PREPARATORY COMMISSION FOR
THE COMPREHENSIVE NUCLEAR-TEST-BAN TREATY
ORGANIZATION
(CTBTO)

PROVISIONAL TECHNICAL SECRETARIAT

INTERNATIONAL DATA CENTRE
SCIENTIFIC METHODS AND DATA FUSION SECTION
SEISMIC, HYDROACOUSTIC AND INFRASOUND METHODS
UNIT

Assessment of Hydroacoustic Processing in the CTBT Release One Monitoring Software

Peter Gerstoft

TECHNICAL REPORT
IDC-1999/10

DECEMBER 1999

Table of Contents

Executive summary

The first delivery, Release 1 (R1), of the seismic, hydroacoustic and infrasound monitoring software at the International Data Centre (IDC), Comprehensive Nuclear-Test-Ban Treaty Organization (CTBTO) relies mainly on the seismic network. The IMS hydroacoustic network routinely monitored at IDC during R1 consists of one hydrophone station, Wake Island (comprising hydrophones WK30 and WK31), and one T-phase station at Queen Charlotte Island, VIB (with one vertical geophone). A fully functional hydroacoustic monitoring capability will be provided at a later stage.

The processing of T-phase and hydrophone station time series is identical. Due to the better data quality from hydrophone stations, the examples in this report are mainly based on the Wake Island hydrophones. The signals from these have, however, a high level of 60-Hz electrical noise that is not removed before processing, and the signal is often clipped. It is expected that these problems will be removed during a station upgrade in summer 1999. Removal of external noise before data analysis should be considered, as this will improve the performance.

Having two hydrophone stations (WK30 and WK31) in close proximity (240 km) to each other offers an excellent opportunity to compare the consistency of the processing. However, the phase identification is inconsistent. At Wake Island it was found that detection of an H-phase on one hydrophone station was not well correlated to the detection of an H-phase on the other hydrophone.

Only about 15% of T-phases automatically associated with an event were accepted by the analysts, and all H-phases automatically associated with an event were rejected.

This report confirms that hydroacoustic processing in R1 is still in a development phase. T-phases are not used in location and all automatic associated H-phases during the R1 life at IDC were disassociated by analysts. Later software releases will make some of the problems identified in this report less severe [Han98, Han99]. The use of triplets of hydrophones in 1.5-2 km triangle should improve the performance, producing fewer misidentifications and providing a unique station-to-event angle will reduce misassociations.

1 Summary and recommendations

Summary of Findings:

1. Signals from WK30 and particularly WK31 have loud electrical noise. Signals at VIB have a high level 36-Hz and a lower level 18-Hz generator noise. [Not discussed in report].
2. Waveforms from WK30 may be clipped [It is observed that most larger signals are clipped, but not discussed in report].
3. The estimated arrival-time will likely be wrong for waveforms containing a spike.
4. The error estimates of the arrival-time (DELTIM) are unacceptably small.
5. The identification of phases is inconsistent. At Wake Island it was found that detection of an H-phase on one hydrophone station was not correlated to detecting an H-phase on the other hydrophone.
6. All automatically associated H-phases were rejected by the analysts. The automatic processing identified 744 events with H-phases from WK30 or WK31 associated. Misassociation of H-phases is a problem.
7. The R1 uncertainty ellipse is always too small. The number of degrees of freedom should be increased.
8. H-phases constrain the solution more tightly than the seismic phases for events with one or more H-phases associated. This is partly related to 4.
9. The blockage files for WK30 and WK31 are identical. There is no local blockage for Wake Island in the files; the closest blockage is about 30 deg away. The blockage files are not consistent with the 2D travel-time tables.
10. WK31 and VIB use a 1D travel-time table based on a constant sound speed of 1485 m / s. The modelled travel-time used in the location is based on time of flight (travel-time based on a constant sound speed) and not on probability weighted arrival-time, as the measured travel-times. This inconsistency causes a systematic error between measured and modelled travel-time.
11. WK30 uses a 2D travel-time table based on ocean sound speed profiles and an ocean acoustic propagation code. The 2D travel-time table contains large errors.

Preliminary recommendations:

Except for recommendations 3, 6, 12, 13 and 14, these are all long term recommendations which will require some more research before they can be implemented.

1. (DFX) Include a band rejection filter to remove external noise as part of a QC system. While the current problems at Wake and VIB should be fixed at the stations, there will likely always be some external noise to be removed.

It has also been observed that the received signal at Wake can be contaminated with short duration arrivals every 1-2 minute [not shown in report]. A matched filter in a QC system can likely remove these pings.

2. (DFX) Use signal envelopes in part of the processing. This might be used in feature extraction and determination of arrival-time. It is expected that this will make the arrival-time estimation less prone to errors and peaks grouped together will be assigned a higher weighting than a single peak. Additionally, the arrival-time measurement errors are expected to become more realistic.
3. (DFX) The hydroacoustic SNR in the arrival table is different from the seismic SNR as it depends on the mean squared signal and noise. Change the hydroacoustic SNR so that it similarly to seismics is based on an STA/LTA ratio.
4. (DFX) Base the arrival-time measurement on several frequency bands to increase the stability of the estimate.
5. (DFX) The arrival-time determination only includes the uncertainty in arrival-time due to ambient noise. It should also include uncertainty due to difficulties in estimating the true arrival-time in a noise-free environment. The uncertainty due to “signal error” is important and should be combined with the ambient noise error, as shown in Eq. (2). Signal error refers to uncertainties due to small variations in the frequency bands and the implementation of the filters. Including signal error gives more stable values of the arrival-time and the arrival-time measurement error, as it is less sensitive to the peak values of the waveform. The arrival-time measurement error will be larger when using this approach.
6. (DFX) T-phase stations are seismic stations. In order to also find seismic phases on T-phase stations, run both hydroacoustic and seismic DFX on these stations.
7. (StaPro) H-phases should be identified by specific features.
8. (GA) Study automatically built events with associated T-phases. It is not clear if the association works well. This could lead to a better understanding of T-phases and an implementation where T-phases are included in location but with a large travel-time modelling error.
9. (GA) Study automatically built events with three or more associated time defining H-phases (3 H-phases can build an event). During R1 there were 11 such events built with H-phases from WK30, WK31, NZL01 and NZL06.
10. (GA) Study automatically built events with both associated time defining H-phases and seismic phases. These events are likely spurious (see Figure 31), indicating that the seismic and hydroacoustic technology is not combined well.
11. (GA) Recompute the 2D travel-times using a modified approach. The provided 2D travel-time curves used presently are not smooth. Include earth velocities for the paths known to be through the solid earth.

12. (GA) Base the 1D travel-time table on an ocean sound-speed of 1482 m / s and not 1485 m / s. The 1482 m / s is closer to the average ocean sound-speed in the 2D travel-time tables as seen from WK30. Newer stations will be placed in areas with colder, and thus slower water. An ocean sound-speed of 1482 m / s is close to the default ocean sound-speed.
13. (GA) Blockage files should have about the same resolution as the 2D travel-times, the narrow paths should be removed, and for WK30 and WK31 the blockage files should be different.
14. (GA) Introduce a database field (probably in the `assoc` table) containing the total *a priori* timing error for each associated arrival to make the error calculation transparent (and similar fields for azimuth and slowness). This error depends on both the arrival-time measurement error and the travel-time modelling error. Only the first is currently stored in the database.
15. Develop a database of hydroacoustic events and associated waveforms. This database should be used for evaluation and development. It should contain all known explosions, T-phases with well known locations, numerically modeled explosions and arrivals with special features. There should be a data field to indicate the purpose of each signal in the data base.
16. Future designs of hydrophone stations will employ triplets of hydrophones. Consider how these will improve identification and reduce misassociations.
17. The use of T-phase stations is still on a research level. Several issues must be considered for their design: benefits of 3 component seismometer, benefit of having 2-3 seismometers separated by 10-300 km, distance of the sensor from the coast.

2 Introduction

There is no single reference guide to CTBT related hydroacoustics. For computational issues the excellent book by Jensen et al. [Jen94] is recommended. This book combined with Medwin and Clay [Med97] cover issues related to physics. A good description of CTBT hydroacoustics is in [IMS98]. Finally, hydroacoustic processing is described in [Han98].

The hydroacoustic part of the monitoring network is primarily concerned with monitoring events in-water but also, to a lesser degree atmospheric and terrestrial events. The detection, feature extraction and phase identification are specific for the hydroacoustic technology, while association and location include related seismic and infrasound detections.

Hydroacoustic phases are currently identified as N (Noise), T (Terrestrial source) and H-phases (in-water source). A typical scenario is illustrated in Figure 1. On a hydrophone the received signal for an H-phase is generally mostly affected by the source characteristics and the SOFAR wave propagation. A typical in-water source will be of short duration and contain both low and high frequencies (zero-100 Hz). The SOFAR waveguide will prolong the signal (this depends on range, typically it is about 5-10 s). In the SOFAR channel the phases can propagate long distances with very little attenuation, whereas energy in the seabottom will be strongly attenuated and will only propagate short distances.

For T-phases the waves initially propagate through the earth (see Figure 1). This causes the high frequencies to be filtered out. The coupling into the ocean and the SOFAR channel occurs over a wide area, mostly within a 2 deg horizontal range from the origin. The coupling is expected to mostly occur at sloping bottoms in the propagation direction where the energy couples into modes which have higher energy close to the seabottom. As these modes propagate to deeper water the energy will be transferred from the seabottom to the water column. Modes that have their main energy density bound to the water column will experience little attenuation and can thus propagate for long distances in the SOFAR channel. The T-phases generated in this way tend to be low frequency with a peak energy at frequencies below 10 Hz.

For T-phase stations (a geophone) the coupling from the SOFAR channel to the earth and the propagation through the earth to the geophone must additionally be considered. Relative to a hydrophone station this will attenuate the signal strongly filtering out high frequencies.

The hydroacoustic phase identification and travel-time estimation assume that the phases are horizontally propagating. For large events it is also possible for hydrophones to record signals that arrive at steeper angles, due to conversion of the seismic waves close to the hydrophones (the green dashed path in Figure 1). Mainly due to the small transmission coefficient at the sea bottom hydrophones will only detect larger events and due to the near vertical incidence the seismic

S-wave will usually not couple into the ocean. A T-phase station is a seismic station with a high sampling rate and will detect some seismic arrivals. The automatic hydroacoustic processing is not tuned for detecting and analysing these seismic arrivals. The detection (STA) window is longer than a typical seismic pulse length, so only large signals will be detected. And hydroacoustic station processing is different from seismic station processing, so if detected they will mostly be identified as noise. These seismic arrivals are detected by analysts and they arrive before the hydroacoustic arrival due to slower sound speed in the ocean.

In approaching a hydrophone all H and T-phases and most N-phases will propagate through the SOFAR channel. The modelling of the acoustic propagation in the SOFAR channel is important for understanding hydroacoustic arrivals. This can be done using a ray-theoretic approach, as illustrated in Figure 2. The preferred approach in ocean acoustic is to use normal modes [Jen94, Por91,Bin99]. The normal modes are the eigen functions for a range independent environment and a single frequency. The modes for the SOFAR environment are shown in Figure 3. These modes are very similar to the modes of a vibrating string, with the “natural frequencies” of the string vibration given by the horizontal wave numbers associated with the modal propagation. At a given range the complex-valued pressure at a single frequency is found by summing up the contribution from each of the modes of the product of the complex-valued amplitude and mode shape. By Fourier transforming the frequency response the time domain solution is found, Figure 4. The method can be generalized to slightly range dependent environments.

Relative to a seismic network, the hydroacoustic network has the advantage that the energy is generally travelling in the horizontal SOFAR waveguide where the sound speed is known to within a few metres per second and there is little attenuation. Disadvantages are the multi-pathing and the temporal variation of the sound speed. The multi-pathing makes it more difficult to estimate the travel-time. The temporal variations consist mainly of seasonal variations, for Wake giving timing differences of up to about 20 s for the northern paths at a range of about 50 deg, but usually it will be about 2 s [LeB98]. Daily fluctuations might also be important both for energy distribution within the waveform and timing errors (probably less than 1 s).

The labelling of hydroacoustic phases is unusual. Traditionally, T-phases received their name from the Tertiary phases and were labeled according to propagation path, see e.g. [Kul90]. In R1 there is an attempt to label phases according to the source type. While sufficient at present the approach in R1 might be too simple in the future. For example, it is speculated that the propagation path from the event to the ocean in the future could be classified as P or S propagation. The travel-time from the ocean to the T-phase station is of the order of a few seconds and thus the propagation paths from the ocean to the

T-phase station can likely not be identified for each arrival. For T-phases observed on stations further inland this path identification could likely be made, an example of this is in [Kul90].

T-phase stations are designed to record both H- and T-phases, see Figure 1. For modelling the propagation to a T-phase station, the H/T/N waveforms in the SOFAR channel is decomposed into a set of normal modes. As the modes from offshore approach the sloping bottom the higher order modes will cease to propagate in the water column first and couple into the seabottom, while the lower order modes couple into the seabottom closer to the shore. The T-phase station on land will then receive the signal via the seabottom. This coupling from the ocean to the receiver has not been studied in detail, and the quality of T-phase stations is not yet clear to the hydroacoustic community.

The New Zealand hydrophone station is not considered here as it is not part of the IMS and records T-phases poorly as it is located in shallow water [Han98]. The modes propagating in the SOFAR channel in deep water get stripped off and couple into the seafloor as they approach shallower water. In order to record a T-phase on a shallow-water hydrophone this seafloor-bound energy has to re-radiate into the water column. In fact, a geophone on land or on the seabottom is expected to observe T-phases better than a shallow water hydrophone. This is also confirmed by observations [Han98].

The automatic processing system consists of three main modules: Detection and Feature eXtraction (DFX), Station Processing (StaPro) and Global Association (GA). DFX analyses time series in several frequency bands to detect an arrival. When an arrival has been detected relevant features of the arrival are extracted based on processing in seven frequency bands. Using these extracted features, StaPro identifies the arrival as a N, T or H-phase. The N-phases are not used in subsequent processing. In GA hydroacoustic (both T and H-phases) and seismic phases are associated and H-phases can also be time defining, i.e. they may be used in building and locating events. For both automatic processing and processing by analysts three H-phases are sufficient to build an event. However, due to the small number of hydroacoustic stations the phases are normally combined with seismic phases, the specific rules for this are given in Table 13 and 15 of [IDC5.2.1].

During the automatic processing relevant output from the three modules is stored in the ORACLE database. This allows the next module to access the information and the processing can easily be assessed. The final output of the automatic processing is, for R1, stored in the `SEL1` account (Standard Event List 1). The following tables are used in this report, for further details see [IDC5.1.1]:

- `hydro_features`: written by DFX. Contains waveform features extracted in several frequency bands, see Appendix A.
- `arrival`: written by DFX and StaPro. Contains travel-time features and phase type.
- `assoc`: written by GA. Contains arrivals associated to events.

- `origin:` written by GA. Contains the origins of events.
- `origerr`: written by GA. Contains parameters related to the uncertainty ellipse for events.

After the automatic processing the detections and associations are reviewed by analysts to build the Reviewed Event Bulletin. In this process a large number of events are rejected or modified and a few new events are built. The Reviewed Event Bulletin, the REB account, uses the same table names as the automatic processing.

In the `arrival` table H-phases are always labelled ‘O’ whereas they are labelled ‘H’ in the `assoc` table. This seems confusing and in R2 both tables refer to an H-phase as ‘H’ [Bea98R2]. In this report H-phases are only referred to as ‘H’! (O-phases used to indicate an in-water-source determined with more confidence than an H-phase [Han98]. This was not reflected in R1.)

Events and arrivals in the period from 15 May 1998 to 28 May 1999 when R1 was operational (several weeks it was off-line) are considered in the following.

Hydroacoustic processing operates at present independently on each hydrophone station. Often it is convenient to refer to the stations in groups: Wake refers to WK30 and WK31, Ascension to ASC19, ASC21, ASC26, ASC27 and ASC29, and New Zealand to NZL01 and NZL06. In the future part of the hydroacoustic processing will be based on groups of stations.

3 Case study

In the period from 6-11 December 1998 there were two H-phase arrivals at WK30 and none at WK31. The WK30 arrival detected by DFX on Dec. 6, 1998 4:48:26 was identified by StaPro as an H-phase, see Figure 5. On WK31 a phase was detected by DFX at 4:50:00 and identified as a T-phase by StaPro, see Figure 5. Examination of the arrival-time for WK31 reveals that DFX selected a “spike” as the main arrival (the “spike” is discussed later). This detected arrival-time is about 20 s after the arrival of the main part of the T-phase. A 20 s error corresponds to a range error of about 30 km.

From the observed signal DFX forms 7 bandpassed signals by applying a 3 pole Butterworth filter (The bands are: 2-4, 3-6, 4-8, 8-16, 16-32, 32-64, 2-80 Hz). Time series of the band filtered signals are plotted in Figure 6. For each signal, relevant features from the `hydro_features` table are shown.

For each trace an onset time and termination time (both green dashed in Figure 6) are determined based on STA / LTA ratios. Both onset and termination time were well determined for the signals below 32 Hz. DFX could not detect a signal above 32 Hz for WK31. The reason for this is a high level of 60-Hz (and 120-Hz) electrical noise on WK31 [the sampling frequency was 320 Hz]. By applying a band rejection filter in the high frequency bands the arrival is quite evident, see Figure 6, and it is now feasible to detect the arrival in the high-frequency band. Thus it would be beneficial to implement a band rejection filter.

It might be possible to produce more robust estimates by using the envelope of the signal in each band. The envelope is computed as the magnitude of the analytic signal. The analytic signal is a complex signal with the measured signal as the real part and its Hilbert transform as the complex part. For efficient processing it can be computed as a short time average. The envelopes of the signals in Figure 6 are shown in Figure 7a. These envelopes appear as noisy as the original time series. To make the envelopes smoother and reduce the effect of spikes the envelopes are low pass (0.6 Hz) filtered as shown in Figure 7b. The time series of the low pass filtered envelopes appear smoother than the original time series. Use of envelope based processing requires further developments and testing.

It is useful to analyse the arrivals using spectrograms, see Figure 8. The 60-Hz electrical noise is clearly seen on WK31 and to a lesser degree on WK30. There have been some unsuccessful attempts to remove the electrical noise by installing a proper ground on Wake Island. The mean frequency f_m and the standard deviation σ_f of the normalized spectrum $S(f)$ are also displayed. They are defined as

$$f_m = \int_0^\infty fS(f)df \tag{1}$$

$$\sigma^2{}_f = \int_0^\infty (f - f_m)^2 S(f)df . \tag{2}$$

In order to avoid the 60-Hz noise the above integration was only performed in the 0-55 Hz band. It is seen that close to the arrival, both the mean frequency and the standard deviation of the frequency changes. For WK30 a broad frequency (1-90 Hz) signal spectrum is evident at about 250 s in the spectrogram. This is due to clipping of the signal, see Figure 9. A clipped signal can be represented as the un-clipped signal minus the clipped part of the signal. The clipped part of the signal, and thus also the clipped signal, will have a broad frequency spectrum. This causes relatively more energy in the high frequency part of the spectrum.

The peak-to-peak amplitude measured for the unfiltered waveform was 10,867,303 counts for WK30 and a factor ten lower for WK31 (1,109,545 counts). These values does not contain any correction for instrument response. A higher number of counts for WK30 has also been observed for other arrivals. Thus the problem of clipping could likely be eliminated by decreasing the gain for WK30. The cables have an attenuating effect at frequencies greater than about 5 Hz, and increases with both frequency and cable length [Mcc93] (the cable length is a factor two longer for WK31 than WK30).

Obviously, DFX did not recognize the peak around 800 s as a noise spike. The spike on WK31 is not present in the unfiltered time series, see Figure 8. In fact, as this spike is only present in the 8-16 and 16-32 Hz bands, the QC system needs to run on each band filtered signal. A 4 s time interval around the peak for the 16-32 Hz band is plotted in Figure 10.The arrival-time algorithm is formulated so that only the 3 peaks from 800.3 to 800.5 s will influence the arrival-time (see Section 4.2). Detecting / removing such a spike (which might be related to the real arrival) by a QC system is not easy and could harm the real waveform. Preferably, the effect of such a spike should be reduced by signal processing, during the arrival-time estimation.

As the hydroacoustic processing identifies phases according to source type, both waveforms should be identified as the same phase type, but they are identified differently on WK30 (H-phase) and WK31 (T-phase). For WK31 there were no detections in the 32-64 Hz band, preventing an H-phase identification (rule 2 in Appendix C). WK30 has more high-frequency energy due to the clipping, increasing the likelihood of an H-phase identification. It is expected that these arrivals should have been identified as a T-phase on both stations [Mike Clark, personal communication]. This identification is based only on the extracted features. Many T-phases are observed at Wake that are not associated with other phases (likely due to the higher sensitivity of hydrophones and the sparsity of the network).

If the same phase type was identified at both stations it is possible to estimate two possible station-to-event azimuths. The calculation procedure for this is indicated in Figure 11.

Finally, the arrivals at WK30 and WK31 were associated by GA with arrivals at the seismic array at ASAR (Norway) and CMAR (Thailand) in SEL1 at IDC. This combination of stations seems unlikely. The arrivals on WK30 and WK31 were not associated by GA with arrivals from other stations in the `SEL1` database at the pIDC (where a later version of the processing software is used) and the event was not created.

4 Detection and Feature eXtraction (DFX)

DFX performs first a QC on the trace, then generates band filtered time series, performs the detection on these and then extracts the signal features from the waveform.

The hydroacoustic QC system is similar to the seismic QC [IDC5.2.1], but contrary to seismic processing only the basic QC is activated for hydroacoustics. Based on the unfiltered data the basic QC checks for single-sample spikes and data gaps. The conditions that are checked for in the extended QC, as used in seismics, might not be applicable or relevant to hydroacoustics.

A detection of an arrival is declared when, based on eleven frequency bands, the STA/LTA ratio exceeds a certain threshold (2.6-3.0 for Wake) in one of these bands. The bands are: 1-2, 1.5-3, 2-4, 3-6, 4-8, 6-12, 8-16, 12-24, 16-32, 24-48, 32-80 Hz (based on /vobs/idc/config/app_config/DFX/beam/detection/WK30-beam.par). This is in contrast to [IDC5.2.1 p 48] where the detection bands are said to be similar to the seven processing bands (see next paragraph). It does not appear necessary to have different sets of bands for the detection and feature extraction. The STA/LTA approach is used only in the detection and is illustrated in Figure 12. For hydroacoustics three values of the short term window is used: 10 s (detection), 4 s (onset time) and 15 s (termination time). The use of different STA window is reasonable as the STA values are used to extract different physical quantities. The long term time constant is 150 s and is only used in the detection. The LTA values are based on an exponentially tapered window applied to the STA values. As this tapering in principle extends infinitely, the LTA values depend on the starting time of the processing. The detection algorithm has not been investigated.

An arrival is declared if a detection is on any of the eleven bands. DFX attempts to extract features from the signal based on seven frequency bands. The bands are: 2-4, 3-6, 4-8, 8-16, 16-32, 32-64, 2-80 Hz. All bands where onset and termination time can be determined are processed. The onset time is determined from 15 s before to 90 s after the detection time and is first determined where a 4-s STA window exceeds the noise level by a factor 1.2. Thereafter it is improved using a AIC approach [IDC5.2.1]. In each band, features of the signal are extracted in order to identify the phase and to fill the `hydro_features` table. The attributes in this table are given in Appendix A. For a precise description of how these are computed see [Lan97a].

It should be noted that a hydrophone records pressure and a seismometer records velocity. Thus the measured amplitudes on a hydrophone station and a T-phase station cannot be directly compared (i.e. section 4 in Appendix A). In the IDC documentation [IDC5.1.1, IDC5.2.1] it is not made clear that the measured signal for a T-phase station is not pressure. Thus while the processing is identical the measured physical quantities are not identical. The identification parameters (Appendix C) currently used in StaPro are independent of the physical units of amplitudes.

Also the hydrophones on Wake are about 30 years old and are not well calibrated. Thus the absolute levels received at different hydrophones cannot be compared. All comparisons should be based on relative measures as for example done in the empirical identification (Appendix C), where the amplitudes are normalized with the average noise, or relative amplitudes between two frequency bands are used.

The relative number of detections in each frequency band is displayed in Figure 13. When using all arrivals (including 75% N-phases) it is seen that WK31 produces relatively fewer arrivals in the 32-64 Hz and 2-80 Hz band. The reduced number of detections are probably due to the high electrical noise source at 60 Hz. For empirical identification of H-phases a detection in both the 3-8 and 32-64Hz band is required and thus the detections rate is 1.0 in these bands. For the associated H-phases (top in Figure 13) WK31 has a peculiarly low number of detections in the 16-32 Hz band, the reason for which is not clear.

For the extraction of signal features it is important that all noise sources are reduced so that only the relevant signal remains. Hydroacoustics involves higher frequencies than used in seismics. The hydroacoustic stations might experience problems with electrical noise and the Wake hydrophones have this problem. DFX should only be concerned with the physical part of the signal. Before the signals are analysed by DFX a narrow band rejection filter should be used to remove the electrical noise at multiples of 60 Hz (50 Hz). This could be done by permanently having a band rejection filter for the Wake hydrophones. An alternative approach could be to implement an adaptive noise removal filter. It is claimed, that this noise is not a practical problem for processing [Dysart and Hanson, personal communication, 1999]. While this noise might not affect the identification performance directly it will certainly disguise the physics and thus hamper improvements in identification performance. It should not be a problem for detection of an arrival as the detection is done in several bands [Han98].

It is expected DFX would give more stable arrival-time estimates if based on the envelope of the signal instead of the signal itself when timing the hydroacoustic arrival using probability weighted times. The envelopes were computed for the example in Section 3 (Figure 7a). Envelope based processing requires further development and testing.

The `SNR` in the `arrival` table is computed based on the tuple (a row) in the `hydro_features` table that has the maximum *total_energy* for that arrival (see Appendix E) using the following formula:

$$SNR=10\hat{}[(total\_energy - 10*\log(duration) - ave\_noise)/10], \qquad (3)$$

where *duration* = *termination_time* - *onset_time*. For an explanation of the variables see Appendix A. As it is difficult to estimate onset and termination time duration is not a stable measure, it would be preferable to have the SNR independent of the duration. The definition of SNR Eq. (3) depends on the mean squared value of signal and noise components. This definition of SNR is common in narrow band processing. In seismics the SNR depend on the mean absolute value of signal and noise (STA / LTA ratio). Thus the hydroacoustic SNR Eq. (3) and the seismic SNR in the `arrival` table cannot be compared. Only seismic SNR is described in the Database Schema [IDC5.2.1].

