Pure Appl. Geophys. 173 (2016), 3671–3692

DOI 10.1007/s00024-016-1357-1

Pure and Applied Geophysics

Generating Random Earthquake Events for Probabilistic Tsunami Hazard Assessment

Randall J. LeVeque,[1] Knut Waagan,[2] Frank I. González,[3] Donsub Rim,[1] and Guang Lin[4]

Abstract—To perform probabilistic tsunami hazard assessment for subduction zone earthquakes, it is necessary to start with a catalog of possible future events along with the annual probability of occurrence, or a probability distribution of such events that can be easily sampled. For near-field events, the distribution of slip on the fault can have a significant effect on the resulting tsunami. We present an approach to defining a probability distribution based on subdividing the fault geometry into many subfaults and prescribing a desired covariance matrix relating slip on one subfault to slip on any other subfault. The eigenvalues and eigenvectors of this matrix are then used to define a Karhunen-Loève expansion for random slip patterns. This is similar to a spectral representation of random slip based on Fourier series but conforms to a general fault geometry. We show that only a few terms in this series are needed to represent the features of the slip distribution that are most important in tsunami generation, first with a simple one-dimensional example where slip varies only in the down-dip direction and then on a portion of the Cascadia Subduction Zone.

Key words: Probabilistic tsunami hazard assessment, seismic sources, Karhunen-Loève expansion, subduction zone earthquakes.

1. Introduction

Computer simulation of tsunamis resulting from subduction zone earthquakes can be performed using a variety of available software packages, most of which implement the two-dimensional shallow water equations and require the vertical seafloor motion resulting from the earthquake as the input to initiate the waves. For recent past events, this can be approximated based on source inversions; one example is shown in Fig. 1. However, there are several situations in which it is desirable to instead generate hypothetical future earthquakes. In particular, recent work on probabilistic tsunami hazard assessment (PTHA) has focused on producing maps that indicate the annual probability of flooding exceeding various depths and can provide much more information than a single "worst considered case" inundation map (for example, Adams et al. 2015; Geist and Parsons 2006; Geist et al. 2009; Goda et al. 2015; González et al. 2009; Jaimes et al. 2016; Løvholt et al. 2012; Witter et al. 2013). This requires running tsunami simulations for many potential earthquakes and combining the results based on the annual probability of each, or using a Monte Carlo approach to sample a presumed probability density of potential earthquakes. Generating a large number of hypothetical events can also be useful for testing inversion methods that incorporate tsunami data, such as the current DART buoy network (e.g., Dettmer et al. 2016), or that might give early tsunami warnings for the nearshore (e.g., Melgar et al. 2016a). Both probabilistic seismic hazard assessment (PSHA) and PTHA are also fundamental tools in the development of building codes that are critical in the design of structures able to withstand seismic and tsunami forces (e.g., Chock 2015).

The primary goal of this paper is to introduce a general approach to generating hypothetical rupture scenarios for possible future earthquakes, by producing random slip patterns on a pre-specified fault geometry. Similar techniques have been used in past studies, particularly for the generation of seismic waves in PSHA, which has a longer history than PTHA. A variety of techniques have been proposed for generating random seismic ruptures, see for example studies by Anderson (2015), Dreger et al. (2015), Frankel (1991), Graves and Pitarka (2010), Goda et al. (2014), Guatteri et al. (2003), Lavallée

[1] Department of Applied Mathematics, University of Washington, Seattle, WA, USA. E-mail: rjl@uw.edu
[2] Forsvarets Forskningsinstitutt, Oslo, Norway.
[3] Department of Earth and Space Sciences, University of Washington, Seattle, WA, USA.
[4] Department of Mathematics, Purdue University, West Lafayette, IN, USA.

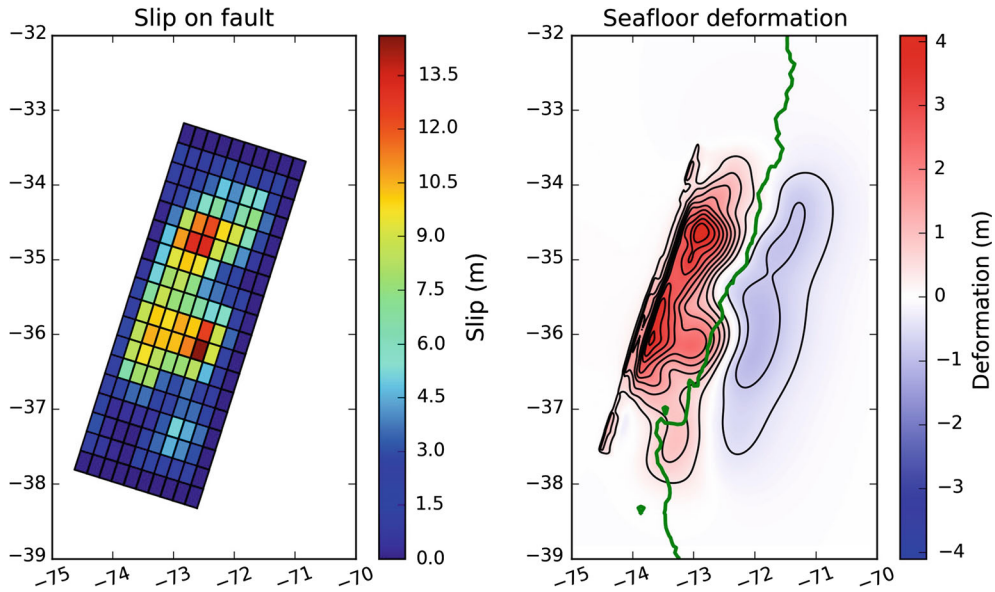


Figure 1
An example of slip distributed on a fault plane, from the USGS inversion of the 27 February 2010 event off Maule, Chile (USGS 2010). The plot on the *right* shows the resulting sea floor deformation computed using the Okada solution to the elastic half-space problem, with the coast line in *green*

et al. (2006), Mai and Beroza (2002). One approach is to use a spectral representation of the slip pattern as a Fourier series with random coefficients that decay at a specified rate based on the desired smoothness and correlation length of the slip patterns, e.g., as estimated from past events in the work of Mai and Beroza (2002). Different correlation lengths can be specified in the strike and slip directions, if these directions are used as the horizontal coordinates in the Fourier representation and the fault is roughly rectangular, as done for example by Mai and Beroza (2002), Graves and Pitarka (2010).

Our approach is very similar on a rectangular fault but generalizes easily to other fault geometries using a Karhunen–Loève expansion. This work was motivated in particular by the need to model events on the curving Cascadia Subduction Zone (CSZ), which lies offshore North America and runs nearly 1200 km from Northern California up to British Columbia, see Fig. 2.

The fault is subdivided into many rectangular subfaults and a value of the slip s_i is assigned to the ith subfault. If here are N subfaults, then this defines a vector $\boldsymbol{s} \in \mathbb{R}^N$. Initially we assume that the moment magnitude M_w of the earthquake (which depends on the total slip summed over all subfaults) has been prescribed, and also that the desired mean slip $\boldsymbol{\mu} \in \mathbb{R}^N$ and covariance matrix $\hat{\boldsymbol{C}} \in \mathbb{R}^{N\times N}$ are known. The mean slip is a vector with components $\mu_i = E[s_i]$, the expected value of the slip on the ith subfault, and the $N \times N$ covariance matrix $\hat{\boldsymbol{C}}$ has components $\hat{C}_{ij} = E[(s_i - \mu_i)(s_j - \mu_j)]$, which can also be expressed as the outer product $\hat{\boldsymbol{C}} = E[(\boldsymbol{s} - \boldsymbol{\mu})(\boldsymbol{s} - \boldsymbol{\mu})^T]$, where T denotes transposing the vector.

The Karhunen–Loève (K-L) expansion (e.g., Ghanem and Spanos 1991; Huang et al. 2001; Karhunen 1947; Loève 1977; Schwab and Todor 2006) is a standard approach to representing a Gaussian random field as a linear combination of eigenvectors of the presumed covariance matrix $\hat{\boldsymbol{C}}$. If the matrix $\hat{\boldsymbol{C}}$ has eigenvalues λ_k (ordered with $\lambda_0 > \lambda_1 > \cdots > 0$) and corresponding eigenvectors $\boldsymbol{v}_k$, then the K-L expansion expresses the slip vector $\boldsymbol{s}$ as

$$\boldsymbol{s} = \boldsymbol{\mu} + \sum_{k=1}^{N} z_k \sqrt{\lambda_k} \boldsymbol{v}_k. \tag{1}$$

where the z_k are independent normally distributed random numbers $z_k \sim \mathcal{N}(0, 1)$ with mean 0 and

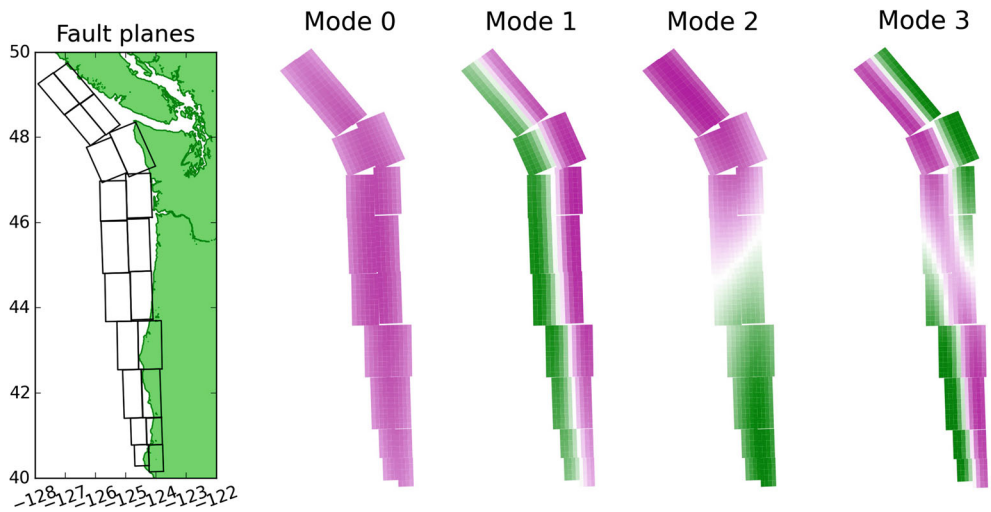


Figure 2
Subdivision of the Cascadia Subduction Zone into 20 subfaults, following Pollitz et al. (2010). These are further divided into 865 subfaults to compute the modes shown, which are the first four eigenvectors of the 865 × 865 covariance matrix as might be used in a Karhunen–Lòeve expansion. *Magenta* and *green* are used to indicate positive and negative entries in the eigenmodes

standard deviation 1. This is described in more detail in Sect. 2 where we explain why this gives random slip patterns with the desired mean and covariance. This expansion makes it easy to generate an arbitrary number of realizations using standard software to generate $\mathcal{N}(0,1)$ random numbers.

Figure 2 shows an example of the first four eigenmodes for the CSZ using this approach, where the N components of each eigenvector are represented on the fault geometry using a color map in which magenta is positive and green is negative. Note that Mode 0 is roughly constant over the fault, so adding a multiple of this mode modifies the total slip and hence the magnitude M_w. On the other hand, the other modes have both positive and negative regions and so adding a multiple of any of these tends to redistribute the slip (e.g., up-dip/down-dip with Mode 1 or between north and south with Mode 2). As with Fourier series, higher order eigenmodes are more oscillatory.

If the presumed correlation lengths are long and the covariance is a sufficiently smooth function of the distance between subfaults, then the eigenvalues λ_k decay rapidly (there is little high-frequency content) and so the K-L series can often be truncated to only a few terms, greatly reducing the dimension of the stochastic space that must be explored.

The K-L series approach could also be used to generate random slip patterns for generating seismic waves, e.g., for performing PSHA or testing seismic inversion algorithms. In this case, high-frequency components of the slip are very important and the K-L expansion may not decay so quickly. However, for tsunami modeling applications the slip pattern on the fault is only used to generate the resulting seafloor deformation. This is a smoothing operation that suppresses high frequencies. In this paper, we also explore this effect and show that truncating the expansion to only a few terms may be sufficient for many tsunami applications. Reducing the dimension of the stochastic space is important for efficient application of many sampling techniques that could be used for PTHA analysis.

