



# Testing five of the simplest upper mantle anisotropic velocity parameterizations using teleseismic S and SKS data from the Billings, Montana PASSCAL array

Huaiyu Yuan,<sup>1,2</sup> Ken Dueker,<sup>1</sup> and Derek L. Schutt<sup>1</sup>

Received 4 April 2007; revised 15 October 2007; accepted 6 December 2007; published XX Month 2008.

[1] Five of the simplest parameterizations of upper mantle anisotropy are tested and ranked for a data set collected on a dense temporary PASSCAL seismic array located 100-km NE of Yellowstone. These hexagonal symmetry anisotropy models possess either one or two layers with either flat or dipping fast velocity axis (FVA). Recordings from fifteen high quality direct-S and SKS arrivals are stacked to provide accurate waveform and error estimates. Source normalization is accomplished using the cross-convolution technique. A direct Monte Carlo Neighborhood Algorithm is used to map the posteriori model probability density (PPD) volume. Using the  $F$  test, we find that models with purely flat FVA can be rejected at >97% probability. Our best model (P5) is a two layer dipping FVA parameterization, albeit the two layer model with one flat and one dipping FVA can only be rejected at 80% probability. The best model has a lower layer with a N65°E FVA strike and a  $-12^\circ$  dip (down to the southwest), and an upper layer with a N20°W FVA strike and a  $-47^\circ$  FVA dip (down to the southeast). The bottom asthenospheric layer FVA strikes parallel to North America's absolute plate motion direction and dips opposite to what passive plate shear of the asthenosphere would predict. The upper lithospheric layer is consistent with LPO accretion associated with north directed drift of the North American plate during the Mesozoic. Comparison between the SKS-and direct S-wave data sets shows that the direct S waves improve resolution of the double layer anisotropic model parameters.

**Citation:** Yuan, H., K. Dueker, and D. L. Schutt (2008), Testing five of the simplest upper mantle anisotropic velocity parameterizations using teleseismic S and SKS data from the Billings, Montana PASSCAL array, *J. Geophys. Res.*, 113, XXXXXX, doi:10.1029/2007JB005092.

## 1. Introduction

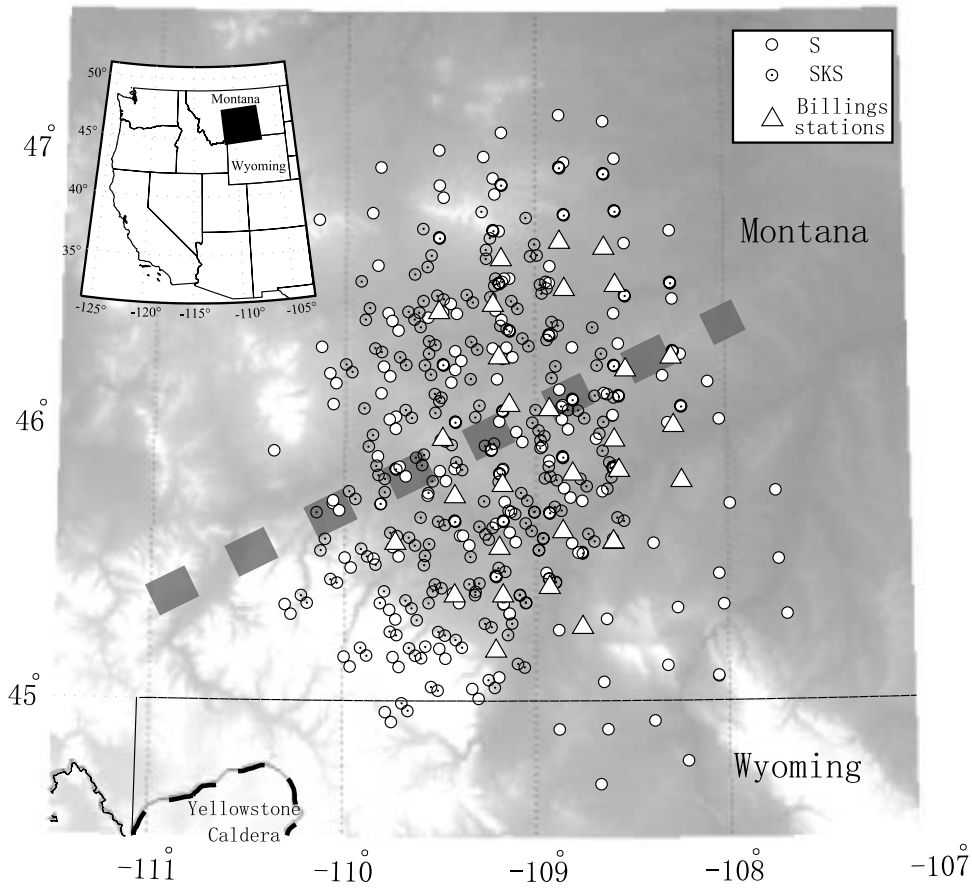
[2] Resolving anisotropic velocity variations between the crust, mantle lithosphere, and asthenosphere is important to constrain a variety of processes that shape our Planet's evolution: e.g., lower crustal flow [Bank and Bostock, 2003; Sherrington *et al.*, 2004], lithospheric accretion [Babuska and Plomerova, 1993; Fox and Sheehan, 2005], slab-induced flow [Fischer *et al.*, 2000; Anglin and Fouch, 2005], and plume-lithosphere interaction [Rumpker and Silver, 2000; Behn *et al.*, 2004]. Yet, issues remain that strongly limit the constraints provided by shear wave anisotropy imaging that utilizes shear wave birefringence [Fischer *et al.*, 2005; Fouch and Rondenay, 2006]: e.g., the non-linearity of the data kernels makes model appraisal difficult, the lack of available high-fold data sets from seismic arrays makes accurate isolation of anisotropic signals from signal generated noise difficult, and olivine

LPO symmetry development in the crust, mantle lithosphere and upper mantle can be complex [e.g., Jung and Karato, 2001; Kaminski and Ribe, 2002; Holtzman *et al.*, 2003; Katayama *et al.*, 2004].

[3] Many shear wave splitting methods use an asymptotic low frequency ray theoretical normal-incidence approximation [Vinnik *et al.*, 1989; Silver and Chan, 1991; Gledhill and Gubbins, 1996; Savage, 1999; Davis, 2003] that extracts apparent splitting parameters (ASP)—either the fast or slow velocity axis strike and splitting time—via a grid search. In choosing whether to use a fast or slow velocity axis parameterization, the fast axis parameterization has been chosen because pyrolytic composition upper mantle typically displays fast axis LPO. The ASP uncertainties are approximated by fitting a 2-D Gaussian function, centered on the model misfit minimum. If the ASP is invariant with respect to shear wave polarization, then a single layer of anisotropy with a flat FVA is the simplest and presumably best model. However, if multiple anisotropic layers with flat FVA [Silver and Savage, 1994; Ozalaybey and Savage, 1994; Levin *et al.*, 1999] or a single layer of dipping FVA exist [Hartog and Schwartz, 2000; Levin *et al.*, 2002], then the ASP are modulated by lower order sinusoidal terms. With this methodology, the analysis devolves into assessing the statistical fits through the ASP error estimates.

<sup>1</sup>Department of Geology and Geophysics, University of Wyoming, USA.

<sup>2</sup>Now at Berkeley Seismological Lab, UC Berkeley, Berkeley, California, USA.



**Figure 1.** Billings array topography, seismic broadband station locations, and shear wave raypath piercing points at 150 km depth. The SKS and direct-S wave piercing points are shaded as shown in the legend. The gray dashed line denotes the strike of the Archean-aged Madison Mylonite shear zone exposed in the Mountains just north of the Yellowstone Caldera.

[4] In this study, a new method is developed to discriminate between multiple layer models of upper mantle anisotropy. We use the waveform data set from the high density Billings array (Figure 1) to test and statistically rank five of the simplest upper mantle anisotropy models. As a sub-project of the Continental Dynamics Yellowstone array project, the Billings array was designed to record teleseismic data in a region 100 km NE of the current location of the Yellowstone hot spot located within Yellowstone Park. The relatively small 150 km by 110 km aperture of the array permits coherent stacking of teleseismic shear wave signals from all 30 broadband stations to accurately isolate anisotropic S-wave signals with robust error estimates. The stacking process also attenuates the signal generated noise and small variations in the upper mantle anisotropy within the Billings array. To source normalize the different events, the cross-convolution technique is used [Menke and Levin, 2003]. A direct Monte-Carlo search method, the Neighborhood Algorithm (NA) [Sambridge, 1999a, 1999b], is used to map the posterior model probability density volumes (PPD) from which maximum likelihood models may be found. This permits the 1- and 2-D marginal distribution functions to be assessed quantitatively. Model robustness is tested using synthetic data sets, geo-

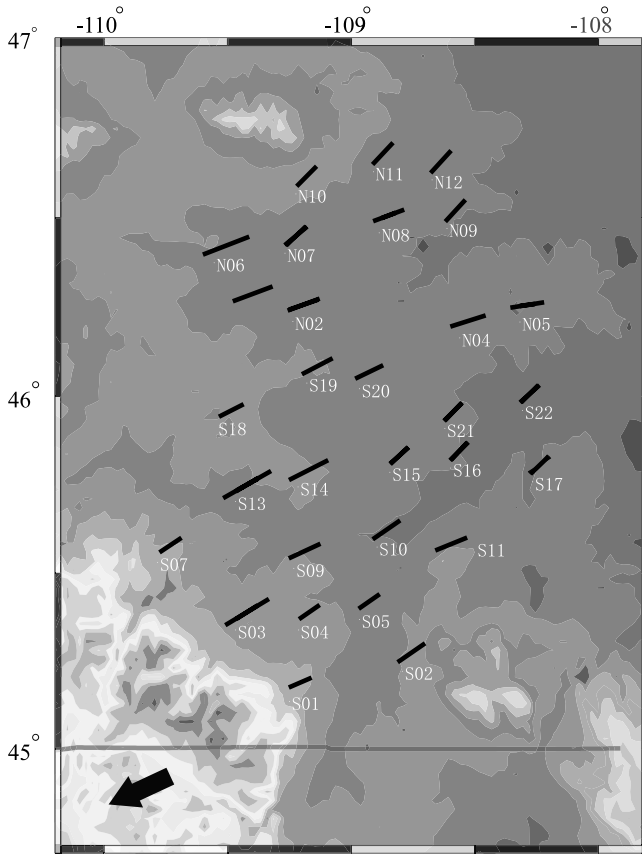
graphic sub-setting of our data set, and comparison of models using the direct-S and SKS data subsets.

[5] Shear wave splitting measurements using Silver and Chan [1988] eigenvector analysis show a nearly uniform FVA strike ( $51^\circ$ ) and split time (0.8 s) across the array (Figure 2). Some systematic variations in the FVA orientation at the northern end of the array are found, but the maximum deviation from the mean FVA is only  $12^\circ$ . On the basis of the relative uniformity of the shear wave splitting analysis, our further analysis of these ten SKS/SKKS waves and an additional five direct S-waves is performed by stacking the entire array to one stack waveform pair for each of our 15 events.

## 2. Data and Methods

### 2.1. Shear Wave Data Windowing and Polarization

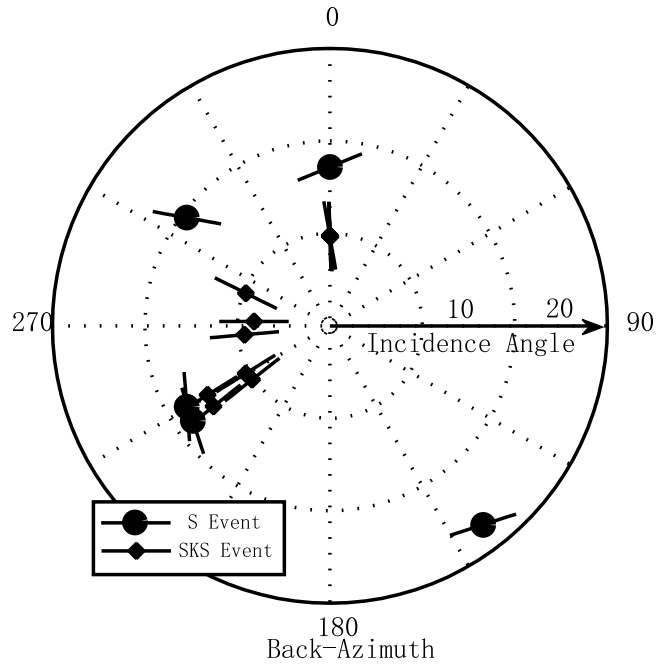
[6] Our teleseismic data set consists of ten SKS/SKKS and five direct-S arrivals (Table 1) that are reasonably well distributed with respect to the events back-azimuth, incidence angle, and wave polarization direction (Figure 3). The addition of the direct-S events significantly increases the diversity of incidence angle and polarization sampling (Figure 3). The increased incidence angle of the direct-S waves is particularly useful to improve the resolution of



**Figure 2.** Billings array shear wave splitting measurements using eigenvector analysis [Silver and Chan, 1988] for the SKS/SKKS events in Table 1. The line segment lengths are proportional to the split time and the line segments are parallel to the fast velocity axis. The peak-to-peak split time is 0.8 s. The thick black arrow shows the absolute plate motion direction of the North America Plate [Gripp and Gordon, 2002].

dipping FVA axis [Chevrot and van der Hilst, 2003]. The waveforms are zero-phase band-pass filtered with a two-pole Butterworth filter with corner frequencies of 0.015 and 0.15 Hz.

