Ryberg, T., Hole, J. A., Fuis, G. S., Rymer, M. J., Bleibinhaus, F., Stromeyer, D., Bauer, K. (2012): Tomographic Vp and Vs structure of the California Central Coast Ranges, in the vicinity of SAFOD, from controlled-source seismic data. - Geophysical Journal International, 190, 3, pp. 1341—1360.

DOI: http://doi.org/10.1111/j.1365-246X.2012.05585.x

Geophysical Journal International

Geophys. J. Int. (2012) **190,** 1341–1360 doi: 10.1111/j.1365-246X.2012.05585.x

Tomographic V_p and V_s structure of the California Central Coast Ranges, in the vicinity of SAFOD, from controlled-source seismic data

T. Ryberg,[1] J. A. Hole,[2] G. S. Fuis,[3] M. J. Rymer,[3] F. Bleibinhaus,[4] D. Stromeyer[1] and K. Bauer[1]

[1] *Helmholz-Zentrum Potsdam, Deutsches GeoForschungsZentrum GFZ, Telegrafenberg,* 14473 *Potsdam, Germany. E-mail: trond@gfz-potsdam.de*
[2] *Department of Geosciences, Virginia Tech,* 4044 *Derring Hall, Blacksburg, VA* 24061*, USA*
[3] *U.S. Geological Survey,* 345 *Middlefield Road MS* 977*, Menlo Park, CA* 94025-3591*, USA*
[4] *Institute of Geosciences, University of Jena, Burgweg* 11*, D-*07749 *Jena, Germany*

Accepted 2012 June 20. Received 2012 April 26; in original form 2011 December 23

SUMMARY

A seismic reflection/refraction survey across the San Andreas fault near Parkfield, California, has refined our knowledge of the upper crustal structure of the central California Coast Ranges at the San Andreas Fault Observatory at Depth (SAFOD). The survey consisted of a 46-km-long line of seismographs (25–50 m spacing) and 63 explosions (25–200 kg; nominal spacing of 500 m, with some gaps). The traveltimes of refracted *P* and *S* waves from the explosions constitute independent data sets of relatively high quality that were inverted to produce *P*- and *S*-wave velocity models (V_p, V_s) along the profile, extending to as much as 5 km depth. The V_p and V_s models show a prominent lateral drop in velocities a few hundred metres northeast of SAFOD, between the drill hole and the San Andreas fault. The V_p model shows particularly well a southwest-dipping velocity inversion beneath SAFOD, the top of which correlates with a fault penetrated by the drill hole that separates granitic rocks above from sedimentary rocks below. In addition to V_p and V_s models, a V_p/V_s model was derived. A V_p/V_s ratio lower than 1.73 is seen only at depth, in a narrow zone beginning at the target earthquakes for SAFOD and extending downward and northeastward into the North America Plate. Clusters in the parameter space spanned by V_p/V_s ratios and V_p can be identified by two different methods, one more intuitive analytical method and one more abstract method based on neural network techniques. These clusters are correlated to different rock types, based on laboratory and *in situ* data. These clusters are remapped back into *x–z* plane along the profile. Prominent features mapped this way include Salinian granitic rocks beneath and west of SAFOD, and a body of sedimentary rocks faulted beneath these granitic rocks along what we and others interpret to be a branch of the Buzzard Canyon Fault (BCF) system. These sedimentary rocks extend from this fault to the San Andreas fault system. Unfortunately, our cluster analysis shows no significant discontinuity at the San Andreas fault, owing presumably to the fact that the San Andreas fault is located within sedimentary rocks having similar elastic properties. This paper is an attempt to 'downward' continue a geological map by geophysical means based on elastic properties of rock samples from the region.

Key words: Controlled source seismology; Body waves; Seismic tomography; Transform faults; Continental margins: transform; North America.

1 INTRODUCTION

The San Andreas Fault (SAF) near Parkfield, California, a major plate boundary strike-slip fault system, has been intensively studied for decades (Bakun *et al.* 2006). Within the framework of the EarthScope project (http://www.earthscope.org), the SAF Observatory at Depth (SAFOD) is a deep borehole observatory designed to sample SAF zone material, investigate fault zone properties and monitor seismic and creeping activities directly at depth. This location has been chosen for scientific drilling since repeating shallow earthquakes occur frequently, thus providing a target location to study and understand the physics of earthquakes.

In the past, several geophysical investigations have been carried out in the vicinity of the SAFOD location to characterize the structure and composition of the area surrounding the drill location. Detailed structural images also helped to guide the drilling operations. The results of preliminary site characterization studies are summarized in Hickman *et al.* (2004). Most relevant to our studies are the investigation of the V_p (compressional wave velocity) and V_s (shear wave velocity) models derived from local earthquakes (Thurber *et al.* 2003, 2004a,b; Roecker *et al.* 2005), the strong electrical conductivity anomaly imaged by magneto-telluric investigations (Unsworth & Bedrosian 2004; Zhang *et al.* 2009) and the shallow seismic studies which had shown a steeply dipping (∼83°) SAF within the uppermost kilometre (Hole *et al.* 2001).

Shallow subsurface structure in conjunction with the surface geology is important for understanding tectonic processes. Prior study of the V_p structure at SAFOD with the data set presented here was carried out by Hole *et al.* (2006) using traveltime tomography and by Bleibinhaus *et al.* (2005, 2007) using waveform tomography. Zhang *et al.* (2009) used a combined inversion of controlled and earthquake sources to do double-difference seismic tomography for V_p and V_p/V_s. They also used electrical resistivity models. While Zhang *et al.* (2009) investigates the 3-D velocity and resistivity structures around SAFOD, our study focusses on the analysis of high-resolution, 2-D velocity models across SAF. In our study, we invert traveltimes of refracted *P* and *S* phases, and create high-resolution images of the shallow V_p and V_s structures down to a depth of as much as 5 km. A high-resolution image of the V_p/V_s ratio is derived by dividing the V_p and V_s models, as it is commonly done in controlled-source seismics. Then both, V_p and V_p/V_s ratio models are analysed by cluster analysis techniques similar to those of Bauer *et al.* (2003), to produce a non-unique geological model of the region near SAFOD. The results of the inversion and cluster analysis techniques are compared to results of Hole *et al.* (2006), Thurber *et al.* (2004b) and Zhang *et al.* (2009).

2 GEOLOGICAL AND TECTONIC SETTING

The region of our seismic profile (Fig. 1) consists of three main lithologic blocks or terranes; from east to west these are the

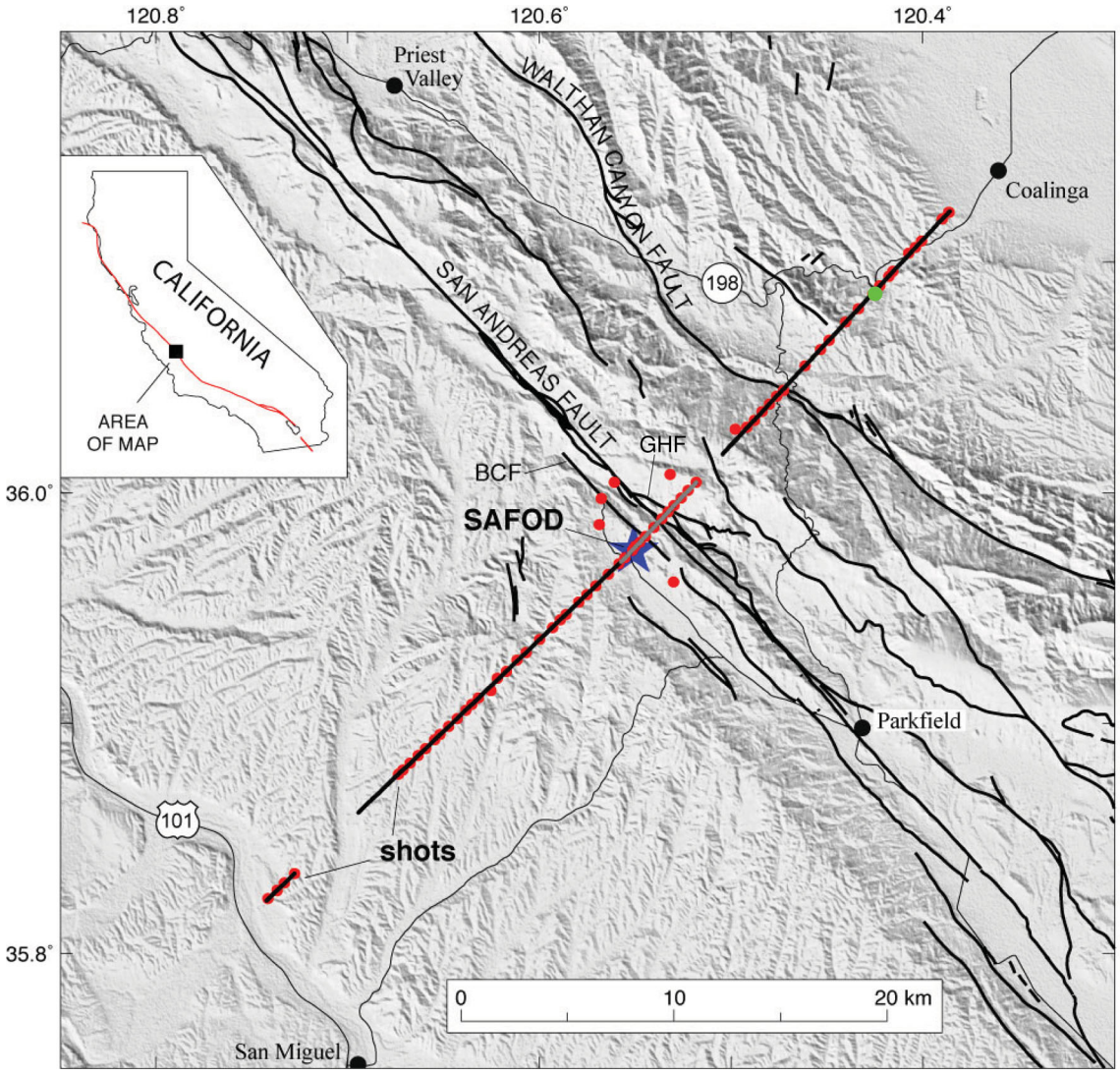

Figure 1. Shaded relief map of study area showing the location of the SAFOD drill site (blue star), faults (black lines) and the seismic profile (heavy black line) and shotpoints (red dots). Faults from Dickinson (1966), Dibblee Jr. (1971), Rymer (1981), Sims (1988), Sims (1990), Rymer *et al.* (2003), Rymer *et al.* (2006), Thayer (2006). BCF, Buzzard Canyon fault; GHF, Gold Hill fault. Green-filled circle is the location of SP 263 (see Fig. 3).

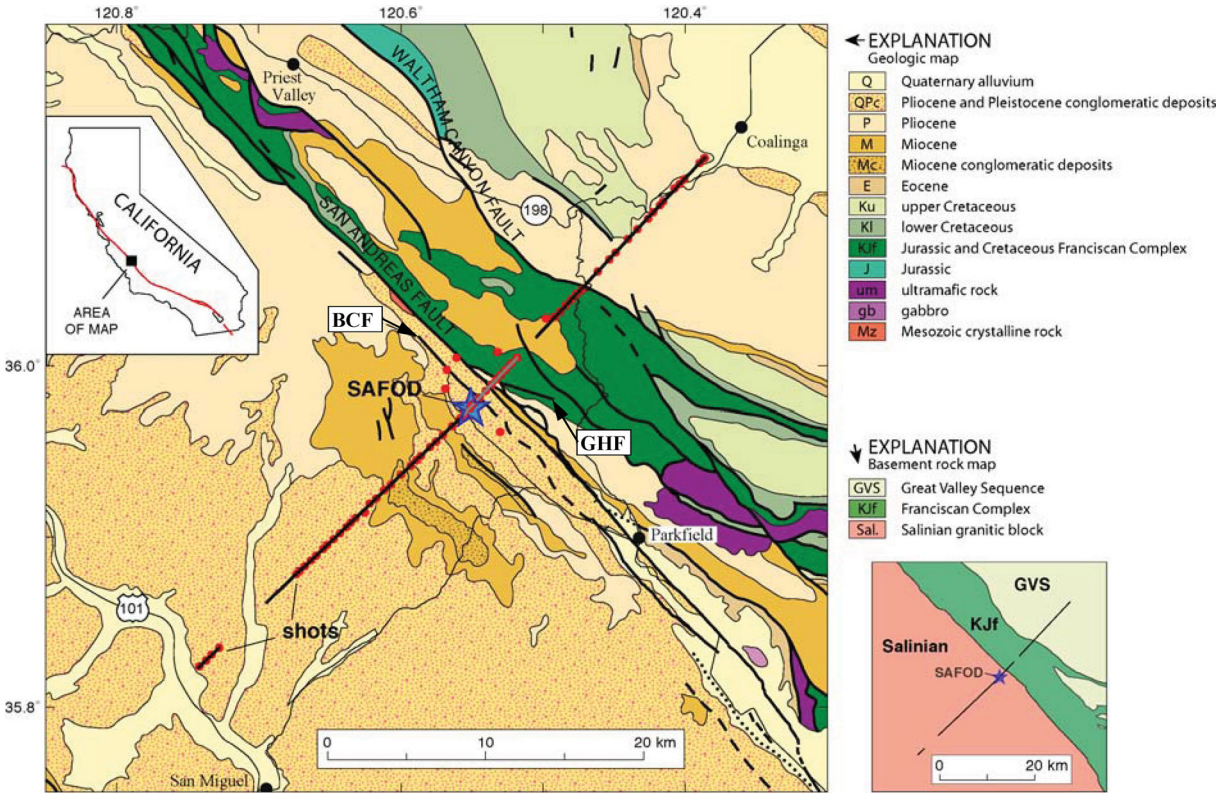

Figure 2. Geological map of study region (modified from Dibblee Jr. 1971, 1980; Bradbury *et al.* 2007). The seismic line (black line) and shot locations (red circles) are shown. Buzzard Canyon fault (BCF), Gold Hill fault (GHF), SAFOD and SAF are indicated.

Great Valley sequence, the Franciscan Complex and the Salinian block (Fig. 2). These assemblages are all chiefly Late Mesozoic rocks. Separating these assemblages are major faults: between the Great Valley sequence and the Franciscan Complex is the Coast Range Fault (locally expressed as the Waltham Canyon fault) and between Franciscan rocks and the Salinian block is the SAF.

The Great Valley sequence consists of stratified silicic-clastic sediment derived from the Sierra Nevada magmatic arc. In the region of the seismic line (Figs 1 and 2), most sediments were deposited in deep-sea fans and consists of sandstone, siltstone, shale and less abundant conglomerate. Unconformably overlying the Great Valley sequence are marine and non-marine sandstones, siltstones and mudstones of Eocene to Pliocene age.

The Franciscan Complex consists of a mixture of strata derived from deep-sea and terrigenous sediment, oceanic basalt and other parts of oceanic crust. Rock types include chert, greenstone, greywacke, blueschist and conglomerate. These rock types are present as coherent blocks or as a melange, where local exposures include a pervasively sheared argillaceous matrix. The Franciscan Complex is locally overlain by Tertiary and Quaternary sedimentary strata, including sandstones, siltstones, mudstones and conglomerates.

The Salinian block represents a Mesozoic magmatic arc that has been displaced hundreds of kilometres from its place of origin (Page 1981). Included within the Salinian block are metasedimentary rocks (schist of Sierra de Salinas) that were once thought to predate plutons of the arc, but now appear to be younger. Their location below the plutons may be a result of tectonic wedging in the latest Cretaceous and Palaeocene (Barth *et al.* 2003; Grove *et al.* 2003).