In this paper, we focus on explaining the key ideas in the context of a one-dimensional fault model (with variation in slip only in the down-dip direction) and a two-dimensional example using the southern portion of the CSZ. However, we do not claim to have used the optimal parameters for modeling this particular fault. We also do not fully explore PTHA applications here, and for illustration we use some quantities of interest related to a tsunami that are easy to compute from a given slip realization, rather than performing a full tsunami simulation for each. This

allows us to explore the statistics obtained from a large number of realizations (20,000) to illustrate some possible applications of this approach and explore the effects of truncating the K-L series. Work is underway to model the CSZ in a realistic manner and to couple this approach with a full tsunami model.

The K-L expansion as described above generates a Gaussian random field, in which each subfault slip s_i has a normal distribution with mean μ_i and variance $\hat{C}_{ii}$ and together they have a joint normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\hat{\boldsymbol{C}}$. A potential problem with this representation is that when the variance is large it is possible for the slip s_i to be negative on some subfaults. Since we assume the rake is constant (e.g., 90° for a subduction thrust event), this would correspond to subfaults that are slipping in the wrong direction. The same issue arises with Fourier series representations and can be dealt with by various means, for example, by simply setting the slip to zero anyplace it is negative (and then rescaling to maintain the desired magnitude). This naturally changes the statistics of the resulting distributions.

Another approach is to instead posit that the random slip can be modeled by a joint lognormal distribution, for which the probability of negative values is zero. Random slip patterns with a joint lognormal distribution can be generated using the K-L expansion to first compute a Gaussian field and then exponentiating each component of the resulting vector to obtain the slip on each subfault. By choosing the mean $\boldsymbol{\mu}^g$ and covariance matrix $\hat{\boldsymbol{C}}^g$ for the Gaussian field properly, the resulting lognormal will have the desired mean $\boldsymbol{\mu}$ and $\hat{\boldsymbol{C}}$ for the slip. This is discussed in Sect. 4 and used in the two-dimensional example in Sect. 5.

2. Expressing Slip Using a Karhunen–Loève Expansion

If the earthquake fault is subdivided into N small rectangular subfaults, then a particular earthquake realization can be described by specifying the slip on each subfault, i.e., by a vector $\boldsymbol{s} \in \mathbb{R}^N$ where s_i is the slip on the ith subfault. Note that we are assuming that only the slip varies from one realization to another; the geometry and rake (direction of slip on each subfault) are fixed, and the slip is instantaneous and not time dependent. These restrictions could be relaxed at the expense of additional dimensions in our space of realizations.

Initially, assume we wish to specify that the slip is a Gaussian random field with desired mean slip $\boldsymbol{\mu} \in \mathbb{R}^N$ and covariance matrix $\hat{\boldsymbol{C}} \in \mathbb{R}^{N\times N}$, which we write as $\boldsymbol{s} \sim \mathcal{N}(\boldsymbol{\mu}, \hat{\boldsymbol{C}})$. Then we compute the eigenvalues λ_k of $\hat{\boldsymbol{C}}$ and corresponding normalized eigenvectors $\boldsymbol{v}_k$ so that the matrix of eigenvectors $\boldsymbol{V}$ (with kth column $\boldsymbol{v}_k$) and the diagonal matrix of eigenvalues $\boldsymbol{\Lambda}$ satisfy $\hat{\boldsymbol{C}} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^T$. Note that the covariance matrix is symmetric positive definite, so the eigenvalues are always positive real numbers and the eigenvectors can be chosen to be orthonormal, $\boldsymbol{V}^{-1} = \boldsymbol{V}^T$.

Then the K-L expansion (1) can be written in matrix-vector form as

$$\boldsymbol{s} = \boldsymbol{\mu} + \boldsymbol{V}\boldsymbol{\Lambda}^{1/2}\boldsymbol{z}, \tag{2}$$

where $\boldsymbol{z} \in \mathbb{R}^N$ is a vector of independent identically distributed $\mathcal{N}(0,1)$ random numbers. Realizations generated via the K-L expansion have the right statistics since we can easily compute that $E[\boldsymbol{s}] = \boldsymbol{\mu}$ (since $E[\boldsymbol{z}] = \boldsymbol{0}$) and

$$\begin{aligned} E[(\boldsymbol{s}-\boldsymbol{\mu})(\boldsymbol{s}-\boldsymbol{\mu})^T] &= E[\boldsymbol{V}\boldsymbol{\Lambda}^{1/2}\boldsymbol{z}\boldsymbol{z}^T\boldsymbol{\Lambda}^{1/2}\boldsymbol{V}^T] \\ &= \boldsymbol{V}\boldsymbol{\Lambda}^{1/2}E[\boldsymbol{z}\boldsymbol{z}^T]\boldsymbol{\Lambda}^{1/2}\boldsymbol{V}^T \\ &= \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^T = \hat{\boldsymbol{C}} \end{aligned} \tag{3}$$

using the fact that $\boldsymbol{V}$ and $\boldsymbol{\Lambda}$ are fixed and $E[\boldsymbol{z}\boldsymbol{z}^T] = \boldsymbol{I}$. Note that the $\boldsymbol{z}$ could be chosen from a different probability density with mean $\boldsymbol{0}$ and covariance matrix $\boldsymbol{I}$ and achieve the same covariance matrix $\hat{\boldsymbol{C}}$ with the K-L expansion, although the $\boldsymbol{s}$ would not have a joint normal distribution in this case.

3. One-Dimensional Case: Down-Dip Variation

We first illustrate this technique on a simplified case, a rectangular fault plane that is essentially infinitely long in the strike direction and with uniform slip in that direction, similar to the test case used by Løvholt et al. (2012). The slip will only vary in the

down-dip direction, reducing the problem to a single space dimension. The fault width is 100 km, a typical width for subduction zone faults, and is assumed to dip at 13° from horizontal, with the upper edge at a depth of 5 km below the sea floor.

For the tests we perform here, we will focus on events of a single specified magnitude. The moment magnitude M_w is a function of the total slip integrated over the entire fault plane, and also depends on the rigidity of the rock. For typical rigidity parameters, an average of 10 m of slip distributed over a fault that is 100 km wide and 1000 km long would result in a magnitude $M_w \approx 9.0$ and so we fix the total slip to have this average. If the fault were only half as long, 500 km, then this would be a $M_w \approx 8.8$ event and 20 m average slip would be required for a magnitude 9 event. With the exception of the potential energy, the quantities of interest considered in this paper are all linear in the total slip, however, it does not really matter what value we choose.[1]

An important aspect of PTHA analysis is to consider possible events of differing magnitudes as well, and take into account their relative probabilities. For smaller earthquakes, the Gutenberg–Richter relation approximately describes their relative frequency, but for large subduction zone events that may have a recurrence time of hundreds of years, there is generally no simple model for the variation of annual probability with magnitude. There may be a continuous distribution of magnitudes or there may be certain "characteristic earthquakes" that happen repeatedly after sufficient stress has built up. The lack of detailed data for past events over a long time period makes this difficult to assess.

For the purposes of this paper, we assume that an earthquake of a particular magnitude occurs and we wish to model the range of possible tsunamis that can arise from such an event. We thus discuss the relative probability of different slip patterns and tsunamis given that an event of this magnitude occurs, and so the probability density should integrate to 1. This could then be used as one component in a full PTHA analysis by weighting these results by the probability that an event of this magnitude occurs and combining with similar results for other magnitudes. Alternatively, one could introduce the magnitude as an additional stochastic dimension and assume some probability density function for this.

We use x to denote the distance down-dip and split the fault into N segments of equal width Δx, where $N\Delta x$ is the total width of the fault in the dip direction. We then specify N slips s_i for $i = 1, 2, \ldots, N$. In our one-dimensional experiments, we take $N = 200$. This is finer than one would use in two dimensions and much finer than is needed to represent slip at an adequate level for either seismic or tsunami modeling. (For example, note from Fig. 1 that the seismic inversion for this event represents the slip as piecewise constant on a 18×10 grid with only 10 segments in the down-dip direction.) One could certainly reduce the dimension of the stochastic space below $N = 200$ using fewer subfaults. However, we will show that the dimension can be drastically reduced by instead using the K-L expansion reduced to only a few terms (e.g., 3, for this one-dimensional model the parameter choices below). By starting with a fine discretization of the fault, the eigenmodes used are smooth and perhaps better represent actual slip patterns than piecewise constant functions over large subfaults.

We assume that the N slips are to be chosen randomly from a joint normal distribution with mean $\boldsymbol{\mu} = [\mu_1, \mu_2, \ldots, \mu_N]^T$. The mean is chosen to be the desired taper, scaled to have the desired total slip. As an illustration of taper, we use the function

$$\tau(d) = 1 - \exp(-20\,|d - d_{\max}|/d_{\max}) \qquad (4)$$

where d is the depth of a subfault and $d_{\max} = 22,500$ m is the maximum depth of the fault. This function is close to 1 over most of the fault but tapers toward the down-dip edge. This taper, after scaling to give the mean slip, is shown as the dashed line in Fig. 3b. Other tapers can be used instead, e.g., the taper proposed by Wang and He (2008).

We set the desired covariance matrix to be $\hat{C}_{ij} = \sigma_i \sigma_j C_{ij}$ where $\sigma_i = \alpha\mu_i$ for some scalar $\alpha \in \mathbb{R}$ and $\boldsymbol{C}$ is the desired correlation matrix. Since α scales the standard deviation relative to the mean, the larger this is the more likely the slip is to be negative in some

[1] We follow http://earthquake.usgs.gov/aboutus/docs/020204mag_policy.php and use $M_w = \frac{2}{3}(\log_{10}(M_o) - 9.05)$ where the seismic moment M_o =length × width × (average slip) × (rigidity) and set the rigidity to 3.55×10^{10} N-m for this calculation.

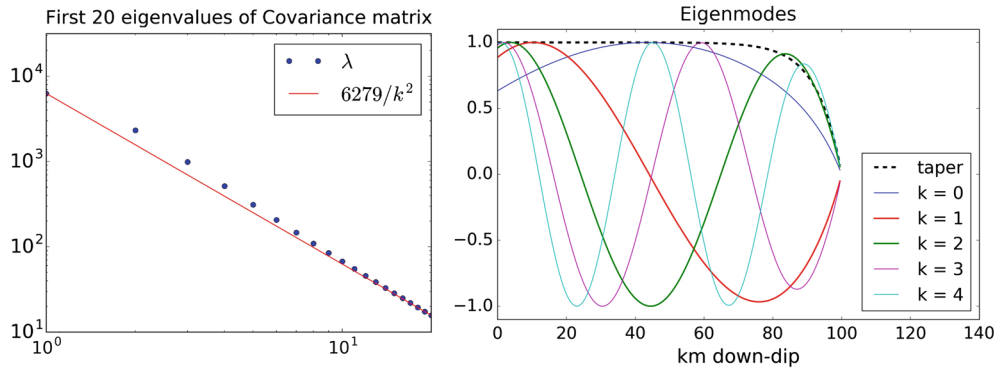


Figure 3
Eigenvalues decay like $1/k^2$ when the exponential autocorrelation function is used. The corresponding eigenvectors are similar to Fourier modes, shaped by the taper. The taper is shown as a dashed line and the eigenmodes are normalized to have max-norm equal to 1

regions. We take $\alpha = 0.75$, which tends to keep the slip positive everywhere, as desired, while still giving reasonable variation in slip patterns. This is similar to the value 0.85 used by Graves and Pitarka (2010) with the Fourier approach. The correlation matrix is given by $C_{ij} = \texttt{corr}(|x_i - x_j|)$ in terms of some autocorrelation function (ACF) $\texttt{corr}(r)$, and we choose

