ANNALI DI GEOFISICA, VOL. XXXVII, N. 3, June 1994

# Automated detection and association of surface waves

**Robert G. North and Catherine R.D. Woodgold**
*Geophysics Division, Geological Survey of Canada, Ottawa, Ontario, Canada*

**Abstract**
An algorithm for the automatic detection and association of surface waves has been developed and tested over an 18-month interval on broad-band data from the Yellowknife array (YKA). The detection algorithm uses a conventional STA/LTA scheme on data that have been narrow-band filtered at 20 s periods and a test is then applied to identify dispersion. An average of 9 surface waves are detected daily using this technique. Beamforming is applied to determine the arrival azimuth; at a non-array station this could be provided by polarization analysis. The detected surface waves are associated daily with the events located by the short-period array at Yellowknife, and later with the events listed in the USGS NEIC Monthly Summaries. Association requires matching both arrival time and azimuth of the Rayleigh waves. Regional calibration of group velocity and azimuth is required. Large variations in both group velocity and azimuth corrections were found, as an example, signals from events in Fiji-Tonga arrive with apparent group velocities of 2.9-3.5 km/s and azimuths from −5 to +40 degrees clockwise from true (great circle) azimuth, whereas signals from Kuriles-Kamchatka have velocities of 2.4-2.9 km/s and azimuths off by −35 to 0 degrees. After applying the regional corrections, surface waves are considered associated if the arrival time matches to within 0.25 km/s in apparent group velocity and the azimuth is within 30 degrees of the median expected. Over the 18-month period studied, 32% of the automatically detected surface waves were associated with events located by the Yellowknife short period array, and 34% (1591) with NEIC events; there is about 70% overlap between the two sets of events. Had the automatic detections been reported to the USGS, YKA would have ranked second (after LZH) in terms of numbers of associated surface waves for the study period of April 1991 to September 1992.

**Key words** *surface wave – detection – association - automated detection – azimuth determination*

## 1. Introduction

The quantification of earthquakes is a problem that has been largely solved for bigger events with routine and mostly automatic procedures such as centroid moment tensor inversion. The sheer number of smaller events has so far precluded a similar approach and they are still largely measured by magnitude. In a typical (recent) year, the USGS NEIS locates some 20000 events and for about one third of these, a teleseismic body wave magnitude ($m_b$) is given. For many of the events without $m_b$, local agency magnitudes are often reported, but there is a bewildering variety of these, based variously on the peak amplitudes or durations of specific phases or the whole recorded seismogram, using unknown or inconsistent formulae, and their relationship to seismic moment or to teleseismic magnitudes is generally unknown.

For those events that are large enough to be recorded to teleseismic distances, long period surface wave magnitude ($M_s$) is reported for only 20% of those for which $m_b$ is given. To some extent this is because surface waves are hard to observe for smaller events, and, even if observed at a single station, generally difficult to associate with the $P$ wave observed for the same event. Current practice at the USGS NEIS and at the International Seismological Centre (ISC) is to calculate $M_s$ only when a surface wave has been reported by a station as being associated to a reported $P$ wave. Many fewer stations report long-period amplitudes than short period $P$ wave amplitudes, and often only the amplitude of the long period waves and not the time at which they were observed is given. Thus it would be difficult for international agencies such as the ISC to associate surface waves with a known event. If arrival azimuth of surface waves, which would be an additional useful factor in associating them with events, is reported, it is not given in the bulletins (neither are array observations of $P$ wave slowness or azimuth, reported to these agencies by a number of arrays worldwide).

The low level of reporting of surface waves must to some extent be due to the difficulties of observing and associating them. An extensive literature search has failed to reveal any published material on the detection or automated association of surface waves. It is known that long-period arrays were operated both in Alaska and Iran in the late 1960s and in the early 1970s, and that these used long-period beamforming, but no publications of the detection techniques employed have been identified. The purpose of this report is to describe a method for the automatic detection of surface waves that takes into account not only their amplitude but also their dispersive character. It also proposes the routine reporting of both the time and arrival azimuth, and outlines an automatic method whereby this additional information can be used to associate the surface waves with known events. These techniques have been tested using data from the Yellowknife array, but are equally applicable to three-component broad band or long period stations. Increased reporting of surface waves would help not only in the quantification of seismic events, but also prove useful for the verification of a comprehensive test ban. A reliable report of surface waves, even at a single station, that is consistent both in arrival time and azimuth with that expected from an event, may be sufficient to unambiguously identify that event as an earthquake rather than an explosion.