It is suggested to change the hydroacoustic SNR so that it is also based on a STA / LTA ratio. The algorithm for this has not been investigated, a possibility is to select the maximum STA / LTA value between onset and termination time based on a 15-s STA window. 15 s is suggested as this will extract a significant portion of the energy in most hydroacoustic signals. This will remove the dependency on duration and the SNR will compare better to seismics.

Contrary to seismics STA / LTA is not used for the SNR calculation and the hydroacoustic SNR value is not used in computation of DELTIM, and is thus not as important as the seismic SNR. (The hydroacoustic SNR is used to disable the lowest quality H-phases as drivers [IDC5.2.1 p 110], see also Appendix F).

4.1 Probability Weighted Times (PWT)

For defining a unique arrival-time several options are in general available. The standard definitions of arrival-times all have major drawbacks. Onset time is not easy to define for emerging signals, the peak arrival-time is very sensitive to noise and the intensity weighted mean is difficult to pick for the analysts. The peak time is the most interesting as the peak is easy to observe. In R1 a measure, probability weighted time, related to the peak arrival-time is used. Here an arrival-time is formed based on how much each time sample is above the estimated noise background, the Probability Weighted Time (PWT), as described in [Gue98, IDC5.2.1 p 227] and summarized in Appendix B. A feature of PWT algorithm is that it obtains both the arrival-time and the arrival-time measurement error.

The principle of the PWT approach is illustrated in Figure 14. Based on an observed signal (red) the probability (blue) that a sample is the true peak is calculated. The true peak is not known as the signal is contaminated with noise. This probability depends only on the ambient noise. For high SNR (Figure 14a) the actual peak is very likely to be the true peak, but for low SNR (Figure 14b) the four peaks are as likely to be the main peak. The average of these probabilities is the probability weighted time and defines the arrival-time. The standard deviation of these probabilities defines the arrival-time measurement error. Logically the arrival-time measurement error becomes larger for lower SNR. The problem with this approach, as shown later, is that all other sources of error are neglected and for high SNR the arrival-time measurement error approaches zero and the arrival-time will approach the time corresponding to the peak of the waveform.

The relation between the absolute value of the observed data $p(t_i)$ on a sensor and the absolute value of the deterministic and unknown signal $s(t_i)$ at a time t_i is assumed to be

$$p(t_i) = s(t_i) + e(t_i) , \tag{4}$$

where $e(t_i)$ is the error term and the errors are assumed only to be additive. Independent of the time sample t_i the error $e(t_i)$ is assumed to be Gaussian distributed with zero mean and variance σ_e^2. Thus for each time sample t_i the observed data $p(t_i)$ are also Gaussian distributed with mean signal $s(t_i)$ and variance σ_e^2. The principle of this model is illustrated in Figure 15. Note, the model does not assume any relation between neighbouring points. Only the largest samples in the waveform, where the signal $s(t_i)$ is much larger than the standard deviation of the error σ_e is of interest in the following. The lower tail of the error distribution corresponding to a negative right hand side of Eq. (4) [$e(t_i) < -s(t_i)$] can thus safely be neglected.

Based on the model in Eq. (4) the probability that a given time sample contains the peak of the signal $s(t_i)$ is computed based on the observed time series. The probability that sample n with an amplitude $p(t_n)$ is the peak is [Gue98] (see also Appendix B):

$$P(t_n) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} dx\ e^{-x^2/2}\prod_{i \neq n} N\left(x + \frac{p(t_n) - p(t_i)}{\sigma_e}\right), \tag{5}$$

where N is the standard normal probability distribution and σ_e is the standard deviation of the errors. The mean of this distribution is the probability weighted arrival-time and it is defined as the arrival-time and the standard deviation as the arrival-time measurement error.

The errors $e(t_i)$ stem from many sources but only two separate error terms are considered here, ambient noise and “signal errors”. Signal errors are intended to capture all errors except ambient noise. In R1 only the ambient noise is taken into account.

The standard deviation of the ambient noise σ_a is estimated from a signal free part of the data before the arrival and is currently computed based on the `ave_noise` data field in the `hydro_features` table [σ_a =10^(*ave_noise*/20)].

Signal errors refer to that the band filtered response depends on the frequency band, filter length, instrument response and other factors. Slight changes in these values will change the observed data. The data with the largest magnitude is most important in the following, as it is most likely to contain the peak. The standard deviation of the signal error is estimated to be $\sigma_s = peak/Z$, where *peak* is the peak of the signal $p(t_i)$. Z is an arbitrary scaling factor that depends somehow on the actual data. Based on the data in the next sections, it has empirically been found that Z=10 is a reasonable choice. It is expected that this value is reasonable for most waveforms.

Combining the standard deviations for the ambient noise σ_a and signal error σ_s the standard deviation of the error σ_e is

$$\sigma_e = \sqrt{\sigma_a^2 + \sigma_s^2}. \tag{6}$$

In R1 only the ambient noise is taken into account and thus

$$\sigma_e = \sigma_a \quad . \tag{7}$$

For many signals the signal error is larger than the ambient noise (when $\sigma_s = peak/10 > \sigma_a$), and thus the standard deviation of the error σ_e is underestimated. This causes, as shown in the next sections, 1) that the peak of the observed data is given too much weight and often defines the arrival-time. 2) The estimated arrival-time measurement error (DELTIM) is underestimated.

The PWT approach does not truncate the waveform to a specified SNR ratio. Thus the strongest part of an arrival will always be most likely to define the arrival. The normalization with σ_e, Eq. (6), as opposed to σ_a, Eq. (7), alone in Eq. (5) will change the probability that a given point is a maximum. Use of σ_e will avoid too much emphasis on the peak of the signal.

4.2 PWT for two T-phase arrivals

Computation of the standard normal probability distribution is an integral part of the PWT approach. The approximations in connection with computing the standard normal probability distribution are sufficiently accurate, see Figure 16 where the method of [Gue98] is compared with the high-order approximation in the Matlab implementation of the standard normal probability distribution. [Gue98] is using a rational approximation based on a six-term denominator [Abr70, formula 7.1.28]. Further, the error function is tabulated to avoid repeated calculations, this is not either expected to be a source of error.

The PWT and the standard deviation as computed by DFX for two T-phase arrivals on WK31 are shown in Table 1 and Table 2. The corresponding band passed time series for the arrivals are shown in Figure 17. The first T-phase arrival was analysed in Section 3 (referred to as the Dec-6 arrival), is an example of a small amplitude T-phase. The second T-phase arrival, from the 20/9/1999 Taiwan earthquake (referred to as the Taiwan arrival) is an example of a large amplitude T-phase.

Table 1: *Arrival-times for the Dec-6 arrival. The time is measured from 912919000 s. Peak is the peak sample amplitude measured in noise levels* σ_a. *A and B refers to PWT estimation using version A and B of the Matlab code.*

Freq. (Hz)	DFX PWT (s)	DFX Std. dev. (s)	A and B PWT (s)	A Std. dev. (s)	A Peak (σ_a)	A K (samples)	B Std. dev. (s)
2-4	782	6	784	4.3	11	544	4.6
3-6	779	3.9	779	3.5	5.5	3137	4.2
4-8	779	.34	779	.009	17	24	.012
8-16	800	.8	790	11.3	21	60	11.3
16-32	777	.65	777	.63	23	15	.64

Table 2: *Arrival-times for the Taiwan arrival. The time is measured from 937852000 s. Peak is the peak sample amplitude measured in noise levels* σ_a. *A and B refers to the PWT estimation using version A and B of the Matlab code.*

Freq. (Hz)	DFX PWT (s)	DFX Std. dev. (s)	A and B PWT (s)	A Std. dev. (s)	A Peak (σ_a)	A K (samples)	B Std. dev. (s)
2-4	954	0.0042	954	0.0025	294	4	0.0025
3-6	962	0.055	961	0.0021	307	4	0.0021
4-8	911	0.0042	911	0.0260	319	4	0.0261
6-12	911	0.0042	911	0.0015	291	2	0.0015
8-16	908	0.0042	908	0	203	1	0.0001
16-32	909	2.26	907	0.0072	26	5	0.078
2-80	911	0.06	919	15.5	27	24	15.6

The noise levels are indicated on the time-series in Figure 17, the maximum signal amplitude measured in noise levels is given in Table 1 and Table 2. For the Dec-6 arrival the signal level is up to 20 times the noise level and for the Taiwan arrival the maximum signal level is about 300 times the noise level. The consideration of noise levels is important as only samples with a signal between the peak and the peak minus four noise levels contribute to the arrival-time determination as implemented in R1.

To analyse the processing further, the PWT algorithm as implemented in DFX was duplicated in Matlab (Version A). The time series was read into Geotool and then bandpass filtered. These bandpass filtered time series could be slightly different from the DFX time series due to differences in implementation of the filtering. The arrival-time estimation should also be so robust that a small change in the filter bands should not change the arrival-time.

A more precise implementation of the version A code was also implemented. This version, version B, does not exclude any samples and the integration bounds [x in Eq. (5)] are much larger, from -6 to +5 standard deviations from the peak (DFX and version A used -4 to +3 samples from the peak). It is necessary to integrate above the observed peak as the true peak might be larger than the observed peak. As an indication of the accuracy the sum of the probabilities for all samples in version A was about 0.98, but in version B a value of 0.9999 was typically obtained. Both sums are quite close to the maximum of 1. Version B requires substantially more computations.

Comparison of the two simulations, Table 1 and Table 2, indicates that the approximations in [Gue98] seem reasonable. The nature of the approximations causes the standard deviations to be slightly underestimated, but based on the comparison of the standard deviation from version A and B this error is not large and it is not expected to be a problem. Other error sources are more important.

K in Tables 1-4 is the number of samples between onset and termination time that is used in the estimation of the standard deviation. In the approach [Gue98] only the K samples 4 noise levels from the peak are considered in computing the PWT, as indicated in Table 1 20-3000 samples out of 16000 samples are sufficient to calculate the PWT.

For the Taiwan arrival the maximum signal level is about 300 times the noise level, see Table 2. Thus the noise is negligible, and only a few samples are needed to estimate the PWT. As only uncertainty due to ambient noise is considered in the PWT-model, a standard deviation close to zero is obtained in all bands, except the 2-80 Hz band where the 60-Hz noise increases the noise level. The arrival-time and standard deviation were obtained using just 1-24 samples out of 55000 samples.

For four of the seven bands the standard deviation of the arrival-time is exactly one sampling interval (0.0042 s), see Table 2. As implemented a standard deviation lower than the sampling time is truncated to the sampling time. In the 8-16 Hz band the amplitude of the maximum sample was 6 ambient noise levels σ_a above the second largest sample, causing just the peak sample to define the arrival-time. Thus, by having such a low noise level σ_a the PWT approach scales the probability of the largest sample so much that the standard deviation of the arrival-time becomes equivalent to the one from a single-sample spike (K=1). Version A uses then just one sample (K=1) to estimate the arrival-time and then the standard deviation becomes exactly 0.

There is a large difference in the standard deviation for the Dec.-6 8-16Hz band pass filtered arrival based on DFX (0.8 s) and the Matlab implementation (11.3 s), see Table 1. The main feature affecting the PWT is the peak at 780 s and at 800 s, see Figure 17a. The band filtered timeseries from DFX and Geotool could be somewhat different, due to differences in implementation. Possibly the difference could be explained by a somewhat higher peak at 800 s for the DFX filtered timeseries. The spike is about 60 samples wide (0.2 s), see Figure 10, but few of the samples are at peak-values. To investigate the influence of the magnitude of the peak, the 60 samples are multiplied with a factor, as indicated in Table 3. Increasing this factor causes the PWT to move from 790 s to 800 s and the standard deviation to decrease from 11 s to 0.13 s. A factor 1.12 gives the same values for the DFX signal as the Matlab signal, and could be a reason for the difference.

Table 3: *The PWT and standard deviation from the version A code for the Dec-6 arrival. 60 samples around t=800s for the 8-16 Hz band are multiplied with a factor as indicated in the first column.*

Factor	PWT (s)	Std.dev (s)	Peak (σ_a)	K (samples)
1.0	790	11	21.1	60
1.05	798	6	22.1	46
1.1	800	1.4	23.2	26
1.12	800	0.6	23.6	26
1.15	800	0.1	24.3	20

Table 4: *The PWT and standard deviation from the version A code for the Taiwan arrival. 3 samples around t=911s for 2-80 Hz band are multiplied with a "factor" as indicated in the first column.*

Factor	PWT (s)	Std.dev (s)	Peak (σ_a)	K (samples)
1.0	919	16	27.3	24
1.05	912	6	28.6	7
1.1	911	1.2	30.0	4
1.12	911	0.6	30.6	3
1.15	911	0.002	31.4	3

There is also a large difference in the standard deviation for the Taiwan 2-80Hz band pass filtered arrival based on DFX (0.06 s) and the Matlab implementation (15.5 s), see Table 2. The main feature affecting the PWT is the peak at 911s and at 950 s, see Figure 17b. This difference in standard deviation could be due to a higher peak at 911 s for the DFX filtered timeseries. To investigate the influence of the magnitude of the peak, just 3 samples around the peak are multiplied with a factor, as indicated in Table 4. Increasing this factor on the 3 samples causes the PWT to move from 919 s to 911s and the standard deviation to decrease from 15.5 s to 0.0018 s.

Thus based on Table 3 and Table 4 it is seen that a change of 15% in the peak levels can change the standard deviation by several decades and the travel-time by many seconds. The present approach is not robust.

Due to the nature of the hydroacoustic signals both onset and termination time are difficult to estimate precisely and a good estimator of arrival-time should be independent of these. Most samples used in the PWT calculation are significantly above the noise level, and the samples around the onset and termination time are usually sufficiently below the peak so that the PWT does not depend much on onset and termination time. The independence of onset and termination time is a good feature of the PWT algorithm.

4.3 Including signal error for the two T-phases

In this section the approach of [Gue98] is augmented by including the signal error σ_s, using Eq. (6), in the Matlab processing.

The total error σ_e when signal error σ_s is included is indicated in Figure 18 for the Dec-6 arrival and the Taiwan arrival. The arrival-time and standard deviation for the two arrivals are given in Table 5. It is seen that the standard deviations are increased. For the two bands (Dec-6 8-16 Hz and Taiwan 2-80 Hz) where the signal had two nearly identical sized peaks the arrival-time is now between the peaks as opposed to biased on one of the peaks. Naturally, an arrival-time between two peaks are not correct in this case. The influence of the second peak could for example be reduced by estimating PWT from a low pass filtered envelope (an example is given in Figure 7). The signal could also be weighted according to where in the signal the true peak arrival is expected to be. It is known that the peak cannot be in the first few seconds after onset and that for longer duration signals it should not be too late in signal (this has not been tested).

For the other bands the PWTs have only changed slightly. This is because the strongest part of the signal is still most likely to define the arrival-time and the bi-modal structure is not too strong. The difference between the PWT with and without signal error is an indication of the stability of the method. For all the bands this difference is within the standard deviations for the PWT with signal error included.

For the Dec-6 arrival a test similar to the one in Table 3 was performed but with the included signal error, Eq. (6). The result of this, Table 6, shows much more stability, the arrival-time changes slowly from 790 to 800 s and the standard deviation decreases only from 11 to 4 s.

In summary, including the signal error reduces the influence of spikes in a physically acceptable way, increases the standard deviation and makes the PWT approach more robust. The method requires further testing.

Table 5: *Including also the signal error in the PWT calculation for the version B code. Peak is measured in error-levels* σ_e.

Freq. band (Hz)	Taiwan PWT (s)	Taiwan Std. dev. (s)	Taiwan Peak (σ_e)	Dec. 6 PWT (s)	Dec. 6 Std. dev. (s)	Dec. 6 Peak (σ_e)
2-4	952	3.5	10	779	5	7.3
3-6	959	8	10	784	4	4.8
4-8	911	1.1	10	779	.2	8.6
6-12	911	0.5	10	788	11	9
8-16	908	0.3	10	777	.7	9.1
16-32	907	4	9.3			
2-80	929	23	9.3			

Table 6: *The PWT and standard deviation from the version A code with also signal error included. The 60 samples around t=800 s for Dec-6 8-16 Hz band are multiplied with a "factor".*

Factor	PWT (s)	Std.dev (s)	Peak (σ_e)	K (samples)
1.0	790	11	9	344
1.05	793	10	9.5	242
1.1	793	7	9.9	180
1.12	798	6	9.9	163
1.15	799	4	9.9	127

4.4 The arrival-time measurement error

The measurement error (DELTIM) is the standard deviation as obtained by from PWT algorithm. Specifically DELTIM is the SIGMA_TIME from the tuple in `hydro_features` table which has the highest energy, TOTAL_ENERGY. Physically, and empirically, it is expected that DELTIM should be larger than 0.5 s and that a typical value is about 5 s for a real T-phase, which will contain many samples of about the same amplitude.

The histograms of the measurement error for each phase type and station are displayed in Figure 19. Most of these arrivals have a value significantly less than 1 s. 380 arrivals have DELTIM =0 in the `SEL1.arrival` table, indicating no uncertainty in the arrival-time, 60 of these arrivals were T or H-phases. According to the PWT algorithm DELTIM cannot be less than the sampling time (Appendix B, approximation 5), so the reason for the DELTIM=0 is not clear.

It is concluded that the currently implemented arrival-time measurement error is unacceptably small.

4.5 Averaging the arrival-time estimates

In R1 the arrival-time and the associated measurement error are defined as the PWT and standard deviation for the band with the highest energy. It is expected that more accurate and robust measures can be found if the arrival-time is based on the observed PWT arrival-times in multiple bands. It could be based on a simple mean of the arrival-time in each band. This might favour low-energy frequency-bands too much. Alternatively, the arrival-time could then be estimated based on a weighted average. This weighting could be based on the energy in each band, or the observed standard deviation in each band. The latter is in close agreement with the PWT approach (Appendix B) and is chosen here. In this approach the mean travel-time t_m and standard deviation σ_m are given by:

$$t_m = \left(\sum_f \frac{t_f}{\sigma_f^2}\right)\sigma_m^2 \qquad \text{and} \qquad \sigma_m^2 = \left(\sum_f \frac{1}{\sigma_f^2}\right)^{-1} . \tag{8}$$

The above procedure was applied to the WK31 observations in Figure 6. The estimated PWT and standard deviation (DELTIM) for each band, as computed by DFX, are given in Table 7. For finding the arrival-time of the phase (to be inserted in the `arrival` table) DFX R1 uses only the frequency band with maximum energy (TOTAL_ENERGY in the `hydro_features` table). That is here the 8-16 Hz band with an arrival-time of 800 s. In all other bands the estimated arrival-time is about 780 s, and based on the time series this estimate is reasonable. By using only the 8-16Hz band an error in the arrival-time estimate of 20 s is obtained. This corresponds to 30 km in range for a sole arrival.

Using Eq. (8) a weighted average is formed as indicated in Table 7. Inspection of Figure 6 indicates that this arrival-time (781 s) is reasonable. The example shows that by averaging it is possible to get a better estimate for the peak in the case of a “spike”. The approach requires further testing.

Table 7: *Arrival-times for WK31. The “modified” refers to the weighting as specified in Eq. (8) and “R1” to the arrival-time as estimated in R1. The time is measured from 912919000 s*

Frequency band (Hz)	PWT (s)	Std. dev. (s)
2-4	782	6.2
3-6	779	3.9
4-8	779	.34
8-16	800	.8
16-32	777	.65
R1	800	.8
modified	781.2	0.28

5 Station Processing (StaPro)

The neural network phase identification is only available for Point Sur (which is not an R1 station). All R1 hydroacoustic stations use empirically derived phase identification rules.

As opposed to seismic phases, which are named according to propagation path, hydroacoustic H and T-phases are named according to the medium of the source. All hydrophones should therefore usually detect the same phase type for the same event. By treating all hydrophones in one group together (Wake hydrophones, and all Ascension hydrophones), it may be avoided that a group detects signals from the same event and they are identified as an H- and T-phase on different hydrophones. As shown in Table 8, it happens quite often.

Figure 20 shows examples of the received waveform at Point Sur and WK30 for the Nov. 10 1997 C4 explosions. 60 C4 SUS-like charges (4 kg Sounding Under Sea charges) were deployed as part of an anti submarine warfare exercise and the exact location is not known [D'Spain, personal communication], but the range to PSUR is about 1 deg while the range to WK30 is about 60 deg and VIB about 30 deg. The upper-right wavelet for WK30 is a classic example of an H-phase. The slow buildup is due to the faster propagating high-order modes and the more energy in the late arrival represents low-order modes with the major energy close to the sound-axis (compare to Figure 2 and 4).

Of the 60 charges 20 are recorded in the REB at pIDC with detections at PSUR, WK30 and VIB (3 hydroacoustic stations needed to build an event). In the automatic processing, SEL1 account, none of the REB detections were at WK30 or VIB. PSUR detected automatically all 20 and classified 18 as H-phases and 2 as N-phases. Examination of the time series revealed that many more H-phases could easily be detected at PSUR. For WK30 17 T-phases were automatically detected in this time interval. Likely some of these T-phases were related to the SUS-charges.

Visually the features at PSUR and WK30 are quite different. It is expected that the phase identification rules should be range dependent. Longer ranges will tend to prolong the signal duration and the signal will have more low frequency energy. Thus a detected arrival could be either a “short range” T-phase or a “long range” H-phase. If these features are important it could be taken into account by letting GA rename phases with certain features and ranges (StaPro does only know the features).

5.1 Algorithm for empirical phase identification

The empirical phase identification (described in Appendix C) is used for all hydroacoustic stations where neural network phase identification is not available, thus the same rules are applied for Ascension, New Zealand, Wake and VIB. The effect of a missing detection in some bands (rule 2 Appendix C) is not clear in [IDC5.2.1, Han99]. In order to identify an H-phase a detection in both the 3-6 Hz and 32-64 Hz band is required, whereas to identify a T-phase a detection only in the 3-6 Hz band is required.

In the present identification rules, H-phases are phases that are not identified as a T or N-phase. If possible, it would be preferable if H-phases are identified by some specific features (like presence of bubble pulse, duration, presence of high frequencies,...).

Due to electrical noise on WK31 only phases with significant energy in the high frequency, 32-64 Hz, band will have detections in that band. Thus rule 2 in Appendix C often determines if the phase is a T or an H-phase. The probability for identifying an H-phase given that there is a T or H detection in the 32-64 Hz band is 0.92. The low detection ratio in the 32-64 Hz band for T-phases in Figure 13 is an indication of this.

5.2 Observed H and T-phases

On VIB no H-phases have been automatically identified (neither by analysts).There was no known H-phase events during R1 (May 1998 to May 1999).

Signals observed on a T-phase station have little energy in the 32-64 Hz band, due to attenuation when propagating from the ocean coupling point to the station. The identification rules are identical for T-phase and hydrophone stations, and require a detection in the 32-64 Hz band to identify an H-phase, making an H-phase identification unlikely on VIB.

The number of H and T-phases for the Wake hydrophones are given in Table 1, based on the `SEL1.arrival` and `SEL1.assoc` table. In addition there were numerous N-phases (15343 on WK30 and 22451 on WK31) and 33 arrivals that were not identified with a phase type (i.e. iphase=”-”). Based on Table 8, it is three times more likely to observe an H-phase on WK30 than on WK31. It would be expected that there should be about the same number of H-phases detected for each station. A reason for detecting fewer H-phases could be the higher level of 60-Hz noise on WK31. This produces fewer detections in the 32-64Hz band. The empirical identification rules do not permit an H-phase to be declared without detection in the 32-64Hz band. A few H-phase signals from WK30 have been examined and similar to the case study, Section 3, they contain a clipped signal. Clipping will increase the high frequency energy and thus increase the likelihood of identifying a phase as an H-phase.

Table 8: *Phase identification and associations of arrivals at either WK30 or WK31 for automatic processing (SEL1). It is compared to the analyst reviewed (REB) events.*

Station	Phase	SEL1.arrival	SEL1.assoc	REB.assoc
WK30	T	5507	1886	304
WK30	H	784	527	0
WK31	T	7535	1916	297
WK31	H	291	227	0

Table 9: *Phase association of arrivals at both WK30 and WK31 for automatic (SEL1) and analyst reviewed (REB) events.*

Station	Phase	SEL1.assoc	REB.assoc
WK30 and WK31	T	945	266
WK30 and WK31	H	12	0

For the analyst reviewed events (right hand column in Table 8) about 15% of the automatic associated T-phases have been used by the analysts. The T-phases are not allowed to be used to locate events. All H-phases are disassociated by the analysts.

Given that the events are far from both hydrophones and the local blockage is minor [See Section 6.2], the same phase should in most cases be detected on both WK30 and WK31. In Table 9 the number of arrivals associated on both WK30 and WK31 is given. It is seen that for automatic identification half (945 / 1886 and 945 / 1916) of T-phases detected at one station are also detected on the other station. For the analyst reviewed events this ratio is about 90% (266 / 304 and 266 / 297). For H-phases only 12 events were automatically identified on both stations.

It is concluded that in the automatic processing the H-phases are inconsistently identified at Wake. There are also inconsistencies in identifying T-phases, but these are relatively fewer than for H-phases (Table 8 and Table 9). The measure of performance used in this report (same phase type on both hydrophones) deviates from the ones reported in [Han98, Han99 p 43], where the performance is based on phases identified by analysts versus automatically identified phases and about 70% (!) were automatically identified correctly.

No special study has been carried out to explain the identification performance, but several factors contribute to the inconsistencies [in expected order of importance]: 1) The identification rules; 2) Equipment characteristics such as clipping and electronic noise; 3) Partial blockage. Blockage will affect the performance, as an arrival will not be identified at both stations.

5.3 Modelling of waveforms

In order to calibrate the empirical identification rules a large data set is necessary and also synthetic arrivals are necessary, as not all propagation conditions are available in the observed data. While the empirical identification rules are so simple that synthetic data is not of major importance, they are necessary in order to train the neural network. Further, such synthetic data are excellent to assess the automatic processing.

In R1 only Point Sur (which is not an IMS station) used neural network phase identification. In the hydroacoustic network the shape of the waveform will vary for each station as they are placed in different ocean environments, giving rise to different propagation conditions. Thus each hydroacoustic station will have its own neural weights. Only data from Point Sur has been used to train the neural network for Point Sur [Lan97b], using both real and synthetic data. The synthetic data were generated by KRAKEN, an adiabatic normal mode acoustic propagation code [Por91].

A database of hydroacoustic arrivals (HydroCAM, [Pul98]) has been created. They seem to use much physical insight and such a database modelling capability seems useful. The synthetic data were also generated by KRAKEN. Such a database could prove useful in training the network. It will be difficult to train a neural net for the T-phase stations due to the limited data available and the difficulty in modelling the propagation. It would require a hybrid modelling approach to model the response at a T-phase station. To compute the long-range propagation for a weakly range-dependent environment an adiabatic normal mode code should be used and for the geometric complex environment near the island a finite difference code could be used. Some effort would be required in interfacing the codes. This computation could easily be done on a workstation.

