



RESEARCH ARTICLE

10.1029/2020JB019553

Is Coulomb Stress the Best Choice for Aftershock Forecasting?

Key Points:

- Our tests for many mainshocks indicate that considering the variability of aftershock mechanisms significantly improves Coulomb forecasts
- A distance-slip probabilistic model outperforms all stress-based metrics concerning the forecast of the aftershock area
- The best forecasting metric for aftershock forecasting is simple shear scalars

Supporting Information:

- Supporting Information S1
- Data Set S1

Correspondence to:

S. Sharma,
sharma@gfz-potsdam.de

Citation:

Sharma, S., Hainzl, S., Zöeller, G., & Holschneider, M. (2020). Is Coulomb stress the best choice for aftershock forecasting? *Journal of Geophysical Research: Solid Earth*, 125, e2020JB019553. <https://doi.org/10.1029/2020JB019553>

Received 7 FEB 2020

Accepted 2 SEP 2020

Accepted article online 7 SEP 2020

Shubham Sharma^{1,2} , Sebastian Hainzl² , Gert Zöeller³ , and Matthias Holschneider³

¹Institute of Geosciences, University of Potsdam, Potsdam, Germany, ²GFZ German Research Centre for Geosciences, Potsdam, Germany, ³Institute of Mathematics, University of Potsdam, Potsdam, Germany

Abstract The Coulomb failure stress (CFS) criterion is the most commonly used method for predicting spatial distributions of aftershocks following large earthquakes. However, large uncertainties are always associated with the calculation of Coulomb stress change. The uncertainties mainly arise due to nonunique slip inversions and unknown receiver faults; especially for the latter, results are highly dependent on the choice of the assumed receiver mechanism. Based on binary tests (aftershocks yes/no), recent studies suggest that alternative stress quantities, a distance-slip probabilistic model as well as deep neural network (DNN) approaches, all are superior to CFS with predefined receiver mechanism. To challenge this conclusion, which might have large implications, we use 289 slip inversions from SRCMOD database to calculate more realistic CFS values for a layered half-space and variable receiver mechanisms. We also analyze the effect of the magnitude cutoff, grid size variation, and aftershock duration to verify the use of receiver operating characteristic (ROC) analysis for the ranking of stress metrics. The observations suggest that introducing a layered half-space does not improve the stress maps and ROC curves. However, results significantly improve for larger aftershocks and shorter time periods but without changing the ranking. We also go beyond binary testing and apply alternative statistics to test the ability to estimate aftershock numbers, which confirm that simple stress metrics perform better than the classic Coulomb failure stress calculations and are also better than the distance-slip probabilistic model.

1. Introduction

Large earthquakes are almost always followed by a sequence of aftershocks in the first months to years, which might themselves be destructive as, for example, in the case of the 2011 M_w 6.2 Christchurch event, which was triggered by the M_w 7.1 Darfield mainshock (Stramondo et al., 2011). It is generally accepted that aftershocks result from stress changes induced by the mainshock. In particular, the Coulomb Failure Stress (CFS) is commonly used as scalar quantification of the stress state. It is defined as

$$CFS = \tau - \mu(\sigma - p) \quad (1)$$

where τ is the shear stress in slip direction on the fault plane, σ is the normal stress (positive for compression), μ is the coefficient of friction, and p is the pore fluid pressure. Positive CFS changes, $\Delta CFS > 0$, indicate areas of potential aftershock activity, while no aftershocks are expected in regions with $\Delta CFS < 0$. Many studies have demonstrated a clear correlation of the spatial aftershock pattern with static CFS changes calculated based on slip models (Harris, 1998; King et al., 1994; Steacy, Gomberg, et al., 2005; Stein, 1999). However, the applicability is still ambiguous, especially because of the lack of observational evidence for seismic quiescence in stress shadow areas associated to $\Delta CFS < 0$ (Felzer & Brodsky, 2005; Harris & Simpson, 2002) and the missing effect of dynamic stress triggering (Felzer & Brodsky, 2006).

In general, ΔCFS calculations rely on information that contains large uncertainties, such as nonunique inversion of slip models (Hainzl et al., 2009), secondary stress triggering (Helmstetter et al., 2005), and unknown receiver fault mechanisms. CFS calculations require a definition of the fault geometry and slip direction to calculate the Coulomb stress. This is typically done by resolving stress (1) on faults with known geometry or (2) on optimally oriented planes (OOP) having maximum total Coulomb stress. Both approaches are limited either due to poorly constrained fault geometries, ignoring blind faults that could pose significant threat, or to an unknown background stress field. To account for these problems, Steacy, Nalbant, et al. (2005) suggested to fix the strike according to the regional fault trends and vary dip and rake

©2020. The Authors.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

to maximize the total stress tensor, while Hainzl et al. (2010) proposed to use a distribution of receiver fault orientation to estimate the Δ CFS net effect. As a result, stress shadows are less pervasive in agreement with observations. Statistical space-time seismicity models have also been developed based on Δ CFS calculations (so-called hybrid models), for example, Bach and Hainzl (2012) implemented Δ CFS maps as spatial kernel in the epidemic-type aftershock sequence (ETAS) model, Steacy et al. (2013) combined the spatial constraints from Coulomb stress with the short-term earthquake probability (STEP) model (Gerstenberger et al., 2005), and Cattania et al. (2014, 2015) implemented Δ CFS in the rate-state model of Dieterich (1994) accounting for uncertainties. Those models are shown to be successful in explaining the observed trends in seismicity (Cattania et al., 2018).

Recently, the use of CFS has been questioned in general by the studies of Meade et al. (2017) and DeVries et al. (2018) showing that CFS is clearly outperformed by alternative stress metrics and deep neural network (DNN) techniques. They used receiver operating characteristic (ROC) analysis to assess the ability to forecast aftershock areas. The results suggest that alternative stress metrics such as maximum shear and von-Mises stress are more accurate and reliable than CFS. Mignan and Broccardo (2019) recently questioned the results of the DNN approach and stress metrics, proposing a distance-based approach, which is simpler and superior. In this study, we challenge the previous results by repeating the analysis with more appropriate stress calculations as well as alternative tests. In particular, we use all available slip inversions from SRCMOD database to calculate more realistic CFS values assuming layered-half spaces and variable receiver mechanisms. Furthermore, we explore the previously ignored effect of different magnitude cutoffs, grid size variations, and aftershock durations to verify the use of ROC analysis for ranking the stress metrics. We also perform a sensitivity test for aftershock locations using Monte Carlo simulations of catalogs with spatial uncertainties drawn from a Gaussian distribution around the real hypocenter. Because of limitations of the binary forecasts and the ROC analysis, we finally perform additional tests of forecasts of aftershock numbers based on the different stress metrics.

2. Data

We use finite-fault rupture models from the SRCMOD (<https://equake‐rc.info/SRCMOD/>) database by Mai and Thingbaijam (2014). As of 27 November 2019, the database consist of 406 models from 188 earthquakes. However, in this study, we use only a subset of 289 models related to 130 distinct earthquakes for which reviewed aftershock data were available. The slip models are based on single or joint inversion of seismic, geodetic, and other available data.

