LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

LLNL-JRNL-522935

Signal Processing in Nuclear Test Ban Treaty Verification

D. H. Harris, S. J. Gibbons, A. J. Rodgers, M. E. Pasyanls

January 12, 2012

Institute of Electrical and Electronics Engineers

Disclaimer

This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.

Signal Processing in Nuclear Test Ban Treaty Verification

D. B. Harris[1], S. J. Gibbons[2], A. J.Rogers[3], M. E. Pasyanos[3]

[1]*Deschutes Signal Processing LLC,* [2]*NORSAR,* [3]*Lawrence Livermore National Laboratory*

One branch of signal processing in geophysics has undergone significant, long-term development due to the requirements of nuclear test ban monitoring. There was a burst of activity in the 1960s and 1970s due to the Limited Test Ban Treaty (1963), which banned tests in the atmosphere, underwater and in outer space, and the Threshold Test Ban Treaty (1974), which placed a cap on explosive yield at 150 kilotons. These treaties drove testing underground and created requirements for detecting, locating and identifying explosions, and estimating their yield through observations of seismic waves.

These requirements gave rise to spectacular development of large seismic observatories – for example, the Large Aperture Seismic Array (LASA) with 525 instruments spread over a 200 kilometer (km) aperture in Montana and the Norwegian Seismic Array (NORSAR) with 132 sensor locations deployed over a 120 km aperture. There were matching developments in sophisticated array signal processing [e.g. 1]. At this point monitoring was conducted principally at teleseismic distances – distances greater than 2000 km. At such great ranges the signals are relatively simple, typically with most signal energy concentrated at frequencies around 1 Hz for body waves (compressional and shear waves that travel in the interior of the Earth) and at much lower frequencies (< 0.05 Hz) for guided waves (Rayleigh, Love) traveling along the Earth's surface.

However, as analysis of LASA and NORSAR data made clear, the strong heterogeneity of the Earth places severe limits on processing gains attainable under simple signal models. The defining characteristic of signal processing in seismic verification is the lack of an accurate model for predicting observables. This statement may seem at odds with the great success enjoyed by seismology in defining the interior structure of the Earth. But models of Earth structure generally are insufficiently detailed to predict the structure of signals at frequencies above 0.05 Hz.

The problem became acute in the 1980s and 1990s with the shift to monitoring at regional distances (< 2000km). The transition was driven by the need to monitor smaller explosions, and especially by the Comprehensive Nuclear Test Ban Treaty (CTBT, 1996) which focused attention on explosions around one kiloton yield. At regional distances, the paths that seismic signals take are confined to the upper mantle and crust, which are vastly more heterogeneous than the deep interior of the Earth. Although attenuation tends to suppress high-frequency components of the signal, regional seismic signals often have significant

energy between one and eight Hz (and higher). At these higher frequencies and with the lower velocities of wave propagation found in the crust and upper mantle, the dominant wavelengths of regional seismic signals are consistent with the scale length of heterogeneities and surface topography, leading to strong scattering.

The next generation of seismic arrays consequently had significantly smaller apertures. The NORES, ARCES, FINES and GERES arrays had as many as 25 elements deployed within a 3 km aperture. Current seismic arrays deployed under the CTBT continue this trend with typically 9 sensors deployed within a 3-5 km aperture.

Modern approaches to verification signal processing are being driven by revolutions in the availability of data and the speed of processing resources. Recent decades have seen the deployment of very large networks of stations with real-time digital data transmission by satellite and over the internet. The CTBT International Monitoring System (IMS) consists of 321 monitoring stations in four technologies: seismic, infrasound, hydroacoustic and radionuclide. The seismic network alone (Figure 1) is comprised of 50 real-time primary stations, many of which are arrays, and 120 auxiliary stations supplying data on demand. The Incorporated Research Institutes in Seismology (IRIS) maintains the Global Seismic Network of over 150 stations, and the USArray. The USArray consists of more than 400 stations deployed in stages at 2000 sites across the continental US to image Earth structure in unprecedented detail.

Large-scale application of database technology makes data from these networks available over the internet to interested parties. The Comprehensive Test Ban Treaty Organization maintains an International Data Center (IDC) in Vienna, Austria which supplies data to states party to the treaty. The Incorporated Research Institutes in Seismology (IRIS) maintains an open-access Data Management Center in Seattle, Washington where continuous data from the GSN, the USArray and stations operated by other contributing networks are available freely to anyone over the internet with a variety of access mechanisms, including real-time feeds. Similar data centers exist in northern (NCEDC) and southern (SCEDC) California, the Netherlands (ORFEUS), and Germany (GEOFON).

The convergence of large volumes of organized data, sophisticated algorithms for model development and high-speed computing resources sets the stage for possibly dramatic improvements in signal processing in verification seismology. In Figure 1 we give a high-level sketch of the state of the art in network data processing systems and indicate a possible future extension to these architectures. Current systems

acquire continuous or on-demand data from 100-200 stations globally and run power detectors on individual station waveforms. Beamforming is performed on array station data streams before power detection, usually on a "beam recipe" consisting of some tens of beams directed uniformly in azimuth and slowness.

Following detection, the seismogram is parsed into separate seismic arrivals ("phases"). Figure 1 shows a representative seismogram: the recording of the 11 March 2011 magnitude 9.0 Japan earthquake at station ULN in Mongolia. Three phases are evident: those labeled P and S are compressional and shear phases that travel through the interior (body) of the Earth. The Rayleigh wave is a guided wave trapped along the Earth's surface. It usually exhibits strong dispersion, which is a function of the velocity structure with depth. The processing pipeline makes a large number of measurements on each phase, chief among which are the arrival time (used in location) and amplitude (used in magnitude estimation and event identification).

In most pipelines, subsequent processing is performed on parameter data (the measurements). The phase arrivals across the network are used to define events and are associated with these events. This is a difficult and (especially for automated systems) error-prone task – events frequently are incorrectly built especially when many events are producing hundreds of intermingled phase arrivals across a network. Following association, the phase arrival times are inverted (by generalizations of triangulation) for the event location (including depth and origin time). Finally the amplitudes of seismic phases are used to identify the event type – earthquake or explosion. For example, the mb-Ms ratio is a typical discriminant: it compares the relative log amplitude of the P wave (mb) to that of the Rayleigh wave (Ms). Earthquakes excite Rayleigh waves preferentially and explosions are efficient P wave generators. Location and identification functions make extensive use of calibrations – models correcting parameter data developed from millions of historic measurements.

This paper will discuss a recent trend in using more of the waveform data than the parameters typically extracted following detection. We will focus on waveform matching strategies, in which the continuum of source locations and types is discretized and each discrete source in the collection that results is characterized by a full-waveform template suitable for correlation detection or parameter estimation. The benefit of the template approach is that the detection, association, location and identification pipeline operations that now are distinct functions can be wrapped into a single pattern matching operation. Waveform templates can be extracted from historic observations in an empirical approach or, more controversially, obtained by forward calculation through Earth models. We first describe the empirical

approach, then discuss developments in tomography that offer some reason for optimism about model-based methods.