6 Global Association (GA)

In forming an event the two Wake hydrophones are treated as separate stations. By treating the two stations as a cluster the arrival angle relative to the hydrophone axis can be determined, except for the left-right ambiguity (two azimuths are estimated). To obtain a unique angle three hydrophones must be used.

Due to the uncertainty in how the T-phases couple into the ocean waveguide they are not used in location, but they can be associated with an event. In order to treat all arrivals similarly it might be possible to also include T-phases in the location procedure, though with a large modelling travel-time error. It is expected that as the knowledge improves this modelling travel-time error can be reduced. The benefits will be better location performance and a better understanding of the modelling of T-phases.

6.1 Modelled travel-time

H-phases and T-phases currently use the same travel-time and are based only on propagation in the oceans. Thus for modelling T-phase travel-times the depth of the event is always in the ocean and the initial propagation through the solid earth is neglected even for onshore events. The time difference for propagation though the solid earth is also neglected for T-phase stations.

All hydroacoustic stations except WK30 use the velocity model referred to as "hydro_1d" in the `vmodel` field in the `assoc` table. The corresponding modelled travel-times are in the travel-time tables `iasp91.H` and `iasp91.T`, and they are based on a constant acoustic sound speed of 1485 m/s. The tables are used similarly to seismics. This sound speed is larger than both the default velocity (1482 m/s) and the average sound speed (1482 m/s) based on the 2D travel-time for WK30. The default velocity is used when no travel-time table is available.

For WK30 the velocity model is "hydro_saic". This velocity model is not available at IDC. The corresponding travel-times are the 2D modelled travel-times in /cmss/config/earth_specs/TT/hydro_saic/AUG/WK30.

The 2D travel-time table has been generated using the same spatial resolution as in [Kee98]. They are based on normal mode calculations with the ASPM code [Bur94] and a US Navy water sound speed database discretized at the resolution of 5 nautical miles [IDC5.2.1]. The ASPM code, the sound speed and bathymetry database is not available at the IDC, this complicates the assessment of the travel-times. The travel-times are computed at one sole frequency, 10 Hz. The normal modes are determined based on the ocean sound speed profile and the sediment sound speed, thus the full waveguide is taken into account when computing the travel-times.

The received waveform is determined based on the group velocity and transmission loss for each mode. For every half degree in range and azimuth the modelled travel-time is estimated based on probability weighted travel-times, see Section 4.1. This approach gives both the mean probability weighted travel-time and the standard deviation. In R1 the travel-time modelling error for H-phases is linearly increasing to 5 s at 5 degrees range and is constant at 5 s thereafter (solid red line in Figure 22, the figure is discussed later). In Release 2 this error is the standard deviation of the probability weighted travel-times [LeB98, Bea98R2] (this is both range and azimuth dependent, an example is the solid blue line in Figure 22). In both cases the travel-time modelling error is in the 2D travel-time files.

The reduced travel-time based on a reduction velocity of 1482 m/s (and 1 deg=111.12 km) is displayed in Figure 22. This is similar to plotting the difference between the 2D travel-time and the travel-time based on a constant velocity of 1482 m/s. The maximum reduced travel-time difference is 71 s. The average sound speed is found to be close to 1482 m / s and thus lower than the ones used in the 1D travel-time tables. The average is computed as an un-weighted average over both range and azimuth. This corresponds to the average velocity as seen from WK30. To obtain a global velocity each point should be weighted with the corresponding area that cell covers. From the figure, it is seen that especially the northern and southern travel-times are slower. This is reasonable as they propagate through colder, and thus slower, water.

The overall variations in Figure 23 seem reasonable. To investigate the variations in more detail the reduced travel-times along azimuths 117.5, 118 and 118.5 deg are extracted and plotted in Figure 24. These curves should be reasonably smooth as a function of range and the variation between the curves (variation in azimuth) should also be smooth. There are large fluctuations in the reduced travel-time both in range and in azimuth. It is expected that azimuth 118 deg (green curve) has a glitch around range 75 deg, where the reduced travel-time suddenly drops 10 s. At about range 60 deg, the reduced travel-time for azimuths 118 deg and 118.5 deg (green and blue) move quite suddenly away from the reduced travel-time for azimuth 117.5 deg, this seems unlikely to be correct.

From the 2D travel-time table the travel-time increase as a function of range (step 0.5 deg) is computed, see Figure 25a. Based on these travel-time differences the local velocity between each range cell is computed, Figure 25b, the local velocity is the 0.5 deg range step divided by the travel-time in the 0.5 deg cell and thus appears similar to Figure 25a. It is seen that the large fluctuations are mostly clustered in certain areas. The computed velocity along azimuths 117.5, 118, 118.5 deg are extracted and plotted in Figure 26. Some of this variation can be attributed to variations in the wavefrom between neighbouring ranges, due to different travel path and interference for each mode. For ocean acoustics this scatter of velocities is unlikely.

Based on Figure 25a the maximum, average and minimum travel-time for a 0.5 deg range sector for each azimuth is computed in Figure 25c. The average is un-weighted, and it is observed that it does not fluctuate much with azimuth. The maximum and minimum are nearly mirror images. Often a large fluctuation in travel-time in one direction is followed by one in the opposite direction. These variations are clearly too large, indicating an error in the approach.

It should be noted, that for Release 2 the travel-time error has been reduced by about a factor two, see e.g. Figure 27 which should be compared to Figure 24, or the computed velocities, Figure 28 which should be compared to Figure 26. However, they are still large and more stable travel-times should be obtained.

There could be several reasons for the fluctuations in travel-time:

- The source is located at 10 m depth, and thus outside the main energy in the SOFAR channel. The sound speed in the top ocean will show larger variability than the deeper ocean. Because of the low energy level, the waveforms can be quite sensitive to the small fluctuation in ocean sound speed. Variations in the waveform could then cause travel-time fluctuations.
- The travel-time is only computed at 10 Hz. It is expected that a broad band approach based on Fourier synthesis will increase the stability. This is because small errors at one frequency will be averaged out by the other frequencies.

For cases where the 1D or 2D travel-time tables are not available a default travel-time corresponding to a slowness of 75 s / deg is used for the whole propagation path. This corresponds to an ocean sound speed of 1482 m / s. This applies to both offshore and inshore events. The default travel-time model is referred to as "-" in the `vmodel` field in the `assoc` table. There were 9 associated T-phases from WK30 and 17 from WK31 that used the default travel-time.

There is a large difference between the slowness in the ocean (75 s / deg) and the slowness for the upper earth crust (typically 25 s / deg for P-waves). Thus when part of a path is though the earth this part of the travel-time should preferably be computed based on sound speeds in the earth. For T-phases generated inland this could easily be done.

6.2 Blockage factors

Blockage has two meanings:

- *Full blockage*: No signal will be able to be received beyond a zone of full blockage. For hydroacoustic signals this would be a sufficiently large land mass. Too much energy will be lost propagating through such a land mass.
- *Partial blockage*: For some directions the transmission loss will be so high that few signals will have sufficient energy to be detected. A typical scenario for this could be a seamount sticking into the SOFAR channel. This effect will likely be frequency dependent. Depending on the strength and frequency content of the source signal it might be possible to detect a signal from this direction.

It has been observed that explosions from northwest are observed on WK30, but not on WK31. This was observed for a Japanese refraction experiment on 25 and 26 September 1998 [Han99a]. Examination of the local bathymetry, Figure 29, revealed that no land masses were blocking the signal, but the seamounts north of WK31 will likely cause *partial blockage*. A signal from same direction but with higher energy or different frequency content might be detected on WK31.

In order to remove unrealistic arrivals the code GA checks for *full blockage* during association and after the iterative location. If the uncertainty ellipse of an event has no clear path to a hydroacoustic station with an associated phase, this phase is disassociated.

Blockage is modelled for each hydroacoustic station as the maximum distance of propagation along a set of azimuths. For H-phase this distance is determined based on the coastline from landmasses where small islands have been removed. This distance is stored in the blockage files. An onshore event sufficiently close to the shore can create T-phases and thus for T-phases an additional range of 2 deg is added to the above blockage [Hanson, personal communication]. For stations with 1D travel-time tables this travel-time will be used. The 2D travel-time tables do not cover inland events and the default slowness of 75 s / deg will be used for the entire path.

GA cannot associate hydroacoustics arrivals to an event if the travel-time is larger than 11400 s (`end_time_hydroacoustic` in GA.par). This corresponds to a maximum range of 152 deg, thus all paths are blocked past 152 deg.

GA uses the blockage files in the directory / cmss / config / earth_specs / BLK_OSO / . Here files for each hydrophone at Ascension and New Zealand differ, but the files for WK30 and WK31 are identical. WK30 and WK31 are so far apart (240 km) that they should have different blockage.

Figure 30 shows the blockage maps for ASC26, Wake and PSUR. For Wake the closest blockage is about 30 deg away. There are several narrow paths. For example, for Wake the paths are south of South America, between the New Zealand islands, East of Australia and south of Java.

The correctness of these narrow paths are questionable. Firstly, many of these are through very shallow water, indicating a signal from these paths will be strongly attenuated and are very unlikely to be observed. Secondly, signals for these paths will be strongly affected by horizontal refraction, and thus to be on the safe side the paths should be much wider. Thirdly, these paths are so complicated to model that the modelled travel-times for these paths will likely be wrong. To simplify the problem it is suggested to remove these paths.

As currently implemented, the maximum extent of the 2D travel-time tables should correspond to the distance to the line of blockage. Comparison of the limits for the 2D travel-time table and the blockage grid (red line) in Figure 23 reveals that the blockage grid seems to have been calculated on a coarser grid than the travel-time,

- The azimuths for a few propagation paths based on the 2D travel-time tables are blocked.
- Several (azimuth, range) points, with a non-blocked path, have no 2D-travel-time (when this is the case a default travel-time is used for the whole path).

During the iterative location process there is no check of blockage[IDC5.2.1]. It will likely cause some instability if the travel-times changes from a 2D travel-time to a default travel-time. Thus it is preferable to also have 2D travel-times past the blockage. Interpolated travel-times could be calculated for the shadow of smaller islands.

6.3 A priori timing error

The *a priori* timing error of an H-phase is used in the association (Section 6.4), the iterative location and in the computation of the uncertainty ellipse (Section 6.5). As a T-phase can only be associated, the *a priori* timing error for a T-phase is only used in the association.

The *a priori* σ_T timing error depends on the arrival-time measurement error σ_A and the travel-time modelling error σ_M:

$$\sigma_T = \sqrt{\sigma_A^2 + \sigma_M^2} \quad . \tag{9}$$

The arrival-time measurement error corresponds to DELTIM. The travel-time modelling error is distance dependent. The T-phase travel-time modelling error depends on the H-phase travel-time modelling error [Bea98R1],

$$\sigma_M(\mathrm{Tphase}) = \delta_{\mathrm{coupling}} + \sigma_M(\mathrm{Hphase}) \quad . \tag{10}$$

The coupling term $\delta_{\mathrm{coupling}}$ is set to 40 s and is included to account for the larger travel-time uncertainty due to the unknown location of the coupling into the SOFAR channel for T-phases. Assuming the event depth to be just below the sea bottom and an ocean sound speed of 1.5km/s and a bottom sound speed of 4 km/s the maximum distance of the coupling point from the event is found to be about 40(4-1.5)=100 km.

For the 1D travel-time tables (see Section 6.1) the H-phase modelling error are given by [Bea98R1]:

$$\sigma_M(\mathrm{Hphase}) = \delta t_0 \sqrt{\frac{\Delta}{180}} \quad , \tag{11}$$

where Δ is the distance in degrees. The square root is based on the assumption that the travel-time can be expressed as a sum of N travel-times for N range segments, and each of these travel-times are uniformly identically distributed stochastic variables. Thus the total standard deviation increases as $\sqrt{N}$. δt_0 is the modelling error at 180 deg range and is arbitrarily set to 30 s.

For T-phase stations the unknown coupling point from the oceans into the solid earth will cause increased uncertainty. This uncertainty is not taken into account.

For the 2D travel-time tables the modelling error is in the travel-time files, as described in Section 6.1. The travel-time modelling error for both 1D (green) and 2D (red) is shown versus range in Figure 22. Relative to the 1D travel-time tables the more accurate 2D travel-times has a smaller error for ranges greater than 5 deg.

For reference an example of the range dependent modelling error in Release 2 is also shown (blue) in Figure 22. These are computed based on standard deviation of the probability weighted time for the modelled travel-time [LeB98]. These errors will in a later release also contain two additional terms in order to take uncertainty in water sound speed and intra-seasonal travel-time uncertainty into account [LeB98].

The arrival-time measurement error is the DELTIM attribute in the `arrival` table and has been discussed in Section 4. Most arrivals have a DELTIM value significantly less than 1 s. Due to the sum squared in computing *a priori* time error the values of DELTIM are insignificant relative to the travel-time modelling error (for ranges larger than 5 deg the travel-time modelling error is 5 s).

6.4 Location performance

Phases are associated to an event based on a Chi-square test [IDC5.2.1] p 247. The time residuals for each phase are normalized with the *a priori* timing error. If a detection does not satisfy this test after a location, it is disassociated from the event. The threshold for that test is set by the parameter chi_outlier, which is 0.99 in GA.par. This criterion can be used for both time, slowness and azimuth, and a simple residual limit is not used.

Thus phases with a large *a priori* timing error will more likely be associated with an event. If the *a priori* timing error for a phase type is increased the number of associated phases for that phase type will also increase. This might be an important consideration in modifying the arrival-time measurement error and the travel-time modelling error.

The estimated origins for all events where an H-phase were detected on either WK30 or WK31 are shown in Figure 31 based on the associations in the `SEL1` account. The distribution of the origins of the H-phase events appears to be random.

The distribution of locations for events involving H-phases at Wake as a function of range and azimuth is shown in Figure 32. The solid line indicates the blockage i.e. distances from the receiver to land. The events must have an uncertainty ellipse with nonblocked path to the hydroacoustic station.

The T-phases are not used in the location of an origin, due to the uncertainty in the location where they couple into the ocean waveguide. They can, however, be associated with an event if the time residual is not too large. For automatic processing this is determined by the Chi-square criterion and analysts rejects T-phase association if the absolute value of the residual exceeds 120 s. The time residual is the difference between observed arrival-time and expected arrival-time, where the latter is the origin time plus the travel-time as estimated from the travel-time table. The distribution of all T-phase time residuals is given in Figure 33. It is based on T-phases form WK30, WK31, NZ01, NZ06 and VIB. The residuals are biased towards positive residuals. This was also analysed in [Bea98R1] and they attributed the positive bias to the fact that the 1D travel-times are based on a too fast average phase speed (1485 m / s) and use of a slower average phase speed in the 2D travel-time table reduces this bias. All stations except WK30 use 1D travel-times. The average phase speed for WK30 are closer to 1482 m / s (based on the 2D travel-times).

Figure 34 shows the automatic locations for all origins with associated T-phases. Clearly, the majority of the origins are coinciding with known seismic zones, though some events scattered in the Mid-Pacific are probably spurious. Whether the proper association of T-phases has been achieved is still to be investigated.

The distribution of automatic locations for events with associated T-phases at Wake as a function of range and azimuth is shown in Figure 35. Many of the origins of events with associated T-phases are located close to regions of high seismic activity. Histograms of the station to event azimuth for all T-phase arrivals at Wake are displayed in Figure 36. As WK30 and WK31 are separated by 240 km the azimuth to the origins differ somewhat for the two stations. From this figure, there does not seem to be a particular direction from which arrivals to WK30 or WK31 are blocked.

In the `SEL1` account there are 55 events that are located based on arrivals on just one seismic station and H-phases from two or more hydrophones (WK30, WK31, NZ01 and NZ06). While the phases from the hydroacoustic station belong together, it is in most cases doubtful if the seismic arrival is related to the same source as the H-phases. These mis-associations can be reduced by several methods: 1) Use of either maximum reachable distance for a given seismic magnitude or simple transmission loss rules could eliminate the spurious associations [for the hydroacoustic arrivals cylidrical spreading will be assumed (no radiation out of the waveguide) while for the seismic arrivals spherical spreading should be assumed]. 2) By requiring the same phase to be detected on neighbouring hydrostation 3) By combining the arrival travel-time on the two neighbouring hydrostations to compute an azimuth “interval” which the origin must be located in. [How to compute the azimuth is illustrated in Figure 11]. It should be noted that use of triplets of hydrophones the estimated azimuth can be determined uniquely. Point 2) and 3) can only be applied when both stations detect an event, and this will not happen in case of local blockage.

6.5 Uncertainty ellipse

Of hydroacoustic arrivals, only H-phases contribute to location and the error ellipse. Thus this section is entirely devoted to H-phases. First a qualitative discussion of timing errors is given and then the error ellipse for automatically built events (based on SEL1) with H-phases associated is discussed.

The uncertainty ellipse is the 90% spatial confidence interval, with 90% probability is the true solution within this ellipse. Some background for computing the ellipse for only horizontal arrivals and no azimuthal information is given in Appendix D. When computing the uncertainty ellipse it is assumed that the *a priori* timing errors are perfectly known. In such an approach the value of the time-residuals for the associated phases do not influence the error ellipse. The size of the uncertainty ellipse depends on angular coverage and on the *a priori* timing errors divided by the local horizontal slownesses for all arrivals.

The local horizontal slowness for an arrival is the spatial derivative of the travel-time and for a hydroacoustic phase it is about 1 / 1500 s / m. The local horizontal slowness for a seismic phase is a factor 2-10 smaller than for a hydroacoustic arrival due to difference in sound speed and, in addition a seismic phase will propagate with a large angle to the horizontal. Thus a hydroacoustic arrival would have to have a timing error at least a factor 2-10 larger than a seismic arrival in order that they have about the same contribution to the location error. As discussed below this factor will mostly be in the range 1-4.

The H-phase arrival-time measurement error is significantly less than 1 s, see Figure 19 (it is discussed in Section 4.1). For seismics this value is computed based on SNR and phase type and is in the range 0.7-1.7 s. Taking into account the weighting with slowness the hydroacoustic range-error contribution is much less. The H-phase travel-time modelling error is given in Figure 22, a typical value is 5 s. For seismics the travel-time modelling error is 1-3.5 s. The total timing error is the root mean square of the two contributions. For seismics the timing error will be 1.2-4 s and for H-phases a typical value is 5 s. Taking into account the difference in slowness this indicates that H-phases have a smaller range-error contribution.

In Figure 37 the uncertainty ellipses are plotted for all events with associated H-phases from WK30 and WK31. The minor axis is mostly in the radial direction from the hydrophone stations. This indicates that the H-phases constrain the solution more than the seismic arrivals.

Histograms of uncertainty ellipse area are displayed in Figure 38 for events where T-phases were associated and where H-phases were used in the location. It is observed that distribution of the uncertainty ellipse area for origins involving H-phases are more compact than for T-phases. This is mainly due to the fact that for “H-phase events” the number of defining phases is always small, but for “T-phase events” they vary considerably. Other factors such as azimuthal coverage and distance to event will also influence the area of the uncertainty ellipse. The uncertainty ellipse area for origins where arrivals from either station WK30 and WK31 were associated seem to have similar characteristics.

References

[Abr70] Abramowitz M and I.A Stegun, "Handbook of mathematical functions", Dover Publications Inc., New york, 1970.

[Bra88] Bratt, S. R. and Bache, T. C., "Locating events with a sparse network of regional arrays," *Bull. Seism. Soc. Am*. 78, 780-798, 1988.

[Bea98R1] G. Beall, L. Al-Rawi, W. Fox, R. Le Bras, W. Nagy, M O'Brian, D. Wahl, J. Wuester, and J. Zhang, *PIDC 6.0 Automatic processing of seismic and hydroacoustic data*, CCB-PRO-98/06 1998.

[Bea98R2] G. Beall, D. Brown, D. Brumbaugh, J. Guern, R. Hobbs, H. Israelsson, C. Katz, R. Le Bras, W. Nagy, *PIDC 6.1 Automatic processing of seismic and hydroacoustic data*, CCB-PRO-98/38 1998.

[Bin99] F. Bini-Verona, P.L. Nielsen and F.B. Jensen, *Prosim broadband normal mode model*, SACLANTCEN, to be published, 1999.

[Bru98] D.A. Brumbaugh and R.J. LeBras, *Hydroacoustic signal from a ground-truth data set of marine explosions*, Technical Report SAIC-98 / 3008 Science Application International Corporation 1998.

[Bur94] M.D. Burger, C.E. Boucher, E.M. Daley, W.W. Renner, V.L. Pastor, L.C.Haines and A.I. Eller, *Acoustic system performance model ASPM 4.0A: Users Guide*, Technical Report SAIC-94 / 1000 Science Application International Corporation 1994.

[Gue98] J.J. Guern, *Probability weighted time for hydroacoustic detections*, Center for Monitoring Research, memorandum 12-3-97, also Appendix I of CCB Memo CMR-PRO 98/06 March 1998.

[Han98] J. Hanson, R.LeBras, P. Dysart, X. Yang, D. Brumbaugh and A. Gault "Hydroacoustics at the Prototype international Data Center", in *20th Annual Research Seismic Symposium On monitoring a Comprehensive Test Ban Treaty*, edited by Fantroy *et al.*, *Santa Fee, New Mexico* (1998).

[Han99] J. Hanson, "Hydroacoustic Processing", Presentation at IDC, March 22, 1999.

[Han99a] J. Hanson, E-mail 6 October 1999. "As for blockage at Wake, there are two observations. First T-phases are sometimes observed on one hydrophone but not the other. This varies as a function of azimuth. For WK31, blockage is most pronounced to the northwest. The document you refer to (SAIC-3008) covers the 1996 experiment, but a similar experiment was conducted in September 1998 when WK31 was operational."

[IDC5.1.1] Science Applications International Corporation, Pacific-Sierra Research Corporation, *Database Schema (Part 1, 2, and 3), Revision 1*, SAIC-99/3009, PSR-99/TN1142, 1999.

[IDC5.2.1] Science Applications International Corporation, *IDC processing of seismic, hydroacoustic and infrasonic data*, SAIC-99/3023, 1999.

[IMS98] International Monitoring System Division, *Technical Training Program*, Comprehensive Nuclear-Test-Ban Treaty Organization, Vienna 1998.

[Jen94] F.B. Jensen, W.A. Kuperman, M. Porter and H. Schmidt, *Computational ocean acoustics*, American Institute of Physics, New York,1994.

[Jor81] T.H. Jordan and K.A. Sverdrup, “Teleseismic Location Techniques and their Application to Earthquake Clusters in the South-central Pacific,” *Bull. Seism. Soc. Am.,* 71, 1105-1130, 1981.

[Kee98] R.E. Keenan and P. Dysart: *Probability weighted travel-time delivery,* Center for Monitoring Research, Appendix A of CCB-PRO-98/34 July 1998.

[Kul90] O. Kulhanek, *Anatomy of Seismograms,* Elsevier, Amsterdam/New York, 1990.

[Lan97a] H. Laney, *Hydroacoustic feature extractor (DFX-H versions 1), Technical description*, Technical report CMR-97/06, 1997.

[Lan97b] H. Laney, *StaPro Neural Network training data base description: Point Sur California*, Technical report CMR-97/07, 1997.

[LeB98] R. Le Bras, J. Guern, P. Dysart, D. Brumbaugh and P. Dysart, *Seasonally-varying hydroacoustic travel-time tables*, CCB-PRO-98 / 34 July 1998.

[Mcc93] C.S. McCreery, F.K. Duennebier and G.H. Sutton,“Correlation of deep ocean noise (0.4-40 Hz) with wind and the Holu spectrum--A worldwide constant,” J. *Acoust. Soc. Am*. 93, 2639-2648, 1993

[Med97] H. Medwin and C.S. Clay, Fundamentals of Acoustical Oceanography. Academic Press, San Diego, 1997.

[Por91] M. Porter, *The Kraken normal mode program*, SACLANTCEN SM--245, 1991.

[Pul98] J. Pulli, E. Dorfman, T. Farrell and J. Angell “Modelling hydroacoustic waveform envelopes using normal modes”, in *20th Annual Research Seismic Symposium On monitoring a Comprehensive Test Ban Treaty,* edited by Fantroy et al., *Santa Fee, New Mexico*, 1998.

Appendix A: Hydro_features table

Below is a short description of the `hydro_features` table. For a description of how the features are computed see [Lan97a]. T-phase stations measure velocities (vertical or 3 component). Thus the physical quantity is not pressure. The reported quantity for T-phase stations is not known, but for seismic stations the amplitude is reported in displacements in units of nm. If this is correct all places where pressure or μPa are mentioned they should be replaced with displacement and nm**,** respectively**.**

1 Administrative

1. ARID: Arrival ID.
2. LDDATE: Load date.

2 Filter type

3. LOW_CUT: The low frequency cutoff value.
4. HIGH_CUT: The high frequency cutoff value.
5. FORD: Filter order.
6. FT: Filter type (usually a Butterworth).
7. FZP: Filter causality. Usually a zero-phase causal filter is used.

3 Pulse window

8. ONSET_TIME: Time, in epoch seconds, where the signal is estimated to begin. It is first based on a sliding average power window. This onset time is further refined using an Akaike Information Criterion (AIC) adjustment.
9. TERMINATION_TIME: Time, in epoch seconds, where the signal is estimated to end. It is also based on a sliding power window, however, with many refinements [Lan97a].

4 Energy and amplitude measures

10. PEAK_TIME: Identical to PROB_WEIGHT_TIME.
11. PEAK_LEVEL: The level in dB re 1 μPa at the max peak.
12. TOTAL_ENERGY: Time integral of the pressure squared between ONSET_TIME and TERMINATION_TIME. It is the sum of squares of the trace corrected for the estimated noise and divided by the sampling rate. Expressed in dB re $1(\mu\mathrm{Pa})^2\mathrm{s}$.
13. NUM_CROSS: Number of times the pressure squared exceeds the AVE_NOISE by a factor noise_onset. The noise_onset is defined in the hydro-rec.par file and is currently 1.2
14. AVE_NOISE: The average per sample of the squared pressure in the noise segment (15s long), expressed in dB re 1 μPa. Computed for a part of the signal 30 s before the ONSET_TIME.