[7] To estimate the dominant polarization of the direct-S and SKS waves, a principle component analysis of the 30

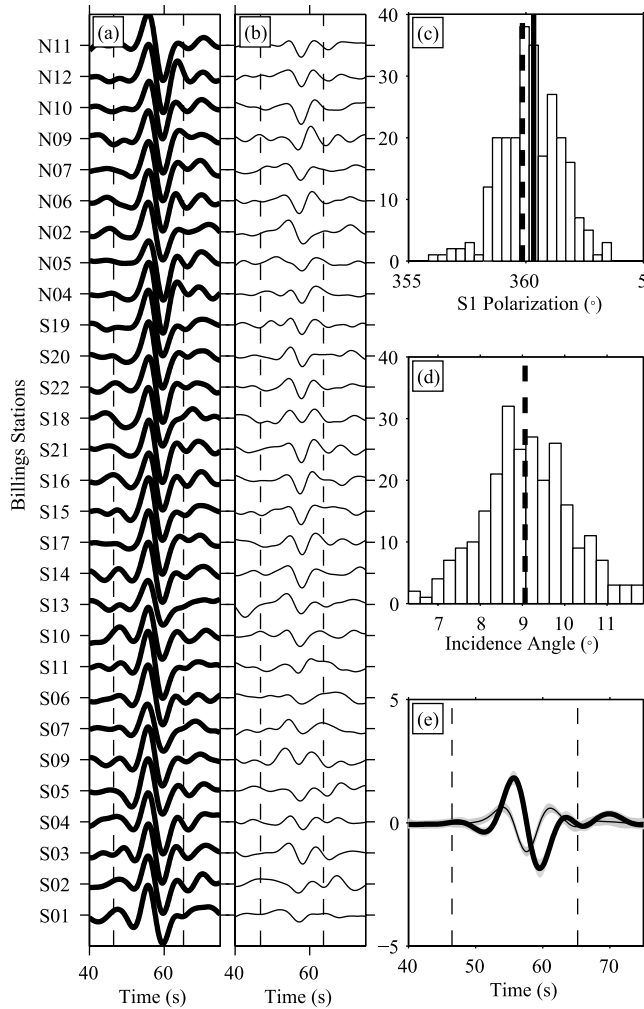


**Figure 3.** SKS and direct-S wave event distribution with respect to incidence angle (at 150 km depth) and back-azimuth. The orientation of the line segments for each event denotes the polarization of the principle component (S1) of the wave. For the SKS events this observed polarization is within 3° of the theoretical back-azimuth (radially polarized). Most of the direct S-events are dominantly SH polarized.

stations recording each event is used. The wavefield parameters constrained via the principle component analysis are: the back-azimuth, incidence angle, and the dominant polarization direction. The dominant polarization direction is denoted as S1 with the secondary polarization direction S2 defined as orthogonal to S1. Principle component parameter uncertainties are estimated by bootstrap resampling [Efron and Tibshirani, 1986] of the 28–30 three-component stations that recorded each event (see an example in Figure 4). The measured polarization of the S1 component of the SKS waves is found to have a mean value within 3° of the theoretically radially polarized SKS-

**Table 1.** Event Table

Event BAZ, (°)	Event Year-Day	Phase	Latitude, (°)	Longitude, (°)	Depth, km	Polarization, (°)	Slowness, s/km
0.3	2000-133	SKS	35.98	70.66	108	357	0.046
141.4	1999-325	S	-21.75	-68.78	101	72	0.097
235.3	2000-228	SKS	-31.51	179.73	358	231	0.064
235.3	2000-228	SKKS	-31.51	179.73	358	232	0.043
235.3	2000-228	S	-31.5	179.73	358	162	0.075
240.5	2000-166	SKS	-25.52	178.05	605	238	0.045
240.5	2000-166	SKKS	-25.52	178.05	605	238	0.065
240.5	2000-166	S	-25.52	178.05	605	355	0.075
264.1	2000-008	SKS	-9.81	159.81	33	265	0.046
273.3	2000-037	SKS	-5.84	150.88	33	270	0.043
291.2	2000-057	SKS	13.8	144.78	132	296	0.052
307.1	2000-161	S	30.49	137.73	307	280	0.089
359.9	2000-199	SKS	36.28	70.82	141	350	0.046
359.9	1999-312	SKS	36.52	71.24	228	358	0.046
359.9	1999-312	S	36.52	71.24	228	248	0.076

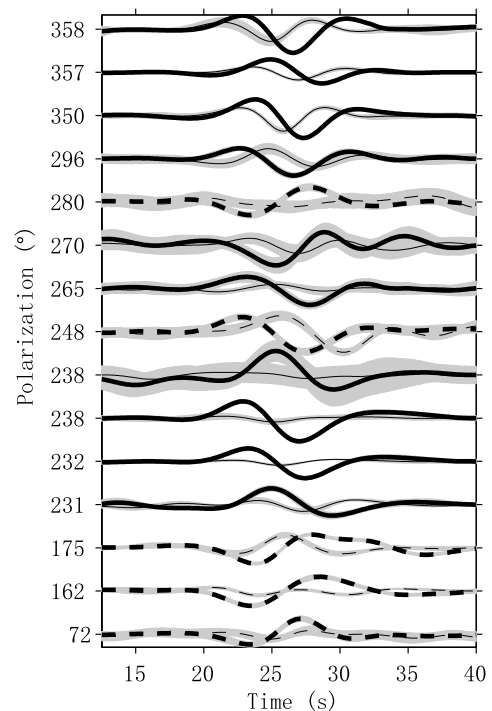


**Figure 4.** SKS event 2000-133 (Table 1) recording and polarization estimation. (a) Traveltime residual adjusted S1 component waveforms. (b) Residual adjusted S2 component waveforms. (c) S1 polarization direction histogram derived from all recording stations. (d) Incidence angle histogram. (e) The S1 and S2 stack traces with one-sigma error bar shaded in light gray. The thin dashed lines in (a), (b) and (e) define the signal measurement window. The polarization measurements are derived from 200 bootstraps of the principle component analysis for all the stations. The thick black line in (c) is the theoretical event back-azimuth and the dashed line is the mean estimate of the S1 polarization direction. For the SKS/SKKS arrival, the event back-azimuth is very close to the S1 direction (Table 1). The dashed line in (d) is the mean estimate of the incidence angle from the histogram. Note that the S1 component has high signal coherence while the split energy present on the S2 components is less coherent. The relative incoherence of the S2 component is attributed to signal generated noise created by the crust and overlying sedimentary layer.

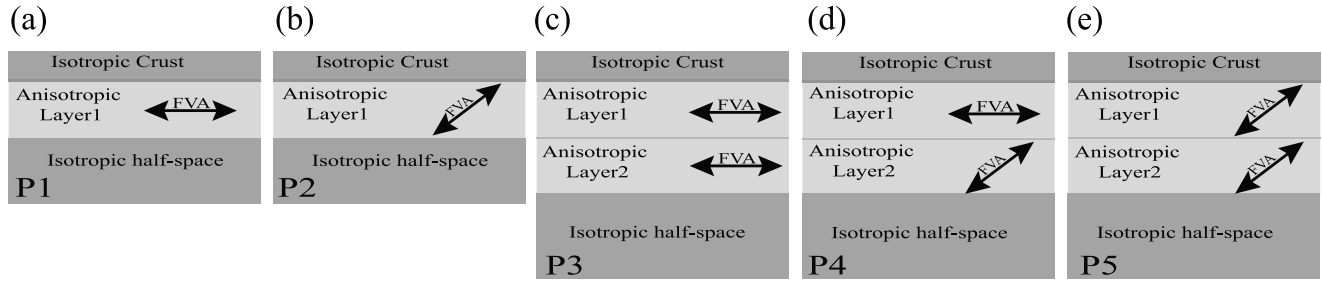
waves (Table 1). Noteworthy is that for dipping FVA anisotropy, the polarization directions of the fast and slow wave components are no longer orthogonal and a systematic angular deviation (polarization anomaly) of the transverse wave is predicted [e.g., *Chevrot and van der Hilst, 2003*].

This polarization anomaly increases as the angle between the FVA and the raypath departs from perpendicularity. For the mean  $19^\circ$  incidence angle of our direct S-rays at 150 km depth (Figure 3), and the  $12^\circ$ – $28^\circ$  FVA dip found by model parameterization P4 and P5 (described in section 2.4) for the bottom layer, the transverse wave polarization anomaly is predicted to be  $4^\circ$ – $8^\circ$ . Thus this is a small error that we have ignored.

[8] The waveforms are visually windowed to minimize the influence of secondary converted and reverberated phases from the crust-mantle boundary. For each individual event, a common window length is found that starts at the shear wave signal onset to avoid S-P precursors and the length of the waveform segment is 10–20 s depending on the magnitude of the earthquake. This window length excludes all free surface Moho reflection except the  $Sp_m$ s phase which is predicted to arrive 16–19 s after the S-wave arrival. This arrival time is estimated based on combined receiver function [*Yuan et al., 2006*] and diffusive/ballistic surface wave imaging beneath this array (*Stachnik et al., in prep.*). To account for relative traveltimes variations associated with small isotropic velocity variations beneath the array, the multichannel cross-correlation method [MC, *vanDecar and Crosson, 1990*] is used to measure S-wave arrival times. Relative traveltimes are measured on the S1 component, and the maximum relative traveltimes variation for the S1 component is 1.6 s. After applying the MC calculated time-shifts to the S1 and S2 components recorded



**Figure 5.** Direct-S and SKS waveform pairs. The five direct-S waveform pairs are shown as dotted lines and the ten SKS waveform pairs as solid lines. The thick and thin lines are the S1 and S2 components, respectively. The shaded region about each trace is one standard deviation estimated from stacking all the stations recording each event.



**Figure 6.** Five simple anisotropy models. The P1 and P2 parameterizations are single layer anisotropy models with P1 having a flat FVA and P2 having a potentially dipping FVA. Model P3 is a two layer model with a flat FVA in both layers. Model P4 has the upper layer with flat FVA and a potentially dipping FVA in the lower layer. Model P5 has a potentially dipping FVA in both layers. Note all anisotropic layers are fixed to be 100 km thick. The crustal thickness is fixed to the array mean of 52 km (Yuan *et al.*, in prep.).

by each event, the waveforms are linearly stacked to obtain 15 event stack S1-S2 waveform-pairs that are subsequently modeled (Figure 5).

[9] Strategies to minimize potential source-side anisotropy signal in direct S-arrivals include: explicit correction for known source-side anisotropy [e.g., Yang and Fischer, 1994; Anglin and Fouch, 2005]; or use of deep-focus events [Savage *et al.*, 1990; Fischer and Yang, 1994; Hartog and Schwartz, 2000; Long and van der Hilst, 2005]. For this study, a total of 12 direct-S events with depths >100 km are initially selected for analysis. To cull these direct-S arrivals, the waveforms are source normalized (see below) with our best one layer anisotropy model found using the SKS-only data set. Then, event stack waveforms with large misfits are assumed to have large source-side anisotropy contamination and discarded. This culling produces a final data set of five direct-S stack waveforms with source depths >228 km except one event at 101 km depth (Table 1).

## 2.2. Source Normalization and Synthetic Seismograms

[10] Our 15 event waveform pairs all have high SNR and small error estimates; therefore we first normalize the amplitudes of the waveform pairs by the sum of squared amplitudes of S1 and S2 components. This ensures that every event waveform pair will have equal weight in the minimization of the cost function. Given that each event has a different source time function, we source normalize our waveform pairs using the cross-convolution method [Menke and Levin, 2003; Levin *et al.*, 2006]. The method is predicated on the reasonable assumption that the source wavelet is well described as a linearly polarized wave before it splits into two approximately orthogonal waves during transit through an anisotropic medium. An approximate cross-convolved relation between the observed S1 and S2 polarized S-waves (Figure 5) and synthetic s1 and s2 waveforms is

$$S1*s2(\mathbf{m}) \approx S2*s1(\mathbf{m}) \quad (1)$$

where  $\mathbf{m}$  is the synthetic anisotropic velocity model and \* indicates convolution. This equation simply states that when  $\mathbf{m}$  is near the true model, the convolution of the observed S1 trace with its orthogonal synthetic is approximately equal to the convolution of the observed S2 trace with its orthogonal synthetic.

[11] Following Frederiksen *et al.* [2003], we assess the misfit associated with equation (1) using cross correlation and the  $L^1$  and  $L^2$  norms. We find that these different misfit norms produce very similar results. The  $L^2$  norm is used in this research because this norm can be used as input to the  $F$  test to compare the significance of the variance reductions between model parameterizations that possess different number of degrees of freedom. Search for the best models are quantified by minimizing the  $L^2$  misfit of the cross-convolved traces,

$$\min \left\{ [s2(\mathbf{m}) * S1 - S2 * s1(\mathbf{m})]^2 \right\}. \quad (2)$$

[12] The synthetic seismograms are calculated using ray theory for a transverse anisotropic medium [Frederiksen and Bostock, 2000]. The 3-D traveltime equation of Diebold [1987] is adopted in this method, and the phase amplitudes are computed using the reflection and transmission matrices for planar interfaces separating homogeneous anisotropic media. This ray-based approach allows us to selectively model the first-order free-surface Moho reverberations and to assess the influence of these small secondary arrivals upon our model fits. The anisotropy is parameterized as the percent variation in P- and S-velocity perturbation ( $dV_p$  and  $dV_s$ ) and  $\eta$ , in the manner described by Farra *et al.* [1991]. In this study,  $dV_p$  is set equal to  $dV_s$  and  $\eta$  is set to 1.03 where  $\eta$  is the anisotropic parameter that controls the velocity variation for directions that are not parallel or perpendicular to the symmetry axis (it creates small dimples on the prolate spheroid velocity representation). Thus only one anisotropic parameter, the percent variation of velocity

**Table 2.** Synthetic Test Models

Models	Layer 1			Layer 2		
	Anis., (%) <sup>a</sup>	FVA Strike, °	FVA Dip, °	Anis., (%) <sup>a</sup>	FVA Strike, °	FVA Dip, °
SD1	4	65	-	-	-	-
SD3	4	-5	-	4	70	-
SD5	4	-20	-60	6	60	-10

<sup>a</sup>For 100-km thick layer. The FVA strike is rendered positive clockwise with respect to North. The dip is positive downwards with respect to horizontal along the FVA strike direction. Negative dip means it dips towards 180° of the strike.