$$\texttt{corr}(r) = \exp(-r/r_0), \tag{5}$$

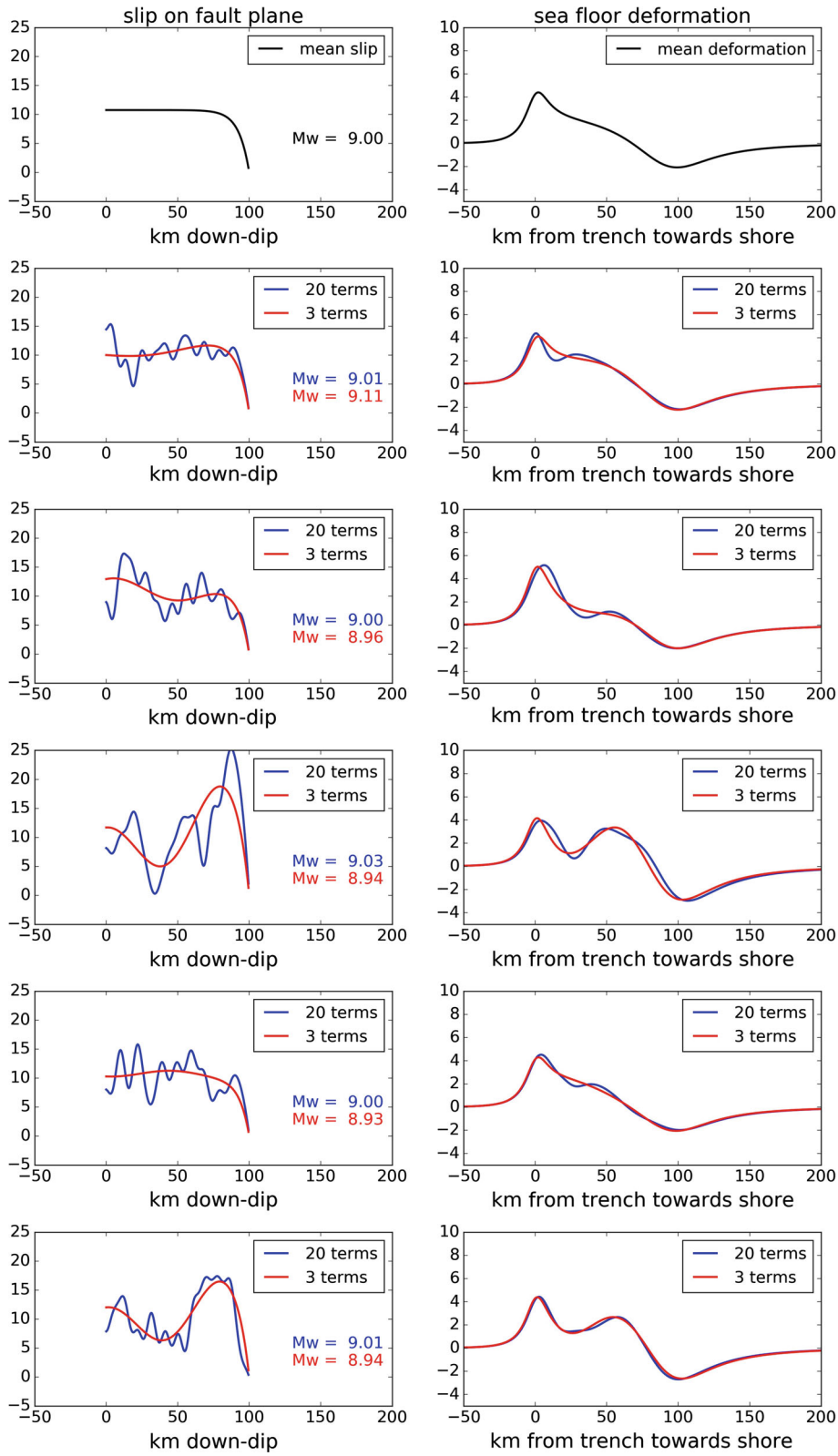
where the correlation length is set to $r_0 = 0.4W = 40$ km, i.e., 40 % of the fault width as suggested by the work of Mai and Beroza (2002). Figure 3a shows the first 20 eigenvalues on a logarithmic scale, showing that they decay like $\lambda_k \sim 1/k^2$. This is the same spectral decay rate as would be observed from taking the Fourier transform of the exponential ACF in a Fourier series-based approach. If the discretization is refined (i.e., the number of subfaults N grows), the spectral properties of the correlation matrix C_{ij} converge to those of the continuous ACF (5). For the untapered case on an finite interval, the analytic expressions for the eigenvalues of the ACF are known in terms of transcendental equations (Van Trees et al. 2013) and the decay can be shown to asymptotically approach $1/k^2$. More generally a different autocorrelation function could be chosen with a minor change to the code; for example, a von Karman ACF as considered in Mai and Beroza (2002), Graves and Pitarka (2010).

For our choice of ACF, Fig. 3b shows the taper along with the first several eigenvectors of the covariance matrix $\boldsymbol{C}$, ordered based on the magnitude of the eigenvalues. For comparison purposes, they are scaled to have maximum amplitude 1 and to be positive at the up-dip edge. Note that the lowest mode 0 looks very similar to the taper. Adding in a multiple of this mode will modify the total slip and hence the magnitude, so we drop this mode from the sum. The higher modes are orthogonal to mode 0 and hence do not tend to change the total slip. They look like Fourier modes that have been damped near the down-dip boundary by the taper.

To create a random realization, we choose a vector $\boldsymbol{z}$ of N i.i.d. Gaussian $\mathcal{N}(0,1)$ values z_k for $k = 0, 1, \ldots, N-1$. If we neglect the 0-mode and truncate the expansion after m terms, then this amounts to setting $z_0 = 0$ and $z_k = 0$ for $k > m$. We will denote such a $\boldsymbol{z}$ vector by $\boldsymbol{z}^{[m]}$. The slip pattern can then be written as

$$\boldsymbol{s} = \boldsymbol{\mu} + \boldsymbol{V}\boldsymbol{\Lambda}^{1/2}\boldsymbol{z}^{[m]}. \tag{6}$$

The left column of Fig. 4 shows the mean slip in the top plot, followed by several random realizations generated by the K-L expansion using 20 terms, with the $\boldsymbol{z}^{[20]}$ coefficients chosen as i.i.d. $\mathcal{N}(0,1)$ values. These are the blue curves in each plot. In each case, the slip is also shown when only 3 terms in the series are used (i.e., $\boldsymbol{z}^{[3]}$ is computed by leaving z_1, z_2, z_3 unchanged from $\boldsymbol{z}^{[20]}$ but with the higher terms dropped, equivalent to truncating the expansion at an earlier point). These slip patterns, shown in red, are smoothed versions of the 20-term slip patterns since


slip on fault plane
mean slip
Mw = 9.00
km down-dip
sea floor deformation
mean deformation
km from trench towards shore
20 terms
3 terms
Mw = 9.01
Mw = 9.11
Mw = 9.00
Mw = 8.96
Mw = 9.03
Mw = 8.94
Mw = 9.00
Mw = 8.93
Mw = 9.01
Mw = 8.94

◀Figure 4
The *left column* shows slip on the fault plane of width $W = 100$ km. The *right column* shows the resulting seafloor deformation if the up-dip edge of the fault plane is 5 km below the surface and it dips at 13°. The *top row* shows the mean slip and resulting deformation. The *remaining rows* show random realizations using 20 terms of a K-L expansion (*blue*) and the same sum truncated to 3 terms (*red*)

the higher wave number components have been suppressed. In many cases there appears to be quite a large difference between the 3-term and 20-term slips. This is a reflection of the fact that the eigenvalues do not decay all that quickly in this case. There would be faster decay if a longer correlation length were chosen, and much more rapidly if the Gaussian autocorrelation function were chosen instead of the exponential.

In spite of the differences in the slip patterns, for tsunami modeling the 3-term series may still be adequate. The right column of Fig. 4 shows the sea floor deformations $\Delta \boldsymbol{B}$ that result from the slips shown on the left. These are computed using the Okada solution (Okada 1985) to the homogeneous elastic half plane problem consisting of uniform dislocation on a rectangular subfault. Okada's result is the surface deformation, which can be evaluated on a fine grid at the surface. Applying this to each subfault, weighting by the corresponding slip, and summing the results produces (by linearity) the resulting sea floor deformation $\Delta \boldsymbol{B}$ in Fig. 4. The blue curves show the deformation due to the 20-term sum while the red curves show the deformation resulting from the truncated 3-term sum. Note that high wavenumber oscillations in the slip pattern are highly damped in the resulting seafloor deformations. (The degree of damping increases with depth of the fault below the surface.) If the $\Delta \boldsymbol{B}$ are sufficiently similar between the 3-term and the full K-L expansion, then there is no reason to use more terms. In this case, we have reduced the stochastic space that needs to be explored down to 3 dimensions. There is much greater similarity for some realizations than for others, and so below we examine the statistical properties of this approximation using a sample of 20,000 realizations.

Note that using only 3 modes may be too few for this particular set of fault parameters—the comparisons shown in Fig. 4 would look more similar if a few more terms were retained—but we will see that good statistical properties are obtained even with this severe truncation. How many terms are required depends on various factors: not only the correlation structure of the slip as discussed above, but also the depth of the fault plane. The deeper it is, the more damping takes place when the elastic wave model is used to obtain the Okada solution for seafloor deformation. Here, we placed the top of the fault plane at 5 km depth.

To examine statistical properties of the 20-term sum and the 3-term approximation, we generate 20,000 samples of each and compare some quantities that are cheap to compute but that are important indicators of the severity of the resulting tsunami. Running a full tsunami model based on the shallow water equations is not feasible for this large number of realizations, but the quantities we consider will stand in as proxies for the quantities one might actually want to compute, such as the maximum depth of flooding at particular points onshore. Moreover, the distribution of these proxy values can be used in a later stage to help choose particular earthquake realizations for which the full tsunami model will be run. It is desirable to run the model with judiciously chosen realizations for which the proxy values are well distributed over the range of possible values. The computed densities of the proxy values can also be used to weight the results from the full model runs to accurately reflect the probabilities of such events. This will be explored in detail in a future paper.

Computations for a large number of realizations can be sped up substantially by realizing that the Okada solution is linear in slip, i.e., if the slip vector is given by $\boldsymbol{s}$ then the resulting sea floor deformation can be written as $\Delta \boldsymbol{B} = \boldsymbol{\Theta s}$ for a matrix $\boldsymbol{\Theta} \in \mathbb{R}^{N_B \times N}$, where N_B is the number of grid points at which the deformation $\Delta \boldsymbol{B}$ is evaluated (in our experiments we use $N_B = 1001$ over an interval that extends 100 km on either side of the fault region). The Okada solution implemented in the GeoClaw `dtopotools` Python module is used, which directly computes $\Delta \boldsymbol{B}$ from $\boldsymbol{s}$ and so we do not actually compute the matrix $\boldsymbol{\Theta}$, but

it is useful conceptually. In particular, if the K-L expansion (6) is to be used to compute $\boldsymbol{s}$ then we find that

$$\Delta \boldsymbol{B} = \boldsymbol{\Theta\mu} + \boldsymbol{\Theta V \Lambda}^{1/2} \boldsymbol{z}^{[m]}. \tag{7}$$

The vector $\boldsymbol{\Theta\mu}$ is obtained by applying Okada to the mean slip vector. The matrix $\boldsymbol{\Theta V}$ can be computed by applying Okada to each column of $\boldsymbol{V}$ to compute the columns of the product matrix. Since the sum only involves m nonzero terms, we need to only apply Okada to columns 1 through m of $\boldsymbol{V}$ (i.e., the first m K-L modes $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_m$ used to express s). Hence, if we plan to use at most 20 modes of the K-L expansion then we need only apply the Okada solution to 21 slip vectors and we can then take linear combinations of the resulting sea floor deformations, rather than applying Okada to 20,000 slip realizations separately.

In practice, this can be simplified further. Applying Okada to a mode $\boldsymbol{v}_k$ actually requires applying Okada to each of the N subfaults, weighting by the corresponding element of $\boldsymbol{v}_k$, and summing over all the subfaults. So applying Okada to m modes in this way actually requires applying Okada mN times. Instead, we can first apply Okada to N unit source scenarios in which the slip is set to 1 on the jth subfault and to 0 on all other subfaults. Call this slip vector $\boldsymbol{s}^{[j]}$. Applying Okada to this gives a resulting $\Delta \boldsymbol{B}^{[j]} = \boldsymbol{\Theta s}^{[j]}$. Now for any slip vector $\boldsymbol{s}$ we can compute $\boldsymbol{\Theta s}$ as

$$\boldsymbol{\Theta s} = \sum_{j=1}^{N} s_j \Delta \boldsymbol{B}^{[j]}. \tag{8}$$

In particular, taking $\boldsymbol{s} = \boldsymbol{v}_k$ would give $\boldsymbol{\Theta v}_k$, but (8) can be used directly to compute the seafloor deformation $\Delta \boldsymbol{B} = \boldsymbol{\Theta s}$ for any slip realization. This approach can also be used in the lognormal case described in Sect. 4 and employed in Sect. 5.

