ISSN 1069-3513, Izvestiya, Physics of the Solid Earth, 2011, Vol. 47, No. 3, pp. 176–185. © Pleiades Publishing, Ltd., 2011.
Original Russian Text © A.A. Gusev, 2011, published in Fizika Zemli, 2011, No. 3, pp. 24–33.

Statistics of the Values of a Normalized Slip in the Points of an Earthquake Fault

A. A. Gusev[a,b]

[a] *Institute of Volcanology and Seismology, Russian Academy of Sciences, bulv. Piip 9, 683006 Petropavlovsk-Kamchatskii, Russia*

[b] *Kamchatka Branch, Geophysical Service, Russian Academy of Sciences, bulv. Piip 9, 683006 Petropavlovsk-Kamchatskii, Russia*

Received September 23, 2009

Abstract—In order to develop a statistical description for the values of a final slip (dislocation) in different points of an area of an earthquake source fault, we analyzed the recently compiled collection extended-fault descriptions, obtained by solution of the corresponding inverse problem. The probability distribution function was studied for the normalized average slip in a subfault (an element of a large fault). Each large fault yielded by an individual inversion was normalized by division of the slip value in a subfault to the mean value averaged across the subfaults of a given fault. The processing included the following steps: informal rejection of less reliable inversions, normalization, and construction of empirical distribution functions for individual earthquakes and for the pooled sample. We estimated the parameters of the empirical functions and approximated them by simple standard distribution laws. The statistical structure of the values of the slip is found to be rather stable. The individual empirical samples have the coefficient of variation 0.98 ± 0.28 and, generally, resemble those with the exponential distribution law. The upper tail of the distribution rather sharply tapers off, following, on average, the power law with the exponent α being approximately from -3.5 to -4. The composite distribution has a noticeable atom at zero with a weight of about 10%. The presence of this atom impedes approximation of the observed distribution by a simple law. We proposed a reasonable approximation by a modified lognormal law; the modification includes shifting to the left and winsorization at zero. The appearance of the atom at zero is possibly due to the unbreakable barriers at propagation of a rupture; at the same time, we cannot rule out the possibility for it to be an artifact generated by the procedure of inversion. Our results provide good grounds for practical simulations of the scenario earthquakes; they also are of interest for the physics of the earthquake source.

Keywords: earthquake source, fault, slip, dislocation, statistic, distribution law.

DOI: 10.1134/S1069351310101015

INTRODUCTION

In recent decades, dozens of records of strong motions were obtained in the close vicinity of earthquake sources for a variety of strong earthquakes worldwide. Although the results of kinematical inversion for the source, based on these data, are still not quite reliable, they, however, provide a sound basis for a first approximation of the properties of the real sources. A number of important results has been obtained on this way. For example, Heaton [1990] showed that the earthquake rupture behaves as a "slip pulse", which has a limited duration and propagates in the wake of the front of the rupture. Tsai [1997] found that the two-dimensional (2D) spatial spectrum of the final slip function is close to the power law spectrum. This inference has confirmed the fractal structure of this function hypothesized in [Andrews, 1980]. It was recognized that the spatial distribution of the final slip over the area of a source fault is essentially nonuniform: it exhibits local spots of increased values, the so-called "asperities" (see, e.g., [Somerville et al., 1999]). To make the exposition clear and to avoid the confusion, the term "distribution" in the further text will be used only in a probabilistic sense, and the final slip (dislocation) as a function of coordinates on the area of the source fault will be referred to as the "discrete slip model" (DSM). Several important regularities in the distribution of the values of the final slip function were first established in [Somerville et al., 1999]; later, they were successfully applied by Irikura [2006] in the strong motion simulations. The common approach in these works is to describe the discrete slip model as a set of individual objects in the form of rectangular "asperities", superimposed on the uniform background. The question regarding the structure of the distribution function for the values of a slip remains poorly studied. Based on the evident qualitative agreement and a few tests (not published), in our works on modeling the source [Gusev and Pavlov, 2006; 2009; Gusev et al., 2008], we assumed this distribution to be lognormal. This topic is worth a detailed analysis.

Lately, Martin Mai (USA–Switzerland) has created an online database of finite-source rupture mod-

els (SRCMOD, [Mai, 2004]). This collection includes over 150 solutions of the inverse problem for a source. The solutions are represented in a common format. The use of this valuable data source, which incorporates both the models from the published papers and the authors' models contributed by a number of researchers, allowed us to solve the problem formulated above with reasonable efforts. The key challenge here is the selection of admissible discrete slip models based on poorly formalizable criteria rather than the analysis itself of the distribution function, which is relatively easy. In this situation, the results of analysis are unavoidably tentative. However, even such analysis is of interest since the statistic of a slip contains information on the mechanics of sliding on the rupture. This statistic is also necessary for predicting the strong motions of the soil. The further description is organized in the following way. First, we describe the selection and preprocessing of the most reliable slip inversions. Then we construct the empirical distribution functions, primarily for individual events and then for a merged sample (in fact, by averaging across the events). Finally, we carry out parameterization, approximation and analysis of these functions.

DATA SELECTION AND ANALYSIS OF THE INDIVIDUAL DISTRIBUTION FUNCTIONS

The latest (as of 2007) version of SRCMOD database includes 152 solutions of inversions (kinematics of a rupture, final slip or DSM, etc.) for 80 earthquakes. These data are extremely heterogeneous as to their reliability and quality. Obtaining rather reliable results of inversion requires, first, that the input data include a sufficient number of seismograms recorded close to the earthquake source (i.e., "near-source" records). In fact, a part of solutions is derived from the inversion of teleseismic seismograms alone, and only half of DSM is constructed using a considerable volume of "near-source" records. Second, the inversion, which usually invokes the strong motion velocities as the input data, often underestimates the value of the total seismic moment relative to that known from the low-frequency surface wave data, which are rather reliable. Some authors try to recover the lost fraction of the total seismic moment by adding it to the already reconstructed DSM in the form of a constant background, an operation that radically distorts the statistical structure of the values of a slip. Third, some researchers apply a dense grid of unknowns in the inversion, which implies that the estimates of the values in the points of such a grid are strongly correlated for the neighboring subsources. This introduces drastic changes in the statistic for the tails of distribution; besides, it artificially exaggerates the amount data. Therefore, the following data types were rejected straight away: those for events earlier than of 1968, those without the "near-source" data, those with definitely nonzero background, and those with the number of subsources (DSM grid points) over 350 (with one exception).