(referred to as anisotropy magnitude hereinafter) is modeled. The  $S1$  and  $S2$  polarization responses are calculated via rotation of the  $SV$  and  $SH$  impulse responses. This approximation assumes that the  $SV$  and  $SH$  raypaths are the same which is a good weak-anisotropy approximation [e.g., *Crampin, 1977; Rumpker and Silver, 2000*].

### 2.3. Neighborhood Algorithm Sampling and Appraisal

[13] The Neighborhood Algorithm (NA) [*Sambridge, 1999a, 1999b*] is used to map the posterior model probability density volume (PPD) associated with our five anisotropic velocity model parameterizations. NA is a direct Monte Carlo search method which has been applied to many non-linear inversion problems. Readers are referred to *Sambridge [1999a, 1999b]* for the details of the method. The NA methodology consists of two stages. First, a search stage is guided by a random “walk” through the nearest neighbor Voronoi cells that discretize the model parameter volume. Two controlling parameters,  $n_s$ , the number of models generated in each iteration, and  $n_r$ , the number of lowest misfit models in which the random walks are performed, control the spatial uniformity of the model search. These search parameters have been both set to 5, 10, and 100, and we find they all can provide an extensive search of the model space so that the global minimum is not missed [*Sambridge, 2001*]. The total number of models generated by the NA is between  $10^3$ – $10^6$  depending on the dimensionality of the model space which varies between 2 and 6.

[14] The reduced chi-square misfit between the cross convolved waveforms is defined as:

$$\chi_v^2(\mathbf{m}) = \frac{1}{dof} \sum_{i=1}^N \frac{(S1^{i*}s2(\mathbf{m})^i - S2^{i*}s1(\mathbf{m})^i)^2}{(\sigma_x^i)^2 + 2\sigma_x^i \times \sigma_y^i + (\sigma_y^i)^2}, \quad (3)$$

where  $i$  indexes the events number,  $N$  is number of events,  $dof$  is the number of degrees of freedom, the star operator is convolution,  $S1$ ,  $S2$  and  $s1$ ,  $s2$  are the stacked waveforms and synthetics, respectively, and a sum over the time series is implicitly assumed. The standard deviation of the two cross-convolved traces is defined as

$$\sigma_x^i = s1(\mathbf{m}) * \sigma_{s1}^i \text{ and } \sigma_y^i = s2(\mathbf{m}) * \sigma_{s2}^i, \quad (4)$$

where  $\sigma_{s1}^i$  and  $\sigma_{s2}^i$  are the standard errors of the  $S1$  and  $S2$  components. The errors for the two cross-convolved terms are considered perfectly correlated and hence the error cross term in the denominator of equation (3) is present. The  $dof$  for each event waveform pair is calculated using the spectral bandwidth of the data [*Silver and Chan, 1991*]. This

analysis shows that one dof requires  $\sim 0.8$  s of S-wave signal. Thus for 15 events with an average signal length of 10 s, our data set contains 120 dof.

[15] In the second stage of the NA modeling, quantitative model information is extracted using the Bayesian approach [*Sambridge, 1999b*]. In this approach, the information contained in the model ensemble is represented by the posterior probability density functions (PPD). The PPD functions are then used to calculate quantities such as the model expectation (mean) and the 1- and 2-d posterior marginal probability density functions. Assuming a multivariate Gaussian probability function with a uniform prior probability distribution  $\rho(\mathbf{m})$ , the PPD is given by

$$P(\mathbf{m}) = k\rho(\mathbf{m})\exp\left(-\frac{v}{2}\chi_v^2(\mathbf{m})\right) \quad (5)$$

where  $k$  is a constant that normalizes the total probability to unity,  $v$  is  $dof$ , and  $\chi_v^2$  is the reduced chi-square value (equation (3)). N-dimensional Bayesian integrals are used to calculate the 1- and 2-D PPD marginals (note: we use the short-hand term “marginals” to refer to the set of 1- and 2-D marginal PPD herein). To improve the accuracy of the N-dimensional integrations, a Gibbsian resampling of the probability space is used [*Sambridge, 1999b*].

### 2.4. Five Model Parameterizations

[16] The main goal of this paper is to test our 15 SKS- and direct S-wave waveform pairs against five of the simplest models of upper mantle anisotropy. Given that our error estimates are well-characterized by the standard deviation statistic, the  $F$  test is used to rank the significance of the relative variance reductions between the five different model parameterizations. For our anisotropic velocity model parameterizations, we assume a hexagonal anisotropy system with a fast velocity axis, which is a widely acceptable representation for olivine lattice preferred orientation in the upper mantle [e.g., *Park and Levin, 2002*]. The five anisotropic model parameterizations are denoted P1–P5 (Figure 6) and can be described as follows: P1 and P2 are single layer models with either a flat or potentially dipping FVA; P3 is a double layer model with a flat FVA in both layers; P4 has an upper flat FVA layer and lower potentially dipping FVA layer; P5 has two layers with potentially dipping FVA.

[17] The thickness of each layer is fixed to 100 km. The layer thickness is fixed because our data set has very limited sensitivity to the trade-off between layer thickness and anisotropy magnitude. Therefore the anisotropy magnitude found for each layer is only meaningful if one believes that the 100 km thick layers are close to their true values. We

**Figure 7.** Synthetic model PPD marginals constructed using input parameters in Table 2. (a) SD1 data set. (b) SD3 data set. (c) SD5 data set. Each subplot of the PPD marginals is labeled by the model parameterization (P1–P5). The synthetic model parameter input mimics the real data inversion results (shown in section 3.2). The diagonal of each lower-half “matrix” subplot shows the 1-D PPD marginals for the model parameters. The 1-D PPD marginals are found by integrating over the other model parameters and hence properly accounts for correlated model errors. The lower “matrix” shows the 2-D PPD marginals, which show the correlation of any pair of the model parameters. For the 2-D marginals, the darker shading corresponds to higher probabilities and a contour line outlines the 80% probability contour. The model parameter names and values are labeled along the bottom and left side of the matrices. The strike of the FVA is rendered in degrees positive clockwise with respect to North.

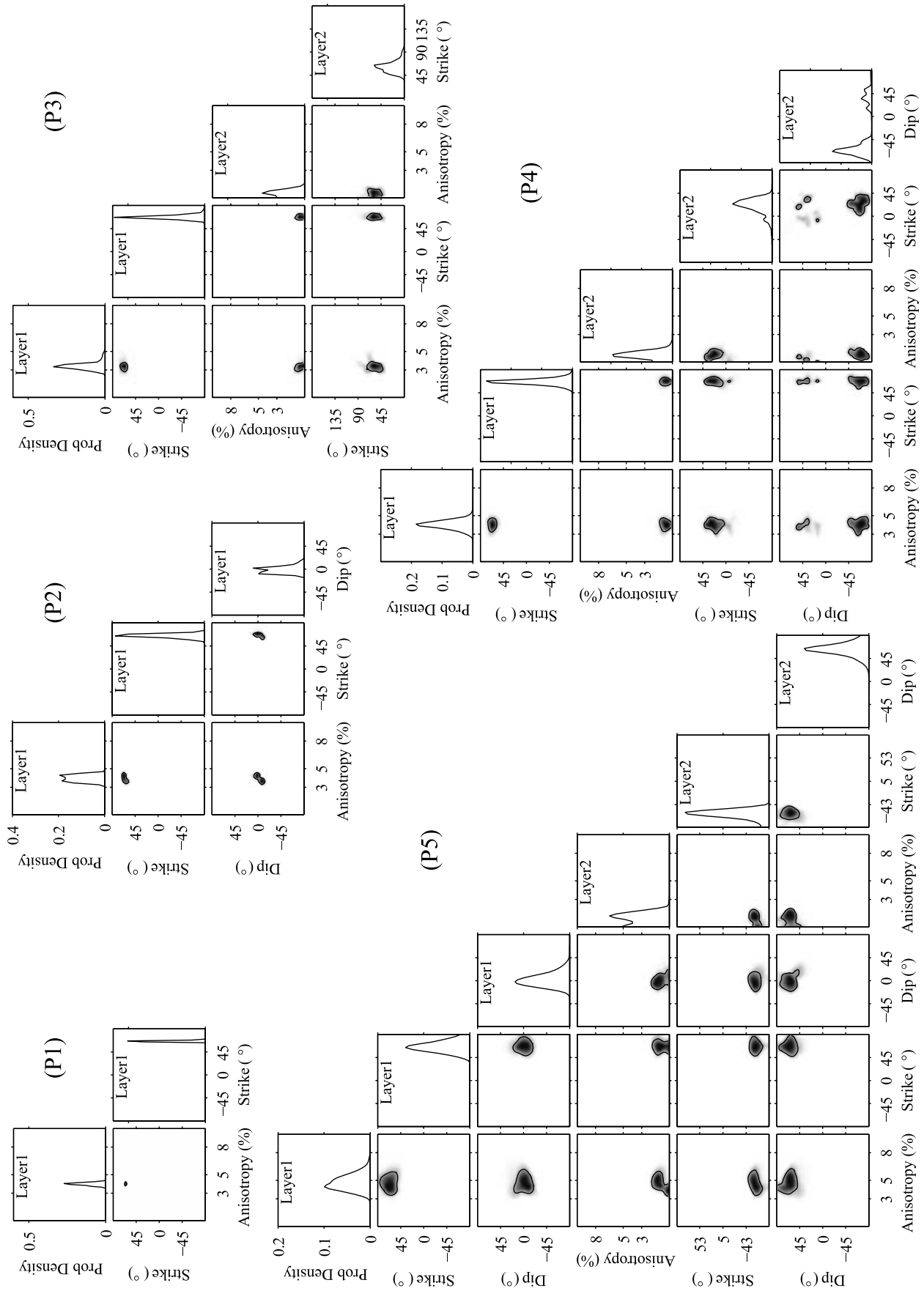


Figure 7

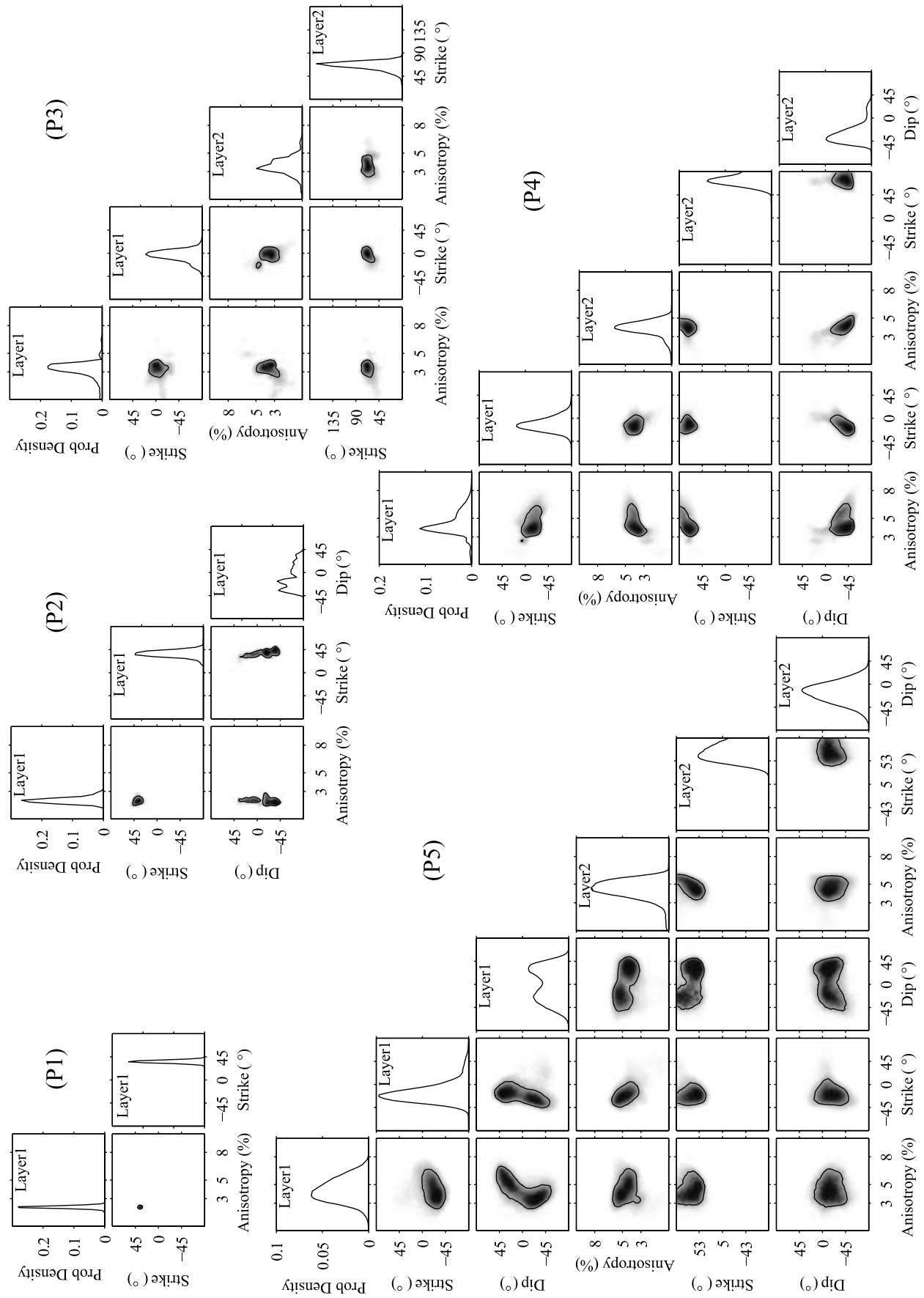


Figure 7. (continued)



