AGU PUBLICATIONS

Geophysical Research Letters

RESEARCH LETTER
10.1002/2017GL075875

Key Points:
- What are good stress based predictors of aftershock locations?
- Coulomb failure stress is a poor predictor, but the third invariant of the stress tensor is quite good
- Large-scale analysis of mainshocks and aftershocks suggests very different patterns than commonly assumed

Supporting Information:
- Supporting Information S1

Correspondence to:
B. J. Meade,
meade@fas.harvard.edu

Citation:
Meade, B. J., DeVries, P. M. R., Faller, J., Viegas, F., & Wattenberg, M. (2017). What is better than Coulomb failure stress? A ranking of scalar static stress triggering mechanisms from 10^5 mainshock-aftershock pairs. *Geophysical Research Letters*, *44*. https://doi.org/10.1002/2017GL075875

Received 28 SEP 2017
Accepted 12 NOV 2017
Accepted article online 16 NOV 2017

What Is Better Than Coulomb Failure Stress? A Ranking of Scalar Static Stress Triggering Mechanisms from 10^5 Mainshock-Aftershock Pairs

Brendan J. Meade[1,2], Phoebe M. R. DeVries[1], Jeremy Faller[2], Fernanda Viegas[2], and Martin Wattenberg[2]

[1]Department of Earth and Planetary Sciences, Harvard University, Cambridge, MA, USA, [2]Google, Inc., Cambridge, MA, USA

Abstract Aftershocks may be triggered by the stresses generated by preceding mainshocks. The temporal frequency and maximum size of aftershocks are well described by the empirical Omori and Bath laws, but spatial patterns are more difficult to forecast. Coulomb failure stress is perhaps the most common criterion invoked to explain spatial distributions of aftershocks. Here we consider the spatial relationship between patterns of aftershocks and a comprehensive list of 38 static elastic scalar metrics of stress (including stress tensor invariants, maximum shear stress, and Coulomb failure stress) from 213 coseismic slip distributions worldwide. The rates of true-positive and false-positive classification of regions with and without aftershocks are assessed with receiver operating characteristic analysis. We infer that the stress metrics that are most consistent with observed aftershock locations are maximum shear stress and the magnitude of the second and third invariants of the stress tensor. These metrics are significantly better than random assignment at a significance level of 0.005 in over 80% of the slip distributions. In contrast, the widely used Coulomb failure stress criterion is distinguishable from random assignment in only 51–64% of the slip distributions. These results suggest that a number of alternative scalar metrics are better predictors of aftershock locations than classic Coulomb failure stress change.

1. Introduction

Several studies have demonstrated that stress changes due to large earthquakes may influence the location of subsequent events (e.g., Deng & Sykes, 1996, 1997a, 1997b; King et al., 1994; Nostro et al., 1997; Reasenberg & Simpson, 1992). Static stress changes have been invoked to explain earthquake interactions over thousands of kilometers and decadal timescales (e.g., Stein et al., 1997), but aftershocks provide the most ubiquitous demonstrations of near-field (<100 km) and near-term (<1 year) earthquake triggering and interactions (Freed, 2005).

The Coulomb failure stress criterion (Table 1) (e.g., King et al., 1994) is one of the most common quasi-static metrics used to explain spatial patterns of aftershocks. Regions of increased Coulomb failure stress have been correlated with locations of aftershocks after large earthquakes in Japan (e.g., Toda et al., 1998), California (e.g., Parsons et al., 1999; Reasenberg & Simpson, 1992), and many other locations worldwide (e.g., Jacques et al., 1996; Nostro et al., 1997). Despite these examples, the Coulomb failure stress criterion is not successful in every case; after the 1994 Northridge earthquake, Coulomb failure stress changes could not explain the locations of aftershocks significantly better than a set of randomly distributed events (Hardebeck et al., 1998). Furthermore, after the 1992 Landers and 1995 Kobe earthquakes in California and Japan, seismicity rates increased everywhere (Mallman & Zoback, 2007), suggesting that decreases in Coulomb failure stress did not inhibit failure within stress shadows in the vicinity of these mainshocks.

A number of studies have proposed small variations on the classic Coulomb failure stress criterion, such as changing the effective coefficient of friction (e.g., Kagan & Jackson, 1998; King et al., 1994; Reasenberg & Simpson, 1992). However, it is difficult to rigorously evaluate the effectiveness of the Coulomb failure stress criterion and alternatives to it, because the vast literature on aftershock triggering is dominated by studies that focus on a small number of mainshocks (<3) and their aftershocks (e.g., Hardebeck et al., 1998; King et al., 1994; Parsons et al., 1999; Reasenberg & Simpson, 1992). Large earthquakes and their aftershocks can be viewed as repeated natural experiments: each mainshock induces stress changes, and subsequent aftershocks record the response of the crust. From this perspective, one mainshock and its aftershocks are analogous to only one trial run in a repeated experiment and the results may not be representative. In

Table 1
A List of Scalar Fields Considered in This Study, Where χ Represents Either the Full Stress Tensor σ or the Deviatoric Stress Tensor σ', χ_i are the Principal Stresses, $\mathbf{n}_\parallel$ Is a Unit Vector Aligned With the Average Mainshock Coseismic Slip Vector, $\mathbf{n}_\perp$ Is a Unit Vector Normal to the Mean Orientation of the Mainshock Coseismic Rupture Plane, μ Is the Coefficient of Friction, x and y Are the Locations of the Grid Cell, and x_f and y_f Are the Locations of the Mainshock Rupture Plane