The DSMs remaining after such a selection still are clearly inhomogeneous: some appear to be "undersmoothed" by the inversion, which can make the tails of distribution artificially heavier, and the others seem to be "oversmoothed", which could artificially narrow down the width of a peak of distribution. However, no formal criteria for rejection of such cases were formulated, and the selection of suitable DSM was therefore inevitably subjective. We picked out 37 DSM overall. After selection, if there was more than one DSM for one event, each DSM was assigned a weight $1/K$, where K is the number of DSM for the given event. (Otherwise, the greater the number of publications, the heavier the weight of the event would be.) Then, the weights were used in the calculations of mean values. A brief summary of characteristics of the selected DSM are presented in the table. The detailed data for each event are available on http://www.seismo.ethz.ch/srcmod/Events.html, and the relevant references can be found on http://www.seismo.ethz.ch/srcmod/References.html.

The main input data for the analysis are M DSM matrices (the slipSPL arrays) containing the values of a slip in the subsources $S_{pq,k}$, where $k = 1, 2, ..., m$ is the serial number of the event, pq are indices of the matrix entries. Before processing, the initial matrices $S_{pq,k}$ were trimmed by discarding the marginal rows and columns if these rows or columns contained only zero entries. Then each matrix $S_{pq,k}$ containing N_k elements was reshaped into a vector S_{ik}, where $i = 1, 2, ..., N_k$, and normalized:

$$s_{ik} = S_{ik} \Big/ \left(\frac{1}{N_k} \sum_{m=1}^{N_k} S_{ik} \right) \qquad (1)$$

thereby, we obtained M sets of normalized slips s_{ik} with unit average.

Figure 1 presents the example of DSM and processing results for two very thoroughly characterized earthquakes (one crustal and one subduction). Here, the cumulative empirical distribution functions for a normalized slip (the common cumulative distribution function (cdf) ($P(s)$ (and the complementary cumulative distribution function (ccdf)) $Q(s) = 1 - P(s)$) as well as the histogram are shown. As seen from the examples, even after elimination of the zero columns and rows, the matrices contain a considerable number of zero entries (about 9% overall); however, in fact they are present only on a part of DSM. The analysis of the initial DSM suggests that there are two types of zero entries. Zeros of the first type do not occur as a continuous border; rather, they are interspersed with positive $S_{pq,k}$. They may be caused by various reasons. First, an additional constraint of non-negative $S_{pq,k}$, is often imposed in the inversion, therefore the true small values of $S_{pq,k}$ may appear as zeros in the real noised

The discrete slip models used and the estimates of the parameters of its probability distributions[1]

Date	Event name[2]	M_w	Source [weight]	n_L	n_W	n	n_{z0}	n_1	L	W	CV_k	$s_{k,1\%}$	α_k	$\sigma_{\log}$	k	γ
04/01/1968	Hyuga-nada (JP)	7.53	Yagi et al. (1998)	8	7	56	0	56	72	63	0.77	3.03	6.75	0.68	1.71	1.33
05/16/1968	Tokachi-oki (JP)	8.35	Nagai et al. (2001)	12	6	72	0	72	240	120	0.76	3.78	2.82	0.68	1.72	1.33
10/15/1979	ImperialValley (CA)	6.53	Archuleta (1984) [0.5]	15	14	210	9	204	37.5	14	0.89	3.48	6.87	0.77	1.24	1.11
10/15/1979	ImperialValley (CA)	6.58	Hartzell and Heaton (1983) [0.5]	14	4	56	7	50	42	10.4	0.82	3.03	6.4	0.71	1.48	1.23
04/24/1984	MorganHill (CA)	6.28	Beroza and Spudich (1988) [0.5]	61	11	671	158	671	30.5	11	1.14	4.07	5.35	0.91	0.77	0.88
04/24/1984	MorganHill (CA)	6.07	Hartzell and Heaton (1986) [0.5]	27	6	162	69	158	27	11.5	1.51	5.76	5.5	1.09	0.46	0.68
03/03/1985	CentralChile (Chile)	8.16	Mendozaetal (1994)	17	11	187	7	184	255	165	0.93	3.69	4.36	0.79	1.13	1.07
09/19/1985	Michoacan (Mexico)	8.01	Mendoza and Hartzell (1989)	12	10	120	4	120	180	139	0.90	4.45	3.09	0.77	1.23	1.11
12/23/1985	Nahanni2 (Canada)	6.66	Hartzell et al. (1994)	18	9	162	80	122	48	21.2	1.94	9.35	2.48	1.25	0.30	0.56
07/08/1986	N. PalmSprings (CA)	6.21	Hartzell (1989)	11	8	88	9	84	22	15.2	0.80	3.07	5.56	0.70	1.58	1.27
11/24/1987	Superstit.Hills (CA)	6.51	Wald et al. (1990)	20	10	2	0	200	20	11.5	1.01	4.84	2.09	0.83	0.97	0.99
10/18/1989	LomaPrieta (CA)	6.94	Wald et al. (1991)	12	7	84	5	84	40	17.5	0.84	3.82	4.52	0.73	1.41	1.19
06/28/1991	SierraMadre (CA)	5.59	Wald (1992)	7	10	70	27	54	4.9	6	1.38	5.51	3.85	1.03	0.52	0.74
06/28/1992	Landers (CA)	7.28	Wald and Heaton (1994)	26	6	156	11	152	78	15	0.85	3.28	5.23	0.73	1.39	1.18
07/12/1993	Hokkaido area (JP)	7.60	Mendoza and Fukuyama (1996)	20	7	140	9	140	2	70	1.08	5.52	2.29	0.87	0.85	0.93
01/17/1994	Northridge (CA)	6.73	Hartzell et al. (1996) [0.5]	14	14	196	13	185	20	24.9	0.70	2.99	5.42	0.63	2.04	1.46
01/17/1994	Northridge (CA)	6.8	Wald et al. (1996) [0.5]	14	14	196	10	196	18	24	0.66	2.71	4.9	0.60	2.36	1.56
12/28/1994	Sanrikuki (JP)	7.7	Nagai et al. (2001)	11	14	154	0	154	110	140	1.05	5.29	2.25	0.86	0.90	0.95
10/09/1995	Colima (Mexico)	7.96	Mendoza and Hartzell (1999)	20	10	2	15	200	2	1	0.92	3.89	3.52	0.78	1.18	1.09
01/17/1995	Kobe (JP)	7.02	Sekiguchi et al. (2002) [0.5]	31	10	310	0	310	63.5	20.5	0.77	3.96	3.68	0.68	1.7	1.32
01/17/1995	Kobe (JP)	6.92	Wald (1996) [0.5]	18	8	144	0	144	60	20	0.75	4.49	2.70	0.66	1.8	1.36
12/02/1996	Hyuga-nada2 (JP)	6.68	Yagi and Kikuchi (1999)	10	10	1	0	100	29.2	29.2	0.68	3.85	2.65	0.61	2.19	1.49
03/26/1997	Kagoshima (JP)	6.04	Miyakoshi et al. (2000)	9	6	54	9	51	18	12	0.99	4.16	3.73	0.82	1.02	1.01
06/25/1997	Yamaguchi (JP)	5.82	Miyakoshi et al. (2000)	8	7	56	19	49	16	14	1.55	7.04	2.70	1.11	0.45	0.67
09/20/1999	ChiChi (Taiwan)	7.69	Ma et al. (2001) [0.33]	21	8	168	34	163	105	40	1.11	4.87	2.98	0.89	0.80	0.9
09/20/1999	ChiChi (Taiwan)	7.68	Chi et al. (2001) [0.33]	32	10	320	18	309	112	35	0.88	4.27	3.14	0.75	1.3	1.14