The SAF, a known Holocene fault, is the main plate-boundary fault in the central Coast Ranges and, since its development on land about 28 Ma, has right-laterally displaced rocks about 315–320 km (Mathews 1976; Ross 1984; Irwin 1990). Geological mapping of the SAF indicates a complex zone of faulting that is 5–8 km wide (Dickinson 1966; Dibblee Jr. 1980; Rymer 1981; Sims 1988, 1990; Rymer *et al.* 2003; Thayer 2006). Near the SAFOD drill site, faults within the SAF zone have juxtaposed tectonic slivers of contrasting rock types by tens to hundreds of kilometres from their original positions (Rymer *et al.* 2003). High-resolution reflection and refraction profiling across the SAFOD site indicate the fault zone is composed of at least two, but probably three, upwardly flaring flower structures, with the more recent, youngest development below the current SAF (Catchings *et al.* 2002; Rymer *et al.* 2003).

Two of the faults of structural importance in our seismic profile are the Gold Hill fault (GHF) and the Buzzard Canyon fault (BCF) (Figs 1 and 2). The BCF, part of a flower structure, dips moderately west and the seismic imaging of Catchings *et al.* (2002) and Rymer *et al.* (2003), and this study indicate that a subsurface branch of this fault may be the observed contact in the SAFOD drill hole at ∼1.8 km depth between Salinian granitic rock and sedimentary rocks. The Gold Hill Fault (GHF) is the east boundary of a block of marble and schist that dips steeply to moderately west toward the SAF and was imaged by Ryberg *et al.* (2005).

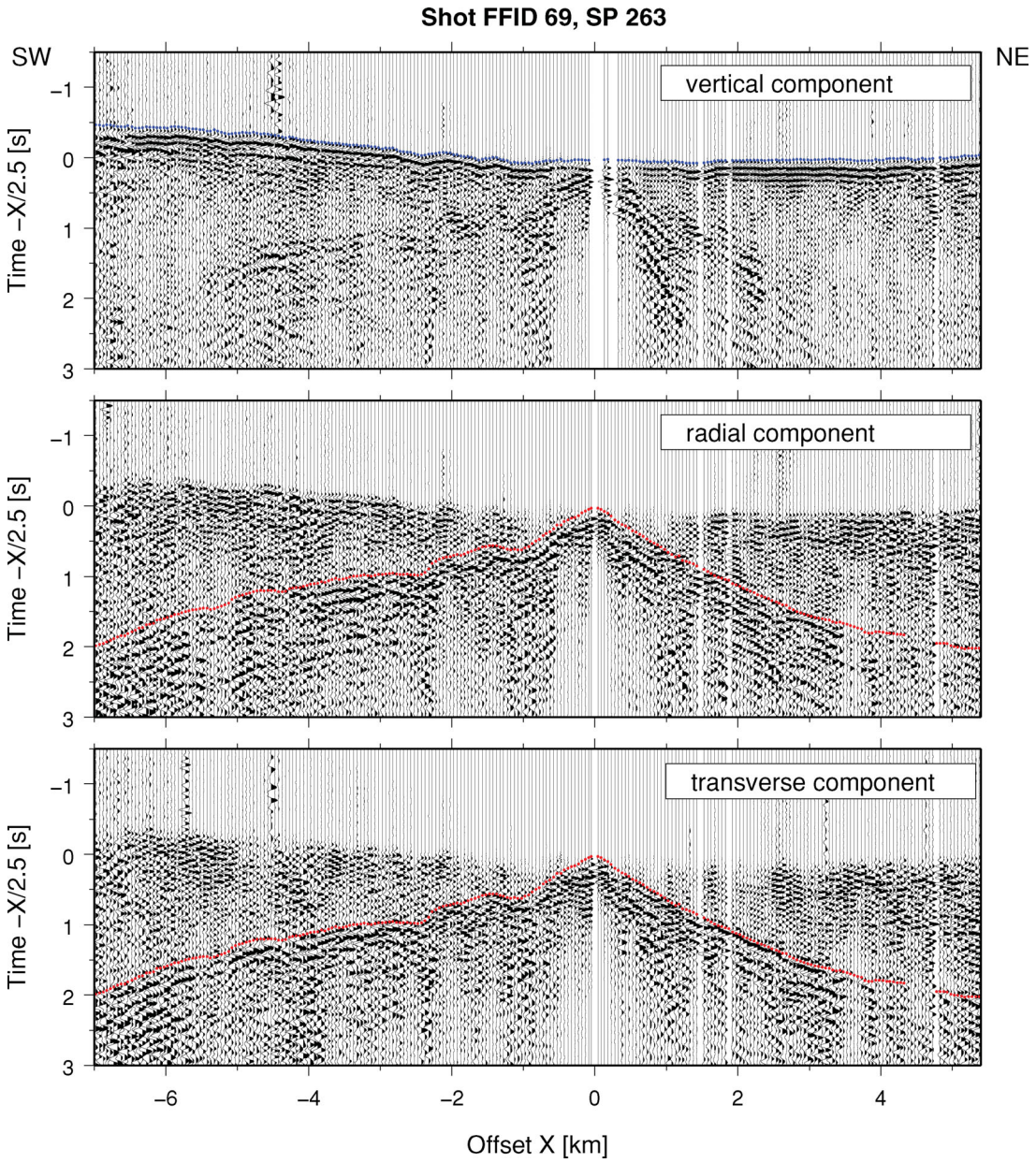

Figure 3. An example of an average-quality seismic record section, from shot SP 263 (northeastern part of the line, indicated as green-filled circle in Fig. 1). Data are trace-normalized and time-reduced. Blue and red dots, traveltime picks of the *P* and *S* waves, respectively; the *S*-wave picks are from the transverse component. Picking of *S*-wave traveltimes in both horizontal components resulted in very similar values. Only short-offset part of record section is shown for the vertical, radial and transverse component of the wavefield.

3 DATA ACQUISITION

In autumn 2003, seismic data were acquired along a 46-km-long line (Figs 1 and 2) that extended through SAFOD and crossed the SAF perpendicularly. Along this line, 912 three-component seismic recorders were deployed at 50 m spacing. In the central part (5 km) of the line near SAFOD the station spacing was 25 m. The sensors were oriented in a way that the vertical, radial and transverse components of the wavefield could be recorded. The sampling rate was 2 ms and the trace length 30 s. Although the recorders were deployed along a straight line (with a slight bend at SAFOD), owing to significant topography, large static shifts were observed and these had to be taken into account during reflection data processing and interpretation. Sensors could not be deployed along two short segments of the line, resulting in two gaps of several kilometres. 63 explosions were shot in boreholes along the line. A combination of 100 kg shots (nominal spacing 1 km), 25 kg shots (spacing 500 m) in the central part of the line and 200 kg shots at both line ends were used. Generally, the signal-to-noise ratios were favourable, and refracted arrivals (direct phases) could be observed for the 100 kg shots to a typical distance of >20 km. The 200 kg shots produced visible refractions to >40 km. Fig. 3 shows an example of the vertical, radial and transverse component of shot SP 263 near the NE end of the line (see Fig. 1 for location).

First-arrival traveltimes were picked manually for refracted *P* and *S* waves and inverted independently for 2-D velocity models. Altogether, 45 630 traveltime picks for *P* phases and 20 027 picks for *S* phases were made. Picking of refracted *S* waves was done

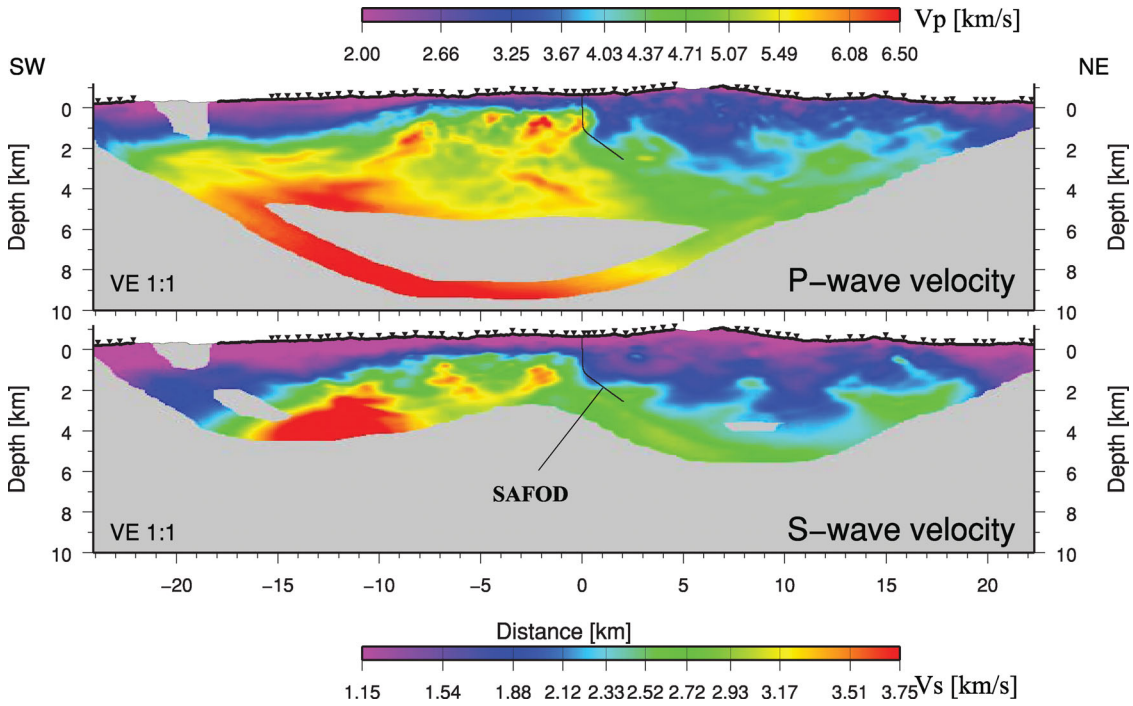

Figure 4. Complete tomographic models for V_p and V_s velocities. The shallow parts of the velocity models (down to ~2 km depth) have a resolution better than 0.5 km × 0.25 km (width/depth). The resolution quickly degrades with depth. Black triangles indicate the position of the shots. The V_s velocity colour table is scaled from the V_p velocity table by $\sqrt{3}$. Cells with less than 100 rays passing through are shown as grey. The black line shows the SAFOD drill hole.

using the transverse component recording, although *S*-wave energy was equally well observed on the radial and transverse component recordings. Picking of *S*-wave traveltimes on both horizontal components resulted in very similar values. Picking *S*-wave traveltimes at short offset was sometimes difficult due to the presence of surface waves and/or ground roll. Generally, the picking accuracy was high, for *P*-wave arrivals <0.02 s and for *S*-wave arrivals ~0.04 s.

4 TOMOGRAPHIC INVERSION

The inversion code FAST by Zelt & Barton (1998) in its 2-D version was used to derive a velocity model based on these traveltime data. Shot and receiver locations were projected onto a straight line (Zelt 1999). The coordinate system was centered at SAFOD; elevations are negative above sea level.

Our goal was to achieve a high-resolution velocity image along the seismic line. The inversion code by Zelt & Barton (1998) allows one to invert for models composed of uniform rectangular blocks. Inversion models with small blocksizes producing maximum resolution have problems related to inversion stability, choice of starting model and the development of strong inversion artefacts in regions with poor ray coverage. To minimize these problems, the inversion procedure of Zelt & Barton (1998) was modified by adding a special iterative step (Ryberg *et al.* 2007). For both, the V_p and V_s models, we started the inversion using a simple starting model and large blocks (1600 m × 800 m). This inversion grid size is quite coarse, given the density of shots and sensors. After this inversion step, we used the inversion result as a starting model to invert for a more finely gridded model, that is, 800 m × 400 m blocksize. By decreasing the inversion block size down to 100 m × 50 m, we produced the final inversion model. The final V_p-model is characterized by an rms traveltime misfit of ~0.017 s (45 630 traveltime picks), which is significantly lower than the value of 0.04 s found by Hole *et al.* (2006) and compares to our estimated picking accuracy. This inversion procedure, with decreasing blocksize, makes the final inversion result stable with respect to the original starting model, that is to say, the final tomographic model does not depend on the choice of a reasonable starting model. Another advantage is that this approach significantly reduces the appearance of tomographic artefacts, such as streaks. This same inversion technique was applied to the set of 20 027 *S*-wave traveltime picks, and a stable V_s model was obtained. Given the poorer accuracy of the *S*-wave picks (phase onsets are hidden in the *P* coda, ground roll, multiples, etc.), the final rms misfit of 0.039 s is higher than for the V_p model, and the maximum depth resolution for the model is shallower.

As the inversion is non-unique, we have conducted extensive testing of model and inversion parameters. By varying the model parameters (starting block size and shape) and the inversion parameters (number of iterations, vertical and horizontal smoothing parameters, etc.), a small range of stable final V_p and V_s models could be found; our preferred models are shown in Fig. 4. The final inverted models consist of 465 × 225 blocks, measuring 100 m × 50 m in size, resulting in an model of 46.5 km in length and 10 km in depth; however, only blocks with ray coverage are inverted and shown. The depth penetration along the line ranges from 4 to 6 km (below surface) for the V_p model and 3 to 5 km for the V_s model. The highest ray density (>1000 per block) and the highest model resolution is achieved in a depth range down to about 1 km below the surface. Deeper parts of the model are less well resolved (see Figs 4–7). To further test the stability of our inversion result we performed inversion runs with randomly selected data subsets (picks) for the V_p model. Even with only 10–20 per cent of the picks, decimated at all offsets, the inversion resulted in a stable velocity model, differing in regions sampled by only few rays (deepest regions).