5 Time measures

15. PROB_WEIGHT_TIME: The probability weighted time, based on [Gue98]. Measured in seconds.
16. SIGMA_TIME: The standard deviation for the PWT time, based on [Gue98]. Measured in seconds.
17. MEAN_ARRIVAL_TIME: mean arrival-time of the estimated signal energy (first moment, in time).
18. TIME_SPREAD: The RMS time spread of the estimated signal energy (second moment, in time).
19. TOTAL_TIME: Total time, in seconds, where the signal exceeds the noise_onset threshold.
20. SKEWNESS: Skewness a of the estimated signal energy (third moment, in time).
21. KURTOSIS: Kurtosis of the estimated signal energy (fourth moment, in time).

6 Cepstrum analysis

22. CEP_VAR_SIGNAL: variance of the cepstrum.
23. CEP_DELAY_TIME_SIGNAL: Time in seconds to the largest value in the cepstrum
24. CEP_PEAK_STD_SIGNAL number of standard deviations from the mean to the largest amplitude.
25. CEP_VAR_TREND variance of the detrended cepstrum
26. CEP_DELAY_TIME_TREND: Time in seconds to the largest cepstrum value for the detrended spectrum
27. CEP_PEAK_STD_TREND number of standard deviations from the mean to the largest amplitude.

Appendix B: Algorithm for probability weighted time

The first part of this appendix is based on [Gue98, IDC52.1 p.227]. From the band passed time signal the following observations are available

1. $p_i = i^{th}$ sample absolute magnitude of pressure

2. Δt = sampling interval

3. σ_e standard deviation of the observed data, which in R1 is identical to the ambient noise level $\sigma_a = \sqrt{Q}$, where Q is the average squared noise level per sample as defined in Eq. (E1).

The probability that any particular sample n is the peak is [Gue98]:

$$P(t = n\Delta t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} dx\ e^{-x^2/2}\prod_{i \neq n} N\left(x + \frac{p_n - p_i}{\sigma_e}\right) \quad , \tag{1}$$

where N is the standard normal probability distribution. Eq. (1) is identical to [Gue98, Eq. (13)], except that his “errf” has been replaced with the more common notation “N”. They are both defined as (see also [Gue98, Eq. (13)])

$$N(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} dx\ e^{-x^2/2} \quad ,$$

which is related to the error function *erf* as follows [Abm70, formula 7.1.1 and 26.2.2]

$$N(x) = \frac{1}{2}\left(1 + \mathrm{erf}\left(\frac{x}{\sqrt{2}}\right)\right). \tag{2}$$

The peak of P(t) will be the sample with the greatest measured pressure. But this peak has a problem very similar to mode jumping. Therefore, [Gue98] proposes to use a time averaged over P(t)

$$\text{Probability averaged peak} = \langle t\rangle_P = \int_{-\infty}^{\infty} tP(t)dt \tag{3}$$

$$= \sum_n P_n t_n \tag{4}$$

$$= t_0 + \Delta t\sum_n P_n n \tag{5}$$

Note that if two samples are near to one another in intensity, the corresponding P_i s will be nearly equal, and so $\langle t\rangle_P$ will be between the samples. Also, if sample j is clearly dominant, $P_j \approx 1$ and so $\langle t\rangle_P \approx t_j$.

It is straightforward to calculate the standard deviation, which is used as the measurement error.

$$\sigma_{\langle t\rangle} = \sqrt{\sum_n P_n t_n^2 - \langle t\rangle_P^2} \tag{6}$$

Note that for a high SNR, if any one sample, i, dominates, then $P_i \to 1$ and so the measurement error becomes very small. By contrast, for low SNR, the P_n s become more evenly distributed and the measurement error grows larger. So, even for identical signals, different values of the measurement error will be calculated for different noise levels, which is appropriate.

Approximations:

The approach described above is quite CPU intensive. The following approximations are used for in computing the probability in Eq. (1):

1. The peak, $\max(e_i)$, of the time series between onset and termination time is found. Only samples $e_j > \max(e_i) - 4\sqrt{Q}$ are used, i.e. only the samples with values greater than four noise levels from the peak.
2. The error function is tabulated, and saved in memory for the next calculation.
3. The infinite integral is evaluated in the interval x= [max(0, -4$+\max(e_i)$), 3$+\max(e_i)$)].
4. The approximations will result in a set of probabilities that sum up to slightly less than 1.0 (in the cases examined 0.98 was typically obtained). The probabilities are renormalized so that the sum becomes 1.0.
5. If the standard deviation, Eq. (6), becomes less than the sampling time it is set to the sampling time.

Appendix C: Algorithm for empirical phase identification

The algorithm below is used for all hydroacoustic stations where neural network phase identification is not available, thus the same rules are applied for Ascension, New Zealand and Wake hydrophone stations and for the VIB T-phase station. The algorithm is extracted from the R1-source code and corresponds, except for rule 2, to the description in [Han99, IDC5.2.1]. The limiting values are, at present, hard coded (no parameter file is used).

1) For a given arrival the low frequency band (LFB) is the 3-6 Hz band and high frequency band (HFB) is 32-64 Hz band. The following measures are then computed based on the `hydro_features` table (for definition of the variables from the `hydro_features` table see Appendix A):

- EnergyRatio: total_energy(HFB) -total_energy(LFB)
- Duration: termination_time(LFB)-onset_time(LFB)
- TimeSpread: total_spread(LFB)
- FractionalTime: total_time(LFB) / duration
- CrossingDensity: num_cross(LFB) / duration

2) If there are no HFB detection the arrival cannot be an H-phase (EnergyRatio is set to a large negative number). If there is no LFB detection the arrival is a N-phase.

3) Based on these five measures the arrival is a T-phase if ALL the following rules are satisfied

- EnergyRatio < -15.5 dB
- TimeSpread>5 s
- CrossingDensity>12

4) If not a T-phase the arrival is a N-phase if ANY of the following rules are satisfied

- EnergyRatio < -15.5 dB
- Duration<6 s
- TimeSpread>35s
- FractionalTime<0.4
- CrossingDensity>20
- Or features missing so that the 5 measures could not be computed.

5) The remaining unidentified phases are identified as H-phases.

Appendix D: The uncertainty ellipsoid

The uncertainty ellipsoid as introduced in [Jor80] and [Bra88] combined both *a priori* and *a posteriori* error estimates. Thus the size of the uncertainty ellipsoid was F-distributed. The *a priori* errors represent the errors before localization is attempted and consist of arrival-time measurement error and travel-time modelling error. The *a posteriori* errors are the based on the actual timing residuals for the obtained location. In R1 the *a posteriori* errors are neglected in computing the uncertainty ellipsoid and the F-distribution reduces to a Chi-squared distribution. An uncertainty ellipsoid computed assuming only *a priori* errors is called a coverage ellipsoid. The 90-% coverage ellipse is used.

In R1 the points on the p per cent confidence ellipsoid are computed as

$$(\mathbf{x} - \mathbf{x}_e)^T \mathbf{V}_x^{-1} (\mathbf{x} - \mathbf{x}_e) = \chi_p^2(M) \tag{1}$$

where $\mathbf{x}_e$ and $\mathbf{x}$ are the points on the ellipsoid and the solution, respectively. For fixed depth solution they contain the [x y t] coordinates and when the depth is included in the location they contains additionally the z coordinate. $\chi_p^2(M)$ is the Chi-square distribution with M number of degrees of freedom, see later. $\mathbf{V}_x$ is the parameter correlation matrix and is given by (for further details see e.g [Jor80])

$$\mathbf{V}_x^{-1} = \sum_{i=1}^{N} a_i^T a_i \qquad a_i^T = \sigma_i^{-1} \begin{bmatrix} \frac{\partial}{\partial x} \\ \frac{d}{dy} \\ \frac{d}{dt} \end{bmatrix} t_i \tag{2}$$

where N is the number of arrivals used in the location, t_i and σ_i are the arrival-time and the *a priori* error for arrival i. It is seen that $\mathbf{V}_x^{-1}$ depends on the slowness and the *a priori* error for each arrival.

M is the number of degrees of freedom for the location problem. For fixed depth 3 independent parameters are determined [x y t] and thus M=3. When depth is included in the location 4 parameters are determined.

Often only the confidence interval for each parameter is of interest. This is the marginal confidence interval and it is obtained by projecting the uncertainty ellipsoid onto each axis. This marginal interval can be determined based on only the diagonal value of $\mathbf{V}_x$ for that parameter, i.e. $\sqrt{\chi_p^2(M) V_{x,ii}}$. Note, that the degree of freedom M is unchanged.

Similarly, the spatial coverage ellipse can be determined by using the block diagonal of the parameter correlation matrix. Again, the degrees of freedom should not be changed.

For fixed depth the R1 uncertainty ellipsoid is always decoupled into a spatial ellipse and time uncertainty. This seems reasonable, but when doing this decoupling the degrees of freedom is also reduced, i.e. M=2 for the coverage ellipse and M=1 for the time error. This will give an optimistic spatial uncertainty ellipse as the number of degrees of freedom should not be changed. This gives that the area of the ellipse should be a factor larger $\chi^2(3)/\chi^2(2) = 1.34$ than the one used in R1.

Appendix E: Total energy, average noise and SNR

The total energy is used first to define the channel on which the travel-time is estimated and later to identify the signal type.

The average squared noise level per sample, Q, is calculated for a time window (by default 15 seconds long) 30 s prior to the onset_time:

$$Q = \frac{\sum_{i=1}^{N_{noise}} x_i^2}{N_{noise}} \tag{1}$$

x_i is the pressure samples and is measured in μPa. Q is thus measured in $(\mu Pa)^2$.

The average noise level in dB is defined as

$$ave\_noise = 10\log Q . \tag{2}$$

The unit of *ave_noise* is dB re 1 μPa.

Total energy is the integral of the signal squared and is estimated by calculating the sum of the received waveform squared minus the average noise:

$$total\_energy = 10\log\left(\sum_{i=1}^{N}(x_i^2 - Q)\Delta t\right) , \tag{3}$$

where x_i is the *i*th sample of the N data samples between onset and termination time. This is similar to the time integral of pressure square (tips) commonly used in sonar processing [Med98]. Using the definition of

$$duration = onset\_time - termination\_time = N\Delta t , \tag{4}$$

$$total\_energy = 10\log\left(\frac{\sum_{i=1}^{N}(x_i^2 - Q)}{N} duration\right) . \tag{5}$$

The unit of *total_energy* is thus dB re 1 $(\mu Pa)^2 s$. While the *total_energy* is proportional to energy it is more correct to term this tips. The unit is the IDC descriptions [IDC5.1.1, IDC5.2.1] is wrong.

The signal-to-noise ratio SNR is defined as the ratio of mean-squared values of signal and noise components. Using the assumption that the signal and noise are uncorrelated, the following expression can easily be derived

$$\mathrm{SNR} = \frac{\left(\sum_{i=1}^{N}(x_i^{\,2} - Q)\right)/N}{Q}. \tag{6}$$

Expressed in terms of *total_energy* and *ave_noise* this becomes

SNR=10^[(*total_energy*-10*log(*duration*)- *ave_noise*)/10] (7)

This is the value written to the data base, it is a dimensionless variable. Similar to the seismic SNR it is not expressed in dB.

It is common in ocean acoustics to use dB. But since the seismic processing does not use dB and the units and variables used here are unusual, it might be easier to not use dB.

In seismics and in detection processing the signal-to-noise ratio is defined as the ratio between the STA and LTA values. The definition used in hydroacoustic feature extraction, Eq. (6), is quite different as it depends on amplitude squared.

Both SNR and *total_energy* depends on *duration*. As duration is not well defined for hydroacoustics, it would be preferable to make the definition independent on this.

Appendix F: SNR limit for H-phase drivers

From the hydroacoustic arrivals only arrival-time information can currently be extracted, no azimuth information is available. Thus an exhaustive search over all grid points is required when building an event with an H-phases as a driver. If many H-phases are detected this can potentially cause an overload. In order to avoid this the following rules are applied in GA_DBI (see man pages and code).

1. Only H-phases with SNR>0.1 (hydro_snr_tresh in GA_DBI.par) can be drivers. The SNR is defined in Eq. (E6). The threshold is set quite low, so that only false H-phase are eliminated.
2. If there is more than 5 (hydro_max_H_per_hour in GA_DBI.par) H-phases with SNR>0.1 from a station then only the 5 arrivals per hour with highest SNR can be drivers.

The above does only apply to H-phase drivers, all H-phases can be associated and time-defining.

There were 51 H-phases with SNR<0.1 in R1. The formula for calculating the hydroacoustic SNR is given in Eq. (3). The SNR<0.1 is typically obtained when the estimated duration of the signal is much larger than the actual signal. This was observed when signals from two arrivals are combined into one arrival.

According to Le Bras (personal communication): “We implemented this feature after testing with various amounts of synthetic H phases and in anticipation of the case of seismic reflection or refraction surveys (for which we had actual examples). It would also apply to bogus H phases if there are examples of large densities of them, but I am not aware of such cases.”

These rules becomes mute when azimuth information is present. In fact this was one of the incentives behind the R3 hydro_azimuth requirements and the fact that there will be triplets of hydrophones (Le Bras, personal communication)

FIGURES

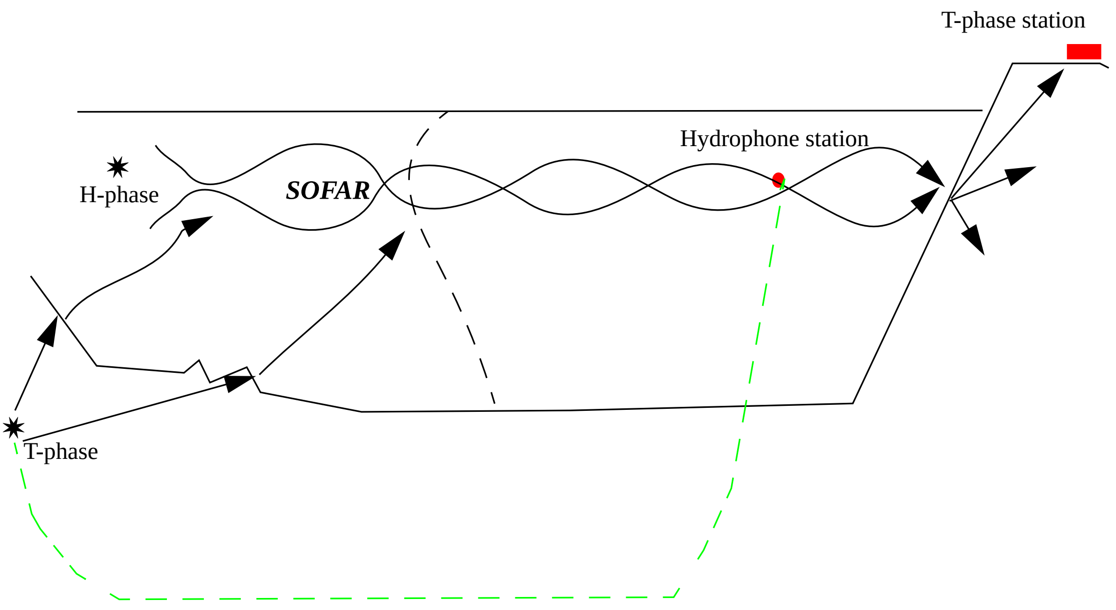

Figure 1: *H and T-phase propagation and their detection on a hydrophone or T-phase station. The green dashed line indicates an arrival that has not propagated through the SOFAR channel. Such an arrival is possible, but automatic processing is not tuned to detect it.*

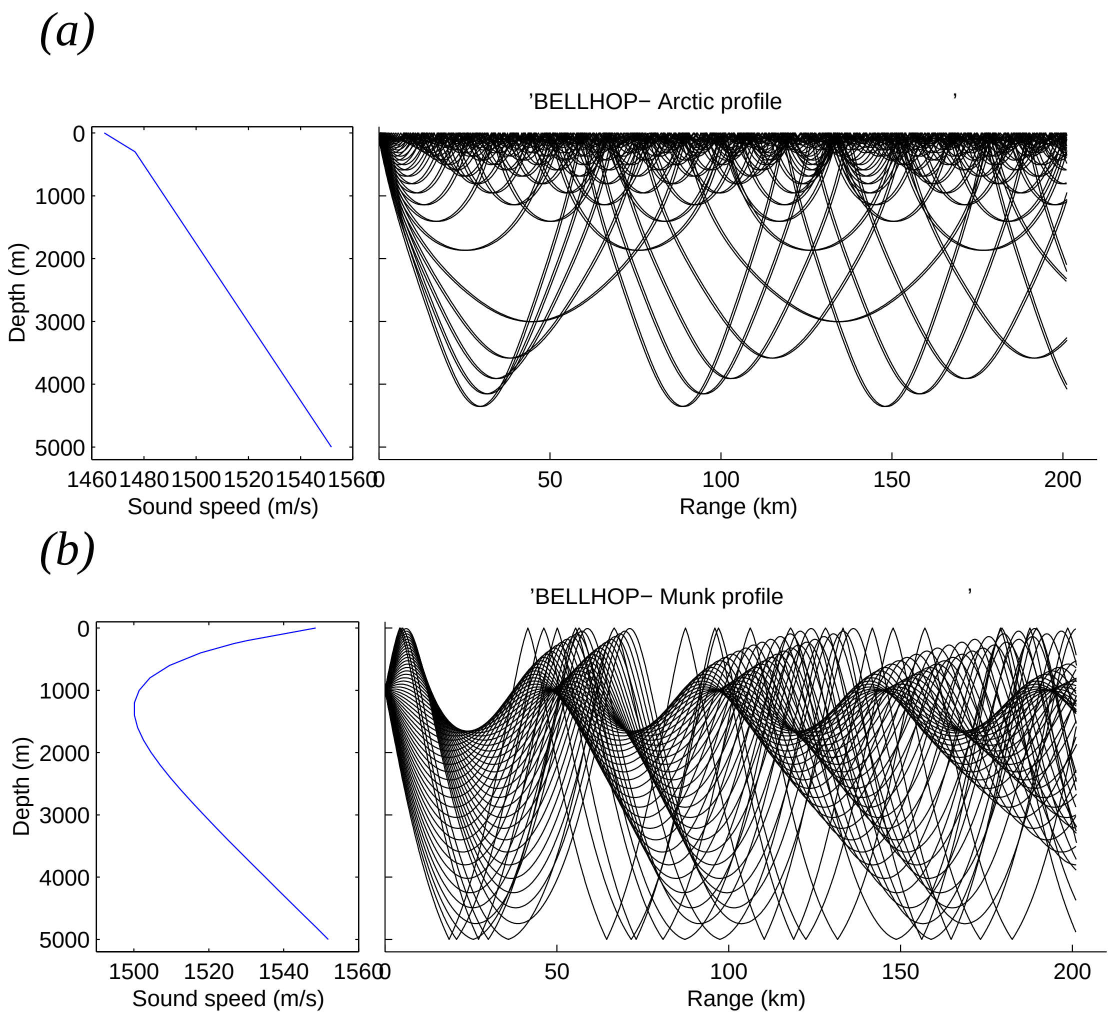

Figure 2: *Examples of raytracing in (a) Arctic environment and (b) SOFAR environment. The ocean sound speed profiles are to the left. The rays are computed using Bellhop [Por91]. In the arctic environment the sound speed is minimum at the surface and increases linearly with depth, mainly due to increased pressure. In the arctic environment all rays are trapped close to the surface. Due to the interaction with the surface the waveforms will be more attenuated than in a SOFAR environment. In the SOFAR environment the sound speed initially decreases, due to decreasing temperature. For the deeper ocean the temperature is constant and the sound speed increases due to increasing pressure. The SOFAR environment has typically a minimum sound speed close to 1000 m depth. The rays in the SOFAR channel can travel for long distance without interaction with the ocean boundaries, and thus the received signal will not be much attenuated. Even though the steeper rays travel longer distance they arrive before the shallower rays, as they spend more time in faster media.*

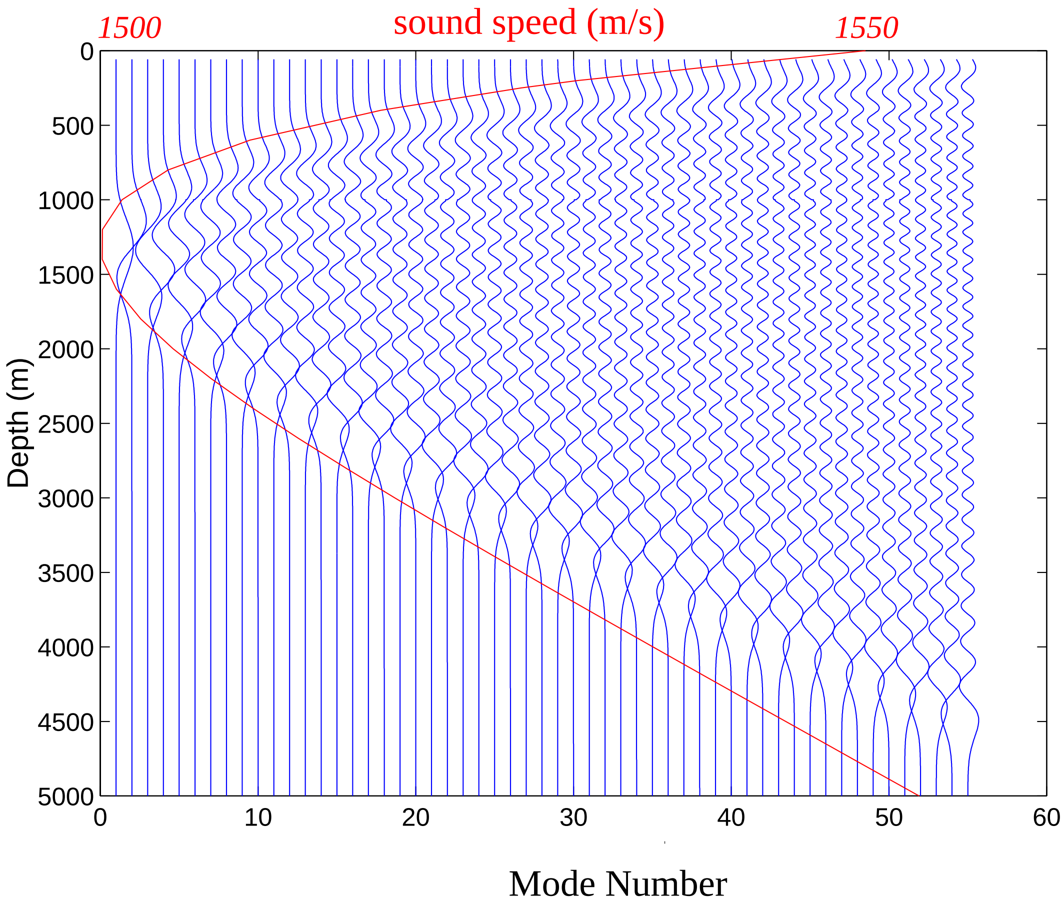

Figure 3: *Normal modes for a SOFAR environment. The ocean sound speed (red curve) illustrates how the modes are centered around the minimum sound speed. Each mode has a different group velocity and the lower order modes travel slower. The modes are computed using Kraken [por91].*

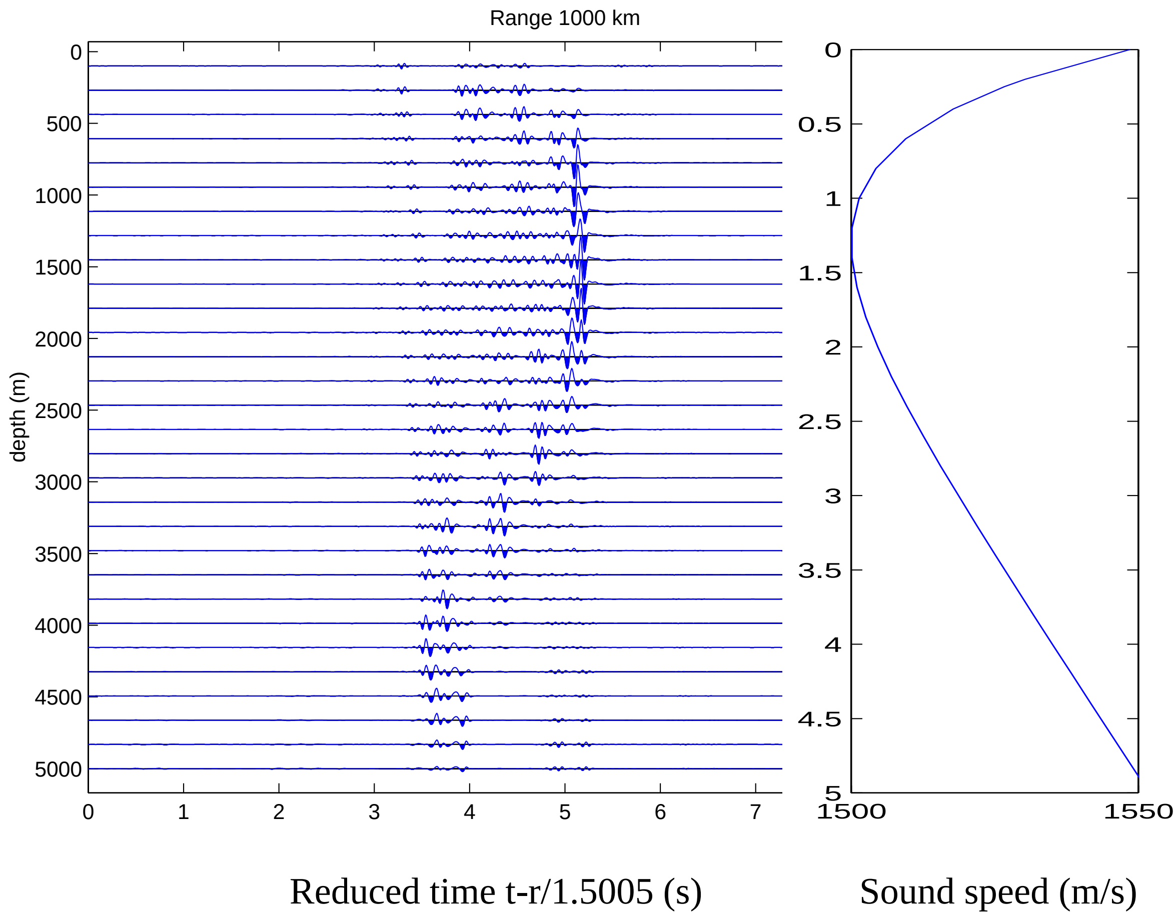

Figure 4: *Time series of the modelled received signal for a vertical array in a SOFAR environment. The source is at 1000 m depth, giving maximum energy in the waveguide. First the high order modes arrive, they contain energy across most of the ocean depth. The lower order modes travel slower, are trapped closer to the sound-speed axis and contain most of the energy. The slow build up of energy for the depth 700-2500 m is characteristic of a source close to the ocean sound-speed axis in a SOFAR environment. The normal mode solution is computed using the prosim acoustic propagation code [Bin99]. The time series are obtained by Fourier synthesis of the single frequency components.*