Quantity	Symbol	Evaluation
Nearest distance[a]	r	$r = \min\left(\sqrt{(x-x_f)^2 + (y-y_f)^2}\right)$
Maximum shear	$\Delta\tau_{max}(\chi)$	$\Delta\tau_{max}(\chi) = \lvert\chi_1 - \chi_3\rvert/2$
First invariant	$\Delta I_1(\chi)$	$\Delta I_1(\chi) = \chi_1 + \chi_2 + \chi_3$
Second invariant	$\Delta I_2(\chi)$	$\Delta I_2(\chi) = \chi_1\chi_2 + \chi_2\chi_3 + \chi_1\chi_3$
Third invariant	$\Delta I_3(\chi)$	$\Delta I_3(\chi) = \chi_1\chi_2\chi_3$
Coulomb failure criteria μ = 0.0, 0.2, 0.4, 0.6, 0.8	$\Delta CFS(\chi,\mu)$	$\Delta CFS(\chi,\mu) = (\mathbf{n}_\perp \cdot \chi)\cdot\mathbf{n}_\parallel - \mu(\mathbf{n}_\perp \cdot \chi)\cdot\mathbf{n}_\perp$
Coulomb failure criteria normal only, μ = 0.4	$\Delta CFS_n(\chi)$	$\Delta CFS_n(\chi) = -\mu(\mathbf{n}_\perp \cdot \chi)\cdot\mathbf{n}_\perp$
Coulomb failure criteria total shear	$\Delta CFS_\tau(\chi)$	$\Delta CFS_\tau(\chi) = (\mathbf{n}_\perp \cdot \chi)\cdot\mathbf{n}_\parallel + (\mathbf{n}_\perp \cdot \chi)\cdot(\mathbf{n}_\parallel \times \mathbf{n}_\perp)$
Coulomb failure criteria total, μ = 0.4	$\Delta CFS_{total}(\chi,\mu)$	$\Delta CFS_{total}(\chi,\mu) = (\mathbf{n}_\perp \cdot \chi)\cdot\mathbf{n}_\parallel + (\mathbf{n}_\perp \cdot \chi)\cdot(\mathbf{n}_\parallel \times \mathbf{n}_\perp) - \mu(\mathbf{n}_\perp \cdot \chi)\cdot\mathbf{n}_\perp$

Note. We consider the quantities below as well as their magnitudes. In total, 38 scalar stress fields are included in the analysis, although they are of course not all linearly independent.
[a]Note that in the ROC analysis (e.g., Figure 2), the quantity (150 km-nearest distance) is used to transform distance from a negative predictor into a positive predictor.

order to gain a more comprehensive understanding of aftershock triggering and advance aftershock prediction around the world, multiple mainshocks should be considered.

In this paper, we examine a suite of 213 mainshock slip distributions (123 distinct mainshocks) and their aftershocks. For each of these slip distributions, we compare the spatial patterns of 38 alternatives to and variants of classic Coulomb failure stress change (e.g., maximum shear stress change, invariants of the stress change tensor, and magnitude of Coulomb failure stress change; Table 1) to aftershock locations using receiver operating characteristic (ROC) analysis. ROC analysis is commonly used to evaluate the accuracy of diagnostic medical tests; here we use these methods to assess rates of true-positive and false-positive classification of regions with and without aftershocks. This approach allows us to fully evaluate the predictive ability of each stress metric—in other words, the ability of each metric to accurately discriminate between, or classify, regions with and without aftershocks. We focus on static stress changes although other sources of stress such as dynamic stress changes (e.g., Kilb et al., 2000; Felzer & Brodsky, 2005, 2006), secondary stress changes (e.g., Meier et al., 2014), and postseismic stress changes (e.g., Pollitz & Sacks, 2002) may also be important. ROC analysis reveals 23 static stress metrics that perform significantly (α = 0.005) better than random assignment in over 76% of the slip distributions, including maximum shear stress change and the magnitudes of the second and third invariants of the stress change tensor. The remaining stress metrics, including the classic Coulomb failure stress criterion, are distinguishable from random guessing at a significance level of 0.005 for only 10–64% of the slip distributions.

2. Data and Models

The SRCMOD (http://equake-rc.info/SRCMOD/) online database of finite-fault rupture models contains 334 models from 169 earthquakes since 1906 as of 1 August 2017. These mainshock slip distributions were obtained from inversion and modeling of seismic, geodetic, and other types of geophysical data; a full list of references can be found at http://equake-rc.info/SRCMOD/searchmodels/allevents/. For our analysis, we used 220 of these slip distributions; the remaining 114 rupture models could not be read in due to formatting irregularities.

