Geophys. J. Int. (2004) **157,** 589–594 doi: 10.1111/j.1365-246X.2004.02276.x

FAST TRACK PAPER

The Source-Scanning Algorithm: mapping the distribution of seismic sources in time and space

Honn Kao and Shao-Ju Shan

Pacific Geoscience Centre, Geological Survey of Canada, Sidney, British Columbia, V8L 4B2, *Canada. E-mail: hkao@nrcan.gc.ca*

Accepted 2004 February 25. Received 2004 February 25; in original form 2004 January 12

SUMMARY

We introduce a new method, which we call the Source-Scanning Algorithm (SSA), for imaging the distribution of seismic sources in both time and space. Using trial locations and origin times, the method calculates the 'brightness' function by summing the absolute amplitudes observed at all stations at their respective predicted arrival times. The spatial and temporal distribution of sources is then identified by a systematic search throughout the model space and time for the maximum brightness. The greatest advantages of this method are that: (1) it exploits waveform information (both arrival times and relative amplitudes) without the need to calculate high-frequency synthetic seismograms; and (2) it requires neither pre-assembled phase-picking data nor any *a priori* assumptions about the source geometry. A series of tests using synthetic data have shown that this method is robust and can faithfully recover the input source configuration to within 1 grid interval. Finally, we demonstrate the value of the algorithm by locating a typical tremor event with emergent waveforms that occurred during the recent episodic tremor and slip (ETS) sequence in the northern Cascadia subduction zone.

Key words: ETS tremors, seismic source distribution, Source-Scanning Algorithm.

1 INTRODUCTION

The precise determination of the origin time and hypocentre of a seismic source is probably the most fundamental issue in earthquake seismology. Traditionally, this is achieved by minimizing the difference between the observed and predicted arrival times of various phases (for examples, *P* and/or *S*) at a number of seismic stations. Despite significant advances in both theoretical seismology and computational facilities, most seismic networks still rely on phase-picking methods to locate seismic sources.

There are at least two major drawbacks to phase-picking methods that use arrival times only. First, phase picks can be in error if the onset of the arrival is misidentified, especially when the signal-to-noise ratio is poor. Secondly, and probably even more seriously, the proper correlation of individual phases from the same source among different stations becomes extremely difficult when multiple events are closely spread in space and time. In general, phase-picking methods are most effective for individual events that are well separated in time and that generate clear arrivals at seismic stations with a relatively low level of background noise.

There have been major efforts to improve the precision of relative locations within a cluster of seismic events using the travel-time differences between pairs of events (Waldhauser & Ellsworth 2000) or stations (Zhou 1994). However, these methods are subject to the same difficulties as mentioned above. Inversion of seismic waveforms, on the other hand, can provide better estimates of the source distribution in both time and space. The basic principle is that the complete source configuration can be recovered from the constructive and/or destructive interference observed at stations along different azimuths/distances (Hartzell & Heaton 1983). In reality, however, the waveform inversion method has been applied only to a very limited number of relatively large earthquakes ($M_w > \sim 7$). This is at least partially due to our limited ability to calculate synthetic seismograms accurately at short wavelengths (say, less than a few kilometres), where the effects of small-scale heterogeneity become more apparent. Moreover, nearly all waveform inversion methods require the user to make *a priori* assumptions about the geometry and dimension of the fault or faults involved, and this information is often unavailable for smaller events.