We use the International Seismological Center (ISC) catalog and select all events occurred within 1 year and within 100 km horizontal distance to the mainshock fault, with a depth range from 0 to 50 km. The catalog is obtained in the form of a pickle file (binary format) taken from the released data of DeVries et al. (2018). These events are called aftershocks, despite the fact that some of them were probably not be related to the mainshock. The catalog covers the period between 1 January 1964 and 30 November 2012. Out of 1,689,845 total events in the reviewed catalog, selection yields 410,064 aftershocks for the analysis.

3. Stress Metrics and Distance Model

The finite-fault rupture models are used to calculate stress changes in a region up to 100 km away from the rupture plane and from 0 to 50 km depth. We use the PSGRN + PSCMP tool by Wang et al. (2006) to calculate the stress tensor in a $5 \times 5 \times 5$ km gridded volume. Given the stress tensor, we calculate five different scalar quantifications. Additionally, we also use the distance-slip probabilistic model suggested by Mignan and Broccardo (2019) comparing the performance of stress based metrics:

1. Δ CFS on master fault orientation (MAS), where Δ CFS is calculated for a receiver mechanism identical to mainshock mechanism.
2. Δ CFS on optimally oriented planes (OOP), assuming a background stress field with principal stress components $\sigma_1 = 1$, $\sigma_2 = 0$, and $\sigma_3 = -2$ MPa, that is, a differential stress of 3 MPa, which is in agreement to the average stress drop of interplate earthquakes (Allmann & Shearer, 2009). The orientation of the principle components is in a way that the stress field is optimally oriented for the mainshock rupture.

3. ΔCFS assuming fault variability (VM): Here the average ΔCFS value is calculated for receiver faults with positive stress changes (no aftershocks are expected on other faults) according to

$$\overline{\Delta CFS} = \frac{1}{N_p} \sum_{i=1}^{N_p} \Delta CFS_i \cdot H(\Delta CFS_i) \quad (2)$$

where H is a Heaviside function and $N_p = 1,500$ is the number of random planes which are selected from a Gaussian distribution centered around the mainshock mechanism, with an assumed standard deviation of 30° for strike, dip, and rake.

4. Maximum Shear (MS)

$$\Delta \tau_{max}(\chi) = |\chi_1 - \chi_3|/2 \quad (3)$$

where χ is the stress tensor and χ_1 and χ_3 are corresponding eigenvalues of the deviatoric stress tensor.

5. von-Mises stress (VMS)

$$\sqrt{3J_2} = \sqrt{I_1^2(\sigma) - 3I_2(\sigma)} \quad (4)$$

where I_1 and I_2 are first and second invariant of deviatoric stress tensor. VMS is a scaled version of second invariant of the deviatoric stress change tensor (DeVries et al., 2018). If VMS of a material under a load is equal or greater to the yield limit, then the material will yield.

6. Distance-slip probabilistic model (R), which was introduced by Mignan and Broccardo (2019) by a logistic regression based on average slip, d , and the minimum distance, r , between the fault and grid node. The probability $Pr(r, d)$ of earthquake occurrence in each grid is given by

$$Pr(r, d) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 \log r + \beta_2 \log d)}} \quad (5)$$

with parameters $\beta_0 = 10.18 \pm 0.07$, $\beta_1 = -2.32 \pm 0.02$, and $\beta_2 = 1.16 \pm 0.01$, which were obtained by a fit to 75% of the data set.

In general, the stress metrics are calculated for a layered half-space, where the layering is based on the CRUST 2.0 (Bassin, 2000) velocity model. However, as a reference model, we use the ΔCFS values calculated for a homogeneous half-space (MAS_0) with assumed Lamé's parameter $\lambda = \mu = 30$ Gpa. For the Coulomb stress calculations, we use the constant apparent friction model (Cocco & Rice, 2002), according to which the Coulomb-stress changes can be written as $\Delta CFS = \Delta \tau - \tilde{\mu} \Delta \sigma$ with the effective friction coefficient $\tilde{\mu} = (1 - B)\mu$, where B is the Skempton's coefficient. In our study, we use a value of $\tilde{\mu} = 0.4$ (King et al., 1994).

4. Methods

To evaluate the forecasting capability, we use the same binary classification method, that is, receiver operating characteristic (ROC) analysis, which has been introduced in previous studies to rank the performance of metrics (DeVries et al., 2018; Meade et al., 2017). In ROC analysis, a two-by-two table (Figure 1) is defined to compare results of model outcomes with observed aftershocks. For different thresholds, the number of true positive (TP), false positive (FP), true negative (TN), and false negative (FN) cases are counted, and the true positive rate (TPR) and false positive rate (FPR) are calculated as

$$TPR = \frac{TP}{TP + FN} \quad (6)$$

$$FPR = \frac{FP}{TN + FP} \quad (7)$$

The cutoff thresholds are defined by stress change values, where stress values are first arranged in ascending order before each stress value is used as a cutoff to calculate TPR and FPR. A test with random classification of binary data has equal rates of true-positive and false-positive classification. For such a case, the area under the curve (AUC) value of a ROC curve is equal to 0.5. A model that has $AUC > 0.5$ is better than a random classifier. A model performing no better than a random classifier ($AUC < 0.5$) can be rejected. Therefore, we use the ROC analysis to quantify the accuracy of our metrics for classifying areas with or without aftershocks.

		Observed	
		Earthquake	No Earthquake
Model	Earthquake	True Positive (TP)	False Positive (FP)
	No Earthquake	False Negative (FN)	True Negative (TN)

Figure 1. Table defining the true positive, true negative, false positive, and false negative classes.

The ROC results can be biased by the inhomogeneity of the earthquake catalog due to varying completeness over time and space, associated location uncertainties, and the occurrence of background activity, which is not related to the mainshock. To test the potential effect of these issues, we also calculated AUC for different aftershock durations (aftershocks up to 1 year excluding first 24 hours, aftershocks in the first 3, 9, and 12 months after the main shock). Furthermore, we explore the potential dependence of the results on the grid size (2.5, 5, and 10 km). We also perform a sensitivity test concerning the effect of location uncertainties by repeating the analysis for 25 randomized catalogs, where the original earthquake location is perturbed in each direction by a Gaussian distribution with standard deviation of 3 km.