Stress changes due to all 220 mainshock ruptures (M_w = 4.4 to M_w = 9.2) were calculated in a volume extending 100 km from the mainshock rupture plane horizontally and 50 km vertically (e.g., Figures 1a–1c and 1e–1g), gridded into 5 × 5 × 5 km cells. For each mainshock rupture, the stress tensors in each grid cell centroid were transformed into 38 different scalar stress metrics (Table 1). These metrics were selected based on previous studies (e.g., Coulomb failure stress change and variations in the coefficients of friction μ' (King et al., 1994; Reasenberg & Simpson, 1992)), as well as a number of alternative scalarizations of the stress tensor (e.g., first, second, and third invariants) which to our knowledge have not been considered in earthquake triggering studies to date. All the stress tensor scalarizations (Table 1) are evaluated using both the deviatoric and full stress tensors. Certain scalar stress metrics (e.g., classic Coulomb failure stress) depend

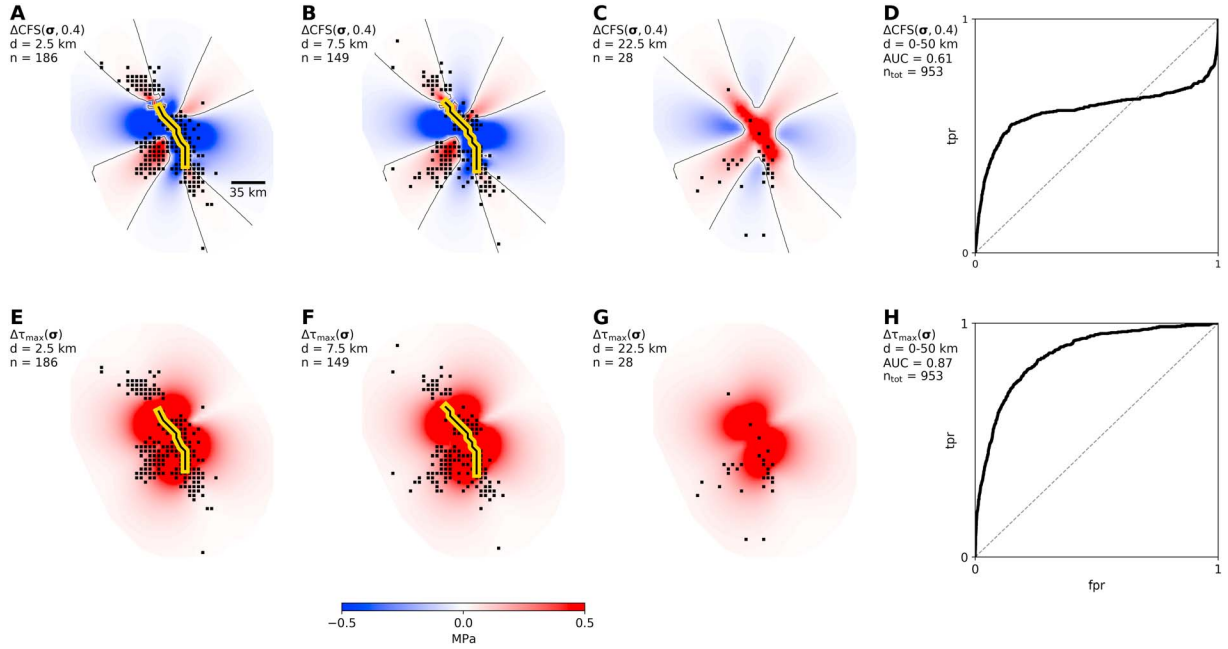

Figure 1. Examples of two static stress fields considered in this analysis and corresponding ROC curves for one slip distribution (Cohee & Beroza, 1994) from the 1992 M_w = 7.3 Landers earthquake in California. (a–c) Map view of $\Delta CFS(\boldsymbol{\sigma}, 0.4)$ values within 100 km of the fault at 2.5 km, 7.5 km, and 22.5 km depth, respectively. Black squares represent grid cells in which one or more aftershocks occurred within 1 year of the mainshock. Thick yellow and black line represents extent of the mainshock rupture at each depth. Scale bar is shown in Figure 1a as a thick black line. (d) ROC curve for this particular slip distribution (Cohee & Beroza, 1994) and $\Delta CFS(\boldsymbol{\sigma}, 0.4)$, including all grid cells and all aftershocks within a year after the mainshock. Black dotted line is a 1:1 line for reference. (e–h) Analogous to Figures 1a–1d for a different static stress field, $\Delta\tau_{max}(\boldsymbol{\sigma})$.

on both the orientation of the mainshock source and the receiver plane; for these metrics, we assume aftershock receiver plane orientations equal to the source (mainshock) average fault orientation. Due to possible strike and dip ambiguities, we calculate four versions of classic Coulomb failure stress change and use the best performing sign convention for $\Delta CFS(\boldsymbol{\sigma}, 0.4)$ (Table 1) for each slip distribution.

Aftershocks within 1 year of the mainshocks (116,814 aftershocks in total) were compiled from the International Seismological Center (ISC) event catalogue. Seven of the 220 slip distributions had no aftershocks within the volume ($<$100 km from mainshock horizontally and $<$50 km vertically) and time window ($<$1 year from mainshock) considered and were excluded from our analysis. The remaining 213 slip distributions (Table A1; 123 distinct mainshocks) form the basis of this study.

3. Methods

To assess the ability of each stress metric to explain aftershock locations, we use receiver operating characteristic (ROC) analysis. To build an ROC curve, the true-positive rate is plotted against the false-positive rate for the full range of possible cutoff thresholds. In ROC space, a diagnostic test (or binary classification method) that is no better than random guessing would plot close to a 1:1 line (equal rates of false-positive and true-positive classifications). A test that is more effective than random guessing would plot above a 1:1 line, as effective tests will have higher true-positive rates than false-positive rates.

