Geophysical Journal International

Geophys. J. Int. (2016) **205,** 856–863
Advance Access publication 2016 February 16
GJI Seismology

doi: 10.1093/gji/ggw050

Model for small-scale crustal heterogeneity in Los Angeles basin based on inversion of sonic log data

W.H. Savran[1,2] and K.B. Olsen[1]

[1]*Department of Geological Sciences, San Diego State University, San Diego, CA, USA. E-mail: wsavran@ucsd.edu*
[2]*Scripps Institution of Oceanography, University of California, San Diego, CA, USA*

Accepted 2016 January 29. Received 2016 January 28; in original form 2015 August 16

SUMMARY

High-frequency seismic ground motion (10+ Hz), as needed for earthquake engineering design purposes, is largely controlled by the metre-scale structure of the earth's crust. However, the state-of-the-art velocity models poorly resolve small-scale features of the subsurface velocity and density variation. We invert 35 sonic logs (up to 3000 m in depth) in and near Los Angeles basin, CA, to obtain a statistical description of the small-scale heterogeneities of the basin. Assuming a von Karman autocorrelation function, our analysis finds that Hurst numbers, ν, between 0.0 and 0.2, vertical correlation lengths, a_z, of 15–150 m and standard deviations of about 5 per cent characterize the variability in the borehole data. We report average parameters for Los Angeles basin of $\nu = 0.064\,(0.058, 0.069) \pm 0.01\,(0.006, 0.012)$ and $a_z = 54\,(51.1, 57.6) \pm 5.9\,(1.79, 9.53)$ m with 95 per cent confidence intervals listed in the parentheses. Despite the large depth range of the logs, there is no significant variation of the statistical parameters with depth. Our analysis of 371 depth-averaged shear wave velocities in the upper 30 m, Vs_{30}, provides only an upper bound of basin scale-length estimates due to the coarse sampling distance, with a Hurst number of about 0.3 and lateral correlation lengths, a_x, of 5–10 km.

Key words: Time-series analysis; Fractals and multifractals; Earthquake ground motions; Statistical seismology.

1 INTRODUCTION

Gaining a solid understanding of the expected range of ground motions from future large earthquakes is imperative so structures can be properly engineered to withstand this shaking. For example, crucial lifelines such as railroads, bridges and pipelines span large horizontal distances across the ground where they experience significant variability in seismic velocities down to the smallest scale lengths. Local site effects can drastically amplify the ground motion from an earthquake, with prominent examples from the 1989 Loma Prieta earthquake in the Marina District of San Francisco and the 1985 Michoacan earthquake in Mexico City. Understanding this small-scale variability is necessary as simulations begin to resolve seismic wavelengths on the scale length of several metres. Until recently, fully 3-D physics-based simulation of ground motions has been limited to around 1 Hz (Olsen *et al.* 2006, 2008, 2009; Graves *et al.* 2008; Roten *et al.* 2011), while higher frequencies are modelled by stochastic or semi-stochastic approaches (Frankel 2009; Anderson 2015; Crempien & Archuleta 2015; Graves & Pitarka 2015; Olsen & Takedatsu 2015). Due to a surge in available powerful supercomputing resources and parallel/GPU programming, fully deterministic seismic wave propagation to frequencies of 10 Hz+ is now computationally within reach (Cui *et al.* 2013; Taborda & Bielak 2013). Along with better understanding source complexity (Shi & Day 2013) and nonlinearity (Roten *et al.* 2014), another limiting factor in realistic high-frequency ground motion simulations remains in characterizing the small-scale variation (from metres to tens of metres) of crustal seismic velocities. Current velocity models, for example, the Southern California Earthquake Center (SCEC) Community Velocity Model (CVM)-S 4.0 (Magistrale *et al.* 2000; Kohler *et al.* 2003), CVM-H 11.9 (Süss & Shaw 2003) and CVM-SI 4.26 (Lee *et al.* 2014) insufficiently resolve the small-scale variability known to exist in the earth's crust.

Due to large data acquisition costs at metre-scale resolution, it is not feasible to resolve the small-scale structure using direct measurements. Instead, the small-scale heterogeneities may be described by their two-point statistics (i.e. their spatial variability). While statistical representations of small-scale heterogeneities naturally do not include the actual physical locations of the short-wavelength velocity and density variations, ensembles of ground motion simulations computed using such distributions should reflect the correct average and variability observed in strong-motion recordings. A requirement for the simulated ground motion level and variability to be realistic is to use constraints from the statistics of actual small-scale variation observed in the earth's crust.

© The Authors 2016. Published by Oxford University Press on behalf of The Royal Astronomical Society.

Downloaded from http://gji.oxfordjournals.org/ at San Diego State University on March 17, 2016

Frankel & Clayton (1986) showed that the von Karman autocovariance function is well suited for characterizing the statistical variation of crustal heterogeneities. The functional form of the von Karman autocovariance function

$$\Phi_{\nu,a}(r) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{r}{a}\right)^{\nu} K_{\nu}\left(\frac{r}{a}\right), \quad (1)$$

where ν is the Hurst exponent, a is the correlation length, K_{ν} is the modified Bessel function of order ν, $\Gamma(\nu)$ is the gamma function and σ^2 is the variance with Fourier transform

$$P(k) = \frac{\sigma^2 (2\sqrt{\pi}a)^E \Gamma(\nu + E/2)}{\Gamma(\nu)(1 + k^2 a^2)^{\nu+E/2}}, \quad (2)$$

where E is the Euclidean dimension. Previous works have constrained the values of the von Karman parameters—the correlation length a, Hurst exponent ν and standard deviation σ from different data sets, such as sonic logs, digitized geological maps and seismic reflection data (Holliger & Levander 1992; De *et al.* 1994; Levander *et al.* 1994; Wu *et al.* 1994; Holliger 1996, 1997; Pullammanappallil *et al.* 1997; Dolan *et al.* 1998; Nakata & Beroza 2015). These studies report a wide range of Hurst exponents, 0.0–0.5, and correlation lengths between 30 and 300 m in the vertical direction, and horizontal to vertical anisotropy (H/V) between about 2 and 5. We attribute the wide range of parameters to the type of data from which the von Karman parameters were derived, and their analysis. We find upper crustal sonic logs tend to produce lower Hurst exponent estimations (Holliger 1996; Dolan *et al.* 1998) as compared with lower crustal data (Holliger & Levander 1992). There remains some disagreement between correlation lengths observed in the data, which appears to be related to whether the authors modelled residuals or velocities (Pullammanappallil *et al.* 1997; Dolan *et al.* 1998) or used cross-correlation versus spectral methods (Wu *et al.* 1994; Pullammanappallil *et al.* 1997; Dolan *et al.* 1998; Nakata & Beroza 2015). In this study, we aim to quantify the stochastic small-scale variability observed after removing a depth-dependent mean. For our purposes, we assume the CVMs to contain the deterministic or background portion of the velocity structure.