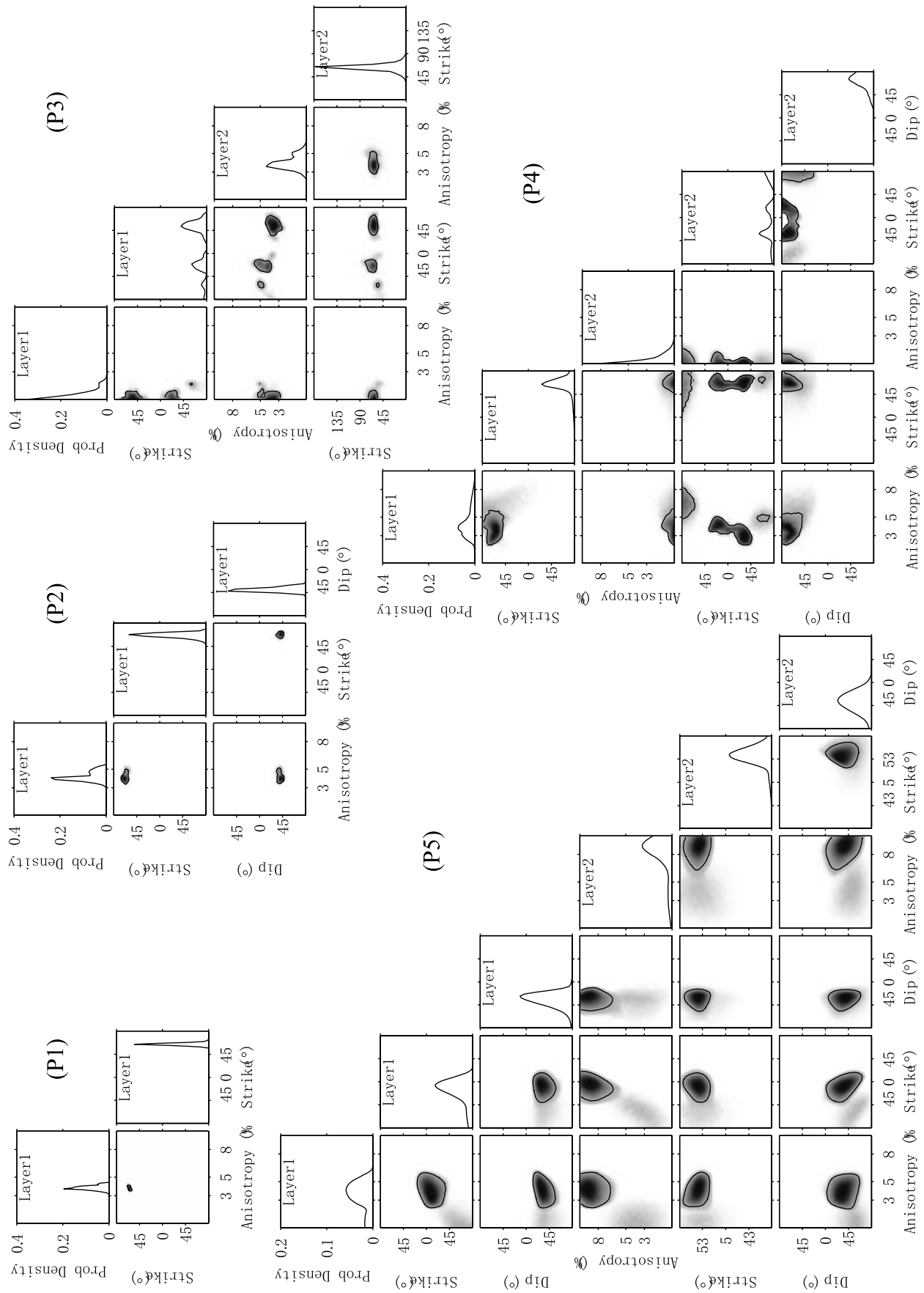


Figure 7. (continued)

**Table 3.** Synthetic Test: Results of SD1

	Layer 1						Layer 2								
	Anis., (%) <sup>a</sup>		FVA Strike, (°)		FVA Dip, (°)		Anis., (%) <sup>a</sup>		FVA Strike, (°)		FVA Dip, (°)		DOF	$\chi_v^2$	RP <sup>b</sup>
P1	-4.0	[3.9, 4.2]	66	[65, 67]	-	-	-	-	-	-	-	-	78	1.50	-
P2	4.3	[3.6, 4.4]	65	[62, 68]	1	[-10, 3]	-	-	-	-	-	-	77	1.51	52.2%
P3	3.3	[3.1, 3.7]	67	[64, 70]	-	-	0.5	[0.2, 0.8]	65	[49, 68]	-	-	76	1.53	53.8%
P4	4.1	[3.6, 4.5]	67	[62, 72]	-	-	0.8	[0.4, 1.6]	24	[23, 38]	-68	[-81, 37]	75	1.55	56.6%
P5	4.3	[3.8, 5.5]	65	[55, 76]	-2	[-13, 11]	1.2	[0.0, 1.6]	-59	[-68, -52]	64	[52, 75]	74	1.57	58.4%

<sup>a</sup>For 100-km thick layer.<sup>b</sup>Rejection probability. Brackets show the 80% probability region. P 1 is the correctly parameterized model.

note that mantle xenoliths from eastern Montana suggest a 120–140 km thick lithosphere [Carlson *et al.*, 2004]. Therefore given the 48–54 km thickness of the crust beneath the Billings array [Yuan *et al.*, 2006] at least 70–90 km of lithospheric mantle is present based on the xenolith studies. If the anisotropic layers beneath the array have sharp anisotropic velocity contrasts, then modeling of the S-P precursors could provide useful information to constrain the trade-off between layer-thickness and anisotropy magnitude.

[18] The crustal thickness and shear velocity has been constrained in our model parameterization using a combination of teleseismic P- and S-wave traveltimes [Yuan and Dueker, 2005], P wave receiver function analysis [Yuan *et al.*, 2006], ballistic Rayleigh wave imaging [Schutt *et al.*, 2008] and combined diffusive and ballistic Rayleigh wave imaging [Stachnik *et al.*, in prep]. These new seismic constraints show that a relatively uniform crustal thickness of 48–54 km is present beneath the array with an average shear wave velocity of 3.74 km/s and a  $V_p/V_s$  value of 1.78 [Yuan *et al.*, 2006]. These observations are consistent with the Deep Probe refraction line that sampled beneath the Billings array [Henstock *et al.*, 1998; Gorman *et al.*, 2002]. This well-constrained crustal thickness and velocity information means that the relatively long-period  $Sp_{ms}$  free surface reverberation present in our signal time window is modeled.

### 3. Results

#### 3.1. Synthetic Data Tests

[19] Synthetic tests using the same event parameters (back-azimuth, incidence angle, and polarization) as the true data set (Table 1) are conducted to investigate the ability of our data set coverage to constrain anisotropic models. Three synthetic waveform data sets (SD1, SD3, and SD5) are created using the parameters in Table 2. Note SD1

and SD3 correspond to the one- and two-layer models with flat FVA anisotropy and SD5 corresponds to the two-layer model with dipping FVA anisotropy. We choose to test these three models because SD1 and SD3 are most commonly tested anisotropic model parameterizations, while SD5 is our most complex model parameterization. The anisotropy parameters of the synthetic models (Table 2) are chosen to simulate the real-data modeling results (shown in section 3.2). A 4% and 6% anisotropy magnitude is used for the upper and lower anisotropic layers. The synthetic waveforms are generated using *Frederiksen and Bostock* [2000] ray theoretical code. Random white noise filtered to the bandwidth of the S-wave amplitude spectrum is scaled to match the observed data signal-to-noise ratio and added to the synthetic waveforms. The sum of the number of degrees of freedom for each synthetic data set is about 80.

[20] To show how these three synthetic data sets are fit by our five model parameterizations, the PPD marginals for each of the synthetic data sets are calculated. To evaluate these synthetic results, three relations between the synthetic model from which a data set is calculated and the model parameterization used, are defined: 1) a correct-parameterized model where the synthetic data model and the model parameterization are the same; 2) an under-parameterized model where the synthetic data model has higher dimensionality than the model parameterization; 3) an over-parameterized model where the synthetic data model has lower dimensionality than the model parameterization. To assess the significance of variations in the data misfit between each of the synthetic data sets, rejection probabilities are calculated via the F-test which uses the reduced chi-square values of each synthetic model. The rejection probabilities are calculated with respect to the lowest misfit model and a 50% rejection probability means that a particular model is no more probable than the lowest misfit model.

[21] Inspection of the synthetic test results (Figure 7 and Tables 2, 3, 4, 5) reveals the following conclusions. As

**Table 4.** Synthetic Test: Results of SD3

	Layer 1						Layer 2								
	Anis., (%) <sup>a</sup>		FVA Strike, (°)		FVA Dip, (°)		Anis., (%) <sup>a</sup>		FVA Strike, (°)		FVA Dip, (°)		DOF	$\chi_v^2$	RP <sup>b</sup>
P1	2.1	[2.0, 2.2]	35	[33, 38]	-	-	-	-	-	-	-	-	78	2.0	87%
P2	2.0	[1.7, 2.2]	35	[32, 43]	-16	[-41, -9]	-	-	-	-	-	-	77	2.0	89%
P3	3.9	[3.0, 4.0]	-2	[-12, 7]	-	-	3.3	[2.9, 4.4]	69	[61, 75]	-	-	76	1.54	-
P4	3.9	[3.2, 5.3]	-16	[-27, -2]	-	-	3.9	[3.4, 4.7]	72	[63, 83]	-42	[-52, -22]	75	1.56	52%
P5	3.9	[3.0, 5.6]	-20	[-36, -6]	31	[-39, 43]	4.5	[3.8, 5.6]	62	[50, 87]	-13	[-35, 5]	74	1.57	54%

<sup>a</sup>For 100-km thick layer.<sup>b</sup>Rejection probability. Brackets show the 80% probability region. P3 is the correctly parameterized model.

**Table 5.** Synthetic Test: Results of SD5

	Layer 1					Layer 2					DOF	$\chi_v^2$	RP <sup>b</sup>		
	Anis., (%) <sup>a</sup>		FVA Strike, (°)		FVA Dip, (°)	Anis., (%) <sup>a</sup>		FVA Strike, (°)		FVA Dip, (°)					
P1	3.7	[3.6, 4.0]	65	[62, 67]	-	-	-	-	-	-	78	2.3	94%		
P2	4.1	[3.6, 4.7]	68	[63, 71]	-42	[-44, -36]	-	-	-	-	77	1.7	61%		
P3	0.0	[0.0, 0.7]	54	[-31, 65]	-	-	3.6	[3.1, 4.3]	65	[59, 68]	-	76	2.3	94%	
P4	3.9	[2.5, 4.9]	65	[56, 76]	-	-	0.0	[0.0, 0.6]	-31	[-42, 90]	75	[66, 89]	75	2.1	88%
P5	4.1	[2.7, 5.2]	-6	[-27, 8]	-27	[-42, -19]	8.5	[7.5, 10]	61	[46, 77]	-34	[-54, -17]	84	1.6	-

Anis.: Anisotropy.

<sup>a</sup>For 100-km thick layer.

<sup>b</sup>Rejection probability. Brackets show the 80% probability region. P5 is the correctly parameterized model.

expected, the SD1 data set (single layer with flat FVA) can be equally well fit by all five model parameterizations (Figure 7a and Table 3). This equally good model fitting is quantified by the F-test which shows that the rejection probabilities for the over-parameterized models are negligible (52% to 58%). Remarkable is that the over-parameterized models (P2–P5) remained stable and do not produce spurious results such as large anisotropy magnitudes or large FVA dip. The SD3 data set (two-layer model with both flat FVA, Figure 7b and Table 4) is best fit by the correctly parameterized model P3. As expected, the under-parameterized models have high rejection probabilities (87% and 89%) and the over-parameterized models have low rejection probabilities (52% and 54%). The SD5 data set (two layers with dipping FVA) is best fit by the correctly parameterized model P5, and parameterizations P1–P4 can be rejected at 94%, 61%, 94%, and 88% probability (Figure 7c and Table 5).