Subsidence or uplift One quantity that has a significant impact on the severity of tsunami flooding is the vertical displacement of the seafloor at the coast. If this displacement is negative and the land subsides at the shore, then flooding may be much worse than if uplift occurs. The behavior seen for a particular event depends on how far offshore the subduction zone is, which is generally directly related to the width of the continental shelf offshore from the community of interest (since the top edge of the fault is usually located near the trench at the edge of the shelf, which is part of the continental plate beneath which the oceanic plate is subducting). This distance can vary from only a few km (e.g., along the Mexico coast) to 200 km (e.g., along the coast of Japan where the Tohoku event occurred). In our model, the top of the plate is at $x = 0$ and we choose the coast line location to be at $x = 75$ km, which gives a wide range of subsidence and uplift values, as can be observed for the realizations shown in Fig. 4.

The displacement at one particular point is easy to determine from each realization, it is just one entry in the vector of sea floor deformation obtained from the Okada solution, say $\Delta B_j = \boldsymbol{e}_j^T \Delta \boldsymbol{B}$ for some j, where $\boldsymbol{e}_j$ is the unit vector with a 1 in position j. (We have evaluated $\Delta \boldsymbol{B}$ on a fine grid so we assume we do not need to interpolate). As such, this particular quantity is in fact easy to compute directly from $\boldsymbol{z}$ for any given realization, as

$$\Delta B_{\text{shore}} = \Delta B_{\text{shore}}(\boldsymbol{\mu}) + \boldsymbol{b}^T \boldsymbol{z}, \tag{9}$$

where $\Delta B_{\text{shore}}(\boldsymbol{\mu}) = \boldsymbol{e}_j^T \boldsymbol{\Theta\mu}$ is the shoreline displacement resulting from the mean slip and the row vector $\boldsymbol{b}^T$ is

$$\boldsymbol{b}^T = \boldsymbol{e}_j^T \boldsymbol{\Theta V \Lambda}^{1/2}, \tag{10}$$

i.e., the vector consisting of the jth component of the sea floor displacement resulting from applying Okada to each K-L mode, scaled by the square root of the corresponding eigenvalue. From (9) it follows immediately that ΔB_{shore} is normally distributed with mean $\Delta B_{\text{shore}}(\boldsymbol{\mu})$ and variance $\sigma^2 = \sum_{k=1}^{m} b_k^2$ (in the Gaussian case considered here, not in the lognormal case considered below). Hence, for this particular quantity of interest in the Gaussian case, we do not need to estimate the statistics based on a large number of samples. We can immediately plot the Gaussian density function for the "full" expansion with 20 terms and compare it to the density for the truncated series with only 3 terms. These are seen to lie nearly on top of one another in Fig. 5. This plot also shows the density that would be obtained with only 1 or 2 terms in the K-L expansion, which are substantially different. This confirms that, in terms of

this particular quantity of interest, it is sufficient to use a 3-term K-L expansion (but not fewer terms).

Figure 5 also shows the density as estimated using 20,000 samples, using a kernel density estimate computed using the Python function `gaussian_kde` from the package `scipy.stats`, version 0.16.1. With either 3 terms or 20 terms, the estimated density lies nearly on top of the true density, giving confidence that the sampling has been programmed properly and that 20,000 samples are sufficient since the true density is known in this case.

Potential energy From the samples, it is possible to also estimate the densities for other quantities of interest for which it is not possible to compute the true density. We consider two additional quantities that have relevance to the magnitude of the tsunami generated. One is the potential energy of the initial perturbation of the ocean surface, which is one measure of its potential for destruction. For example, a recent study of 44 ocean-bottom earthquakes by Nosov et al. (2014) showed that tsunami intensity is highly correlated with potential energy. The potential energy is given by

$$E = \frac{1}{2}\int\int \rho g \eta^2(x,y)\,\mathrm{d}x\,\mathrm{d}y \qquad (11)$$

where $\rho = 1000$ kg/m^3 is the density of water, $g = 9.81$ m/s^2, and $\eta(x,y)$ is the initial perturbation of the surface. With our assumption that the sea surface moves instantaneously with sea floor deformation generated from the slip, η is equal to the sea floor displacement in the ocean, while onshore we set $\eta = 0$ since displacement at these points does not contribute to the potential energy of the tsunami. For the one-dimensional problem considered here, we sum the square of the displacement over $x < 75$ km and scale by $\rho g L \Delta x$ to define E, taking $L = 100$ km. Finally, we multiply by 10^{-15} so that the results are order 1, with units of PetaJoules. We plot the density (again estimated using `gaussian_kde`) obtained with 20,000 samples, using 20 terms, or truncating further to 1, 2, or 3 terms. The results in Fig. 6 again show that 3 terms are sufficient to obtain very similar results to 20 terms.

Maximum wave height The maximum positive seafloor displacement gives the maximum amplitude of the tsunami at the initial time. We expect this to be positively correlated with the amplitude of the wave that approaches shore (although the wave propagation can be complicated by multiple peaks, the location of the $\eta_{\max}$ relative to the shore, or various other factors that can only be studied with a full tsunami model). The right plot of Fig. 6 shows the kernel density estimates of this quantity $\eta_{\max}$.

Joint probability densities It is also interesting to plot the joint probability density of pairs of quantities to better explore the ability of 3 terms to capture the variation. This is illustrated in Fig. 7, where the top row shows kernel density estimates for E vs. $\eta_{\max}$ and the bottom rows shows ΔB_{shore} vs. $\eta_{\max}$. In each case, the left figure shows the density computed from 20,000 realizations of the 20-term K-L expansion while the right figure shows the density estimated

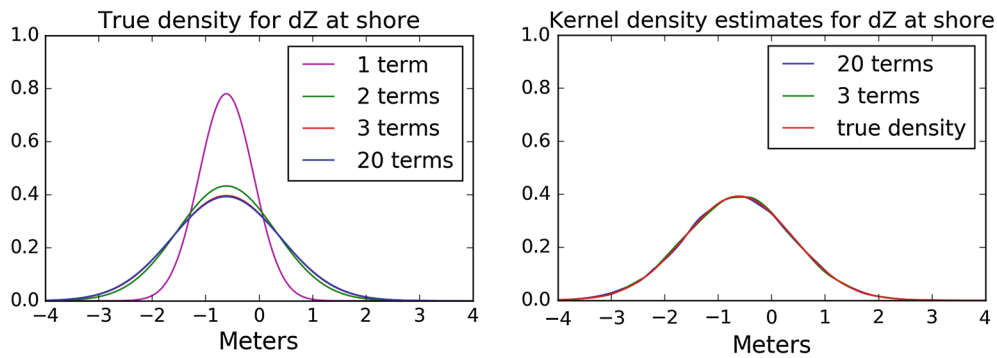


Figure 5
The figure on the *left* shows the true Gaussian density for the shoreline displacement for K-L expansions with 1, 2, 3, or 20 terms. The figure on the *right* shows the kernel density estimate from 20,000 samples using 3 terms or 20 terms, together with the true density for 20 terms

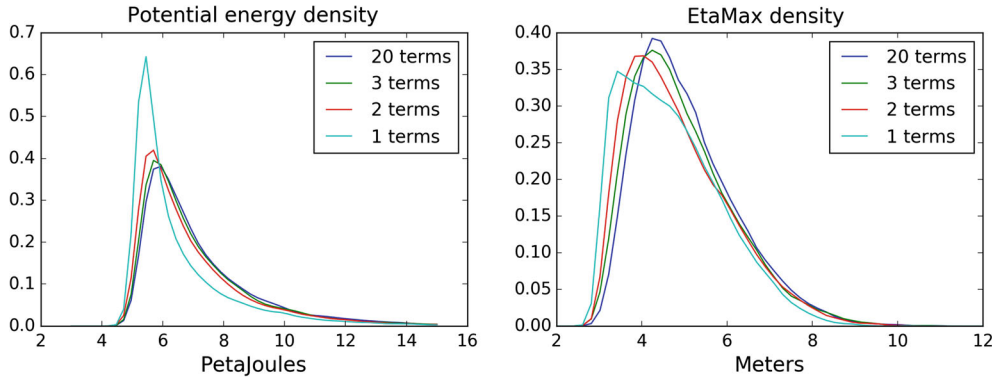


Figure 6
Kernel density estimates based on 20,000 samples, using 1, 2, 3, or 20 terms in the K-L expansion. The *left figure* shows the potential energy (11) and the *right figure* shows the maximum amplitude of deformation (sea surface elevation)

from an equal number of 3-term expansions. In each case, it appears that the 3-term expansion captures the bulk of the variation.

The joint distribution of $\eta_{\max}$ and ΔB_{shore} is of particular interest since the most dangerous events might be those for which $\eta_{\max}$ is large while ΔB_{shore} is most negative (greatest subsidence of the coast). The fact that the joint distributions look quite similar gives hope that the 3-term model will adequately capture this possibility.

Depth proxy hazard curves The goal of a full-scale PTHA exercise is often to generate hazard curves at many points onshore or in a harbor. A hazard curve shows the probability that the maximum flow depth (or some other quantity of interest) will exceed some value as a function of that "exceedance value". Construction of these curves is discussed, for example, in the appendices of the study by González et al. (2014). The curves may vary greatly with spatial location due to the elevation of the point relative to sea level, and also due to the manner in which a tsunami interacts with nearby topography. Hazard curves must thus be computed using fine-grid simulations of the tsunami dynamics and cannot be computed directly from the sea floor deformation alone.

The general idea is to choose a quantity of interest, such as the maximum depth D observed over all time at some particular spatial point, and then for a set of exceedance values D_e, we need to calculate

$$P[D > D_e] = \int \rho(z) I(z; D_e)\, dz \tag{12}$$

where the integral is over the m-dimensional stochastic space of coefficients z of the K-L sum (assuming m terms are used) and $I(z; D_e)$ is an indicator function that is 1 at points $z \in \mathbb{R}^m$ where the corresponding realization gives a tsunami that exceeds D_e and 0 elsewhere (or it could take values between 0 and 1 to incorporate other uncertainties, e.g., if the approach of Adams et al. (2015) is used to incorporate tidal uncertainty). The function $\rho(z)$ in (12) is the probability density for z. In the K-L approach, z is a vector of i.i.d. normally distributed values so $\rho(z)$ is known; for the m-term expansion it takes the form

$$\rho(z) = \frac{1}{\sqrt{(2\pi)^m}} \exp\left(-\frac{1}{2}\sum_{i=1}^{m} z_i^2\right). \tag{13}$$

A brute force approach to estimate this is to use a simple Monte Carlo method in which the integral in (12) is replaced by

$$P[D > D_e] \approx \frac{1}{n_s}\sum_{j=1}^{n_s} I_e(z^{[j]}), \tag{14}$$

with a large number of samples n_s, where $z^{[j]}$ now represents the jth sample, drawn from the joint normal distribution with density $\rho(z)$. More sophisticated techniques would be needed in general to reduce the number of samples required, since running the full