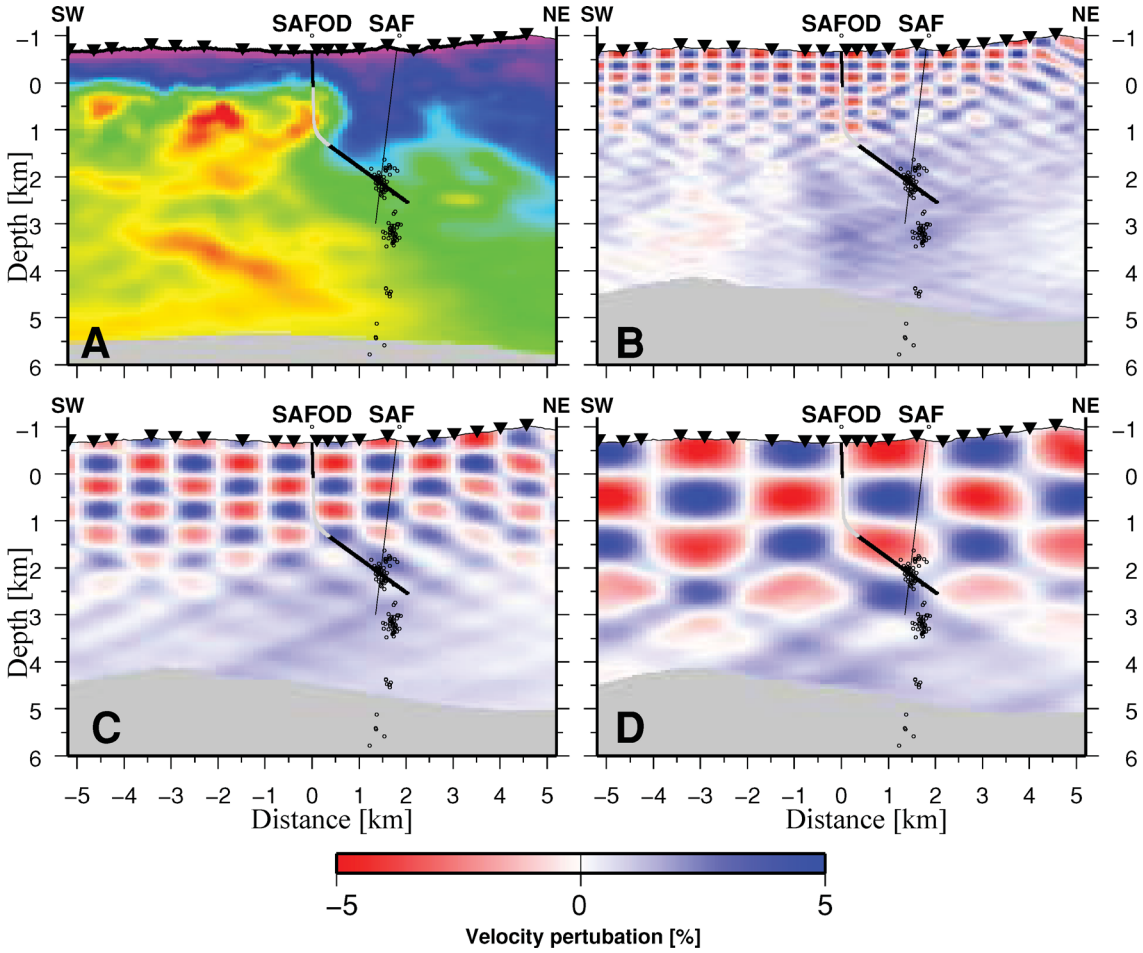

Figure 5. Checkerboard test for our V_p model (a). Models (b–d) show the inversion results of checkerboard anomalies measuring 500 × 250, 1000 × 500 and 2000 × 1000 m, respectively. A ±5 per cent velocity perturbation was superposed on a smoothed version of the final V_p model (Fig. 4). The checkerboard tests were carried out for the entire model, although we show only the region in the vicinity of SAFOD and the SAF. Black circles show locations of local earthquakes projected onto the line within 1 km horizontal distance (from Thurber *et al.* 2004a; Roecker *et al.* 2005). The black and grey line shows the SAFOD drill hole; the grey part indicates the drilled granite and the thin black line indicates the inferred SAF.

We performed several checkerboard tests (Fig. 5) to determine the depth resolution of our inversion. In these tests, rectangular velocity anomalies having amplitudes of ±5 per cent of the background model (a checkerboard pattern) were overlain on an average velocity model (strongly smoothed version of the final velocity model). Synthetic traveltimes were calculated for this model, and a 15 ms (rms) random time jitter was added to the data, similar to an estimated picking accuracy. The inversion method described above was then applied. In the result, features of several hundred metres in size can be easily resolved down to 1.5 km below surface (Fig. 5b). At 4 km below surface, only features larger than ∼1 km can be imaged properly (Figs 5d and 6d). Below the above depths of resolution, the rectangular velocity anomalies cannot be recovered in full and the shapes begin to distort. Slightly poorer resolution is seen for the corresponding V_s model since the number and quality of the *S*-wave traveltime picks is lower, shown in Fig. 6. Finally, we tried to recover checkerboard patterns for V_p/V_s ratios in Fig. 7. The recovery of these V_p/V_s ratio patterns is intermediate in appearance between the corresponding V_p and V_s checkerboards, as one might expect.

In addition to the checkerboard test, we compared our velocity model around SAFOD with V_p velocity models derived by other methods and/or traveltime data sets (Fig. 8). All models show significant similarities, including the low-velocity sedimentary cover, and the high-velocity block SW of SAFOD and low-velocity block NE. Partly due to the dense observations (surface shots and receivers) our inversion model (Fig. 8a) has a much higher resolution in the first few kilometres depth than the model of Thurber *et al.* (2004b) (Fig. 8b). Both show that along the SAFOD drill hole (grey line) low-velocity material is penetrated a second time, near the deviation of the borehole from the vertical. As expected, the inversion result of Hole *et al.* (2006) (Fig. 8c), which emphasizes minimum structure shows significantly less structure in the shallower part of the model. We also include a second inversion of our data set, using the method of Zelt & Barton (1998) with inversion parameters equivalent to high damping (Fig. 8d) to better compare them with the result of Hole *et al.* (2006) (Fig. 8c). The resolution tests of Fig. 5 indicate that the small details (<0.5 km in dimension) seen in the upper 1.5 km of our model (Fig. 8a) and not seen in Figs 8(c and d) are real. The same is true for larger features (<1 km in dimension) in the upper 2.5 km.

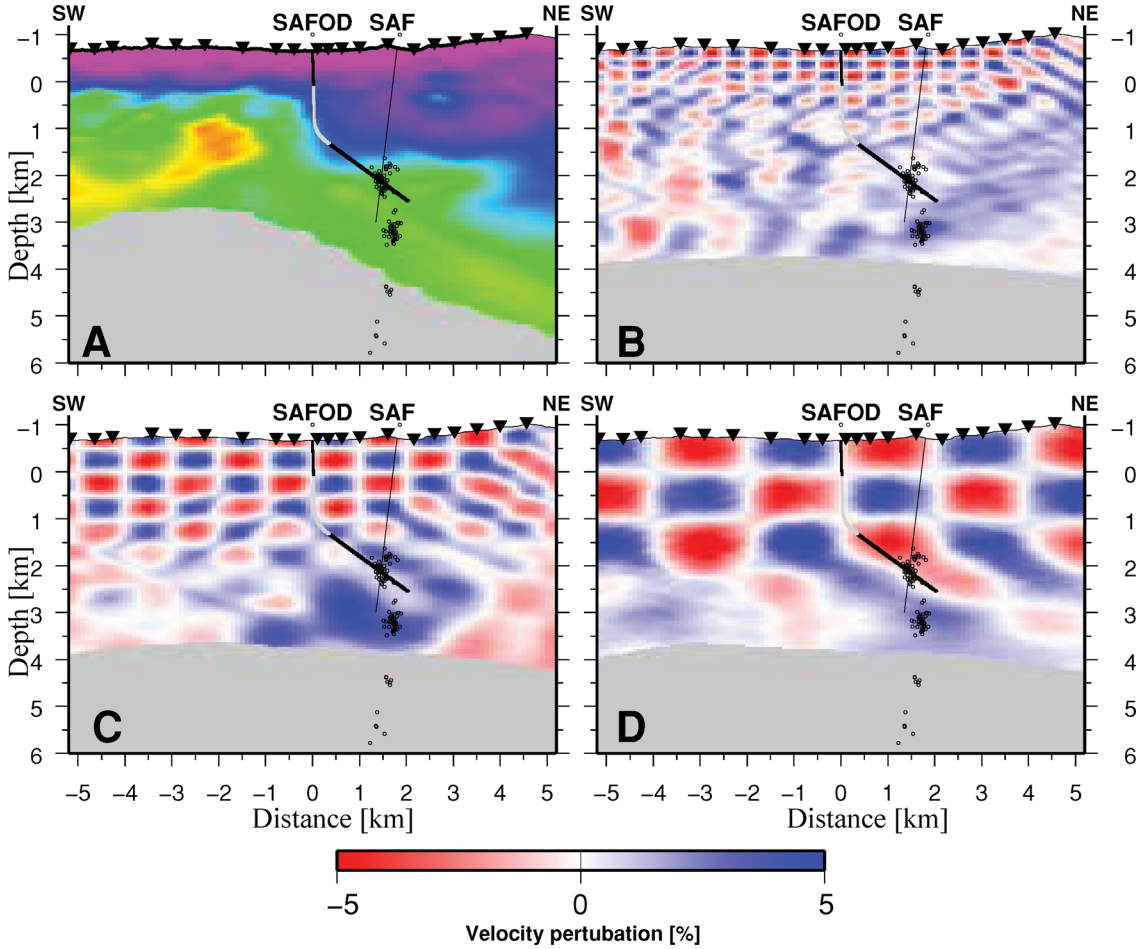

Figure 6. Checkerboard test for our V_s model (a), similar as Fig. 5. Models (b–d) show the inversion results of checkerboard anomalies measuring 500 m × 250 m, 1000 m × 500 m and 2000 m × 1000 m, respectively. A ±5 per cent velocity perturbation was superposed on a smoothed version of the final V_s model (Fig. 4). Note that the resolution of checkerboard details is similar but slightly poorer than the resolution for the V_p model due to the smaller number and different distribution of *S*-wave traveltime picks.

5 ANALYSIS OF THE V_p AND V_s MODELS

Near the SAF, our final V_p and V_s models both contain a shallow low-velocity layer that averages 0.6 km in thickness, extending from the surface to approximately sea level (0 km depth; Fig. 9). Below sea level, both models have relatively high velocities in the southwest and relatively low velocities in the northeast (northeast of SAFOD) in the depth range of 0–3 km. The V_p model shows a strong velocity inversion near a point at 1.3 km depth where the SAFOD drill hole bends. This feature is less pronounced for the *S*-wave model.

We calculated a V_p/V_s-ratio model by simply dividing the V_p and V_s models in all locations where both are well resolved (Fig. 9, bottom) as it is done in controlled-sources seismic investigations (e.g. Mechie *et al.* 2004; Raileanu *et al.* 2005; Mechie *et al.* 2005; Hauser *et al.* 2008; Murphy *et al.* 2010; Mechie *et al.* 2012). We chose not to invert directly for V_p/V_s, as is commonly done in earthquake tomography for the following reasons: (1) Our V_p and V_s models are independent of each other, unlike V_p and V_s models from earthquake tomography, which depend jointly on inverted earthquake locations and origin times. (2) Our V_p and V_s models are constructed from ray paths that are truly independent of one another, in that V_p/V_s can vary from point to point. (3) Our V_p and V_s models are of comparable quality, and checkerboard tests for V_p, V_s and V_p/V_s models (Figs 5, 6 and 7) demonstrate that the derived V_p/V_s model is as well resolved as the V_s model (as expected).

5.1 Comparison of results with downhole velocity logs

Downhole measurements of V_p and V_s in the SAFOD drill hole (Zoback *et al.* 2005) are compared to tomographic V_p and V_s values taken from our model along the path of the drill hole (Fig. 10). Tomographic V_p values (Fig. 10a, blue line) are generally higher than downhole values in the interval 0.5–1.0 km below sea level but generally lower in the interval 1.0–2.0 km below sea level; on average, they are ∼0.1 km s^{-1} lower in the total interval from 0.0 to 2.0 km depth. Above 2 km depth, tomographic V_s values (Fig. 10b, blue line) are consistently lower than the downhole V_s measurements by an average of 0.6 km s^{-1}. Because of strong lateral velocity variations in the tomographic V_p and V_s models, we decided to compare the downhole logs with a slightly shifted borehole track. Further details are given in Appendix A. Tomographic V_p values from 500 m southwest of the drill hole (Fig. 10a, green line) appear to agree a bit better with the downhole V_p values. 500 m southwest, V_s values (Fig. 10b, green line) agree a bit better with the downhole V_s values but are still consistently low. Agreement between downhole and tomographic V_p and V_s values improves markedly below 2 km depth. We note that V_p values obtained from explosion waveform

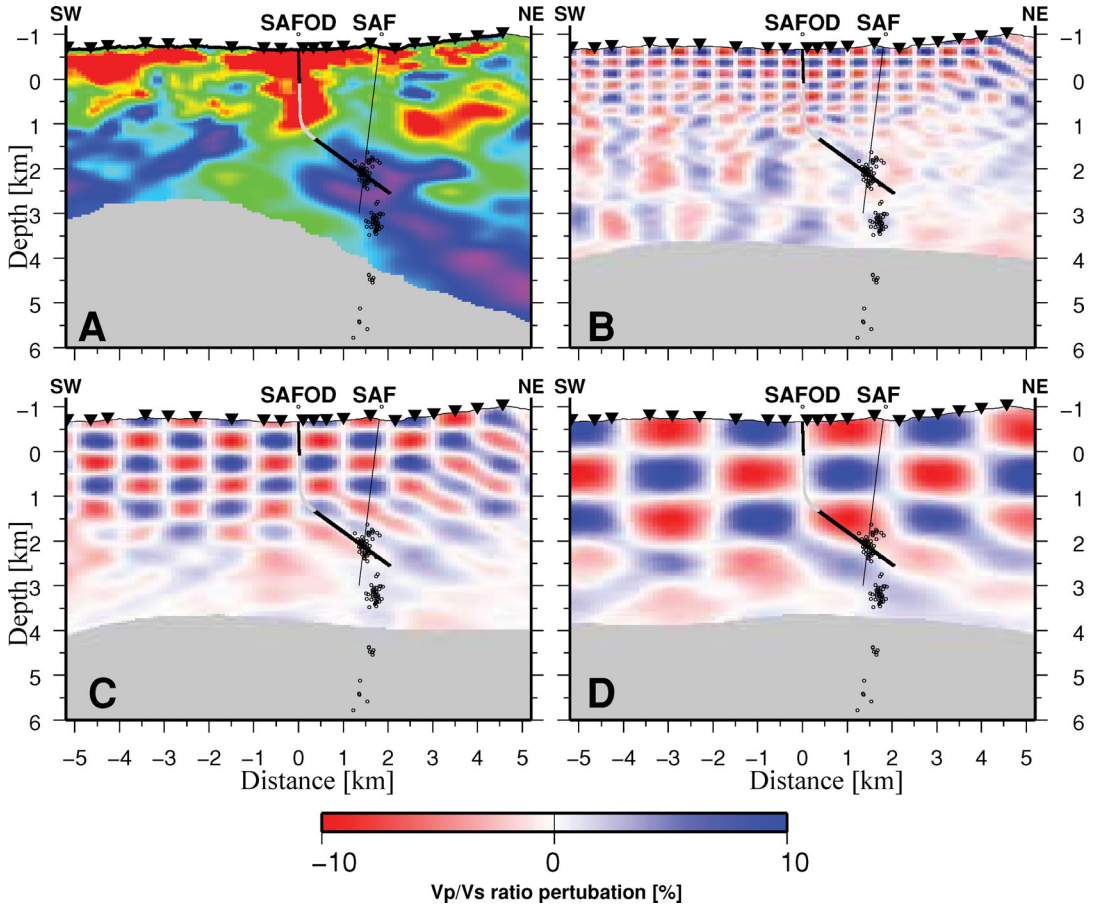

Figure 7. Checkerboard test for our V_p/V_s-ratio model (a), similar to Figs 5 and 6. Note that the same colour scale is used as in Fig. 9. Models (b–d) show the inversion results of checkerboard anomalies measuring 500 m × 250 m, 1000 m × 500 m and 2000 m × 1000 m, respectively. A ±10 per cent V_s/V_s-ratio perturbation was superposed on a smoothed version of the final V_p and V_s models (Fig. 4). Note that the resolution of checkerboard details is similar to the resolution for the V_s model.

tomography, as opposed to our explosion traveltime tomography, show a very similar comparison to downhole V_p values (see Bleibinhaus *et al.* 2007). A very similar discrepancy in V_s velocity is observed by Zhang *et al.* (2009, their Fig. 8b.) between V_s determined from traveltime tomography and the SAFOD downhole data.