Table (Contd.)

Date	Event name[2]	M_w	Source (weight)	n_L	n_W	n	n_{z0}	n_1	L	W	CV_k	$s_{k,1\%}$	α_k	$\sigma_{\log}$	k	γ
09/20/1999	ChiChi (Taiwan)	7.63	Sekiguchi et al. (2002) [0.33]	26	13	338	0	338	78	39	0.60	2.71	5.78	0.55	2.8	1.74
10/16/1999	HectorMine (CA)	7.14	Salichon et al. (2003)	18	6	108	0	108	54	18	1.1	4.94	3.71	0.89	0.8	0.91
08/17/1999	Izmit (Turkey)	7.44	Sekiguchi and Iwata (2002)	42	8	336	15	331	126	23.3	0.84	4.02	2.84	0.73	1.4	1.18
09/30/1999	Oaxaca (Mexico)	7.47	Hernandez et al. (2001)	12	6	72	1	72	90	45	0.97	3.8	4.48	0.81	1.06	1.03
10/06/200	Tottori (JP)	6.73	Semmane et al. (2005)	16	10	160	34	160	32	20	1.04	4.15	3.75	0.85	0.93	0.96
03/24/2001	Geiyo (JP)	6.79	Sekiguchi and Iwata (2002)	10	7	70	0	70	30	21	0.80	3.54	4.21	0.70	1.57	1.27
05/21/2003	Boumerdes (Algeria)	7.25	Semmane et al. (2005)	16	8	128	@	@	64	32	0.549	2.56	7.13	0.513	@	@
01/22/2003	Colima (Mexico)	7.5	Yagi et al. (2003)	14	17	238	0	238	70	85	1.08	4.87	3.15	0.88	0.85	0.92
09/25/2003	Tokachi-oki (JP)	8.16	Yagi (2004)	13	17	221	0	221	130	170	0.85	3.71	4.81	0.74	1.38	1.18
09/28/2004	Parkfield (CA)	6.06	Custodio et al. (2005) [0.5]	21	9	189	0	189	40	15	1.31	6.4	2.38	1.	0.58	0.77
09/28/2004	Parkfield (CA)	6.0	Dreger et al. (2004) [0.5]	15	7	105	22	89	28.3	12.1	1.22	6.44	2.63	0.95	0.67	0.82
03/20/2005	Fukuoka (JP)	6.67	Asano and Iwata (2006)	13	9	117	0	117	26	18	0.80	3.58	3.41	0.70	1.57	1.27
			Average								**0.98**	**4.37**	**3.94**	**0.80**	**1.27**	**1.11**
			standard deviation								*0.28*	*1.34*	*1.43*	*0.16*	*0.66*	*0.29*
			coefficient of variation								0.29	0.31	0.36	0.20	0.52	0.26
			MSE[23]								**0.91**	**4.37**	**3.25**	**0.77**	**1.21**	**1.10**

[1] For the full references to the sources and other details see http://www.seismo.ethz.ch/srcmod/References.html.

[2] The region is indicated in brackets; CA standing for California, and JP standing for Japan).

[3] Merged sample estimate.

Designations: n – total number of subsources, n_L – size of the subsource grid along the source length, n_W – same, along the source width, n_{z0} – number of sources (among n) with zero values of the slip, n_1 – number of subsources within a convex hull; L – source length, W – source width; Parameters of the slip in the subsource fault normalized to unity mean: CV – coefficient of variation, $s_{1\%}$ – upper 1% quantile, α – exponent in the approximation of empirical $P(s)$; parameters of the approximating model distributions $P(s)$: $\sigma_{\log}$ – for lognormal distribution, k – for gamma distribution, γ – for Weibull distribution.

inversion. In addition to these artifacts, we may expect that there are as well true spots of zero slips (the "barriers") and some portion of true negative values of $S_{pq,k}$. All these zero values reflect the reality, and it is inadmissible to discard them. Zeros of the second type arise when the entire rectangular area that represents the source in the inversion is treated as consisting of possibly non zero unknown elements while the real source occupies only a part of this rectangle (see the example in Fig. 1d) and corresponds to a certain domain of nonzero $S_{pq,k}$ outlined by a rather sharp boundary. The zero values outside this contour are pure artifacts, and it is desirable to eliminate them from a sample.

For approximate solution of this problem, we selected a subset of nonzero entries of the matrix S_{pq} and considered the centers of the corresponding cells–subsources on a plane. For the set of the centers, we constructed a convex hull; all cells whose centers fell outside this hull were discarded. The boundaries of the real sources are not necessarily convex, therefore a certain portion of false zero values is still retained. Nevertheless, the procedure is, generally, rather efficient: it reduced the total number of zero values from 9% to 6.5%.

