

Geophysical Research Letters®



RESEARCH LETTER

10.1029/2022GL098344

Key Points:

- We develop a probabilistic formulation of earthquake rupture predictability and estimate magnitude distributions given early observations
- The final magnitude of an event can not be determined before the peak moment release from moment rate functions or teleseismic waveforms
- In multi-asperity ruptures, the occurrence of future asperities can not be foreseen

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

J. Münchmeyer,
munchmej@gfz-potsdam.de

Citation:

Münchmeyer, J., Leser, U., & Tilmann, F. (2022). A probabilistic view on rupture predictability: All earthquakes evolve similarly. *Geophysical Research Letters*, 49, e2022GL098344. <https://doi.org/10.1029/2022GL098344>

Received 24 FEB 2022
Accepted 17 JUN 2022

A Probabilistic View on Rupture Predictability: All Earthquakes Evolve Similarly

Jannes Münchmeyer^{1,2} , Ulf Leser², and Frederik Tilmann^{1,3} 

¹Deutsches GeoForschungsZentrum GFZ, Potsdam, Germany, ²Institut für Informatik, Humboldt-Universität zu Berlin, Berlin, Germany, ³Institut für Geologische Wissenschaften, Freie Universität Berlin, Berlin, Germany

Abstract Ruptures of the largest earthquakes can last between a few seconds and several minutes. An early assessment of the final earthquake size is essential for early warning systems. However, it is still unclear when in the rupture history this final size can be predicted. Here we introduce a probabilistic view of rupture evolution - how likely is the event to become large - allowing for a clear and well-founded answer with implications for earthquake physics and early warning. We apply our approach to real time magnitude estimation based on either moment rate functions or broadband teleseismic P arrivals. In both cases, we find strong and principled evidence against early rupture predictability because differentiation between differently sized ruptures only occurs once half of the rupture has been observed. Even then, it is impossible to foresee future asperities. Our results hint toward a universal initiation behavior for small and large ruptures.

Plain Language Summary Earthquakes are among the most destructive natural hazards known to humankind. While earthquakes can not be predicted, it is possible to record them in real-time and provide warnings to locations that the shaking has not reached yet. Warning times usually range from few seconds to tens of seconds. For very large earthquakes, the rupture itself, which is the process sending out the seismic waves, can have a similar duration. Whether the final size of the earthquake, its magnitude, can be determined while the rupture is still ongoing is an open question. Here we show that this question is inherently probabilistic - how likely is an event to become large? We develop a formulation of rupture predictability in terms of conditional probabilities and a framework for estimating these from data. We apply our approach to two observables: moment rate functions, describing the energy release over time during a rupture, and seismic waveforms at distances of several thousand kilometers. The final earthquake magnitude can only be predicted after the moment rate peak, at approximately half the event duration. Even then, it is impossible to foresee future subevents. Our results suggests that ruptures exhibit a universal initiation behavior, independent of their size.

1. Introduction

It is a longstanding question at which time during an earthquake rupture its final size can be constrained. Answering this question would have direct implications for early warning systems (Allen & Melgar, 2019) and would provide insights into the underlying physical processes. Accordingly, its formulation spanned a series of studies over the last decades. However, so far results have been contradictory - some argue for early predictability, others against.

A common theory implying predictability is the preslip model (Ellsworth & Beroza, 1995), in which failure starts aseismically until the process reaches a critical size and becomes unstable. Here, the final moment of the earthquake might be derivable at the event onset time from properties of the nucleation zone, that is, its size or its magnitude of slip. Other models also suggest early predictability, but only after several seconds. For example, Melgar and Hayes (2017) argue that ruptures of large events propagate as self-healing pulses, and that pulse properties, that is, rise time and pulse width, allow identification of very large events after ~15 s. Support for such theories has been provided by the analysis of, for example, waveform onsets (Ellsworth & Beroza, 1995), moment rate functions (Danré et al., 2019), early ground motion parameters (Colombelli et al., 2020), or geodetic observations (Melgar & Hayes, 2017).

The opposing hypothesis, often termed cascade model (Ellsworth & Beroza, 1995), suggests a universal initiation behavior: small and large earthquakes start identically and are differentiated only after the peak moment release, which occurs approximately at half of the rupture duration. Rupture evolution is controlled by heterogeneous local

© 2022. The Authors.

This is an open access article under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

conditions, such as the pre-event stress distribution or the presence of mechanical barriers. Studies supporting this theory also analyzed properties like moment rate functions (Meier et al., 2017), waveform onsets (Ide, 2019), or peak displacement (Goldberg et al., 2019; Trugman et al., 2019).