## 2. Data and pre-processing

As part of the 1989 modernization of the Yellowknife array (YKA), four Streckeisen STS1-VBB seismometers were installed in tunnel vaults with a spacing of 10-12 km. In addition to the continuous recording of the broad band channels, LP data at 1 sample/s are produced by filtering and decimation. Rayleigh wave detection is carried out on the vertical component LP data in batch mode, with one data day being processed at a time. Half an hour of data from the previous day is included to initialize the long-term average and to avoid missing surface waves which overlap the day boundary. Data drop-outs and quality problems are sufficiently common that it is worth routinely checking for their presence. The day is divided into 10 min periods and in each such interval a data channel is zeroed if its maximum value exceeds 5 times the maximum of the other three channels. The data are then filtered with a three-pole Butterworth band pass filter with corner frequencies at 0.04 and 0.06 Hz.

## 3. Beamforming and detection

For each of 19 beams, a beam trace is formed as the median (average of the two middle values) of four time-shifted traces at each sampling time. The median, rather than the mean, is used to reduce the effects of very large signals at one station due to calibrations or malfunctions, or zeroes at one station due to outages. Eighteen of the beams correspond to a phase velocity of 3.6

km/s and azimuths at 30 degree intervals. The corresponding time shifts are comparable to the 1 s sampling interval and quadratic interpolation is used between samples. The nineteenth beam is a straight sum of the traces, with no time shifts, and is used to identify body waves – if the amplitude is greater on this beam than any of the others, the detection is discarded.

The absolute value of each beam trace is summed over 1 min intervals, producing a short-term average (STA). If the beam with maximum STA is the nineteenth, the STA values on all beams are set to zero for 2 min. A long-term average (LTA) is computed for the same minute, as the mean of 20 consecutive STAs with a lag of 10 min (*i.e.* from 10 to 30 min ago). This lag improves the detectability of emergent signals.

A detection is declared if STA > 1.7 LTA for at least two consecutive minutes. It ends if any of several criteria are met: STA < 1.25 oldLTA where oldLTA is the value of the LTA at the trigger time; STA < 0.15 max(STA) where max(STA) is the greatest value of STA reached during the detection; or the length of the detection exceeds 60 min. The azimuth of the signal is obtained by interpolation between the two beams with the highest values of STA.

For non-array stations the beamforming stages described above cannot be carried out, but after signal detection, arrival azimuth can be estimated at stations with three-component data. The beamforming carried out on the Yellowknife broad band data does not greatly improve the signal to noise ratio for surface waves as the small aperture means that the background long period noise is almost as coherent as the signals; the chief function of the beamforming is simply to obtain an estimate of arrival azimuth.

## 4. Dispersion test

Many of the detected signals are not clearly dispersed wave trains, and a simple test for dispersion is applied. The unfiltered data for the detection, plus 10 min before and after, are selected. A beam is formed with delays corresponding to the azimuth determined, and filtered in seven different bands by windowing with a gaussian bell in the frequency domain as

$$w(f) = e^{-\alpha\left(\frac{f - f_c}{f_c}\right)^2} \tag{4.1}$$

where $\alpha = 10$ and $f_c$ is the centre frequency. The centre frequencies used are 0.02, 0.025, 0.03, 0.035, 0.04, 0.05 and 0.06 Hz. In each frequency band a beginning and an end of the signal is found as follows from the signal envelope. A threshold $d$ is calculated as $d = 0.7\,max + 0.3\,av$ where $max$ is the peak value of the envelope and $av$ is the average value of the envelop. The beginning of the signals is the end of the last 5 min interval before the peak in which the envelope stays below $d$. The end of the signal is the start of the first 5 min interval after the peak during which the envelop stays below $d$. The midpoint of the signal, in each frequency band, is simply defined as the average of the start and end points. This rather complicated scheme is found to give more consistent results than just using the peak values of the envelopes.