The relatively low tomographic V_p and V_s values can arise from at least three sources, including (1) lateral smearing of velocity values across a steep velocity discontinuity, (2) anisotropy and (3) sampling size. In the vicinity of SAFOD, these effects may conspire to produce the low model values. Certainly, a steep velocity discontinuity is present, just northeast of SAFOD (see Fig. 9, top). (See Appendix A for further discussion.) Our lower tomographic velocities in the upper 2 km near SAFOD produce V_p/V_s ratios that are too high compared to ratios that can be calculated from the downhole logs (see Fig. 9, bottom). One feature in our V_p/V_s model is the onset of low V_p/V_s values where the drill hole encounters the SAF and hypocentres of small earthquakes. Thurber *et al.* (2003) and Zhang *et al.* (2009) found the hypocentres in a region of a transition of high to low V_p/V_s values. The moderate northeastward dip of this feature is, however, currently unexplained.

5.2 2-D histograms of V_p versus V_p/V_s

To derive rock types from V_p or V_s models alone is not straightforward. The velocity V_p or V_s values from an inversion cell do not uniquely identify a specific rock type; the same rock can have different velocity values which depend for instance on the weathering state, metamorphic grade, *in situ* pressure and temperature condition. The simultaneous interpretation of V_p or V_s values of a given inversion cell can have the potential of resolving different rock types. Bauer *et al.* (2003), Bedrosian *et al.* (2004), Haberland *et al.* (2003), Maercklin *et al.* (2004) and Zhang *et al.* (2009) describe a method of interpreting and mapping rock types in cross sections using V_p, V_s, V_p/V_s and electrical resistivity. This method first creates a 2-D histogram of V_p–V_p/V_s values, for example, from each x and z point of the tomographic velocity models. The histogram generally shows well-defined clusters. Next, the clusters are selected in various ways and assigned arbitrary colours. Finally, each point on the histogram is mapped back into x–z plane, with its assigned colour. This method is based purely on statistical correlations and does not depend on any empirical link between V_p and V_s velocities.

Using this method for the region around SAFOD (see Fig. 9), we created a histogram of 5674 inversion cells for which both V_p and V_s values were well determined tomographically (Fig. 11a). We decided for the V_p–V_p/V_s instead of V_p–V_s space following Holbrook *et al.* (1992), White *et al.* (1992), Musacchio *et al.* (1997) and Kern *et al.* (1999). The bin width for the V_p velocity was 0.1 km s^{-1}, the width for the Poisson's ratio was 0.025. This histogram does not, of course, take into account any error in the tomographic velocities. As discussed above, the tomographic models have a velocity resolution which depends strongly on the location of the inversion

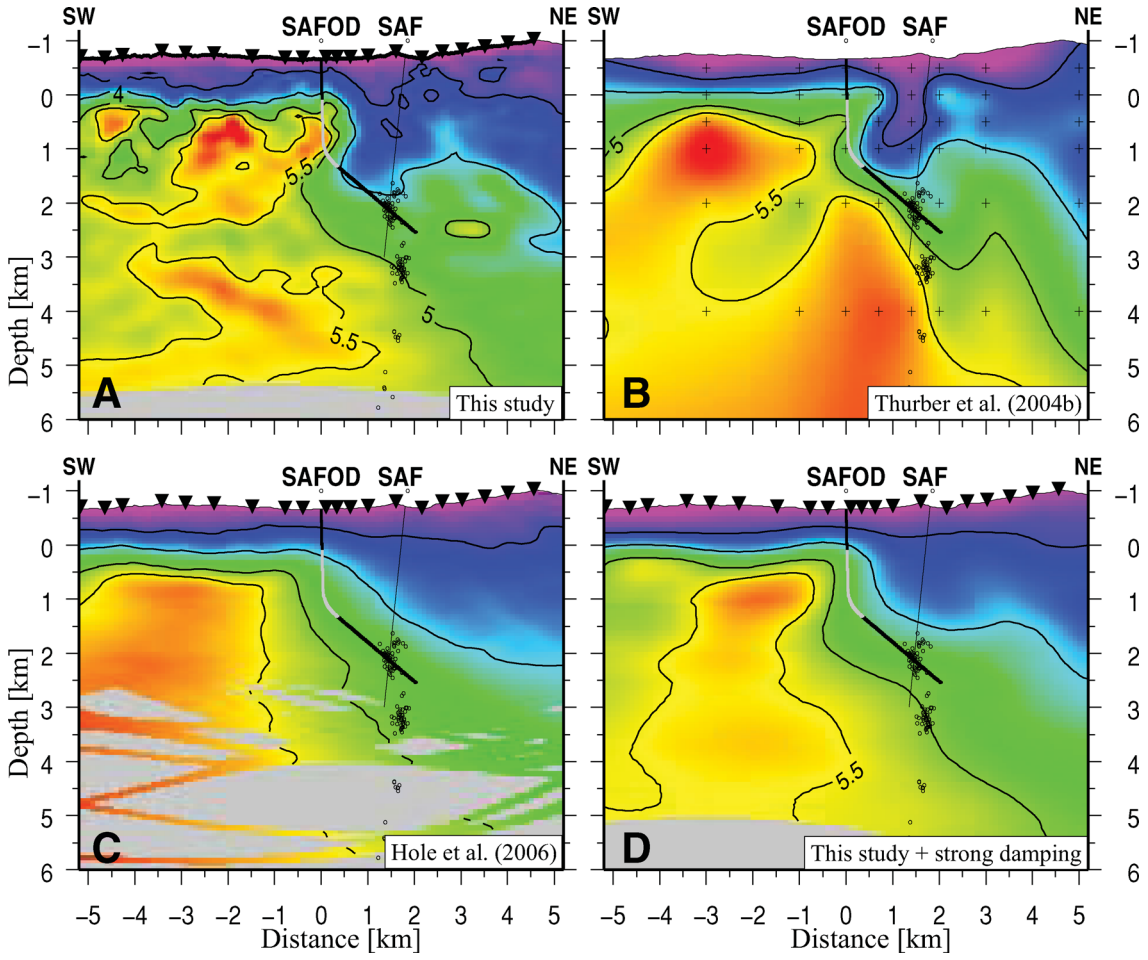

Figure 8. Comparison of V_p models, derived from various data sets and inversion methods. All models are centred at SAFOD, roughly 1.8 km SW of SAF, units are in kilometre. (a) Part of the V_p model of Fig. 4 in the vicinity of SAFOD. Only cells having more than 100 rays passing through are shown. (b) 2-D cross section of the 3-D inversion result of Thurber *et al.* (2004b) based on the joint inversion of traveltimes from local seismicity and calibration shots (crosses indicate the inversion nodes; the distribution of these nodes is very similar to that in the models of Zhang *et al.* 2009). (c) Inversion of the same traveltime picks as (a), using the inversion algorithm of Hole (1992) involving strong smoothing. (d) Inversion of the data set in (a) with FAST (Zelt & Barton 1998) with inversion parameters equivalent to high damping. The black and grey line shows the SAFOD drill hole: grey part indicates the drilled granite.

cell: deeper structures, regions close to the model edges and regions distant from a source can not be imaged with high resolution. In an attempt to reduce the effect of these velocity uncertainties in our analysis, we use a probability density function (PDF) approach (Bauer *et al.* 2003; Bedrosian *et al.* 2007). Eq. (1) defines the PDF of the V_p velocity and V_p/V_s ratio for the inversion cell (i, j) assuming a normal error distribution.

$$PDF_{ij}\left(\left[\frac{V_p}{V_s}\right], V_p\right) = \frac{1}{\sqrt{2\pi\,\delta\left[\frac{V_p}{V_s}\right]_{ij}\,\delta V_{p,ij}}}\exp -\frac{1}{2} \times \left[\frac{\left(\left[\frac{V_p}{V_s}\right]-\left[\frac{V_p}{V_s}\right]_{ij}\right)^2}{\delta\left[\frac{V_p}{V_s}\right]^2_{ij}} + \frac{(V_p - V_{p,ij})^2}{\delta V^2_{p,ij}}\right], \quad (1)$$

where $\left[\frac{V_p}{V_s}\right]_{ij}$ is the V_p/V_s-velocity ratio and $V_{p,ij}$ the V_p velocity in inversion cell (i, j), while the $\delta\left[\frac{V_p}{V_s}\right]_{ij}$ and $\delta V_{p,ij}$ represent their respective uncertainties. Unfortunately, there is no direct measurement of the uncertainties of the inverted velocities with the FAST code from Zelt & Barton (1998). Therefore, we use a proxy for the velocity uncertainty. The velocity uncertainty for a given inversion cell strongly depends, but not exclusively, on the number of rays passing through that cell. So, we constructed the relative velocity uncertainty as

$$\delta V_{ij} = C_V\left[\frac{\langle\log(n)\rangle}{\log(n_{ij})}\right], \quad (2)$$

where n_{ij} is the number of ray hit counts for cell (i, j), $\langle\log(n)\rangle$ denotes the average of the logarithmic hit count for the entire model and C_V is the mean velocity uncertainty. This downweights cells having a lower hit count. Fig. 12 shows the distribution of V_p and V_p/V_s uncertainties in the region of SAFOD. These velocity uncertainties were determined for every inversion cell for the V_p and V_s models, using an assumed mean velocity uncertainty C_V of 0.1 km s^{-1} for V_p and 0.2 km s^{-1} for V_s velocities. Under the assumption that all data points (V_p and V_s velocities) are independent, the joint PDF can be constructed by simply summing all individual PDF_{ij}.

Fig. 11(b) shows the joint PDF. Instead of the simple histogram count (Fig. 11a) it is much smoother, and the occurrence of specific clusters is much clearer. Two main visually identified clusters

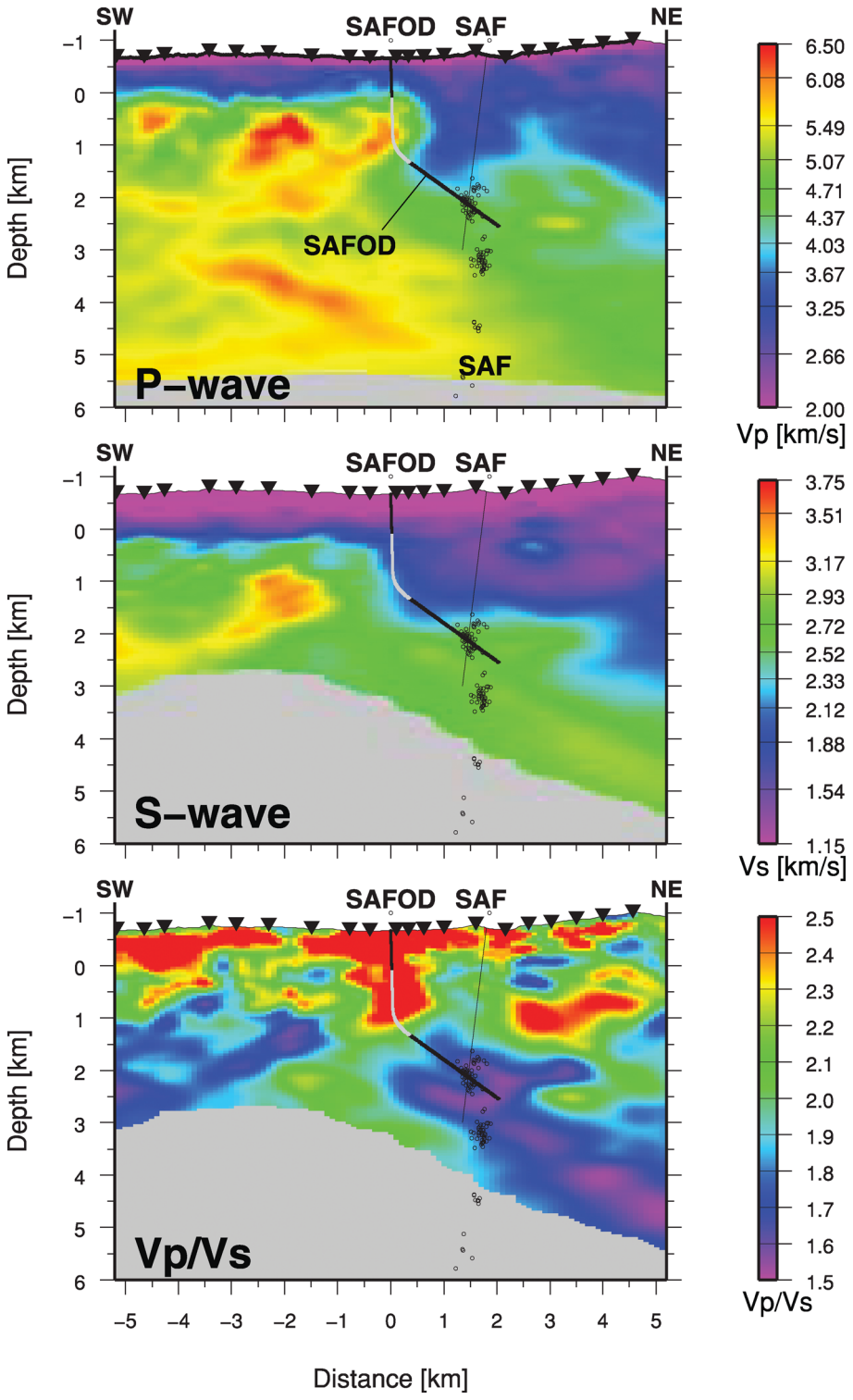

Figure 9. V_p, V_s models and V_p/V_s ratio in the vicinity of SAFOD and the SAF. All models are centred at SAFOD, roughly 1.8 km SW of SAF; units are in kilometre. Only cells having more than 100 rays passing through them are coloured. Other regions are grey. The low V_p/V_s feature at 2 km depth on the SAF which dips moderately northeastward is not obvious in V_p and V_s diagrams alone. The black and grey line shows the SAFOD drill hole: the grey part indicates the drilled granite.

centred at $V_p/V_s = 2.1$ and $V_p = 3.1$ km s^{-1} and $V_p/V_s = 1.7$ and $V_p = 4.7$ km s^{-1} can be seen. In addition to these dominant clusters, we identified several less-pronounced clusters with higher and lower velocities.

6 CLUSTER DETERMINATION

The appearance of clusters is not surprising, since rock types with different elastic properties (V_p and V_s velocities) crop out in the

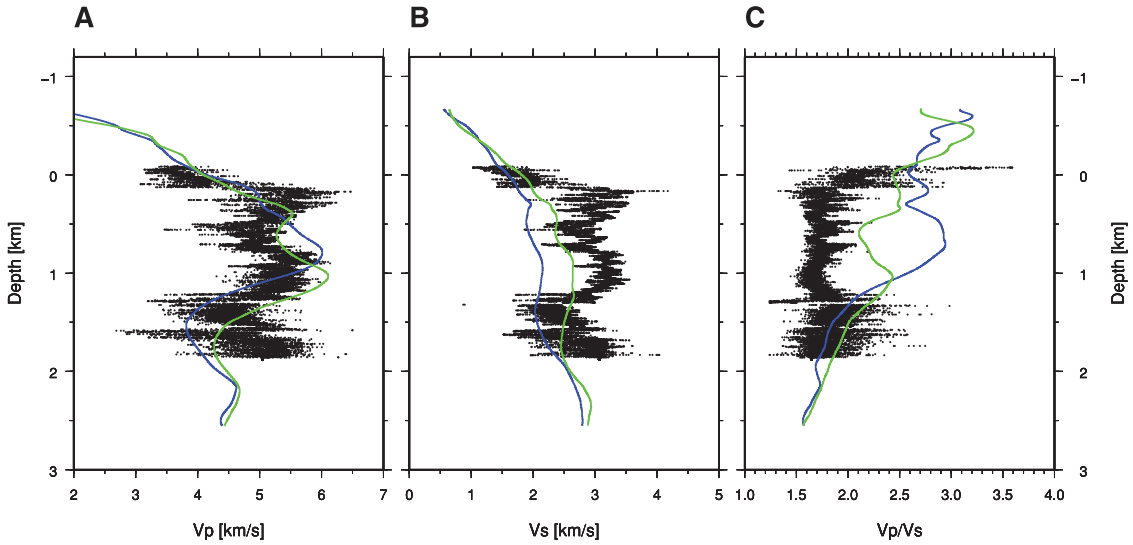

Figure 10. Diagram showing comparison of downhole V_p (a), V_s (b) and V_p/V_s-ratio (c) measurements in main SAFOD hole (data available at http://www.icdp-online.de/contenido/icdp/front_content.php?idart=1033) and the tomographic velocities. Black dots represent downhole velocities. Blue lines are tomographic velocities taken from V_p, V_s and V_p/V_s models (Fig. 9) along trace of the SAFOD drill hole. Green lines show velocities along drill track shifted by 500 m to southwest. See text and Appendix A for discussion.