ROC curves have many advantages over a statistic like the misclassification rate. First and foremost, ROC curves visualize the efficacy of a given test for all possible thresholds, while a misclassification rate is the error rate for only one threshold. The area under an ROC curve (AUC) quantifies the overall performance of a test across all thresholds; a perfect test would correspond to an AUC = 1, while a test that was no better than random guessing would correspond to an AUC = 0.5. In probabilistic terms, the AUC is the probability that given randomly chosen positive and negative samples, the test will rank the negative sample lower than the positive sample.

In our case, we have a binary classification problem: our goal is to assess the accuracy with which each stress metric (Table 1) can be used to classify the 5 × 5 × 5 km grid cells in the volume around the mainshock as

either "containing aftershocks" or "not containing aftershocks." Statistics like accuracy rate or misclassification rate could be extremely misleading in our application because we have ~74,000 grid cells containing aftershocks, and over 2 orders of magnitude more grid cells (~6.6 million) without aftershocks. ROC analysis is appropriate for this application because it is insensitive to class imbalance; false-positive and true-positive classification rates do not depend on the absolute number of "positive" (with aftershocks) or "negative" (without aftershocks) grid cells.

4. Results

For each of the 213 slip distributions (Table A1) and for each stress metric considered (Table 1), we build an ROC curve and calculate the corresponding AUC value (38 metrics × 213 slip distributions = 8094 ROC curves and AUC values in total). With AUC analysis, the stress metrics can be ranked in order of their ability to correctly classify whether or not an aftershock occurred in a grid cell. Consider the ROC curves for one of the 1994 $M_w = 7.4$ Landers earthquake slip distributions (Figure 1) (Cohee & Beroza, 1994): for maximum shear stress (Figure 1h), the ROC curve is above the 1:1 line with AUC = 0.87, suggesting that this stress metric can be used to discriminate grid cells with and without aftershocks in the volume around the Landers' mainshock more accurately than a random classifier. By contrast, the corresponding ROC curve for classic Coulomb failure stress (AUC = 0.61; Figure 1d) suggests that using this scalar mechanical metric to classify grid cells is little better than random assignment.

The results across all slip distributions are summarized below. To be as concise as possible, the accuracy with which each stress metric can be used to classify grid cells in the volume around the mainshock is referred to as the discriminating ability or performance of the metric.

A rank order of the performance of the stress metrics. We rank the overall discriminating ability of the stress metrics from best performing to worst by combining every slip distribution-stress metric pair across all slip distributions and calculating the merged AUC values (A_m; Figure 2). Maximum shear stress ($\Delta\tau_{\max}(\boldsymbol{\sigma})$) and the absolute values of the second and third invariants of the deviatoric and full stress tensors ($|\Delta I_2(\boldsymbol{\sigma})|$, $|\Delta I_2(\boldsymbol{\sigma}')|$, $|\Delta I_3(\boldsymbol{\sigma})|$, and $|\Delta I_3(\boldsymbol{\sigma}')|$; Table 1) ranked among the best ($A_m = 0.754$–0.758). Deviatoric Coulomb failure stress change $\Delta\mathrm{CFS}(\sigma', \mu)$, the second and third invariants of the full stress tensor, $\Delta I_2(\sigma)$ and $\Delta I_3(\sigma)$, ranked among the worst ($A_m = 0.411$–0.580).

As stated above, these rankings are based the AUC value of the merged average (A_m) of the ROC curves, but there are other ways of averaging ROC curves (threshold averaging, A_t, or vertical averaging, A_v, in Figure 2) (Fawcett, 2006). Although the rankings change slightly depending on the averaging method used, the stress metrics can always be divided into two consistent and distinct categories. The 23 Category I stress metrics (Figure 2; red labels) are characterized by A_m values greater than 0.72. The remaining 15 stress metrics are defined as Category II (Figure 2; black labels) and have A_m values between 0.41 and 0.62. Notably, classic Coulomb failure stress falls into Category II regardless of the assumed effective coefficient of friction; the highest $A_m$value for classic Coulomb failure stress is 0.62. When grid cells with nearest distances less than 5 km from the mainshock are discarded from the analysis, these categories remain distinct and the six best performing stress metrics remain top ranked.

Statistical significance. We perform permutation tests to assess the statistical significance of the performance of the static stress fields. For each mainshock slip distribution, we generated 500 random realizations of the aftershock locations. Each of the resulting 'random' ROC curve realizations yields an AUC value. The observed AUC value is compared with the empirical distribution of 'random' AUC values from the permutation tests to obtain a one-sided empirical p-value for every stress metric-slip distribution pair. To effectively summarize the results of these permutation tests, we define Ψ as the fraction of p-values for each stress metric that are less than $\alpha = 0.005$ (Benjamin et al., 2017) across all slip distributions. For example, for maximum shear stress ($\Delta\tau_{\max}(\boldsymbol{\sigma})$; Figure 2), $\Psi = 0.808$ represents the fraction of slip distributions for which maximum shear stress can be used to classify grid cells significantly better than random assignment at a significance level of 0.005.