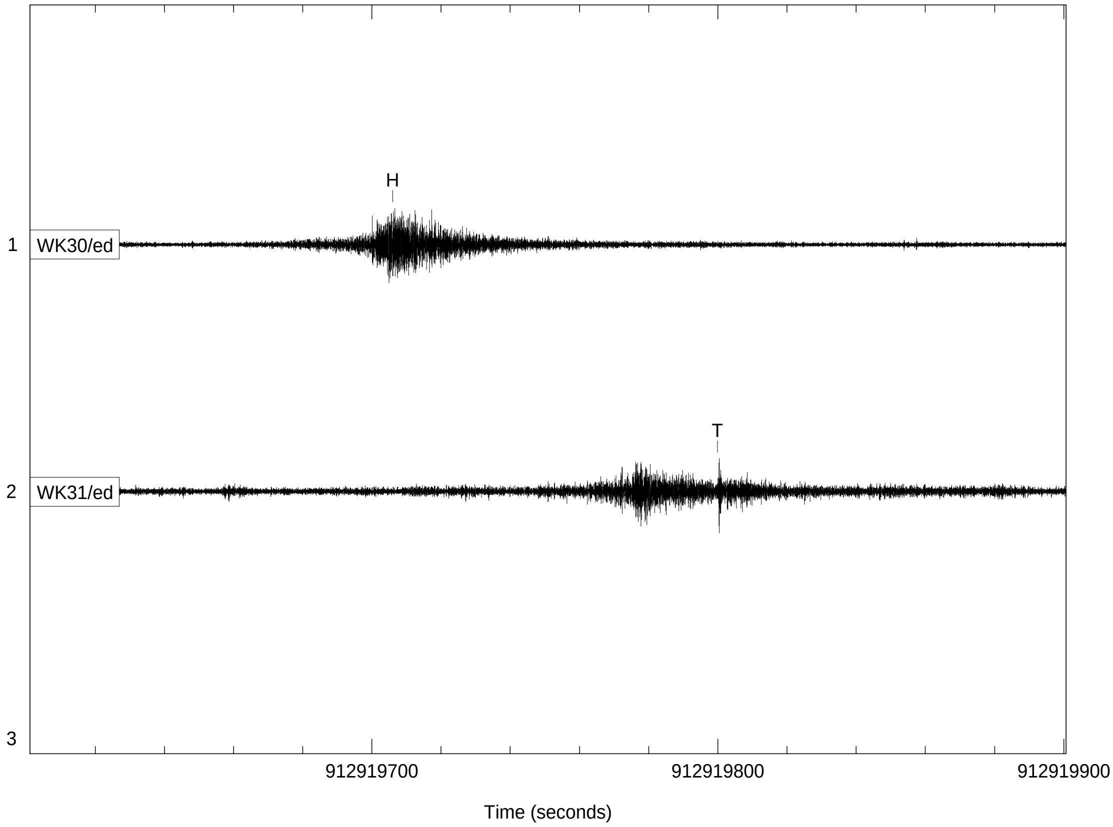

Figure 5: *The arrivals at WK30 and WK31 on 6 Dec. 1998 (orid 100717). A 3-20 Hz 3 pole Butterworth filter is used. StaPro identified the arrival on WK30 as an H-phase and the arrival on WK31 as a T-phase.*

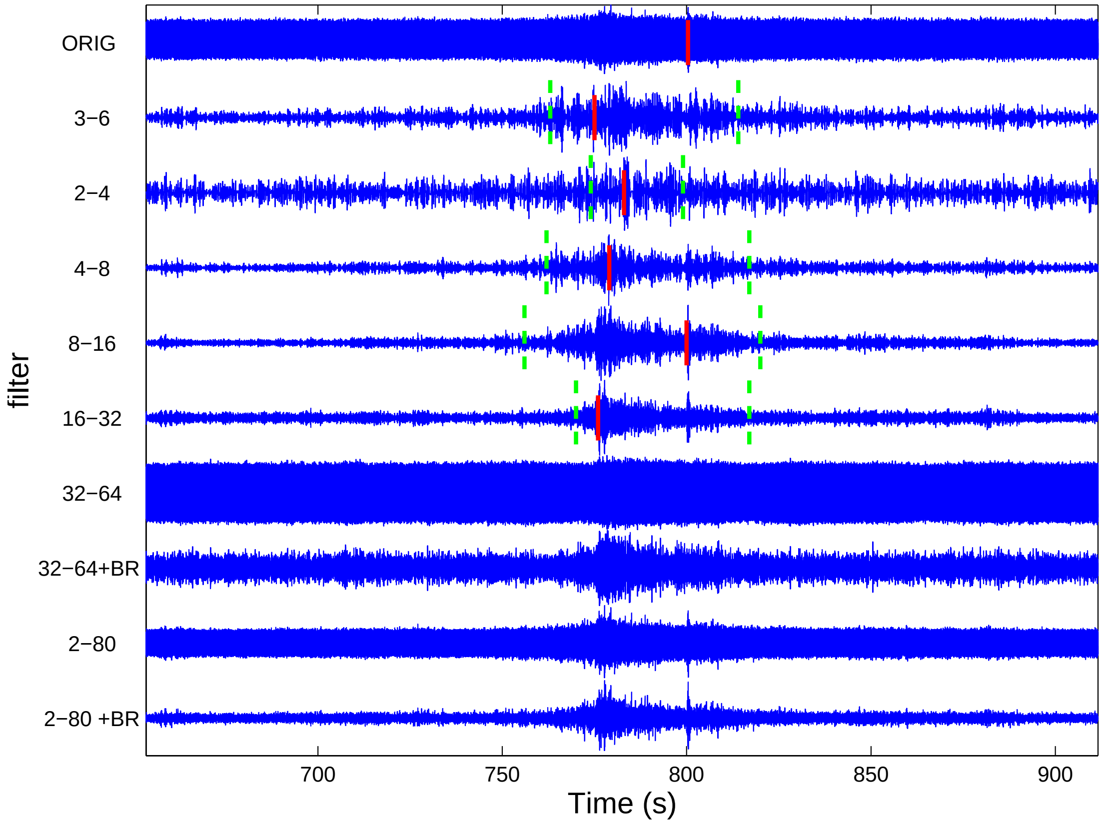

Figure 6: *Filtered signal from WK31 for the 6 Dec. 1998 arrival (arid=1573660). The Y-axis label indicate the lower and upper bandpass filter frequency, "ORIG" refers to the unfiltered trace and "BR" indicates that a band rejection filter additionally is used around 60 Hz. The green dashed lines indicate the onset and termination times for each frequency band as determined by DFX. The solid red lines are the PWT arrival-times for each frequency band as determined by DFX.*

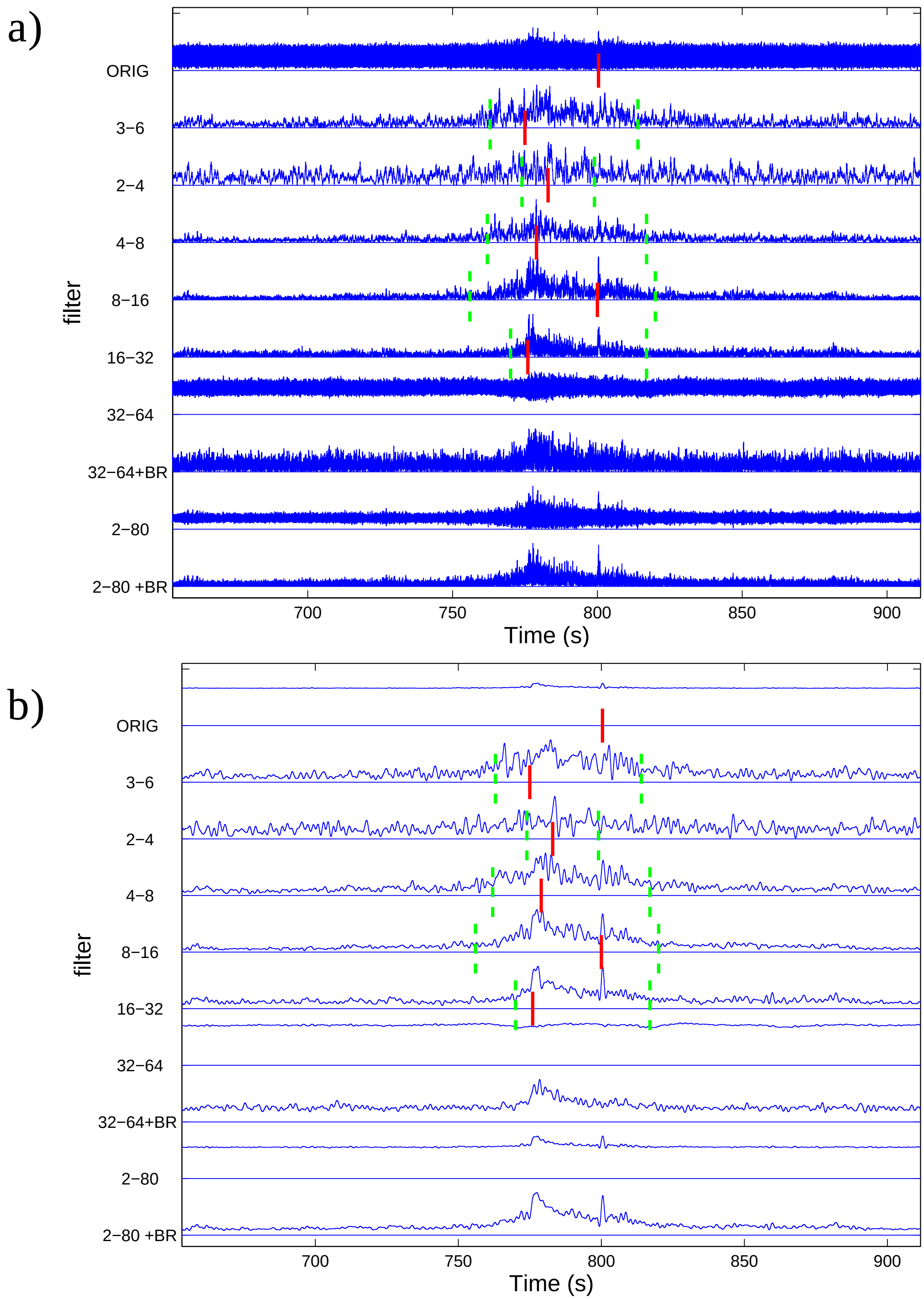

Figure 7: *Envelopes of filtered time series from WK31 for the 6 Dec. 1998 arrival, a) unfiltered envelopes and b) low pass (0.6 Hz) envelopes. The Y-axis label indicate the lower and upper bandpass filter frequency, "ORIG" refers to the unfiltered trace and "BR" indicates that a band rejection filter additionally is used around 60 Hz. The green dashed lines indicate the onset and termination times for each frequency band as determined by DFX. The solid red lines are the PWT arrival-times for each frequency band as determined by DFX.*

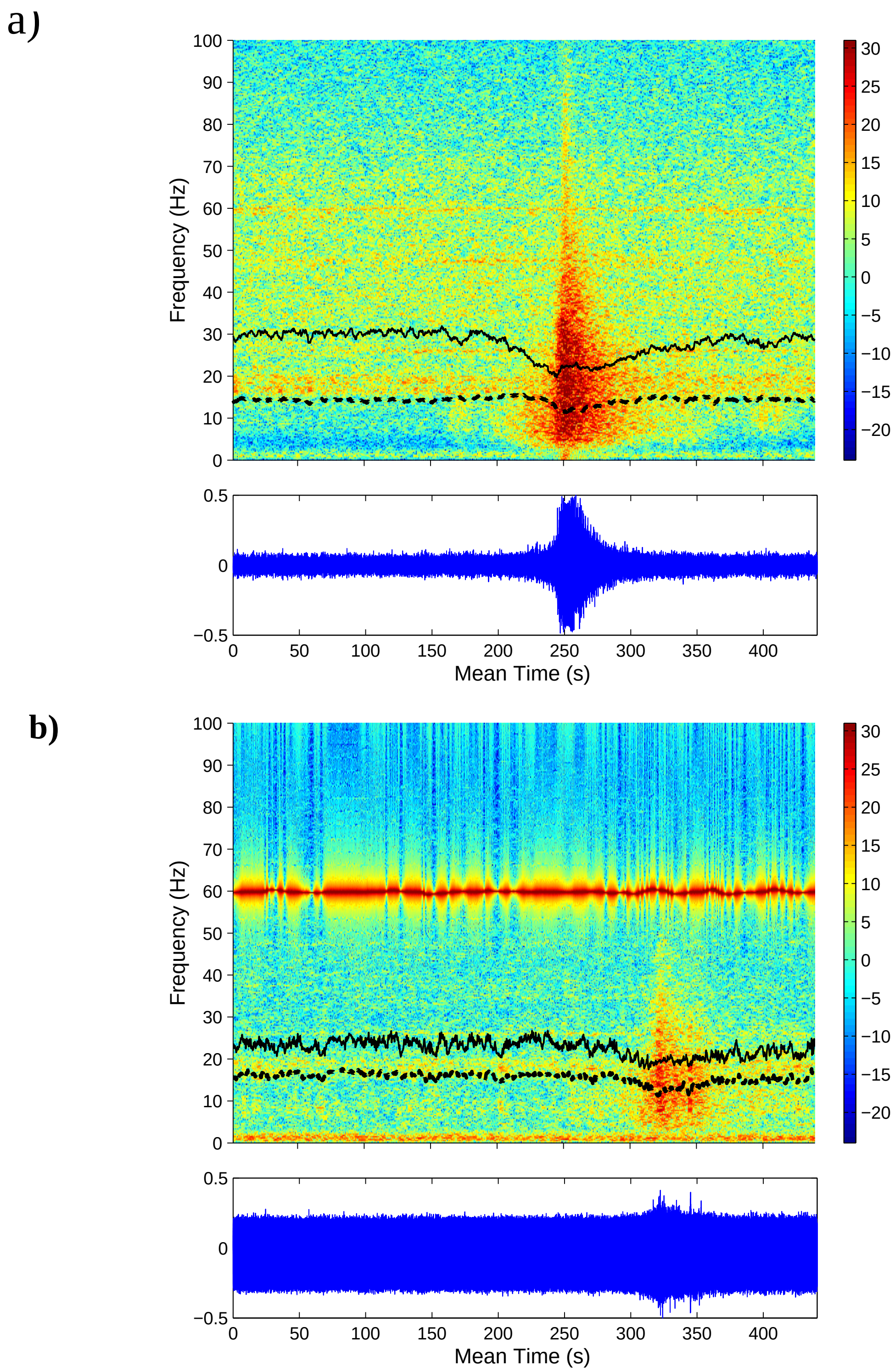

Figure 8: *Spectrograms for Wake (6 Dec. 1998). The mean frequency (solid line) and the standard deviation (dotted line) of the normalized spectrum are shown. a) WK30 (arid=1573661), b) WK31 (arid=1573660). The standard deviation is plotted from zero to appreciate the variation in standard deviation across a waveform.*

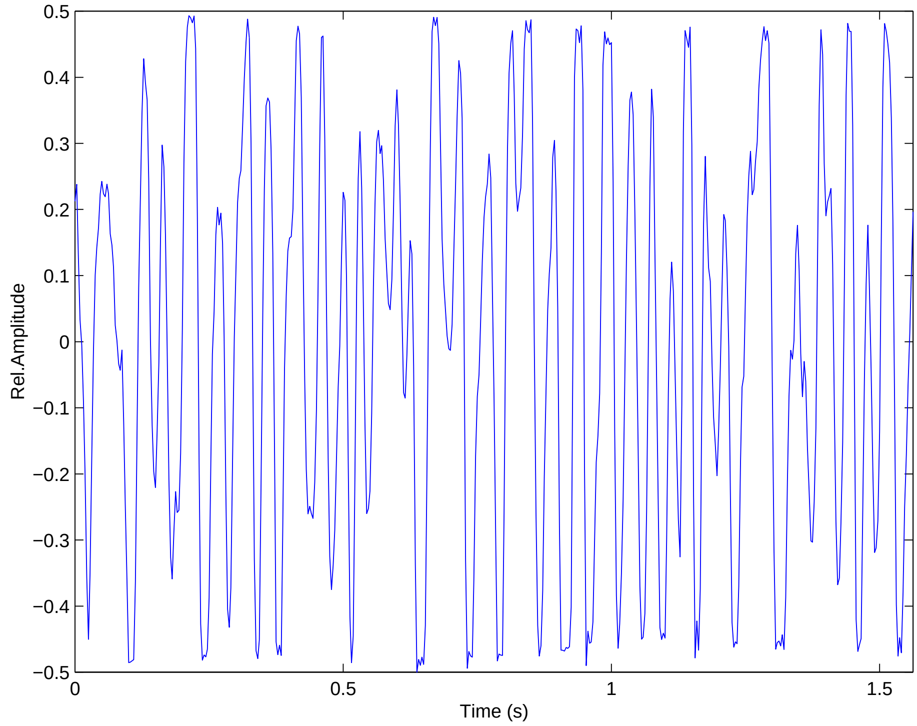

Figure 9: *The clipped signal on WK30, 6 Dec. 1998.*

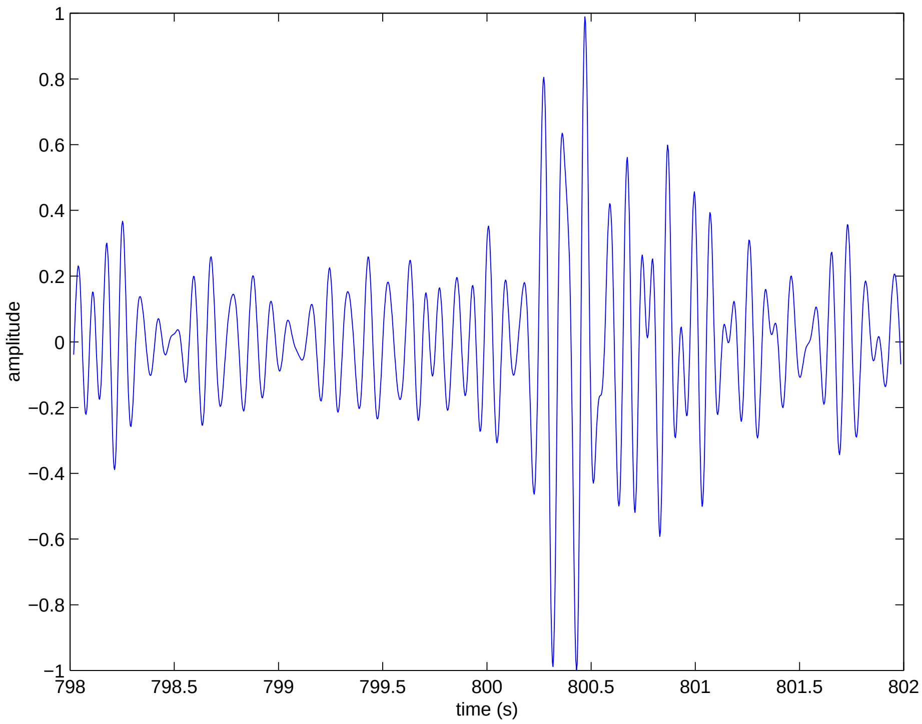

Figure 10: *The peak around 800 s for the signal on WK31, 6 Dec. 1998.*

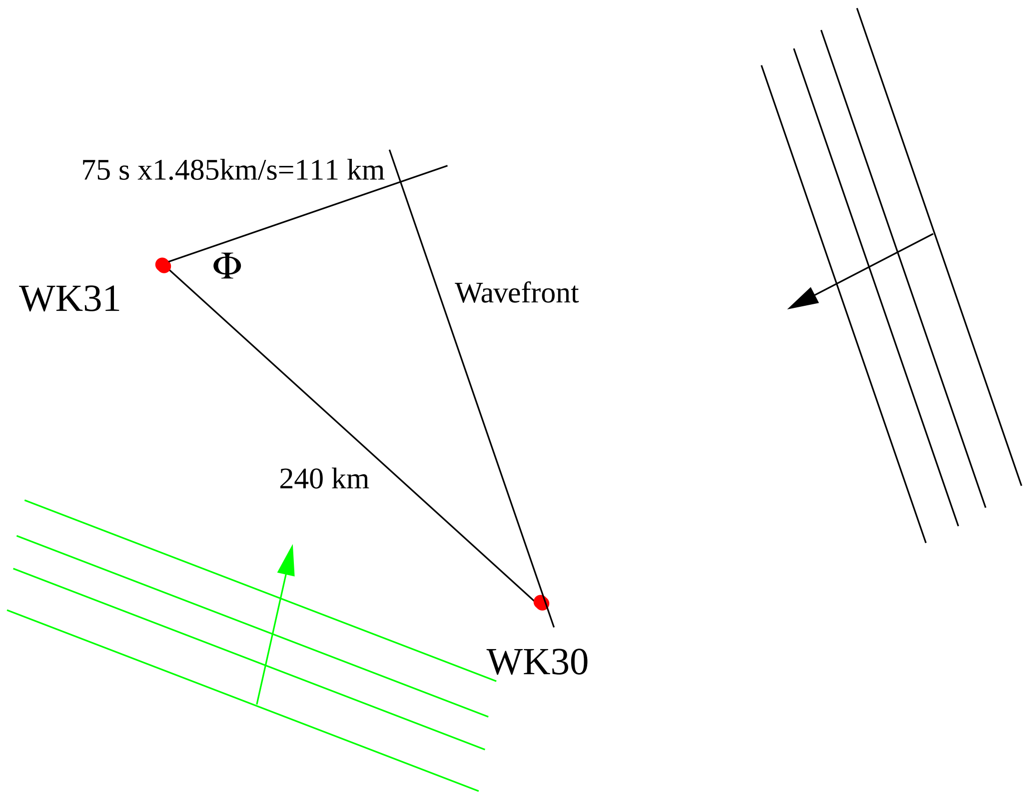

Figure 11: *If the same phase type was detected on both stations (WK30 and WK31) the angle of arrival could be determined from the measured time delay (here 75 s). By simple geometry it is found that Φ = acos(111/240)=62 deg. Based on time difference alone, both the "left" and "right" arrival are possible (the so-called left/right ambiguity). As the azimuth from WK31 to WK30 is 133 deg the station-to-event azimuth is 133+/- 62= 71 or 195 deg.*

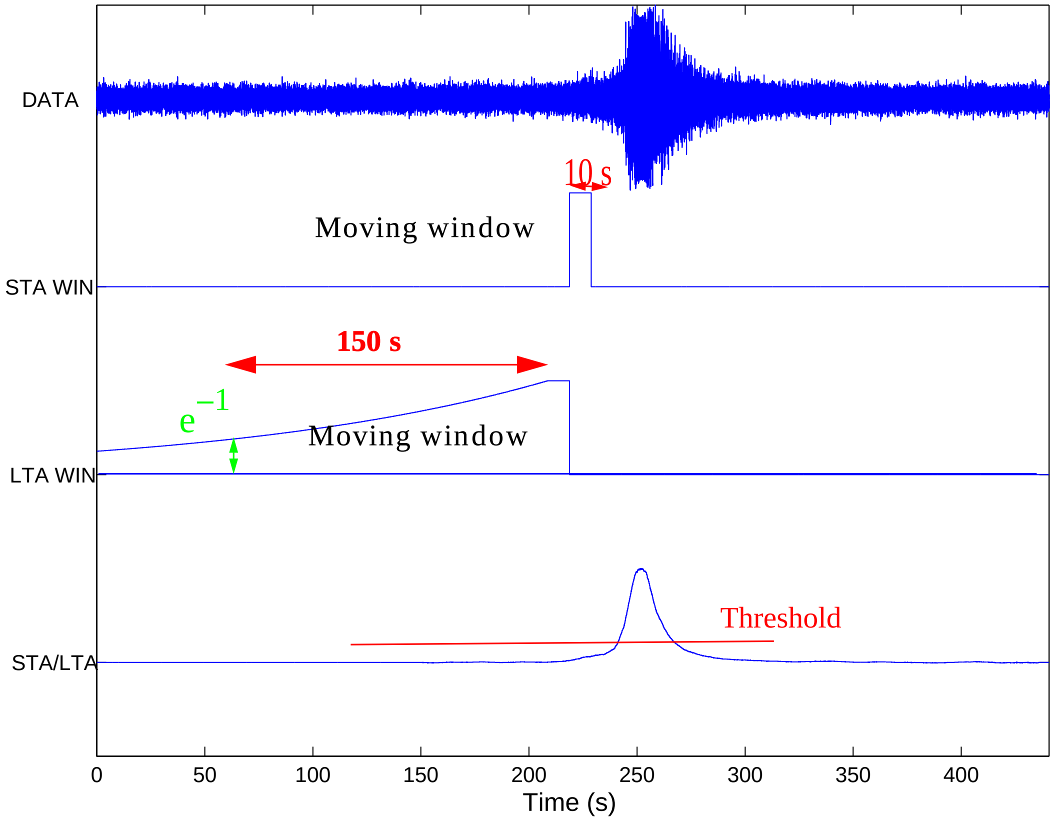

Figure 12: *Illustration of calculating the STA/LTA for a hydroacoustic arrival. 1st row: The raw data. 2nd row the STA moving window, this is applied directly to the data. The 10 s is the duration of the STA window. 3rd row: the LTA moving window, this window is applied to the STA values. The 150 s is the value when the weighting has decreased from 1 to exp(-1). 4th row: The STA/LTA ratio computed for each time sample.*