Parsons (2020) recently discussed the fact that the imbalance of typical aftershock distributions with most areas lacking events inhibits resolving power of the ROC analysis. The binary ROC test generally suffers from the fact that the test is dependent on the magnitude cutoff. This can be easily seen by considering the end-member case of an earthquake catalog, which is complete to very low magnitudes. In this case, earthquakes would likely be recorded in all subvolumes due to ongoing background activity which make the ROC results insensitive to the tested metric. The same holds for the opposite case that the threshold is too high and no aftershock is found. For intermediate cases, the test has some statistical power which depends on the cutoff value. For illustration of the dependence, we present in section 5 the ROC results for different magnitude cutoffs in the case of the 1999 Chi-Chi sequence. Later on we also present the results of a systematic test of the dependence of the AUC values on aftershock duration and magnitude cutoff for all slip models to find the best possible combination of both choices for forecasting aftershocks. The four chosen combinations are (1) aftershocks within the first year and a cutoff magnitude $M_m - 3$ with M_m being the mainshock magnitude, (2) aftershocks within the first 3 months without cutoff magnitude, (3) aftershocks within the first 3 months and cutoff magnitude $M_m - 3$, and (4) all aftershocks within 1 year without any cutoff magnitude.

Because of the general flaws of the ROC analysis, we additionally perform tests of the number forecasts as used in CSEP tests for seismicity models (Jordan, 2006). In this way, we test both the ability to forecast the activated area and the strength of the activation, which is important for real hazard assessment. For this, we transform the spatial stress values into a spatial probability map of aftershock occurrences, which is based on the assumption of a linear dependence of the triggering potential on the stress change (if positive). In particular, the number of events in each grid cell (λ_n) is assumed to be proportional to the positive stress metric change ($S_n > 0$)

$$\lambda_n = c \cdot S_n \cdot H(S_n) \quad (8)$$

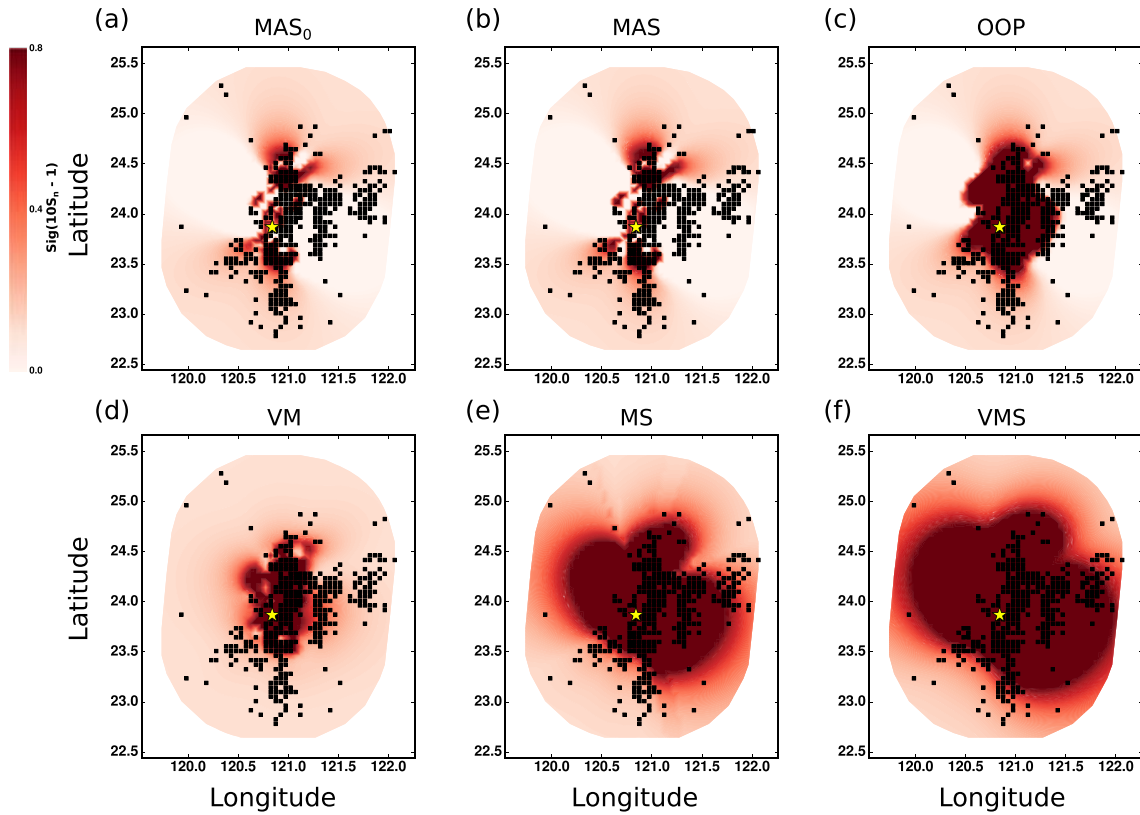


Figure 2. Stress maps calculated for 1999 Chi-Chi earthquake at the hypocentral depth of 7.5 km and passed to Sigmoid filter. (a) shows the reference model for comparison; (b)–(f) are stress metrics MAS, OOP, VM, MS, and VMS, respectively. Black squares indicate areas, where one or more aftershocks occurred in the ± 2.5 km depth interval. The yellow star refers to the mainshock epicenter.

where H is the Heaviside function. The normalization factor c is determined by the condition that summing over the whole region results in the total number of observed aftershocks

$$\sum_{n=1}^{N_g} \lambda_n = c \cdot \sum_{n=1}^{N_g} S_n \cdot H(S_n) = N_{obs} \quad (9)$$

where N_{obs} is the total number of observed aftershocks within 1 year after the mainshock and N_g is the number of grids cells in the region. Equations (8) and (9) can be used to determine the number of events in each grid cell. The likelihood of the observation in each cell is calculated by assuming a Poissonian process with an average rate λ_n . However, this approach would lead to an immediate falsification of a model, if the model predicts 0 activity in a grid cell, where one or more events occurred. While this is theoretically correct, real aftershock catalogs are potentially contaminated by background activity and incorrect locations. To minimize this problem, we distribute a fraction f of aftershocks homogeneously in space, that is,

$$\lambda_n = \frac{N_{obs}}{\sum S_n \cdot H(S_n)} \cdot S_n \cdot H(S_n)(1 - f) + \frac{N_{obs}}{N_{grid}} f. \quad (10)$$

Here we choose a fraction ($f = 0.01$) which is uniformly distributed. The probability that N_n events will occur in a given time period and in the n th grid with predicted rate λ_n is described by to the Poisson model with

$$P_n(N_n | \lambda_n) = \exp(-\lambda_n) \frac{\lambda_n^{N_n}}{N_n!} \quad (11)$$

and the joint log-likelihood for all grid cells becomes

$$LL = \sum_{n=1}^{N_g} \log(P_n(N_n | \lambda_n)). \quad (12)$$

The joint log-likelihood has a negative value, and the values closer to zero indicate that forecasts are close to the observations.