In this paper, we introduce a new method of mapping the source distribution without any *a priori* knowledge of the orientation or geometry of the actual fault plane. The method, named the Source-Scanning Algorithm (SSA), exploits waveform information, including both relative amplitudes and arrival times, from an array of seismic stations, to determine whether or not a seismic source is present at a particular time and location. By systematically scanning through a range of trial source locations and origin times, we are able to recover the entire distribution and sequence of seismic sources without the need either to pick arrival times of seismic phases accurately or to calculate synthetic seismograms. A further merit of SSA is that it can take full advantage of a known (and presumably more accurate) 3-D velocity model, reducing the effect of lateral velocity

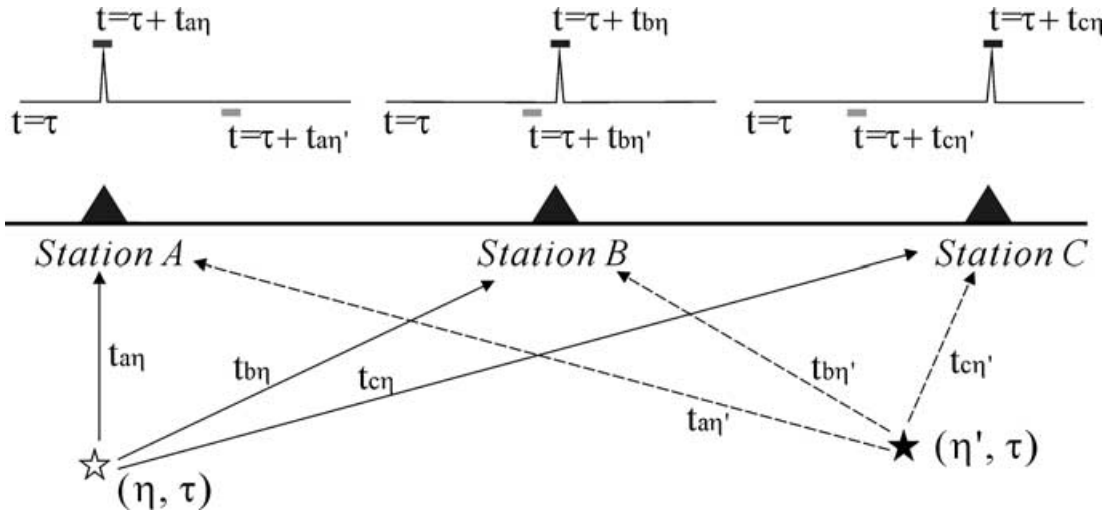

Figure 1. A schematic diagram to illustrate the concept of SSA. The 'brightness' of a point η at time τ is calculated by summing the normalized amplitudes from all stations at the predicted arrival times (i.e. τ plus the respective traveltimes $t_{a\eta}$, $t_{b\eta}$, and $t_{c\eta}$, as marked by the solid short bars). A bright spot (the white star) is found if its location and time are consistent with the arrival of the largest amplitude at each station. A point without any seismic source (η', τ) will have little brightness (the black star) due to the lack of amplitude at the predicted arrival times ($\tau + t_{a\eta'}$, $t_{b\eta'}$, and $t_{c\eta'}$, as marked by the grey short bars).

heterogeneity. We present three examples using synthetic data to demonstrate the robustness of SSA. Finally, SSA is applied to a tremor event in northern Cascadia, illustrating the advantage of this powerful tool over the conventional approach.

2 THEORY

Let us assume that a seismic event is recorded by a seismic array of N stations. We first normalize the amplitude of each digitized seismogram. We then calculate the 'brightness' of a point (η) at a specific time (τ), defined as

$$\mathrm{br}(\eta, \tau) = \frac{1}{N}\sum_{n=1}^{N} |u_n(\tau + t_{\eta n})|, \tag{1}$$

where u_n is the normalized seismogram recorded at station n, $t_{\eta n}$ is the predicted traveltime from point η to station n of a particular phase with the largest observed amplitude (on a regional scale, this means the S phase). If all the largest amplitudes originate from a source at point η and time τ, then $\mathrm{br}(\eta, \tau) = 1$ (Fig. 1). Likewise, $\mathrm{br}(\eta, \tau) = 0.1$ means that a seismic source at point η and time τ would generate, on average, only 10 per cent of the largest observed amplitude at each station. By systematically searching through all η and τ for the local maxima of the brightness function, we effectively reconstruct the spatial and temporal distribution of the seismic sources.