Several studies have also assessed the statistical parameters based on wave propagation simulations. Hartzell *et al.* (2010) used $\nu = 0.0$ with a = 5–10 km and σ = 5–10 per cent based on results from Frankel & Clayton (1986). Jacobsen & Olsen (2011) included statistical distributions with ν = (−0.5, 0, 0.5), a = 250 m and σ = 5–10 per cent. Imperatori & Mai (2013) and Bydlon & Dunham (2015) used $\nu = 0.3$ with correlation lengths on the order of hundreds of metres to kilometres, also including multiscale models created by the superposition of autocorrelation functions with multiple correlation lengths. Frankel & Clayton (1986) generated small-scale heterogeneities with three distinct autocorrelation functions, Gaussian, exponential (corresponding to $\nu = 0.5$) and self-similar (corresponding to $\nu = 0.0$). By testing end-member cases (i.e. $\nu = 0.0$, 0.5), Frankel & Clayton (1986) generated von Karman models of small-scale heterogeneities with a preferred value of $\nu = 0.0$ and correlation lengths between 5 and 10 km. The considerable variation in the preferred parameters of the autocorrelation found in the studies reflects the poor constraints from the underlying data. For example, the correlation lengths suggested from Frankel & Clayton (1986) (>10 km) were obtained using teleseismic sources, which provide limited constraints at near-field stations. Other studies included limited or no data constraints in their modelling.

The primary goal of this study is to understand the spatial statistics of the small-scale variation in the seismic velocity structure in order to provide the community with a data-derived model of small-scale heterogeneities in Los Angeles basin. Towards this goal we use variogram analysis on two separate data sets consisting of Vs_{30} measurements and borehole sonic logs to estimate the parameters needed to model the small-scale variability for von Karman autocorrelation functions (Hurst exponent, standard deviations and correlation length), so we perform our statistical analysis on the sonic log slowness residuals.

2 METHODS

To estimate the parameters necessary to characterize the lateral variability observed in Vs_{30} and the vertical variability observed in sonic log measurements we introduce two geostatistical methods: the classical variogram (Matheron 1963) and the moving-window variogram (Li & Lake 1994). In the following we briefly summarize these two methods, along with the Monte Carlo inversion that we use to estimate the von Karman parameters from the sonic logs.

2.1 Classical variogram

We analyse the Vs_{30} data in the time-domain using the classical estimate of semi-variance given by Matheron (1963) in eq. (3).

$$\tilde{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i) - Z(x_i + h)], \quad (3)$$

where $N(h)$ is the number of data points at each lag distance, h is the lag distance, and $Z(x_i)$ is spatial random field at location x_i. This variogram estimator has a vetted history in geostatistics since its derivation and is frequently used in kriging algorithms where developing a model of spatial correlation is paramount. The factor of 2 appears in the denominator accounting for the double representation of field values (x_1, x_2) and (x_2, x_1) in the summation. The estimate requires binning of the data to specify lag distances, and we find this estimate performs best when there are near-equal representations of the data in each bin. We can directly interpret the correlation length of the velocity structure given an estimate of the variogram, while we must fit a model to the variogram to determine the Hurst exponent. Here, we model the variogram (eq. 4) using a fractional Brownian motion (fBm) form (Li & Lake 1994; Mela & Louie 2001), based on the results from previous studies (Frankel & Clayton 1986; Wu *et al.* 1994; Holliger 1996; Pullammanappallil *et al.* 1997; Nakata & Beroza 2015). Following eq. (4), the Hurst exponent can be estimated as the half-slope of the linear regression in logarithmic space.

$$\begin{aligned} \tilde{\gamma}(h) &= \tfrac{\sigma^2}{a^{2\nu}} h^{2\nu} \quad h < a \\ \tilde{\gamma}(h) &= \sigma^2 \quad h \geq a \end{aligned} \quad (4)$$

where $\gamma(h)$ is the semi-variance and a is the correlation length. This model follows a power law for lags smaller than the correlation length and tends to a semi-variance equal to σ^2 after the correlation length. The constant $\sigma^2/a^{2\nu}$ implies some interdependence of the two parameters on the resulting semi-variance estimate.

Downloaded from http://gji.oxfordjournals.org/ at San Diego State University on March 17, 2016

2.2 Moving-window variogram

We introduce an alternative variogram estimator (moving-window formulation) derived by Li & Lake (1994) to analyse the sonic log measurements:

$$\hat{\gamma}_{N1}(h) = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{1}{2m}\sum_{j\in D_{i,h}}[Z(x_i) - Z(x_j)]^2\right\}. \quad (5)$$

Here, m is the number of data points in $D_{i,h}$ which is the index set of the data points contained within the moving window, $\Delta_{i,h}$, which has size h. n is the total number of data points in the random field, Z. $Z(x_i)$ represents the midpoint of $\Delta_{i,h}$, and $Z(x_j)$ represents each data point contained within $D_{i,h}$ excluding $Z(x_i)$. The inner summation is over the moving-window while the outer sum is over the entire random field. For a more detailed explanation and derivation of the moving-window variogram, please refer to Li & Lake (1994). The moving-window formulation (eq. 5) provides an estimate utilizing all available data pairs at all lag distances, whereas other variogram estimates (such as the Matheron (1963) method, discussed above) provide fewer data pairs as the observation distance increases resulting in larger variance as the lag distance increases (for example, see Fig. 1c and Supporting Information).

We choose to use the moving-window variogram method for the borehole logs as it reduces the variance at large lag distances which aids our inversion, because the moving-window estimate provides more consistent mean-squared residual values. The choice of variogram method does not change the results of the inversion (see Electronic Supplement). Also, the moving-window variogram does not always provide a more accurate direct estimate as compared to the Matheron (1963) method. For example, we find the classical variogram provides a more stable result for the Vs_{30} data set.