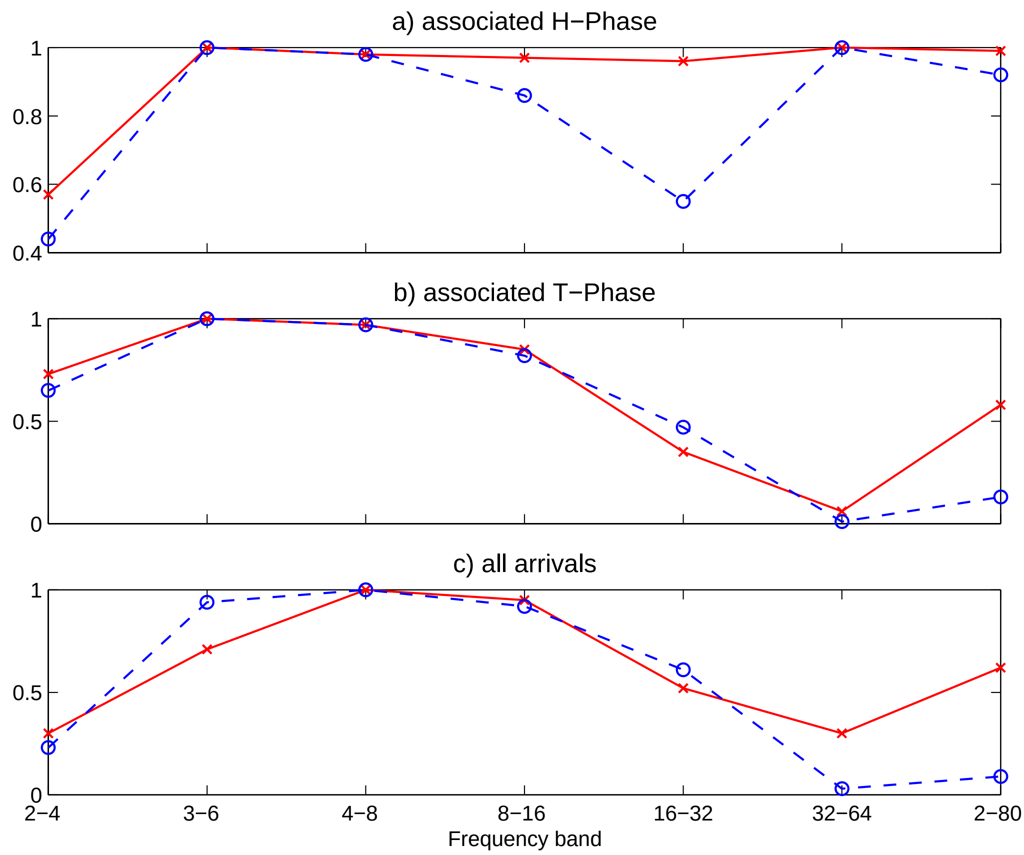

Figure 13: *Relative number of detections for each frequency band for WK30(solid red) and WK31 (dashed blue). a) only associated H-phases, b) only associated T-phases, and c) all arrivals (H, T and N phases). From c) it is seen that WK31 has few detections in the 32-64 Hz and the 2-80 Hz bands. This is due to the high level of electrical noise in these bands. The empirical identification rules require a detection in both the 3-6 Hz and the 32-64 Hz bands to identify an H-phase, thus H-phases have a detection rate of 100% in these bands. The same applies to T-phase detections in the 3-6 Hz band. It is not clear why WK30 has few H-phase detections in the 16-32Hz band.*

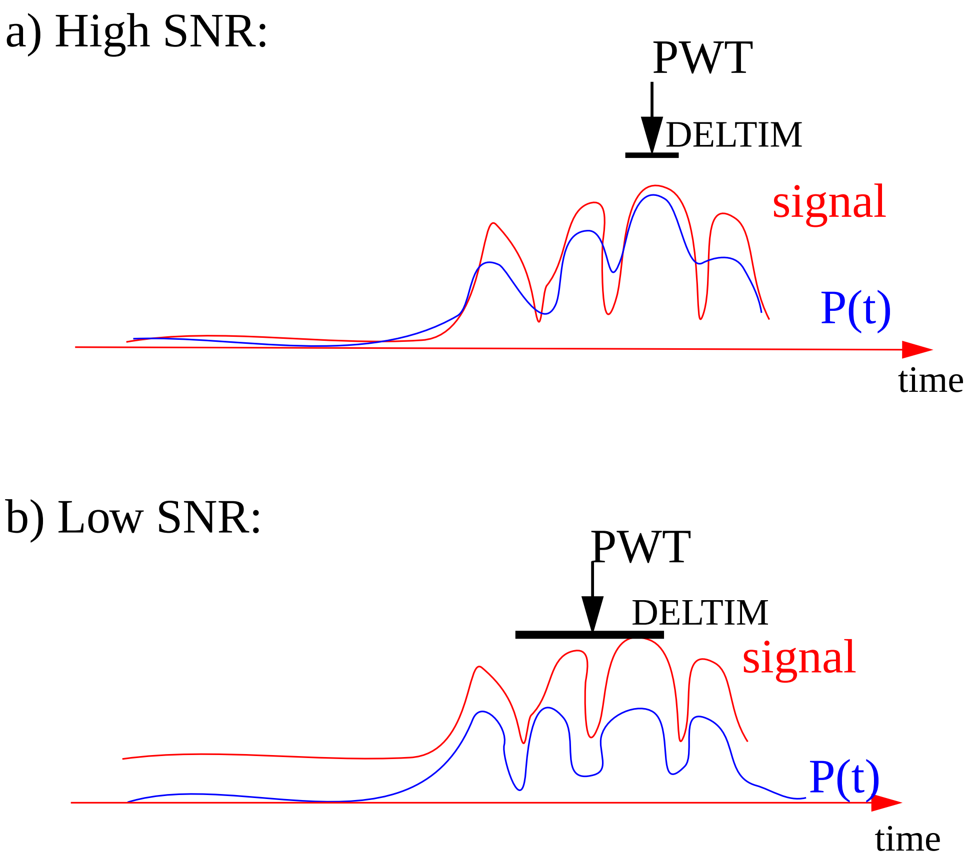

Figure 14: *Hydroacoustic arrival-time is determined by Probability Weighted Time (PWT) as illustrated for a) high SNR and b) low SNR [schematic drawing]. Based on an observed signal (red) the probability (blue) that a sample is the true maximum is calculated. This probability depends on the ambient noise. For high SNR (a) the actual peak is very likely to be the true peak, but for low SNR (b) the four peaks are as likely to be the main peak.*

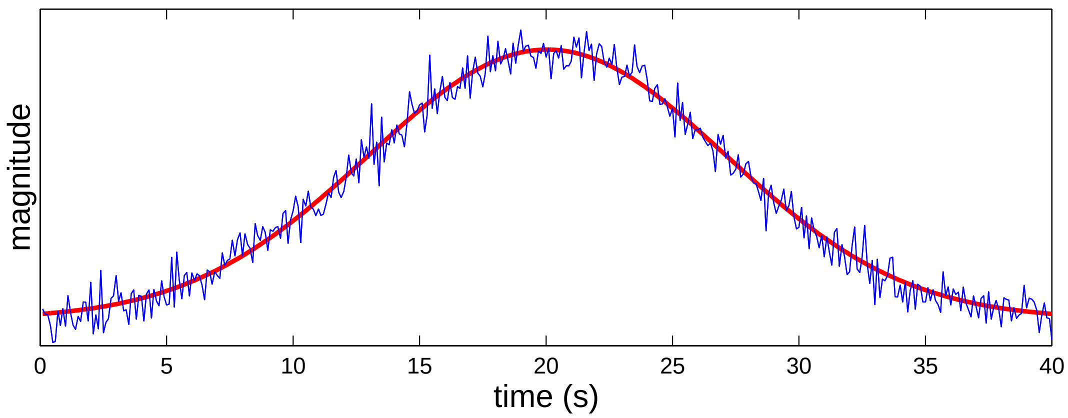

Figure 15: *Principle in the model for the PWT calculation. The magnitude of the received data (blue) is a sum of an unknown signal (red) and a random component.*

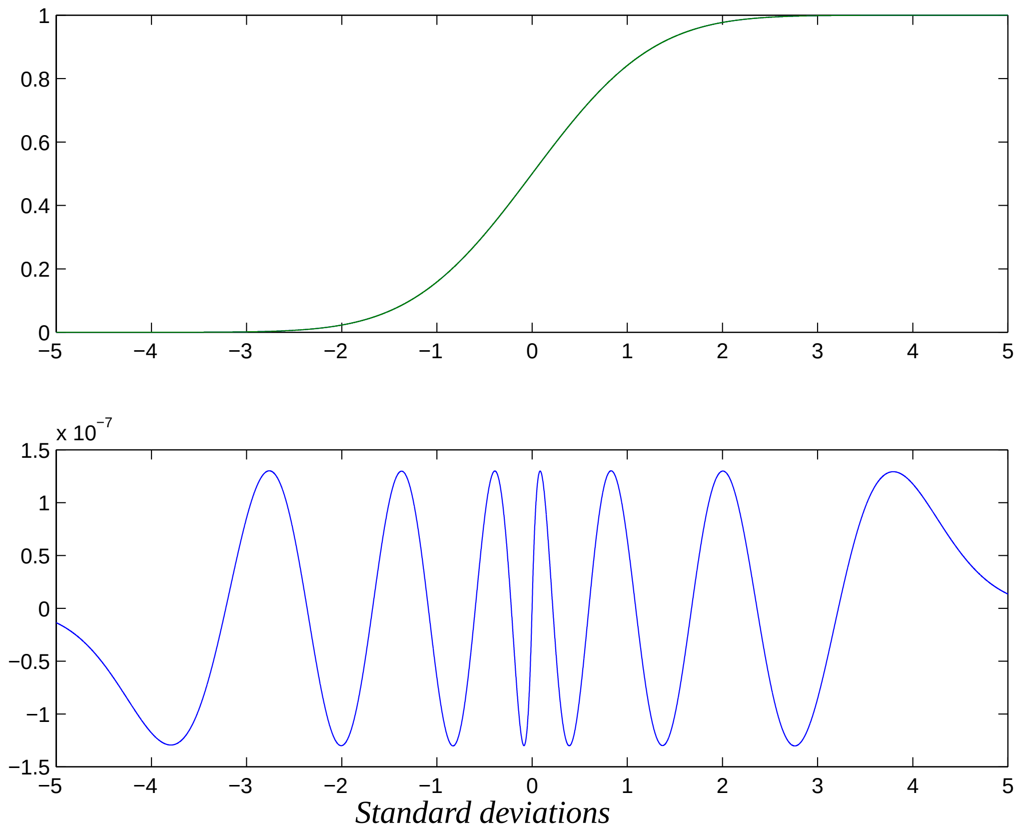

Figure 16: *Computation of the standard normal probability distribution by (a) a Matlab function and (b) the difference between this and the approximation in [Gue98]. The difference is of the order 1E-7.*

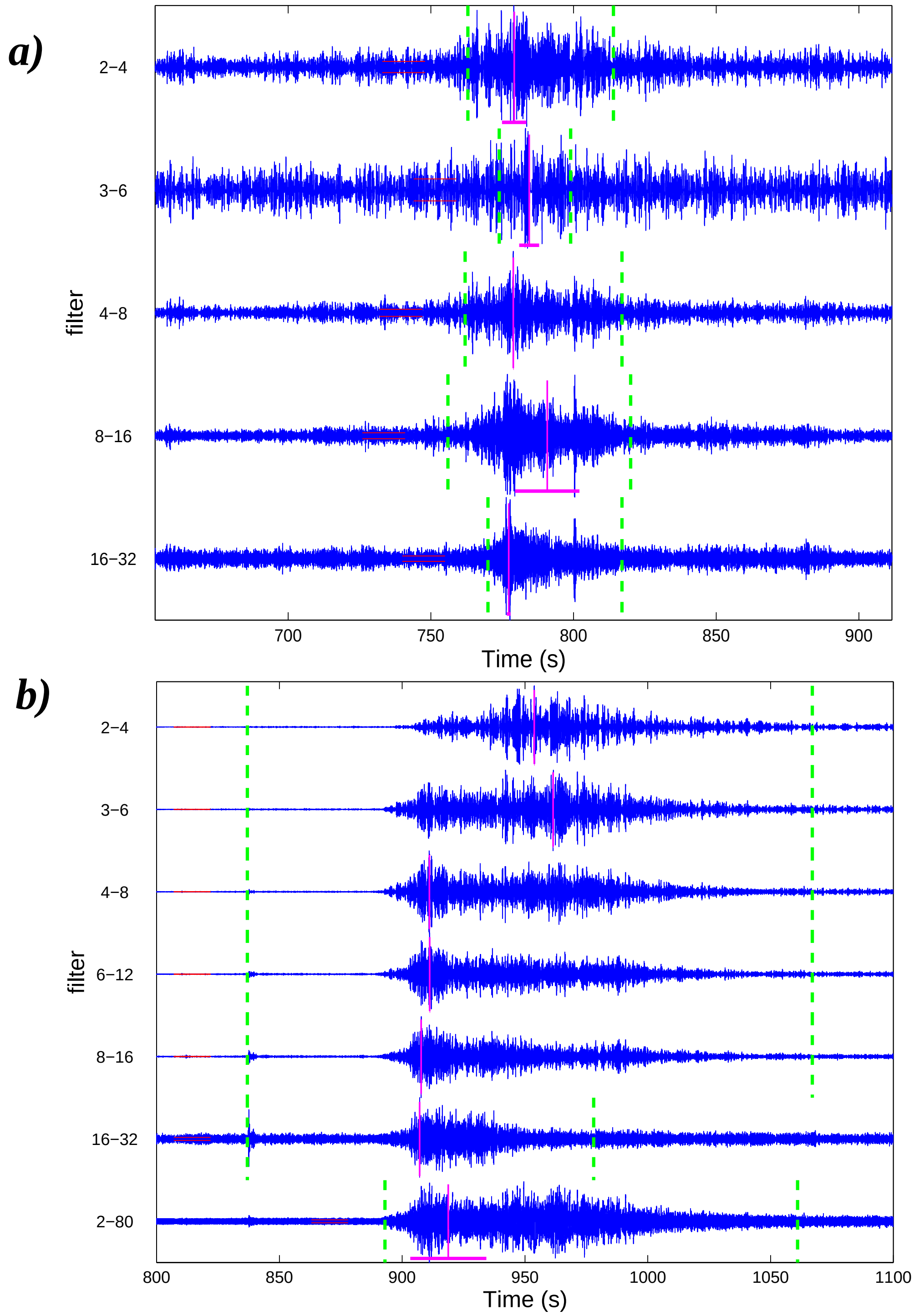

Figure 17: *PWT determined using the DFX-approach a) Dec-6 arrival and b) Taiwan arrival. The version A estimated arrival-time (magenta) and two times the standard deviation (magenta) is shown. Plus/minus the noise level σ_a (red) is indicated before the onset time.*

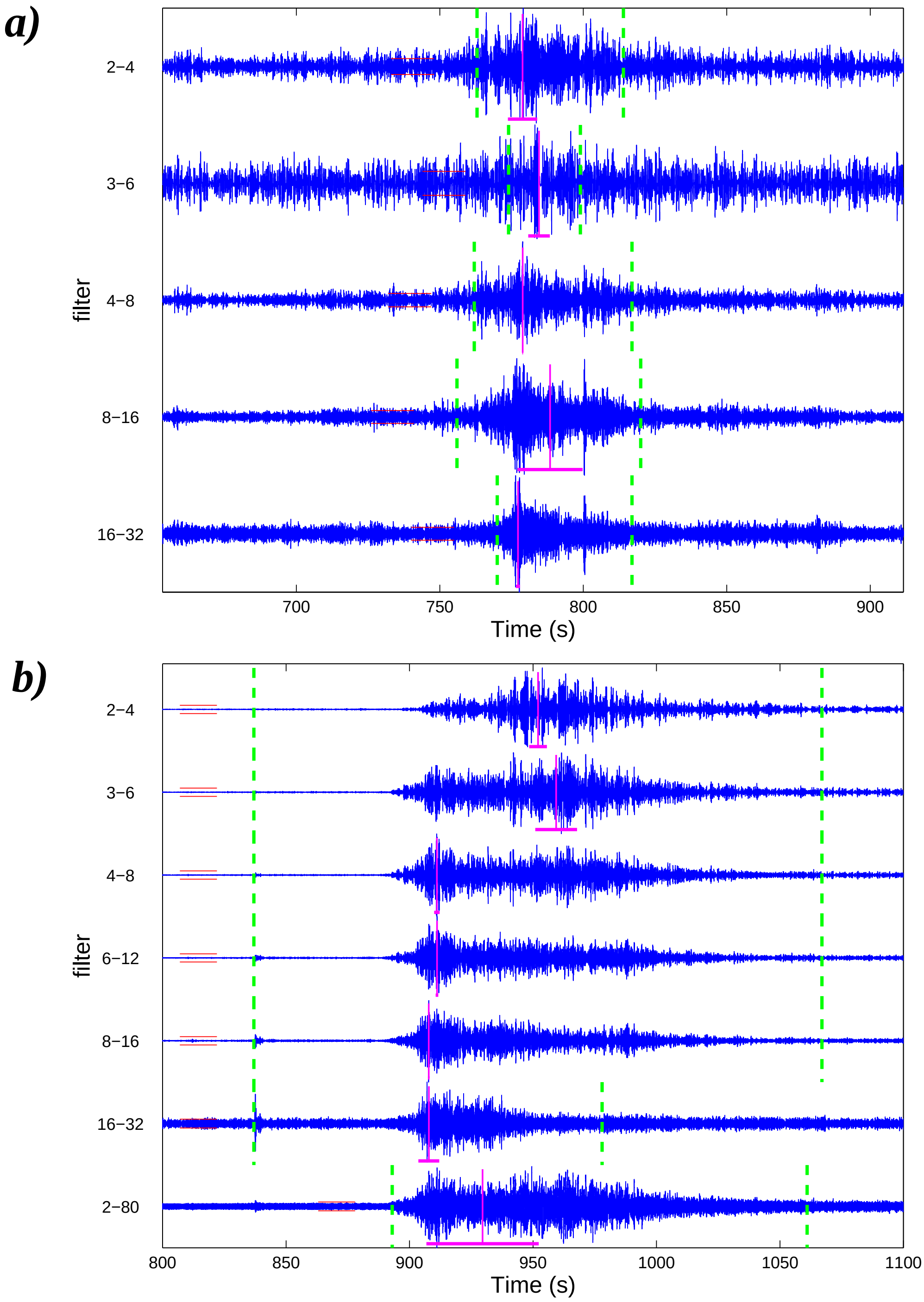

Figure 18: *PWT for a) Dec-6 arrival and b) Taiwan arrival with also signal error* σ_s *included in the processing. The version A estimated arrival-time (magenta) and standard deviation (magenta) is shown. Plus/minus the error level* σ_e *(red) is indicated before the onset time. In a) the noise level* σ_a *are comparable to the signal error* σ_a*, the total error* σ_e *is not much changed, the standard deviation in each band has increased slightly. In b) the first five filter bands are virtually noise free, thus the signal error* σ_s *dominates and the standard deviation is now much larger.*

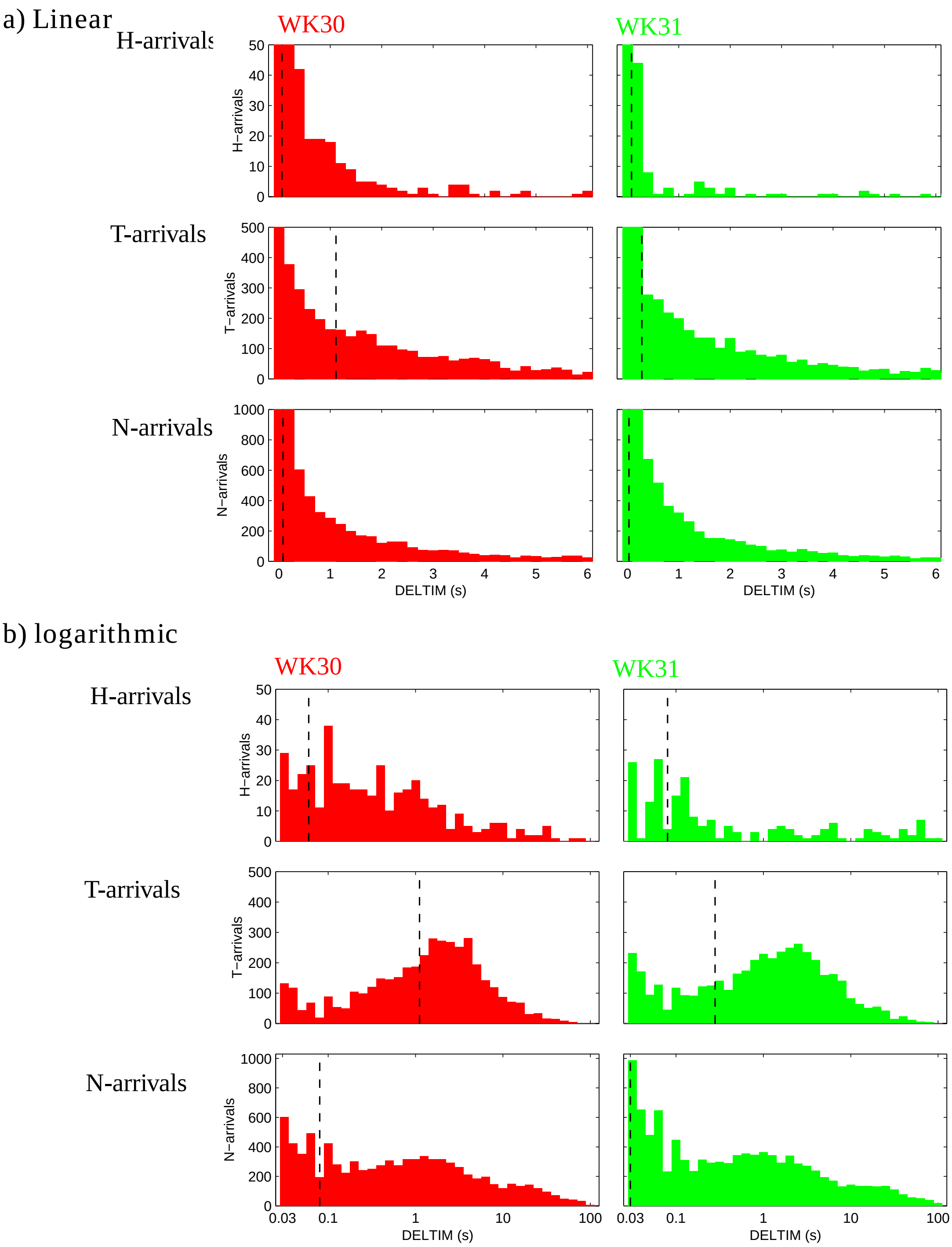

Figure 19: *Distribution of arrival-time measurement error (DELTIM) for hydroacoustic H-/T-/N-arrivals for WK30 (left) and WK31 (right) with a) linear and b) logarithmic time axis. The dashed vertical line indicates the mean value for each arrival type. For the linear time scale the size of each bin was 0.2 s. Many of the arrivals were in the 0-0.2 s bin: for WK30 (H, T, N) = (376, 1124, 6880) arrivals and for WK31 WK30 (H, T, N) = (136, 280, 11967) arrivals. This bin and the 0.2-0.4 s bin are dominating, but scale of figure is such that these bins are not shown in full. The logarithmic scaling does not show arrivals less than 0.03 s and de-emphasize the problems with small DELTIM.*

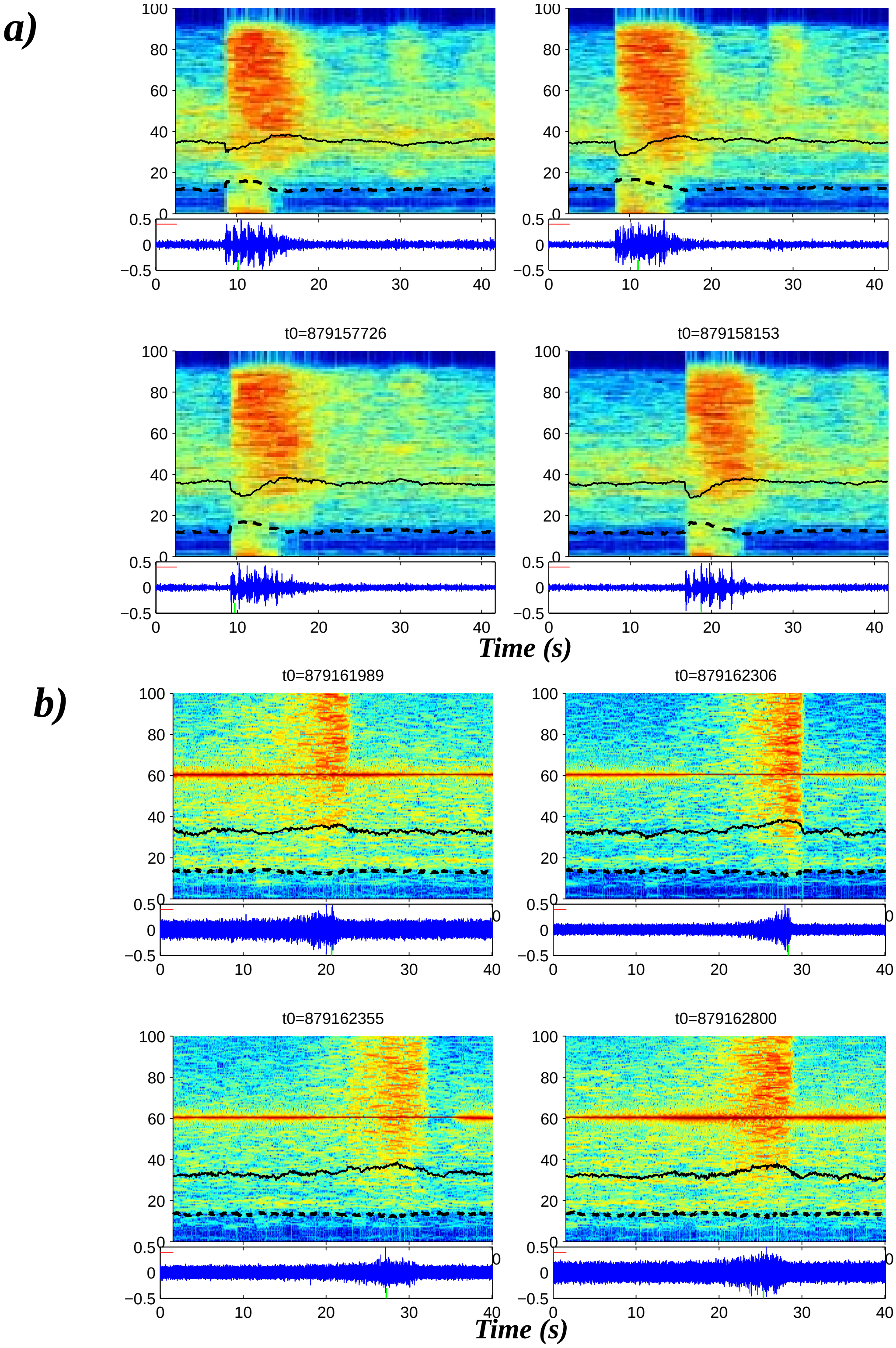

Figure 20: *Examples of the received waveform at a) PSUR and b) WK30 for the C4-shots off California (Nov. 10 1997). Range to PSUR is about 1 deg, the range to WK30 about 60 deg and the range to VIB about 20 deg. On WK30 the amplitudes are quite low and the signals was not clipped. Data for WK31 were not available for Nov. 10 1997.*