[22] The synthetic tests reveal that the correctly parameterized models can be recovered by the full data set. In all cases, the under-parameterized models can be rejected with high probability while the over-parameterized models are always acceptable (Tables 2–5). The PPD marginals for the synthetic tests (Figure 7) demonstrate that the double layer model parameterizations have 1- and 2-D PPD marginals that are less compact and in some cases multimodal. This increased uncertainty of the model parameters is due to the fact that two layer models with dipping FVA produce greater waveform complexity with respect to single layer models [e.g., *Ozalaybey and Savage, 1994*]. In practice, this data kernel complexity produces non-uniqueness of the model space: i.e., there are a range of models that produce nearly identical waveforms within the error bars of our observed waveforms. The conclusions from our synthetic tests are that: 1) the input synthetic models for the noisy synthetics can be recovered without model parameter instabilities, and 2) the under-parameterized models can be rejected in most cases using the  $F$  test.

### 3.2. Real-Data PPD Results

[23] NA mapping of the PPD marginals with the combined direct-S and SKS data set is performed for all five model parameterizations (P1–P5). Inspection of the PPD marginals for the different parameterizations shows that the

marginals are generally compact and uni-modal, indicating good resolution of the model parameters with little model parameter correlation (Figure 8). Most notable is that the 1- and 2-D marginals widen for the more complex models consistent with the increasing non-linearity and non-uniqueness associated with the more complex anisotropic models.

[24] Ranking of the models is done using the  $F$  test whose utility depends on the accuracy of our waveform error estimates (Figure 5). Inspection of the chi-square values and associated  $F$  test based model rejection probability levels (Figure 9) shows that the P5 model provides the best  $L^2$  fit to the cross-convolved waveforms (Figure 10). The reduced chi-square value of this model is 1.9, indicating a slight under-fitting of the waveforms (Table 6). This under-fitting of the data is expected given the approximations in our forward problem, for instance source-side contamination of the direct S-wave data set (addressed below). The P1–P3 model parameterizations have larger reduced chi-square values and these models can be rejected at 98%, 89%, and 97% probability levels. Given that the P4 model can only be rejected at 80% probability, we consider this model to be acceptable. Note that the P4 and P5 model parameter uncertainty bounds do overlap.

[25] Model parameter uncertainties are estimated by calculating the 80% probability bounds from the 1-D PPD marginals (Figure 8). Noteworthy is that these 1-D PPD marginals fully account for correlation between the model parameters. Inspection of Table 6 shows that commonalities exist between the different models. The most consistent result is the existence of a layer with its FVA directed toward the ENE (i.e.,  $N64^\circ$ – $N75^\circ$ ). This direction is  $19^\circ$  larger than the mean FVA of  $51^\circ$  found by our SKS eigenvector analysis (Figure 2), but is consistent with the absolute plate motion direction [*Gripp and Gordon, 2002*]. We suggest this FVA difference results from the biasing effects of the more complex anisotropy models found by our waveform analysis upon the eigenvector SKS analysis. For the models that permit a dipping FVA, the bottom layer FVA consistently dips at  $-12^\circ$  to  $-28^\circ$  (down to the SW) with respect to horizontal. The anisotropy magnitude range for the upper 100 km thick layer is 2.9–3.6% and the 100 km thick bottom layer has a larger anisotropic magnitude range of 5.9–7.5%.

**Figure 8.** Real-data 1- and 2-D PPD marginals. See Figure 7 for figure layout description. Each subplot of the PPD marginals is labeled by the model parameterization (P1–P5). The model values and uncertainty bounds for the P1–P5 model parameterizations are listed in Table 6.

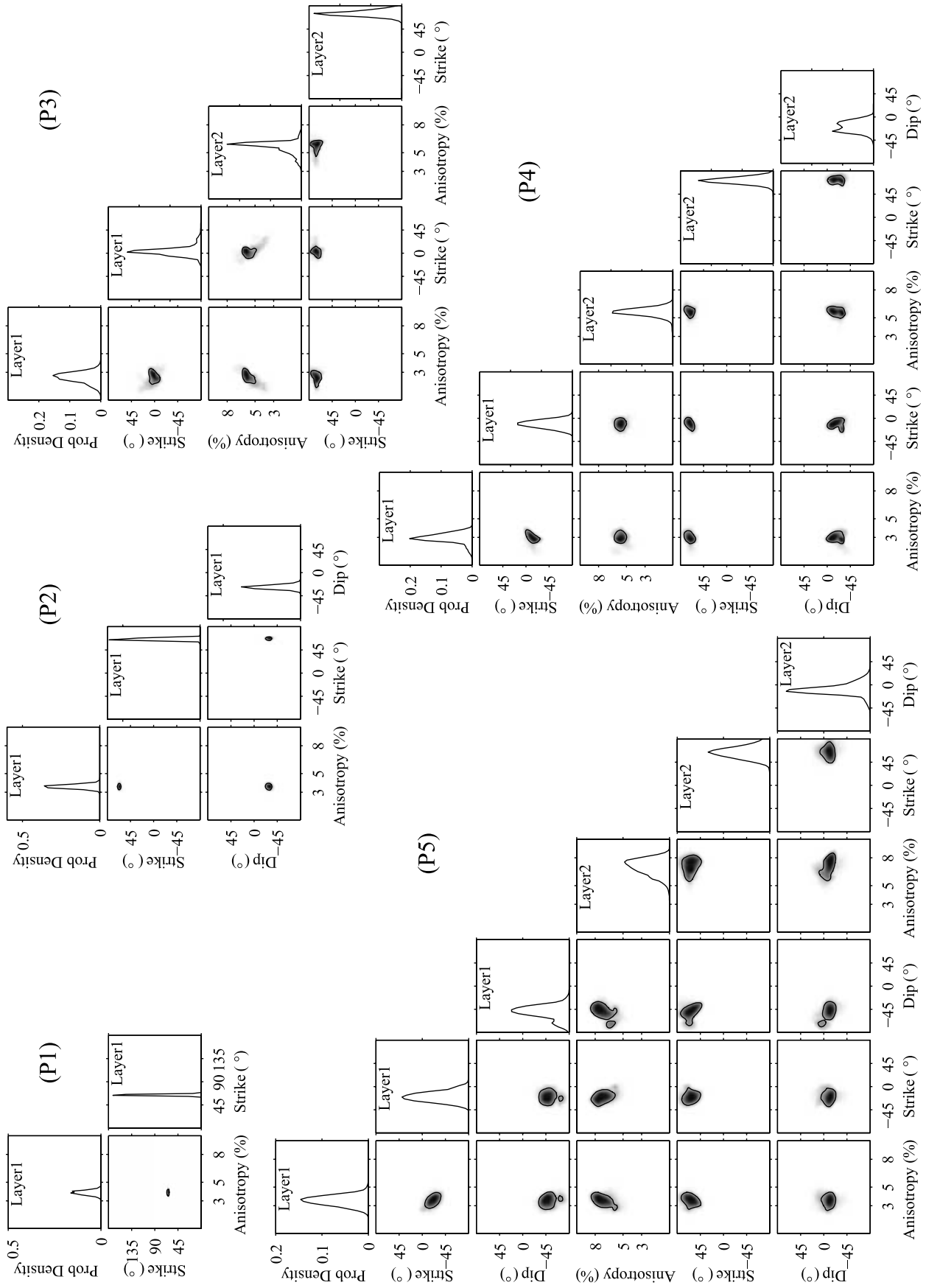
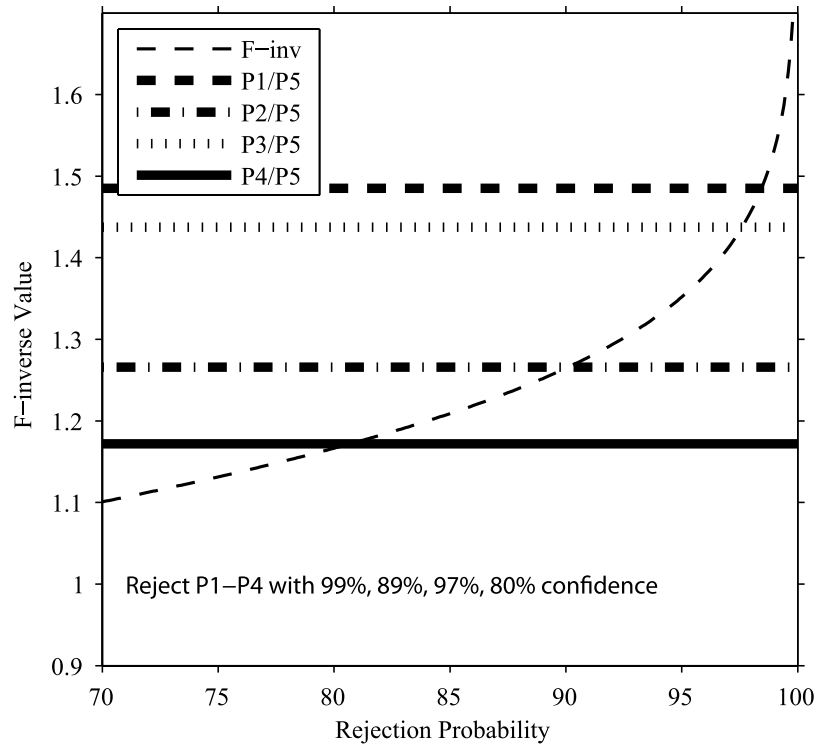


Figure 8



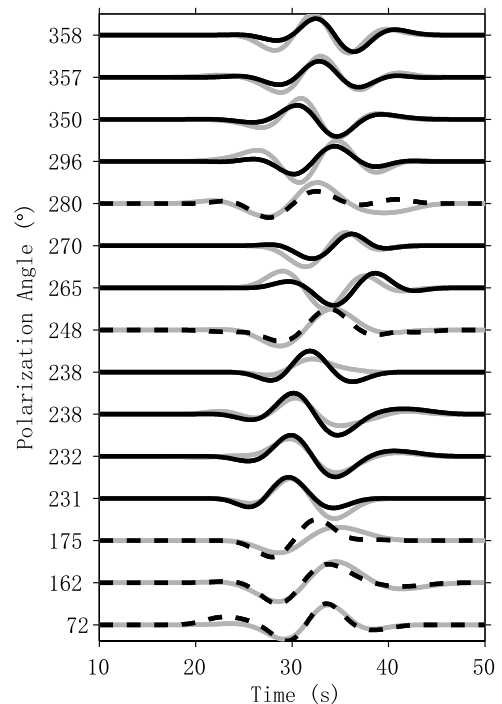
**Figure 9.**  $F$  test results showing the model rejection probabilities with respect to the lowest misfit model P5. The thin dashed line plotted is the  $F$ -inverse curve using the average number of degrees of freedom of the five parameterizations. The two flat FVA model (P1 and P3) can be rejected at  $>97\%$  probability. Because the  $F$  test is a comparison of two models misfit variances, a rejection probability of 50% mean that a model is not a statistical improvement with respect to the lowest misfit model.

[26] For our lowest misfit model parameterization (P5), the upper layer has an anisotropic velocity magnitude of 3.1%–4.1% with an FVA dip of  $-36^\circ$  to  $-59^\circ$  (down to the SSE) and a FVA strike of N29°W–N10°W. The lower layer has an anisotropic velocity magnitude of 6.4%–8.0% with a FVA strike of  $52^\circ$ – $74^\circ$  and a FVA dip between  $-20^\circ$  and  $-1^\circ$  (down to the SW). This best model is very similar to the P4 model which is also an acceptable model. Our most robust conclusion is that model parameterizations without a dipping FVA do not provide acceptable fits to our waveform data set.

### 3.3. Direct S-Wave Resolution Improvement

[27] An important question with respect to our analysis is how much the five direct S-events contributed to the resolution of our model parameters. Of most concern is whether source-side anisotropy has significantly contaminated our direct-S waveforms. Theoretical analysis demonstrates that the greater incidence angles of direct S arrivals with respect to SKS arrivals (Figure 3) provide greater sensitivity to dipping FVA anisotropy [Chevrot and van der Hilst, 2003]. In addition, the direct S-waves provide a range of SH-like polarizations not provided by the SV-polarized SKS waves. As mentioned previously, our analysis finds that the P1 and P3 model parameterizations without a dipping FVA can be rejected at high probability (Table 6).

[28] Figure 11 shows the one and two dimensional PPD marginals for three subsets of our waveforms: the five direct



**Figure 10.** Maximum likelihood model (Table 6) predicted cross-convolved waveform pairs. The dotted lines denote the five direct-S waveform pairs and solid lines are the SKS waveform pairs. The direct S-waves are fit as well as the SKS waves.

**Table 6.** Maximum Posterior Probability Density (PPD) Models

	Layer 1					Layer 2					DOF	$\chi^2_v$	RP <sup>b</sup>		
	Anis., (%) <sup>a</sup>		FVA Strike, (°)		FVA Dip, (°)	Anis., (%) <sup>a</sup>		FVA Strike, (°)		FVA Dip, (°)					
P1	3.8	[3.6, 4.0]	64	[63, 65]	-	-	-	-	-	-	118	2.8	98%		
P2	3.7	[3.4, 3.8]	65	[63, 67]	-28	[-32, -25]	-	-	-	-	117	2.4	89%		
P3	2.7	[2.1, 3.1]	2	[-5, 8]	-	-	5.9	[5.3, 6.2]	75	[69, 80]	-	116	2.7	97%	
P4	2.9	[2.6, 3.4]	-10	[-18, -3]	-	-	5.5	[5.3, 6.1]	71	[66, 77]	-28	[-30, -9]	115	2.2	80%
P5	3.6	[3.1, 4.1]	-20	[-29, -10]	-47	[-59, -36]	7.5	[6.4, 8.0]	65	[52, 74]	-12	[-20, -1]	114	1.9	-

<sup>a</sup>For 100-km thick layer.