2.3 Monte Carlo inversion

We incorporate the moving-window variogram method into an inversion routine which performs a grid search on the parameter space $\nu \in [0.0, 0.3]$ and $a_z \in [15\,\text{m}, 150\,\text{m}]$, while minimizing the mean-squared residual between each sonic log variogram and variogram estimated from a synthetic borehole with pre-specified, ν and a_z. Here, we assume that each sonic log is a 1D realization of a single random field of velocity fluctuations representing the entire Los Angeles basin, and the statistics are stationary. We invert each borehole independently resulting in an estimate of the von Karman parameters for each borehole. This allows us to quantitatively assign uncertainty to our estimation.

To pre-processes the logs before inversion, we first detrend the data using a long-period median filter providing a zero-mean profile. We find the choice of demeaning filter can significantly affect the estimation of the correlation length. For example, choosing too short of a long-period median filter can force the correlation length to be proportional with the window size of the median filter. Due to large variability in the data, we find choosing long period median filters proportional to the borehole depth extent provide more stable estimates of the correlation length than low-order polynomial regression (see Electronic Supplement or Fig. 1(c) for all variogram estimates). Next, we apply a short-period median filter to remove artefacts resulting from the downhole logging procedure using a median filter with window size of 1.5 m, corresponding to the approximate resolution of the sonic logging tool. Also, we performed a sensitivity analysis on the effect of short-period median filtering the logs and find the results are robust with filter lengths <5 m. We

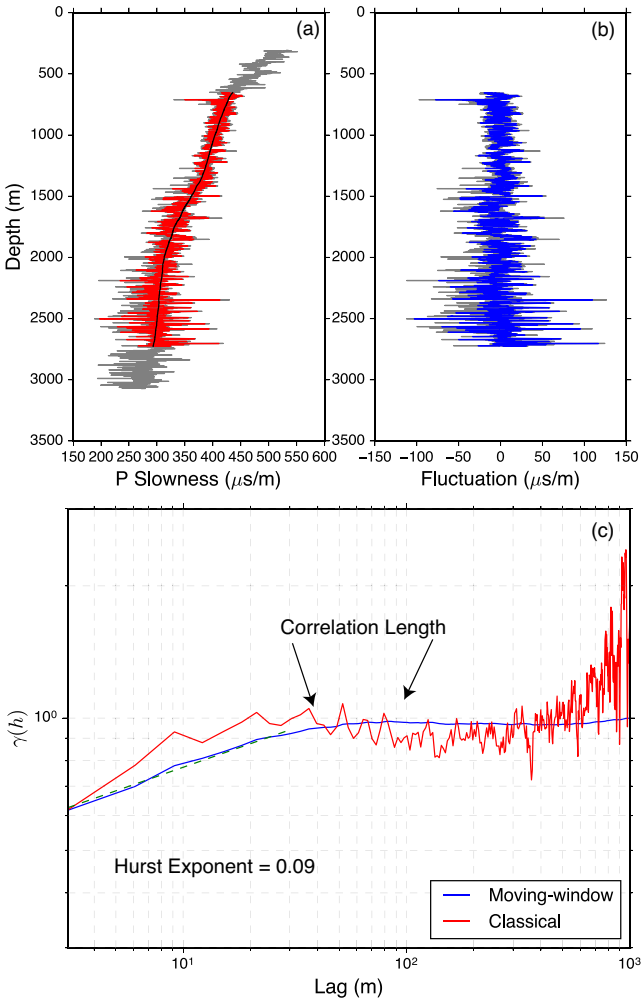

Figure 1. (a) Apal1 Vp sonic log (grey) with long-period median filter (black), and 1.5 m smoothed log (red). (b) De-trended fluctuation profile (blue) generated by subtracting the long-period median filter (black) from the red and grey logs in panel (a) and used to estimate variograms. The fluctuation profile (blue) becomes shorter due to edge effects from the median filter. We show logs in terms of slowness because all data analysis is performed in the slowness domain. (c) Variogram estimate from the fluctuation profile (blue) shown in panel (b). The dashed green line is the logarithmic regression used to estimate the Hurst exponent. We estimate the correlation length to be between 40 and 90 m, and $\nu = 0.09$. Note the much larger variance associated with the classical estimate of (Matheron 1963) as opposed with the moving-window estimate of Li & Lake (1994).

assume that the large spikes seen in the logs are due to measurement errors resulting from either decoupling with the logging tool and the host rock or cycle skips, a phenomenon resulting in anomalously high transit times or low velocities. Figs 1(a) and (b) shows the pre-processing method performed on one borehole log, namely apal1 located approximately 5 km north of Long Beach, CA, along with the classical and moving-window variogram (Fig. 1c) estimates for this single borehole. We perform all statistical analysis in the slowness domain. Notice the larger variance associated with the classical estimate of semi-variance. For the remainder of the manuscript we refer to the pre-processed zero-mean logs as fluctuation profiles.

For the inversion, (1) we estimate the semi-variance using eq. (5) for each borehole. (2) We simulate 500 realizations for each combination of parameters (ν, a_z) and estimate the semi-variance for each realization using eq. (5). (3) We compute the mean-squared

Downloaded from http://gji.oxfordjournals.org/ at San Diego State University on March 17, 2016