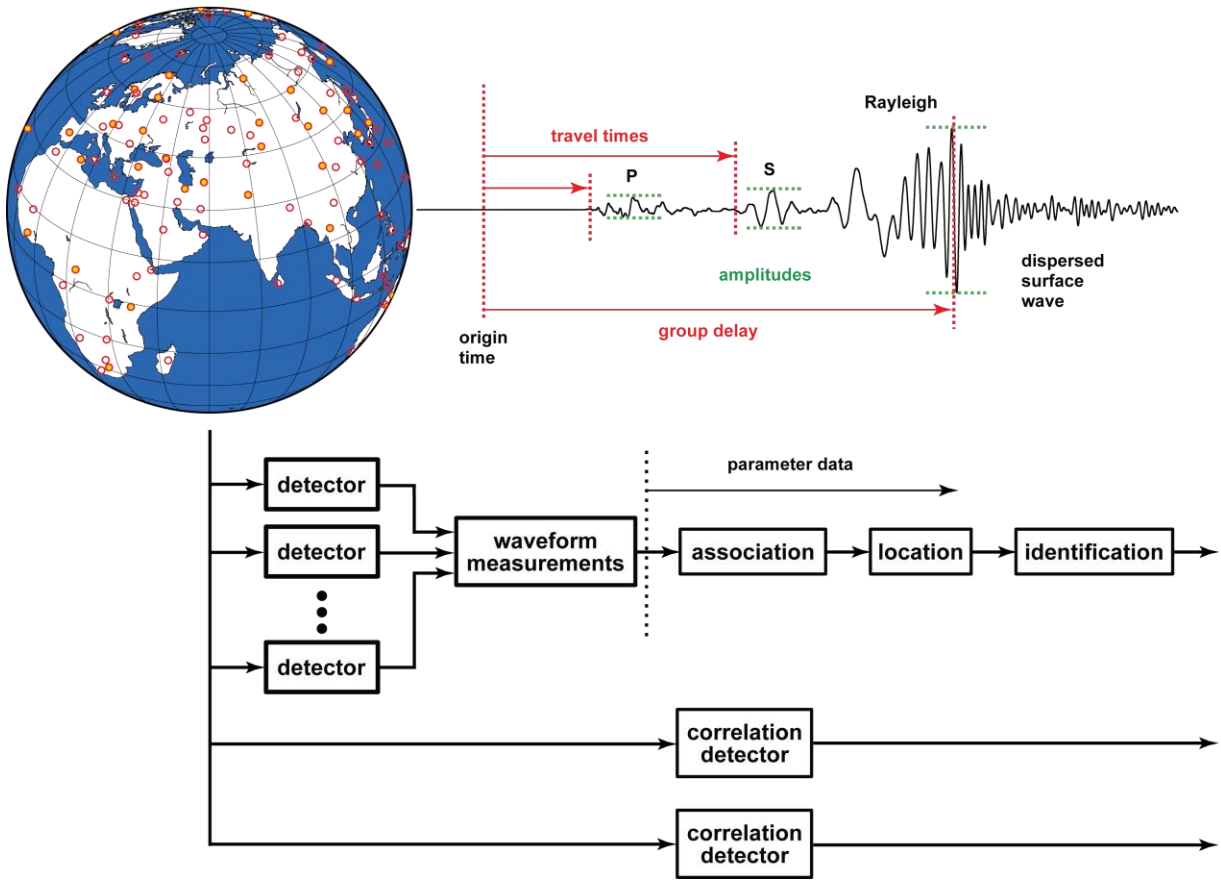

Figure 1 IMS seismic network (Eurasia and Africa shown: the western hemisphere is covered at a similar density) and high-level architecture of a processing pipeline. The processing pipeline reduces waveform observations to parametric data, principally seismic wave arrival times and amplitudes, which are subsequently processed to locate and identify events in separate, consecutive operations. Correlation detectors are a possible future addition to pipelines that would roll detection, association, location and identification functions into a single operation. The waveform is the mb 9.0, 11 March 2011 Japan earthquake recorded at ULN (Ulaanbaator, Mongolia). About 1500 sec of data are shown. IMS seismic stations are shown in the map as solid circles (primary stations) and open circles (auxiliary stations).

Empirical Signal Processing

The "stamp collecting" strategy of cataloging historic waveform observations for use as processing templates is feasible because of two factors. Natural and man-made seismicity is not uniformly distributed over the Earth's surface, but is concentrated in somewhat limited geographic regions and, consequently, often produces recurrent signals. The second factor is the large-scale collection and archival of seismic data.

The use of a waveform template in a detection operation is illustrated in Figure 2 for observations of the 2006 and 2009 North Korean tests made by the Matsushiro array, Japan, about 956 km from the source. This array has 14 sites spread over a 12 km aperture. Waveforms from five individual sensors are shown in the top panel of the figure. The signals from the 2006 (red) and 2009 (black) tests are superimposed. Considering signals from just one event first, the waveforms vary substantially across the array aperture even accounting for propagation delays (this incoherence is worse than at most sites). This situation is markedly different from the type of propagation encountered in radar and sonar applications. In those applications, the spatial structure of wavefields can safely be assumed to be planar over apertures many tens of wavelengths across. The waveforms are identical apart from propagation delays, which can be predicted and removed with a simple model. In that case, optimal detection consists of beamforming to exploit the spatial structure of the signal for processing gain followed by matched filtering of the resulting scalar beam to exploit its temporal structure (if known). That approach is less successful in the seismic application.

In the figure, it is apparent that the waveforms from the 2009 test (in black) are very similar to those from the 2006 test. For this situation, the optimal detection approach is the reverse of that for radar and sonar: temporal matched filters are applied first, channel for channel, followed by beamforming (summing) of the resulting correlation traces [2]. The detection statistic is defined by:

$$\chi\ t\ = \frac{1}{N_s}\ {}_{j=1}^{N_s}\ \frac{u\ \mathbf{x}_j, \tau\ \ d\ \mathbf{x}_j, t+\tau}{d\ \mathbf{x}_j, t+\tau}$$

$$v\ \tau\ \ w\ \tau\ \ =\ \ {}_0^T\ v\ \tau\ w\ \tau\ \ d\tau \qquad\qquad v\ \tau\ \ =\ \ v\ \tau\ \ v\ \tau\ \ {}^1{}_2 \tag{1}$$

$$u\ \mathbf{x}_j, \tau\ \ =\ s\ \mathbf{x}_j, \tau\ \ \ \ s\ \mathbf{x}_j, \tau$$

Here the $s\ \mathbf{x}_j, t$ are the template waveforms recorded at sensor locations $\mathbf{x}_j;\ \ j = 1, \cdots, N_s$ and $d\ \mathbf{x}_j, t$ is the continuous data stream. The second trace in the top panel of Figure 2 shows the individual channel correlations superimposed. The effect of correlation with the 2006 template waveforms is to compress the energy in the 2009 waveforms into relatively narrow pulses that are aligned across the array aperture. These can be stacked (summed) to enhance the post-processing signal-to-noise ratio (SNR) as

can be seen in the first trace of the top panel. The bottom panel of the figure shows the correlation traces (stack on top and individual channels beneath) over a longer time period to emphasize the reduction in noise obtained by stacking.

Correlating the array waveform against a master event template corrects the effects of refraction and scattering across the array aperture in a manner reminiscent of adaptive optics, allowing coherent signal processing at frequencies much higher than on the uncorrected waveforms. In principle, it should allow coherent operations across apertures of arbitrary size, reversing the historic decline in the scale of processing apertures in this discipline. In this example, the signals were large enough to detect on individual array channels. Consequently, the principal value of the correlation operation is to demonstrate that the 2009 event occurred very close (within a wavelength) to the 2006 event and had a similar source mechanism (they were both explosions). For weaker signals, correlation detectors permit detections of events a full magnitude unit smaller than can be detected by beamforming followed by incoherent power detection [3, 4].

As a source-region calibration, the waveform template has a rather limited geographical "footprint", on the scale of the dominant signal wavelength. This concept is illustrated in Figure 3, which shows the locations of events found by six correlation detectors constructed from aftershocks of the 2003 mb 6.5 San Simeon earthquake. This moderately-large earthquake produced thousands of aftershocks, which were well-located by a dense network in the source region (coastal California). The observing station, a primary IMS array (NVAR), was 390 km to the northeast in Nevada. Each waveform template was constructed from 110 sec of an aftershock recording at nine sensor locations. The map of Figure 3 shows the locations of all aftershocks in a 10-day period reported by the dense local network (light grey). Superimposed upon the cloud of catalog aftershock locations are the six smaller clusters of events found by the correlation detectors (various colors), which are about 3 km across on average. As many as one hundred detectors may be required to "cover" this moderate aftershock sequence; many thousands would be required to characterize broader regions of seismicity.

Several variations on the empirical matched filtering concept are under study. To increase the geographic footprint or reduce the sensitivity to source mechanism or time history variations, subspace detectors [5] can be used. In the particular adaptation to seismic problems, [e.g. 6] subspace detectors replace the single multichannel template with an orthonormal basis $u_i \ \mathbf{x}_j, t \ ; \ i = 1, \cdots, N_d$ derived from an ensemble of events characterizing a particular source. The detection statistic measures the projection of a sliding data window into the subspace spanned by the waveform basis:

$$\chi\ t\ =\ \sum_{i=1}^{N_d}\ \sum_{j=1}^{N_s}\ u_i\ \mathbf{x}_j, \tau\ \ d\ \mathbf{x}_j, t+\tau\ \ ^2\ \sum_{j=1}^{N_s}\ d\ \mathbf{x}_j, t+\tau\ \ ^2 \qquad (2)$$