<sup>b</sup>Rejection probability. Brackets show the 80% probability region. The FVA strike is rendered positive clockwise with respect to North. The dip is positive downwards with respect to horizontal along the FVA strike direction. Negative dip means it dips towards 180° of the strike.

S-waveforms, the ten SKS waveforms, and the combination of both waveforms sets. These three data sets are fitted to our five different model parameterizations. For the simplest model parameterization (P1), all three data sets produce 1- and 2-D PPD marginals that overlap. This demonstrates that the direct-S and SKS data sets individually possess coherent anisotropic signal with which to constrain the single anisotropic layer. For the single layer dipping FVA parameterization (P2), the FVA strike found is the same as the P1 model (Table 6), but the anisotropy magnitude found by the direct-S-waves is 2% greater than the SKS or combined data sets. For the double layer model with flat FVA parameterization (P3), the strike of the bottom layer FVA is well defined by all three data sets, but the direct S-wave data set shows variable marginals.

[29] The two most complicated model parameterizations (P4 and P5) reveal the following general characteristics with respect to their PPD marginals. First, the marginals for the combined direct-S and SKS data sets are always more peaked and compact than the marginals from either of the individual data sets. Second, the 80% probability bounds overlap between the individual SKS and direct-S waveforms subsets for ten out of the fifteen 2-D marginals. Third, the five direct-S waveform data set do have the least compact PPD marginals. We suggest that the marginals for the direct S-wave data set are less compact because of the existence of the possible source-side anisotropy and the limited resolving power provided by this small data subset (without the SKS/SKKS waveforms). Yet, the five direct-S events do add significant information as demonstrated by the compactification of the 1- and 2-D PPD marginals using the full data set with respect to the SKS-only data set.

### 3.4. Geographic Variations and Model Misfit

[30] As a test of robustness of our models with respect to the geographic size of our array, the Billings array is divided into a northern and southern sub-array of stations. This division results in a northern and southern sub-array with 11 and 19 stations, respectively. From these two geographical sub-arrays, a set of event stack waveforms are produced and

NA modeled. Comparison of the PPD associated with these two sub-arrays shows that the 1- and 2-D PPD marginals strongly overlap (Figure 12). This suggests that the anisotropic structure beneath the Billings array is reasonably spatially uniform within the sampling volume.

[31] The synthetic waveform pairs associated with the best model for each of the five anisotropic model parameterizations are shown in Figure 13. This plot shows that the S1 components are only marginally affected by the different anisotropy models while the S2 components are greatly affected by the different anisotropy models. Obviously, modeling these S2 response variations requires an accurate set of waveforms with well characterized error estimates.

[32] To assess how the misfit between the data and five best models is distributed with respect to the different events, the cross-convolved waveforms and their misfits are presented (Figure 14). Because all the events are initially equal-weighted in our data fitting, the chi-square values have been normalized by the chi-square value associated with the P1 model parameterization. This permits the trend in the chi-square values with respect to the P1–P5 models to be easily assessed. Inspection of the normalized chi-square values for each event and model shows that the more complicated anisotropy models generally have better fits to the data (Table 6). Noteworthy is that some event waveforms are more poorly fit by the more complicated anisotropy models. We suggest that signal exists in our data set that is not predicted by our forward model. For example, the true anisotropic variations could be more complicated than two layers of transverse symmetry anisotropy, perhaps due to modest lateral heterogeneity in anisotropy. Also, source side anisotropy for the five direct-S events could contaminate these waveforms to some degree. Yet, we note that the cross-convolved direct-S waveforms are not systematically misfit with respect to the SKS arrivals (Figure 14).

## 4. Discussion and Conclusions

[33] The orientation of the two anisotropic layers' strike and dip for the best anisotropy model (P5) is superimposed

**Figure 11.** PPD marginals for the SKS-only, direct-S-only, and full data set (Figure 8). See Figure 7 for figure layout description. Each subplot of the PPD marginals is labeled by the model parameterization (P1–P5). The marginals for the SKS, direct-S, and combined data sets are color coded as blue, green and red. Note that both the 1- and 2-D marginals are made more compact by the addition of the five direct-S waveform pairs. The 2-D marginals show that the direct-S wave data set alone is insufficient to constrain the P3–P5 model parameters well.

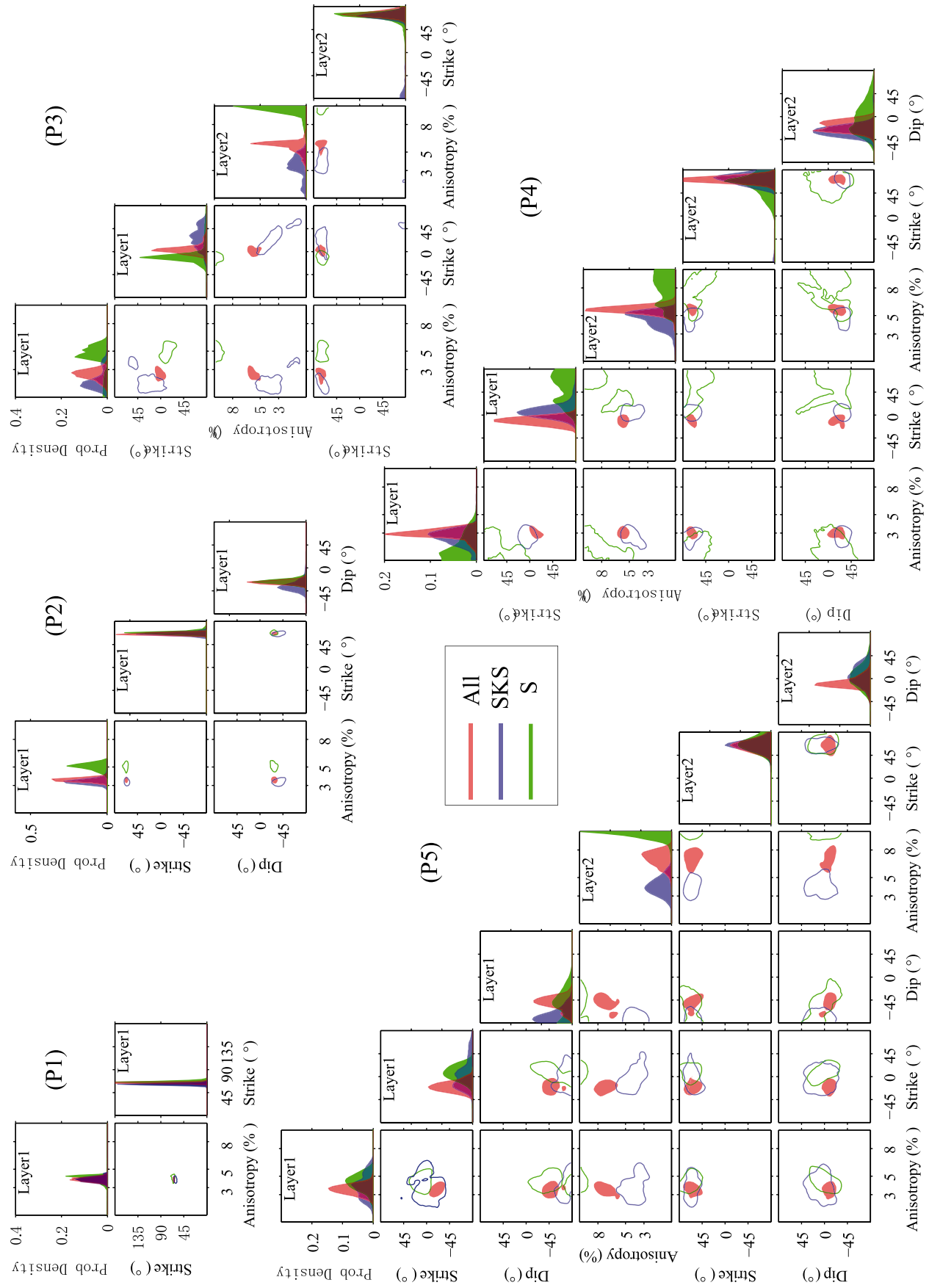


Figure 11

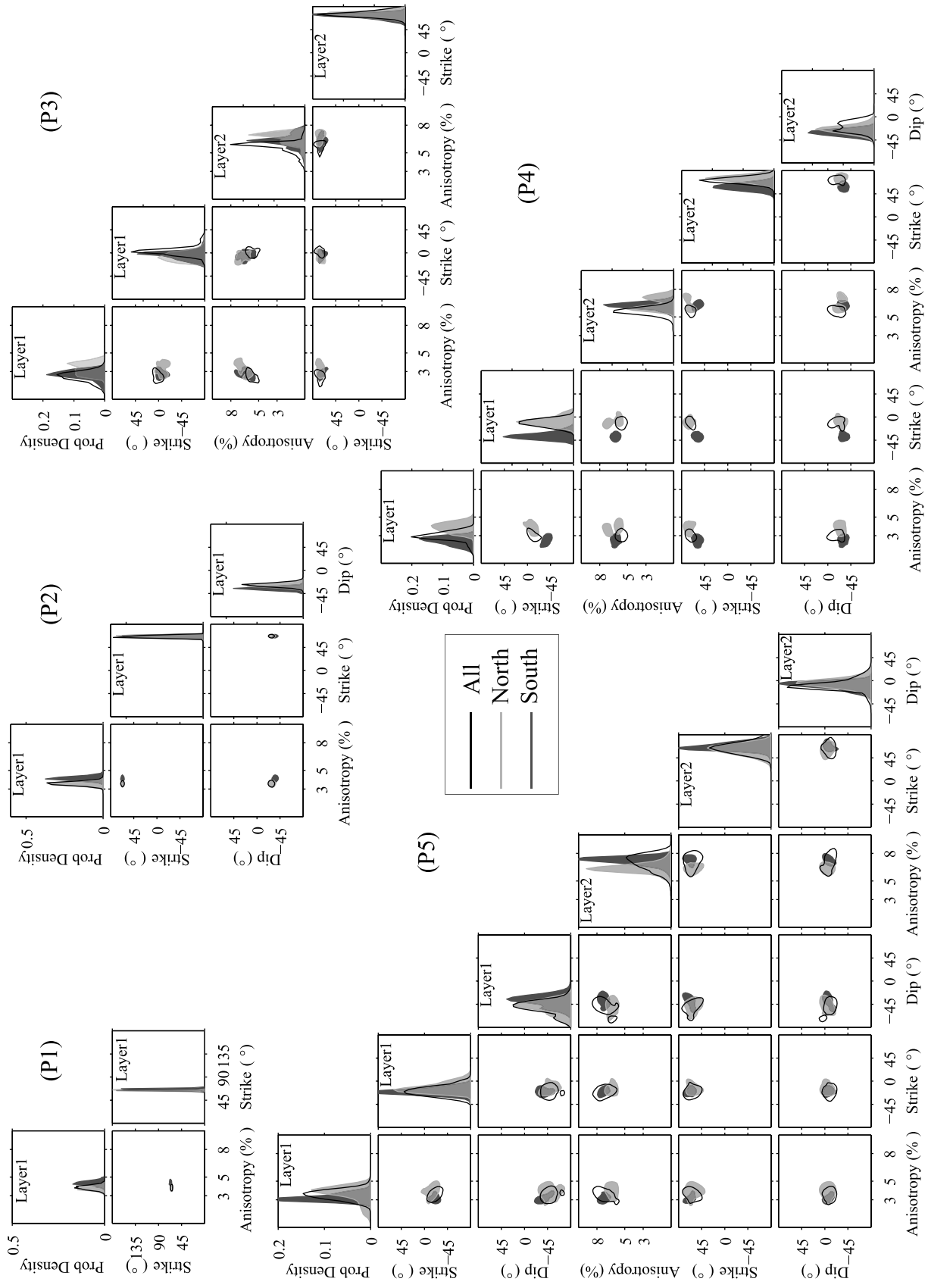
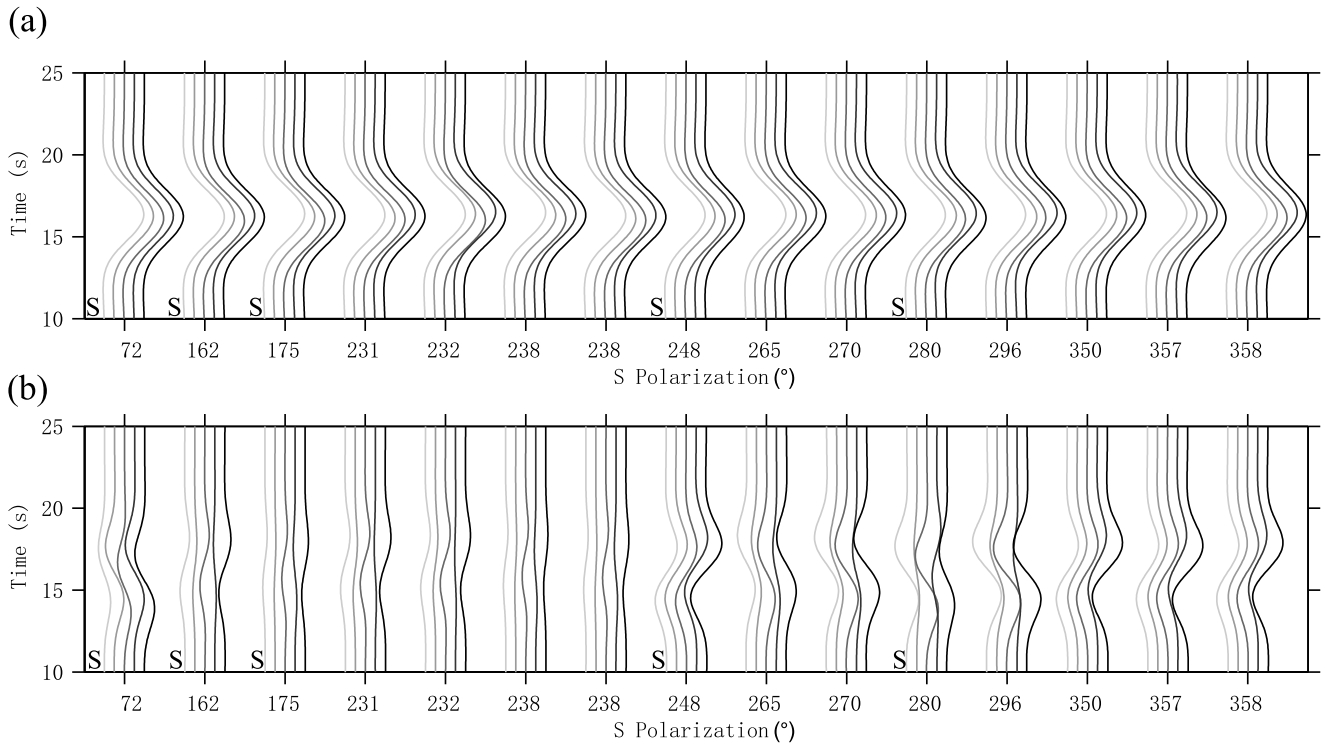