In practice, however, due to our imperfect knowledge of the velocity model, the predicted arrival time of the largest amplitude at each station may be slightly different from that observed. Therefore, instead of using only the amplitude at the predicted arrival time, that is, $u_n(\tau + t_{\eta n})$, we modify eq. (1) to include a contribution from surrounding points within a chosen time window, thus:

$$\mathrm{br}(\eta, \tau) = \frac{1}{N}\sum_{n=1}^{N} \left\{ \frac{\sum_{m=-M}^{M} W_m |u_n(\tau + t_{\eta n} + m\delta t)|}{\sum_{m=-M}^{M} W_m} \right\}, \tag{2}$$

where $2M$ is the number of points within the time window centred around the predicted arrival time, δt is the sampling interval, and W_m is a weighting factor which varies according to how far the arrival time of the energy departs from the predicted arrival time.

Even with the computing power of today's workstations, significant resources are required to calculate the brightness for all points and times at the resolution we might desire, which might be as little as 1 km and 0.1 s when the dimension of the interested region is of the order of a few hundred kilometres. We therefore adopt a practical stratagem to make the scanning more efficient. We first examine the recorded waveform of a particular station (referred to as the reference station, $n = 1$), where the level of background noise is relatively low, to identify the arrival time of the largest amplitude, t_{ar}. The brightness function of all (η, τ) pairs that can satisfy the condition $t_{\mathrm{ar}} = \tau + t_{\eta 1}$ is then calculated, and the pair with the maximum brightness is identified (say, at $\eta_{\max}$ and $\tau_{\max}$). Thus, it is now no longer necessary to search through the entire model space and time, and the calculation of $\mathrm{br}(\eta, \tau)$ is performed only for a subvolume and a time window centred at $\eta_{\max}$ and $\tau_{\max}$, respectively. Depending on the size of the subvolume and the length of the chosen time window, this approach can reduce the scanning time by as much as 75 per cent.

3 CONTROLLED STUDIES USING SYNTHETIC DATA

Fig. 2(a) shows the configuration of our model studies. An array of nine stations is used with a spatial interval of 50 km along both the X- and Y-directions. The first experiment is designed to test the resolution of SSA in space by placing two sources at the same time, but 4 km apart in distance. In the second experiment, to test the resolution of SSA in time, we place two consecutive sources, 1 s apart, at the same location. The time function of each source is 0.5 s in duration with a simple triangular shape. The seismograms calculated for each station and for each of the two situations described above are shown in Figs 2(b) and (c). It is apparent that different source distributions indeed generate subtle differences in waveforms at different stations.

Figs 2(b) and (c) show the results of applying SSA to these suites of synthetic seismograms. We search the model space at grid and time intervals of 1 km and 0.1 s, respectively. In both experiments, sharp bright spots appear at locations consistent with the input sources in terms of epicentres, depths, and times. Some points show