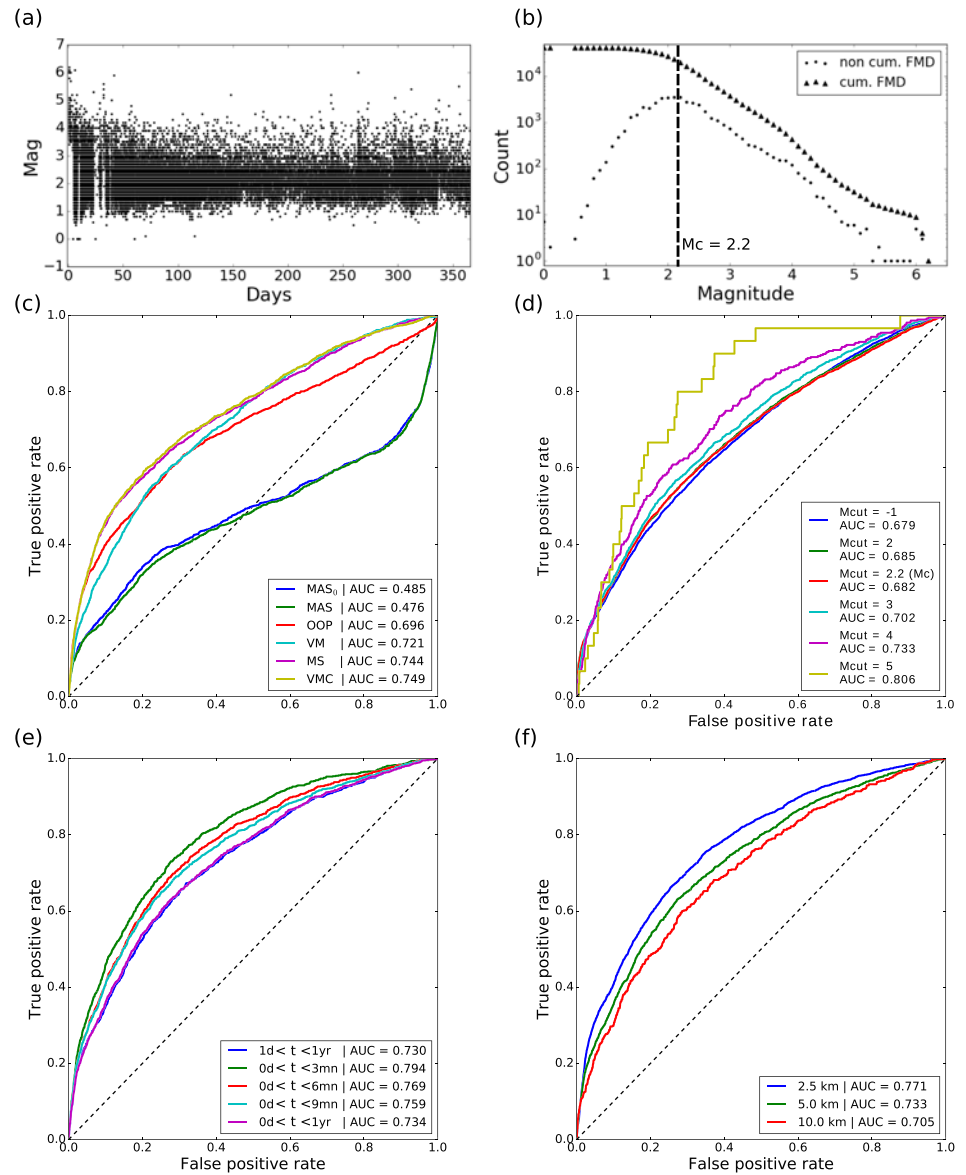


Figure 3. ROC analysis of 1999 Chi-Chi sequence: (a) shows the aftershocks within the first year after the mainshock; (b) cumulative and noncumulative frequency-magnitude distribution, where the estimated completeness magnitude M_c is marked at 2.2; (c) shows the ROC curves for the different stress metrics, and (d)–(f) are the ROC curves for the VM stress metric with different M_{cut} , aftershock duration, and grid size, respectively.

5. Results

The stress tensors generated from the PSGRN + PSCMP (Wang et al., 2006) program are used to calculate the stress metrics described in section 3. While our quantitative analysis is always done for the calculated stress in the whole gridded volume extending ± 100 km horizontally and from 0 to 50 km in depth, we first illustrate the calculations by selecting a specific depth level for one specific case, namely, the slip distribution of 1999 Chi-Chi earthquake derived by Ma et al. (2001). In Figure 2, we compare the stress maps generated by the different scalar stress calculations. Classical Coulomb stress calculations have positive and negative values which mark the regions with and without possible aftershocks, respectively. Correspondingly, the values of MAS_0 , MAS, and OOP are negative and positive. In contrast, VM, MS, and VMS are only positive. In order to compare the results on a same scale we use a Sigmoid filter $sig(10S_n - 1)$ (where $sig(x) = \frac{1}{1 + e^{-x}}$). The stress maps are computed at the depth of the mainshock hypocenter (7.5 km), where the epicenter of the mainshock is marked by the yellow star. The stress maps are compared to areas (black squares) where

aftershocks occurred within 1 year after the mainshock in the depth interval of 7 ± 2.5 km. Δ CFS for MAS_0 and MAS (Figures 2a and 2b, respectively) show very little to no difference in their maps. However, the OOP-type Coulomb stress map (Figure 2c) is significantly different from the former as verified in previous studies. There are few regions with high sigmoid values and no aftershocks and few aftershocks occurring in the stress shadows. Figure 2c, related to Δ CFS calculated on distributed planes, shows maximum sigmoid values in the near-fault region with decreasing values in the far field. Figures 2e and 2f, which are related to maximum stress and von-Mises stress, indicate increased stress values in the near as well as the far field.

For the same example, we also performed a detailed analysis of the ROC test. The availability of a large number of aftershocks in the catalog of the Chi-Chi event makes it a suitable case for testing. The catalog is downloaded from ISC and contains 41,351 events. Unlike the catalog mentioned in section 3 this is an updated catalog with significantly lower magnitude cutoff. Figure 3a shows the aftershocks in a volume of $100 \times 100 \times 50$ km plotted as magnitude versus time. We use the frequency magnitude distribution (Figure 3b) to estimate the magnitude of completeness ($M_c \approx 2.2$). The ROC curves using all aftershocks and stress metrics in the gridded $100 \times 100 \times 50$ km region are plotted in Figure 3c. The analysis reveals that the best performing metrics are maximum shear (AUC = 0.744) and von-Mises stress (AUC = 0.749). The AUC value for the VM model with Δ CFS calculated on distributed planes is lower but also performs well (AUC = 0.721). For the specific case of the VM model, we additionally check the dependence of ROC curves on the magnitude cutoff (M_{cut}), aftershock duration, and grid size. The magnitude cutoff test (Figure 3d) was performed for cutoff magnitudes of -1 (complete catalog), 2, 2.2 (M_c), 3, 4, and 5. A clear dependence of AUC on cutoff magnitude is observed, where AUC increases with increasing M_{cut} . This result indicates that larger aftershocks are in better agreement with the calculated stress maps.