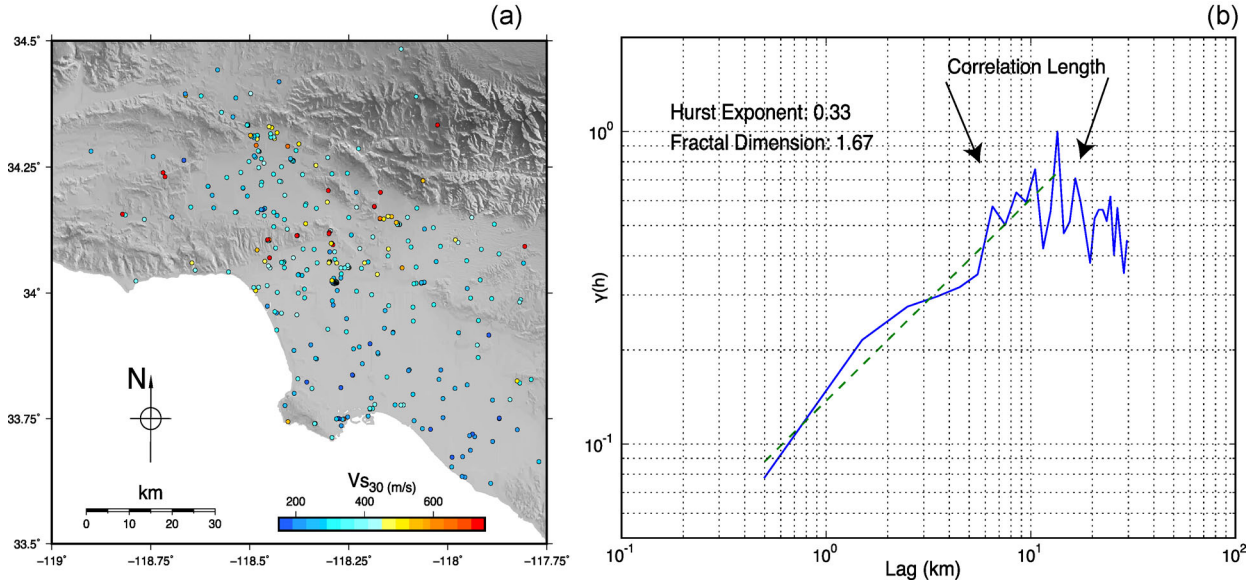

Figure 2. (a) Map showing the 371 Vs_{30} measurements used in our variogram analysis. We colour the Vs_{30} values to provide a visual representation of the spatial variability. (b) Variogram estimated for the Vs_{30} measurements in the Los Angeles basin. The dashed green line shows the logarithmic regression used to estimate the Hurst exponent, $\nu = 0.33$. We estimate lateral correlation lengths between 5 and 10 km.

residual between the borehole variogram and each synthetic variogram. (4) We average the residuals over all realizations to produce a single expected mean-squared residual value for each von Karman parameter pair. (5) We accept any value with a mean-squared residual less than or equal to a tolerance value of 0.01, visually selected from all the inversion results as the value above which the mean-squared residual increases significantly (see Supporting Information Figs S36–S70). This methodology is similar to the inversion routine used by Holliger (1996).

3 STATISTICS OF LOS ANGELES BASIN

First, we analyse the lateral variation of near-surface time-averaged shear wave velocity measurements in greater Los Angeles. We compile 639 Vs_{30} measurements consisting of 350 measurements incorporated in the SCEC CVM-S 4.0 (Magistrale *et al.* 2000), 102 measurements from Louie (2005, 2007) and 187 measurements from Yong *et al.* (2013). Next, we show our inversion results for the von Karman parameters derived from the 35 Vp sonic logs in Los Angeles basin.

3.1 Vs_{30} measurements

Vs_{30} is defined as

$$Vs_{30} = \frac{30\,\mathrm{m}}{\sum \frac{d}{Vs}}, \tag{6}$$

where d is the layer thickness (m) from the 1-D profile below the site of interest, and Vs is the depth-dependent S-wave velocity in (m s^{-1}). We choose a subset of the measurements located near Los Angeles basin resulting in 371 values used in the analysis (see Fig. 2a) to maintain consistency with the spatial extent of the sonic log analysis (see the following section). Since the Vs_{30} measurements are sparse and irregularly located, we assign a relatively large bin spacing for the variogram of 500 m. Also, to obtain a more uniform distribution of observation pairs at all lag distances, we restrict the maximum lag to 30 km.

Fig. 2(b) shows the classical variogram estimated from the 371 Vs_{30} data points. We interpret the correlation length of the heterogeneity in the Vs_{30} values to be approximately 5–10 km (where the variogram approaches a constant value), and the Hurst exponent to be $\nu = 0.33 \pm 0.09$. This analysis provides an understanding of the statistical parameters of larger-scale basin features only, as the Vs_{30} measurements lack the resolution to constrain smaller-scale variation. Incidentally, the relatively large Hurst exponent and correlation lengths estimated from the Vs_{30} data are consistent with the results by Frankel & Clayton (1986), Imperatori & Mai (2013) and Levander *et al.* (1994). We assume the SCEC CVMs have already accounted for these larger-scale variations, and we plan to constrain the parameters of the small-scale variations (on the order of tens of metres) using the densely sampled sonic logs. We note that current state-of-the-art CVMs may artificially taper $Vs30$ values from the near-surface to deeper sections of the models, possibly biasing velocity variability in the deeper sections.

3.2 Sonic log measurements

In order to estimate the small-scale structure in Los Angeles basin, we require much more densely sampled data than that provided by Vs_{30} measurements, as discussed above. Velocity logs provide regularly and densely sampled direct measurements of in situ seismic velocities. We note that our analysis only provides constraints regarding heterogeneity observed in the vertical direction.

We have acquired 38 Vp sonic logs (Andreas Plesch, private communication, 2012) located throughout the Los Angeles basin, extending to more than 3 km depth and containing more than 300 000 measurements shown in Fig. 3. Some logs are missing sections of data, and these values are omitted from the analysis as variograms do not require the data to be evenly spaced. We also discard three boreholes due to significant sections of poor data quality or short sample lengths from which accurate correlation length estimates within our parameter space are not possible. Due to the spatial extent of the boreholes, we assume the boreholes provide a representative sample of the spatial statistics in Los Angeles basin.

Downloaded from http://gji.oxfordjournals.org/ at San Diego State University on March 17, 2016

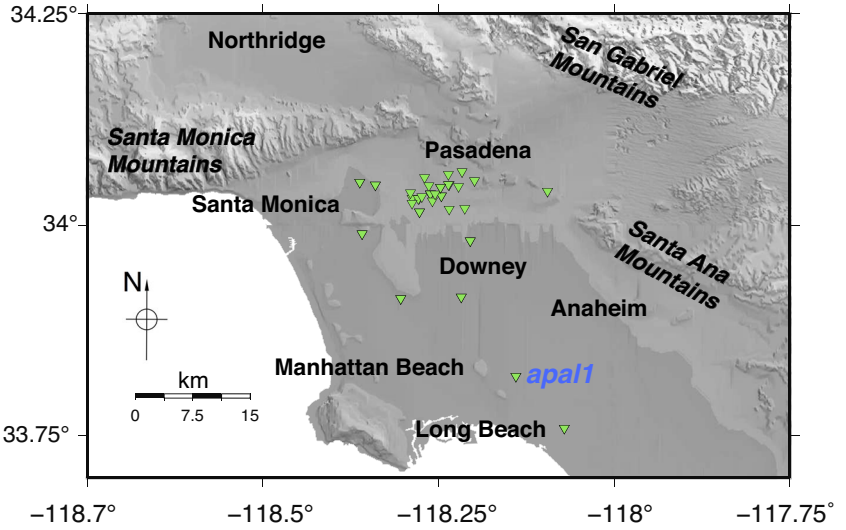

Figure 3. Map showing 35 deep borehole locations within the Los Angeles basin that are used for the inversion. Borehole apal1 is highlighted.