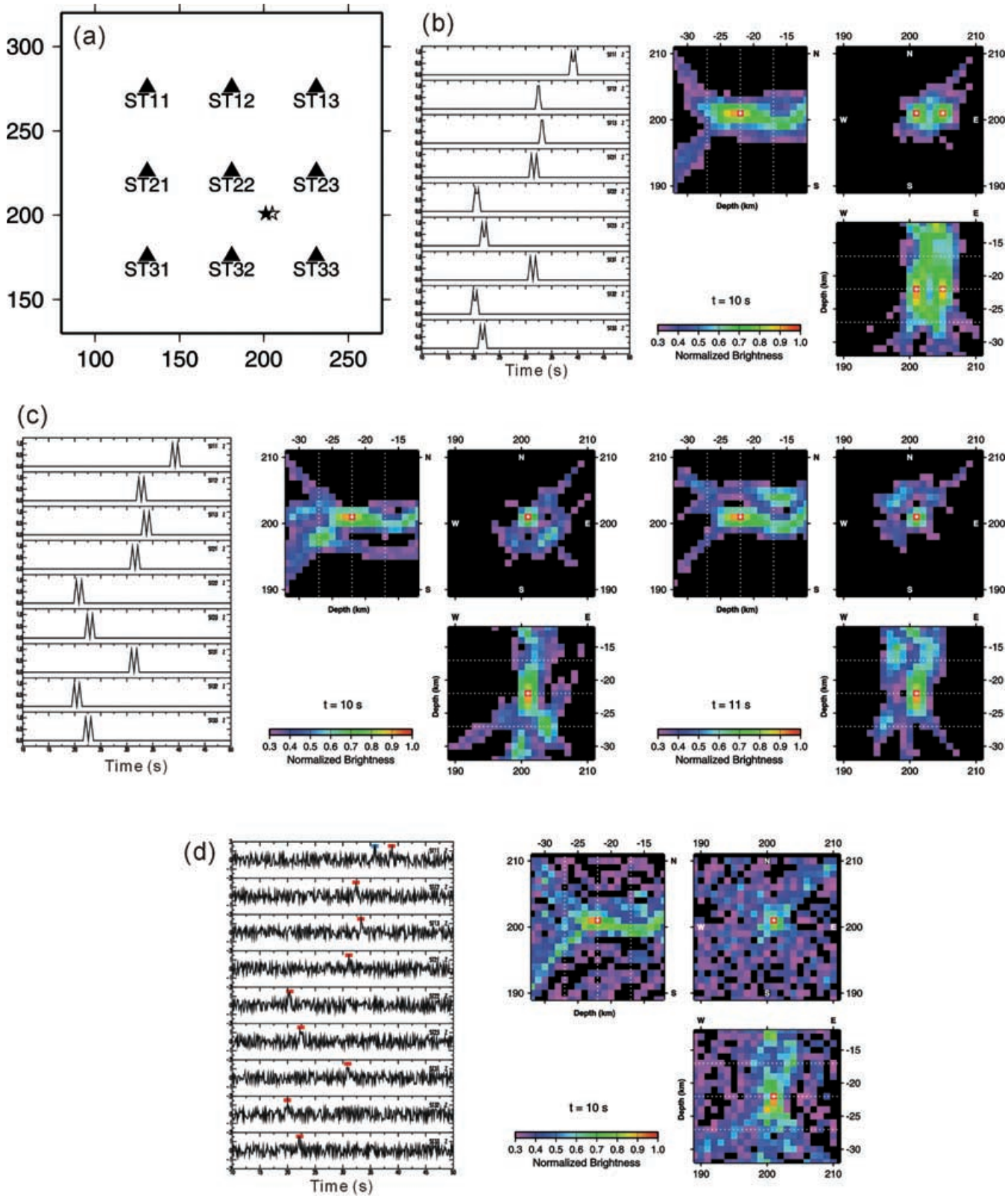

Figure 2. The configuration and results of our controlled experiments. (a) An array of nine stations is placed at 50-km intervals. (b) The first experiment has double sources, both at $t = 10$ s but separated by 4 km [black and white stars in (a)]. The corresponding seismograms and SSA images (a horizontal and two cross-sections passing through the brightest spot) at $t = 10$ s are shown. (c) In the second experiment, a simple impulsive source [black star in (a)] is placed at $t = 10$ and 11 s. The corresponding seismograms are shown in the left panel, whereas the SSA results are shown in the middle and right for $t = 10$ s and 11 s, respectively. (d) The third experiment has a single source at $t = 10$ s, but the noise level in the data set is very high (70 per cent random noise plus a spurious energy peak at ST11, as marked by the short blue bar). The short red bars mark the arrival of signals. In all three cases, the brightest spots match perfectly the original source distribution (white crosses) in both time and space.