The ROC analysis might be significantly biased by background activity, which may begin to dominate with increasing time, as well as catalog incompleteness directly after the mainshock. To explore the potential effects, we also analyzed the ROC curves for different aftershock durations. We find that ignoring the first day, where the catalog is likely incomplete, does not have a significant effect (AUC of 0.696 instead of 0.698). However, ignoring later events does have a clear impact. The results for the first 3, 6, 9, and 12 months after the mainshock show that the AUC value systematically decreases (Figure 3e) with increasing aftershock duration. The maximum AUC is observed for the aftershocks occurring within the 3 months of the mainshock. This indicates that the occurrence of background events in the later phases can significantly blur the test results. The dependence of AUC on grid size is tested by calculating the stress tensors at the centroid of cubes with edge length of 2.5, 5, and 10 km cells. Figure 3f shows the resulting ROC curves for the VM stress metric calculated on different grid sizes. A systematic dependence of AUC on grid size is observed where smaller grid sizes increase the performance of the metric, indicating that the details in the stress maps are useful for aftershock forecasting.

To analyze whether the results for the Chi-Chi event can be generalized, we generated ROC curves for all 289 slip distributions and corresponding aftershock distributions. All resulting ROC curves are plotted in Figure 4 (thin gray lines) for all stress metrics introduced in section 3, where the thick blue line refers to the average curve for each stress metric. The average is calculated by averaging true positive rates in false positive rate bins. We then determined the corresponding AUC for the average curve. We observe no clear difference between the reference metric MAS_0 and MAS calculated for regional layered crust models (Figures 3a and 3b). As expected, Δ CFS resolved on OOP results in a higher AUC (0.659) than MAS_0 and MAS (0.491 and 0.490, respectively). Furthermore, the AUC value (0.718) for Δ CFS calculated for variable mechanism (VM) is significantly higher as compared to the other CFS metrics but lower than the stress scalars MS and VMS (0.743 and 0.746, respectively). However, the maximum AUC value of 0.758 is obtained for the distance-slip probabilistic model (R) for which the figure is included in the supplementary information as Figure S4. Note that the R-model was optimized on a large fraction of the data, while the stress metrics are not optimized. Thus, this comparison might be biased, but the comparison of Mignan and Broccardo (2019) which was performed only on a subset of the test data showed the same tendency.

Figure 5a shows the effect of location uncertainties of 3 km on the AUC values. We find the standard deviation of the resulting mean AUC values is small, and the ranking remains the same. Figure 5b shows a box plot to indicate the distribution of mean, median, quartiles, and extremes of AUC values for all slip models and stress scalars. The metrics OOP, VM, MS, and VMS and distance-slip model (R) have their mean, median, and first quartile all above the AUC threshold (0.5), while the mean and median for MAS_0 and MAS

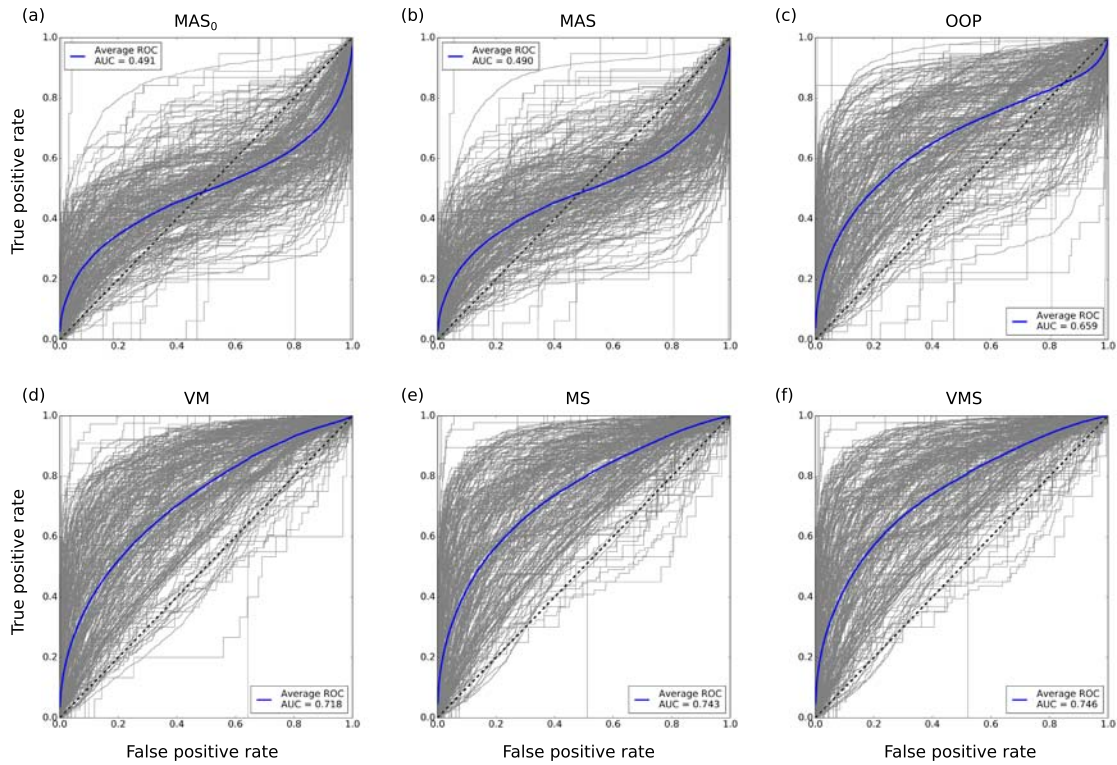


Figure 4. ROC analysis (a–f) for all analyzed slip distributions and stress metrics (a) MAS_0 , (b) MAS, (c) OOP, (d) VM, (e) MS, and (f) VMS, where thin lines are related to the result of individual slip distributions and the thick blue line is the binned average with corresponding AUC value mentioned in the legend. A reference line is plotted as black dotted line, which refers to the result for random data.

do not even cross this threshold. R performs best in terms of mean, median, and quartile range, but the distribution strongly overlaps with those of VMS, MS, and VM as the next best performing metrics. To test the robustness of the ranking, we repeated the calculation for different combinations of M_{cut} and the aftershock period (see section 4). Figure 5c shows the results of the test, which confirms the robustness of the ranking. However, the AUC scores systematically vary for all metrics and for the distance-slip model depending