Figure 12



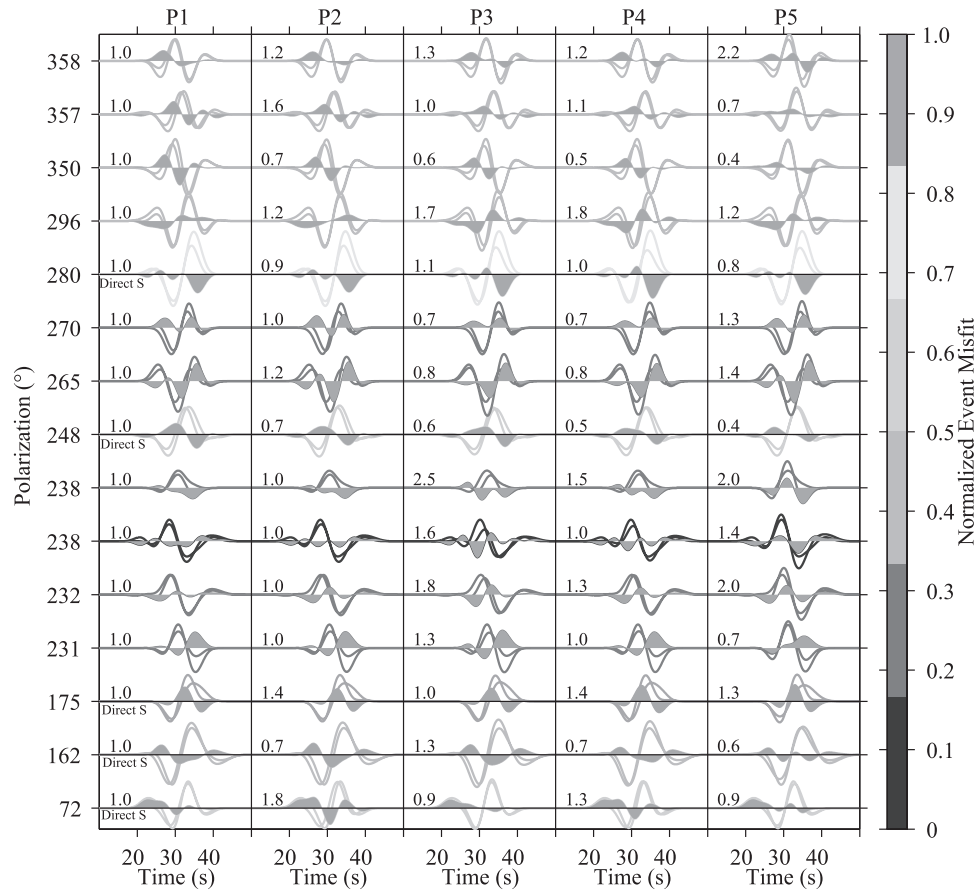


**Figure 13.** Maximum likelihood model synthetic waveform pairs for the five model parameterizations (Table 6). (a) S1 waveforms. (b) S2 waveforms. For each given event polarization values (Table 1), the waveforms from left to right represent the waveform response of model parameterizations P1–P5. The “S” symbol denotes the direct-S waveform.

upon a P wave tomogram that was produced using the Yellowstone and Billings teleseismic P wave traveltimes measurements [Yuan and Dueker, 2005]. Receiver function studies that use local surface wave models to map time to depth (Stachnik *et al.*, in prep) show that the crust beneath the Billings array is 48–54 km thick [Yuan *et al.*, 2006]. The P wave tomogram shows that the crust is underlain by a relatively high velocity mantle lithosphere that extends to 120–140 km depth. This lithospheric thickness is consistent with eastern Montana mantle xenolith studies that find a 120–140 km thick lithosphere from Eocene kimberlite eruptions [Carlson *et al.*, 2004]. Thus the lithospheric mantle layer beneath the Billings array is estimated to be at least 75 km thick which is sufficiently thick to create a significant anisotropic signal. Below 120–140 km depth, the mantle is normal within the volume sampled by our S-wave raypaths, although the westernmost raypaths come within 40 km of the edge of the Yellowstone plume (Figure 15).

[34] The most common element between our five simplest anisotropic velocity models is that one layer (the bottom layer for the two layer models) has a FVA strike between  $64^{\circ}$ – $75^{\circ}$  which is parallel to the North American absolute plate motion direction. This NE-directed FVA strike is consistent with the nearby Yellowstone array single layer shear wave splitting measurements [Waite *et al.*, 2005]. An additional result for our dipping anisotropy models is that the lower layer FVA dips at  $-28^{\circ}$  to  $-12^{\circ}$  down to the SW with respect to horizontal. Unique interpretation of the SW dipping FVA in the lower layer is not possible until potential dipping anisotropy is characterized around the Yellowstone region. At present, only one other analysis requires a dipping FVA axis for the National seismic network station HLID near the Yellowstone hot spot track [Walker *et al.*, 2004]. The most noteworthy statement to be made with respect to the SW dip of the FVA is that this dip is the opposite of what a passive plate shear model of LPO evolution predicts [Bokelmann, 2002]. One could also speculate that plume entrained mantle

**Figure 12.** PPD marginals for the geographically divided north and south array sub-data set. See Figure 7 for figure layout description. Each subplot of the PPD marginals is labeled by the model parameterization (P1–P5). For the 1-D PPD marginals, the combined data set PPD marginals are drawn as unfilled lines and the north and south sub-data set PPD marginals are filled with light gray (north data set) and dark gray (south data set). For the 2-D PPD marginals, the 80% probability contour for the north and south data sets are filled with light/dark gray and the combined data set probability is an unfilled contour.



**Figure 14.** Maximum likelihood model cross-convolved traces and residual traces. The columns labeled P1–P5 are the five model parameterizations. The cross-convolved traces are the thin lines and the gray filled traces are the difference between each cross convolved trace pairs (residual trace). The Chi-square error between each of the cross-convolved waveform-pair, normalized by its P1 Chi-square value, is given to the left of each trace. The contribution of each cross convolved trace pair to the sum of the P1 Chi-square value is shaded in gray scale.

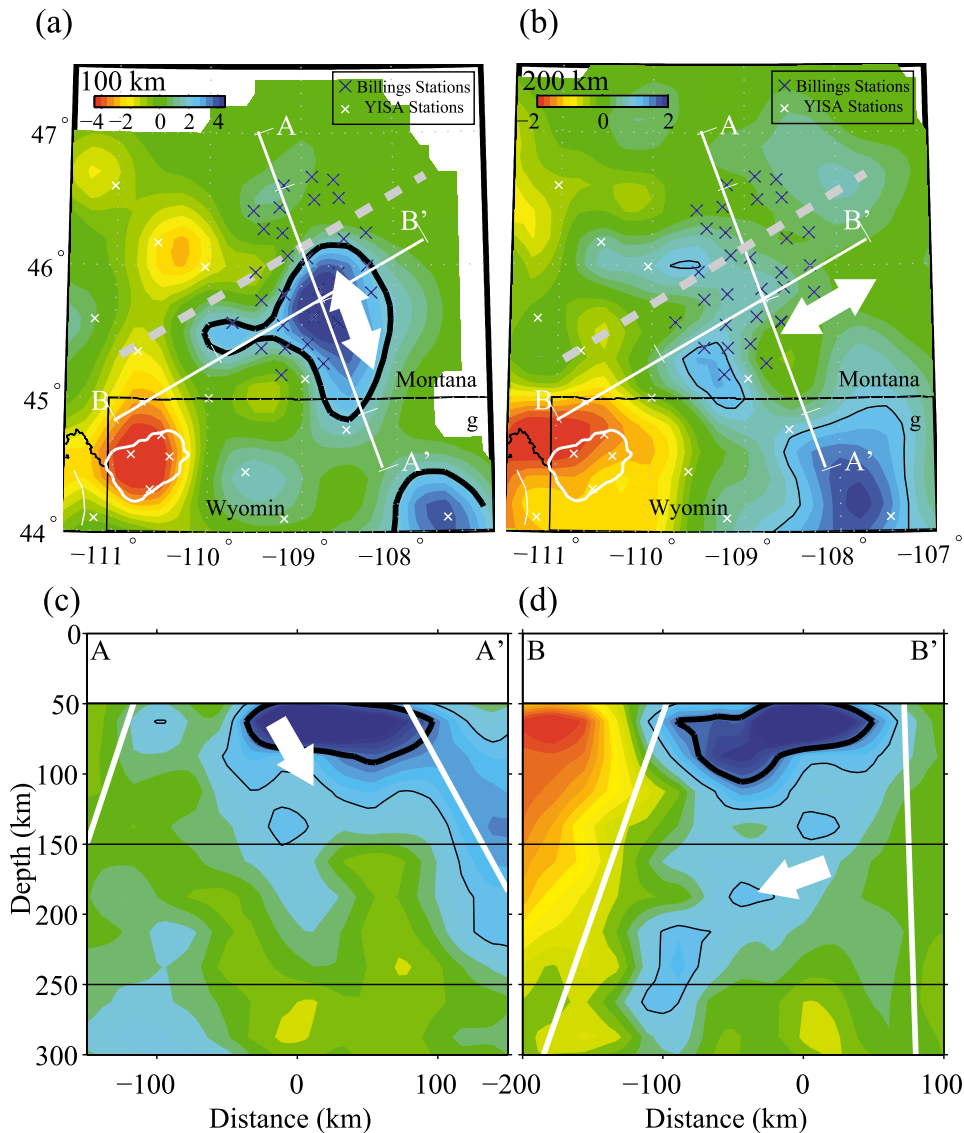
associated with the Yellowstone plume may be perturbing the FVA dip in the asthenospheric layer beneath the Billings array.

[35] The  $N10^{\circ}$ – $20^{\circ}W$  FVA strike of the upper layer for the P4 and P5 models is broadly consistent with the FVA strike of the upper layer found by several other anisotropic studies. Love and Rayleigh wave imaging finds that most of the mid-continental region of the United States and Canada has a northerly trending FVA in the upper 100 km [Marone and Romanowicz, 2007]. More specifically, beneath the Billings array the Marone and Romanowicz upper anisotropic layer FVA points at  $N10^{\circ}W$  within the error of our upper layer FVA orientation. Rayleigh wave imaging beneath the east-central North American continent and the Slave craton also finds an upper lithospheric layer with a north trending FVA [Deschamps et al., 2006; Chen et al., 2007; Snyder and Bruneton, 2007]. In addition, two-layer anisotropy modeling in the Slave craton, southern Alberta and southern Wyoming find an anisotropic lithospheric layer with a northerly striking FVA underlain by a lower layer that is approximately parallel to absolute plate motion [Snyder et al., 2003; Currie et al., 2004; Fox and Sheehan,

2005]. Deschamps et al. [2006] speculates that the north trending lithospheric LPO was frozen into the lithosphere during northward drift of the North American plate during Mesozoic times.

[36] Given that the P4 model can only be rejected at 80% probability, we consider the  $-47^{\circ}$  south directed dip of the upper lithospheric layer for model P5 to not be a robust feature. With regard to the geologic perspective, the Madison Mylonite zone (Figure 15) has been interpreted as either an Archean age suture zone [Hoffman, 1989] or an intracratonic shear zone [Erslev and Sutter, 1990]. Given the uncertainties in whether the upper layer requires a dipping FVA, we simply note that the deformation along the Madison Mylonite zone could have produced a dipping FVA.