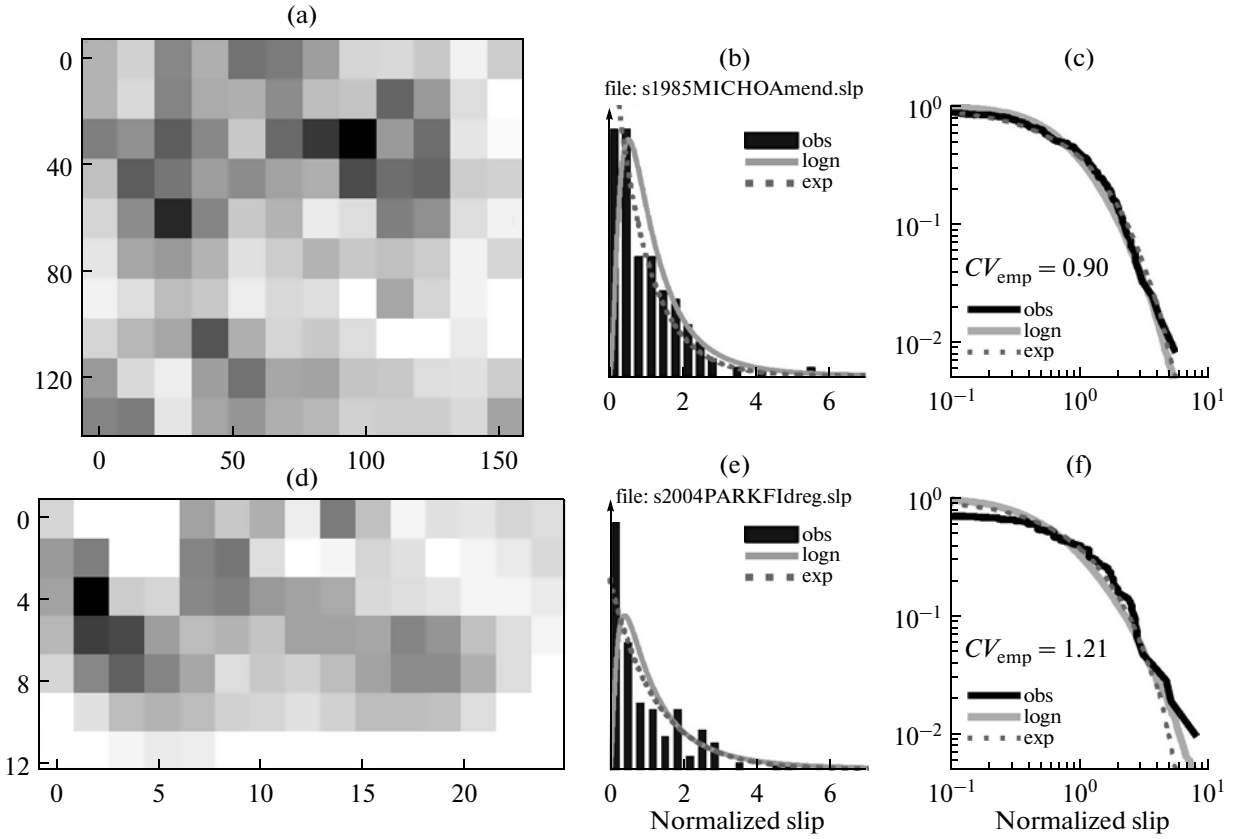

Fig. 1. The examples of individual discrete slip models and their distribution functions: (a) distribution of a slip across the subsources in the source fault of the Michoacan, 1985 earthquake, according to the inversion [Mendoza and Hartzell, 1989]; the value of the slip is shown by the intensity of gray, white color corresponds to zero slip. The abscissa indicates the distance along the strike, in kilometers; the ordinate is the distance along the dip, km; (b) and (c) are the histogram and the complementary cumulative distribution function for the normalized slip compared to two model distributions; (d), (e), and (f) are similar graphs for the Parkfield earthquake of 2004, according the inversion [Custodio et al., 2005].

From a set of s_{ik} values of a particular inversion, we constructed the variational series and calculated the parameters of samples using these series. The key parameter for the studied distributions of non-negative quantities is the coefficient of variation $CV_k = \sigma(s_{ik})/\mu(S_{ik}) \equiv \sigma(s_{ik})$, where $\sigma(\cdot)$ denotes the standard deviation and $\mu(\cdot)$ the mean value. We also calculated the upper F-percent quantiles $s_{k,F}$ of s_{ik} distribution (so as $Q(s_{k,F}) = 0.01F\%$ for $F = 1,2,10$ and 25%. Using these values, for $F = F1 = 2\%$ and $F = F2 = 10\%$ we constructed estimations of the exponent α for the power-law (hyperbolic) approximation of the upper tail of the distribution:

$$\alpha = -\log(s_{k,F1}/s_{k,F2})/\log(F1/F2). \quad (2)$$

The individual values of a number of the individual statistical characteristics of DSM are presented in the Table, where the (weighted averages) and the standard deviations of the individual values are indicated as well. Figure 2 displays the histograms of the studied parameters. The values of CV_k, $s_{k,1\%}$ and α_k are indicated in the top row. We see that the values of CV_k and $s_{k,1\%}$ exhibit a considerable scatter (their coefficient of variation is about 0.3), although they are clearly one-humped, whereas the distribution of α_k is closer to uniform, and its coefficient of variation is higher (0.36). The second row from the top presents the parameters studied in [Sommerville et al., 1999], namely the value of Q (1.5) (the probability for the value of s_{ik} to exceed the threshold of 1.5); the value of the upper 25% quantile (hereinafter referred to as $s_{25\%}$), and the mean across the values of s_{ik} exceeding this threshold ($s_{25\%}$) (hereinafter referred to as $s_{aver25\%}$).

The empirical distribution functions $P(s_{ik})$ were approximated by the theoretical distribution functions. We have analyzed the lognormal distribution with the probability density in the following form

$$p(s) = \frac{1}{(2\pi)^{0.5} s\sigma_{\log}} \exp\left[-\frac{(\ln s - \mu)^2}{2\sigma_{\log}^2}\right], \quad (3)$$

the gamma distribution with the density defined as

$$p(s) = \frac{\beta^k}{\Gamma(k)} s^{k-1} \exp(-\beta s) \quad (4)$$

and the Weibull distribution with the density

$$p(s) = \gamma a^{-\gamma} s^{\gamma-1} \exp\left[-\left(\frac{s}{a}\right)^{\gamma}\right]. \quad (5)$$