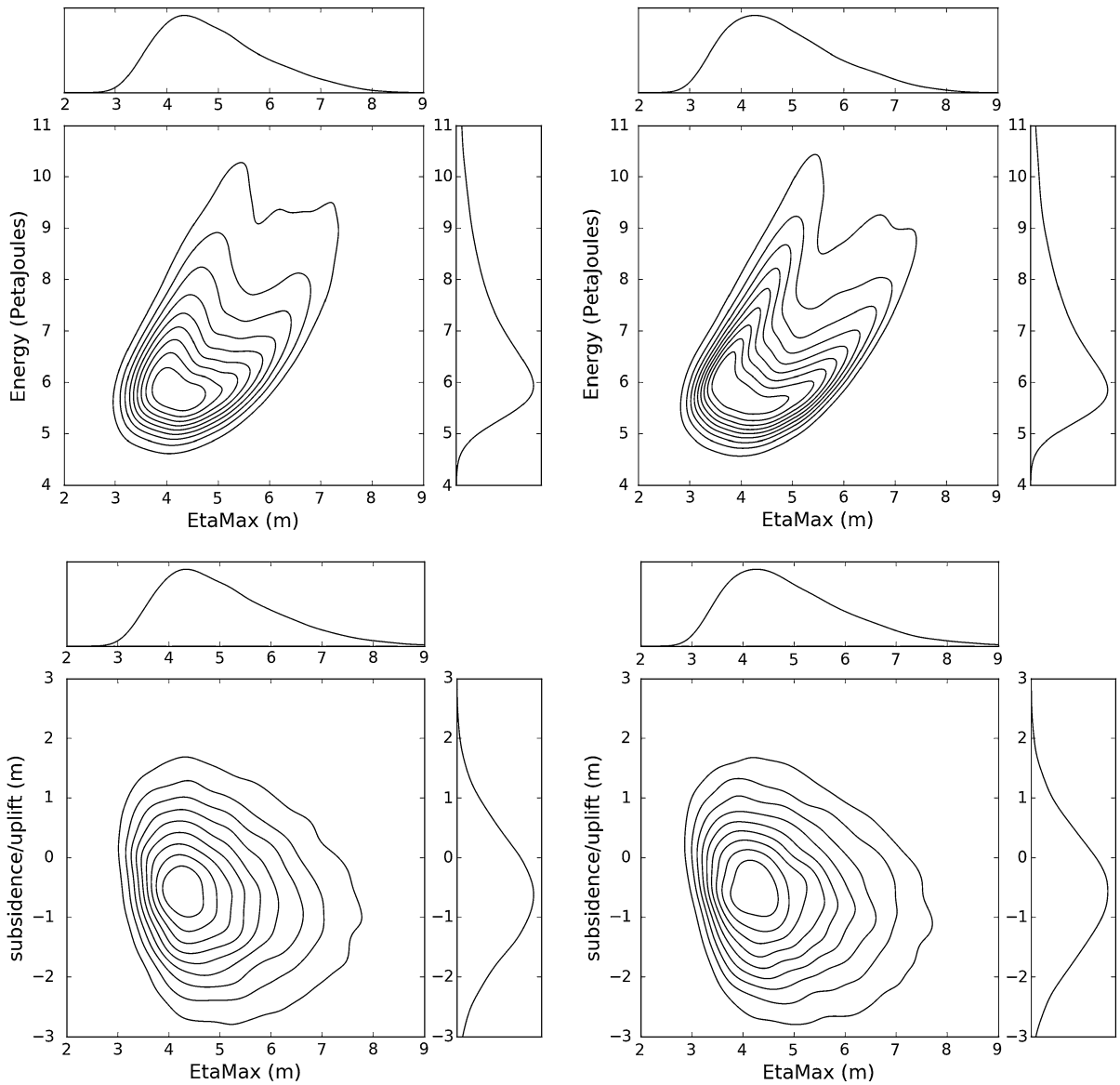


Figure 7
Joint and marginal probability densities for different quantities, comparing the densities estimated using the 20-term expansion (*left column*) and the 3-term expansion (*right column*). The *top row* shows the joint density of $\eta_{\max}$ with potential energy E of the tsunami generated. The *bottom row* shows the joint density of $\eta_{\max}$ with ΔB_{shore}, the vertical displacement at the shore

tsunami model may take hours of computing time for each sample. This is discussed briefly in Sect. 6.

For the purposes of exploring the effects of truncating the K-L expansion, here we use a simple proxy for maximum flooding depth that is cheap to compute for each realization: $D = \eta_{\max} - \Delta B_{\text{shore}}$, the maximum offshore sea surface elevation augmented by any subsidence that occurs at the shore. We do not claim that this is a good estimate of the actual maximum water depth that will be observed at the shore, but computing hazard curves for this quantity provides another test of how well the 3-term K-L expansion captures the full probability distribution described by the 20-term expansion. This curve is

obtained by computing D for each sample and determining the fraction of samples for which this is above ζ_i, for each exceedance level ζ_i on a fine grid covering the range of D observed. Figure 8 shows the resulting hazard curve obtained with the 20-term expansion using $n_x = 20,000$ samples. The curve obtained with the 3-term expansion and an independent set of $n_x = 20,000$ samples is also shown, and lies nearly on top of it.

Exploring parameter space One advantage of describing the probability space of possible events in terms of a small number of stochastic parameters is that it may be possible to identify structure in this stochastic space, which can be important in developing a cheap surrogate model to use in estimating probabilities and computing hazard curves for practical quantities of interest. For example, we can ask what parts of parameter space lead to the worst events. The left figure in Fig. 9 shows the events (projected to the z_1–z_2 plane) from the above tests with the 3-term K-L expansion for which the proxy depth is greater than 8 m. The contours of the bivariate normal distribution are also plotted. A scatter plot of all 20,000 events would cluster in the middle, but we observe that the events giving this extreme depth tend to have $z_1 > 1$. From Fig. 3, we see that positive z_1 redistributes slip from the down-dip to the up-dip portion of the fault. This agrees with common wisdom from past events that concentration

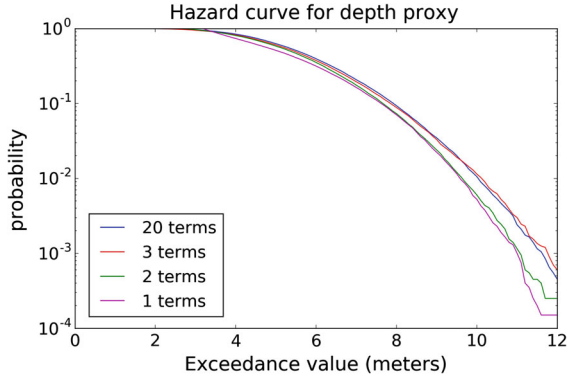


Figure 8
Hazard curves based on the proxy for flooding depth given by $\eta_{\max} - \Delta B_{\text{shore}}$ based on 20,000 samples using the full 20-term K-L expansion, compared with the hazard curves obtained using only 1, 2, or 3 terms in the expansion. Note that 3 terms are sufficient to obtain the hazard curve to high precision

near the up-dip edge gives particularly severe tsunamis (as in the case of the 2011 Tohoku event). The right figure in Fig. 9 shows a similar scatter plot of z_1–z_2 values for which the potential energy was above 9.5 PetaJoules. In this case, most of the extreme events have z_1 either very positive or very negative. In the latter case, slip is concentrated toward the down-dip portion of the fault, which leads to a smaller maximum surface displacement but the displacement spreads out further spatially for a deep rupture, which can lead to large potential energy since this is integrated over space.

4. Lognormally Distributed Slip

If we wish to instead generate slip realizations that have a joint lognormal distribution with a desired mean and covariance matrix, we can first generate realizations of a joint Gaussian random field and then exponentiate each component. This approach will be used in the two-dimensional example below in Sect. 5.

In this case, we first choose the desired mean $\boldsymbol{\mu}$ and covariance matrix $\hat{\boldsymbol{C}}$ for the slip, and then compute the necessary mean $\boldsymbol{\mu}^g$ and covariance matrix $\hat{\boldsymbol{C}}^g$ for the Gaussian to be generated by the K-L expansion, using the fact that if $\boldsymbol{g}$ is a random variable from $\mathcal{N}(\boldsymbol{\mu}^g, \hat{\boldsymbol{C}}^g)$, then $\exp(\boldsymbol{g})$ is lognormal with mean and covariance matrix given by:

$$\mu_i = \exp(\mu_i^g + \hat{C}_{ii}^g/2), \qquad \hat{C}_{ij} = \mu_i \mu_j (\exp(\hat{C}_{ij}^g) - 1). \tag{15}$$

Hence, we can solve for

$$\begin{aligned} \hat{C}_{ij}^g &= \log(\hat{C}_{ij}/\mu_i\mu_j + 1), \\ \mu_i^g &= \log(\mu_i) - \frac{1}{2}\hat{C}_{ii}^g. \end{aligned} \tag{16}$$

We now find the eigenvalues λ_k and eigenvectors $\boldsymbol{v}_k$ of $\hat{\boldsymbol{C}}^g$. To generate a realization we choose N values $z_k \sim \mathcal{N}(0,1)$ and then form the K-L sum

$$\boldsymbol{s}^g = \boldsymbol{\mu}^g + \sum_{k=0}^{N} z_k \sqrt{\lambda_k} \boldsymbol{v}_k. \tag{17}$$

We then exponentiate each component of $\boldsymbol{s}^g$ to obtain the slip values, which then have the desired joint lognormal distribution (see e.g., Ghanem 1999).

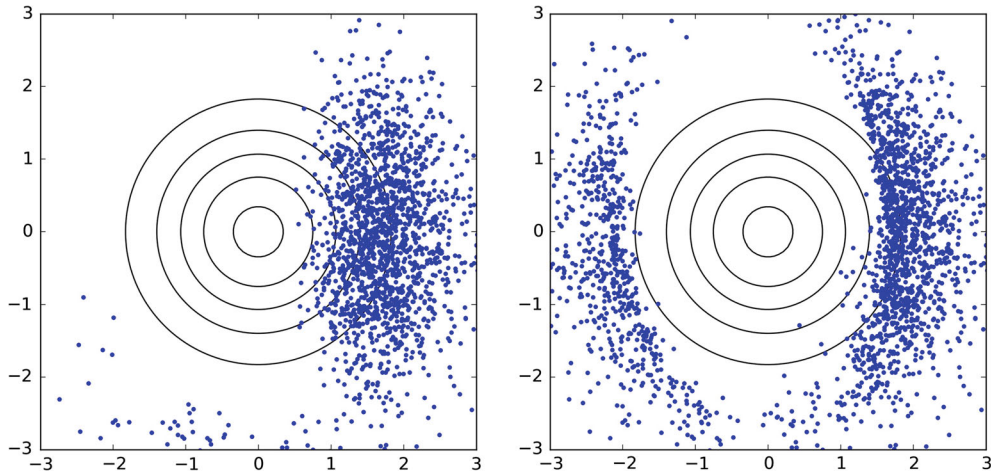


Figure 9
Scatter plots in the z_1-z_2 plane of the subset of 3-term events for which the proxy depth is greater than 8 m (*left*) or for which the potential energy is greater than 9.5 PetaJoules (*right*)

As described, this will generate realizations with total slip (and hence magnitude M_w) that vary around the mean. As in the Gaussian case, we can drop the nearly constant $\boldsymbol{\nu}_0$ term from the sum to reduce this variation. We can also generally truncate the series to a much smaller number of terms and still capture most of the variation if the eigenvalues are rapidly decaying.

Now consider the special case where we make the same assumptions as in Sect. 2 that $\hat{C}_{ij} = \sigma_i \sigma_j C_{ij}$ where $\boldsymbol{C}$ is the desired correlation matrix and $\sigma_i = \alpha \mu_i$, while the mean μ_i was given by some taper τ_i scaled by a scalar value $\bar{\mu}$. Then computing $\boldsymbol{\mu}^g$ and $\hat{\boldsymbol{C}}^g$ according to (16), we find that:

$$\begin{aligned} \hat{C}^g_{ij} &= \log(\alpha^2 C_{ij} + 1), \\ \mu^g_i &= \log(\bar{\mu}\tau_i) - \frac{1}{2}\log(\alpha^2 + 1) \end{aligned} \tag{18}$$

We see that the covariance matrix in this case depends only on the correlation matrix and the scalar α, not on the mean slip itself (and in particular is independent of the taper). We also find that $\exp(\mu^g_i) = \bar{\mu}\tau_i/\sqrt{\alpha^2 + 1}$ is simply a scalar multiple of the taper.

Using these assumptions and the fact that

$$\exp\left(\boldsymbol{\mu}^g + \sum_{k=1}^{N} z_k \sqrt{\lambda_k}\boldsymbol{\nu}_k\right) = \exp(\boldsymbol{\mu}^g)\exp\left(\sum_{k=1}^{N} z_k \sqrt{\lambda_k}\boldsymbol{\nu}_k\right),$$

it is easy to generate realizations that have exactly the desired magnitude: simply compute

$$\exp\left(\sum_{k=1}^{N} z_k \sqrt{\lambda_k}\boldsymbol{\nu}_k\right), \tag{19}$$

multiply the result by the desired taper, and then rescale by a multiplicative factor so that the area-weighted sum of the slips gives the total slip required for the desired seismic moment.