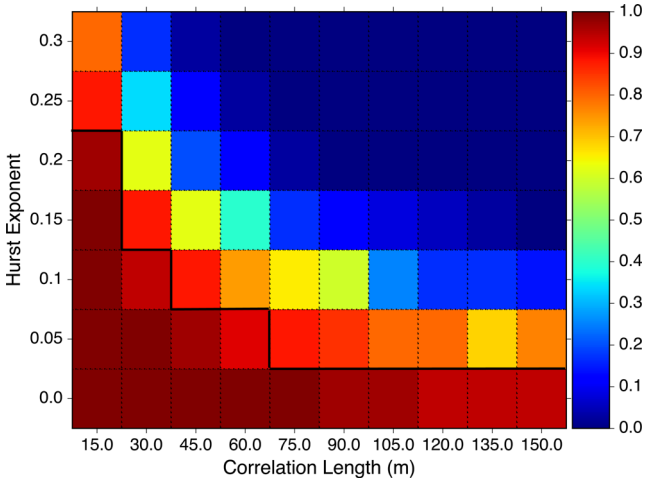

Figure 4. Results shown for the parameter spaced analysed by our inversion routine. Here, we bin all accepted von Karman pairs and normalize by the total number of boreholes. A value of 1.0 indicates that a particular (ν, a_z) pair satisfies the acceptance criteria for every sonic log. Likewise, a value of 0.0 has failed to accept the criterion at every borehole. The von Karman parameters that agree between more than approximately 90 per cent of the borehole sonic logs are defined by (ν, a_z) pairs below the solid line.

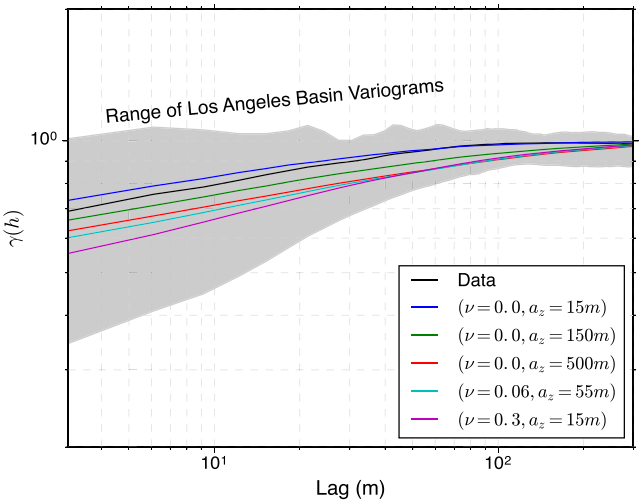

Figure 5. Selected variograms plotted for different combinations of von Karman parameters. The shaded grey region represents the range spanned by all data variograms. We show the variograms representing 100 per cent acceptance as well as the variogram representing the average von Karman parameters for Los Angeles basin. Finally, we show the stacked data variogram computed from an average of each borehole variogram in black. Note the model with $\nu = 0.0$ and $a_z = 500$ m lies within the data variogram range in Los Angeles basin.

Assuming a dense enough network of sonic log measurements, one could also compute horizontal statistics of depth-averaged sections of adjacent boreholes allowing for a much higher resolution lateral study than available from the Vs_{30} data similar to the analysis by Wu *et al.* (1994).

The final inversion result for the 35 boreholes is shown in Fig. 4, which indicates the percentage of boreholes that can be modelled by a given von Karman parameter pair within a mean-squared residual of 0.01. We find that parameters ranging from $\nu = (0.0{-}0.2)$ and $a_z = (15\,\text{m}{-}150\,\text{m})$ represent the variability to an agreement greater than approximately 90 per cent of all 35 Los Angeles Basin sonic logs. Fig. 4 shows a trade-off between ν and a_z meaning that while large a_z are acceptable with low ν and vice-versa, large ν fail to represent the Los Angeles sonic logs even at $a_z = 45$ m. On the other hand, at $\nu = 0.0$, large correlation lengths can provide seemingly reasonable fits to the data. This is because the semi-variance at the nugget (or first lag distance) is approximately equal to the total variance in the sonic logs making the transition from correlated to uncorrelated behaviour hard to distinguish in these variogram estimates. For example, in Fig. 5 we show that a model with $\nu = 0.0$ and $a_z = 500$ m represents the statistics of the sonic logs, where the grey region represents the range of all Los Angeles basin sonic log variograms. This highlights the interdependence between a_z and ν in estimates of semi-variance. Also, we show variograms for several combinations of ν and a_z that could represent reasonable values for Los Angeles Basin. Recent work by Nakata & Beroza (2015) shows that near Long Beach in Los Angeles basin vertical correlation lengths are on the order of 100 m with an anisotropy factor, a_x/a_z of approximately 5. Based on a bootstrapped analysis of the average values calculated for each borehole, we find average values of $\nu = 0.064\,(0.058, 0.069) \pm 0.01\,(0.006, 0.012)$ and $a_z = 54\,(51.1, 57.6)\,\text{m} \pm 5.9\,(1.79, 9.53)\,\text{m}$, where the

Downloaded from http://gji.oxfordjournals.org/ at San Diego State University on March 17, 2016

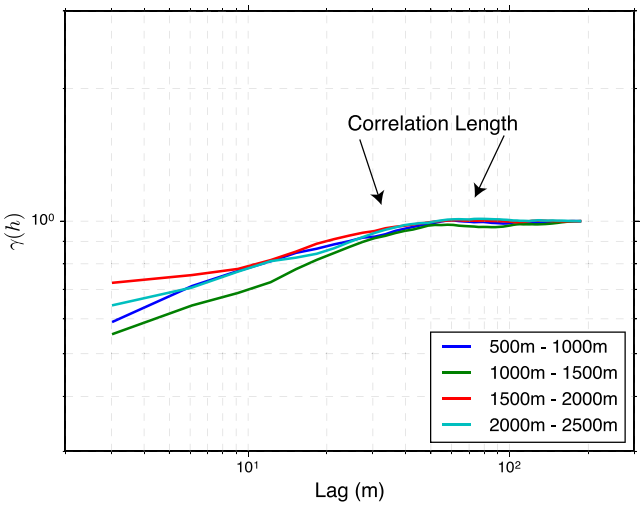

Figure 6. Depth-dependent variogram computed for 500 m windows of the sonic log data set. We find no significant depth-dependence of the von Karman parameters. Each curve represents the stacked variogram over each specified depth interval.