non-zero brightness, especially those around the true locations, at incorrect times. It is inappropriate to interpret these spots as seismic sources with smaller magnitudes, as they are simply artefacts resulting from mapping the maximum amplitudes at one or a few stations onto the wrong combination of locations and times. In other words, we must recognize that the nature of SSA incorporates a trade-off between the spatial/temporal resolution and the relative size of sources. If we take a relatively high threshold of brightness for the results shown in Fig. 2, say >0.85, then we correctly recover the source distribution (in both time and space) that generates the maximum amplitudes at all stations. Any source with a smaller size (that is, one that can generate 85 per cent or less of the maximum amplitudes at the same stations) will not be identified. A lower threshold, on the other hand, enables us to recognize smaller sources, but then we fail to pinpoint the precise location and time of each source.

To test the performance of SSA in a realistic situation, we add a high level of random noise (70 per cent of the largest amplitude, which probably represents the worst-case scenario) and a spurious energy peak to the synthetic data. The synthetic seismograms and the SSA results are shown in Fig. 2(d). The added random noise and false peak result in a high background brightness spread over a wide area around the true location (Fig. 2d). Nevertheless, the original source configuration is correctly recovered if the brightness threshold is set at 0.85, as in the previous two cases, demonstrating that the algorithm is remarkably tolerant of noisy data.

We have conducted several tests for a variety of source distributions, including some outside the recording array. In almost all cases, SSA can successfully recover the trial source configurations even when the station coverage is incomplete. However, the spatial/temporal resolution deteriorates the farther the sources are from the receiver array.

4 APPLICATION TO A TREMOR EVENT IN NORTHERN CASCADIA

Seismic tremor activities are documented in both the southwest Japan (Obara 2002) and Cascadia subduction zones (Rogers & Dragert 2003). In Fig. 3, we show the seismograms of a typical tremor activity that occurred in 2003 March along the northern Cascadia subduction zone. Within an interval of an hour, there are several tremor events that give rise to coherent signals across the region. On examining any of them in greater detail, however, it is found that the seemingly coherent signals are actually spread over a very wide time interval and it becomes extremely difficult to determine with any degree of confidence the arrival times of phases coming from the same source (Fig. 3c). Consequently, large uncertainties are expected if the location is determined by conventional phase-picking methods. Furthermore, because no knowledge is available about the corresponding source configuration (that is, the physical dimension and focal mechanism), waveform inversion cannot be applied.

We apply SSA to locate the tremor event shown in Fig. 3. Specifically, we first convert the *P*-wave tomographic results of Ramachandran (2001) to a 3-D *S*-wave velocity model using a V_P/V_S ratio of 1.73. Then, the traveltimes of *S* waves from all gridpoints to all stations are calculated using an efficient 3-D ray tracing code developed by Hole & Zelt (1995) and stored as 'reference files'. Finally, all brightness values are calculated from trial source locations to receivers using a time window of 2.5 s centred about the predicted arrival time with Gaussian weighting.

The result is shown in Fig. 3(d). The brightest spot is located at 48.7342°N, 123.4220°W with a depth of 28 km. The corresponding origin time is 3:44:43.23, 2003 March 9 (UT). In Fig. 3(c), we have marked the record segments that contribute to the calculation of the brightness of this particular tremor source. As expected, all these segments have relatively large amplitudes even though the association among them is not obvious by visual inspection alone. In essence, SSA gives the optimal location and time of the source that is most consistent with the arrival of large signals.

In Fig. 3(e), we show the maximum brightness of each 1-s time step within the ± 50 s time window around the previously determined origin time, to illustrate the temporal characteristics of this particular tremor. Notice that the maximum brightness quickly drops below the 0.85 threshold when the origin time is in error by more than 2 s, suggesting that the source time function of this particular tremor event is relatively short.