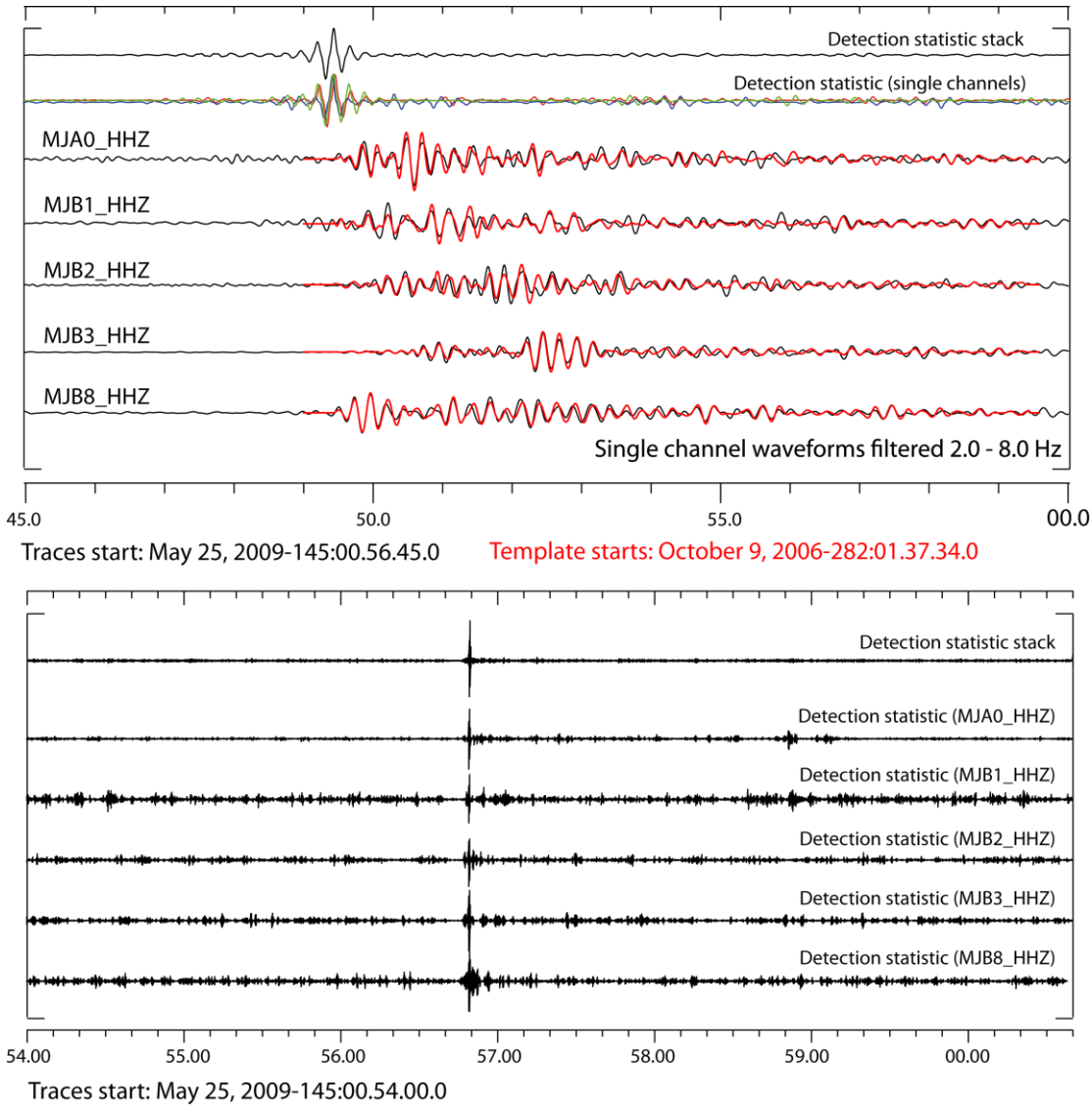

Figure 2 Correlation detection of the 2009 North Korean test using waveforms from the 2006 test as a template. The top panel (time axis in sec), bottom 5 traces are waveforms from individual elements of the Matsushiro array in Japan for the 2009 test in black, and the template constructed from waveforms for the 2006 test from the same elements in red. The second trace shows the corresponding matched filter outputs for the three channels superimposed; the peaks align and when the correlation traces are summed (top), the peak stacks coherently. The bottom panel shows the correlation traces on a larger scale (time axis in minutes).

The basis is obtained with a singular value decomposition of sampled waveforms assembled as columns of a data matrix with the dimension N_d of the representation estimated as the number of significant

singular values. Subspace detectors are most useful in screening mining explosions, which have highly variable waveforms due to the complex spatial distributions of charges used to fracture rock with a minimum of collateral ground motion.

During large aftershock sequences, such as those that followed the 2004 Sumatra and 2011 Japan earthquakes, analytical resources of network operators are overwhelmed by tens of thousands of aftershocks that occur within weeks of the main shocks. This poses a problem for test ban verification, if the resulting "blackout" period makes it difficult to screen events to find explosions. Automated pipeline processes often build events incorrectly when phases from many events are interleaved and superimposed. Analysts must intervene to correctly rebuild the events and can't keep up with the deluge. Correlation detectors offer a potential solution to this problem by organizing aftershocks into large groups that can be reviewed and dismissed efficiently. But the process of template creation, labor-intensive if pursued manually, is too slow.

One possible solution to this problem is to alter pipelines to add correlation detectors dynamically and automatically using the waveforms of valid triggers from power detectors as templates. A prototype system operating from a single array (NVAR) has been successfully tested on ten days of the San Simeon sequence [7], producing the event clusters shown in Figure 3 automatically, in addition to many more. The system ran approximately 100 times faster than real time, produced 676 detections (57% of which were aftershocks) while processing 10 days of data. The detections were clustered into 184 groups, which could have reduced analyst workload by as much as 73%.

Empirical template matching techniques are powerful where applicable, but the extent to which they are applicable is debated. The general approach to answering this question is to estimate the fraction of seismic events that have correlative twins. Estimates of the number of events producing waveforms that correlate with other event waveforms range from a low of 10% to a high of 90-95% [8, 9], depending on the details of window size, bandwidth and correlation threshold. The lower numbers typically are estimated in situations where stations are less dense, limiting observations to large magnitude events at greater ranges. It appears certain that as station density increases and observation histories lengthen (allowing greater linkage among events as event data accumulate), the fraction of correlating events will grow.

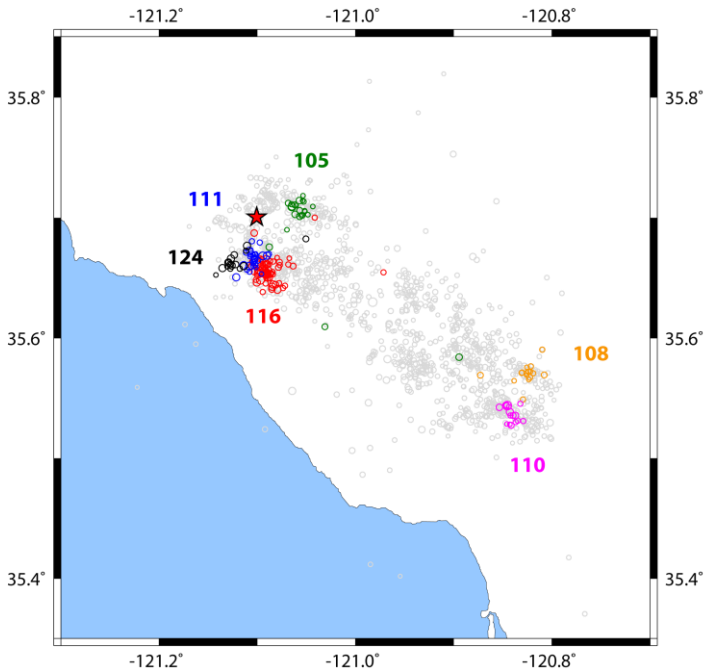

Figure 3 Geographic "footprints" of six correlation or subspace detectors automatically generated for the aftershock sequence associated with the 2003 San Simeon earthquake. The footprints are approximately 2 to 3 km in extent and consist of the locations of events found by the six detectors (various colors represent individual detectors identified by number). The grey symbols indicate the locations of aftershocks reported by a dense local network. Complete coverage of this sequence might require many O(100) detectors. Figure reprinted from Harris and Dodge [7], Copyright © 2011 Bulletin of the Seismological Society of America.

Three-Dimensional Models

As useful as empirical template matching techniques are, they are limited to seismically active source regions with historic observations. Events of interest, particularly underground nuclear tests, may not occur at locations of historic seismicity (tests conducted at established test sites are exceptions). Over most of the Earth's surface waveform observations are unavailable and monitoring techniques are driven by model predictions.

In its early days, seismology enjoyed great success in uncovering Earth structure with relatively simple models represented as a series of homogeneous concentric shells. Two widely used models are PREM, the Preliminary Reference Earth Model [10], derived from observations of long-period surface waves and the Earth's normal modes, and AK135 [11], derived from travel-times of body waves. At regional distances, following an Earth-flattening transformation, Earth structure often is represented as a series of plane layers. In both cases, the models are one-dimensional (1-D), functions only of depth, paramaterized by compressional and shear wave speeds, density and attenuation parameters.

In the monitoring context, Earth models predict body wave travel times used to associate arrivals from a common event, and to locate events. They also predict waveforms from large earthquakes with considerable accuracy in very low frequency bands ($f < 0.05$ Hz) using semi-analytical techniques (e.g. wavenumber integration and normal mode methods). Waveform predictions at low frequencies are used in event location, magnitude estimation and to discriminate earthquakes from explosions. For example, accurate corrections to phase arrival times can be obtained by cross-correlating observed waveforms with model predictions ("synthetics"). And another discriminant for explosions and earthquakes uses direct waveform fits between recorded signals and synthetics. The method chooses parameters of a source model (a moment tensor) to optimize waveform fits with synthetics generated through a tailored 1-D regional Earth model [e.g. 12]. Earthquakes are characterized by two blocks of the Earth slipping past each other along a contact surface (a fault), producing waves that radiate with distinctive cloverleaf patterns. Explosions are nearly isotropic, producing waves that are, in principle, uniform in all directions. Each has a distinctive moment tensor description that allows event identification.

Despite these successes, it is widely anticipated that future developments in verification signal processing will rely upon increasingly more realistic three-dimensional (3-D) Earth models [13]. The new monitoring objective is to detect and interpret smaller events at closer range. As events become smaller, a larger fraction of the declining event energy is generated at higher frequencies. Signal wavelengths become shorter and the effects of lateral refraction and resonance in basins, mountain ranges and other structures become significant. These effects are not captured by 1-D models.