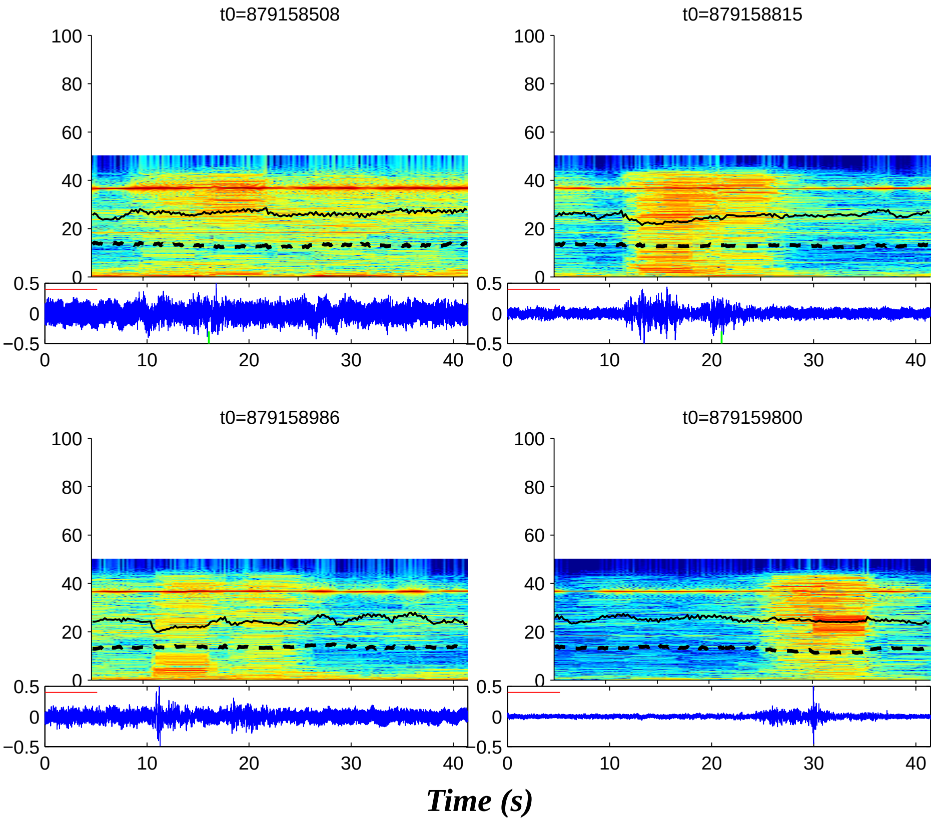

Figure 21: *Examples of the received waveform at VIB for the C4-shots off California (Nov. 10 1997). The generator noise can be seen as a constant line at about 38 Hz. It is questionable if the C4 shots are truly seen at VIB (Gault, personal communication). The distance between the picked arrivals are not the same at VIB as on PSUR and Wake (Hanson, personal communication). Only the two arrivals in the top display were selected as part of the C4-shots.*

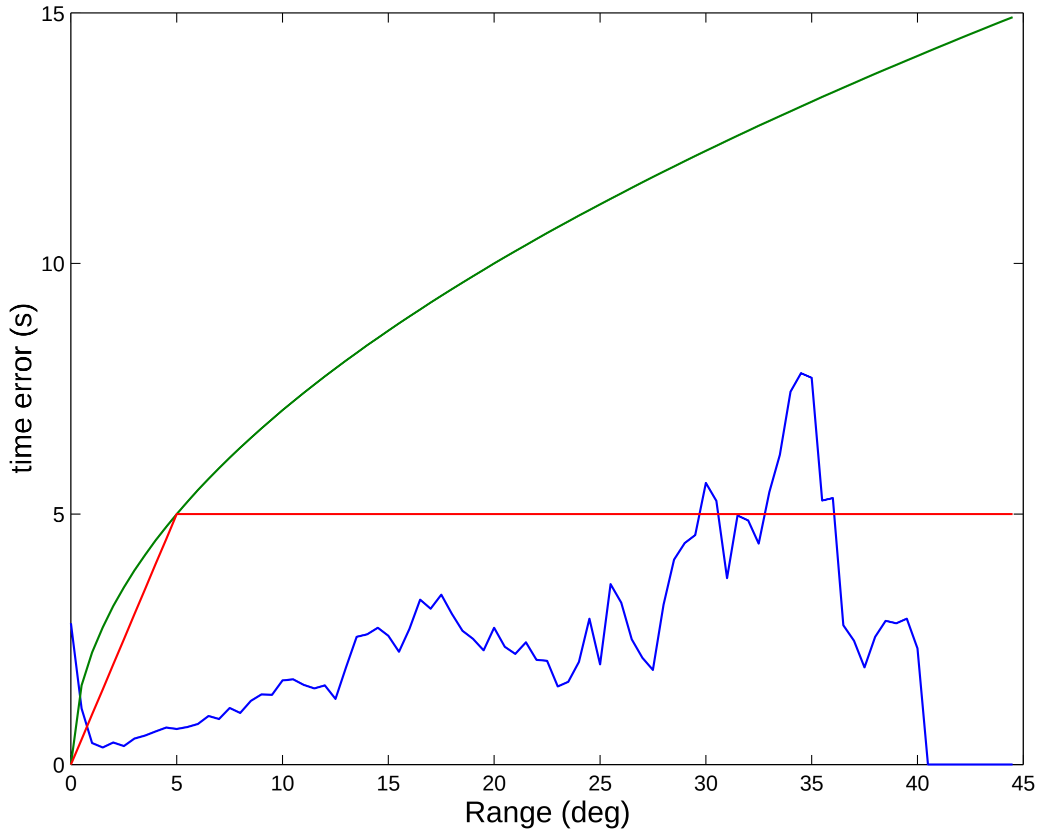

Figure 22: *Travel-time modelling error for H -phases for the 1D travel-times (solid green) and for the 2D travel-times (solid red). In Release 2 the travel-time model error (solid blue) depends on range and azimuth and is here shown for reference for the northern path. Past 41 deg range it is zero as there is land. Overall the increase in modelling error with range seems reasonable. The decease at range 34-41 deg is due to stripping of the modes, as there becomes less propagating modes the travel-time is better determined.*

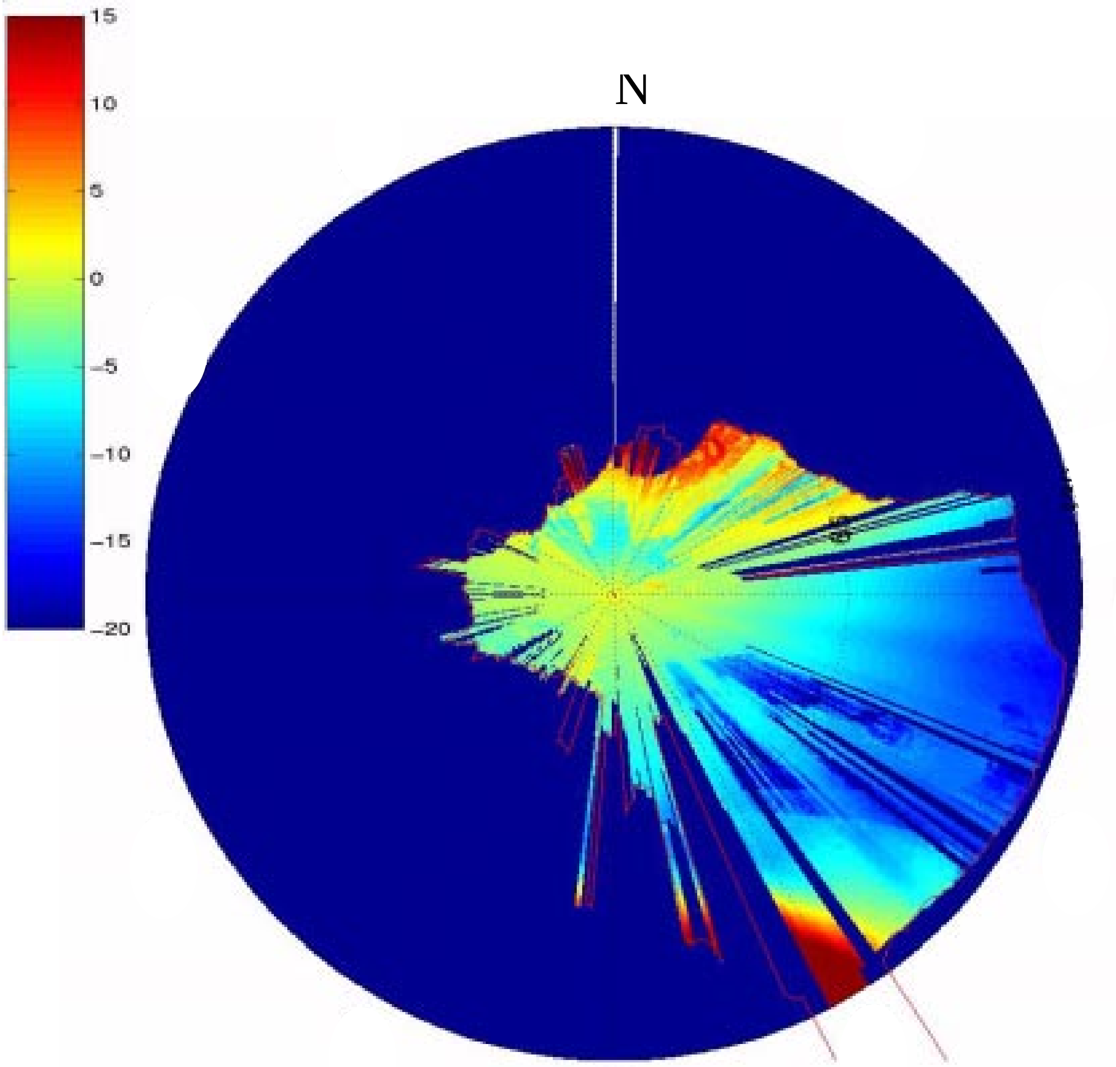

Figure 23: *Polar diagram of the 2D travel-time in seconds reduced with 1482 m/s for WK30, the maximum radial distance in the plot is 130 degrees. The red curve indicates the limits of the blockage grid. Only reduced travel-time in the range 15 to -20 s is shown. The maximum reduced travel-time is 71 s.*

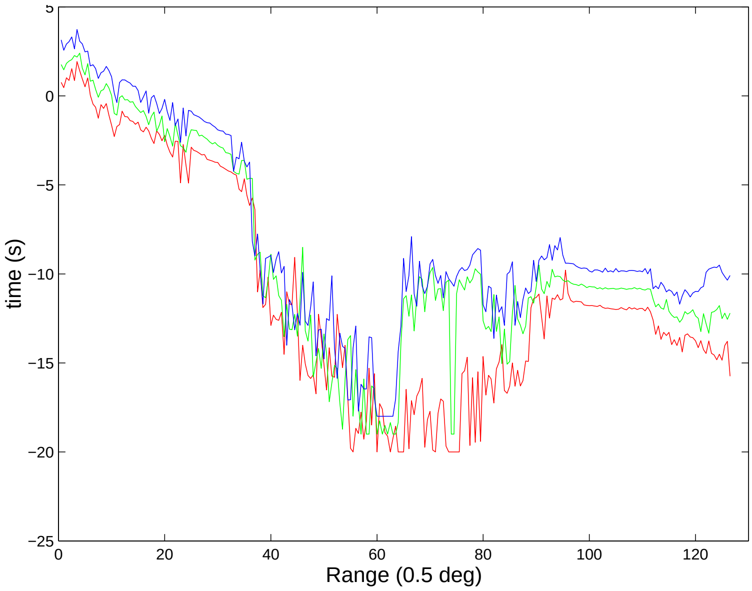

Figure 24: *Reduced travel-time (reduction velocity 1482 m/s) for the 2D travel-time table for WK30, azimuth 117.5 deg (red), 118 deg (green, offset 1 s) and 118.5 deg (blue, offset 2 s).*

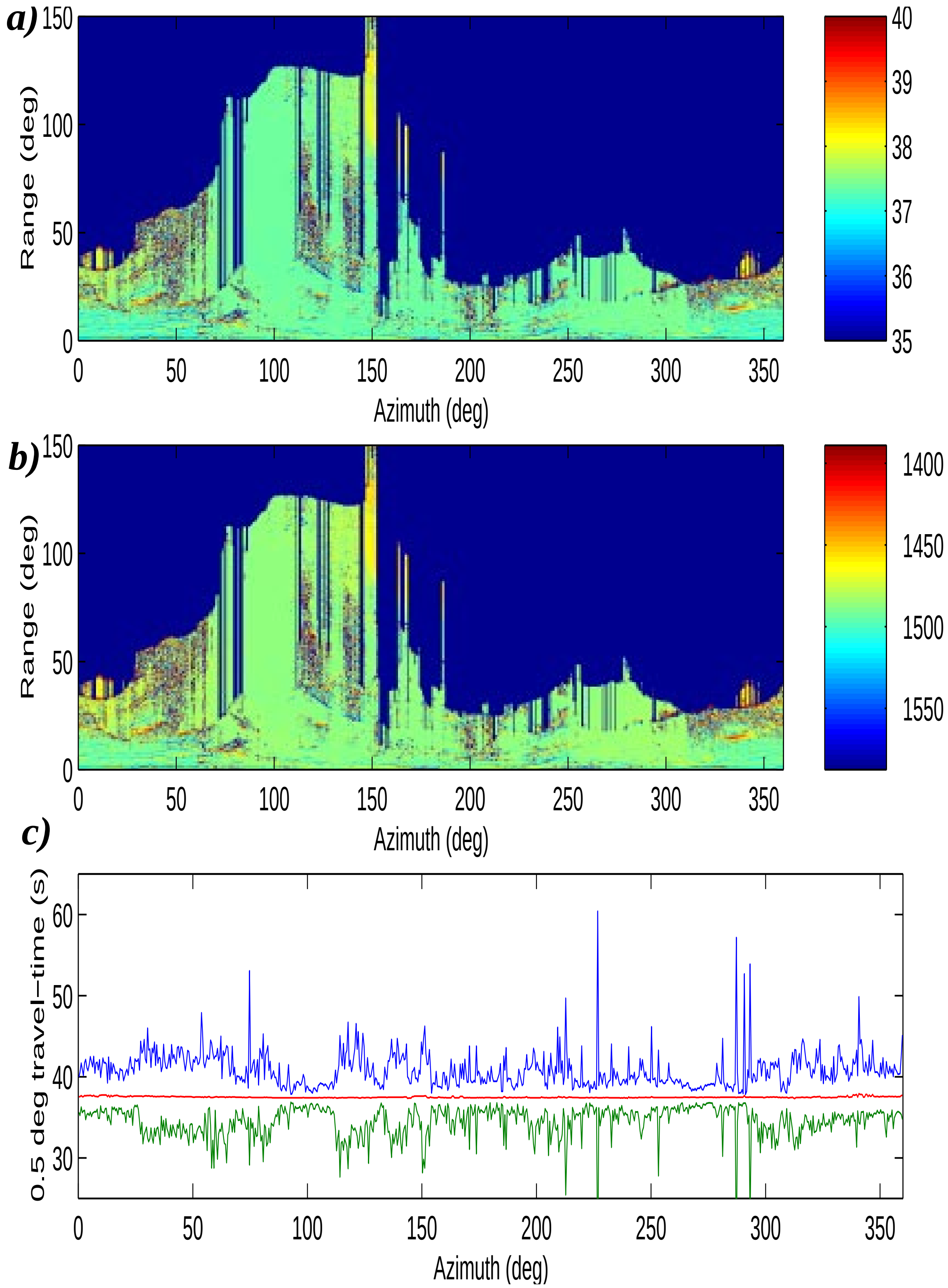

Figure 25: *Variations of travel-times for WK30. a) 0.5 deg travel-time for each 0.5 degrees in range [truncated to 35-40 s], b) The corresponding local velocity [truncated to 1390-1590 m/s] and c) The maximum (blue) mean (red) and minimum (green) travel-time difference for each azimuth computed from a)*

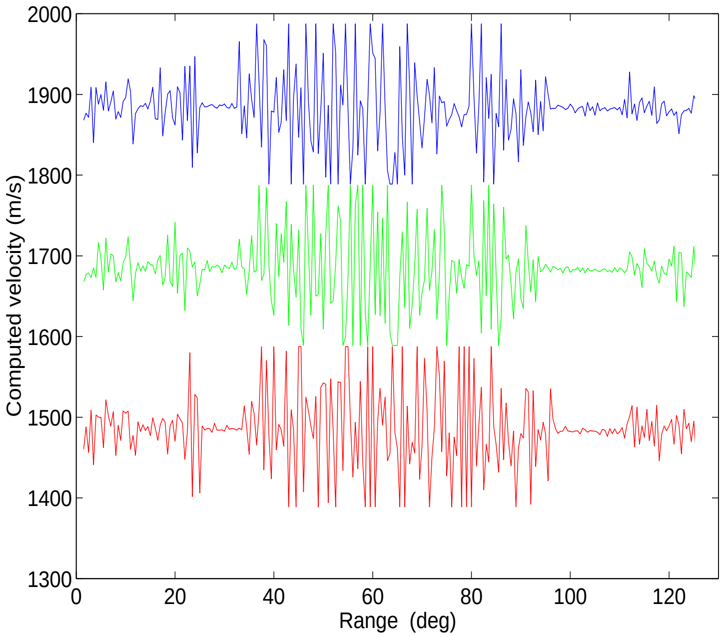

Figure 26: *Computed velocity based on the 2D travel-time table for WK30, azimuth 117.5 deg (red), 118 deg (green, offset 200 m/s) and 118.5 deg (blue, offset 400 m/s) [truncated to 1390-1590 m/s]. The peak at about 75 deg range for 118 deg azimuth in Figure 24 can be seen as first a large positive and then a large negative peak at 75 deg range. If not truncated it would cause variations between 1160 and 1950 m/s.*

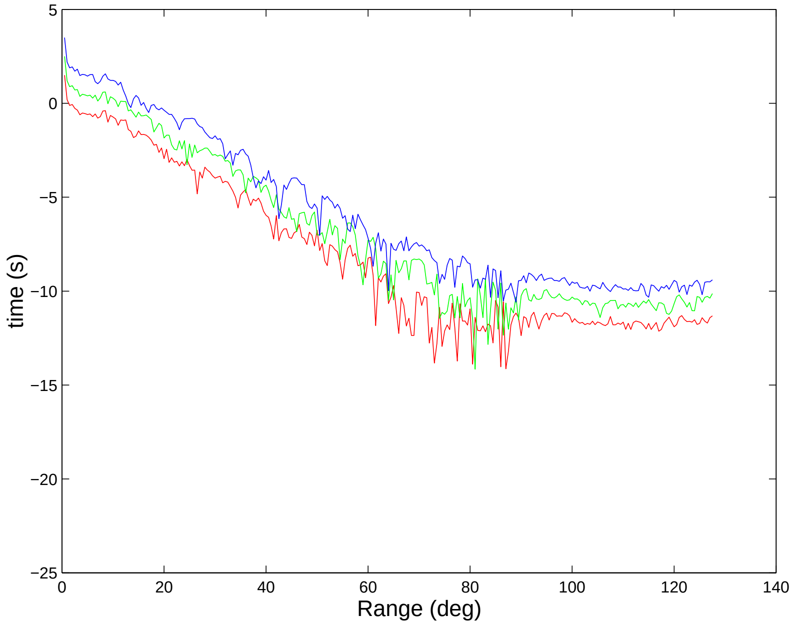

Figure 27: *For Release 2: Reduced travel-time (reduction velocity 1482 m/s) for the 2D travel-time table from WK30 (summer), azimuth 117.5 deg (red), 118 deg (green, offset 1 s) and 118.5 deg (blue, offset 2 s).*

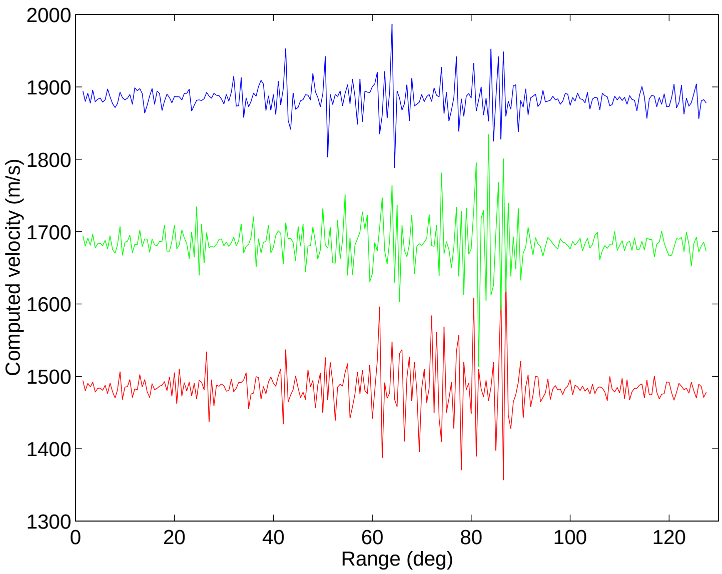

Figure 28: *For Release 2: Computed velocity based on the 2D travel-time table for WK30 (summer), azimuth 117.5 deg (red), 118 deg (green, offset 200 m/s) and 118.5 deg (blue, offset 400 m/s).*

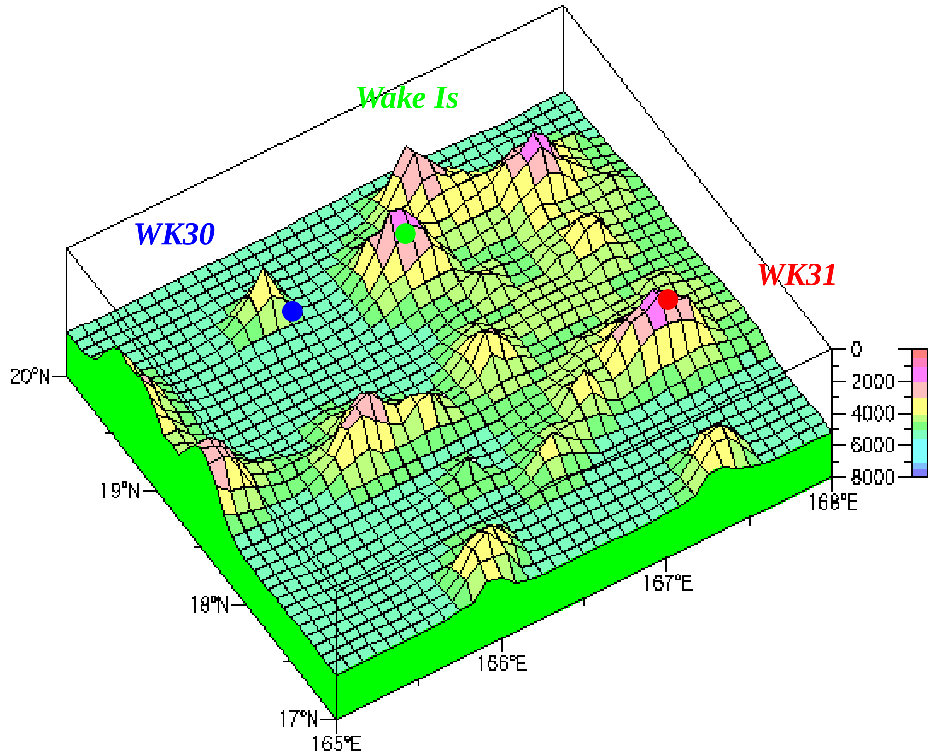

Figure 29: *Bathymetry around Wake. Wake Island is at (19.3N, 166.6E)[Green], WK30 (19,41N,165.86E) [Blue] and WK31 (17.92N, 167.49E) [Red]. Due to resolution the bathymetry at Wake Island does not reach the surface. It is, however, clear that Wake Island is a minor island and thus should there in no full blockage due to the island. To evaluate if there is partial blockage the acoustic propagation conditions must be examined, by modeling the propagation. In general, when a seamount extends into the SOFAR channel some partial blockage would be expected*

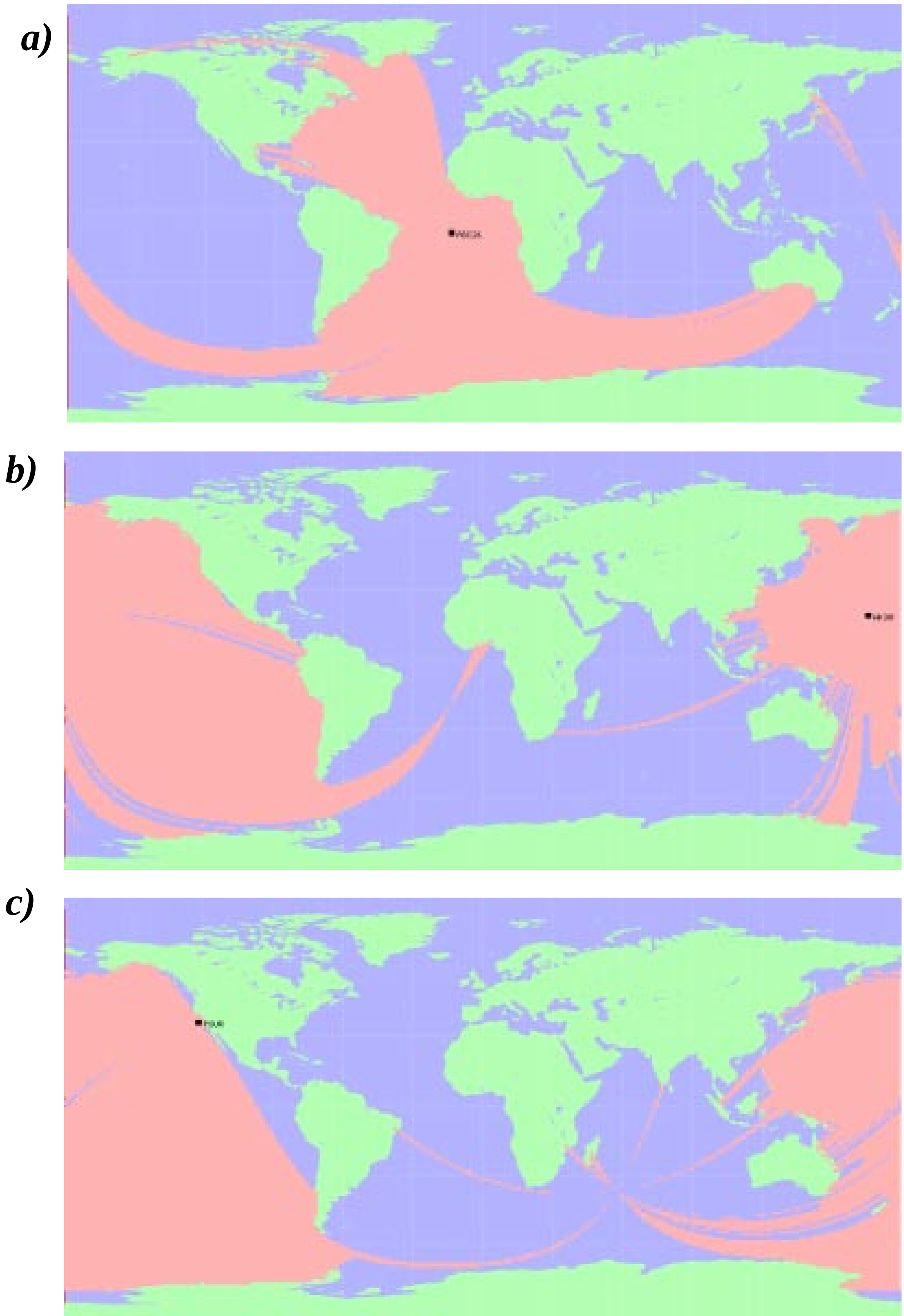

Figure 30: *Blockage maps for a) ASC26, b) WK30 and WK31, and c) PSUR (provided by Mark Fisk).*