95 per cent confidence intervals for the mean and standard deviation are shown in the parentheses. We compute the average values by applying a weighting function to the inversion residuals such that models with lower residual values are weighted higher. The electronic supplement shows the mean-squared residuals computed for each borehole. We also show the raw and processed versions of the boreholes used in the inversion along with moving-window and classical semi-variance estimates.

Finally, we analyse the data set to determine if there are any depth-dependent effects. Towards this goal, we window the data set in 500 m segments and apply traditional variogram analysis to the windowed data, meaning no inversion is performed on the depth-windowed data. For this analysis, we pre-process the logs as described in the Methods section and stack the resulting moving-window variogram estimates to produce a single variogram representative of a given depth-interval. Fig. 6 shows the results from our analysis. This figure suggests the lack of any significant depth dependency for the von Karman parameters. We also find no depth-dependence for σ. Even though there might be depth-dependency for some of the parameters at individual boreholes (e.g. $\sigma(z)$ in apal1 shown in Fig. 1), on average, our analysis suggests limited depth-dependence of the von Karman parameters in Los Angeles basin.

4 CONCLUSIONS

We have used 371 Vs_{30} measurements for Southern California and 35 borehole sonic logs from the Los Angeles basin to constrain the parameters describing von Karman distributions of small-scale heterogeneities. The borehole logs indicate that the ranges of ν and a most representative of the velocity variability are $\nu = 0.0$–0.2, $a_z = 15$–105 m and $\sigma = 5 \pm 2.5$ per cent. We find average values of $\nu = 0.064\,(0.058, 0.069) \pm 0.01\,(0.006, 0.012)$ and $a_z = 54\,(51, 57)\,\text{m} \pm 5.9\,(1.0, 9.5)\,\text{m}$ with 95 per cent confidence intervals listed in the parentheses. The Vs_{30} data provide only an upper bound on the parameters ($\nu = 0.33 \pm 0.09$ and $a_x = 5$–10 km) due to the relatively coarse sampling. We find insignificant depth dependency of the von Karman parameters from 500 m to 2500 m depth. These statistical parameters constitute a data-driven model that can be used for high-frequency simulations of seismic wave propagation in Southern California. These two separate data sets constrain the vertical and lateral features, respectively. We hesitate to infer an anisotropy ratio between the two data sets due to the large differences in resolution between the Vs_{30} and borehole data. We suspect that higher resolution Vs_{30} data, on say a 50 m grid, may help refine the results presented here.

The parameters estimated from the sonic log analysis section of the manuscript only provide information regarding the *P*-wave heterogeneity in Los Angeles basin. Directly applying these parameters to the *S*-wave data implies some correlation between these two measurements. Based on simple elastic theory and analysis of Vp and Vs sonic logs (Dolan *et al.* 1998), we expect some correlation between the two sets of parameters as they are both dependent on the shear modulus and the density of the material. However, without access to full-sonic scanner measurements (i.e. *P*-wave and *S*-wave logs) at the same locations, we cannot definitively comment on the correlation between *P*-wave and S-wave heterogeneity for this data set. The level of correlation does have implications for modelling of the small scale heterogeneities, as uncorrelated or non-perfectly correlated heterogeneity models must simulate separate realizations of the stochastic field for both *P*-wave and *S*-wave velocities.

Our inversion procedure provides limited resolution of a_z for $\nu \approx 0.0$. This means larger correlation lengths may accurately model the spatial statistics observed in Los Angeles basin. Also, we find an interdependence between ν and a_z for all combinations of von Karman parameters. Care must be taken to choose appropriate parameters of ν and a_z that fall within the accepted range for Los Angeles basin. The black line in Fig. 4 shows the values of ν and a_z that represent accepted von Karman parameters for the majority of borehole sonic logs in Los Angeles basin. We recommend choosing parameters that fall within this region, however one could argue for choosing large values of a_z for $\nu = 0.0$.

ACKNOWLEDGEMENTS

We thank Andreas Plesch and John Shaw for providing the deep borehole logs, and to John Louie, Harold Magistrale and Alan Yong for providing us with Vs_{30} measurements throughout Los Angeles basin. We are grateful for stimulating comments from Robert Graves and an anonymous reviewer that helped strengthen the manuscript. This research was supported by NSF awards OCI-1148493 (SI2-SSI), EAR-1135455 (FESD) and EAR-1226343 EAR-1349180, the Southern California Earthquake Center (SCEC) and U.S. Geological Survey award G12AP20015. SCEC is funded by NSF Cooperative Agreement EAR-1033462 and USGS Cooperative Agreement G12AC20038. This is SCEC contribution #6178. Both the Vs_{30} and borehole data used in this publication can be found at http://www-rohan.sdsu.edu/~kbolsen/vonkarman.html. The code used for this report is available upon request from authors.

REFERENCES

Anderson, J.G., 2015. The composite source model for broadband simulations of strong ground motions, *Seismol. Res. Lett.,* **86**(1), 68–74.

Bydlon, S.A. & Dunham, E.M., 2015. Rupture dynamics and ground motions from earthquakes in 2-D heterogeneous media, *Geophys. Res. Lett.,* **42,** 1701–1709.

Crempien, J.G.F. & Archuleta, R.J., 2015. UCSB method for simulation of broadband ground motion from kinematic earthquake sources, *Seismol. Res. Lett.,* **86**(1), 61–67.

Downloaded from http://gji.oxfordjournals.org/ at San Diego State University on March 17, 2016

Cui, Y. *et al.*, 2013. Physics-based seismic hazard analysis on petascale heterogeneous supercomputers, in *Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis,* p. 70, ACM.

De, G.S., Winterstein, D.F. & Meadows, M.A., 1994. Comparison of *P*- and *S*-wave velocities and *Q*'s from VSP and sonic log data, *Geophysics,* **59**(10), 1512–1529.

Dolan, S., Bean, C. & Riollet, B., 1998. The broad-band fractal nature of heterogeneity in the upper crust from petrophysical logs, *Geophys. J. Int.,* **132**(3), 489–507.

Frankel, A., 2009. A constant stress-drop model for producing broadband synthetic seismograms: comparison with the next generation attenuation relations, *Bull. seism. Soc. Am.,* **99**(2A), 664–680.