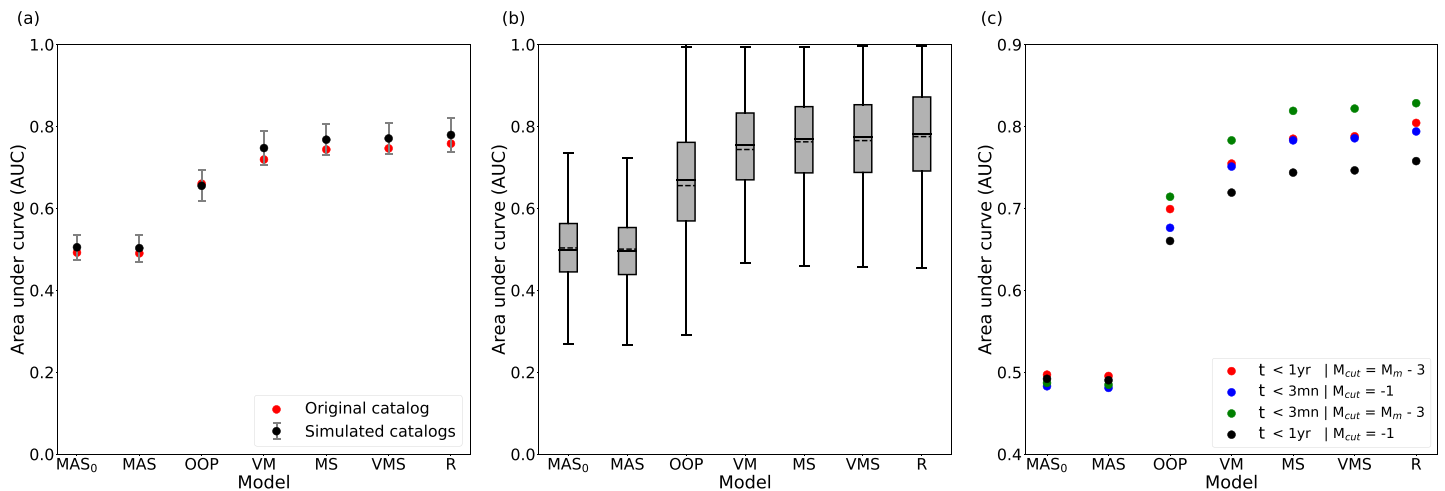


Figure 5. AUC results as a function of the model type: (a) mean AUC values with its uncertainties (\pm on standard deviation) assuming a location error of 3 km (see section 4); (b) AUC results for all slip distributions represented as box plots. The gray part of the box shows the interquartile region (IQR) with first and third quartile as bounds. The solid and dotted horizontal lines inside the box refer to the median and mean values, and the extreme ends (whiskers) indicate the minimum and maximum of all results. (c) shows the average AUC value for the different aftershock sets mentioned in section 4.

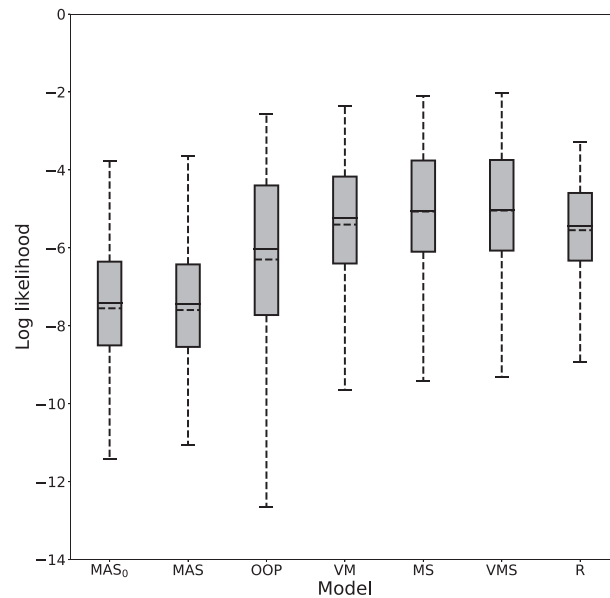


Figure 6. Result of the number forecast test: box plots of the distribution of normalized LL values for all analyzed slip models in dependence of the stress metric. The symbols are defined as in Figure 4a.

on the different settings. In particular, the best performing scenario (Figure 5c green dots) is obtained for aftershocks within the first 3 months after the mainshock and $M_{cut} = M_m - 3$.

Now, we go beyond binary testing and use a statistical test to estimate the aftershock numbers because real aftershock forecasts rely on the event density as discussed in section 4. The result of the log-likelihood (LL) test is shown in Figure 6. For a better comparison of LL values for mainshock-aftershock sequences with different number of aftershocks, the resulting LL value was divided by the number of aftershocks in each case. The results are presented by box plots (whisker diagrams), which are used to study the distribution of LL values. We observe a very similar trend as in the case of the ROC analysis. However, in this case, the distance-slip probabilistic model (R) now ranks below the Δ CFS-based VM metric. Stress metrics VM perform significantly better than the conventional Coulomb stress calculations, but MS and VMS are still better. The best result in terms of number forecasts is obtained for VMS.

6. Discussion

Coulomb failure stress has been largely used to explain earthquake triggering and particularly to describe the locations of aftershocks by separating the region into positive and negative stress change areas. Theoretically, aftershocks are only expected in regions with increased CFS value. However, it has been already recognized that aftershocks frequently occur in regions with calculated, negative stress changes, so-called stress shadows (Felzer & Brodsky, 2005; Harris & Simpson, 2002). It has been suggested that this is related to uncertainties in the slip models (Hainzl et al., 2009; Helmstetter & Shaw, 2006; Marsan, 2006) as well as the variability of the receiver mechanisms (e.g., Hainzl et al., 2010). Taking these uncertainties into account has been previously shown to improve the forecasts significantly (Cattania et al., 2014). A study conducted by Steacy, Nalbant, et al. (2005) suggests in a poorly defined regional stress area, or in a structurally complex area, the strike of the receiver planes should be fixed to that of the mainshock plane and let the dip and rake vary to calculate the Coulomb stress.

The uncertainties and finite resolution of finite-fault models reduce the capability of Coulomb stress to reproduce the on-fault aftershocks (Steacy et al., 2004). With majority of aftershocks occurring in the proximity of the mainshock rupture plane (Felzer & Brodsky, 2006; Gu et al., 2013; Moradpour et al., 2014), the Coulomb stress metric is expected to suffer more strongly than the simple metrics from those limitations. In order to test whether this explains our results, we repeated the ROC analysis by eliminating near-fault grid cells as well as their respective aftershocks (see Figure S3). While MAS_0 and MAS do not significantly improve, the performance of the OOP metric deteriorates when excluding near-fault aftershocks. While this

is counterintuitive, it is likely related to the effect of background activity, which is not associated to the mainshock stress. Excluding the near-fault area also excludes the area with the highest signal-to-noise ratio.

The recent studies of Meade et al. (2017) and DeVries et al. (2018) claimed that simple stress metrics, which do not need any specification of the receiver mechanism, are superior to the Coulomb stress calculations. However, their claim can be challenged because of the unrealistic CFS calculations, which did not account for the known uncertainties in the CFS calculations, as well as the potential artifacts and shortcomings of their ROC analysis. Thus, we performed a systematic reanalysis including (i) previously introduced CFS scalars accounting for receiver fault variability (OOP and VM), (ii) improved stress calculations based on regional, layered velocity models, (iii) different time windows and magnitude cutoffs, and (iv) the LL test quantifying the forecasts of the spatial distribution of aftershock numbers. Countering the approach of DeVries et al. (2018) and Mignan and Broccardo (2019) shows that a logistic regression model using average mainshock slip and measured distance performs better than DNN approach. Hence, we consider the distance-based model as reference model and perform a detailed analysis to compare the performance of stress metrics.