The consequences for traditional monitoring techniques are exemplified by the mb-Ms discriminant. It begins to lose effectiveness for events below Richter magnitude 3.5 because Rayleigh waves become difficult to observe above ambient noise. The problem can be ameliorated by correlating the observed waveform with phase-matched filters, which are all-pass filters designed to have a group delay that matches the dispersion of the Rayleigh wave along the path from source to receiver [14]. Phase-matched filters compress Rayleigh wave energy into a measurable higher-amplitude pulse. The proper group delay is a path integral of Rayleigh wave group velocities predicted by shear-velocity Earth models. At the higher frequencies characteristic of smaller events, higher-resolution models are essential.

All pipeline functions based on full-waveform synthetics are similarly affected, including any future attempts to use synthetics as templates for correlation detectors. Fortunately, there are significant advances in seismic tomography that offer the prospect of higher resolution and extended coverage (both currently very limited) for 3-D models.

Ambient Noise Tomography

For years the dominant approach to seismic tomography was based on observations of transient signals from discrete events. The two principal approaches invert body wave travel-time measurements and surface wave dispersion measurements for Earth structure, usually separately. These approaches suffer limitations because earthquakes, the principal discrete tomographic sources, are not ideally distributed for imaging, and may have uncertain locations and mechanisms. Significant stretches of the Earth (e.g. central Siberia, North Africa) are virtually aseismic. In addition, stations must be deployed for years to record observations to be useful in an inversion.

Ambient noise tomography (ANT) belongs to the class of methods that invert surface wave dispersion measurements for Earth shear-velocity structure. Surface wave group delays are path integrals of surface wave velocities in particular frequency bands. With enough crossing paths (tens of thousands typically) the path integrals can be inverted to produce velocity maps [e.g. 15]. The velocity maps, which are functions of frequency, then collectively are inverted for 3-D shear velocity [16]. ANT differs from other techniques in it source of group delay measurements. These are obtained from large-scale cross-correlations of ambient noise between pairs of stations, which have an asymptotic tendency to reveal the Green's functions describing propagation between the station locations [17-19]. The derivative of the cross-correlation function theoretically tends to the sum of forward and reverse Green's functions between the observing stations [18-19]:

$$\frac{\partial C_{ij}\ \mathbf{x}_1, \mathbf{x}_2, t}{\partial t} = -G_{ij}\ \mathbf{x}_1, \mathbf{x}_2, t\ + G_{ji}\ \mathbf{x}_2, \mathbf{x}_1, -t \tag{3}$$

Here $C_{ij}\ \mathbf{x}_1, \mathbf{x}_2, t$ is the cross correlation of ground motion in direction i at station location $\mathbf{x}_1$ with ground motion in direction j at station location $\mathbf{x}_2$ [18]. $G_{ij}\ \mathbf{x}_1, \mathbf{x}_2, t$ is the Green's function relating a unit impulse displacement at one location to ground motion at the other. The seismic wavefield is a vector (elastic) field with 3 components of motion.

At continental scales, the extraction of Green's functions has been most successful at frequencies below 0.16 Hz (periods above 6 sec). At local scales, in southern California, Green's functions have been extracted in the 0.1-2 Hz band [18]. Ambient seismic noise at low frequencies is largely created by storms at sea, and is dominated by surface waves, though P waves have been observed at short range [20]. ANT has the distinct advantage that it does not require earthquake or explosion sources. The data from

stations can be used almost immediately to produce measurements. It allows model development from temporary deployments, which can involve large numbers of stations, with quite dense spatial sampling.

The USArray presents an exceptional opportunity to apply ANT on a continental scale to produce velocity maps with unprecedented resolution. The station spacing is on the order of 70 km and up to two years of data are available for stations in the array. Lin et al. [19] analyzed data recorded continuously by over 250 stations in the western United States from November 2005 through October 2006 and produced the vertical-motion ambient noise correlation functions shown in Figure 4. The figure displays correlations between one fixed station (MOD on the Modoc plateau in northeastern California) and many other stations at increasing distances from MOD.

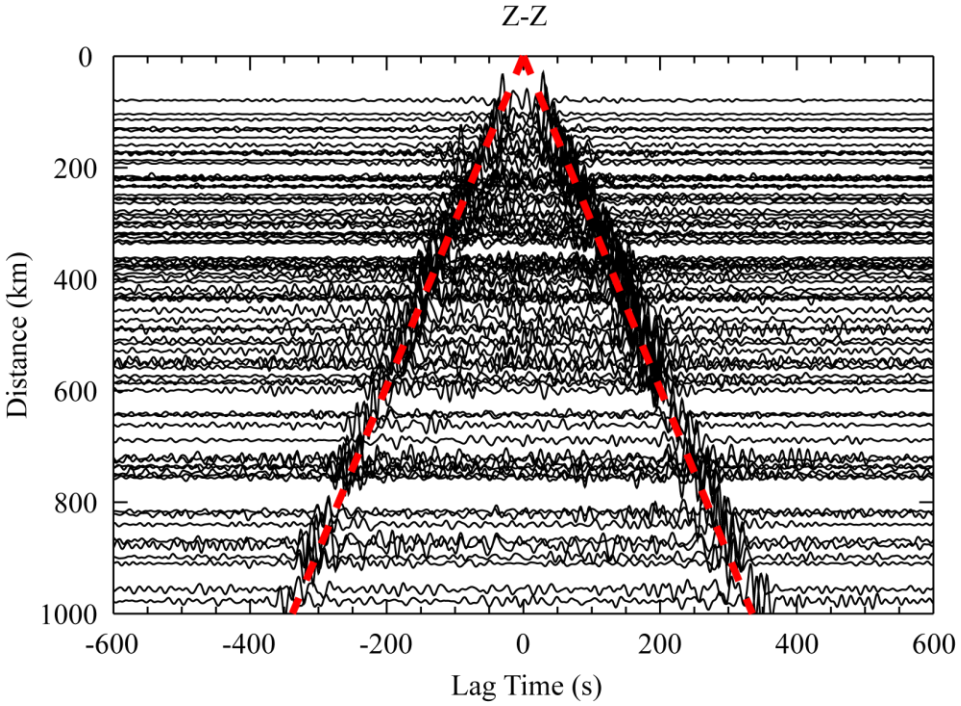

Figure 4 Vertical-motion correlation functions in the 0.02 - 0.1 Hz frequency band (10-50 sec period) centered on station MOD in the northeast corner of California [19] Cross-correlations between the vertical-component sensor at MOD and vertical sensors at other stations are displayed as a function of distance from MOD. The evident signals are Rayleigh waves, and the dashed lines indicate travel-times corresponding to a velocity of 3.0 km/sec. Figure modified from Lin et al. [19] and reprinted by permission of Geophysical Journal International. Copyright © 2008 The Authors, journal compilation copyright © 2007 RAS.

Lin et al. [19] made a total of 31,878 station-pair correlation function calculations using the USArray stations at locations indicated by triangles in Figure 4, and screened the measurements in a selection of frequency (period) bands for signal-to-noise ratio and a criterion ensuring far-field propagation. Phase

delays (the integral of group delays) between station pairs were measured in this case and inverted for Rayleigh velocities, producing the maps shown in Figure 5. These maps show a considerable amount of detail consistent with known Earth structure in the western United States.

Since Rayleigh waves are dispersed, guided waves trapped near the surface of the Earth, the amplitudes of their fundamental modes decline with depth at rates dependent upon frequency. The shorter period (higher frequency) surface waves are sensitive to shallower structure than their longer period (lower frequency) counterparts. Comparison of the 8 sec map to the 20 sec map in Figure 5, gives a sense of the 3-D nature of apparent physical features. The low-velocity structure of the Great Central Valley and the high-velocity structure of the Sierra Nevada mountains in California are obvious in the 8-sec maps, but less distinct in the 20-sec maps. This observation suggests that these structures are shallow.