As a check for dispersion, the signal midpoints are tested to see whether they increase in time as the frequency decreases. Of the seven midpoints, the first or last, and any one other, may be ignored. Of the five remaining, one pair of consecutive midpoints may be simultaneous; in all other cases, each one must be later than that at the next lower frequency. For most real events, all seven midpoints increase in time with decreasing frequency, but many exceptions occur.

## 5. Surface wave detection results and association procedure

The beamforming and detection procedure has been applied to long period data recorded by the four broadband stations of the Yellowknife array between April 1991 and September 1992. A batch process is run early each day on a file containing the long-period data from the previous day. Between 4 and 15 surface waves (average of 8.5) are detected each day; for each the time and period of the peak amplitude and the arrival azimuth are established. The number of surface waves detected daily shows much less seasonal variation than the number of body waves detected by the short period array.

Initially, fairly liberal criteria were used to associate the detected surface waves with the events detected and located by the short period Yellowknife array. A surface wave was considered to be associated if the time of peak amplitude corresponded to an apparent group velocity in the range 2.2-4.0 km/s and the azimuth matched the great circle azimuth to better than 60 degrees. As data accumulated, and as the Yellowknife event list was superseded by the NEIS Monthly Summaries, it became apparent that there were strong variations in azimuth and apparent group velocity from one source region to another. More conservative association criteria can be applied if this regional dependence is taken into account. Figures 1 and 2 show the distribution of azimuth (observed – predicted) and apparent group velocity for events in different seismic regions. It can be seen that the distribution of azimuth residuals is quite strongly peaked (at least 70% of measurements lying within 15° of the peak) but that the peak varies from region to region by as much as 40°. The group velocities also show a strong regional variation, with peak values ranging from 2.5 to 3.2 km/s.

Variations on this scale are not unexpected. Group velocities are the fastest for paths that mostly cross older oceanic lithosphere (*e.g.* to the Southwest Pacific), and slowest for paths that traverse mostly continental structures. The azimuthal variations are broadly consistent with refraction along the paths, at continent-ocean boundaries or as the path crosses oceanic lithosphere of different ages. It should be possible, given the increasingly detailed knowledge of variations in phase velocity across the surface of the earth, to model these changes in azimuth and group velocity in some detail.

As data have accumulated, region-dependent corrections to arrival azimuth and apparent group velocity have been established for surface waves recorded at YKA. For each region, an azimuth and velocity are assigned and a surface wave is considered associated to an event in that region only if it matches these values to within a specified tolerance. The latter varies from region to region, being generally tighter for smaller or more distant regions and looser for larger or closer regions. As an example, the apparent velocities for region 1 (Aleutians-Alaska) are 2.1-2.7, whereas those for region 29 (Western Asia) are 2.6-2.9 km/s.

## 6. Association results

A detected surface wave is considered associated to a particular event only if the corrected azimuth matches the great circle azimuth to better than 30 degrees, and the apparent group velocity lies within the range established for the region in which the event is located. Of the 7369 surface waves detected during the time period April 1991 to September 1992, 1914 were associated with events for which locations were provided by the short period Yellowknife array, and 1753 were associated with events given in the USGS Preliminary Determination of Epicentres (PDE) monthly listings for the same time period.

Values of surface wave magnitude $M_s$ have been calculated from the peak amplitudes of the detected and associated surface waves recorded at Yellowknife. The PDE listings give 2184 $M_s$ values for events in the same time period. Of the 1753 YKA $M_s$ values, 774 are for events for which the PDE listings do not give $M_s$. Figure 3 compares PDE and YKA $M_s$ values for events where both are available. It can be seen that there is no systematic bias between the two; the rms scatter is less than 0.35 magni-