5. Two-Dimensional Case

We now present an example in which the slip is allowed to vary in both directions along a fault surface. For illustration, we use a subset of the Cascadia Subduction Zone from Fig. 2, taking only the southern-most 8 fault segments, as illustrated in Fig. 10. These are subdivided into 540 smaller fault planes for the purposes of defining the slip.

To define the 540×540 correlation matrix, we need to compute the pairwise "distance" between subfault i and subfault j. We can compute the Euclidean distance d_{ij}, but for a long fault geometry it may be desirable to specify a longer correlation length in the strike direction than down-dip, so we use a more general definition

$$C_{ij} = \exp(-(d_{\text{strike}}(i,j)/r_{\text{strike}}) - (d_{\text{dip}}(i,j)/r_{\text{dip}})) \tag{20}$$

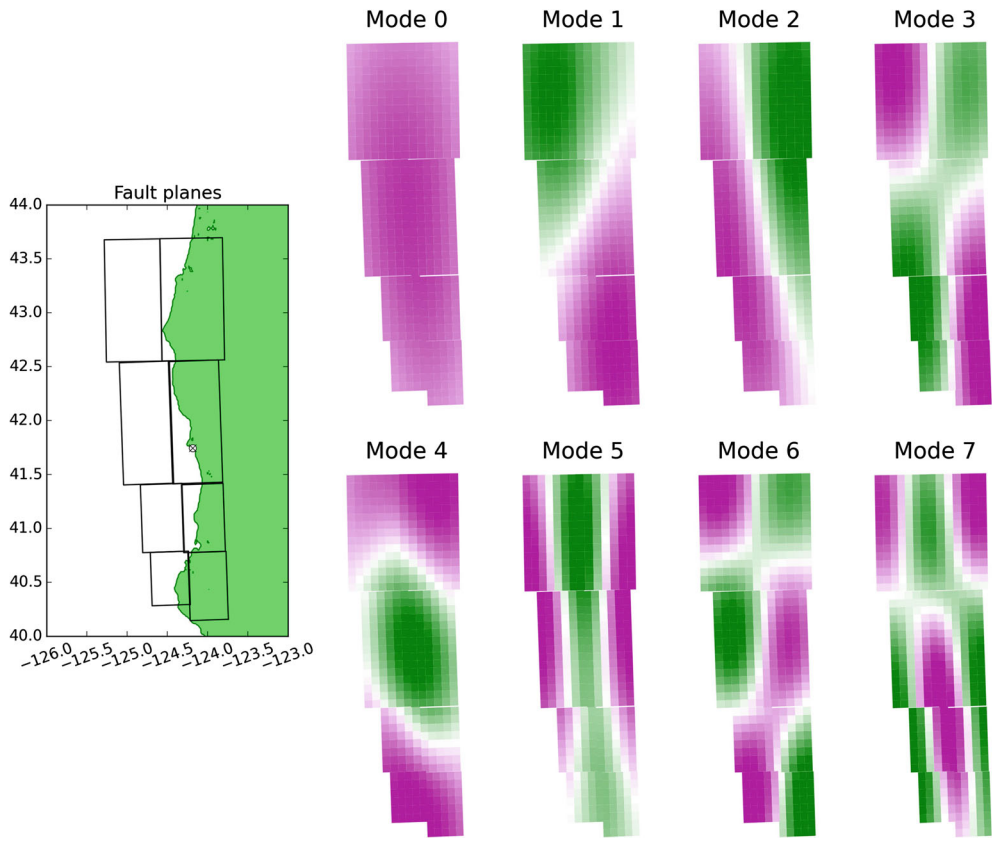
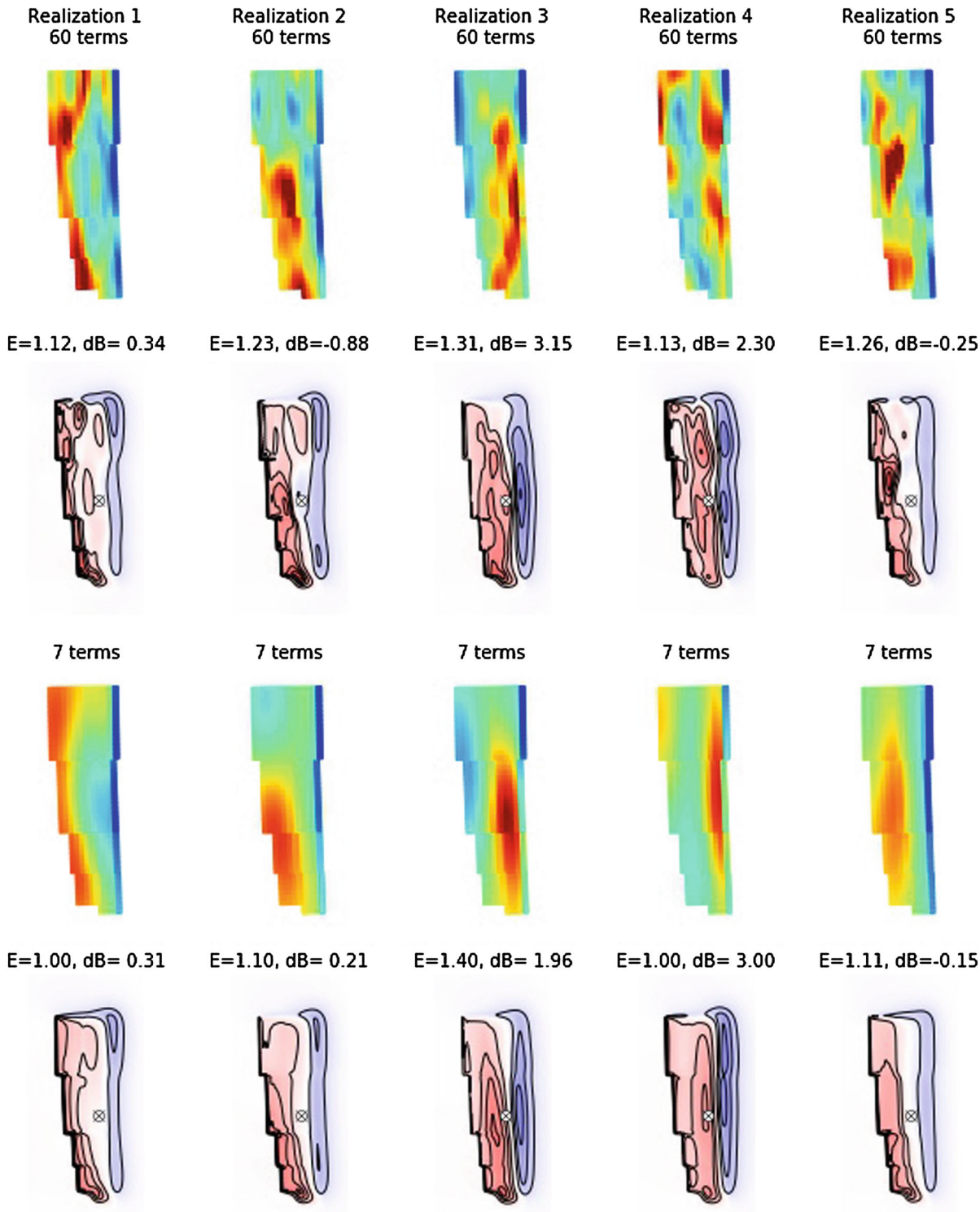


Figure 10
Southern portion of the CSZ fault showing location of Crescent City, CA and the 8 subfaults that are further subdivided into 540 subfaults. The first 7 eigenmodes of the resulting covariance matrix $\hat{\boldsymbol{C}}^g$ are also shown

where $d_{\text{strike}}(i,j)$ and $d_{\text{dip}}(i,j)$ are estimates of the distance between subfaults i and j in the strike and dip direction, respectively, and $r_{\text{strike}}, r_{\text{dip}}$ are the correlation lengths in each direction. We define $d_{\text{dip}}(i,j)$ using the difference in depth between the two subfaults and the dip angle δ as $d_{\text{dip}}(i,j) = d_{\text{depth}}/\sin(\delta)$, setting $d_{\text{strike}}(i,j) = \sqrt{d_{ij}^2 - d_{\text{dip}}(i,j)^2}$. We take the correlation lengths to be 40 % of the fault length and width, respectively, $r_{\text{strike}} = 130$ km and $r_{\text{dip}} = 40$ km. We again use an exponential autocorrelation function as defined in (20), but this could easily be replaced by a different ACF. We use the lognormal approach described in Sect. 4, with parameter $\alpha = 0.5$. Figure 10 shows the first 8 eigenmodes of $\hat{\boldsymbol{C}}^g$. Again we drop Mode 0 from the sum, since this mode is roughly constant over the fault.

To create slip realizations, we use (19) and then apply a tapering only at the down-dip edge, given by (4) with $d_{\max} = 20{,}000$ m (It would also be easy to taper at the other edges of the fault, if desired). We then scale the slip so that the resulting seismic moment gives $M_w = 8.8$. Figure 11 shows 5 typical realizations, comparing the slip generated by a 60-term K-L expansion with the slip generated when the series is truncated after 7 terms. The resulting seafloor deformation in each case is also shown, along with the potential energy and the subsidence/uplift ΔB_{shore} at one point on the coast, the location of Crescent City, CA. Note that in each case, the 7-term series gives a smoother version of the slip obtained with 60 terms, and the seafloor deformations are more similar than the slip patterns, as expected from the one-dimensional analogous case shown in Fig. 4. The potential energy and ΔB_{shore} are also seen to be


Realization 1
60 terms
Realization 2
60 terms
Realization 3
60 terms
Realization 4
60 terms
Realization 5
60 terms
E=1.12, dB= 0.34
E=1.23, dB=-0.88
E=1.31, dB= 3.15
E=1.13, dB= 2.30
E=1.26, dB=-0.25
7 terms
7 terms
7 terms
7 terms
7 terms
E=1.00, dB= 0.31
E=1.10, dB= 0.21
E=1.40, dB= 1.96
E=1.00, dB= 3.00
E=1.11, dB=-0.15

◀Figure 11
The top row shows 5 sample realizations of slip on the southern CSZ fault, as computed with a 60-term K-L expansion. The second row shows the resulting seafloor deformation, with an indication of the potential energy and the vertical displacement at Crescent City, CA, which is indicated by the X in the figures. The *third row* shows the same 5 realizations but with the K-L series truncated to 7 terms, and the *bottom row* shows the resulting seafloor deformations

similar when the truncated series is used to the values obtained with the longer 60-term series.

We can explore the statistical properties by repeating any of the experiments performed above in the one-dimensional case. In the interest of space, we only show one set of results, the same joint and marginal densities examined in the one-dimensional case in Fig. 7. The comparisons for the two-dimensional fault are shown in Fig. 12. To generate each column of figures we computed 20,000 slip realizations and the resulting seafloor deformations [via (8)]. The first column shows statistics when a 60-term KL-expansions is used, producing realizations similar to those shown in the top row of Fig. 11. The second column of figures was produced using an independent set of 7-term realizations (i.e., these were not obtained by truncating the 60-term series from the first set, but rather by generating 20,000 independent samples). Even in this two-dimensional case, less than 10 min of CPU time on a MacBook Pro laptop was required to generate each set of 20,000 realizations, the resulting seafloor deformations, and the kernel density plots.

Finally, as a more quantitative measure of the difference between the density functions shown in Fig. 12, we compute the total variation distance between the probability density obtained with the 60-term series and the density obtained with the same series truncated to N terms. This is computed by evaluating the kernel density estimates of the joint density of $\eta_{\max}$ and ΔB_{shore} on a 200×200 grid of points to obtain P_{ij}^{60} and P_{ij}^{N} and then computing

$$\delta(P^{60}, P^{N}) = \sum_{i=1}^{200} \sum_{j=1}^{200} |P_{ij}^{60} - P_{ij}^{N}|. \qquad (21)$$

Figure 13 shows a plot of how this distance decreases as the number of terms N is increased. We see that it drops rapidly up to $N = 7$ and then continues to decay exponentially (linearly on this semilog scale plot) as the number of terms is increased.