Our comprehensive analysis shows that the results of OOP and VM are significantly better than the previously tested MAS value. This indicates the importance in accounting for the variability of aftershock mechanisms to get more realistic CFS predictions. Further, geological constraints can be used to narrow down the standard deviation in receiver fault distribution to obtain more realistic results. While there are improvements observed in the performance of OOP and VM, scalar stress metrics still performed significantly better than receiver dependent metrics. The underlying reason is not yet clear and needs further research. It might be that MS and VMS are more efficient in accounting for triggering mechanisms, which are not directly considered, such as afterslip, poro/visco-elastic deformations, and dynamic stress triggering. However, our tests show that distance-slip (R) is the best performing model for forecasting the aftershock area (binary forecast). This result indicates that there might not be any need to calculate stress tensors when forecasting the activation area. In contrast, the best stress metrics are found to outperform the R-model in regards to forecasting earthquake numbers, that is, the spatial density of aftershocks which is more important for seismic hazard studies. It should be noted that our analysis of alternative stress metrics is not exhaustive, and other metrics might be even better. For example, Terakawa et al. (2020) just introduced a new energetics-based stress metrics Δ ETS, jointly accounting for coseismic stress changes and the background stress field. While their test for the case of the Landers aftershock sequence shows encouraging results, Mignan and Broccardo (2020) replied that it is likely not better than their distance-based approach. Whether or not Δ ETS systematically improved forecasts might be tested in a future study similar to the present one.

For our analysis, we tried to use some meaningful parameters for the CFS calculations. Using more realistic friction coefficients, background stresses (OOP), or uncertainties in the receiver mechanisms (VM) might improve the results of the CFS metrics. However, such a retrospective optimization of model parameters would bias the comparison, because MS and VMS have no free parameters. Although the background stress field for the OOP model could be set according to alternate rules. For example, Mignan (2020) suggested that the prestress is released by the mainshock stress drop. So, we recalculate the OOP metric for deviatoric background stress, which is calculated for each mainshock individually to equal the estimated stress drop (see details in the supporting information Figure S5). A notable improvement is observed in the OOP results, but it does not change the ranking of stress metrics and distance-slip model.

Our test results might be distorted by background events and aftershocks triggered by aftershocks, so-called secondary triggering. Our analysis already indicates that the results change for different time windows, with best results for aftershocks in the shortest time period after the mainshock. This points to the blurring effect of background activity. Declustering may possibly remove the effect of higher order aftershocks, but no simple method exists which could be applied on our diverse data set. In general, the epidemic type aftershock sequence (ETAS) model accounts systematically for background as well as secondary aftershock triggering. Implementing a spatial kernel based on the stress metrics could be one way to do systematic tests including background and secondary aftershock triggering. However, those studies can only be performed for individual sequences with high data quality.

In our study, we do not consider the effect of dynamic stress changes, which is used to explain aftershock locations within a week of a large mainshock (Prejean et al., 2004). However, our results indicate that the

stress metrics works best for the largest aftershocks triggering in the shortest time period after the main-shock. This is encouraging, because they have the largest impact on seismic hazard. A thorough analysis can be performed like in Figure 5c to specify the forecasting thresholds on the time period and cutoff magnitudes for different tectonic regions.

It is important to note that our study does not discard the use of CFS in general and Coulomb failure theory might still describe the physics for earthquake triggering. It rather indicates that CFS in the case of limited fault information is not the best choice for aftershock forecasting. However, if precise information about the receiver planes are available, CFS might still be the best choice, for example, to evaluate the trigger potential on well-known neighboring fault segments.

7. Conclusion

Despite its frequent application for several decades, Coulomb failure stress calculations have been questioned by recent studies and shown to be outperformed by other stress scalars and state-of-the-art methods like deep neural network in forecasting aftershocks. However, the recent results are also questionable because of an artificial DNN application (Mignan & Broccardo, 2019) as well as simplified CFS calculations. As this has broad implication for this research area, we performed a comprehensive reanalysis of the previous ROC-based study. Here we include CFS metrics accounting for the variability of aftershock mechanisms and additionally taking account of the incompleteness of catalogs as well as the occurrence of background activity. In addition to the previously conducted ROC analysis for binary forecasts, we also tested forecasts of aftershock numbers.

To summarize, we find that the results of the ROC analysis are dependent on the magnitude cutoff, aftershock duration, and grid sizes and that more realistic CFS calculations (OOP and VM) can significantly improve the results. However, our analysis verifies that the stress scalars MS and VMS, and distance-slip probabilistic model (R), all of which do not rely on any specification of receiver mechanisms, outperform on average the CFS metrics in all test setups. While CFS might still be used for the evaluation of the stress changes on well-defined fault segments, our results indicate that spatial forecasts of the aftershock density might be generally improved by using von-Mises stress (VMS) instead of Coulomb stress.

Data Availability Statement

All the slip data are freely available on <https://equake‐rc.info/srcmod/> and International Seismological Center (ISC) event catalog (<https://www.isc.ac.uk/iscbulletin/search/catalogue/>).

References

- Allmann, B. P., & Shearer, P. M. (2009). Global variations of stress drop for moderate to large earthquakes. *Journal of Geophysical Research*, *114*, B01310. <https://doi.org/10.1029/2008JB005821>
- Bach, C., & Hainzl, S. (2012). Improving empirical aftershock modeling based on additional source information. *Journal of Geophysical Research*, *117*, B04312. <https://doi.org/10.1029/2011JB008901>
- Bassin, C. (2000). The Current Limits of resolution for surface wave tomography in North America. *Eos, Transactions, American Geophysical Union*, *81*, F897.
- Cattania, C., Hainzl, S., Wang, L., Enescu, B., & Roth, F. (2015). Aftershock triggering by postseismic stresses: A study based on coulomb rate-and-state models. *Journal of Geophysical Research: Solid Earth*, *120*, 2388–2407. <https://doi.org/10.1002/2014JB011500>
- Cattania, C., Hainzl, S., Wang, L., Roth, F., & Enescu, B. (2014). Propagation of coulomb stress uncertainties in physics-based aftershock models. *Journal of Geophysical Research: Solid Earth*, *119*, 7846–7864. <https://doi.org/10.1002/2014JB011183>
- Cattania, C., Werner, M. J., Marzocchi, W., Hainzl, S., Rhoades, D., Gerstenberger, M., et al. (2018). The forecasting skill of physics-based seismicity models during the 2010–2012 canterbury, new zealand, earthquake sequence. *Seismological Research Letters*, *89*, 1238–1250.
- Cocco, M., & Rice, J. R. (2002). Pore pressure and poroelasticity effects in coulomb stress analysis of earthquake interactions. *Journal of Geophysical Research*, *107*(B2), 2030. <https://doi.org/10.1029/2000JB000138>
- DeVries, P. M. R., Viégas, F., Wattenberg, M., & Meade, B. J. (2018). Deep learning of aftershock patterns following large earthquakes. *Nature*, *560*(7720), 632.
- Dieterich, J. (1994). A constitutive law for rate of earthquake production and its application to earthquake clustering. *Journal of Geophysical Research*, *99*(B2), 2601–2618.
- 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.
- Gerstenberger, M. C., Wiemer, S., Jones, L. M., & Reasenberg, P. A. (2005). Real-time forecasts of tomorrow's earthquakes in california. *Nature*, *435*(7040), 328–331.
- Gu, C., Schumann, A. Y., Baiesi, M., & Davidsen, J. (2013). Triggering cascades and statistical properties of aftershocks. *Journal of Geophysical Research: Solid Earth*, *118*, 4278–4295. <https://doi.org/10.1002/jgrb.50306>

Acknowledgments

We thank DFG - Research Training Group “NatRiskChange” for funding the project. We are also grateful to Shinji Toda, an anonymous reviewer, the Associate Editor, and the Editor for helpful recommendations. I would also like to thank Tim Davis for proofreading the manuscript and providing valuable suggestions. Open access funding enabled and organized by Projekt DEAL.