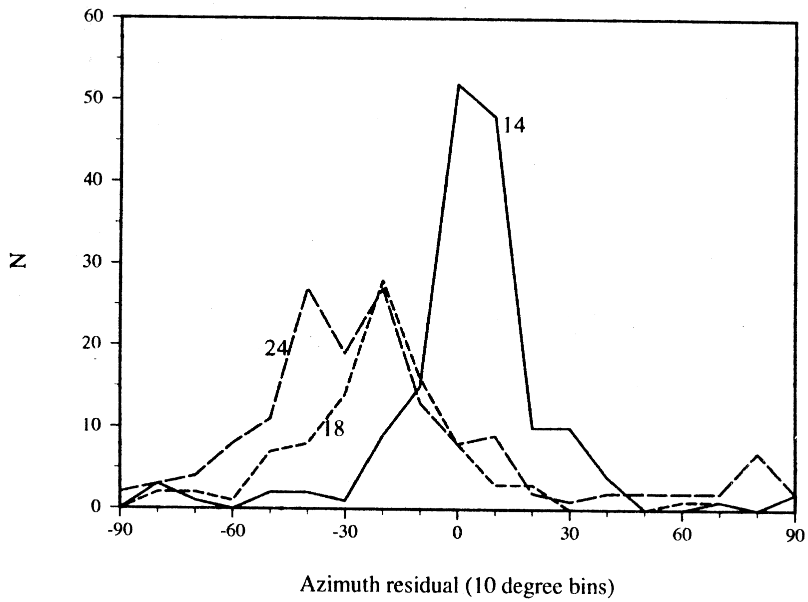


**Fig. 1.** Distribution of azimuth residuals (measured (from beamforming) – great circle value) obtained at YKA for events in seismic regions 14 (New Hebrides), 18 (Guam to Japan) and 24 (Sunda arc). For each region, the number ($N$) of residuals in 10-degree azimuth bins is plotted.

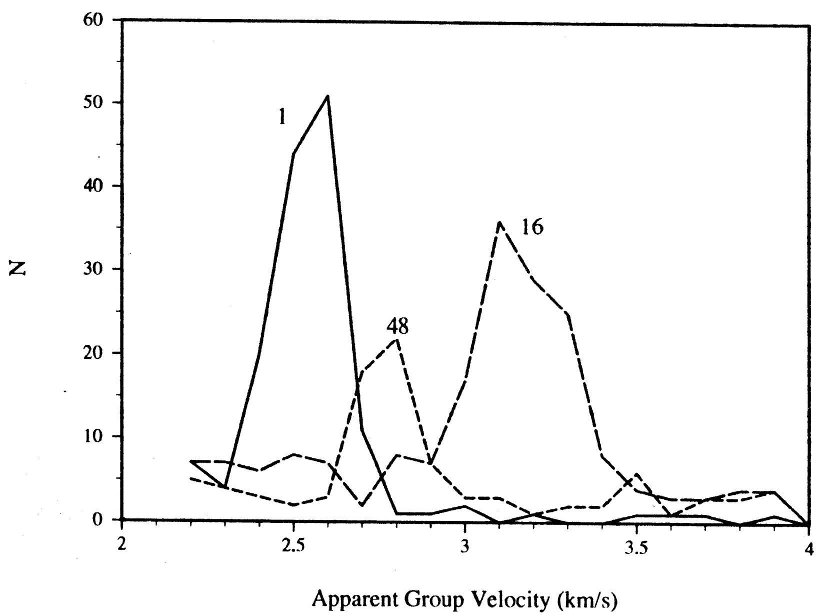


**Fig. 2.** Distribution of apparent group velocities measured at YKA for events in seismic regions 1 (Alaska-Aleutian arc), 16 (New Guinea) and 48 (Hindu Kush and Pamir). The number ($N$) of group velocities in 0.1 km/s bins is plotted for each region.