As a comparison, we tried to locate the tremor event using the differential traveltimes of the observed amplitude bursts (Zhou 1994). The input data set was prepared by cross correlation between pairs of waveform envelopes of the individual seismograms, but the uncertainty is very large ($\sim$5–10 s). Consequently, the location and origin time cannot be determined as precisely as by SSA.

Based on the results of our previous experiments with synthetic data, we take 85 per cent of the largest brightness as the threshold and estimate the uncertainty to be 3 km in all directions (Fig. 3d). Notice that this uncertainty merely reflects the spatial distance at which the brightness drops below the 85 per cent level and does not take into account the possible existence of less energetic sources in the vicinity.

5 DISCUSSION AND CONCLUSION

The idea of using delay-and-sum methods to enhance signal-to-noise ratios in seismic data processing is certainly not new. For example, the beam-forming method and the common-depth-point (CDP) stacking method shift individual seismograms according to the assumed azimuth/slowness of the wavefield propagating across the array and the depth at which the reflected signals originate, respectively (Arnold 1977; Huang *et al.* 1999; Ingate *et al.* 1985; Mayne 1962). The final solution is identified by the maximum beam energy, which is the stacked sum of the time-shifted waveforms. What distinguishes the method proposed here from previous ones is that SSA does not sum over the entire trace to give the brightness. Instead, it looks for the coherent arrivals of the phase with the largest amplitude, assuming that they originate from the same source. This simple conceptual variation is the essence of SSA and it provides the resolution in both time and space to identify the distribution of seismic sources.

There is no doubt that an inaccurate velocity model can affect the performance of SSA. This is mainly because SSA depends on the calculated traveltimes to determine the brightness of all gridpoints in space and time. In general, the less accurately the subsurface velocity model is known, the longer the time window in eq. (2) will need to be. We have repeated the experiments described in Sections 3 and 4 for a variety of *S*-velocity models. The overall patterns remain largely unchanged, but with lower resolution. As a first-order approximation, we conclude that the length of the time window should be comparable with the possible traveltime error caused by an inaccurate velocity model.

The weighting factor in eq. (2) can be flexible, depending on the quality of data and the accuracy of the velocity model used in the calculation of traveltimes. For example, if the average background noise is minimal and the model is sufficiently accurate, we can use a