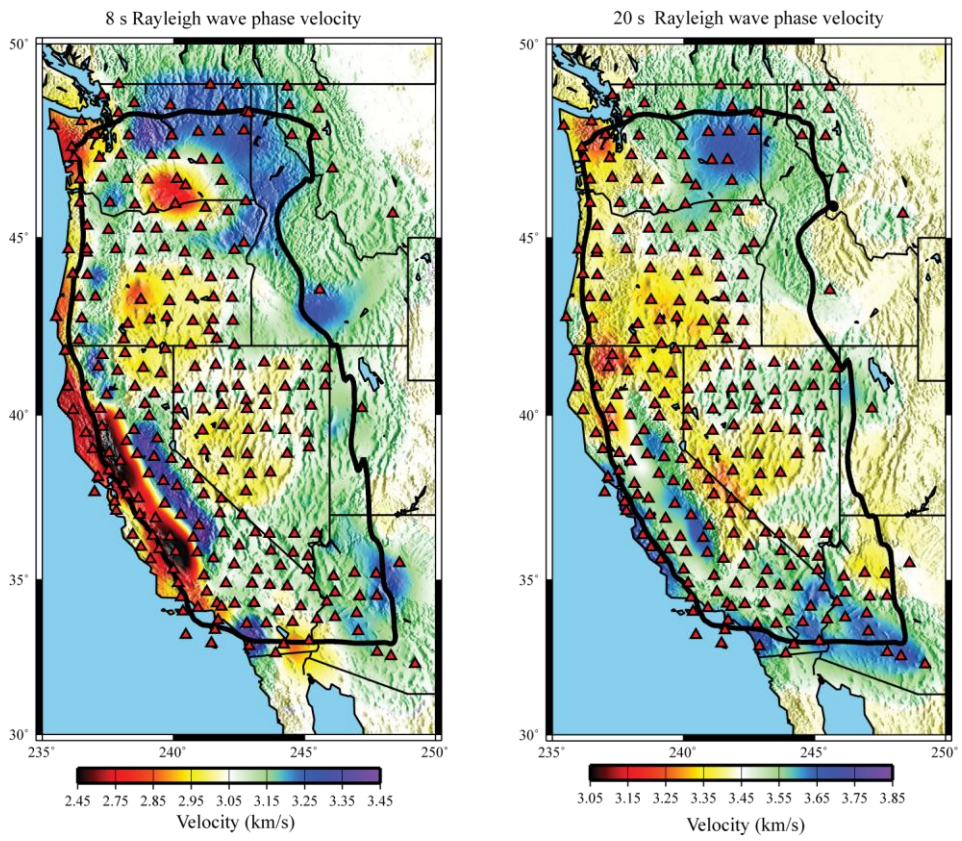

Figure 5 Rayleigh wave velocity maps at periods of 8 sec (left) and 20 sec (right). These maps clearly show slow (red) and fast (blue) features consistent with known Earth structure. The Great Central valley of California is a pronounced slow feature due to thick sediments. Several mountain ranges (e.g. the Sierra Nevada) show as fast features due to their crystalline (high velocity rock) roots. Areas of high heat flow are slower and appear in the volcanic province of south Central Oregon and areas of thinning crust in the Great Basin of Nevada. Figure modified from Lin et al. [19] and reprinted by permission of Geophysical Journal International. Copyright © 2008 The Authors, journal compilation copyright © 2007 RAS.

ANT is being applied widely around the globe and can be expected to improve Earth models to map major 3-D features with resolutions (depending on station density) on the order of hundreds of kilometers.

Adjoint Tomography

Adjoint tomography is one of the more intriguing developments in seismic tomography of the last decade. It differs from classical tomographic methods by employing much more realistic, finite-frequency, full-waveform modeling of seismic waves between sources and receivers. Traditional methods consider propagation in the infinite-frequency limit; the paths that seismic energy takes in this limit are described by ray theory. Often refraction due to horizontal variations in P and S speeds is neglected in traditional methods, so that ray paths are approximated by great circles between sources and observing stations.

Adjoint tomography was introduced to seismology by Tarantola [21]. It relies upon the reciprocity principle, time reversal ideas and the Born approximation. Practical implementations require large-scale, highly accurate simulations of wave propagation in 3-D models. These simulations require high-performance computing using finite difference or spectral element solvers for the wave equation [22] that themselves are the product of years of development.

Following Tromp et al. [23], adjoint methods obtain models as solutions to an optimization problem to minimize an objective function $\Phi\ \mathbf{m}$, describing misfit between waveforms observed by sensors at a collection of locations $\mathbf{x}_j$ and waveforms predicted by an Earth model described by parameter vector $\mathbf{m}$:

$$\Phi\ \mathbf{m} = \int_{j} \int_{0}^{T} \phi\ \mathbf{x}_j, t, \mathbf{m}\ dt \tag{4}$$

The simplest objective measures the direct waveform misfit over a collection of events indexed by i:

$$\phi\ \mathbf{x}_j, t, \mathbf{m} = \frac{1}{2} \int_{i} \mathbf{d}_i\ \mathbf{x}_j, t - \mathbf{s}_i\ \mathbf{x}_j, t, \mathbf{m}\ ^2 \tag{5}$$

Here the $\mathbf{d}_i$ are observed waveforms and the $\mathbf{s}_i$ are their predictions obtained by solving the 3-D wave equation with medium parameters described by $\mathbf{m}$ and a forcing function representing event i . Another frequently used objective measures misfits of travel-time estimates for specific phases obtained by

maximizing the cross correlation between the observed and model waveforms. All such objective functions are strongly nonlinear with respect to **m** and must be minimized by nonlinear optimization (typically conjugate-gradient) methods.

The adjoint method makes possible efficient calculation of the Frechet derivative of Φ with respect to model perturbations $\delta\mathbf{m}$ [24], suitable for use in a conjugate-gradient algorithm. In an isotropic medium, the Frechet derivative can be shown to be an integral of sensitivity kernels $K_{\blacksquare}\ \mathbf{x}$ multiplied by model perturbations over the model volume [22, 23]:

$$\delta\Phi = \int_V K_\rho\ \mathbf{x}\ \ \delta \ln \rho\ \mathbf{x}\ + K_\mu\ \mathbf{x}\ \ \delta \ln \mu\ \mathbf{x}\ \ + K_\kappa\ \mathbf{x}\ \ \delta \ln \kappa\ \mathbf{x}\ \ \ d\mathbf{x} \tag{6}$$

The medium parameter functions $\rho\ \mathbf{x}$ (density), shear modulus $\mu\ \mathbf{x}$, and bulk modulus $\kappa\ \mathbf{x}$ are parameterized by **m**.

The sensitivity kernels are intuitively appealing because, as Frechet derivatives of the objective function, they highlight places in the model where the medium parameters need to be changed to reduce model misfit, i.e. they are physical visualizations of the misfit gradient. Figure 6 shows a remarkable travel-time sensitivity kernel due to Tape et al. [25] for a surface wave traveling between an earthquake and station in southern California. This kernel shows that the path is strongly refracted by horizontal velocity gradients, indicating that traditional tomography algorithms based on great-circle-path assumptions probably would fail to produce adequate models in this region.

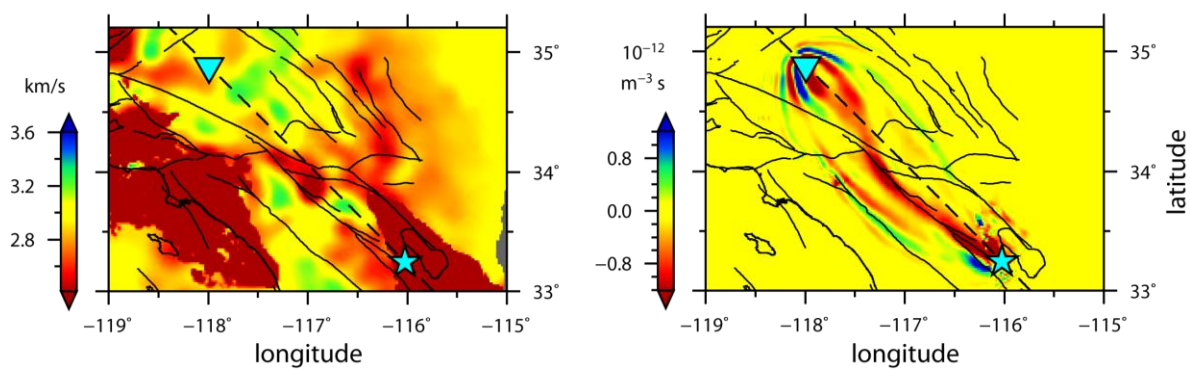

Figure 6 Map (left) of the best shear velocity model obtained by adjoint tomography at 0 km depth. Sensitivity kernel (right) for a Rayleigh wave for a single source-receiver pair in southern California, computed in the 3-30 sec band. The finite-frequency sensitivity kernel shows that the Rayleigh wave is sensitive to velocity structure over a broad region (not an infinitesimal ray) that is not necessarily on the direct path (dashed line) between source and receiver. In this case, the path avoids the region of slow velocity directly between source and receiver. Figure modified from Tape et al. [25] and reprinted by permission of Geophysical Journal International. Copyright (c) 2009 The Authors, journal compilation (c) 2009 RAS.

The sensitivity kernels are obtained through interaction between the forward solution $\mathbf{s}_i$ to the wave equation obtained for physical source i and an adjoint solution $\mathbf{s}_i^\dagger$ obtained by driving the wave equation with a forcing function obtained as a sum of adjoint sources at the receiver locations:

$$\mathbf{f}_i^\dagger\ \mathbf{x}, t \quad = \quad {}_j \quad \nabla_{\mathbf{s}_i}\phi\ \ \mathbf{x}_j, t, \mathbf{m} \quad \delta\ \ \mathbf{x} - \mathbf{x}_j \tag{7}$$

For the waveform misfit objective of equation (5), the adjoint sources are time-reversed residuals between the observed and predicted wavefields at the sensor locations [23]. The interaction occurs as the forward solution propagating outward from the source toward the stations overlaps with the adjoint solution propagating backwards from the stations toward the source, for example:

$$K_\rho\ \mathbf{x} \quad = \ - \quad {}_i \quad {}_0^T \rho\ \mathbf{x}\ \ \mathbf{s}_i^\dagger\ \ \mathbf{x}_j, T - t, \mathbf{m}\ \cdot \frac{\partial^2 \mathbf{s}_i\ \ \mathbf{x}_j, t, \mathbf{m}}{\partial t^2} dt \tag{8}$$

Computation of the sensitivity kernel requires two solutions to the 3-D wave equation for each source i, one to obtain the forward wavefield and the second to obtain the adjoint wavefield. Both solutions must be available simultaneously for the entire time interval of the calculation $0, T$ and spatial domain V. This requirement can be met by storing the forward wavefield for all space and time at a sufficient resolution to allow the integral in (8) to be evaluated as the adjoint field is computed. However, the size of the computed field makes this approach infeasible even on the largest machines. The solution to this problem is to calculate the forward wavefield twice, once to obtain the initial conditions for the adjoint calculation and a second time jointly with the calculation of the adjoint field so that both are available simultaneously [23]. This approach doubles the memory requirement of a single forward wavefield calculation. The computational expense is three times the cost of a single forward calculation for each source used in the inversion, and is nearly independent of the number of stations, as the adjoint field is computed for all stations simultaneously.