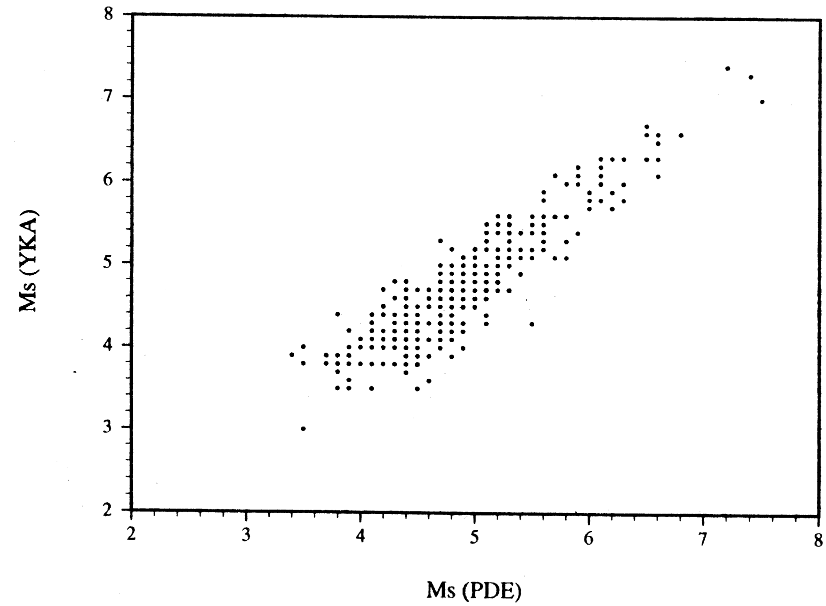


**Fig. 3.** Comparison of $M_s$ values reported by the PDE and determined automatically at YKA. Since the values are reported only to a precision of 0.1 units, a single point generally represents many pairs of values (also applies to figs. 5 and 6).

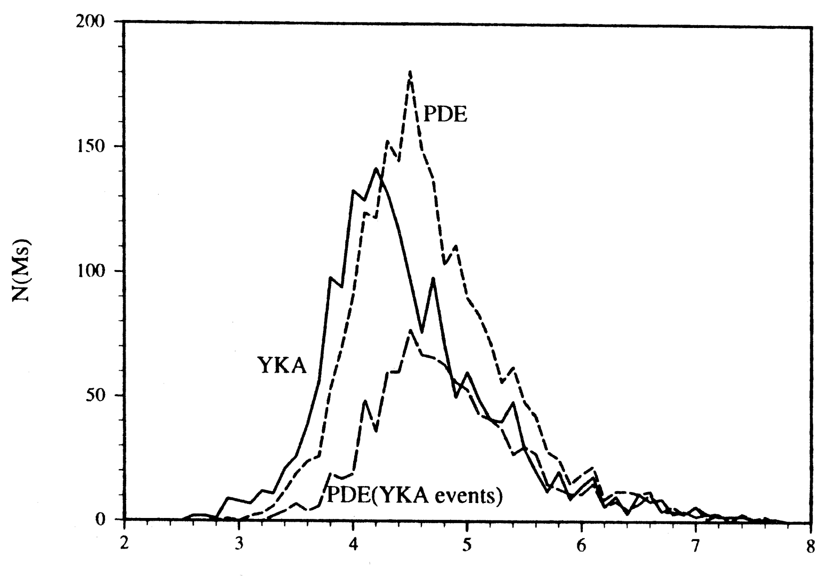


**Fig. 4.** Distribution of $M_s$ values reported for the time period April 1991-September 1992 in the PDE Monthly Summaries, determined automatically at YKA, and reported by the PDE for events for which YKA $M_s$ values were also available.

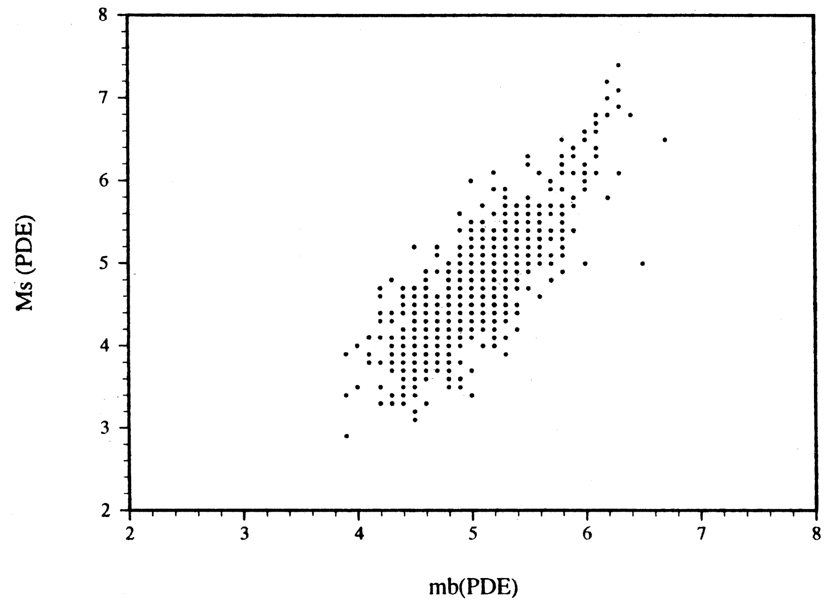


**Fig. 5.** Distribution of the 2190 $m_b$ and $M_s$ values reported in the PDE Monthly Summaries for the time period April 1991-September 1992.