Recall that we assumed correlation lengths of roughly 40 % of the fault dimensions. It is interesting to investigate how things change if we specify a different correlation length in defining the target covariance matrix. Figure 13 also shows the results obtained from this same experiment, but when we specify shorter correlation lengths of 20 % of the fault dimensions. The eigenmodes shown in Fig. 10 change only slightly, but the decay rate of the eigenvalues will be slower. As a result, more terms in the series will be needed to capture the same degree of agreement with the 60-term series. The plot shown in Fig. 13 for the 20 % correlation lengths shows that the total variation distance still decays exponentially, and that for any given level of agreement (as measured by the total variation distance), at most 8 additional terms must be kept. We have computed similar curves for the Kullback–Leibler divergence of the difference in the densities (another common statistical measure) and see similar results.

6. Discussion

We have presented an approach to defining a probability distribution for earthquake slip patterns on a specified fault geometry that has been subdivided into an arbitrary number of rectangular subfaults, with a specified mean and covariance matrix. Slip realizations can be generated that either have a joint normal distribution or a joint lognormal distribution. Once the parameters have been chosen that define the distribution (this is the hard part, see the discussion of epistemic uncertainty below), it is very easy to generate an arbitrary number of sample realizations from the distribution, simply by drawing the coefficients z_k of the K-L series from independent normal distributions.

We have also illustrated that with a realistic choice of correlation length, the K-L series can be truncated to a relatively small number of terms. For tsunami modeling applications, the Okada equations are applied to each slip pattern to generate the seafloor deformation and it was shown that this is a smoothing operation that can further reduce the

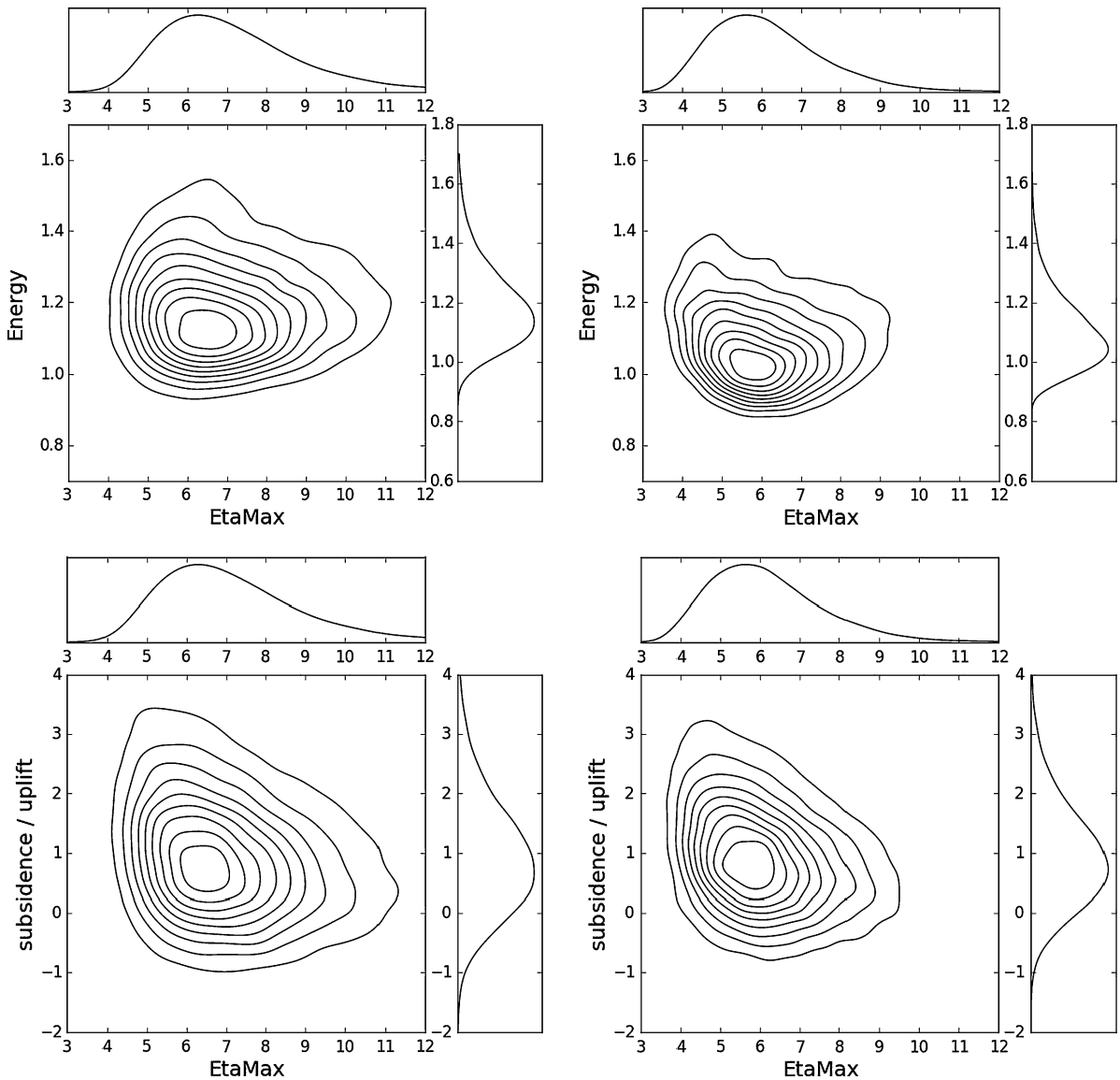


Figure 12
Joint and marginal probability densities for different quantities, comparing the densities estimated using the 60-term expansion (*left column*) and the 7-term expansion (*right column*) for the two-dimensional fault case. The *top row* shows the joint density of $\eta_{\max}$ with potential energy E of the tsunami generated. The *bottom row* shows the joint density of $\eta_{\max}$ with ΔB_{shore}, the vertical displacement at Crescent City, CA. In each case, 20,000 realizations similar to those shown in Fig. 11 were used to create these kernel density estimates

number of terms needed, and hence the dimension of the stochastic space that must be explored in doing PTHA analysis.

We have studied the degree to which the series can be successfully truncated by computing the statistical similarity of ensembles of realizations generated using a series with a large number of terms to ensembles generated from truncated series. To compute statistics based on thousands of realizations, we have restricted our attention to proxy quantities of interest that are easy to compute from slip patterns without running a full tsunami model. Other approaches to comparing the realizations more directly could also be considered. For seismic

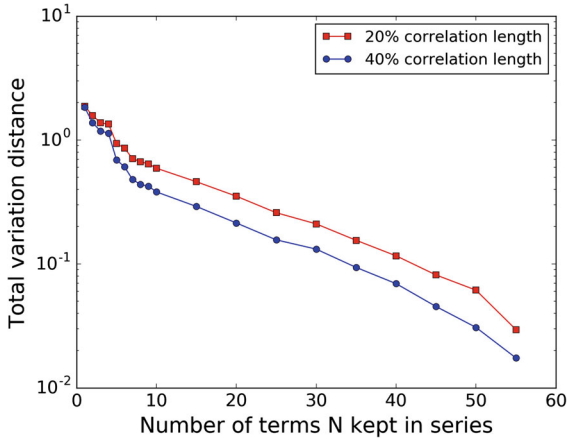


Figure 13
Total variation distance between the joint probability of $\eta_{\max}$ and ΔB_{shore} from Fig. 12 computed using 60 terms with that of the truncated series using N terms, as N varies

applications it might be valuable to compare individual realizations of slip, using, for example, the recent comparison techniques of Razafindrakoto et al. (2015), Zhang et al. (2015), but for tsunami applications we believe that comparison of sea floor deformation and derived tsunami quantities is more relevant.

To use the K-L expansion approach for practical PTHA analysis, two major challenges must be addressed. The first is to tackle the epistemic uncertainty associated with the lack of knowledge about possible future earthquakes. For the purpose of explaining the general methodology in this paper, we have chosen various parameters and the autocorrelation function to give credible examples based on the available literature, but without any claim that these are correct for the Cascadia Subduction Zone or any particular fault. To be optimally useful in practice, we would need to choose the parameters defining the probability distribution in a suitable way for real fault geometries so that it accurately represents the space of possible future earthquakes. Realistic specification of these critical seismic parameters and quantification of the associated uncertainties and geophysical constraints is a major challenge that Stein et al. (2012) have reviewed and summarized; they characterize the problem as a failure of earthquake hazard mapping, in general, and make recommendations regarding improvements. The problem is particularly severe in the case of near-field PTHA studies, because tsunami impact on a coastal community is highly sensitive to details of the seismic deformation (e.g., Geist 2002). Existing expertise and geophysical constraints should at least be incorporated in the choice of these parameters. The ability to generate many realizations and examine statistics of quantities such as those used in this paper may help in this. As one example, the parameters chosen in this paper for the CSZ example tend to give uplift rather than subsidence at Crescent City (as can be seen in the marginal distribution of subsidence/uplift in Fig. 12). If this is viewed as inconsistent with the geological evidence from past events, this could be adjusted, for example by tapering the slip more on the down-dip side. Moving more of the slip up-dip will cause more subsidence at the shore. It would also be possible to explore ways in which the epistemic uncertainty associated with the lack of knowledge about the true probability distribution affects the resulting hazard maps generated by a PTHA analysis, for example by doing the analysis with different parameter choices, and hence different probability distributions, to see how robust the PTHA analysis is to changes in assumptions.

The second major challenge is to deal with the aleatoric uncertainty that is still present even if the parameters defining the probability distribution were known to be correct. We are still faced with a high-dimensional space to sample to perform PTHA analysis. For example, if we wish to compute a hazard curve similar to Fig. 8 for the probability that the maximum depth D at some particular point will exceed various depths, then in practice we wish to perform full tsunami simulations that can take hours to run for a single realization. The brute force approach used in the approximation (14) with $n_s =$ 20,000 samples was feasible with the depth proxy used to produce Fig. 8, but would not be possible if a full tsunami model is used to compute D. The number of simulations required can be reduced by source-filtering techniques that identify a "most-important" subset of realizations that contribute most to the tsunami impact on a particular site, e.g., Lorito et al. (2015). An alternative would be to compute the integral with a quadrature algorithm based on sampling on a grid in z-space, but this is infeasible for high-dimensional m. For example, if $m = 10$ then a

tensor-product grid with only 4 points in each direction has $4^{10} \approx 10^6$ points.

Many other techniques have been developed in recent years to estimate such integrals in high-dimensional spaces, including for example Latin hypercube sampling (e.g., Olsson and Sandberg 2002), sparse grids (e.g., Nobile et al. 2008), and quasi-random grids (e.g., Dick et al. 2013) that have fewer points than uniform tensor-product grids. There are also several Monte Carlo sampling methods that can obtain accurate results with fewer samples than the naive sum of (14), including multi-level or multi-fidelity methods (e.g., Cliffe et al. 2011; Giles 2008; Peherstorfer et al. 2016) that combine results from many simulations that are cheap to compute with a relatively few simulations with the full model on a fine grid. Cheaper approximations might be obtained using some of the proxy quantities from this paper, by computing with a full tsunami model but on coarse grids, or by developing surrogate models or statistical emulators based on relatively few samples (e.g., Bastos and O'Hagan 2009; Benner et al. 2015; Li et al. 2011; Sarri et al. 2012). We are currently exploring several of these approaches for PTHA and will report on them in future publications.

The computer code used to generate all of the figures in this paper can be found in the repository https://github.com/rjleveque/KLslip-paper and is archived permanently on Zenodo (LeVeque et al. 2016). This code repository also includes Jupyter notebooks that present additional documentation of the code and illustration of results. The interested reader can also experiment with changing the parameters to see how this affects the results, and is welcome to build on this code for other projects.

Some of this code has already been adapted to the generation of seismic waveforms in the FakeQuakes module of MudPy (Melgar 2016). A paper describing this work and its application to the generation of synthetic Cascadia events is under review (Melgar et al. 2016b). This work includes several extensions of the approach presented here, including the use of a triangulation of the Slab 1.0 fault geometry as the fault geometry (using the K-L expansion to define slip patterns that are piecewise constant on triangles), the generation of events of smaller magnitude over subsections of the fault, and the incorporation of randomly varying epicenters with time-dependent rupture.