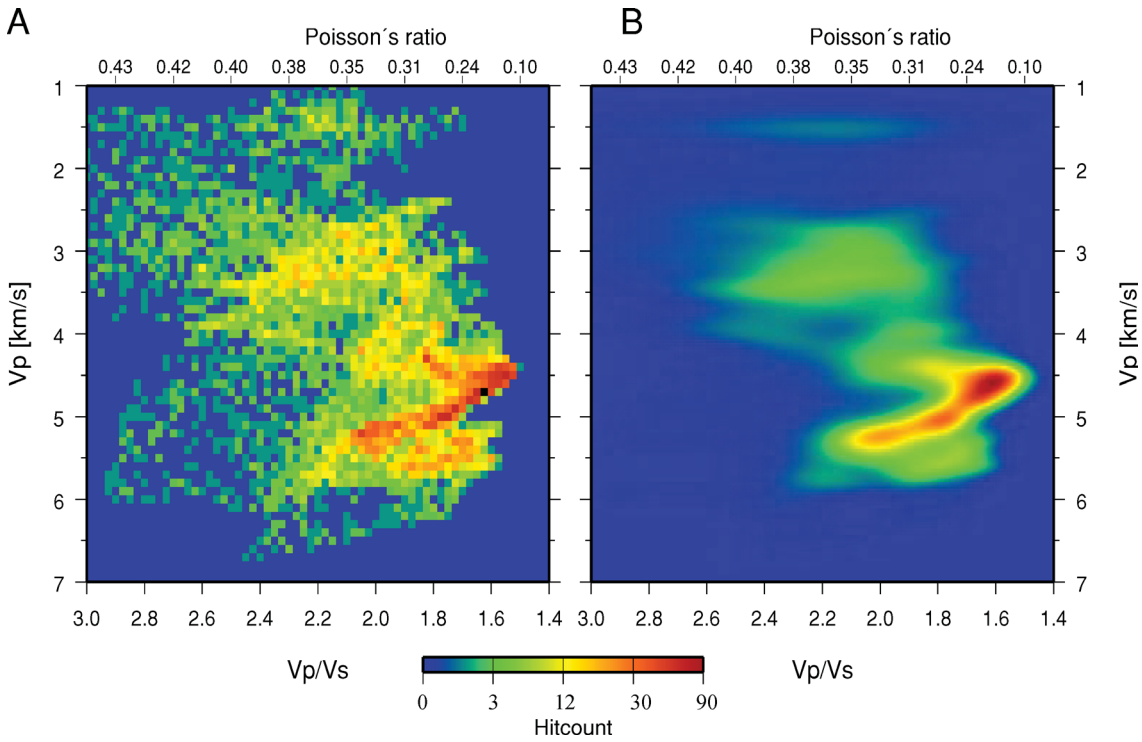

Figure 11. (a) Result of a correlation analysis (histogram plot similar to Bauer *et al.* (2003)) of the V_p and V_s models from Fig. 9. Red colours indicate that a given pair of V_p and V_p/V_s values occur most frequently, blue, least frequently. Generally, V_p/V_s ratios are higher than 1.6 and clustered. (b) Joint probability density function (PDF) of Fig. 11(a), which takes into account the accuracy (uncertainties) of the V_p and V_s values in individual inversion cells.

vicinity of the SAF. Different rocks of homogeneous compositions should appear in Fig. 11(b) ideally as separate points, 'smeared' by the process of tomographic recovery. If we assume that a specific rock (ideally a point in the V_p–V_p/V_s space) is altered by, for instance, cracks, pores and fluid content (due to deformation, weathering, etc.), then its corresponding V_p and V_p/V_s values will be shifted in a systematic way, that is, V_p will be decreased and V_p/V_s will tend to higher values (Babeyko *et al.* 1994; Popp & Kern 1994). This shift means that the same rock type with different crack and pore densities or fluid contents will be imaged in an elongate distribution with the long axis extending toward the upper left corner of the diagrams in Fig. 11 with respect to the unaltered rock. The main yellow, brown and green clusters in Fig. 11(b), with V_p between 4.5 and 5.5 km s^{-1} and V_p/V_s between 1.6 and 2.1, is unlikely to represent a single rock type given the fact that its long axis extends in a direction at an angle to the 'weathering tail' expected for a single rock type, as discussed above. This clustering most likely represents different rock types.

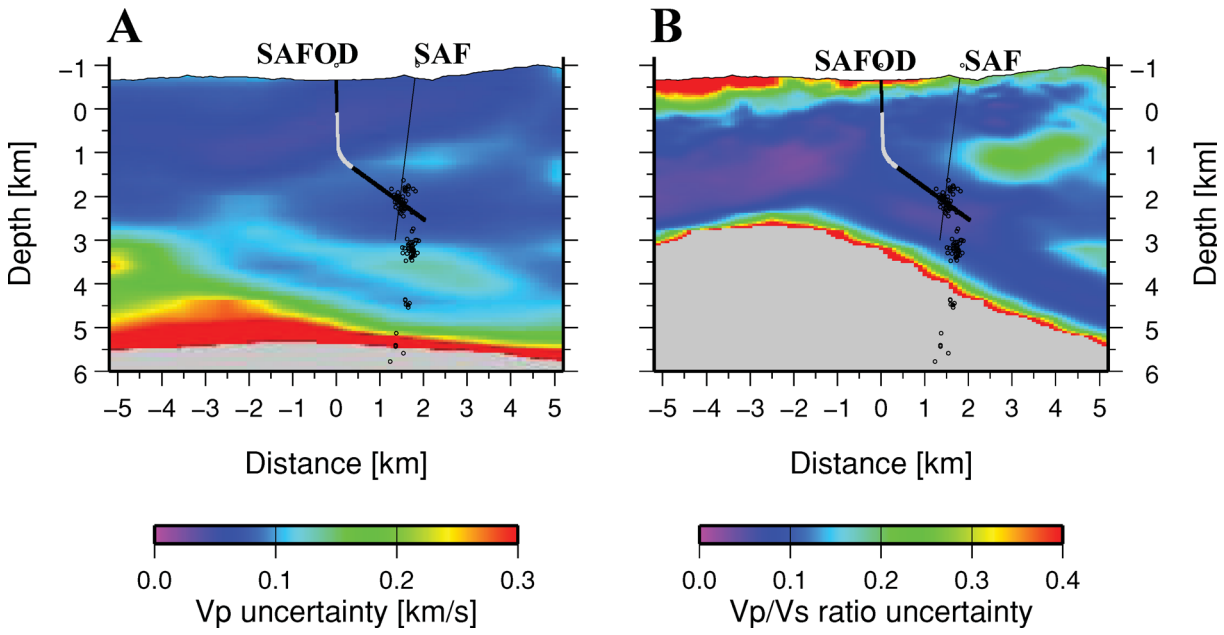

Figure 12. (a) Distribution of V_p uncertainty, based on the ray hitcounts of the tomographic model. (b) Distribution of the V_p/V_s uncertainty.

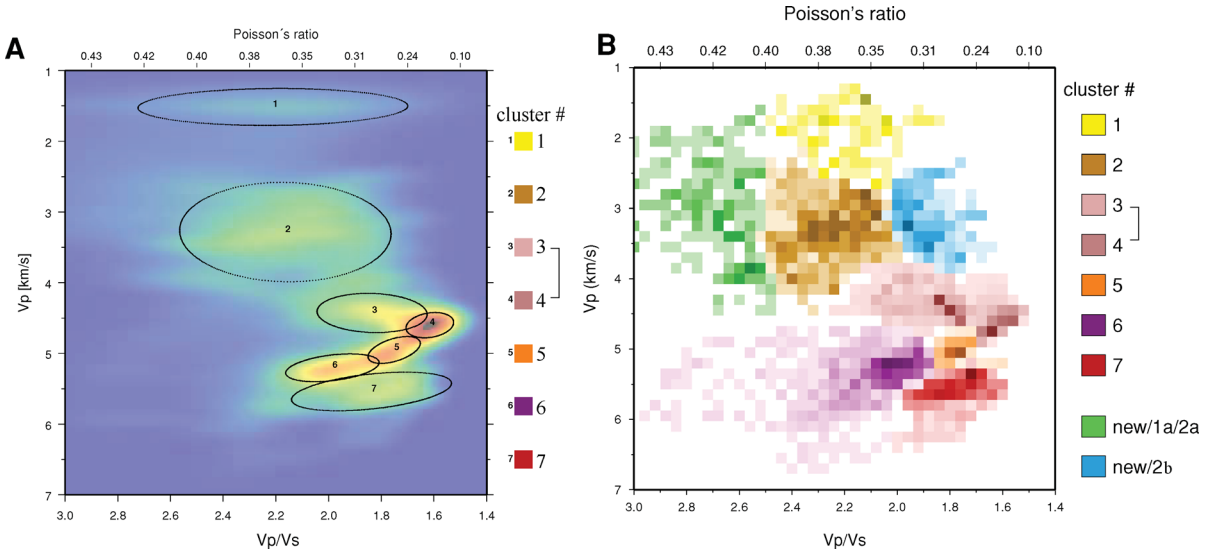

Figure 13. (a) Histogram of Fig. 11(b) with automatically defined, numbered cluster ellipses (see Appendix B). (b) Histogram from cluster analysis using a neural network technique (see Appendix C). Different colours assigned to the clusters and are used in Fig. 14. Colour saturation used here indicates the frequency of occurrence of inversion cells with V_p–V_p/V_s pairs.

To avoid a manual, thus quite subjective cluster assignment, we applied automatic cluster analysis techniques based on two different methods. The application of these analytical techniques give the position and shape of ellipses defining several clusters.

6.1 Automatic cluster determination

We searched for an optimum set of clusters/classes and determine their properties automatically, see Appendix B. From visual inspection of Fig. 11(b) more than five clusters can be identified according to the distribution of the local extrema. In an attempt to differentiate and capture a large number of potentially different rock types and to minimize regions for which no cluster could be assigned we decided to use seven clusters/classes. Fig. 13(a) shows Fig. 11(b) replotted, but this time with the clusters outlined by ellipses. Note that there is some minor overlap of clusters 3–7. Two of the clusters (labelled 3 and 4) are closely related to each other and most likely represent the same rock type, although at different alteration (weathering) states. Generally this cluster analysis and remapping approach is limited by the fact that different rock types naturally may have similar V_p and V_s velocities. This will cause overlapping clusters which might not be identified with this technique. Furthermore, volumetrically small amounts of rock types could be hidden by clusters of other rock types. Note that this automatic cluster determination method leaves significant numbers of V_p–V_p/V_s pairs unassigned to a specific cluster.

6.2 Cluster determination based on neural network techniques

To validate the results of cluster determination from the joint PDF as described in the previous sections, we analysed the data with a fundamentally different, more abstract approach using neural network techniques (Kohonen 1995; Bauer *et al.* 2008). A brief description is given in the Appendix C.

Fig. 13(b) shows the results from the neural network analysis of the SAFOD tomographic images. Each cluster is assigned a particular colour (Fig. 13b). The neural network analysis provides nine clusters.

6.3 Mapping of clusters back into $x-z$ plane

We next mapped the V_p-V_p/V_s pairs that fall in various clusters back to the x–z plane. The cluster colour assignments of Figs 13(a) and (b) are used for this mapping. Light grey areas in Fig. 14(b) indicate that the V_p–V_p/V_s values fall outside of assigned clusters of Fig. 14(a). Unfortunately, the immediate vicinity of SAFOD falls in a large grey area. In Fig. 14(d) much of the x–z plane is covered by cluster assignments and hence coloured units. Nominal rock-type interpretations for each cluster are given in the legends for Fig. 14, but these are, of course, non-unique.

The distribution of the remapped clusters from the two different methods is surprisingly similar. The chief difference are: (1) Two additional clusters were identified by the neural network technique over the automatic technique (clusters 1a/2a and 2b in Fig. 13), (2) the neural network clusters cover more of the x–z plane, including the vicinity of SAFOD (vertical part), which is mapped as cluster 6 (purple).

7 INTERPRETATION

To interpret the rock types in each cluster, we have plotted in Fig. 15V_p and V_p/V_s values determined in the laboratory or in the field (*in situ*) for rocks from SAFOD, the Varian well (∼8 km east of SAFOD) and other areas containing rocks similar to those cropping out near our profile. T.M. Brocher (written communication, 2006) kindly provided the raw data from his summary paper on V_p/V_s (Brocher 2005) and N. Boness and M.D. Zoback (written communication, 2006) kindly provided new data from sedimentary rocks (siltstone, shale, sandstone) penetrated in the deeper part of SAFOD, below the bend in the drill hole. The following is a summary for all clusters of Fig. 15: Cluster 1 occurs in a region of Fig. 15 spanned by data points from *in situ* weathered rocks, within 30 m of the surface, including Tertiary sedimentary rocks and Franciscan rocks. Cluster 2 occurs in a region spanned chiefly by Miocene sedimentary rocks penetrated in the Varian well (8 km SE of SAFOD) and marginally by Franciscan greywackes. Sedimentary rocks penetrated by SAFOD below the bend include chiefly arkoses, between the bend and the SAF, damage-zone rocks bracketed by three chief branches of the SAF, and Great Valley sequence (GVS) sedimentary rocks (Upper Cretaceous) east of the SAF (Bradbury *et al.* 2007; Draper Springer *et al.* 2009; Zoback *et al.* 2010). These sedimentary rocks have a large scatter in physical properties (Fig. 15) spanning clusters 3, 4, 5 and parts of clusters 2, 6 and 7. However, the average value of the GVS rocks and arkoses fall squarely in clusters 3 and 5, respectively. The average value of SAFOD granitic rocks falls within cluster 7, but individual data points scatter into cluster 5 with the arkoses. The arkoses were mostly likely derived from rocks similar to the granitic rocks (Draper Springer *et al.* 2009). Clusters 6 and 7 are overlapped by the sparse data points for serpentinite.