The procedure described above is required for a single conjugate gradient iteration. Figure 7 illustrates the result of applying this approach, using a travel-time misfit objective, to optimization of a 3-D model of the southern California crust using 143 earthquakes and a dense network of stations [25,26]. Sixteen conjugate gradient iterations were required to obtain the final model shown in Figure 7 (top right), which shows significantly greater detail and larger velocity variations than the starting model (top left). The calculation required over 0.8 million CPU hours. From the signal processing standpoint, the most notable

characteristic of this figure is the dramatic improvement in the fit of the model predictions to the observed waveforms as the number of conjugate gradient iterations increased. The bottom panels of the figure show waveform fits between observed and predicted waveforms for a particular earthquake-station pair. The period band represented here is 6-30 sec, which is a significantly higher frequency band for good waveform predictions than normally obtained. This result is the more striking for the fact that data from this earthquake were not used in the inversion process.

Since this kind of fit is observed for the majority of source-receiver paths in this region with the new model, it is likely that coherent processing could be applied across the network with model-based waveform templates in this period band.

Stochastic Tomography

Most Earth models are deterministic, providing no estimate of uncertainty for the parameters describing Earth structure. However an uncertainty model is crucial for monitoring applications, since estimates of errors in the predicted observables are required to put bounds on the errors of estimated event locations, magnitudes and discriminant statistics. The CTBT, for example, allows on-site inspections (OSI) in the event of a suspected violation (clandestine nuclear test), but the region that may be inspected is limited to 1000 km^2. A state party requesting OSI will need an accurate estimate of event location error to know whether this criterion has been met.

Stochastic inverse methods are an attractive alternative to deterministic approaches. The models they produce represent Earth parameters in probabilistic terms, using Bayesian inference to combine prior constraints on model parameters with observed data to obtain posterior probability densities for the parameters. The Markov-chain Monte Carlo (MCMC) method [27] is particularly attractive, since it explores model space without becoming trapped in local minima. It has been used for some time in geophysics [e.g. 28] and recently to obtain stochastic geophysical models in regions of monitoring interest [29,30]. The method produces a representation of the posterior density with a large collection of model realizations sampling model space in a manner proportional to posterior probabilities.

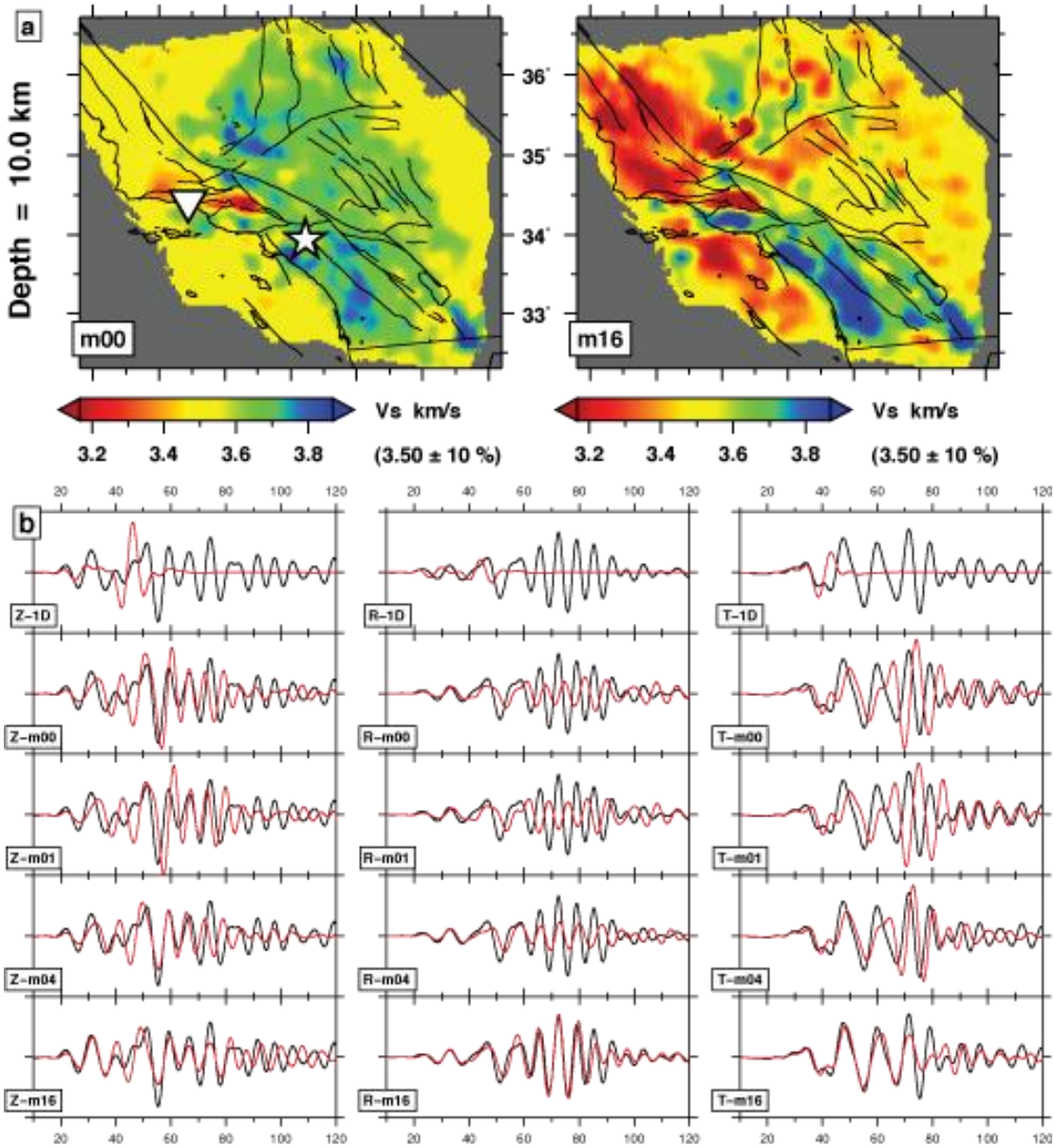

Figure 7 (a) The starting model (shear wave speed) for the full-waveform adjoint travel-time inversion of crustal structure in southern California is shown at left. The adjoint model after 16 iterations is shown at right. Faults are shown as black lines. The adjoint model has significantly greater detail in the coast ranges and basins, and shows stronger velocity contrasts across faults where blocks of different material are in contact. (b) Comparison of observed waveforms (black) and waveforms predicted by various models (red) for the source (star) - receiver (triangle) pair shown on the map at left. Vertical ground motion is shown in the left column, radial motion (along the line between source and receiver) is shown in the center column and transverse motion (perpendicular to the line) is shown in the right column. The top row shows 1-D model results. The 3-D starting model results are shown in the second row and 3-D model results after 1, 4 and 16 iterations of the adjoint conjugate gradient inversion are shown in rows 3-5. Good waveform matches are obtained after 16 iterations. Figure modified from Tape et al. [25] and reprinted by permission of Geophysical Journal International. . Figure modified from Tape et al. [25] and reprinted by permission of Geophysical Journal International. Copyright (c) 2009 The Authors, journal compilation (c) 2009 RAS.

As an example, we take the model of the European Arctic reported by Hauser et al. [30], a region which includes the former Soviet test site on the island of Novaya Zemlya (Figure 8). This model was constrained simultaneously by a variety of data, including Bouguer gravity, surface wave group delay and body wave (P and S) travel-time measurements. The model was parameterized on a set of 592 nodes on a uniform hexagonal grid. At each grid point, the P velocity, S velocity and density were represented as piecewise linear functions of depth in four layers of variable thickness: water, sediment, crystalline crust and upper mantle. The MCMC method simplifies the application of physical constraints such as bounds on the thickness (<10 km) of oceanic crust.

The MCMC algorithm, following Mosegaard and Tarantola [28], consists of a mechanism for proposing models, a series of tests for model acceptance and an archive of accepted models (which form the posterior distribution). The algorithm constructs a chain of models sequentially with a base sampler that proposes new models by perturbing the latest model in the chain subject to a set of rules that define a prior distribution $p\ \mathbf{m}$ on the set of acceptable models. The posterior distribution $P\ \mathbf{m}|\mathbf{d} = L\ \mathbf{m}\ p\ \mathbf{m}$ is constructed by accepting or rejecting models based on the likelihood function:

$$L\ \mathbf{m} = \exp\ -\frac{1}{N}\ {}_{i=1}^{N}\ \frac{d_i - d_i\ \mathbf{m}}{\sigma_i \sigma}^{2} \tag{9}$$

where there are N observations d_i , $d_i\ \mathbf{m}$ is the model prediction for observation i, σ_i is an estimated uncertainty for the observation and σ is an overall uncertainty assigned to the data, which controls the tradeoff between the prior model and the fit to the data. In the Metropolis-Hastings algorithm, model $\mathbf{m}_{j+1}$is accepted if $L\ \mathbf{m}_{j+1} \geq L\ \mathbf{m}_j$ or if $L\ \mathbf{m}_{j+1} < L\ \mathbf{m}_j$ and $L\ \mathbf{m}_{j+1}\ /L\ \mathbf{m}_j$ is larger than a random deviate uniform in $0,1$.