- Hainzl, S., Enescu, B., Cocco, M., Woessner, J., Catali, F., Wang, R., & Roth, F. (2009). Aftershock modeling based on uncertain stress calculations. *Journal of Geophysical Research*, *114*, B05309. <https://doi.org/10.1029/2008JB006011>
- Hainzl, S., Zöller, G., & Wang, R. (2010). Impact of the receiver fault distribution on aftershock activity. *Journal of Geophysical Research*, *115*, B05315. <https://doi.org/10.1029/2008JB006224>
- Harris, R. A. (1998). Introduction to special section: Stress triggers, stress shadows, and implications for seismic hazard. *Journal of Geophysical Research*, *103*(B10), 24,347–24,358.
- Harris, R. A., & Simpson, R. W. (2002). The 1999 m w 7.1 Hector mine, California, earthquake: A test of the stress shadow hypothesis? *Bulletin of the Seismological Society of America*, *92*(4), 1497–1512.
- Helmstetter, A., Kagan, Y. Y., & Jackson, D. D. (2005). Importance of small earthquakes for stress transfers and earthquake triggering. *Journal of Geophysical Research*, *110*, B05S08. <https://doi.org/10.1029/2004JB003286>
- Helmstetter, A., & Shaw, B. E. (2006). Relation between stress heterogeneity and aftershock rate in the rate-and-state model. *Journal of Geophysical Research*, *111*, B07304. <https://doi.org/10.1029/2005JB004077>
- Jordan, T. H. (2006). Earthquake predictability, brick by brick. *Seismological Research Letters*, *77*(1), 3–6.
- King, G. C. P., Stein, R. S., & Lin, J. (1994). Static stress changes and the triggering of earthquakes. *Bulletin of the Seismological Society of America*, *84*(3), 935–953.
- Ma, K.-F., Mori, J., Lee, S.-J., & Yu, S. B. (2001). Spatial and temporal distribution of slip for the 1999 Chi-Chi, Taiwan, earthquake. *Bulletin of the Seismological Society of America*, *91*(5), 1069–1087. <https://doi.org/10.1785/0120000728>
- Mai, P. M., & Thingbaijam, K. K. S. (2014). Srcmod: An online database of finite-fault rupture models. *Seismological Research Letters*, *85*(6), 1348–1357.
- Marsan, D. (2006). Can coseismic stress variability suppress seismicity shadows? insights from a rate-and-state friction model. *Journal of Geophysical Research*, *111*, B06305. <https://doi.org/10.1029/2005JB004060>
- Meade, B. J., DeVries, P. M., Faller, J., Viegas, F., & Wattenberg, M. (2017). What is better than coulomb failure stress? A ranking of scalar static stress triggering mechanisms from 105 mainshock-aftershock pairs. *Geophysical Research Letters*, *44*, 11–409. <https://doi.org/10.1002/2017GL075875>
- Mignan, A. (2020). Forecasting aftershocks: Back to square one after a deep learning anticlimax. *Tembler*. <https://doi.org/10.32858/temblor.053>
- Mignan, A., & Broccardo, M. (2019). One neuron versus deep learning in aftershock prediction. *Nature*, *574*(7776), E1–E3.
- Mignan, A., & Broccardo, M. (2020). Comment on “elastic strain energy and pore-fluid pressure control of aftershocks” by Terakawa et al. [*Earth Planet. Sci. Lett.* 535 (2020) 116103]. *Earth and Planetary Science Letters*, *544*, 116,402.
- Moradpour, J., Hainzl, S., & Davidsen, J. (2014). Nontrivial decay of aftershock density with distance in southern California. *Journal of Geophysical Research: Solid Earth*, *119*, 5518–5535. <https://doi.org/10.1002/2014JB010940>
- Parsons, T. (2020). On the use of receiver operating characteristic tests for evaluating spatial earthquake forecasts. *Geophysical Research Letters*, *47*, e2020GL088570. <https://doi.org/10.1029/2020GL088570>
- Prejean, S. G., Hill, D. P., Brodsky, E. E., Hough, S. E., Johnston, M. J. S., Malone, S. D., et al. (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.
- Steady, S., Gerstenberger, M., Williams, C., Rhoades, D., & Christophersen, A. (2013). A new hybrid coulomb/statistical model for forecasting aftershock rates. *Geophysical Journal International*, *196*, 918–923.
- Steady, S., Gombert, J., & Cocco, M. (2005). Introduction to special section: Stress transfer, earthquake triggering, and time-dependent seismic hazard. *Journal of Geophysical Research*, *110*, B05S01. <https://doi.org/10.1029/2005JB003692>
- Steady, S., Marsan, D., Nalbant, S. S., & McCloskey, J. (2004). Sensitivity of static stress calculations to the earthquake slip distribution. *Journal of Geophysical Research*, *109*, B04303. <https://doi.org/10.1029/2002JB002365>
- Steady, S., Nalbant, S. S., McCloskey, J., Nostro, C., Scotti, O., & Baumont, D. (2005). Onto what planes should coulomb stress perturbations be resolved? *Journal of Geophysical Research*, *110*, B05S15. <https://doi.org/10.1029/2004JB003356>
- Stein, R. S. (1999). The role of stress transfer in earthquake occurrence. *Nature*, *402*(6762), 605.
- Stramondo, S., Kyriakopoulos, C., Bignami, C., Chini, M., Melini, D., Moro, M., et al. (2011). Did the September 2010 (Darfield) earthquake trigger the February 2011 (Christchurch) event? *Scientific Reports*, *1*, 98.
- Terakawa, T., Matsu'ura, M., & Noda, A. (2020). Elastic strain energy and pore-fluid pressure control of aftershocks. *Earth and Planetary Science Letters*, *535*, 116,103.
- Wang, R., Lorenzo-Martin, F., & Roth, F. (2006). Psgm/pscmp—A new code for calculating co- and post-seismic deformation, geoid and gravity changes based on the viscoelastic-gravitational dislocation theory. *Computers & Geosciences*, *32*(4), 527–541.