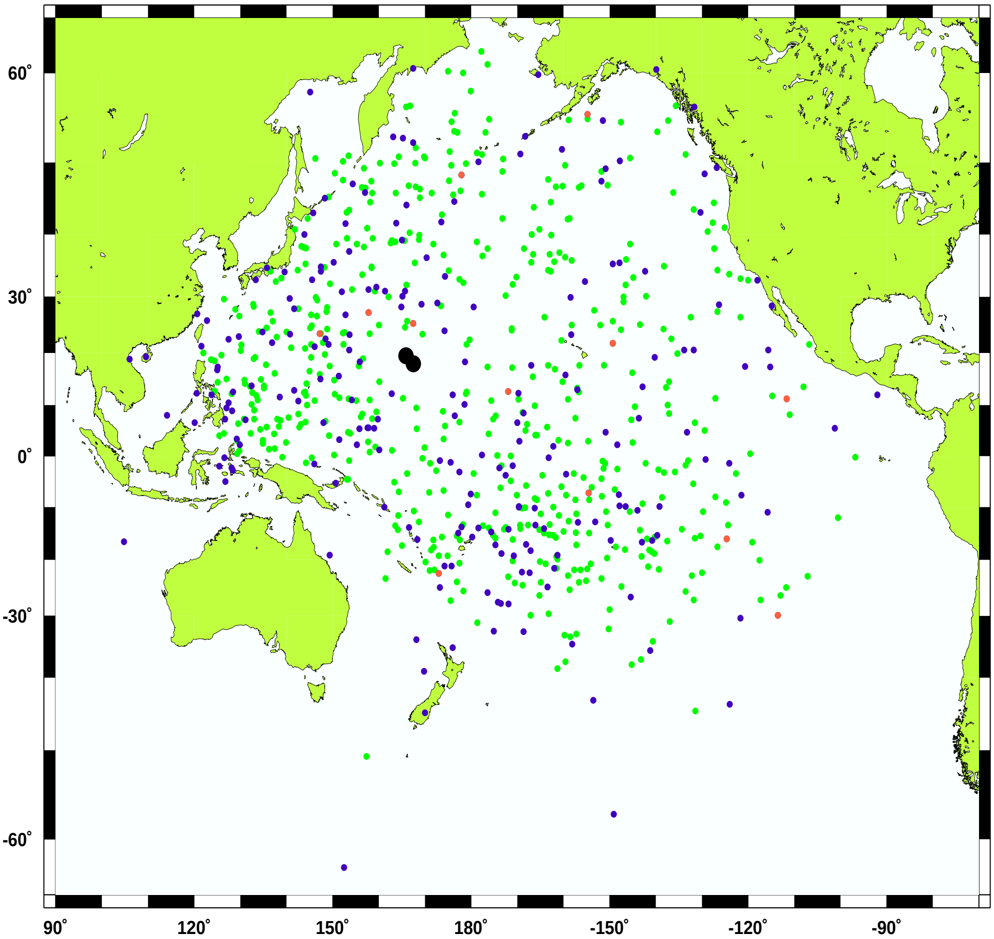

Figure 31: *Origins determined by the automatic processing (SEL1 account) where the H-phase from the hydrophones at Wake (black dot) has been used in the location. Based on H-phase from only WK30 (green, 527 arrivals) or only WK31 (blue, 227 arrivals) and both WK30 and WK31 (red, 12 arrivals).*

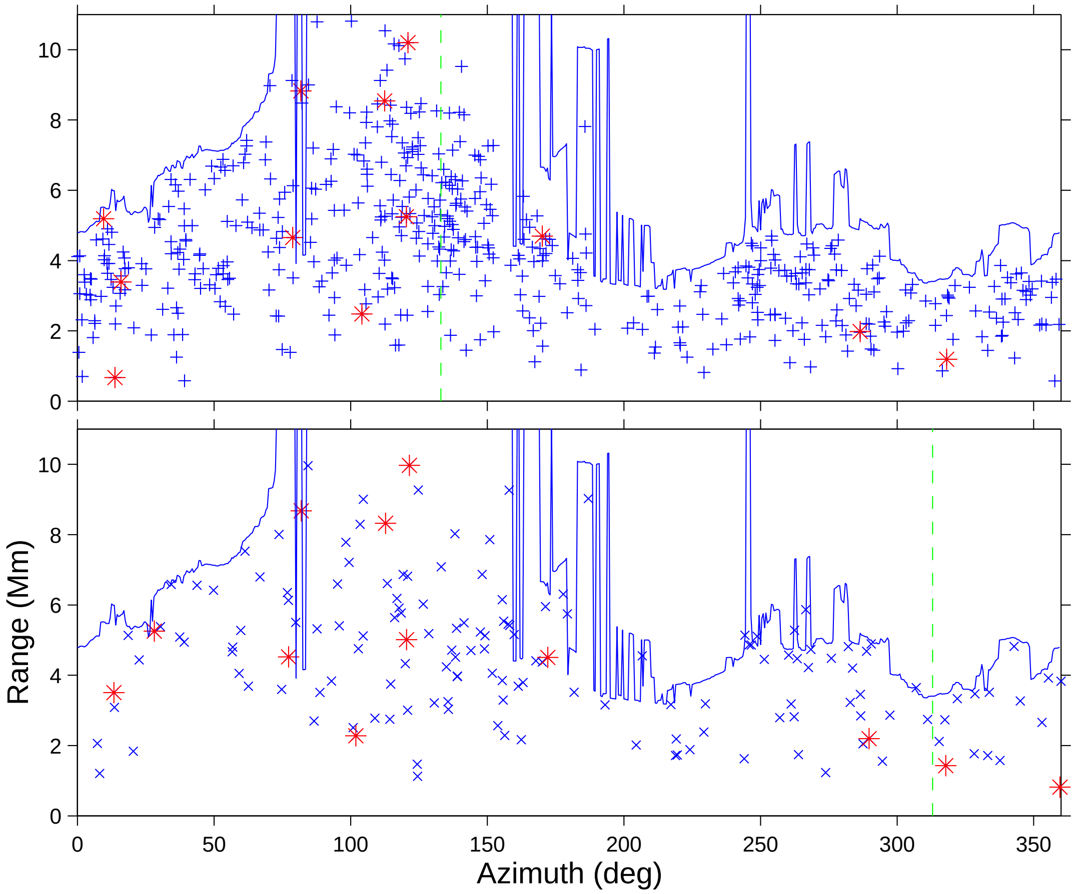

Figure 32: *Distribution of origins for events with a associated H-phases as a function of azimuth and range from WK30 (top) and WK31(bottom) as determined by the automatic processing (SEL1 account). The * indicates events associated to both WK30 and WK31, the +, x indicates events associated to WK30, WK31, respectively. The solid line indicates the limit due to blockage and the vertical dashed line is the azimuth to the other station.*

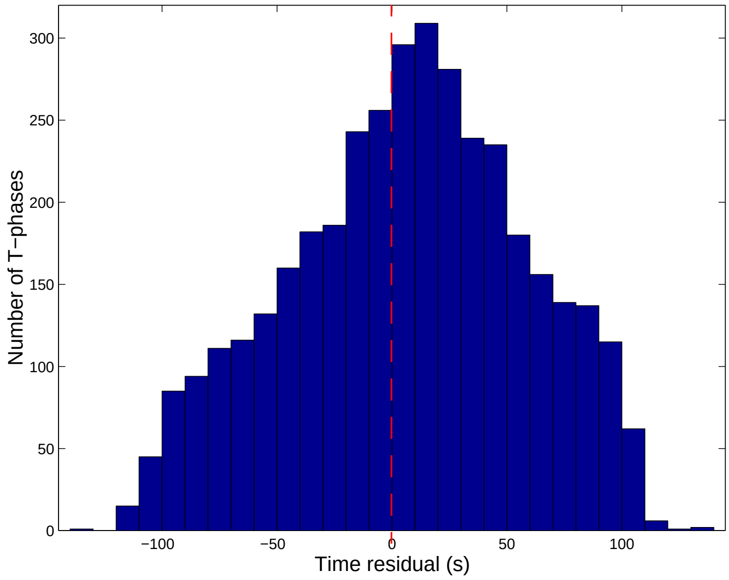

Figure 33: *Time residuals for all T-phases associated to an event in the SEL1 account. The residuals have a slight positive bias and the histogram is not symmetric around the peak.*

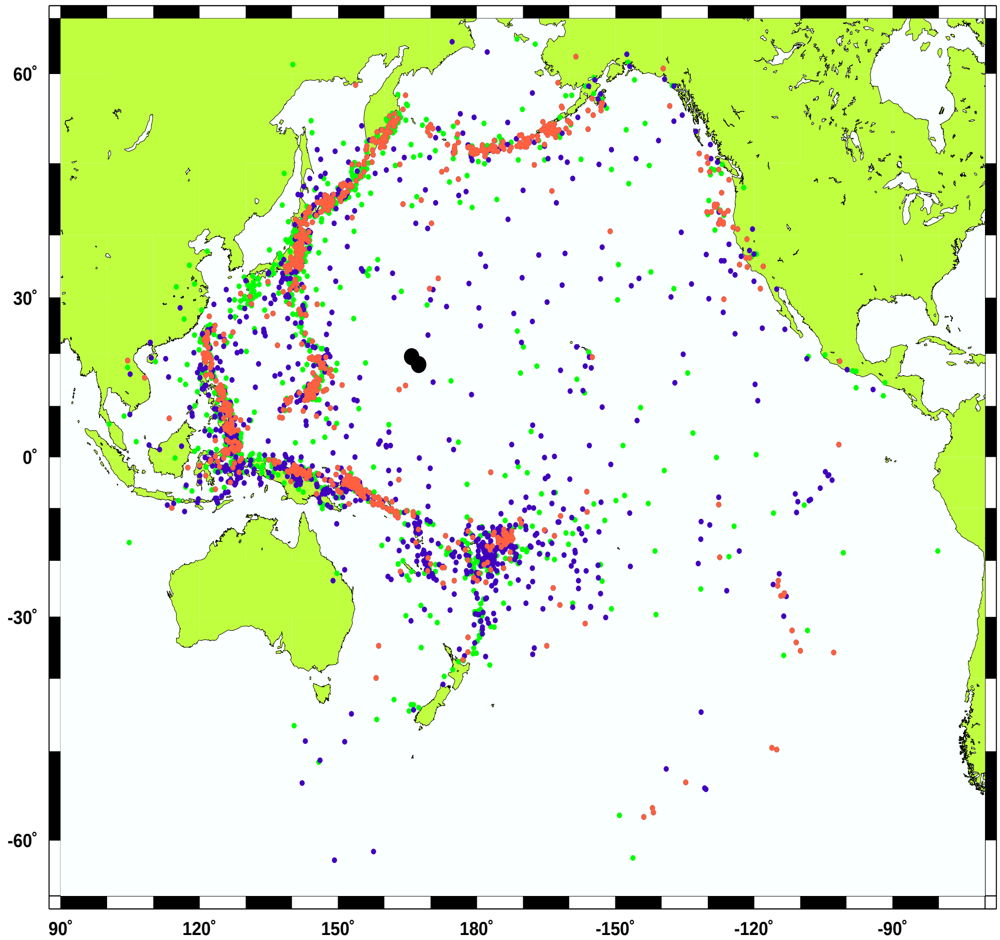

Figure 34: *Origins determined by the automatic processing (SEL1 account) where the T-phase from the hydrophones at Wake (black dots) was associated with the origin. Based on T-phase from only WK30 (green, 1886 arrivals) or only WK31 (blue, 1916 arrivals) and both WK30 and WK31 (red, 945 arrivals). The T-phases are not used in location of events, they are only associated. The origin of T-phase excitation can be onshore.*

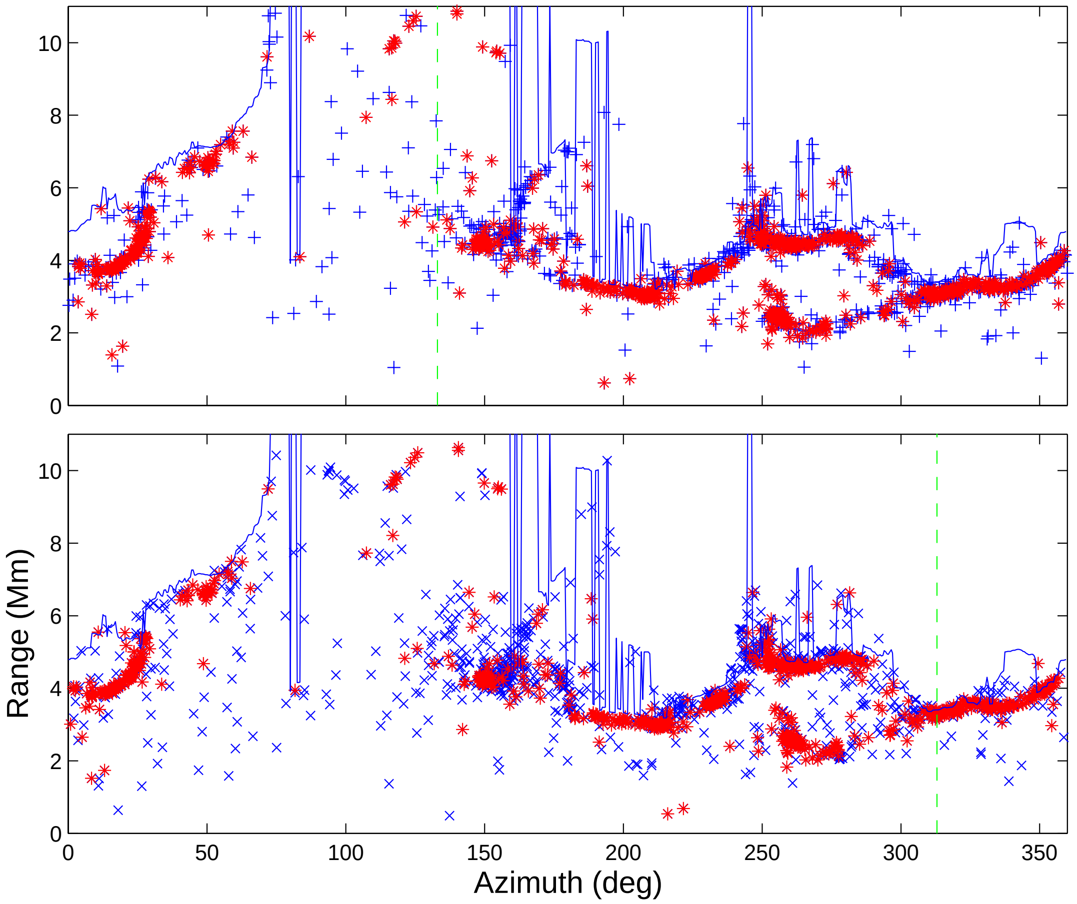

Figure 35: *Distribution of origins for events with associated T-phases as a function of azimuth and range from WK30 (top) and WK31(bottom) as determined by the automatic processing (SEL1 account). The * indicates origins associated to both WK30 and WK31, the +, x indicates events associated to WK30, WK31, respectively. The solid line indicates the limit due to blockage and the vertical dashed line is the azimuth to the other station.*

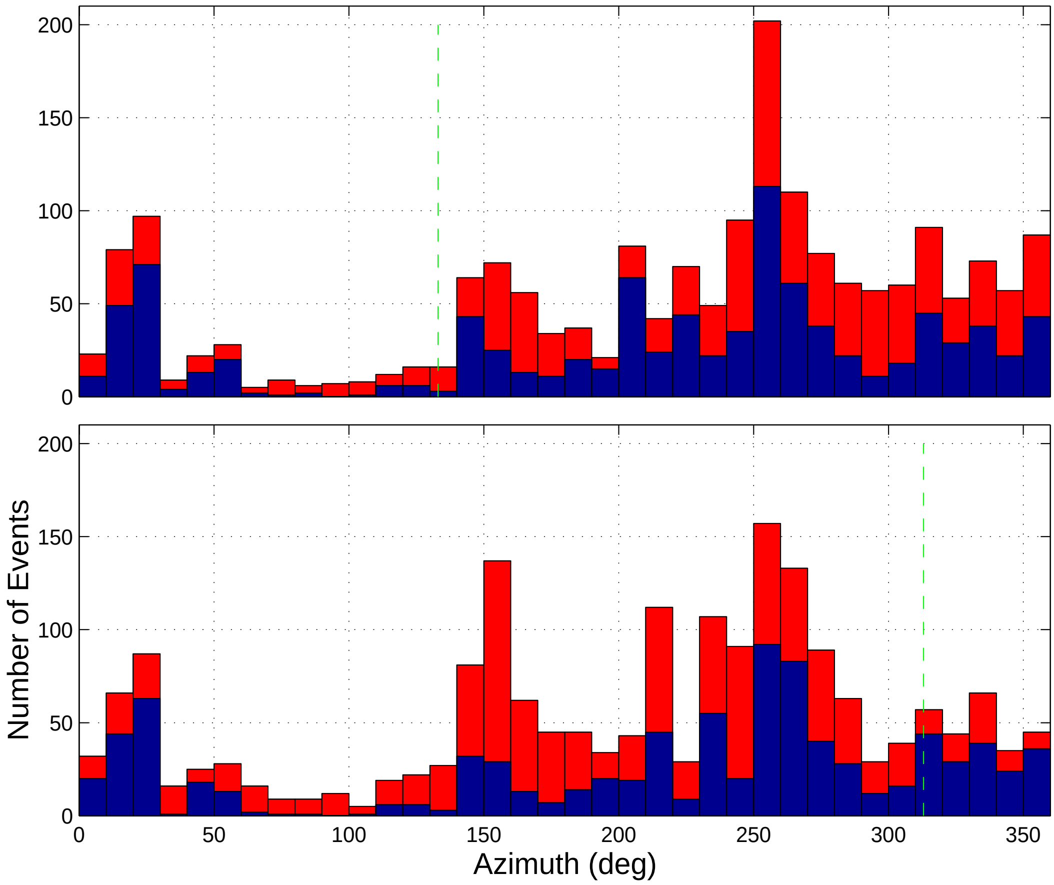

Figure 36: *Histograms of station to event azimuth for T-phases for WK30 (top) and WK31(bottom). The blue histogram indicates origins related to both WK30 and WK31, while the red histogram indicate origins where only the T-phase from one station (WK30 or WK31) is associated.*

a) WK30

b) WK31

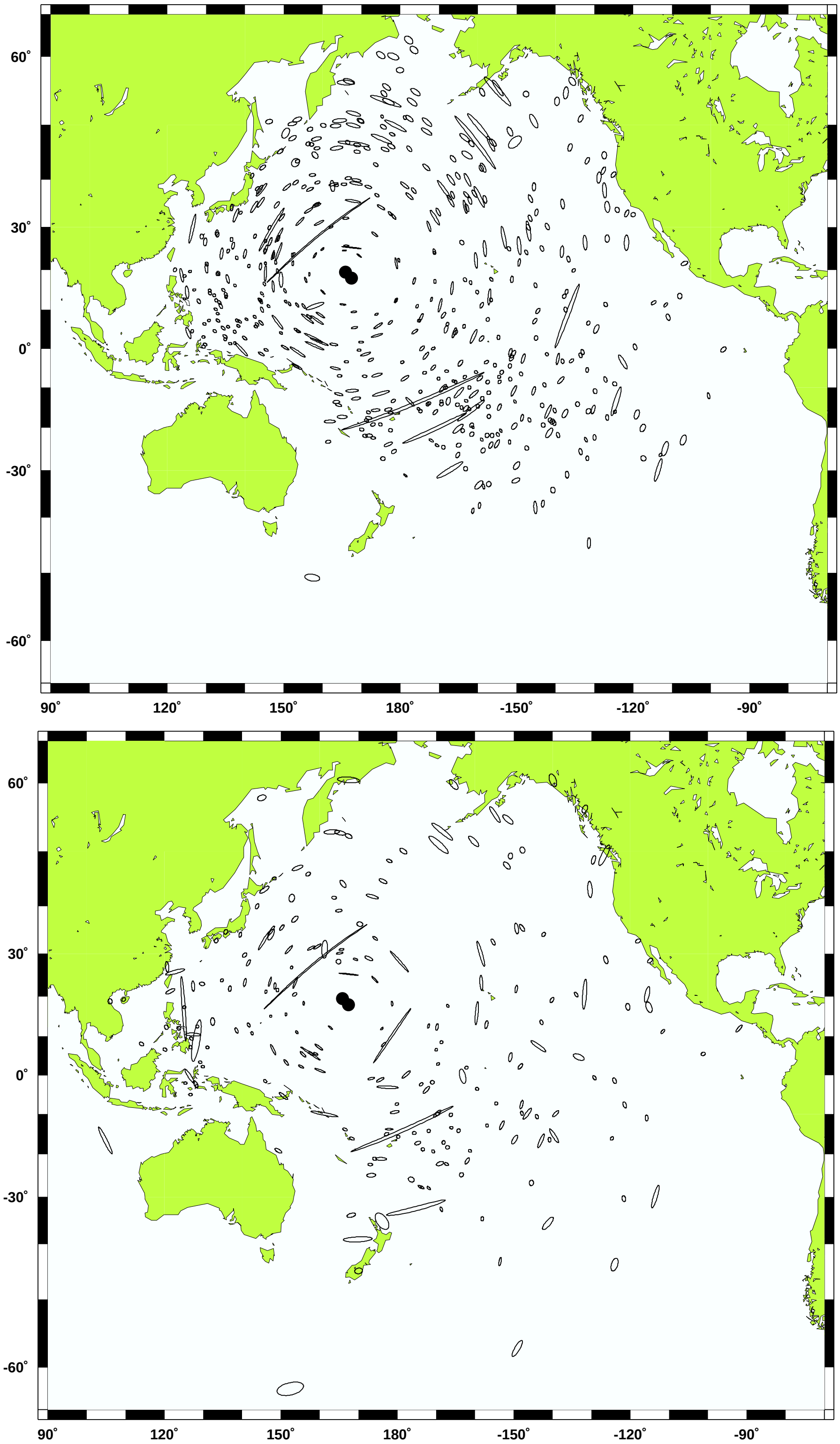

Figure 37: *Uncertainty ellipses for events where the WK30 (a) or WK31 (b) H-phase arrivals participated in the location. The minor axis is mostly in the radial direction from the hydrostations, indicating hydroacoustic arrivals constrain the location more tightly than do other arrivals.*

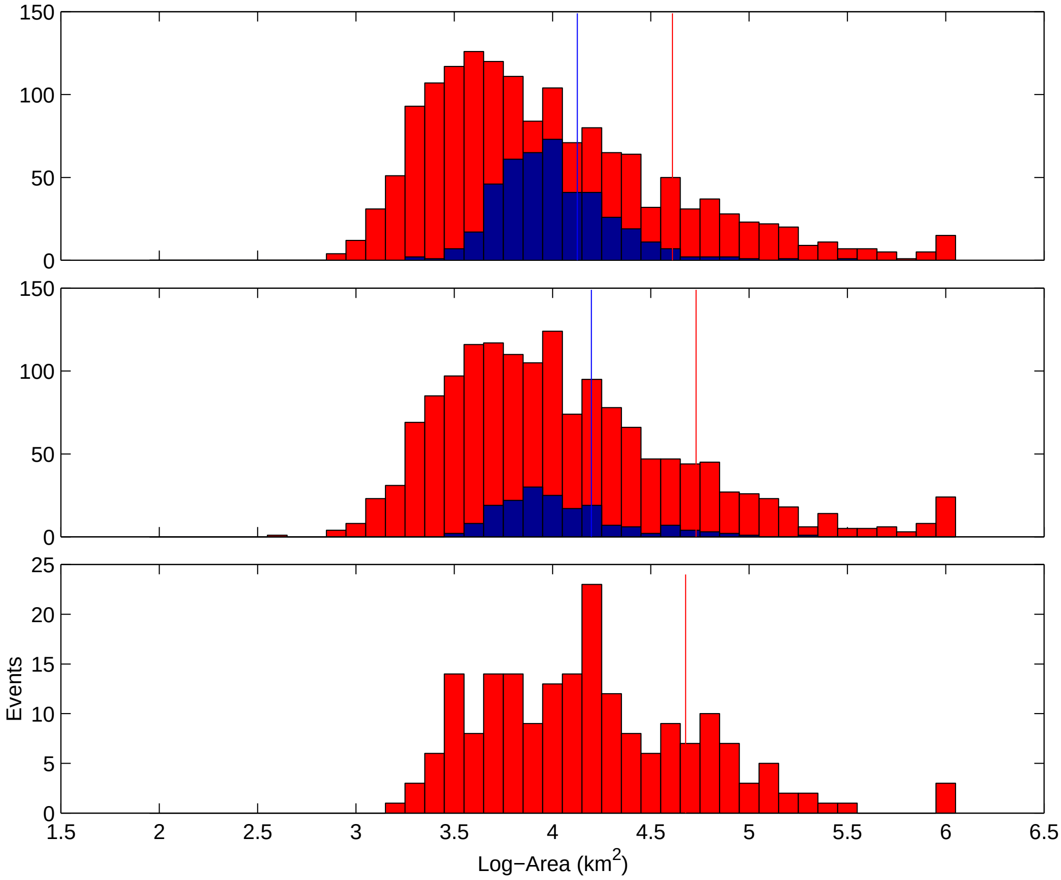

Figure 38: *Histograms of uncertainty ellipse area for events with associated T-phases (Red) and H-phases (Blue) for WK30 (Top) WK31 (middle) and VIB (bottom). The vertical lines indicate the mean for T (blue) and H (red) phase based on a linear (not log) scale.*

.