The empirical p values and Ψ values further emphasize the distinction in discriminatory ability between Category I and Category II stress metrics. The stress metrics in Category I perform significantly better than random guessing at a significance level of 0.005 for between 76 and 82% of the slip distributions. In stark

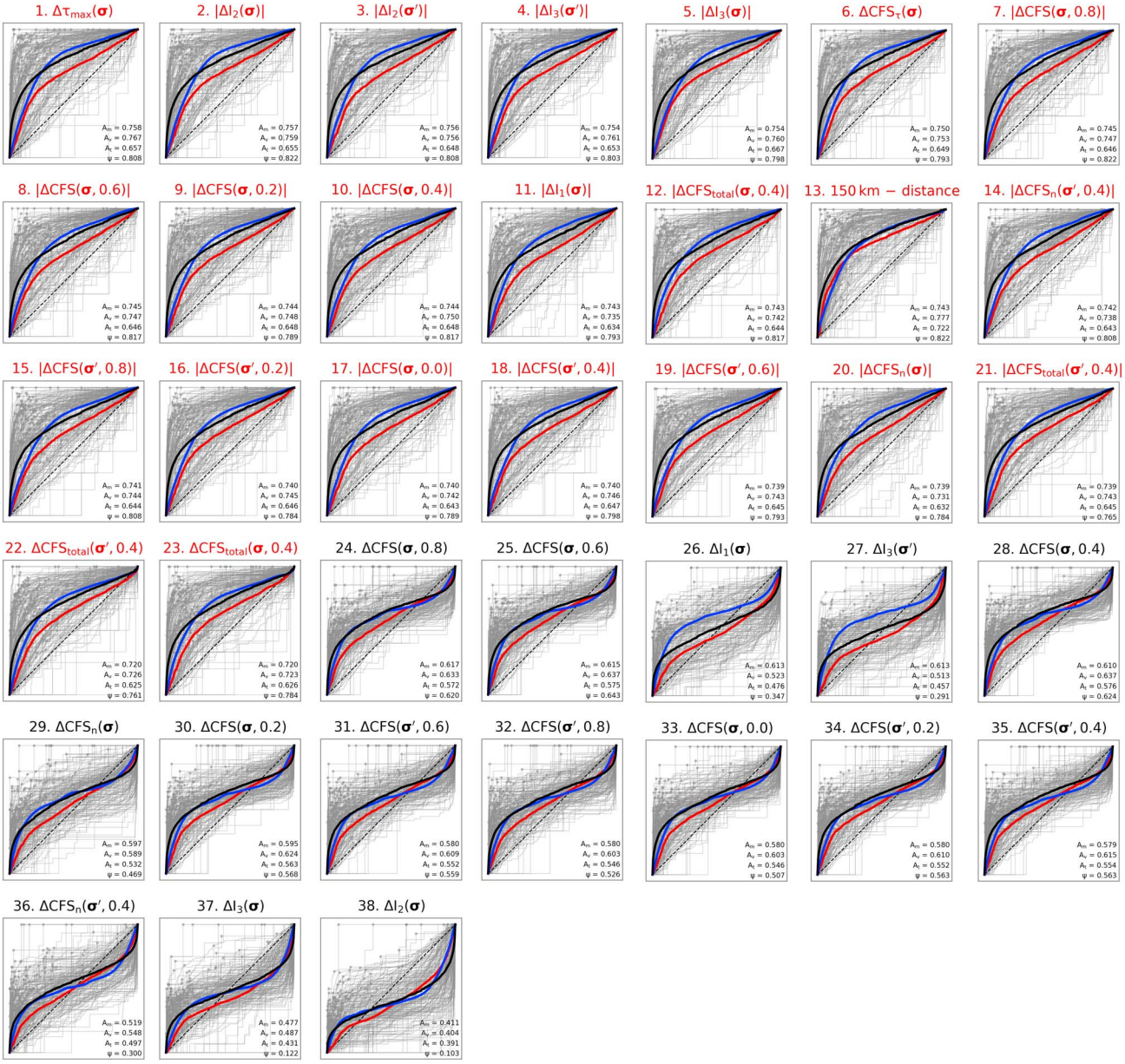

Figure 2. ROC curves for all 38 stress metrics, in ranked order by merged AUC value (A_m). Category I metrics (1–23) have red titles; Category II metrics (24–38) have black titles. See Table 1 for detailed descriptions of the symbols in the titles. Vertically averaged AUC values A_v, threshold averaged AUC values A_t, and the fraction of statistically significant ($\alpha = 0.005$) empirical p values Ψ are included for each metric. ROC curves for every slip distribution (213 in total) are shown with thin gray lines for each stress metric. Gray circles represent the locations on the ROC curves where Youden's index is maximized. Merged, vertically averaged, and threshold averaged ROC curves across all slip distributions are shown in thick blue, black, and red lines, respectively.

contrast, Category II stress metrics can discriminate between grid cells significantly better than random guessing in only 10–64% of the slip distributions considered.

