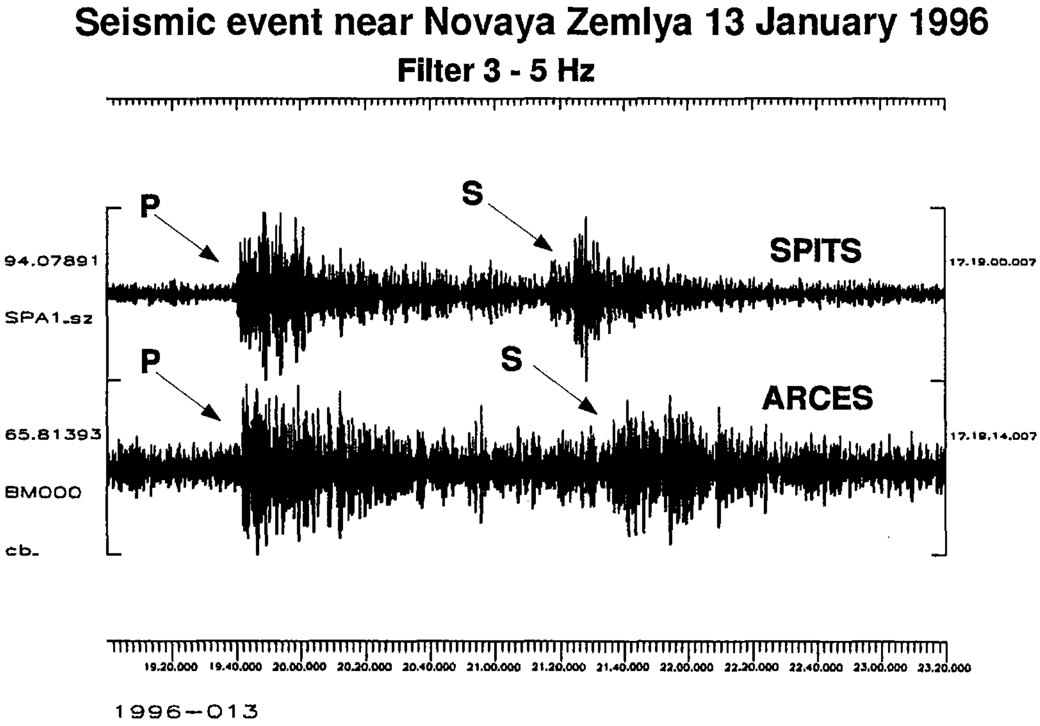

Fig. 7.3.5. SPITS and ARCES recordings of a small seismic event (m_b=2.4) near Novaya Zemlya on 13 January 1996. This event is about 1 magnitude unit smaller than the REB reporting threshold for this region, but can nevertheless be reliably detected and located using these two IMS stations.