It is common to run the algorithm multiple times from different, randomized starting models to test the completeness of model space exploration. Each run produces different chains (12,000 samples long in the Hauser et al. [30] model; each sample model had close to 6,000 parameters). Experience with the MCMC method shows that the early portions of the chains are consumed with moving the models toward that portion of model space where the data are fit well. Only the last 4,000 models of each chain were used to represent the posterior distribution in the European Arctic model. Results of separate runs were checked to ensure that the equilibrium results were statistically similar.

Figure 8 (lower row) shows the equilibrium distribution of model profiles for P velocity in two geographic locations, one south of Spitsbergen in the west where the model is well-constrained by large numbers of observations, and one to the east of Novaya Zemlya where data were sparse. The figure superimposes prior and posterior distributions in both regions. Prior distributions are obtained by running the base sampler in isolation. In both regions, the prior distributions show similar spread, but the region with good data coverage has a much narrower distribution of models than the area with sparse data coverage.

A practical example of using the stochastic model is shown in Figure 9. Here the 4,000 individual models comprising the posterior distribution each were used to estimate the location of an earthquake to the south of Spitsbergen. Five stations were used in the location estimates, as shown in the map at upper left. At bottom left are distributions of predicted average speeds along 2 source-station paths for the 4 observed body phases contributing travel times for estimating location. This plot shows that posterior model uncertainties can be translated into uncertainties on the observables. To the right are the 4,000 location estimates made with the individual models. The resulting cloud of location estimates demonstrates that model uncertainties are translated into estimated parameter uncertainties.

For signal processing applications, stochastic models of this sort provide a convenient mechanism for estimating waveform uncertainty. Synthetics can be generated through models sampled from the posterior distribution to represent a signal subspace predicting waveform observations along a particular path. A subspace representation is a suitable stochastic generalization of a deterministic signal model.

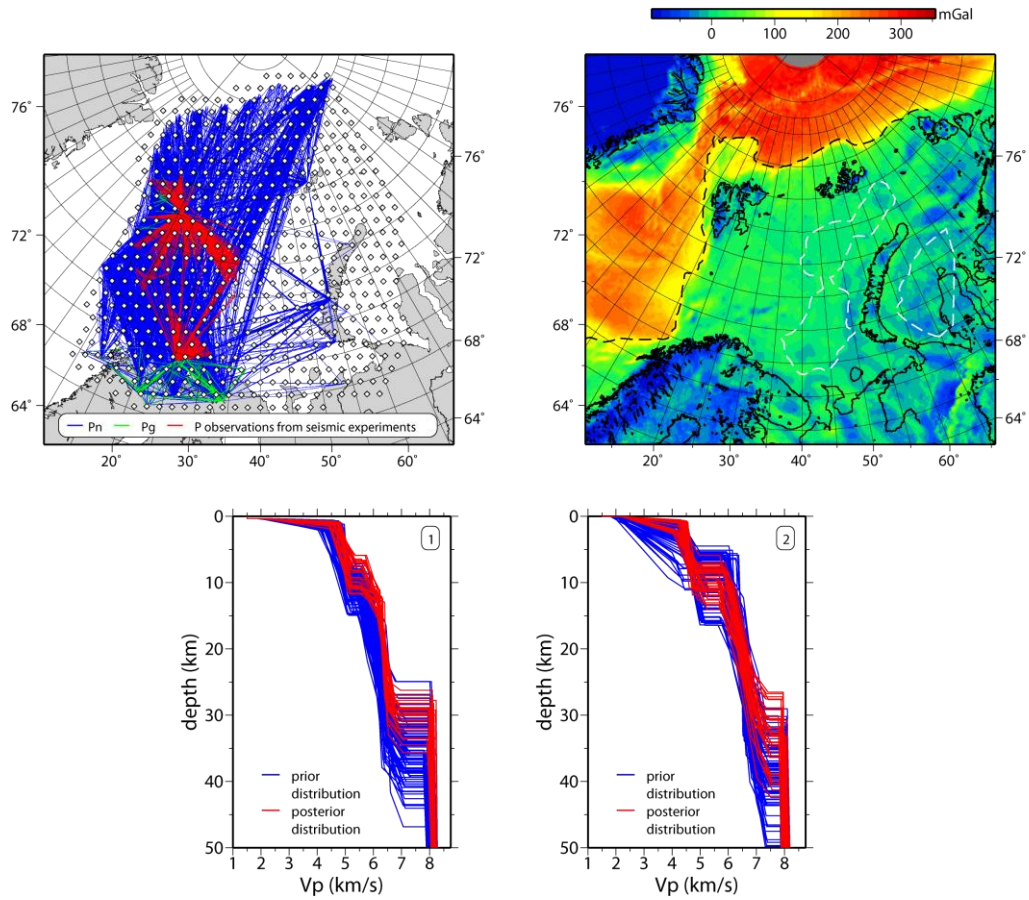

Figure 8 Development of a stochastic model of the European Arctic. Top row: example of data used to constrain the model, P wave paths supplying travel times (left), gravity measurements (right). Each line in the figure at left represents a path along which a travel-time has been measured. Bottom row: P velocity profile realizations with depth at two geographic locations: at a location with good path coverage in the west (left) and poor path coverage in the east (right). Figure modified from Hauser et al. [30] and reprinted by permission of American Geophysical Union. Copyright 2011 American Geophysical Union.

Conclusion

Signal processing for nuclear test ban verification began in the 1960s and 1970s with great optimism for sensitive detection of nuclear tests with the construction of large-aperture arrays and the development of adaptive beamforming techniques. This hope was partially frustrated by the poor signal coherence exhibited across such apertures, and led to the closing of LASA, significant reduction of the NORSAR array and the deployment of large numbers of small-aperture, regional arrays. Now, forty years on, there is interest in reversing this trend through techniques of signal calibration. Empirical methods are the surest approach, but are limited to regions with prior observations of relevant seismicity. Model-based methods work at the very low-frequency end of the spectrum. Where station density is high, even temporarily, there is the possibility that three-dimensional models can be constructed that will expand the

bandwidth over which waveforms can be predicted and coherent processing attempted. Such efforts are underway with adjoint and ambient noise tomography, for example, in the Middle East, though with networks much less dense than those available in North America.

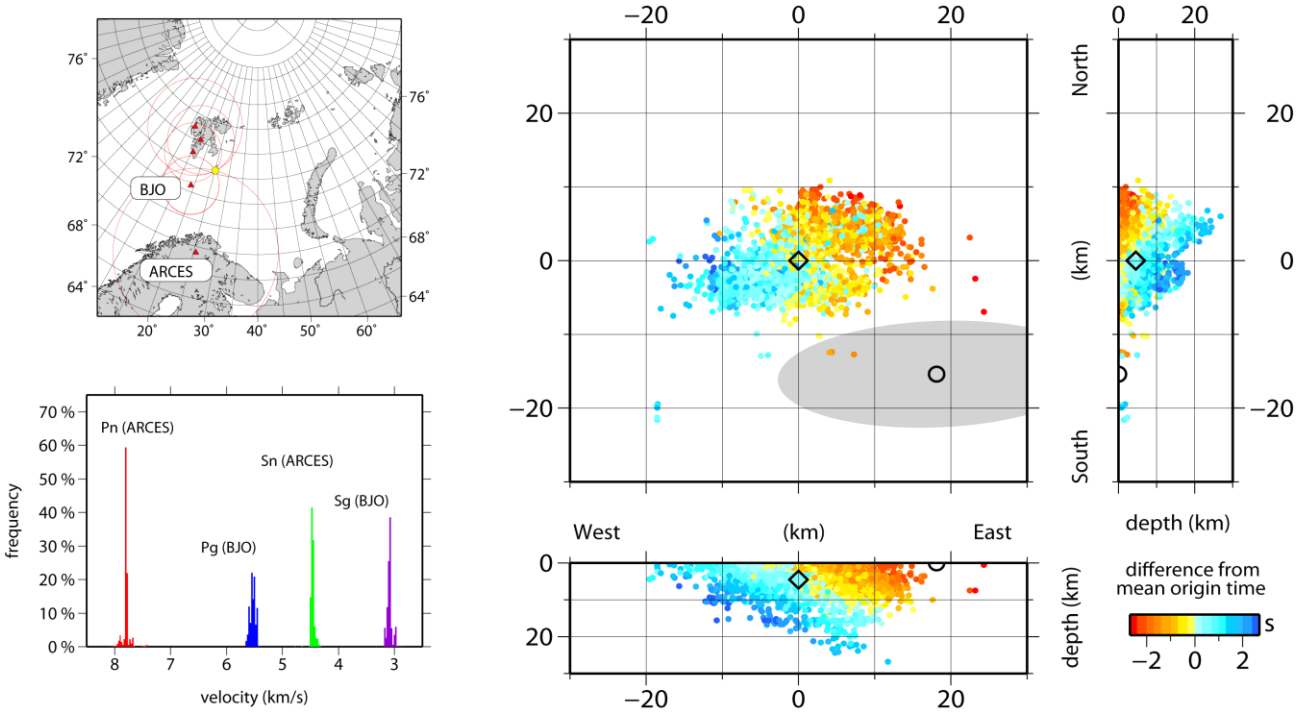

Figure 9 Results of a locating an earthquake southeast of Spitsbergen with the stochastic model. The network geometry is shown at upper left. Three stations (shown as red triangles) are located on Spitsbergen, one (BJO) is on Bjornoye Island and one (ARCES) is on the Norwegian mainland. The distributions of average velocities along the paths to two stations are shown at lower left. To the right are plots of estimates of the earthquake location using the individual models (colored dots) of the posterior distribution and (solid black circle) a location using a 1-D model with a fixed depth of zero. The grey ellipse is a theoretical error ellipse for the 1-D location. Figure reproduced from Hauser et al. [30] and reprinted by permission of American Geophysical Union. Copyright 2011 American Geophysical Union.