Sensitivity to small variations in source geometry. The 1992 Landers, 1995 Kobe, 1999 Chichi, 1999 Izmit, 2007 Tocopila, 2008 Wenchuan, 2009 L'Aquila, and 2011 Tohoku earthquakes each have more than four published and available slip distributions in the SRCMOD catalogue (Table A1). To assess the sensitivity of the performance of the stress metrics to small variations in source geometry, we examine the range of AUC values from different slip distributions with the same mainshock. AUC values for Category I stress metrics vary, on average, by 0.13 or less across all available slip distributions for these eight earthquakes (Table A2). By contrast, the AUC values for Category II stress metrics vary from 0.13 to 0.28 on average across these slip distributions (Table A2). In other words, not only do Category I metrics have significantly ($\alpha = 0.005$) better

discriminatory ability than Category II metrics, but their performance may also be as much as twice as robust to small changes in source geometry.

Variations with mainshock magnitude. The stress metric rankings discussed above are based on 213 slip distributions that include mainshocks varying in size from M_w = 4.4 to M_w = 9.2. To examine the sensitivity of the performance of these stress metrics to mainshock magnitude, we use two-sample Kolmogovov-Smirnov tests to assess the likelihood that the distributions of AUC values for $6 < M_w < 7$, $7 < M_w < 8$, and $M_w > 8$ earthquakes are distinct (Table A3).

Interestingly, for every stress metric (Table 1) except the second invariant of the full stress tensor, the distributions of AUC values for $6 < M_w < 7$ and $7 < M_w < 8$ are not statistically distinct at a significance level of 0.005. However, for all Category I stress fields, distributions of AUC values for $M_w > 8$ earthquakes are significantly different ($\alpha = 0.005$) from distributions of AUC values for $6 < M_w < 7$ and $7 < M_w < 8$ earthquakes. For these high-performing metrics, mean AUC values for $M_W > 8$ earthquakes are between 0.63 and 0.68, while mean AUC values for $6 < MM_w < 7$ and $7 < M_w < 8$ earthquakes are all >0.72 (Table A4). That said, Ψ values are all >0.78 for $M_w > 8$ earthquakes, suggesting that Category I stress metrics still consistently perform better than random assignment at a significance level of 0.005 for the largest events.

5. Discussion

Although classic Coulomb failure stress has been widely used in earthquake triggering studies for decades (e.g., King et al., 1994), the quantity can be used to discriminate between grid cells with and without aftershocks more accurately than random assignment in only 51–64% of the 213 slip distributions considered here. A number of alternatives—the Category I metrics including maximum shear stress, distance to the mainshock rupture, and the magnitudes of the second and third invariants of the stress tensor (Figure 2)—consistently ($\Psi > 0.76$) perform significantly ($\alpha = 0.005$) better than a random classifier. Overall, these results strongly suggest that Category I metrics (Table 1) should be considered in future earthquake triggering and stress transfer studies.

In Coulomb failure stress theory, positive stress changes are interpreted to promote failure, while negative stress changes inhibit failure (e.g., King et al., 1994). Using the ROC analysis described above, we can examine the validity of this hypothesis empirically because ROC curves illustrate the performance of each stress metric across all possible cutoff thresholds. To determine the optimal thresholds—the thresholds that give each stress metric the most discriminating ability – we maximize Youden's index for each ROC curve. This index, defined as the sum of specificity (true-negative rate) and sensitivity (true-positive rate), is widely used to select optimal thresholds for diagnostic tests (e.g., Cho & Wang, 1997; Schisterman et al., 2005). The point on an ROC curve that maximizes Youden's index will correspond to the threshold that gives the stress metric the most discriminating ability when all misclassifications—both false-positive classifications (i.e., false alarms) and false negative classifications (i.e., missed aftershock predictions)—are weighted equally.

The optimal thresholds for the stress metrics vary significantly between slip distributions (Figure A2), but three notable observations about these thresholds are included here: (1) Overall, across all the slip distributions, the average optimal distance threshold is 37.3 km and the median distance is 32.4 km. As might be expected, this optimal distance threshold increases with earthquake magnitude: for $6 \leq M_w < 7$ earthquakes, the mean optimal threshold is 32.1 km, for $7 \leq M_w < 8$ earthquakes it is 36.6 km, and for $M_w \geq 8$ earthquakes it is 49.1 km. (2) The vast majority of stress metrics (37 out of 38) have distributions of optimal thresholds that are skewed to the right, consistent with the hypothesis that for the successful stress metrics, more aftershocks occur in regions of large positive stress values than in regions of large negative stress values. (3) Finally, and perhaps mostly interestingly, the stress thresholds that allow classic Coulomb failure stress change to maximize Youden's index are generally positive and close to zero (median thresholds between 0.01 and 0.04 MPa), regardless of the assumed effective coefficient of friction. In other words, although classic Coulomb failure stress change has a low ($A_m < 0.62$) discriminating ability, using this quantity to predict aftershocks under the hypothesis that large positive values (>0.1 MPa) may promote failure and negative values (<0.0 MPa) inhibit failure (e.g., King et al., 1994) seems to allow it to perform best.

In addition to classic Coulomb failure stress, the nearest distance to the mainshock, *r*, stands out among the metrics (Table 1), because (1) it is not a scalarization of the stress tensor, (2) it has extraordinarily good discriminating ability ($A_m = 0.74$, $\Psi = 0.82$) even compared to other Category I metrics, and (3) it is not a locally

defined mechanical quantity. The 21 top-ranked Category I stress metrics are everywhere positive (Figure A1), and thus, it is possible that distance is the most important control on aftershock distributions, and all the other Category I stress metrics simply correlate well with distance.