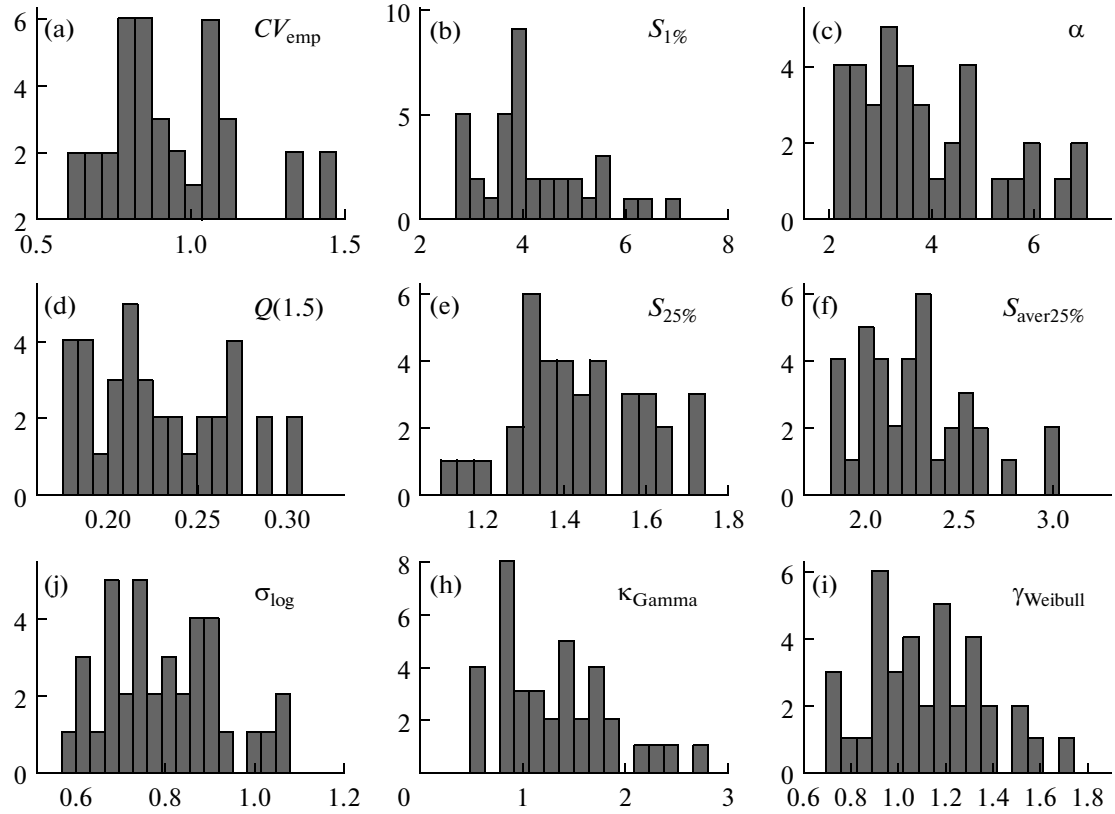

Fig. 2. The histograms of the parameters of distributions for 37 discrete slip models.

The lognormal distribution has been earlier applied for the practical simulation of DSM, and there was a need to verify it. The gamma distribution is attractive by its being infinitely divisible. This feature can considerably simplify the analysis and simulation in the cases when the initial discrete slip model is reproduced in averaged form with fine detail using different sizes of the smoothing window (that is, the size of the subsource) and, correspondingly, different N. The gamma distribution and the Weibull distribution are reduced to the one-parameter exponential distribution at $k = 1$ and $\gamma = 1$, respectively.

In each of the three cases studied, the scale parameter (μ, β, or a), which is of no interest in the given statement of the problem, was specified in such a way as to satisfy the unity mean condition. The shape parameter (s_{log}, k, or γ) was fitted either according to the coincidence condition for the empirical and the theoretical CV, or in a similar way according to the coincidence condition for the upper 2%-quantiles. These two approaches provide only slightly different estimates. The characteristics for the shape parameter are presented in Fig. 2 and in the table. The numerical results indicate that the empirical distributions are surprisingly close to the exponential distribution (which implies that $s_{log} = 0.833$, $k = 1$, $\gamma = 1$).

Each empirical DSM $S_{pq,k}$ has its own level of detail defined by the number N_k of the elementary areas–subsources. It is quite probable, however, that the distribution function for the values of the normalized slip s_{ik} depends on N_k, and it is quite probable that the obtained results are distorted by this factor. We illustrate this by a synthetic example. Let the true (unobservable) slip be a discrete function defined on a dense grid containing a large number N^* of small cells. Let this function be the delta-correlated random function (2D white noise). Then, the slip in each subsource is the mean across N^*/N small cells and, as it can easily be seen, its coefficient of variation CV grows as $N^{0.5}$ as N increases. Here we assumed a slip to be uncorrelated for the pairs of the subsources. This assumption is knowingly invalid, and, surely, there is actually no such a well-pronounced CV dependence on N, although we may still expect CV to slightly increase as the number of cell increases. However, we were not able to reveal any distinct trend of such type. Therefore, from the viewpoint of a possible dependence on N_k, a set of parameters of DSM, yielded by different inversions, can be regarded as a homogeneous sample. Following the same scheme, in the next section of the paper, we merge the values s_{ik} returned by different inversions into a single sample.

THE DISTRIBUTION FUNCTION OF THE MERGED DATA AND ITS ANALYSIS

Construction of the empirical distribution function for a merged sample is simple in the case of equal-

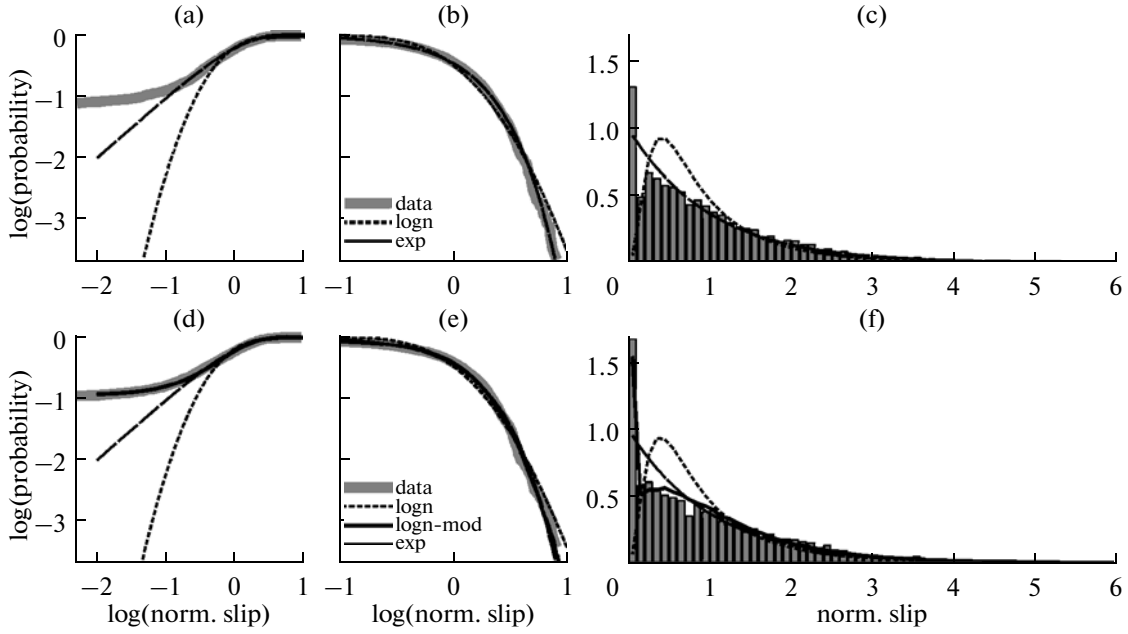

Fig. 3. The approximation of the empirical distributions (merged sample): (a), (b), and (c) for 37 DSM; (d), (e), and (f) for 23 DSM with zero values; (a) and (d) are the log-log plots of the cumulative distribution function $P(s)$ according to observations and for the theoretical models (exponential and lognormal distributions for the both cases and, additionally for the graphs (d), (e), (f), the modified lognormal distribution according to (8)); (b) and (e) are the similar graphs for the complementary cumulative function $Q(s) = 1 - P(s)$; (c) and (f) are the corresponding histograms and graphs for the theoretical density on a natural scale.