Frankel, A. & Clayton, R.W., 1986. Finite difference simulations of seismic scattering: implications for the propagation of short-period seismic waves in the crust and models of crustal heterogeneity, *J. geophys. Res.,* **91**(B6), 6465–6489.

Graves, R. & Pitarka, A., 2015. Refinements to the Graves and Pitarka (2010) broadband ground-motion simulation method, *Seismol. Res. Lett.,* **86**(1), 81–88.

Graves, R.W., Aagaard, B.T., Hudnut, K.W., Star, L.M., Stewart, J.P. & Jordan, T.H., 2008. Broadband simulations for M_w 7.8 Southern San Andreas earthquakes: ground motion sensitivity to rupture speed, *Geophys. Res. Lett.,* **35**(22), L22302, doi:10.1029/2008GL035750.

Hartzell, S., Harmsen, S. & Frankel, A., 2010. Effects of 3D random correlated velocity perturbations on predicted ground motions, *Bull. seism. Soc. Am.,* **100**(4), 1415–1426.

Holliger, K., 1996. Upper-crustal seismic velocity heterogeneity as derived from a variety of P-wave sonic logs, *Geophys. J. Int.,* **125**(3), 813–829.

Holliger, K., 1997. Seismic scattering in the upper crystalline crust based on evidence from sonic logs, *Geophys. J. Int.,* **128**(1), 65–72.

Holliger, K. & Levander, A.R., 1992. A stochastic view of lower crustal fabric based on evidence from the Ivrea Zone, *Geophys. Res. Lett.,* **19**(11), 1153–1156.

Imperatori, W. & Mai, P.M., 2013. Broad-band near-field ground motion simulations in 3-dimensional scattering media, *Geophys. J. Int.,* **192**(2), 725–744.

Jacobsen, B.H. & Olsen, K.B., 2011. Spatial variability of ground motion amplification from low-velocity sediments including fractal inhomogeneities with special reference to the Southern California basins, *EGU General Assembly 2011*.

Kohler, M.D., Magistrale, H. & Clayton, R.W., 2003. Mantle heterogeneities and the SCEC reference three-dimensional seismic velocity model version 3, *Bull. seism. Soc. Am.,* **93**(2), 757–774.

Lee, E.J., Chen, P. & Jordan, T.H., 2014. Testing waveform predictions of 3D velocity models against two recent Los Angeles earthquakes, *Seismol. Res. Lett.,* **85**(6), 1275–1284.

Levander, A.R., England, R.W., Smith, S.K., Hobbs, R.W., Goff, J.A. & Holliger, K., 1994. Stochastic characterization and seismic response of upper and middle crustal rocks based on the Lewisian Gneiss Complex, Scotland, *Geophys. J. Int.,* **119**(1), 243–259.

Li, D. & Lake, L.W., 1994. A moving window semivariance estimator, *Water Resour. Res.,* **30**(5), 1479–1489.

Louie, J.N., 2005. Improving next-generation attenuation models with shear-velocity measurements at all TriNet and strong-motion station in LA, NEHRP Final Technical Report.

Louie, J.N., 2007. Shear-wave velocity map for California: collaborative research with CGS and UNR, NEHRP Final Technical Report.

Magistrale, H., Day, S.M., Clayton, R.W. & Graves, R.W., 2000. The SCEC Southern California reference three-dimensional seismic velocity model version 2, *Bull. seism. Soc. Am.,* **90**(6B), S65–S76.

Matheron, G., 1963. Principles of geostatistics, *Econ. Geol.,* **58**(8), 1246–1266.

Mela, K. & Louie, J.N., 2001. Correlation length and fractal dimension interpretation from seismic data using variograms and power spectra, *Geophysics,* **66**(5), 1372–1378.

Nakata, N. & Beroza, G.C., 2015. Stochastic characterization of mesoscale seismic velocity heterogeneity in Long Beach, California, *Geophys. J. Int.,* **203**(3), 2049–2054.

Olsen, K. *et al.*, 2006. Strong shaking in Los Angeles expected from Southern San Andreas earthquake, *Geophys. Res. Lett.,* **33**(7), doi:10.1029/2005GL025472.

Olsen, K. *et al.*, 2009. ShakeOut-D: Ground motion estimates using an ensemble of large earthquakes on the Southern San Andreas fault with spontaneous rupture propagation, *Geophys. Res. Lett.,* **36**(4), doi:10.1029/2008GL036832.

Olsen, K.B. & Takedatsu, R., 2015. The SDSU broadband ground-motion generation module BBtoolbox version 1.5, *Seismol. Res. Lett.,* **86**(1), 75–80.

Olsen, K.B., Stephenson, W.J. & Geisselmeyer, A., 2008. 3D crustal structure and long-period ground motions from a M9. 0 megathrust earthquake in the Pacific Northwest region, *J. Seismol.,* **12**(2), 145–159.

Pullammanappallil, S., Levander, A.R. & Larkin, S.P., 1997. Estimation of crustal stochastic parameters from seismic exploration data, *J. geophys. Res.,* **102**(B7), 15 269–15 286.

Roten, D., Olsen, K., Pechmann, J., Cruz-Atienza, V. & Magistrale, H., 2011. 3D simulations of M 7 earthquakes on the Wasatch fault, Utah, Part I: Long-period (0–1 Hz) ground motion, *Bull. seismol. Soc. Am.,* **101**(5), 2045–2063.

Roten, D., Olsen, K.B., Day, S.M., Cui, Y. & Faeh, D., 2014. Expected seismic shaking in Los Angeles reduced by San Andreas fault zone plasticity, *Geophys. Res. Lett.,* **41**(8), 2769–2777.

Shi, Z. & Day, S.M., 2013. Rupture dynamics and ground motion from 3-D rough-fault simulations, *J. geophys. Res.,* **118**(3), 1122–1141.

Süss, M.P. & Shaw, J.H., 2003. P wave seismic velocity structure derived from sonic logs and industry reflection data in the Los Angeles basin, California, *J. geophys. Res.,* **108**(B3), 2170, doi:10.1029/2001JB001628.

Taborda, R. & Bielak, J., 2013. Ground-motion simulation and validation of the 2008 Chino Hills, California, earthquake, *Bull. seism. Soc. Am.,* **103**(1), 131–156.

Wu, R.S., Xu, Z. & Li, X.-P., 1994. Heterogeneity spectrum and scale-anisotropy in the upper crust revealed by the German Continental Deep-Drilling (KTB) Holes, *Geophys. Res. Lett.,* **21**(10), 911–914.