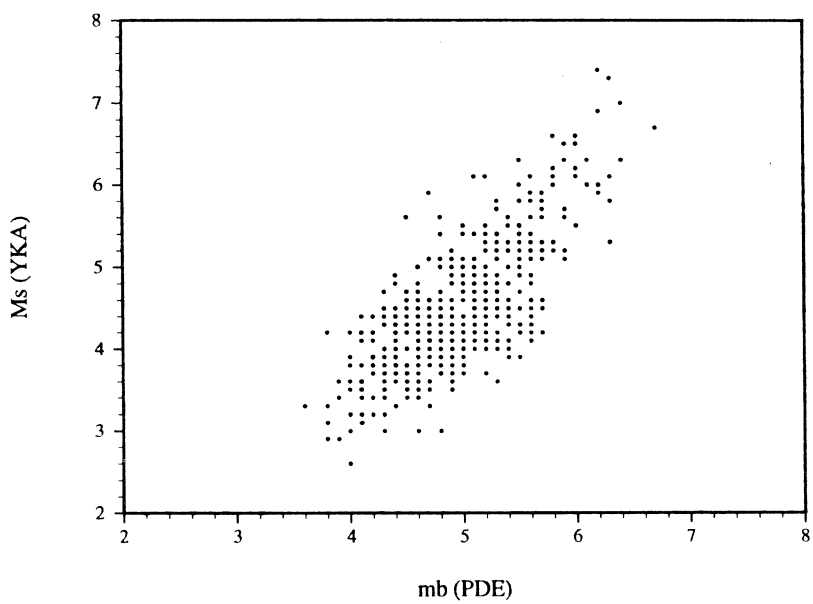


**Fig. 6.** $M_s$ values determined at YKA, plotted against corresponding $m_b$ values reported in the PDE.

tude units from equivalence of the two $M_s$ estimates. Figure 4 shows the distributions of surface wave magnitudes reported in the PDE, calculated from YKA detections, and reported in the PDE for events for which YKA values are also available. Note that the distribution of YKA magnitude peaks at a lower level than that of the PDE magnitudes, indicating that YKA estimate magnitudes for smaller events than does the PDE. It is also clear that YKA has reported surface wave magnitudes for many events for which no PDE estimates are available.

Figures 5 and 6 show PDE and YKA $M_s$ values as a function of PDE $m_b$ values. There is little apparent difference in scatter between these two $m_b$-$M_s$ diagrams but it is again evident that YKA $M_s$ values are available for smaller events than is the case for the PDE $M_s$ values. This provides strong evidence that an automated detection and association scheme can significantly enhance the reporting of $M_s$.

## 7. Conclusions

A conventional STA/LTA detector, supplemented by a test for dispersion, has proved capable of detecting a large number of surface waves at a single station. When combined with estimates of arrival azimuth, the time of arrival of the surface waves can prove very effective in associating them with known events. Azimuth can be determined from beamforming for arrays of long-period or broad band sensors, or from polarization analysis of three-component data. For a given station, region-dependent corrections to both azimuth and apparent group velocity may be significant but, when applied, can allow much tighter constraints on azimuth and arrival time of the surface waves, and thus increase confidence in association to events of known location.

A completely automated scheme based on this approach has been able to determine $M_s$ values for many events listed in the PDE Monthly Summaries for which no $M_s$ values were given in that publication. During the 18-month period, starting April 1991, analysed, the automatic scheme produced $M_s$ values for 1753 of the events listed in the PDE Monthly Summaries. During the same period, only one station, LZH in China, reported more (1968) values to the USGS NEIS, and only 9 stations reported more than 1000 $M_s$ values. The technique developed for YKA, if applied to other long-period stations, would significantly increase the number of $M_s$ values that could be used for the quantification and identification of seismic events.