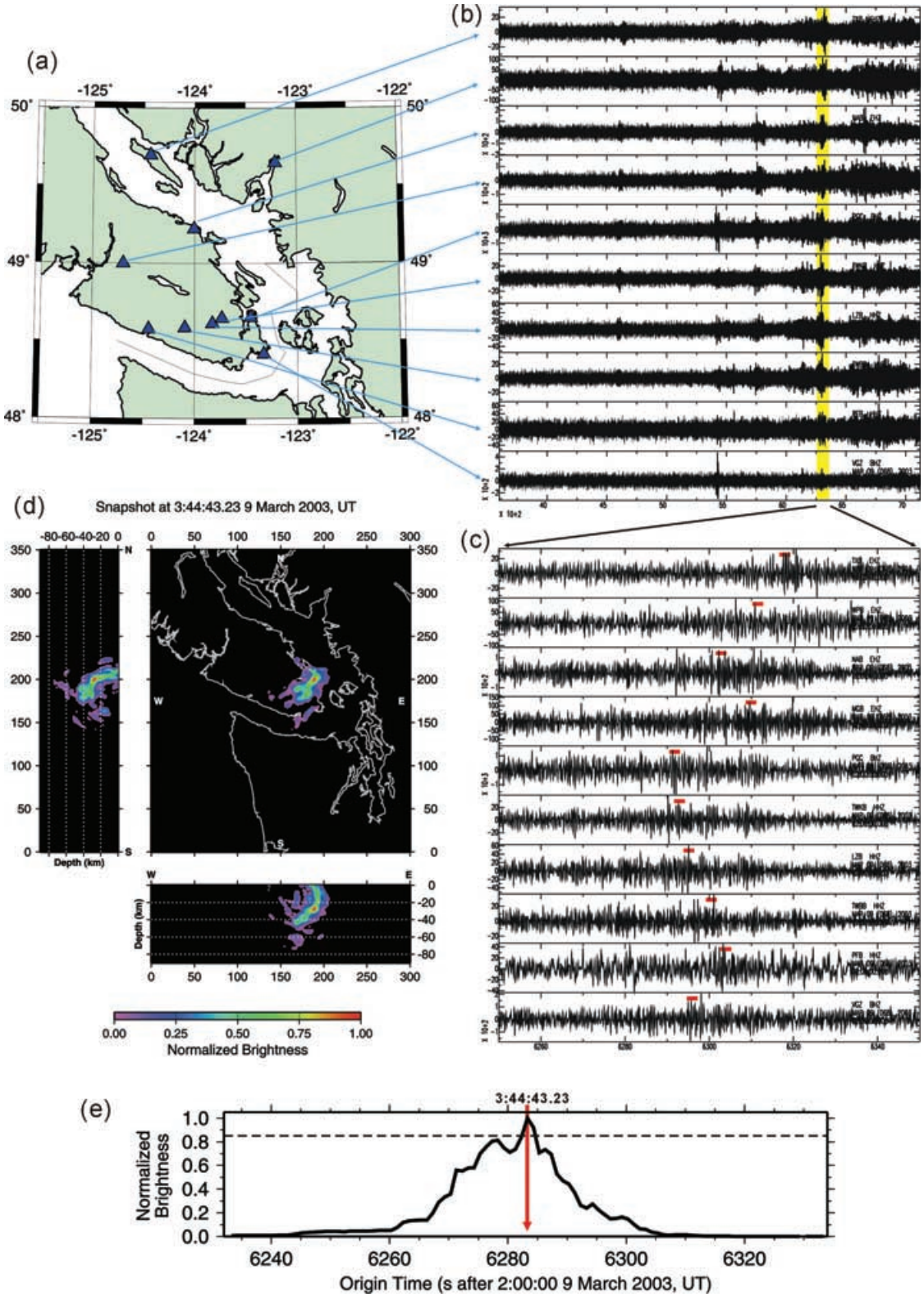

Figure 3. Application of SSA to locating a tremor event in northern Cascadia. (a) Map showing the region and locations of representative seismic stations used in the calculation. (b) Seismograms recorded between 3 am and 4 am, 2003 March 9 (UT). A number of tremors can be clearly recognized during this time window. (c) A zoom-in plot (100 s) showing the detailed waveforms of one particular tremor. The seemingly coherent signals shown in (b) are spread over a very wide time interval, making phase picking extremely difficult. (d) Map and cross-sections showing the SSA result. The coordinates are distances (in km) from the velocity model's origin, as specified in Ramachandran (2001). The brightest spot is located at 48.7342°N, 123.4220°W at a depth of 28 km. The corresponding origin time is 3:44:43.23, 2003 March 9 (UT). The predicted arrival times from this particular tremor source to each station are marked as red bars in (c). (e) Variation of the maximum brightness within the ±50 s time window around the previously determined origin time (marked by the red line). The 0.85 threshold (dashed line) is used to characterize the source duration and uncertainty.

Gaussian weighting function with the half-length of the time window set to twice the expected standard deviation of the arrival time (2σ). On the other hand, if the model is uncertain and/or the noise level is high, then an equal weighting function should be used.

It must be kept in mind that the calculated brightness is not related to the magnitude of the source (in terms of either the released seismic moment or energy) because the individual seismograms are all normalized with respect to their respective maximum amplitudes during data processing. The normalization is required to take care of the geometrical spreading effect during wave propagation so that stations closer to the epicentre will not dominate the calculation of brightness.