Yong, A., Martin, A., Stokoe, K. & Diehl, J., 2013. ARRA-funded Vs30 measurements using multi-technique approach at strong-motion stations in California and central-eastern United States, U.S. Geological Survey Open-File Report.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this paper:

Figure S1: (top left) We show the raw sonic log in grey with long-period moving window filter (black) and short-period median filtered log in red. Here, the red (demeaned and smoothed) log represents the input to the variogram analysis. (top right) We plot the input to the variogram as a stand-alone curve. We find there to be no systematic variability as a function of depth. (middle left) We show histograms computed from the fluctuation profiles and (middle right) a probability plot showing a comparison against a normal distribution. We find the fluctuations to be approximately normally distributed. (bottom) We show the variogram estimates for each borehole, where the black curve shows the moving-window estimate (Li & Lake 1994) and the red curve shows the classical estimate (Matheron 1963).

Figure S2: Same as Fig. S1, except for borehole arsj_11r.

Figure S3: Same as Fig. S1, except for borehole arsj_14.

Figure S4: Same as Fig. S1, except for borehole arsj_16.

Figure S5: Same as Fig. S1, except for borehole arsj_18.

Downloaded from http://gji.oxfordjournals.org/ at San Diego State University on March 17, 2016

Figure S6: Same as Fig. S1, except for borehole arsj_20.
Figure S7: Same as Fig. S1, except for borehole arsj_5.
Figure S8: Same as Fig. S1, except for borehole arsj_8.
Figure S9: Same as Fig. S1, except for borehole arwa_1.
Figure S10: Same as Fig. S1, except for borehole csrc_1.
Figure S11: Same as Fig. S1, except for borehole ope_1.
Figure S12: Same as Fig. S1, except for borehole rowa_2.
Figure S13: Same as Fig. S1, except for borehole soag1-13.
Figure S14: Same as Fig. S1, except for borehole sobc_1.
Figure S15: Same as Fig. S1, except for borehole sobd_1a.
Figure S16: Same as Fig. S1, except for borehole sobudd_1.
Figure S17: Same as Fig. S1, except for borehole socc_1.
Figure S18: Same as Fig. S1, except for borehole soccs_1.
Figure S19: Same as Fig. S1, except for borehole soech_1.
Figure S20: Same as Fig. S1, except for borehole sofc_1.
Figure S21: Same as Fig. S1, except for borehole sofcch_1.
Figure S22: Same as Fig. S1, except for borehole soflc_1.
Figure S23: Same as Fig. S1, except for borehole sog_4a.
Figure S24: Same as Fig. S1, except for borehole sog_6.
Figure S25: Same as Fig. S1, except for borehole sogc_1.
Figure S26: Same as Fig. S1, except for borehole sokc_1.
Figure S27: Same as Fig. S1, except for borehole sokc_1r.
Figure S28: Same as Fig. S1, except for borehole sopc_1.
Figure S29: Same as Fig. S1, except for borehole sosa_1.
Figure S30: Same as Fig. S1, except for borehole souc_1.
Figure S31: Same as Fig. S1, except for borehole sovc_1.
Figure S32: Same as Fig. S1, except for borehole sowc_1.
Figure S33: Same as Fig. S1, except for borehole tohe_1.
Figure S34: Same as Fig. S1, except for borehole tuo_2.
Figure S35: Same as Fig. S1, except for borehole uouspe_1.
Figure S36: We show the average mean-squared residual over 500 realizations for each von Karman parameter pair (ν; a_z). The grey shaded region shows the acceptance criterion for the inversion (0.01), visually selected from all the inversion results as the value above which the mean-squared residual increases significantly. The green curve represents the inversion results using a classical estimate of the semi-variance while the blue curve represents the moving-window estimates. We see the overall trend does not change between the two semi-variance models, but the moving-window variogram provides a more consistent result allowing for the introduction of an acceptance criterion.
Figure S37: Same as Fig. S36, except for borehole arsj_11.
Figure S38: Same as Fig. S36, except for borehole arsj_11r.
Figure S39: Same as Fig. S36, except for borehole arsj_14.
Figure S40: Same as Fig. S36, except for borehole arsj_16.
Figure S41: Same as Fig. S36, except for borehole arsj_18.
Figure S42: Same as Fig. S36, except for borehole arsj_20.
Figure S43: Same as Fig. S36, except for borehole arsj_5.
Figure S44: Same as Fig. S36, except for borehole arsj_8.
Figure S45: Same as Fig. S36, except for borehole arwa_1.
Figure S46: Same as Fig. S36, except for borehole csrc_1.
Figure S47: Same as Fig. S36, except for borehole ope_1.
Figure S48: Same as Fig. S36, except for borehole rowa_2.
Figure S49: Same as Fig. S36, except for borehole soag1-13.
Figure S50: Same as Fig. S36, except for borehole sobc_1.
Figure S51: Same as Fig. S36, except for borehole sobd_1a.
Figure S52: Same as Fig. S36, except for borehole sobudd_1.
Figure S53: Same as Fig. S36, except for borehole socc_1.
Figure S54: Same as Fig. S36, except for borehole soccs_1.
Figure S55: Same as Fig. S36, except for borehole soech_1.
Figure S56: Same as Fig. S36, except for borehole sofc_1.
Figure S57: Same as Fig. S36, except for borehole sofcch_1.
Figure S58: Same as Fig. S36, except for borehole soflc_1.
Figure S59: Same as Fig. S36, except for borehole sog_4a.
Figure S60: Same as Fig. S36, except for borehole sog_6.
Figure S61: Same as Fig. S36, except for borehole sogc_1.
Figure S62: Same as Fig. S36, except for borehole sokc_1.
Figure S63: Same as Fig. S36, except for borehole sokc_1r.
Figure S64: Same as Fig. S36, except for borehole sopc_1.
Figure S65: Same as Fig. S36, except for borehole sosa_1.
Figure S66: Same as Fig. S36, except for borehole sosan_1.
Figure S67: Same as Fig. S36, except for borehole souc_1.
Figure S68: Same as Fig. S36, except for borehole sovc_1.
Figure S69: Same as Fig. S36, except for borehole sowc_1.
Figure S70: Same as Fig. S36, except for borehole tohe_1.
(http://gji.oxfordjournals.org/lookup/suppl/doi:10.1093/gji/ggw050/-/DC1)

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the paper.

Downloaded from http://gji.oxfordjournals.org/ at San Diego State University on March 17, 2016