The existence of large and well-organized archives of event observations, inexpensive, fast computing resources and studies suggesting that a large fraction of seismicity is repetitive encourages the idea that waveform correlation techniques will be applied on a massive scale, using the past history of observation to screen events in real time. This strategy will put a premium on constructing and updating an empirical map of signal space for the network of monitoring stations, and removing the very large amount of redundancy that exists in past observations. As we have seen, it also will be necessary to extend the map in real time to keep pace with the flood of aftershocks that develop after large events like the recent earthquakes in Japan and Sumatra.

The processing pipelines of the future may include banks of tens or hundreds of thousands of correlation detectors, combing the combined streams of continuous data from hundreds of stations to remove recognizable events, and possibly to flag events of high interest. It also is possible that correlation or subspace detectors will be extended to operate across networks of arrays instead of single arrays to solve association problems. However, processing pipelines will continue to have general power detectors operating on array beams to capture new events for which there are no archived waveform templates.

The prospects of model-based signal processing are improving with innovations in tomography. Already it appears that discriminants based upon low-frequency surface waves (mb-Ms and moment-tensor methods) will be improved with 3-D models. However, it is not likely, under foreseeable trends, that tomographic models will push coherent processing into the high-frequency regional monitoring band (0.5-8 Hz). Perhaps the most likely technique for extending the model-based processing into that band is an approach for semi-coherent signal processing wherein the base tomographic model is augmented with a description of diffuse heterogeneity. In this approach, small scale-length medium perturbations not modeled in the tomographic inversion might be described as random fields, characterized by particular distribution functions (e.g. normal with specified spatial covariance). Conceivably, random field parameters (scatterer density or scale length) might themselves be the targets of tomographic inversions of the scattered wave field. Such augmented models may provide processing gain through the use of probabilistic signal subspaces rather than deterministic waveforms.

References

[1] Capon, J., J. Greenfield, and R. J. Kolker (1967), Multidimensional maximum-likelihood processing of a large aperture seismic array, *Proc. IEEE* **55**, 192-211.

[2] Gibbons, S. J. and F. Ringdal (2006), The detection of low magnitude seismic events using array-based waveform correlation, *Geophys. J. Int*, **165**, 149-166.

[3] Gibbons, S. J., , M. B. Sorensen, D. B. Harris and F. Ringdal (2007) The detection and location of low magnitude earthquakes in northern Norway using multi-channel waveform correlation at regional distances, *Phys. Earth and Planet. Int.*, **160**(3-4), 285-309.

[4] Schaff, D. P. and F. Waldhauser (2010), One magnitude unit reduction in detection threshold by cross correlation applied to Parkfield (California) and China seismicity, Bull. Seism. Soc. Am., **100**(6), 3224-3238.

[5] Scharf, L. L. and B. Friedlander (1994), Matched subspace detectors, *IEEE Trans. On Sig. Proc.***42**(8).

[6] Harris, D. B. (2006), Subspace detectors: theory, Lawrence Livermore National Laboratory report UCRL-TR-222758, 48 pp.

[7] Harris, D. B. and D. A. Dodge (2011), An autonomous system for grouping events in a developing aftershock sequence, *Bull. Seism. Soc. Am.*, **101**(2), 763-774

[8] Richards, P. F. Waldhauser, D. Schaff, W-Y. Kim, The applicability of modern methods of earthquake location, *Pure and Applied Geophysics*, **163**(2), 351-372.

[9] Schaff, D. P. and P. G. Richards (2004) Repeating seismic events in China, *Science* **303**, 1176-1178.

[10] Dziewonski, A. M., D. L. Anderson (1981), Preliminary reference Earth model, *Physics of the Earth and Planetary Interiors* **25**, 297–356.

[11] Kennett, B.L.N., E. R. Engdahl, R. Buland (1995), Constraints on seismic velocities in the Earth from travel times, *Geophys. J. Int*, **122**, 108-124.

[12] Ford, S. R., D. S. Dreger and W. R. Walter, (2009), Identifying isotropic events using a regional moment tensor inversion, *Jour. Geophys. Res.*, **114**, *B01306.*

[13] Zucca, J., W. Walter, A. Rodgers, P. Richards, M. Pasyanos, S. Myers, T. Lay, D. Harris and T. Antoun (2009), The prospect of using three-dimensional earth models to improve nuclear explosion monitoring and ground-motion hazard assessment, *Seis. Res. Lett.*, **80**(1), 31-39.

[14] Herrin, E. E. and T. T. Goforth (1977), Phase-matched filters: application to the study of Rayleigh waves, *Bull. Seism. Soc. Am.* **67**, 1259-1275.

[15] Shapiro, N. M., M. Campillo, L. Stehly, M. H. Ritzwoller, (2005), High-resolution surface-wave tomography from ambient seismic noise, *Science*, **307**, 1615-1618.

[16] Moschetti, M. P., M. H. Ritzwoller and N. M. Schapiro (2007) Surface wave tomography of the western United States from ambient seismic noise: Rayleigh wave group velocity maps, *Geochem. Geophys. And Geosystems*, **8**(8), Q08010.

[17] Shapiro, N. M. and M. Campillo (2004), Emergence of broadband Rayleigh waves from correlations of the ambient seismic noise, *Geophys. Res. Lett.*,**31**, L07614*.*

[18] Sabra, K. G., P. Gerstoft, P. Roux, W. Kuperman, and M. C. Fehler (2005), Extracting time-domain Greens function estimates from ambient seismic noise,*Geophys. Res. Lett.*,**32**, L03310.

[19] Lin, F. C., M. P. Moschetti and M. H. Ritzwoller (2008), Surface wave tomography of the western United States from ambient seismic noise: Rayleigh and Love wave phase velocity maps, *Geophys. J. Int*., *doi: 10.1111/j.1365-246X.2008.03720.x*.

[20] Roux, P., K. G. Sabra, P. Gerstoft, W. A. Kuperman, and M. C. Fehler (2005), P-waves from cross-correlation of seismic noise, *Geophys. Res. Lett.*,**32,** L19303.

[21] Tarantola, A., (1984), Inversion of seismic reflection data in the acoustic approximation, *Geophysics*, **49**(8), pp. 1259-1266.

[22] Tromp, J., D. Komatitsch and Q. Liu (2008), Spectral-element and adjoint methods in seismology, *Commun.Comput. Phys.*,**3**(1), 1-32.

[23] Tromp, J., C. Tape and Q. Liu (2005), Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels, *Geophys, J. Int*., **160,** 195-216.

[24] Fitchner, A., H.-P. Bunge, H. Igel (2006), The adjoint method in seismology I. Theory, *Phys. Earth Planet. Int.*, **157**, 86-104.

[25] Tape, C., Q. Liu, A. Maggi and J. Tromp (2010), Seismic tomography of the southern California crust based on spectral-element and adjoint methods, *Geophys, J. Int*., **180,** 433-462. doi: 10.1111/j.1365-246X.2009.04429.x

[26] Tape, C., Q. Liu, A. Maggi, J. Tromp (2009), Adjoint tomography of the southern California crust, *Science*, 325, 988-992.

[27] Metropolis, N., A. Rosenbluth, M. Rosenbluth, A. Teller and E. Teller (1953), Equation of state calculations by fast computing machines, J. Chem. Phys., **21**(6), 1087-1092.

[28] Mosegaard, K., and A. Tarantola (1995), Monte Carlo sampling of solutions to inverse problems, J. Geophys. Res, **100**(B7), 12,431 - 12,447.

[29] Pasyanos, M. E., G. A. Franz and A. L. Ramirez (2006), Reconciling a geophysical model to data using a Markov chain Monte Carlo algorithm: An application to the Yellow Sea-Korean Peninsula region, *J. Geophys. Res.,***111**, B03313.

[30] Hauser, J., K. M. Dyer, M. E. Pasyanos, H. Bungum, J. I. Faleide, S. A. Clark, and J. Schweitzer (2011), A probabilistic seismic model for the European Arctic, *J. Geophys. Res.*, **116,** B01303.

Prepared by LLNL under Contract DE-AC52-07NA27344.