In the remapping of clusters back into the x–z plane (Figs 14b and d), an interpretable picture emerges given the rock-type correlations of Fig. 15. Cluster 1 (yellow; plus green in Fig. 14d) maps into the near-surface region (upper 30 m) throughout the model. Rocks in this region are interpretable as weathered rocks (Fig. 15). Cluster 2 (tan; plus blue in Fig. 14d) maps into the shallow layer, approximately 600–800 m thick, west of SAFOD; Tertiary sedimentary and volcanic rocks and metamorphic equivalents crop out at the surface in this area. Approximately 800 m of such rocks were penetrated in SAFOD, above the granitic rocks. East of SAFOD, Cluster 2 maps into a layer 2–3 km thick. In this region, both Tertiary sedimentary rocks and Franciscan sedimentary and volcanic rocks crop out at the surface (Fig. 2), in agreement with the rock types that overlap this cluster in Fig. 15. The interpretation of cluster 2 in this region might include either of these two rock types at depth: (1) Between SAFOD and a point ∼0.5 km east of the SAF, Tertiary sedimentary rocks crop out and are interpreted to extend to a depth of 1 km or more by Zoback *et al.* (2010); (2) farther east, Franciscan sedimentary and volcanic rocks and metamorphic equivalents are the chief outcrops (except for a thin layer of Miocene sediments between 5 and 8 km east of SAFOD; Fig. 2). Thus, Franciscan rocks may be interpretable in the Cluster 2 region at depth east of SAF. Clusters 3–5 (shades of pink and orange) map into the a region east of the bend in SAFOD and below 1.5 km depth. These clusters correspond to the sedimentary rocks penetrated by SAFOD east of its bend. The clusters do not distinguish the arkoses from the GVS sedimentary rocks in the fashion in which they are observed in SAFOD, but at least sedimentary rocks are mapped in locations where they are observed in SAFOD. There is a suggestion that these sedimentary rocks are offset across the SAF, higher on the east. Cluster 7 (reddish brown) maps to a region below 1 km depth west of SAFOD. This cluster correlates with granitic rocks penetrated in SAFOD.

Cluster 6 (purple) maps to a position below the BCF and west of SAFOD in Fig. 14(b). In fact, we have interpreted the dip of the BCF west of SAFOD to lie along the top of this cluster. However, in the neural-network cluster map, cluster 6 maps to most of the positions in which no cluster was mapped in the automatic-cluster algorithm, namely along the vertical part of SAFOD and also below 1 km depth to the west. In much of this cluster 6 region, one would interpret granitic rocks, based on SAFOD results. We interpret that the mapping of cluster 6 in Fig. 14(d) is more poorly constrained than in Fig. 14(b), based on the inconsistency between the mappings, and we do not interpret the purple regions in Fig. 14(d) as correlative with serpentinite, as would be suggested. We note that in the vicinity of SAFOD, V_s and V_p deviate from the downhole logs, V_s more than V_p (Fig. 10). This 'distortion' in the tomographic velocities would increase V_p/V_s artificially, perhaps moving the points to the vicinity of cluster 6. Since the neural-network algorithm covers more of the V_p–V_p/V_s space than the automatic-cluster algorithm (Figs 13b and 14c), it may contain more of these artefacts for cluster 6 than the automatic algorithm.

One place where we do not see serpentinite, where it is predicted by McPhee *et al.* (2004), is at a depth of about 2 km east of the SAF. Serpentinite is also not penetrated by SAFOD east of the SAF. To resolve this discrepancy, D. McPhee *et al.* (oral communication, 2012) will use the SAFOD and aeromagnetic constraints to remodel this magnetic body.

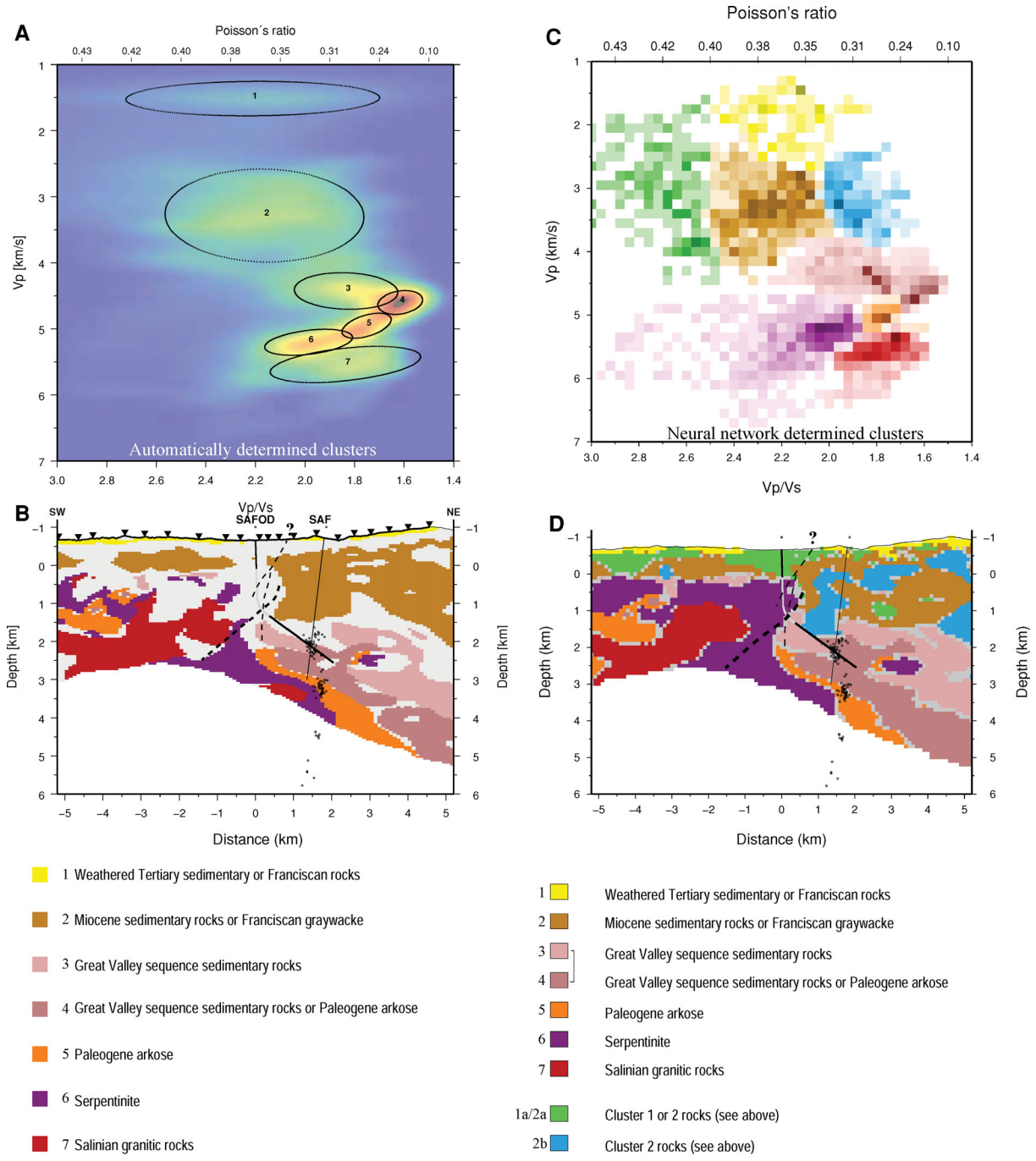

Figure 14. (a) Histogram from Fig. 13(a) with automatically defined, numbered cluster ellipses. In legend below, clusters are assigned colours and nominal rock-type interpretation from Fig. 15. (b) Distribution of clusters of (a) when mapped back to *x*–*z* plane. Note that most of the tomographically inverted cells have a cluster assignment, grey spaces represent unassigned cells. Light dashed lines, interpreted branches of Buzzard Canyon Fault (BCF) from Catchings *et al.* (2005); heavy dashed line, our interpretation of fault along top of pink- and purple-coloured patches that may or may not be related to Buzzard Canyon fault (question mark on upward connection). Our interpreted BCF fault intersects SAFOD drill path at location of fault separating granitic rocks above from sedimentary rocks below (see text). (c) Histogram from Fig. 13(b). (d) Remapping of clusters from (c) into *x*–*z* plane, similar to (b).

8 COMPARISON WITH RESULTS OF ZHANG *et al.* (2009)

Zhang *et al.* (2009) performed a similar cluster analysis using separately V_p and electrical resistivity, V_s and resistivity and V_p/V_s and resistivity. Their results are compared with the results of this study in Fig. 16. Similarities and dissimilarities can be seen. Similarities are as follows: Rocks with low V_p, low V_s, high V_p/V_s and low resistivity (black) are found near the surface, similar to our 'weathered Tertiary sedimentary rocks or weathered Franciscan rocks' (yellow). Rocks with somewhat higher V_p (~3.5–4.0 km s^{-1}), higher V_s (~1.5–2.5 km s^{-1}), lower V_p/V_s (~1.9) and higher resistivity (<20 Ωm) (red) correlate with our 'Miocene sedimentary rocks or Franciscan greywacke' (tan). The former would be interpreted west of SAFOD; the latter, east of SAFOD. Rocks penetrated by SAFOD below the BCF, having intermediate V_p (~4.5–6.0 km s^{-1}),

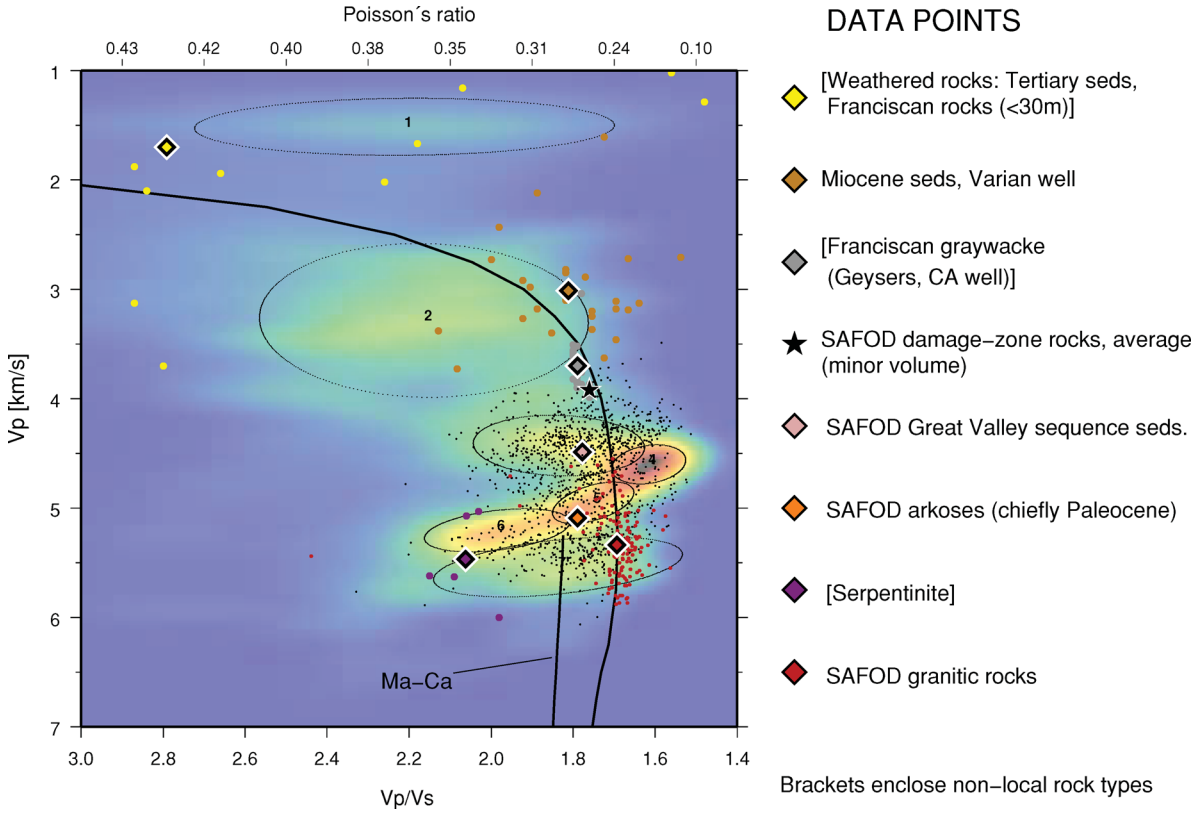

Figure 15. Histogram of Fig. 13(a) with automatically defined, numbered cluster ellipses. We have superposed data points from *in situ* or laboratory measurements of different rock types from SAFOD and other regions. Individual data points are indicated by coloured circles, their respective averages are shown as coloured diamonds. The thick curves are model curves from Brocher (2005). The one labelled Ma–Ca is the model curve for mafic and calcium-rich rocks. Individual data points of SAFOD GVS sedimentary rocks and arkoses are both represented by black dots.

intermediate V_s (∼2.5–3.5 km s^{-1}), low V_p/V_s (∼1.7–1.85) and intermediate resistivity (∼20–125 Ωm) (green and cyan) correlate with our 'Great Valley sequence sedimentary rocks and Palaeogene arkose' (shades of pink and orange). Rocks with high V_p (∼6.0–6.2 km s^{-1}), high V_s (∼3.25–3.75 km s^{-1}), low V_p/V_s (∼1.7) and high resistivity (∼120–250 Ωm) (dark blue) correlate with our 'Salinian granitic rocks' (reddish brown) and also with our 'serpentinite' (purple). Note that in our study, we interpret only a small possible body of serpentinite associated with (and below) the BCF. Finally, rocks with moderately high V_p (5.5–6.0 km s^{-1}), moderately high V_s (3.0–3.5 km s^{-1}), low V_p/V_s (1.65–1.8) and moderately low resistivity (10–60 Ωm) (magenta) seem to correlate east of SAFOD with our 'Great Valley sequence sedimentary rocks and Palaeogene arkose' (shades of pink and orange). And, west of SAFOD, some of this magenta cluster is mixed in with rocks interpreted as Salinian granitic rocks, as is our orange cluster. It is interesting that, as in our study, Zhang *et al.* (2009) map no clusters in the vicinity of the vertical part of SAFOD, as well as along the boundaries of many units, most notably between their red and green units (our tan and dark pink units). In the vertical part of SAFOD, we have hypothesized that tomographic smearing of velocities near the BCF, a steeply dipping boundary between granitic and sedimentary rocks, has led to distortion of V_p, V_s and V_p/V_s in the vicinity of SAFOD that has led to poorly defined clustering in this location. Smearing in tomographic velocities near the boundaries of geological units may similarly explain blank areas on the cluster maps between units. Finally, our interpretation deep extension of the BCF (Fig. 16, heavy dashed line), correlates with a discontinuity, or blank area, in the dark blue cluster of Zhang *et al.* (2009). Here our studies agree on the discontinuity but not on the rock-type contrast. The similarities of cluster mapping between our study and that of Zhang *et al.* (2009) appear to lend credibility to the use of the cluster-mapping technique. Dissimilarities in shape between our study and that of Zhang *et al.* (2009) can be seen. These may arise from the fact that our resolution for V_p/V_s is best in the region from the surface to as much as 3 km depth (Fig. 7), whereas the resolution of Zhang *et al.* (2009) is best at depths below about 2 km.