To examine this hypothesis, we attempt to account for distance from the mainshock in evaluating the performance of the stress metrics. The stress metrics are corrected for distance as $1/r^2$, consistent with the decay expected for a finite source (Okada, 1992). (In the far-field, stresses would be expected to decrease as $1/r^3$ from a point source (Okada, 1992), but because the vast majority aftershocks take place near ($<$~50 km from) the mainshock (e.g., Figure 1), this correction is likely not be appropriate here.) Once corrected for distance as $1/r^2$, the only metrics that consistently ($\Psi > 0.70$) perform significantly better ($\alpha = 0.005$) than random classifiers consistently are the magnitudes of the second and third invariants of the full and deviatoric stress tensor. However, the second and third invariants are defined as the products of two and three components of the stress tensor respectively (Table 1)—and thus may be expected to decay as ~$1/r^4$ and ~$1/r^6$. When corrected by these factors, their discriminating ability disappears. Thus, it may be that distance from the mainshock is the best predictor of aftershock locations and Category 1 metrics may simply serve as proxies for distance.

Of course, there are a number of simplifying assumptions implicit in this analysis. We do not account for time-dependent sources of stress that could influence the spatial distribution of aftershocks over time, such as secondary aftershock triggering (e.g., Meier et al., 2014), dynamic stress changes (e.g. Felzer & Brodsky, 2005, 2006; Kilb et al., 2000; van der Elst & Brodsky, 2010), or postseismic stress changes (e.g., Pollitz & Sacks, 2002). Dynamic stress changes have been used to explain aftershock locations immediately ($<$1 week) after large mainshocks (Prejean et al., 2004), while viscoelastic stress changes and secondary aftershock triggering may be important on longer timescales (days-years) (e.g., Lorenzo-Martín et al., 2006; Pollitz & Sacks, 2002; To et al., 2004). However, the discriminatory ability of Category I static stress change values due to the mainshocks alone (e.g., Figure 2) suggests that while additional sources of stress change might influence aftershock locations, they are likely not the primary controls.

6. Conclusions

Coulomb failure stress changes due to individual mainshocks are often invoked to explain the locations of subsequent aftershocks. Using receiver operating characteristic (ROC) analysis, we assess and compare the ability of 38 alternatives to Coulomb failure stress change to discriminate between regions with and without aftershocks after 123 different mainshocks worldwide (213 slip distributions). Classic Coulomb failure stress change can be used to discriminate between regions significantly better than random assignment ($\alpha = 0.005$) in less than 65% of the slip distributions considered, regardless of the assumed coefficient of friction. By contrast, we identify 23 alternatives to and variations of classic Coulomb failure stress, including maximum shear stress and the magnitudes of the second and third invariants of the stress tensor, that perform significantly better than random assignment ($\alpha = 0.005$) in over 76% of the 213 slip distributions. Although these stress metrics may simply serve as effective proxies for distance, their performance nevertheless suggests that they should be considered in future earthquake triggering studies.

Acknowledgments
All data are freely available from the SRCMOD catalogue (http://equake-rc.info/SRCMOD/) and the International Seismological Center (ISC) event catalogue (http://www.isc.ac.uk/iscgem/). We thank the Editor and two reviewers for thoughtful comments.

References

Benjamin, D. J., Berger, J., Johannesson, M., Nosek, B. A, Wagenmakers, E.-J., Berk, R., … Johnson, V. E. (2017). Redefine statistical significance, PsyArXiv, July 22. Retrieved from psyarxiv.com/mky9j

Cho, D. Y., & Wang, Y. C. (1997). Comparison of the APACHE III, APACHE II and Glasgow Coma Scale in acute head injury for prediction of mortality and functional outcome. *Intensive Care Medicine*, *23*(1), 77–84. https://doi.org/10.1007/s001340050294

Cohee, B. P., & Beroza, G. C. (1994). Slip distribution of the 1992 Landers earthquake and its implications for earthquake source mechanics. *Bulletin of the Seismological Society of America*, *84*(3), 692–712.

Deng, J., & Sykes, L. R. (1996). Triggering of 1812 Santa Barbara earthquake by a great San Andreas shock: Implications for future seismic hazards in southern California. *Geophysical Research Letters*, *23*, 1155–1158. https://doi.org/10.1029/96GL00738

Deng, J., & Sykes, L. R. (1997a). Evolution of the stress field in southern California and triggering of moderate-size earthquakes: A 200-year perspective. *Journal of Geophysical Research*, *102*, 9859–9886. https://doi.org/10.1029/96JB03897

Deng, J., & Sykes, L. R. (1997b). Stress evolution in southern California and triggering of moderate-, small-, and micro-size earthquakes. *Journal of Geophysical Research*, *102*, 24,411–24,435. https://doi.org/10.1029/97JB02127

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, *27*(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

Felzer, K. R., & Brodsky, E. E. (2005). Testing the stress shadow hypothesis. *Journal of Geophysical Research*, *110*, B05S09. https://doi.org/10.1029/2004JB003277

Felzer, K. R., & Brodsky, E. E. (2006). Decay of aftershock density with distance indicates triggering by dynamic stress. *Nature*, *441*(7094), 735–738. https://doi.org/10.1038/nature04799