Notice that we can modify eqs (1) and (2) to include additional phases (*P*, *PP*, *SS*, depth phases, etc.) as long as the respective traveltime tables are provided. Theoretically, this should reduce the artefacts because the other phases can be matched with the correct predicted arrival times, such that a better resolution is expected. In practice, however, such an approach depends even more on an accurate 3-D velocity model, and this is often unavailable for many regions. Furthermore, most phases at local distances either have smaller amplitudes than *S* or have not yet fully developed. Nevertheless, it is always a good precaution to make sure that the traveltime tables used in SSA match the phases with the largest amplitudes on seismograms. As a rule of thumb, special attention is needed when the refracted phases (for example, S_b or S_n, which usually have small amplitude) arrive ahead of the direct S_g phase at an epicentral distance of $\sim$150 km or more.

Finally, it has not escaped our attention that the proposed method can be employed to obtain critical observational constraints on the rupture processes of earthquakes. For example, SSA is capable of distinguishing the true fault plane from the auxiliary plane by mapping the individual pulses from the rupture front. Similarly, detailed slip distributions of large earthquakes can be directly imaged without a predefined fault geometry. Such applications are beyond the scope of this paper but are well underway for several earthquakes, including the 1999 Chi-Chi earthquake (Taiwan) and the 2001 Nisqually earthquake (Washington State, USA), where there is significant debate relating to different source models arising from the use of different data sets and/or assumptions.

ACKNOWLEDGMENTS

We thank Andrew Calvert and Gilberto Saccorotti for their constructive review. We benefited from inspiring discussions with John Cassidy, Ralph Currie, Herb Dragert, Roy Hyndman, David McCormack, Garry Rogers, and Kelin Wang. Seismic Analysis Code (SAC) by W. C. Tapley and J. E. Tull was used in the data processing. Some figures were generated using the GMT software by P. Wessel and W. H. F. Smith. This work is Geological Survey of Canada contribution number 2003248.

REFERENCES

Arnold, M.E., 1977. Beam forming with vibrator arrays, *Geophysics,* **42,** 1321–1338.

Hartzell, S.H. & Heaton, T.H., 1983. Inversion of strong ground motion and teleseismic waveform data for the fault rupture history of the 1979 Imperial Valley, California, earthquake, *Bull. seism. Soc. Am.,* **73,** 1553–1583.

Hole, J.A. & Zelt, B.C., 1995. 3-D finite-difference reflection traveltimes, *Geophys. J. Int.,* **121,** 427–434.

Huang, B.-S., Chen, K.-C., Yen, H.-Y. & Yao, Z.-X., 1999. Re-examination of the epicenter of the 16 September 1994 Taiwan Strait earthquake using the beam-forming method, *Terr. Atmos. Oceanic Sci.,* **10,** 529–542.

Ingate, S.F., Husebye, E.S. & Christoffersson, A., 1985. Regional arrays and optimum data processing schemes, *Bull. seism. Soc. Am.,* **75,** 1155–1177.

Mayne, W.H., 1962. Common reflection point horizontal data stacking techniques, *Geophysics,* **27,** 952–965.

Obara, K., 2002. Nonvolcanic deep tremor associated with subduction in southwest Japan, *Science,* **296,** 1679–1681.

Ramachandran, K., 2001. Velocity structure of S.W. British Columbia, and N.W. Washington, from 3-D non-linear seismic tomography, PhD thesis, Univ. Victoria, BC, Canada.

Rogers, G. & Dragert, H., 2003. Episodic tremor and slip on the Cascadia subduction zone: The chatter of silent slip, *Science,* **300,** 1942–1943.

Waldhauser, F. & Ellsworth, W.L., 2000. A double-difference earthquake location algorithm: Method and application to the Northern Hayward Fault, California, *Bull. seism. Soc. Am.,* **90,** 1353–1368.

Zhou, H.-W., 1994. Rapid three-dimensional hypocentral determination using a master station method, *J. geophys. Res.,* **99,** 15 439–15 455.