weighted data. However, in our study, it is expedient to introduce different weights ($1/K$) for the data. The cumulative empirical distribution function for this case was constructed in the following way. Each s_{ik} is assigned a weight w_{ik} (actually, at fixed k, the weights for all i are identical). We merge all s_{ik} into one sample and sort it ascending, thereby obtaining a variational series $s_j, j = 1, 2, ..., L$, where each s_j is associated with a weight w_j. The value of the empirical function $P(s)$ at $s = s_j$ is assumed to be

$$P(s_j) = \sum_{m=1}^{j} w_m \Big/ \sum_{m=1}^{L} w_m. \qquad (6)$$

The histograms were constructed for the illustrative purposes only; the weights were disregarded.

Figure 3 presents the empirical $P(s)$, $Q(s)$ and the histogram for the merged data. The zero values of s_j produce an atom at zero in the distribution. The weight of the atom, initially attaining 9%, was then reduced to 6.5% by using the convex hull as described above. We have estimated the same parameters for the merged sample as those estimated for the individual inversions. We have also estimated the parameter α in the interval between the remote upper quantiles $F1 = 0.5\%$ and $F2 = 5\%$. In this interval the tail of the merged empirical distribution tapers off as $Q(s) \sim s^{-3.9}$. The estimates of the parameters k and γ for the gamma distribution and the Weibull distribution again are close to unity: $k = 1.18$, $\gamma = 1.09$. For comparison, the theoretical curves for the lognormal and exponential distributions with the same CV values are plotted on the graphs in Fig. 3. We see on the $P(s)$ graph and, even more, in the histogram that the divergences are quite large, especially in the lower tail. An obvious reason for this divergence is the atom at zero, but this explanation is not comprehensive. One may notice that the exponential distribution does not reproduce the distinct hump apparent at approximately $s = 0.25$ in the histogram for empirical data; besides, the upper tail of this distribution is "lighter" than the empirical one. The lognormal distribution seems qualitatively rather more acceptable: it does predict the hump, although a less expressed and appearing at much larger s (≈0.40); the upper tail of this distribution is somewhat heavier than that of the empirical distribution.

Basically, given the limited accuracy of the input data, a pure phenomenological description of the observed distribution is possible (e.g., in a tabulated form). This is sufficient for practical simulations. A mixture of two or more distribution laws seems to be another plausible model for reproducing the empirical distributions. However, it is also important to derive a simple model description for the observed distribution, which is not identical to a standard distribution low or a mixture of any standard lows.

We note primarily that no theoretical concept for the statistic of the value of a slip was proposed in the literature. We conducted our search, that is described above, among such candidates as the lognormal distribution, the gamma distribution and the Weibull distribution without any confidence in their eligibility. Nev-

ertheless, the major difficulty on this way seems to be to account for the atom at zero, while any ready models to do this are absent. We took the tack of modifying the known probability distributions and selected the lognormal law to be the basis for our model, because this distribution at least qualitatively fits the data (except for the atom at zero). Our goal was to find such a model that reproduces the atom at zero and at the same time fits the upper tail, the coefficient of variation and, if possible, the position and the amplitude of the hump. Since the main purpose of the model is to account for the atom at zero, it is reasonable to invoke only the results of those inversions where zero values were admitted. The criterion indicating that the algorithm of inversion rejected the zero values was the absence of zeros in the DSM matrix, and DSM of such inversions were excluded from the consideration. Thereby, 23 DSM matrices from 37 ones studied at the previous stage were picked out for the further analysis; these correspond to the cases $n = n_1$ in the table.

Above, we have listed several factors, which might be responsible for the appearance of zero values. Let us consider these factors. Primarily, this is the influence of the noise of inversion. In the linear inversion, the positive and negative deviations of the estimates from the true value of the parameter are approximately equally probable. If the condition for the estimates to be non-negative is introduced into the algorithm of inversion, the effect of nonlinearity arises, which, roughly speaking, reduces to the fact that the DSM estimates in the cells, where they would have been negative in case of linear inversion, become zero. Consider the case when the true values of the slips are always positive (probably very small), so that the observed zero values are pure effect of noise. Monte-Carlo simulation of this situation includes two steps. First, a large number of random values x_i is generated according to the following rule:

$$x_i = L(0, \sigma_{\log}) + N(0, c\sigma_{\log}), \quad (7)$$

where $L(m, \sigma_{\log})$ and $N(m_1, \sigma_1)$ denote random numbers obeying the lognormal distribution (3) and the standard normal distribution, respectively. The first term in (7) is always positive; it reproduces the true value of the slip. The second term emulates the random error of inversion, which is symmetric about zero. Then the negative x_i are replaced by zeros, and, finally, the result is normalized by dividing it by the mean. By varying the parameters $\sigma_{\log}$ and c, the distribution of x_i is fitted to the observed distribution for s_{ik} so as the coefficient of variation CV_{emp} is 0.98 and the weight of the atom at zero p_z is 0.11. The fitted values are $\sigma_{\log} = 0.79$ and $c = 0.87$. The upper tail of the resultant distribution is somewhat heavier than in the empirical one, and the position of the hump near $s \approx 0.65$ is not quite adequate. Although these discrepancies are marginally acceptable, the model overall can hardly be considered plausible, since the value of c is rather large. At $CV \approx 1$, the value of c being close to unity means that the anticipated random errors are of the same order of magnitude as the estimated values themselves. Since, really, the errors of inversion are not necessarily small, this explanation of the atom at zero cannot be ruled out completely; however, it is unlikely the best one.