Freed, A. M. (2005). Earthquake triggering by static, dynamic, and postseismic stress transfer. *Annual Review of Earth and Planetary Sciences*, *33*(1), 335–367. https://doi.org/10.1146/annurev.earth.33.092203.122505

Hardebeck, J. L., Nazareth, J. J., & Hauksson, E. (1998). The static stress change triggering model: Constraints from two southern California aftershock sequences. *Journal of Geophysical Research*, *103*, 24,427–24,437. https://doi.org/10.1029/98JB00573

Jacques, E., King, G. C., Tapponnier, P., Ruegg, J. C., & Manighetti, I. (1996). Seismic activity triggered by stress changes after the 1978 events in the Asal Rift, Djibouti. *Geophysical Research Letters*, *23*, 2481–2484. https://doi.org/10.1029/96GL02261

Kagan, Y. Y., & Jackson, D. D. (1998). Spatial aftershock distribution: Effect of normal stress. *Journal of Geophysical Research*, *103*, 24,453–24,467. https://doi.org/10.1029/98JB00699

Kilb, D., Gomberg, J., & Bodin, P. (2000). Triggering of earthquake aftershocks by dynamic stress. *Nature*, *408*(6812), 570–574. https://doi.org/10.1038/35046046

King, G. C., Stein, R., & Lin, J. (1994). Static stress changes and the triggering of earthquakes. *Bulletin of the Seismological Society of America*, *84*(3), 935–953.

Lorenzo-Martín, F., Roth, F., & Wang, R. (2006). Elastic and inelastic triggering of earthquakes in the North Anatolian Fault zone. *Tectonophysics*, *424*(3-4), 271–289. https://doi.org/10.1016/j.tecto.2006.03.046

Mallman, E. P., & Zoback, M. D. (2007). Assessing elastic Coulomb stress transfer models using seismicity rates in southern California and southwestern Japan. *Journal of Geophysical Research*, *112*, B03304. https://doi.org/10.1029/2005JB004076

Meier, M. A., Werner, M. J., Woessner, J., & Wiemer, S. (2014). A search for evidence of secondary static stress triggering during the 1992 M_w7.3 Landers, California, earthquake sequence. *Journal of Geophysical Research*, *119*, 3354–3370. https://doi.org/10.1002/2013JB010385

Nostro, C., Cocco, M., & Belardinelli, M. E. (1997). Static stress changes in extensional regimes: An application to southern Apennines (Italy). *Bulletin of the Seismological Society of America*, *87*(1), 234–248.

Okada, Y. (1992). Internal deformation due to shear and tensile faults in a half-space. *Bulletin of the Seismological Society of America*, *82*(2), 1018–1040.

Parsons, T., Stein, R. S., Simpson, R. W., & Reasenberg, P. A. (1999). Stress sensitivity of fault seismicity: A comparison between limited-offset oblique and major strike-slip faults. *Journal of Geophysical Research*, *104*, 20,183–20,202. https://doi.org/10.1029/1999JB900056

Pollitz, F. F., & Sacks, I. S. (2002). Stress triggering of the 1999 Hector Mine earthquake by transient deformation following the 1992 Landers earthquake. *Bulletin of the Seismological Society of America*, *92*(4), 1487–1496. https://doi.org/10.1785/0120000918

Prejean, S. G., Hill, D. P., Brodsky, E. E., Hough, S. E., Johnston, M. J. S., Malone, S. D., … Richards-Dinger, K. B. (2004). Remotely triggered seismicity on the United States west coast following the M_w 7.9 Denali fault earthquake. *Bulletin of the Seismological Society of America*, *94*(6B), S348–S359.

Reasenberg, P. A., & Simpson, R. W., (1992). Response of regional seismicity to the static stress change produced by the Loma Prieta earthquake. *Science*, *255*(5052), 1687–1690. https://doi.org/10.1126/science.255.5052.1687

Schisterman, E. F., Perkins, N. J., Liu, A., & Bondell, H. (2005). Optimal cut-point and its corresponding Youden Index to discriminate individuals using pooled blood samples. *Epidemiology*, *16*(1), 73–81. https://doi.org/10.1097/01.ede.0000147512.81966.ba

Stein, R. S., Barka, A. A., & Dieterich, J. H. (1997). Progressive failure on the North Anatolian fault since 1939 by earthquake stress triggering. *Geophysical Journal International*, *128*(3), 594–604. https://doi.org/10.1111/j.1365-246X.1997.tb05321.x

To, A., Bürgmann, R., & Pollitz, F. (2004). Postseismic deformation and stress changes following the 1819 Rann of Kachchh, India earthquake: Was the 2001 Bhuj earthquake a triggered event? *Geophysical Research Letters*, *31*, L13609. https://doi.org/10.1029/2004GL020220

Toda, S., Stein, R. S., Reasenberg, P. A., Dieterich, J. H., & Yoshida, A. (1998). Stress transferred by the 1995 M_w = 6.9 Kobe, Japan, shock: Effect on aftershocks and future earthquake probabilities. *Journal of Geophysical Research*, *103*, 24,543–24,565.

van der Elst, N. J., & Brodsky, E. E. (2010). Connecting nearfield and far-field earthquake triggering to dynamic strain. *Journal of Geophysical Research*, *115*, B07311. https://doi.org/10.1029/2009JB006681