9 CONCLUSIONS

Seismic traveltime data for *P*- and *S*-wave phases were used to derive velocity models for the central California region near SAFOD. 2-D traveltime tomography from controlled-source seismic data yielded independent V_p and V_s models, and a V_p/V_s model was calculated from these two models. The velocity models have been compared with available SAFOD downhole data. Cluster analysis of the models was performed whereby clusters of data points in a 2-D histogram of V_p and V_p/V_s values were identified, assigned a colour and typical rock type, and mapped back into x–z plane. Two essentially different cluster analysis approaches, one based on an automatic cluster determination and one based on neural network techniques, have been developed and found to provide similar results. After mapping back the clusters, the new x–z cross section is broadly consistent with outcrop, SAFOD and other data, but is inconsistent with the layer of serpentinite inferred from

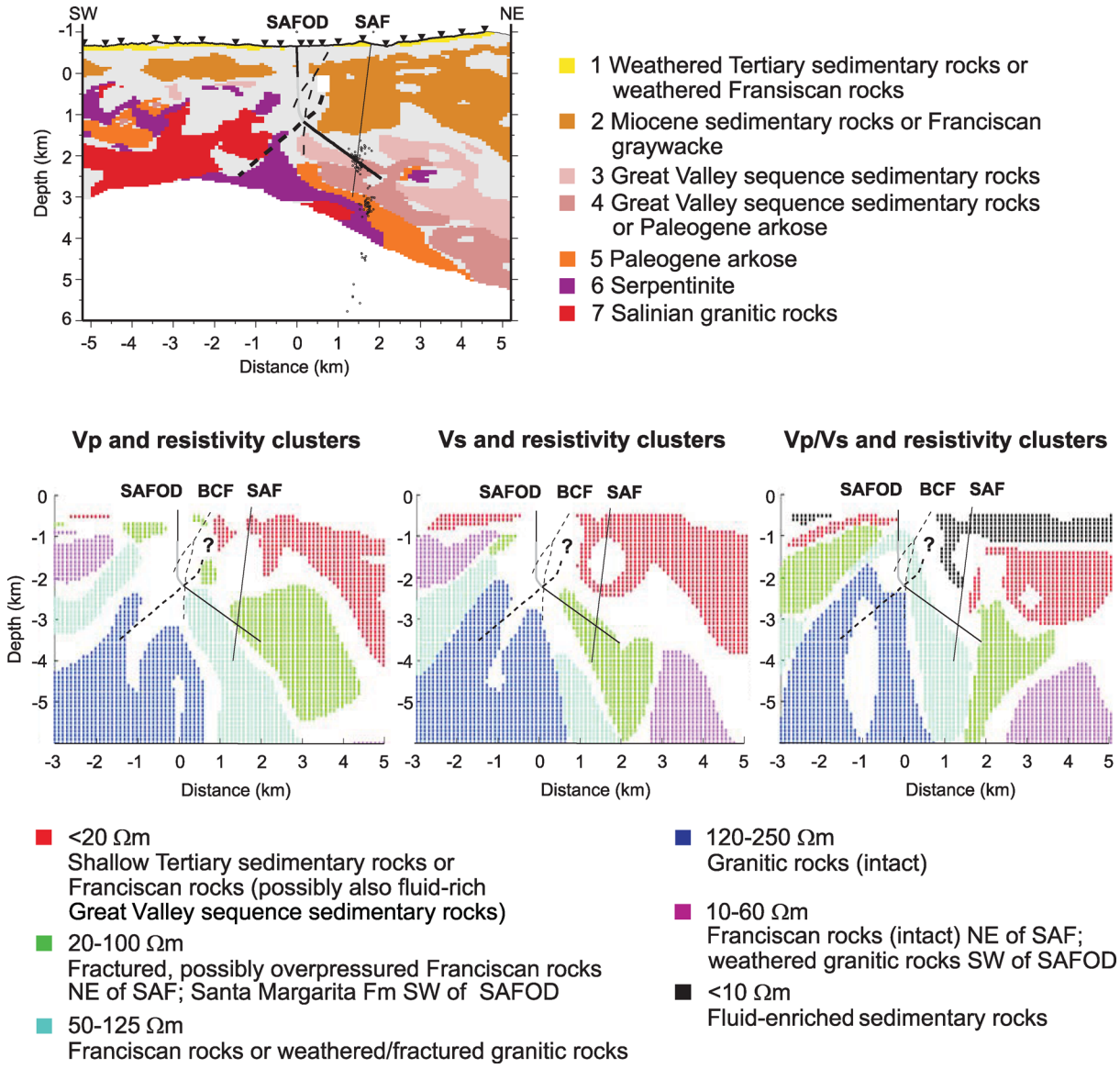

Figure 16. Comparison of our interpretation (top) with those of Zhang *et al.* (2009) (bottom, note that depth is depth below gound level), see text for details.

magnetic data east of SAF. The neural-network algorithm may contain more artefacts than the automatic algorithm. Using this technique, we are able to map the fault intersected by SAFOD, where rock type changes downhole from granitic rocks to arkosic sedimentary rocks. Cluster analysis, thus, provides additional information on rock types at depth, if carefully interpreted using all available data.

ACKNOWLEDGMENTS

The authors would like to thank Tom Brocher, Jeanne Hardebeck and the two anonymous reviewers for valuable comments on the paper. Without the kind cooperation of landowners in the Parkfield region, this project could not have been accomplished. The project was financed by the Deutsche Forschungsgemeinschaft (DFG), NSF, USGS and the GeoForschungsZentrum Potsdam (GFZ). Figures were prepared using the Generic Mapping Tool (GMT) (Wessel & Smith 1995, 1998).

REFERENCES

Babeyko, A.Y., Sobolev, S., Sinelnikov, E.D., Smirnov, Y.P. & Derevschikova, N.A., 1994. Calculation of elastic properties in lower part of Kola borehole from bulk chemical composition of core samples, *Surv. Geophys.,* **15,** 545–573.

Bakun, W.H. *et al.*, 2006. Implications for prediction and hazard assessment from the 2004 Parkfield earthquake, *Nature*, **437,** doi:10.1038/nature04067.

Barth, A.P., Wooden, J.L., Grove, M., Jacobson, C.E. & Pedrick, J.N., 2003. U-Pb zircon geochronology of rocks in the Salinas Valley region of California: a reevaluation of the crustal structure and origin of the Salinian block, *Geology,* **31,** 517–520, doi:10.1130/0091-7613(2003)031<0517:UZGORI>2.0.CO;2.

Bauer, K., Schulze, A., Ryberg, T., Sobolev, S.V. & Weber, M.H., 2003. Classification of lithology from seismic tomography: a case study from the Messum igneous complex, Namibia, *J. geophys. Res.,* **108**(B3), doi:10.1029/2001JB001073.

Bauer, K., Pratt, R.G., Haberland, C. & Weber, M.H., 2008. Neural network analysis of crosshole tomographic images: the seismic signature

of gas hydrate bearing sediments in the Mackenzie Delta (NW Canada), *Geophys. Res. Lett.,* **35,** L19306, doi:10.1029/2008GL035263.

Bedrosian, P.A., Unsworth, M.J., Egbert, G.D. & Thurber, C.H., 2004. Geophysical images of the creeping segment of the San Andreas fault: implications for the role of crustal fluids in the earthquake process, *Tectonophysics,* **385,** 137–158.

Bedrosian, P.A., Maercklin, N., Weckmann, U., Bartov, Y., Ryberg, T. & Ritter, O., 2007. Lithology-derived structure classification from the joint interpretation of magnetotelluric and seismic models, *Geophys. J. Int.,* **170,** 737–748.

Bleibinhaus, F., Hole, J.A. & Ryberg, T., 2005. Seismic reflection and diffraction imaging of the San Andreas Fault at SAFOD, *EOS, Trans. Am. geophys. Un.,* **86**(52), Fall Meeting Supplement, Abstract T21A–0446.

Bleibinhaus, F., Hole, J.A., Ryberg, T. & Fuis, G.S., 2007. Structure of the California Coast Ranges and San Andreas Fault at SAFOD from seismic waveform inversion and reflection imaging, *J. geophys. Res.,* **112,** doi:10.1029/2006JB004611.

Boness, N. & Zoback, M.D., 2006. A multiscale study of the mechanism controlling shear velocity anisotropy in the San Andreas Fault Observatory at Depth, *Geophysics,* **71,** F131–F146.

Boness, N.L. & Zoback, M.D., 2004. Stress-induced seismic velocity anisotropy and physical properties in the SAFOD Pilot Hole in Parkfield, CA, *Geophys. Res. Lett.,* **31,** doi:10.1029/2003GL019020.

Bradbury, K.K., Barton, D.C., Solum, J.G., Draper, S.D. & Evans, J.P., 2007. Mineralogic and textural analyses of drill cuttings from the San Andreas Fault Observatory at Depth (SAFOD) boreholes: initial interpretations of fault zone composition and constraints on geological models, *Geosphere,* **3,** 299–318.

Brocher, T.M., 2005. Empirical relations between elastic wavespeeds and density in the Earth's crust, *Bull. seism. Soc. Am.,* **95,** 2081–2095.

Catchings, R.D., Rymer, M.J., Goldman, M.R., Hole, J.A., Huggins, R. & Lippus, C., 2002. High-resolution seismic velocities and shallow structure of the San Andreas Fault zone at Middle Mountain, Parkfield, California, *Bull. seism. Soc. Am.,* **92,** 2493–2503.

Catchings, R.D., Rymer, M.J. & Goldman, M.R., 2005. The San Andreas Fault at SAFOD: seismic imaging and borehole comparisons, *EOS, Trans. Am. geophys. Un.,* **86**(52), Fall Meeting Supplement, Abstract T21A–0457.

Dibblee T.W., Jr., 1971. Geologic maps of seventeen 15-minute quadrangles (1:62 500) along the San Andreas fault in the vicinity of King City, Coalinga, Panoche Valley, and Paso Robles, California, U.S. Geol. Surv. Open-File Report [OF-71-87], U.S. Geological Survey.

Dibblee T.W., Jr., 1980. Geology along the San Andreas fault from Gilroy to Parkfield, in *Studies of the San Andreas Fault Zone in Northern California,* Special Report 140, pp. 3–18, eds Streitz, R. & Sherburne, R., California Division of Mines and Geology, Sacramento, CA.

Dickinson, W.R., 1966. Structural relationships of San Andreas fault system, Cholame Valley and Castle Mountain Range, California, *Geol. soc. Am. Bull.,* **77,** 707–726.

Draper Springer, S., Evans, J.P., Garver, J.I., Kirschner, D. & Janecke, S.U., 2009. Arkosic rocks from the San Andreas Fault Observatory at Depth (SAFOD) borehole, central California: Implications for the structure and tectonics of the San Andreas fault zone, *Lithosphere,* **1,** 206–226.

Grove, M., Jacobson, C.E., Barth, A.P. & Vucic, A., 2003. Temporal and spatial trends of Late Cretaceous-early Tertiary underplating of Pelona and related schist beneath southern California and southwestern Arizona, in *Tectonic Evolution of Northwestern Mexico and the Southwestern USA,* Geol. Soc. Am. Special Paper 374, pp. 381–406, eds Johnson, S.E., Fletcher, S.R.P.J.M., Girty, G.H., Kimbrough, D.L. & Martin-Barajas, A., Geological Society of America, Boulder, CO.

Haberland, C. *et al.*, 2003. Modeling of seismic guided waves at the Dead Sea Transform, *J. geophys. Res.,* **108**(B7), doi:10.1029/2002JB002309.

Hauser, F., O'Reilly, B.M., Readman, P.W., Daly, J.S. & Van den Berg, R., 2008. Constraints on crustal structure and composition within a continental suture zone in the Irish Caledonides from shear wave wide angle reflection data and lower crustal xenoliths, *Geophys. J. Int.,* **175,** 1254–1272.

Hickman, M., Zoback, M. & Ellsworth, W., 2004. Introduction to special section: preparing for the San Andreas Fault Observatory at depth, *Geophys. Res. Lett.,* **31,** L12S01, doi:10.1029/2004GL020688.

Holbrook, W.S., Mooney, W.D. & Christensen, N.I., 1992. The seismic velocity structure of the deep continental crust, in *Continental Lower Crust,* pp. 257–270, eds Fountain, D.M., Arculus, R. & Kay, R.W., Elsevier, Amsterdam.

Hole, J., Ryberg, T., Fuis, G., Bleibinhaus, F. & Sharma, A.K., 2006. Structure of the San Andreas fault zone at SAFOD from a seismic refraction survey, *Geophys. Res. Lett.,* **33,** L07312, doi:10.1029/2005GL025194.

Hole, J.A., 1992. Nonlinear high-resolution three-dimensional seismic travel time tomography, *J. geophys. Res.,* **97,** 6553–6562.

Hole, J.A., Catchings, R.D., St. Clair, K.C., Rymer, M.J., Okaya, D.A., & Carney, B.J., 2001. Steep-dip seismic imaging of the San Andreas fault near Parkfield, *Science,* **294,** 1513–1515.

Irwin, W.P., 1990. Geology and plate-tectonic development, in *The San Andreas Fault System, California,*U.S. Geol. Surv. Open Prof. Paper 1515 pp. 51–80, ed. Wallace, R., U.S. Geological Survey, Washington, DC.

Kern, H., Gao, S., Jin, Z., Popp, T. & Jin, S., 1999. Petrophysical studies on rocks from the Dabie ultrahigh-pressure (UHP) metamorphic belt, Central China: implications for the composition and delamination of the lower crust, *Tectonophysics,* **301,** 191–215, doi:10.1016/S0040-1951(98)00268-6.

Kohonen, T., 1995. *Self-Organizing Maps,* Springer, New York, NY.

Maercklin, N., Haberland, C., Ryberg, T., Weber, M., Bartov, Y. & DESERT Group, 2004. Imaging the Dead Sea Transform with scattered seismic waves, *Geophys. J. Int.,* **158,** 179–186, doi:10.1111/j.1365-246x.2004.02302.x.

Mathews, V., III, 1976. Correlation of Pinnacles and Neenach Volcanic Formations and their bearing on San Andreas fault problem, *Am. Assoc. Petrol. Geol. Bull.,* **60,** 2128–2141.

McPhee, D.K., Jachens, R.C. & Wentworth, C.M., 2004. Crustal structure across the San Andreas Fault at the SAFOD site from potential field and geologic studies, *Geophys. Res. Lett.,* **31,** doi:10.1029/2003GL019363.

Mechie, J. *et al.*, 2004. Precise temperature estimation in the Tibetan crust from seismic detection of the $\alpha - \beta$ quartz transition, *Geology,* **32,** 601–604.

Mechie, J., Abu-Ayyash, K., Ben-Avraham, Z., El-Kelani, R., Mohsen, A., Rümpker, G., Saul, J. & Weber, M., 2005. Crustal shear velocity structure across the Dead Sea Transform from two-dimensional modelling of DESERT project explosion seismic data, *Geophys. J. Int.,* **160,** 910–924, doi:10.1111/j.1365-246x.2005.02526.x.

Mechie, J. *et al.*, 2012. Crustal and uppermost mantle velocity structure along a profile across the Pamir and southern Tien Shan as derived from project TIPAGE wide angle seismic data, *Geophys. J. Int.,* **188,** 385–407.

Moos, D. & Zoback, M.D., 1983. In situ studies of velocity in fractured crystalline rocks, *J. geophys. Res.,* **88,** 2345–2358.

Murphy, J.M., Fuis, G.S., Ryberg, T., Lutter, W.J., Catchings, R.D. & Goldman, M.R., 2010. Detailed P- and S-wave velocity models along the LARSE II transect, Southern California, *Bull. seism. Soc. Am.,* **100,** 3194–3212.

Musacchio, G., Mooney, W.D., Luetgert, J.H. & Christensen, N.I., 1997. Composition of the crust in the Grenville and Appalachian Provinces of North America inferred from V_P/V_S ratios, *J. geophys. Res.*, **102,** 15 225–15 241.

Page, B.M., 1981. The southern Coast Ranges, in *The Geotectonic Development of California,* pp. 329–417, ed. Ernts, W.G., Prentice Hall, Englewood Cliffs, NJ.

Popp, T. & Kern, H., 1994. The influence of dry and water saturated cracks on seismic velocities of crustal rocks—a comparison of experimental data with theoretical model, *Surv. Geophys.,* **15,** 443–465.

Raileanu, V., Bala, A., Hauser, F., Prodehl, C. & Fielitz, W., 2005. Crustal properties from S-wave and gravity data along a seismic refraction profile in Romania, *Tectonophysics,* **240,** 251–272.

Roecker, S., Thurber, C., Shuler, A., Liu, Y., Zhang, H. & Powell, L., 2005. Refined images of the crust around the SAFOD drill site derived from combined active and passive seismic experiment data, *EOS, Trans. Am. geophys. Un.,* **86,** Fall Meet. Suppl., Abstract T24B–05.