As the atom at zero cannot be fully accounted for by the shortcomings of the inversion, let us consider an alternative explanation. Namely, let the atom at zero reflect the geophysical reality. Ignoring the errors of inversion, we obtain two undistinguishable explanations in case of $s_{ik} = 0$: either the slip is really zero (the "unbreakable barrier," according to [Das and Aki, 1977]), or it is negative, which seems less probable but certainly quite possible. For the simulation purposes, it is convenient to use the second alternative and to assume that the true value of s_{ik} is negative while the observed value is, however, zero. Thus, we assume conventionally that negative s_{ik} are admissible, and generalize (3) in the following way:

$$p(s) = \frac{1}{(2\pi)^{0.5}(s + ds)\sigma_{\log}} \exp\left[-\frac{(\ln(s + ds) - \mu)^2}{2\sigma_{\log}^2}\right], \quad (8)$$

where we introduce a new shift parameter ds, which is in our case always positive. The value of ds is derived from the condition $P(ds) = p_z$, and the parameters $\sigma_{\log}$ and μ are, as earlier, found from CV_{emp} (known from the observations) and from the unity mean condition. In the simulation, we replace the negative model values of s by zeros (i.e., apply winsorization). The fitted values are $\sigma_{\log} = 0.51$ and $ds = 0.53$. The results are presented in Fig. 4. We see that the upper tail of the distribution well agrees with the observations, while the agreement for the position of the hump near $s \approx 0.45$ is fairly acceptable. A similar modification with a shifted distribution has also been tested for the gamma distribution, and the result turned out to be much worse: the upper tail was too light, and the hump was too flat.

We have also tested the descriptions of the empirical distributions by the simple distribution laws and their mixtures. The exponential distribution is a very simple one. Its upper tail rather well agrees with the observed distribution; however, this description is not quite accurate in the vicinity of zero. This obstacle can be easily overcome by using binary mixtures. For example, the upper tail of the mixture of a lognormal distribution with $\sigma_{\log} = 0.745$ and the atom at zero with a weight of 0.11 is sufficiently close to the empirical distribution. At the same time, however, all mixtures like "gamma distribution plus atom at zero", "lognormal distribution plus atom at zero," etc., poorly reproduce the details of the histogram. These details can be roughly described in the context of a three-component mixture like atom at zero plus two variants of lognormal distribution; however, rather exact description of the histogram requires at least four components to be mixed. The gamma mixtures can be used as well. However, formula (8) (plus winsorization) provides a

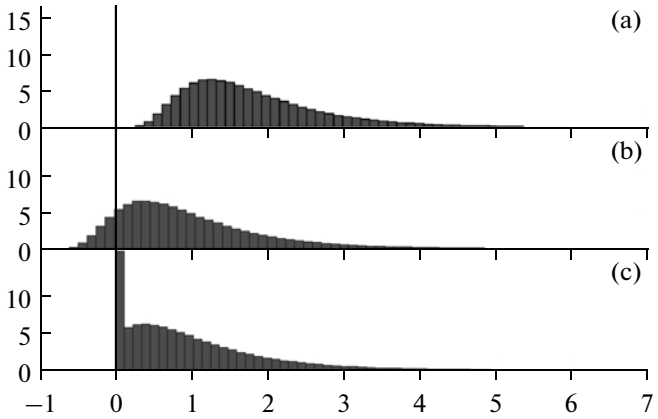

Fig. 4. The illustration for the procedure of modification of lognormal distribution: (a) is the histogram of the values having lognormal distribution with $\sigma_{log} = 0.51$, see formula (3); (b) the distribution is shifted to the left, see formula (8); (c) the negative values of (b) are replaced by zeros (winsorization), the result reproduces the distribution of a slip in the subsources.

much simpler solution to the problem of compact description of s_{ik} distribution. Unfortunately, this formula is conceptually vague: the physical idea behind the notion of the displacement parameter is unclear. Note, however, that, taking into account the normalization to the unity mean, the value of the negative displacement for the studied data is 87% of the mean slip value. Such a considerably large negative slips are barely plausible; rather, they are artifacts. Apparently, replacement of the negative values by zeros (winsorization) reflects the real situation, and a major portion of the discussed 11% of the fault area remains stuck upon the earthquake.

DISCUSSION

Earlier, the authors of [Somerville et al., 1999] studied, among other issues, the empirical distribution functions for the normalized values of a slip, yielded by the inversion. They do not provide any description of the methods used for selection of suitable inversions in that paper, and they present the results in a very brief form. These authors obtained the following estimates: $Q_{1.5} = 25\%$ (correspondingly, $s_{25\%} = 1.5$) and $s_{aver25\%} = 2.0$. Our average estimates for $Q_{1.5}$ and $s_{25\%}$ are very close the these values, whereas our estimate of $s_{aver25\%}$ is somewhat higher (2.4 against 2.0). Fitting a theoretical distribution to the empirical data and study of CV values were beyond the scope of the cited paper.

In [Lavallee and Archuleta, 2003] it was assumed that the distribution function is of most interest for the results of its reduction to a delta-correlated form (whitening) rather than for the values of S_{pq}. To reduce the function to the desired form, the authors constructed a suitable fractional-differentiation whitening filter with the power-law transfer function (in the form k^{δ}, where k is the magnitude of a 2D wave vector) for each S_{pq}; the parameter δ was usually close to unity. After such filtering, it was found that the distribution of the filtered DSM values resembles the Cauchy distribution. If the values of a slip S_{pq} really have the Cauchy or some other stable distribution, the filtering, being a linear operation, will not distort the value of the exponent α in the power-law dependence of the supposed distribution $Q(s)$ ($Q(s) \sim s^{-\alpha}$). In case of the Cauchy distribution, $Q(s) \sim s^{-1}$. From the theoretical viewpoint, the property of "having the stable distribution" is extremely attractive, as it admits using the conceptually transparent models of Levy-stable processes (a random field with stable increments) for the description of the slip function. However, the fact that the tail of the empirical distribution decreases as $Q(s) \sim s^{-3.5-4}$, as is has been established in the present work, completely contradicts the idea of heavy tail with α close to unity. It is premature to infer from this that the concept itself of the stable law is inapplicable for the slip function. One should take into account that at α over 1.6, the analysis of the empirical distribution in the interval of quantiles from 0.1 to 1% gives no ground to regard the tail as certainly obeying the power-law distribution. Besides, it is impossible to estimate the value of α from the slope of the $\log Q(\log s)$ graph (decimal logarithm is meant) in this area even for the case of a knowingly stable distribution [Weron, 2001]. To reveal the stability from the empirical data, one should invoke rarer values, which requires the sample volumes that are by a few orders of magnitude larger than available. It is the easiest to draw misleading conclusions when $\alpha = 1.8–1.9$: in this case, in the vicinity of 1%-quantile the behavior of $Q(s)$ is similar to that for the light-tail (either normal or other) distribution. Thus, up to date, the question remains yet to be answered: whether the idea itself of a stable distribution is applicable to the values of a slip. For the purposes of practical simulation, when it is normally sufficient to confine the description to the interval up to the 0.1%-quantile, the gamma distribution or lognormal distribution can be regarded as the acceptable approximations for the empirical distributions.