[37] In conclusion, we have developed a new method to discriminate between the five simplest models of upper mantle anisotropy using source normalized and stacked teleseismic SKS and direct-S waveforms. The addition of direct-S events has been shown to significantly improve resolution of the anisotropic model parameters. The direct Monte-Carlo NA search has proven to be an efficient



**Figure 15.** Comparison of best anisotropic velocity model (P5) and a teleseismic P wave tomographic model [Yuan and Dueker, 2005]. (a) Map view at 100 km depth. The thick and thin black contours are the 2% and 1% P wave velocity perturbation. The Yellowstone Caldera is outlined in white. The gray dash line denotes the Madison Mylonite shear zone. The two-headed white arrow shows the strike of the FVA strike of the upper anisotropic layer. (b) Map view at 200 km with lower layer FVA strike denoted by two-headed white arrow. (c) Profile along A-A'. The white arrow shows the FVA dip in the top layer. The two white lines show the volume sampled by the S-wave raypaths (Figure 1). (d) Profile along B-B'. The white arrow shows the FVA dip in the bottom layer.

algorithm to map the model parameter PPD. This methodology has permitted full assessment of model parameter uncertainties using the 1- and 2-D PPD marginals. Finally, the robustness of our results is largely dependent upon the well-resolved anisotropic signals that the stacking of 30 broadband stations from the Billings array has permitted.

[38] **Acknowledgments.** We thank the Billings array deployment group, the IRIS PASSCAL data center and the PASSCAL instrument center. Martha Savage, the Associate Editor and an anonymous reviewer provide detailed comments that improved the manuscript. We are also grateful to M. Sambridge and A. Frederiksen for providing codes in original format, and S. Wulff and J. Pierce for useful discussions. This project was funded by NSF Geophysics program grant EAR-0440432.

## References

- Anglin, D. K., and M. J. Fouch (2005), Seismic anisotropy in the Izu-Bonin subduction system, *Geophys. Res. Lett.*, *32*, 1–4, L09307, doi:10.1029/2005GL022714.
- Babuska, V., and J. Plomerova (1993), Lithospheric thickness and velocity anisotropy; Seismological and geothermal aspects, *Tectonophysics*, *225*, 79–89.
- Bank, C., and M. G. Bostock (2003), Linearized inverse scattering of teleseismic waves for anisotropic crust and mantle structure: 2. Numerical examples and application to data from Canadian stations, *J. Geophys. Res.*, *108*(B5), 2259, doi:10.1029/2002JB001950.
- Behn, M. D., C. P. Conrad, and P. G. Silver (2004), Detection of upper mantle flow associated with the African Superplume, *Earth Planet. Sci. Lett.*, *224*, 259–274.
- Bokelmann, G. H. R. (2002), What forces drive North America?, *Geology*, *30*, 1027–1030.

- Carlson, R. W., A. J. Irving, D. J. Schulze, and B. C. Hearn (2004), Timing of precambrian melt depletion and phanerozoic refertilization events in the lithospheric mantle of the Wyoming Craton and adjacent Central Plains Orogen, *Lithos*, 77, 453–472.
- Chen, C.-W., S. Rondenay, D. S. Weeraratne, and D. B. Snyder (2007), New constraints on the upper mantle structure of the slave craton from Rayleigh wave inversion, *Geophys. Res. Lett.*, 34, L10301, doi:10.1029/2007GL029535.
- Chevrot, S., and R. D. van der Hilst (2003), On the effects of a dipping axis of symmetry on shear wave splitting measurements in a transversely isotropic medium, *Geophys. J. Int.*, 152, 497–505.
- Crampin, S. (1977), A review of the effects of anisotropic layering on the propagation of seismic waves, *Geophys. J. R. Astron. Soc.*, 49, 9–27.
- Currie, C. A., J. F. Cassidy, R. D. Hyndman, and M. G. Bostock (2004), Shear wave anisotropy beneath the Cascadia subduction zone and western North American Craton, *Geophys. J. Int.*, 157, 341–353.
- Davis, P. M. (2003), Azimuthal variation in seismic anisotropy of the southern California uppermost mantle, *J. Geophys.*, 109.
- Deschamps, F., S. Lebedev, T. Meier, and J. Trampert (2006), Stratification of seismic anisotropy beneath the east-central United States, *Eos Trans. AGU*, 87(52), Fall Meet. Suppl. Abstr. T53C-1630.
- Diebold, J. B. (1987), Three-dimensional traveltimes equation for dipping layers, *Geophysics*, 52, 1492–1500.
- Efron, B., and R. Tibshirani (1986), Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy, *Stat. Sci.*, 1, 54–77.
- Erslev, E. A., and J. F. Sutter (1990), Evidence for Proterozoic mylonitization in the northwestern Wyoming province, *Geol. Soc. Am. Bull.*, 102, 1681–1694.
- Farra, V., L. Vinnik, B. Romanowicz, G. Kosarev, and R. Kind (1991), Inversion of teleseismic S particle motion for azimuthal anisotropy in the upper mantle: A feasibility study, *Geophys. J. Int.*, 106, 421–431.
- Fischer, K. M., and X. Yang (1994), Anisotropy in Kuril-Kamchatka subduction zone structure, *Geophys. Res. Lett.*, 21(1), 5–8.
- Fischer, K. M., E. M. Parmentier, A. R. Stine, and E. R. Wolf (2000), Modeling anisotropy and plate-driven flow in the Tonga subduction zone back arc, *J. Geophys. Res.*, 105, 16,181–16,191.
- Fischer, K. M., A. Li, D. W. Forsyth, and S. H. Hung (2005), Imaging three-dimensional anisotropy with broadband seismometer arrays, in *Seismic Data Analysis and Imaging With Global and Local Arrays*, edited by A. R. Levander and G. Nolet, American Geophysical Union, Washington, DC.
- Fouch, M., and S. Rondenay (2006), Seismic anisotropy beneath stable continental interiors, *Phys. Earth Planet. Inter.*, 158, 292–320.
- Fox, O., and A. F. Sheehan (2005), Shear wave splitting beneath the CDROM transects, in *The Rocky Mountain region—an evolving lithosphere: tectonics, geochemistry, and geophysics*, edited by G. Randy and K. E. Karlstrom, American Geophysical Union, Washington D. C.
- Frederiksen, A. W., and M. G. Bostock (2000), Modelling teleseismic waves in dipping anisotropic structures, *Geophys. J. Int.*, 141, 401–412.
- Frederiksen, A. W., H. Folsom, and G. Zandt (2003), Neighbourhood inversion of teleseismic Ps conversions for anisotropy and layer dip, *Geophys. J. Int.*, 155, 200–212.
- Gledhill, K., and D. Gubbins (1996), SKS splitting and the seismic anisotropy of the mantle beneath the Hikurangi subduction zone, New Zealand, *Phys. Earth Planet. Inter.*, 95, 227–236.
- Gorman, A. R., et al. (2002), Deep probe: Imaging the roots of western North America, *Can. J. Earth Sci.*, 39, 375–398.
- Gripp, A. E., and R. G. Gordon (2002), Young tracks of hotspots and current plate velocities, *Geophys. J. Int.*, 150(2), 321.
- Hartog, R., and S. Y. Schwartz (2000), Subduction-induced strain in the upper mantle east of the Mendocino triple junction, California, *J. Geophys. Res.*, 105, 7909–7930.
- Henstock, T. J., et al. (1998), Probing the Archean and Proterozoic lithosphere of western North America, *GSA Today*, 8, 1–5.
- Hoffman, P. F. (1989), Precambrian geology and tectonic history of North America, in *The Geology of North America - An Overview*, edited by A. W. Bally and A. R. Palmer, pp. 447–512, Geology Society of America, Boulder, CO.
- Holtzman, B. K., T. Hiraga, J. Hustoft, D. L. Kohlstedt, M. E. Zimmerman, and F. Heidelbach (2003), Melt segregation and strain partitioning: Implications for seismic anisotropy and mantle flow, *Science*, 301, 1227–1230.
- Jung, H., and S. Karato (2001), Water-induced fabric transitions in olivine, *Science*, 293, 1460–1462.
- Kaminski, E., and N. M. Ribe (2002), *Timescales for the evolution of seismic anisotropy in mantle flow*, American Geophysical Union and The Geochemical Society, United States. Pages: 17.
- Katayama, I., H. Jung, and S. Karato (2004), New type of olivine fabric from deformation experiments at modest water content and low stress, *Geology*, 32, 1045–1048.
- Levin, V., W. Menke, and J. Park (1999), Shear wave splitting in the Appalachians and the Urals; A case for multilayered anisotropy, *J. Geophys. Res.*, 104, 17,975–17,994.
- Levin, V., J. Park, M. Brandon, J. Lees, V. Peyton, E. Gordeev, and A. Ozerov (2002), Crust and upper mantle of Kamchatka from teleseismic receiver functions, in *Structure of the continental lithosphere and upper mantle*, edited by I. M. Artemieva et al., Elsevier, Amsterdam.
- Levin, V., A. Henza, and J. Park (2006), Texture of mantle lithosphere along the Dead Sea Rift: Recently imposed or inherited?, *Phys. Earth Planet. Int.*, 158, 174–189.
- Long, M. D., and R. D. van der Hilst (2005), Upper mantle anisotropy beneath Japan from shear wave splitting, *Phys. Earth Planet. Inter.*, 151, 206–222.
- Marone, F., and B. Romanowicz (2007), The depth distribution of azimuthal anisotropy in the continental upper mantle, *Nature*, 447, 198–201.
- Menke, W., and V. Levin (2003), The cross-convolution method for interpreting SKS splitting observations, with application to one and two-layer anisotropic earth models, *Geophys. J. Int.*, 154, 379–392.
- Ozalaybey, S., and M. K. Savage (1994), Double-layer anisotropy resolved from S phases, *Geophys. J. Int.*, 117, 653–664.
- Park, J., and V. Levin (2002), Seismic anisotropy; Tracing plate dynamics in the mantle, *Science*, 296, 485–489.
- Rumpker, G., and P. G. Silver (2000), Calculating splitting parameters for plume-type anisotropic structures of the upper mantle, *Geophys. J. Int.*, 143, 507–520.
- Sambridge, M. (1999a), Geophysical Inversion with a Neighbourhood Algorithm -I. Searching a parameter space, *Geophys. J. Int.*, 138, 479–494.
- Sambridge, M. (1999b), Geophysical Inversion with a Neighbourhood Algorithm -II. Appraising the ensemble, *Geophys. J. Int.*, 138, 727–746.
- Sambridge, M. (2001), Finding acceptable models in nonlinear inverse problems using a neighbourhood algorithm, *Inverse Problems*, 17, 387–403.
- Savage, M. K. (1999), Seismic anisotropy and mantle deformation; What have we learned from shear wave splitting?, *Rev. Geophys.*, 37, 65–106.
- Savage, M. K., P. G. Silver, and R. P. Meyer (1990), Observations of teleseismic shear-wave splitting in the Basin and Range from portable and permanent stations, *Geophys. Res. Lett.*, 17(1), 21–24.
- Schutt, D. L., K. Dueker, and H. Yuan (2008), Crust and upper mantle velocity structure of the Yellowstone hotspot and surroundings, *J. Geophys. Res.*, doi:10.1029/2007JB005109, in press.
- Sherrington, H. F., G. Zandt, and A. W. Frederiksen (2004), Crustal fabric in the Tibetan Plateau based on waveform inversions for seismic anisotropy parameters, *J. Geophys. Res.*, 109, B02312, doi:10.1029/2002JB002345.
- Silver, P. G., and W. W. Chan (1988), Implications for continental structure and evolution from seismic anisotropy, *Nature*, 335, 34–39.
- Silver, P. G., and W. W. Chan (1991), Shear wave splitting and subcontinental mantle deformation, *J. Geophys. Res.*, 96, 16,429–16,454.
- Silver, P. G., and M. K. Savage (1994), The interpretation of shear-wave splitting parameters in the presence of two anisotropic layers, *Geophys. J. Int.*, 119, 949–963.
- Snyder, D. B., and M. Bruneton (2007), Seismic anisotropy of the Slave craton, NW Canada, from joint interpretation of SKS and Rayleigh waves, *Geophys. J. Int.*, 169, 170–188.
- Snyder, D. B., M. G. Bostock, and G. D. Lockhart (2003), Two anisotropic layers in the Slave craton, *Lithos*, 71.
- van Decar, J. C., and R. S. Crosson (1990), Determination of teleseismic relative phase velocity arrival times using multi-channel cross-correlation and least squares, *Bull. Seismol. Soc. Am.*, 80, 150–169.
- Vinnik, L. P., V. Farra, and B. Romanowicz (1989), Azimuthal anisotropy in the Earth from observations of SKS at GEOSCOPE and NARS broadband stations, *Bull. Seismol. Soc. Am.*, 79, 1542–1558.
- Waite, G., D. L. Schutt, and R. B. Smith (2005), Models of lithosphere and asthenosphere anisotropic structure of the Yellowstone hotspot from shear wave splitting, *J. Geophys. Res.*, 110, B11304, doi:10.1029/2004JB003501.
- Walker, K. T., G. H. R. Bokelmann, and S. L. Klemperer (2004), Shear-wave splitting beneath the Snake River Plain suggests a mantle upwelling beneath eastern Nevada, USA, *Earth Planet. Sci. Lett.*, 222, 529–542.
- Yang, X., and K. M. Fischer (1994), Constraints on North Atlantic upper mantle anisotropy from S and SS phases, *Geophys. Res. Lett.*, 21(4), 309–312.
- Yuan, H., and K. Dueker (2005), P-wave Tomogram of the Yellowstone plume, *Geophys. Res. Lett.*, 32, L07304, doi:10.1029/2004GL022056.
- Yuan, H., K. Dueker, and D. Schutt (2006), Synoptic scale crustal thickness and velocity maps along the Yellowstone hotspot track, *Eos Trans. AGU*, 87(52), Fall Meet. Suppl., Abstract S43A-1376.

K. Dueker, D. L. Schutt, and H. Yuan, Department of Geology and Geophysics, University of Wyoming, USA. (huaiyu.yuan@berkeley.edu)