Ross, D.C., 1984. Possible correlations of basement rocks across the San Andreas, San Gregorio-Hosgri, and Rinconada Reliz-King City faults, California, U.S. Geol. Surv. Prof. Paper 1317, p. 37, U.S. Geological Survey.

Ryberg, T., Fuis, G.S., Bauer, K., Hole, J.A. & Bleibinhaus, F., 2005. Upper-crustal reflectivity of the Central California Coast Range near the San Andreas Fault Observatory at depth (SAFOD), USA, *EOS, Trans. Am. geophys. Un.,* **86,** Fall Meeting Supplement, abstract T21A–0441.

Ryberg, T., Weber, M.H., Garfunkel, Z. & Bartov, Y., 2007. The shallow velocity structure across the Dead Sea Transform fault, Arava Valley, from seismic data, *J. geophys. Res.,* **112,** B08307, doi:10.1029/2006JB004563.

Rymer, M.J., 1981. Geologic map along a 12 kilometer segment of the San Andreas fault zone, southern Diablo Range, California, U.S. Geol. Surv. Open File Rep. 81-1173, scale 1:12 000, U.S. Geological Survey, http://pubs.er.usgs.gov/pubs/ofr/ofr811173.

Rymer, M.J., Catchings, R.D. & Goldman, M.R., 2003. Structure of the San Andreas Fault zone as revealed by surface geologic mapping and high-resolution seismic profiling near Parkfield, California, *Geophys. Res. Abs.,* **5,** 13 523.

Rymer, M.J. *et al.*, 2006. Surface fault slip associated with the 2004 Parkfield, California, earthquake, *Bull. seism. Soc. Am.,* **96,** 11–27, doi:10.1785/0120050830.

Sims, J.D., 1988. Geologic map of the San Andreas fault zone in the Cholame Valley and Cholame Hills quadrangles, San Luis Obispo and Monterey Counties, California, U.S. Geol. Surv. Misc. Field Stud. Map MF-2115, scale 1:24 000, U.S. Geological Survey.

Sims, J.D., 1990. Geologic map of the San Andreas fault in the Parkfield 7.5-minute quadrangle, Monterey and Fresno Counties, California, U.S. Geol. Surv. Misc. Field Stud. Map MF-2115, scale 1:24 000, U.S. Geological Survey.

Stiermann, D. & Kovach, R., 1979. An in situ velocity study: the Stone Canyon well, *J. geophys. Res.,* **84,** 672–678.

Thayer, M.R., 2006. Structural geology of the San Andreas fault zone at Middle Mountain, near Parkfield, central California, *MSc thesis,* Arizona State University, Tempe, AZ.

Thurber, C., Roecker, S. & Zhang, H., 2004a. Defining the SAFOD drilling trajectory: locating the target earthquakes, *EOS, Trans. Am. geophys. Un.,* **86,** Fall Meeting Supplement, Abstract T11F–04.

Thurber, C., Roecker, S., Zhang, H., Baher, S. & Ellsworth, W., 2004b. Fine-scale structure of the San Andreas fault zone and location of the SAFOD target earthquakes, *Geophys. Res. Lett.,* **31**(3), doi:10.1029/2003GL019398.

Thurber, C.H., Roecker, S., Roberts, K., Gold, M., Powell, L. & Rittger, K., 2003. Earthquake locations and three-dimensional fault zone structure along the creeping section of the San Andreas fault near Parkfield, CA: preparing for SAFOD, *Geophys. Res. Lett.,* **30,** doi:10.1029/2002GL016004.

Unsworth, M. & Bedrosian, P.A., 2004. Electrical resistivity structure at the SAFOD site from magnetotelluric exploration, *Geophys. Res. Lett.,* **31,** doi:10.1029/2003GL019405.

Wessel, P. & Smith, W., 1995. A new version of the Generic Mapping Tools (GMT), *EOS, Trans. Am. geophys. Un.,* **76,** 329, doi:10.1029/95EO00198.

Wessel, P. & Smith, W., 1998. New, improved version of generic mapping tools released, *EOS, Trans. Am. geophys. Un.,* **79,** 579, doi:10.1029/98EO00426.

White, D.J., Milkereit, B., Salisbury, M.H. & Percival, J.A., 1992. Crystalline lithology across the Kapuskasing uplift determined using in situ Poisson's ratio from seismic tomography, *J. geophys. Res.,* **97,** 19 993–20 006, doi:10.1029/92JB01744.

Zelt, C.A., 1999. Modelling strategies and model assessment for wide angle seismic traveltime data, *Geophys. J. Int.,* **139,** 183–204.

Zelt, C.A. & Barton, P.J., 1998. 3D seismic refraction tomography: a comparison of two methods applied to data from the Faeroe Basin, *J. geophys. Res.,* **103,** 7.187–7.210.

Zhang, H., Thurber, C. & Bedrosian, P., 2009. Joint inversion for V_p, V_s, and V_p/V_s at SAFOD, Parkfield, California, *Geochem. Geophys. Geosyst.,* **10,** Q11002, doi:10.1029/2009GC002709.

Zoback, M.D., Hickmann, S. & Ellsworth, W., 2005. Overview over SAFOD Phases 1 and 2: Drilling, sampling and measurements in the Sand Andreas Fault zone at seismologenic depth, *EOS, Trans., Am. geophys. Un.,* **86,** Fall Meeting Supplement, Abstract T23E–01.

Zoback, M.D., Hickman, S. & Ellsworth, W., 2010. Scientific drilling into the San Andreas fault zone, *EOS, Trans. Am. geophys. Un.,* **91,** 197–204.

APPENDIX A: DOWNHOLE AND TOMOGRAPHIC VELOCITIES

Disagreement between downhole and tomographic measurements of V_p and V_s can arise from at least three sources: (1) strong lateral velocity variation that is not resolved by tomography, (2) anisotropy and (3) sampling size. In the vicinity of SAFOD, all of these effects could conspire to make tomographic V_p and V_s smaller on average than downhole values (Fig. 10). In addition, it is a well known fact, that tomography typically does not fully recover the magnitude of velocity anomalies.

A1 Lateral velocity variation not resolved by tomography

From the checkerboard tests in our study, 500 m blocks in the vicinity of SAFOD were well resolved to approximately the depth of the bend in the main hole (Fig. 5b). There is no significant lateral smearing or smoothing apparent at this scale; however, this test assumes infinite-frequency waves. In fact, at a typical frequency of 20 Hz, a *P* wavelength is 250 m and would, in principle, resolve only features significantly larger than that value. An *S* wavelength, at a typical frequency of 10 Hz, would sample ∼300 m laterally and resolve only features significantly larger than that value. The typical frequencies strongly depend on the actual source-receiver distance, which corresponds to the maximum depth sampled by a ray. So the shallower structure of the model is sampled by rays with higher frequencies and is thus resolved with potential better resolution. One feature to note in our V_p and V_s models (Fig. 4) is that there is a steep velocity discontinuity centred only ∼200 m northeast of SAFOD for V_p and centred actually southwest of SAFOD for V_s (refer to Fig. 4; green colours represent the approximate centre of a gradient from red-orange to blue, in both cases). Thus, one would, in fact, expect the tomographic velocities to be lower than downhole measurements owing to tomographic averaging of velocities at SAFOD with lower velocities to the northeast. The fact that tomographic velocities generally rise to the southwest of SAFOD, as is seen in Fig. 10 (green lines), is consistent with this expectation.

A2 Anisotropy

Anisotropy expected at SAFOD is of the right sense to make both tomographic V_p and V_s velocities lower than downhole measurements. Downhole measurements determine velocities in the vertical direction, whereas tomography measures velocities at the turning points of refracted rays, where the rays are approximately horizontal. Downhole measurements of V_s distinguish, for vertically travelling waves, which horizontal vibration direction produces the fastest speed. Boness & Zoback (2004, 2006) find that the V_s fast vibration direction is N–S at the surface and rotates to NNE–SSW at a depth of 2 km in the SAFOD pilot hole, following closely the azimuth of the maximum horizontal compressive stress. In their interpretation, randomly oriented cracks in the bedrock are preferentially closed when they are not near-vertical

and do not have a strike near the maximum compressive direction. For tomographic V_s, velocities are modelled from traveltimes read from the transverse components of our seismographs, where the vibration direction of the S wave is approximately horizontal and NW–SE, given the orientation of our line. Thus tomographic S waves have a vibration direction different by ∼45° to ∼70° from the downhole S-wave fast-vibration direction. Tomographic V_s is, thus, expected to be lower than downhole V_s. However, the size of V_s anisotropy measured downhole ranges from 10 per cent near the surface to 3 per cent at 2 km depth, and could explain a maximum of 0.15 km s^{-1} of the ∼0.6 km s^{-1} average discrepancy noted above.

For V_p similar arguments can be made. Downhole measurements are made in the vertical direction, parallel to the interpreted open cracks, and are thus expected to yield relatively high values. Tomographic (refracted) P waves, on the other hand, traverse the interpreted open cracks horizontally at an azimuth that differs from the maximum compressive direction by 20°–45°, based on the crack orientations discussed above. Thus tomographic V_p is expected to be somewhat lower than the downhole V_p.

A3 Sampling size

Murphy *et al.* (2010) found a systematic difference between V_p and V_s determined from traveltime tomography along an active-source line in southern California (LARSE Line II) and V_p and V_s determined from laboratory and downhole measurements summarized by Brocher (2005). The difference occurs in the range of V_p = 3–5 km s^{-1} (or V_s ∼ 1.25–2.9 km s^{-1}) and is confirmed by several statistical tests. Among other possible interpretations, Murphy *et al.* (2010) cite the possibility that the difference arises from a difference in sampling size in tectonized rock–rock that contains abundant megafractures. This interpretation is favoured by Stiermann & Kovach (1979) and by Moos & Zoback (1983) from active source and downhole logging studies near the SAF. Laboratory samples are typically centimetres in size and bore-hole sampling lengths are typically on the order of a metre or so, whereas refractions sample a volume on the order of 100–1000 s of metres. In sampling a much larger volume, refractions traverse large fractures that are commonly associated with chemical alteration, that may significantly reduce their speeds.

APPENDIX B: AUTOMATIC CLUSTER DETERMINATION

To extract a reasonable number of separated classes from the calculated joint PDF we approximate it by a sum of n bivariate normal distributions

$$G(\mathbf{x}) = \sum_{k=1}^{n} \frac{a_k}{2\pi |\mathbf{C}_k|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\mu_k)^T \mathbf{C}_\mathbf{k}^{-1}(\mathbf{x}-\mu_k)\right] \qquad \text{(B1)}$$

using a nonlinear least-squares technique. Here $\mathbf{x} = [V_p/V_s,\ V_p]$ is the vector of input coordinates and a_k, $\mathbf{C}_k$, and $\mu_k (= [\mu_k^{V_p/V_s},\ \mu_k^{V_p}])$ are an amplitude factor, covariance matrix and mean value of the kth distribution function. Class boundaries are now defined by the confidence intervals for a given probability p of the individual Gaussian peaks. The appropriate elliptical contours are the zeros of

$$(\mathbf{x}-\mu_k)^T \mathbf{C}_k (\mathbf{x}-\mu_k) = \chi^{-2}(p, 2), \qquad \text{(B2)}$$

where the right-hand side of eq. (B2) stands for the the inverse cumulative χ^2-distribution of probability p and two degrees of freedom.

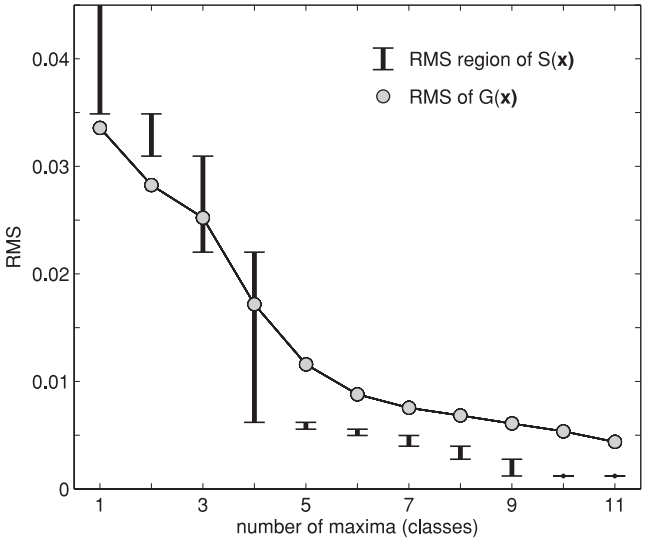

Figure B1. The rms error versus number of maxima for a bicubic spline approximation with continuously decreasing smoothing parameter (black bars) and a increasing set of bivariate normal distributions (grey circles), respectively. The L-shape (concave-up) curve displays the tradeoff between the rms an the number of Gaussian kernels and is used to estimate the optimal number (5–7) of classes.

The crucial point in estimating the 6n parameters (semi-major axis, dip angle, position, magnitude) of $G(\mathbf{x})$ are the starting values of the amplitude a_k and the mean μ_k of the n Gaussian functions.

To objectify this initial guess we perform bicubic smoothing spline interpolations $S(\mathbf{x})$ of the joint PDF. There is a monotonic dependency between smoothing parameter and number of maxima of $S(\mathbf{x})$ which is illustrated in Fig. B1. The 'errorbars' mark the rms-region of the interpolations with a same number of maxima. The locations of maxima for the corresponding mean rms are taken to be the initial guess μ_k. Starting with identity matrices for the covariance, $\mathbf{C}_k = \mathbf{E}$, and assuming that maximal values of $S(\mathbf{x})$ are mainly determined by the peak magnitude of a single Gausssian function the amplitude factor a_k can be expressed by the spline interpolation at μ_k, $a_k = 2\pi S(\mu_k)$. The rms of the resulting approximation G($\mathbf{x}$) as a function of number of classes are shown in Fig. B1.

The optimum number of classes follows from examination of this dependency. The 'knee' of the corresponding L-curve (Fig. B1) occurs between 5 and 7 classes. More classes do not significantly lower the misfit. To separate the regions of this optimal set of classes we choose a common probability $p = 0.5$ in eq. (B2) for all kernels. That means 50 per cent of the power of an individual Gaussian distribution is explained by each ellipse.

APPENDIX C: NEURAL NETWORK CLUSTER DETERMINATION

The self-organizing map (SOM, Kohonen 1995) is a neural network type, which can be used for cluster analysis of multidimensional data sets. For the mathematical and numerical details of the SOM method we refer to Kohonen (1995). In this study, we used the method of Bauer *et al.* (2008), which represents a modified version of the standard SOM. The usage of the method was demonstrated in Bauer *et al.* (2008) for a very similar problem of tomographic multiparameter classification and interpretation.

The principal work flow includes an unsupervised learning phase, application of the learned knowledge and data clustering and remapping of the cluster information. During unsupervised learning, the information of the tomographic images is analysed and mapped onto a 2-D so-called feature map using well-defined learning rules adopted from biological neural systems behaviour. As a result, grid blocks of the tomographic model with similar properties (V_p, V_p/V_s) are mapped onto adjacent regions on the feature map. A watershed segmentation algorithm (Bauer *et al.* 2008) is used to define clusters of similar properties at the feature map. At the final stage, all tomographic grid blocks are classified and assigned to the cluster type with most similar properties, and are mapped back to the tomographic depth section. The clusters can be mapped additionally from the feature map to the V_p–V_p/V_s cross-plot similar to the joint PDF cluster analysis. The advantage of the neural network cluster analysis compared to the earlier described automatic method is that almost all V_p–V_p/V_s pairs are assigned to a specific cluster.