From the viewpoint of reliability of the conclusions, the worrisome fact is that different authors obtain noticeably different results for the same source even in the case of the best quality and the best-volume data. These works are usually based on the different sets of input data. This fact partially allows for the discrepancies, but confirms a low stability of the results. However, there is still some hope that the general pattern of the distribution function of a slip (that is, in simpler words, how prominent are the slip asperities) is estimated with a bit more confidence than the particular details of the map.

The obtained results contain some uncertainty, which is due to the fact that the real sources can and, probably,

do have nonconvex boundaries while our analysis recognizes all zeros in the convex hull as the points of the source. However, from the viewpoint of practical applications, this fact can rather be treated as an advantage, because it admits representation of the boundaries of the model source by simple convex envelopes.

After the issues regarding the "asperities" on the area of the source were lively debatable in the literature, the s_{ik} distribution might have been expected to have a heavy tail (that is, $Q(s)$ at large s to decay slower than exponentially). From the graph of $Q(s)$ and from the estimated values of the parameters of theoretical distributions it is clear that they do not contain any distinct heavy tail, and the decay rate is close to that for the lognormal distribution or is a little slower.

CONCLUSIONS

(1) The character of the input data (the discrete slip models yielded by the inversion of the seismological and other types of data with different level of reliability) does not allow their formal, complete systematic analysis. The rejection of the less reliable inversions is inevitably associated with a deal of subjectivity. In the discrete slip models picked out for the analysis, a noticeable fraction of the values is zeros, and some part of them is likely the artifacts produced by the inversion procedure. Therefore, the obtained results are tentative and not rigorous.

(2) The subject of study is the empirical distribution of the values of a normalized slip in the earthquake source faults. The statistical structure of the values of the normalized slip in the individual subsources is relatively stable. The mean coefficient of variation is 0.98 ± 0.28, and the upper 1%-quantile of the normalized slip is 4.4 ± 1.4. The power-law approximation of the upper tail in the area of upper 1–10% quantiles provides the estimation of the exponent at 3.8 ± 1.3 and prohibits the description by a stable distribution with α below 1.4–1.5. These results characterize a definite, although still obviously limited variability of the variants of probability distributions for a slip.

(3) Generally, the distribution of the values of the normalized slip resembles the exponential distribution. A more exact description requires that the outlier in the distribution density (atom) at zero with a weight of 10% is taken into account. We were not able to find the simplest description in the form of a mixture of a standard distribution and the atom at zero. The approximation by more complex mixtures is of little interest, although possible. We proposed a compact description based on the lognormal law with a shift of the entire distribution to the left and winsorization at zero. The appearance of the atom at zero is probably related to the presence of unbreakable barriers or segments of reverse slip. Over a considerable part of the area of the earthquake source fault (about 10%), the slip can really be zero or even negative. At the same time, the possibility for the zero values to be as well the artifacts generated by the inversion procedure cannot be ruled out completely.

(4) The description of the statistic of a slip in terms of both the parameters of distributions and the proposed approximations of the distribution function can be used in the practical simulations of the earthquake sources. At the same time, this description imposes the constraints on the admissible theoretical models of the sources.

ACKNOWLEDGMENTS

The work was supported by the Russian Foundation for Basic Research, grant no. 07-05-00775.

REFERENCES

Andrews, D.J., A Stochastic Fault Model: 1. Static Case, *J. Geophys. Res.*, 1980, no. 78, pp. 3867–3877.

Custodio, S., Liu, P.C., and Archuleta, R.J., The 2004 $M_w = 6.0$ Parkfield, California, Earthquake: Inversion of Near-Source Ground Motion Using Multiple Data Sets, *Geophys. Rev. Lett.*, 2005, vol. 32, L23312.

Das, S. and Aki, K., Fault Plane with Barriers: A Versatile Earthquake Model, *J. Geophys. Res.*, 1977, no. 82, pp. 5648–5670.

Gusev, A.A., Guseva, E.M., and Pavlov, V.M., Modeling of the Ground Motion for the Petropavlovsk Earthquake of November 24, 1971 (M= 7.6), *Fiz. Zemli*, 2009, no. 5, pp. 29–38 [*Izv. Phys. Earth* (Engl. Transl.), 2009, vol. 45, no. 5, pp. 395–405].

Gusev, A.A. and Pavlov, V.M., Broadband Simulation of Earthquake Ground Motion by a Spectrum-Matching, Multiple-Pulse Technique, *Earthquake Spectra*, 2009, vol. 25, no. 2, pp. 257-276.

Heaton, T.H., Evidence for and Implications of Self-Healing Pulses of Slip in Earthquake Rupture, *Phys. Earth Planet. Inter.*, 1990, no. 64, pp. 1–20.

Irikura, K., Predicting Strong Ground Motions with a "Recipe", *Bull. Earthquake Res. Inst. Univ. Tokyo*, 2006, no. 81, pp. 341–352.

Lavallee, D. and Archuleta, R., Stochastic Modeling of Slip Spatial Complexities for the 1979 Imperial Valley, California, Earthquake, *Geophys. Rev. Lett.*, 2003, no. 30, p. 1245.

Mai, P.M., Database of Finite-Source Rupture Models, http://www.seismo.ethz.ch/srcmod, 2004.

Mendoza, C. and Hartzell, S.H., Slip Distribution of the 19 September 1985 Michoacan, Mexico, Earthquake: near-Source and Teleseismic Constraints, *Bull. Seis. Soc. Amer.*, 1989, vol. 79, no. 3, pp. 655–669.

Somerville, P., Irikura, K., Graves, R., et al., Characterizing Crustal Earthquake Slip Models for the Prediction of Strong Motion, *Seism. Res. Lett.*, 1999, no. 70, pp. 59–80.

Tsai, C.-C.P., Slip, Stress Drop and Ground Motion of Earthquakes: a View from the Perspective of Fractional Brownian Motion, *Pure Appl. Geophys*, 1997, no. 149, pp. 689–706.

Weron, R., Levy-Stable Distributions Revisited: Tail Index >2 Does Not Exclude the Levy-Stable Regime, *Int. J. Modern Phys.*, 2001, no. 12, pp. 209–223.