Acknowledgments

The initial phase of this work was performed when GL was employed at Pacific Northwest National Laboratory (PNNL) and KW was a postdoctoral fellow supported in part by PNNL and by the University of Washington (UW). This work was also supported in part by NSF Grants DMS-1216732 and EAR-1331412, funding from FEMA, and the Applied Mathematics Department of UW. The authors have benefitted from discussions with many applied mathematicians and geoscientists concerning the approach developed in this paper, including in particular Art Frankel, Finn Løvholt, Martin Mai, Siddhartha Mishra, Diego Melgar, and Hong Kie Thio. Numerous suggestions from the referees improved the quality of this paper.

References

Adams, L., LeVeque, R., & González, F. (2015). The pattern-method for incorporating tidal uncertainty into probabilistic tsunami hazard assessment (PTHA). *Natural Hazards*, *76*, 19–39.

Anderson, J. G. (2015). The composite source model for broadband simulations of strong ground motions. *Seismological Research Letters*, *86*(1), 68–74. doi:10.1785/0220140098.

Bastos, L. S., & O'Hagan, A. (2009). Diagnostics for Gaussian process emulators. *Technometrics*, *51*(4), 425–438. doi:10.1198/TECH.2009.08019.

Benner, P., Gugercin, S., & Willcox, K. (2015). A survey of model reduction methods for parametric systems. *SIAM Review*, *57*, 483531.

Chock, G.Y.K. (2015). The ASCE 7 tsunami loads and effects design standard. In N. Ingraffea, M. Libby (Eds.), *Structures congress 2015* (pp. 1446–1456). American Society of Civil Engineers. doi: 10.1061/9780784479117.124.

Cliffe, K. A., Giles, M. B., Scheichl, R., & Teckentrup, A. L. (2011). Multilevel Monte Carlo methods and applications to elliptic PDEs with random coefficients. *Computing and Visualization in Science*, *14*(1), 3–15. doi:10.1007/s00791-011-0160-x.

Dettmer, J., Hawkins, R., Cummins, P. R., Hossen, J., Sambridge, M., & Hino, R., et al. (2016). Tsunami source uncertainty estimation: The 2011 Japan tsunami. *Journal of Geophysical Research: Solid Earth*. doi:10.1002/2015JB012764.

Dick, J., Kuo, F. Y., & Sloan, I. H. (2013). High-dimensional integration: The quasi-Monte Carlo way. *Acta Numerica*, *22*, 133–288. doi:10.1017/S0962492913000044.

Dreger, D. S., Beroza, G. C., Day, S. M., Goulet, C. A., Jordan, T. H., Spudich, P. A., et al. (2015). Validation of the SCEC broadband platform V14.3 Simulation methods using pseudospectral acceleration data. *Seismological Research Letters*, *86*(1), 39–47. doi:10.1785/0220140118.

Frankel, A. (1991). High-frequency spectral falloff of earthquakes, fractal dimension of complex rupture, B value, and the scaling of strength on faults. *Journal of Geophysical Research*, *96*, 6291–6302.

Geist, E. L. (2002). Complex earthquake rupture and local tsunamis. *Journal of Geophysical Research*, *107*(B5), 2086. doi:10.1029/2000JB000139.

Geist, E. L., & Parsons, T. (2006). Probabilistic analysis of tsunami hazards. *Natural Hazards*, *37*, 277–314.

Geist, E. L., Parsons, T., ten Brink, U. S., & Lee, H. J. (2009). In E. N. Bernard, A. R. Robinson (Eds.), *The Sea* (Vol. 15)., Tsunami probability. Harvard University Press.

Ghanem, R. (1999). The Nonlinear Gaussian spectrum of log-normal stochastic processes and variables. *Journal of Applied Mechanics*, *66*(4), 964–973. doi:10.1115/1.2791806.

Ghanem, R. G., & Spanos, P. D. (1991). *Stochastic finite elements: A spectral approach*. New York: Springer.

Giles, M. B. (2008). Multilevel Monte Carlo path simulation. *Operations Research*, *56*, 607–617. doi:10.1287/opre.1070.0496.

Goda, K., Mai, P. M., Yasuda, T., & Mori, N. (2014). Sensitivity of tsunami wave profiles and inundation simulations to earthquake slip and fault geometry for the 2011 Tohoku earthquake. *Earth, Planets and Space*, *66*(1), 1–20. doi:10.1186/1880-5981-66-105.

Goda, K., Li, S., Mori, N., & Yasuda, T. (2015). Probabilistic tsunami damage assessment considering stochastic source models: Application to the 2011 Tohoku earthquake. *Coastal Engineering Journal*, *57*(03), 1550,015. doi:10.1142/S0578563415500151.

González, F. I., LeVeque, R. J., Adams, L. M., Goldfinger, C., Priest, G. R., & Wang, K. (2014). Probabilistic tsunami hazard assessment (PTHA) for Crescent City, CA.

González, F. I., Geist, E. L., Jaffe, B., Knolu, U., Mofjeld, H., Synolakis, C. E., et al. (2009). Probabilistic tsunami hazard assessment at Seaside, Oregon, for near-and far-field seismic sources. *Journal of Geophysical Research*, *114*(C11), 023.

Graves, R. W., & Pitarka, A. (2010). Broadband ground-motion simulation using a hybrid approach. *Bulletin of the Seismological Society of America*, *100*(5A):2095–2123. doi:10.1785/0120100057. http://www.bssaonline.org/cgi/doi/10.1785/0120100057

Guatteri, M., Mai, P. M., Beroza, G. C., & Boatwright, J. (2003). Strong ground-motion prediction from stochastic-dynamic source models. *Bulletin of the Seismological Society of America*, *93*(1), 301–313. doi:10.1785/0120020006.

Huang, S. P., Quek, S. T., & Phoon, K. K. (2001). Convergence study of the truncated Karhunen-Loève expansion for simulation of stochastic processes. *International Journal for Numerical Methods in Engineering*, *52*(9), 1029–1043.

Jaimes, M. A., Reinoso, E., Ordaz, M., Huerta, B., Silva, R., Mendoza, E., et al. (2016). A new approach to probabilistic earthquake-induced tsunami risk assessment. *Ocean and Coastal Management*, *119*, 68–75. doi:10.1016/j.ocecoaman.2015.10.007.

Karhunen, K. (1947). Über lineare Methoden in der Wahrscheinlichkeitsrechnung (vol 37). Universitat Helsinki

Lavallée, D., Liu, P., & Archuleta, R. J. (2006). Stochastic model of heterogeneity in earthquake slip spatial distributions. *Geophysical Journal International*, *165*(2), 622–640. doi:10.1111/j.1365-246X.2006.02943.x.

LeVeque, R. J., Waagan, K., González, F. I., Rim, D., & Lin, G. (2016). Code to accompany this paper. https://github.com/rjleveque/KLslip-paper. doi:10.5281/zenodo.59720.

Li, J., Li, J., & Xiu, D. (2011). An efficient surrogate-based method for computing rare failure probability. *Journal of Computational Physics*, *230*, 8683–8697. doi:10.1016/j.jcp.2011.08.008.

Loève, M. (1977). *Probability theory* (4th ed.). Berlin: Springer-Verlag.

Lorito, S., Selva, J., Basili, R., Romano, F., Tiberti, M. M., & Piatanesi, A. (2015). Probabilistic hazard for seismically induced tsunamis: Accuracy and feasibility of inundation maps. *Geophysical Journal International*, *200*(1), 574–588. doi:10.1093/gji/ggu408.

Løvholt, F., Pedersen, G., Bazin, S., Kühn, D., Bredesen, R. E., & Harbitz, C. (2012). Stochastic analysis of tsunami runup due to heterogeneous coseismic slip and dispersion. *Journal of Geophysical Research*, *117*(C3). doi:10.1029/2011JC007616.

Mai, P. M., & Beroza, G.C. (2002). A spatial random field model to characterize complexity in earthquake slip. *Journal of Geophysical Research*, *107*, ESE10–1–ESE10–21

Melgar, D. (2016). MudPy software. http://www.github.com/dmelgarm/MudPy

Melgar, D., Allen, R. M., Riquelme, S., Geng, J., Bravo, F., Baez, J. C., et al. (2016a). Local tsunami warnings: Perspectives from recent large events. *Geophysical Research Letters*, *43*(3), 2015GL067,100. doi:10.1002/2015GL067100.

Melgar, D., LeVeque, R. J., Dreger, D. S, & Allen, R.M. (2016b). Kinematic rupture scenarios and synthetic displacement data: An example application to the Cascadia Subduction Zone. Submitted

Nobile, F., Tempone, R., & Webster, C. G. (2008). A sparse grid stochastic collocation method for partial differential equations with random input data. *SIAM Journal on Numerical Analysis*, *46*(5), 2309–2345. doi:10.1137/060663660.

Nosov, M. A., Bolshakova, A.V., & Kolesov, S.V. (2014). Displaced water volume, potential energy of initial elevation, and tsunami intensity: Analysis of recent tsunami events. *Pure and Applied Geophysics*, *171*(12), 3515–3525. doi:10.1007/s00024-013-0730-6. http://link.springer.com/10.1007/s00024-013-0730-6

Okada, Y. (1985). Surface deformation due to shear and tensile faults in a half-space. *Bulletin of the Seismological Society of America*, *75*, 1135–1154.

Olsson, A., & Sandberg, G. (2002). Latin hypercube sampling for stochastic finite element analysis. *Journal of Engineering Mechanics*, *128*(1), 121–125. doi:10.1061/(ASCE)0733-9399(2002)128:1(121).

Peherstorfer, B., Cui, T., Marzouk, Y., & Willcox, K. (2016). Multifidelity importance sampling. *Computer Methods in Applied Mechanics and Engineering*, *300*, 490–509. doi:10.1016/j.cma.2015.12.002.

Pollitz, F. F., McCrory, P.A., Wilson, D., Svarc, J., Puskas, C., & Smith, R.B. (2010). Viscoelastic-cycle model of interseismic deformation in the northwestern United States. *Geophysical Journal International*, *181*, 665–696. http://doi.wiley.com/10.1111/j.1365-246X.2010.04546.x

Razafindrakoto, H. N. T., Mai, P. M., Genton, M. G., Zhang, L., & Thingbaijam, K. K. S. (2015). Quantifying variability in earthquake rupture models using multidimensional scaling:

Application to the 2011 Tohoku earthquake. *Geophysical Journal International*, *202*(1), 17–40. doi:10.1093/gji/ggv088.

Sarri, A., Guillas, S., & Dias, F. (2012). Statistical emulation of a tsunami model for sensitivity analysis and uncertainty quantification. *Natural Hazards and Earth System Sciences*, *12*(6), 2003–2018. doi:10.5194/nhess-12-2003-2012.

Schwab, C., & Todor, R. A. (2006). Karhunen-Loève approximation of random fields by generalized fast multipole methods. *Journal of Computational Physics*, *217*, 100–122.

Stein, S., Geller, R. J., & Liu, M. (2012). Why earthquake hazard maps often fail and what to do about it. *Tectonophysics*, *562–563*, 1–25. doi:10.1016/j.tecto.2012.06.047.

USGS. (2010). USGS earthquake archive. http://earthquake.usgs.gov/earthquakes/eqinthenews/2010/us2010tfan/finite_fault.php

Van Trees, H. L., Bell, K. L., & Tian, Z. (2013). *Detection estimation and modulation theory, part I, detection, estimation, and filtering theory* (2nd ed.). New Jersey: Wiley.

Wang, K., & He, J. (2008). Effects of frictional behavior and geometry of subduction fault on coseismic seafloor deformation. *Bulletin of the Seismological Society of America*, *98*, 571–579. doi:10.1785/0120070097.

Witter, R. C., Zhang, Y. J., Wang, K., Priest, G. R., Goldfinger, C., Stimely, L., et al. (2013). Simulated tsunami inundation for a range of Cascadia megathrust earthquake scenarios at Bandon, Oregon, USA. *Geosphere*, *9*(6), 1783–1803. doi:10.1130/GES00899.1.

Zhang, L., Mai, P. M., Thingbaijam, K. K., Razafindrakoto, H. N., & Genton, M. G. (2015). Analysing earthquake slip models with the spatial prediction comparison test. *Geophysical Journal International*, *200*(1), 185–198. doi:10.1093/gji/ggu383.

(Received April 27, 2016, revised July 9, 2016, accepted July 14, 2016, Published online August 16, 2016)