While reaching contradicting conclusions, predictability studies often follow the same principle: analyzing correspondences between earthquake size and real time observables (Colombelli et al., 2020; Danré et al., 2019; Ellsworth & Beroza, 1995; Ide, 2019; Meier et al., 2017; Trugman et al., 2019). Earthquake size is commonly quantified by seismic moment/moment magnitude, as large, high quality catalogs thereof are openly available (Ekström et al., 2012). A common practice is calculating parametric fits between magnitude and observables, and assessing at which time they become significant using standard deviations (Colombelli et al., 2020; Danré et al., 2019; Melgar & Hayes, 2017; Noda & Ellsworth, 2016; Olson & Allen, 2005; Zollo et al., 2006). However, such point-estimator approaches hide the residual distribution and thereby potentially obscure distinct modes of rupture predictability. In particular, non-Gaussian residuals can lead to unexpected results. First, significant fits between magnitudes and observables might still lead to non-point-predictable models (Figure S1 in Supporting Information S1). Second, even without a (significant) fit, there might be some level of predictability in a probabilistic sense (Figure S2 in Supporting Information S1). Notably, exactly such non-Gaussian distributions occur for real observables; examples can be found in Olson and Allen (2005, their Figure 3, the observable is the dominant period of the initial 4 s of the P wave) or Noda and Ellsworth (2016, their Figure 5, the observable is based on early P displacement waveform).

2. A Probabilistic Framework for Rupture Predictability

We argue that a rigorous description of rupture predictability requires a probabilistic approach. To this end, we interpret the magnitude M of an event as a random variable and introduce a stochastic process $(O_t)_{t \in \mathbb{R}}$, the observables at time t . $t = 0$ identifies the event onset. The observables $(O_t)_{t \in \mathbb{R}}$ can be any information describing the event until t , for example, waveforms up to (P travel time + t) or features of moment rate functions until t .

Events with magnitude $M_1 \neq M_2$ differ at time t if the conditional distributions $\mathbb{P}(O_t|M_1)$ and $\mathbb{P}(O_t|M_2)$ differ. However, while describing $\mathbb{P}(O_t|M)$ for scalar O_t is feasible, it becomes intractable for higher dimensional O_t . Furthermore, early warning aims to estimate M from O_t and not vice versa. Therefore, we analyze $\mathbb{P}(M|O_t)$, that is, how the observables constrain the magnitude. While such analysis has been conducted for single-dimensional observables, for example, peak ground displacement (Meier et al., 2017; Trugman et al., 2019) or the amount of slip (Böse & Heaton, 2010), an analysis for higher dimensional observables is still missing. This leaves many promising observables unexplored, for example, seismic waveforms.

There are two distinct aspects of rupture predictability: (a) the future development of the current asperity and (b) the probability of further asperities to rupture. Figure 1d shows an example of $\mathbb{P}(M|O_t)$ with no predictability in the growing rupture, as suggested, for example, by Meier et al. (2017). Before the peak moment release, the distribution equals a Gutenberg-Richter (GR) distribution with lower bound at the currently released moment, accounting for both aspects. After the peak, the distribution becomes Gaussian (a), with a decreasing GR component accounting for potential future asperities (b). Figure 1e shows a skewed GR case: the magnitude cannot be pinpointed, but from early on the event is more likely to become large than the marginal GR distribution. Skewed GR distributions might occur, for example, due to differences in the fault roughness or the amount of slip (Böse & Heaton, 2010). Figure 1f shows the predictable case: the magnitude can be pinpointed early and uncertainties decrease steadily.

The different evolutions of $\mathbb{P}(M|O_t)$ have consequences for early warning: a shifted tail for $\mathbb{P}(M|O_t)$ shifts the estimated distribution of ground shaking and possibly the warning decision (Böse & Heaton, 2010). However, several previous results do not allow a clear distinction of the presented cases. For example, multiple studies (Abercrombie & Mori, 1994; Ide, 2019; Kilb & Gomberg, 1999; Mori & Kanamori, 1996) reported that for most large events, small events with similar onsets exist. While this rules out the predictable case, events might still differ strongly in their likelihood of becoming large.

For practical analysis, $\mathbb{P}(M|O_t)$ needs to be derived from observed samples $\{(M^i, O_t^i)\}_{i=1, \dots, n} \sim_{iid} \mathbb{P}(M, O_t)$. For single dimensional O_t , for example, peak displacement, $\mathbb{P}(M|O_t)$ can be estimated using histograms. However, this direct description is infeasible for high dimensional O_t . Instead, we use a variational approximation

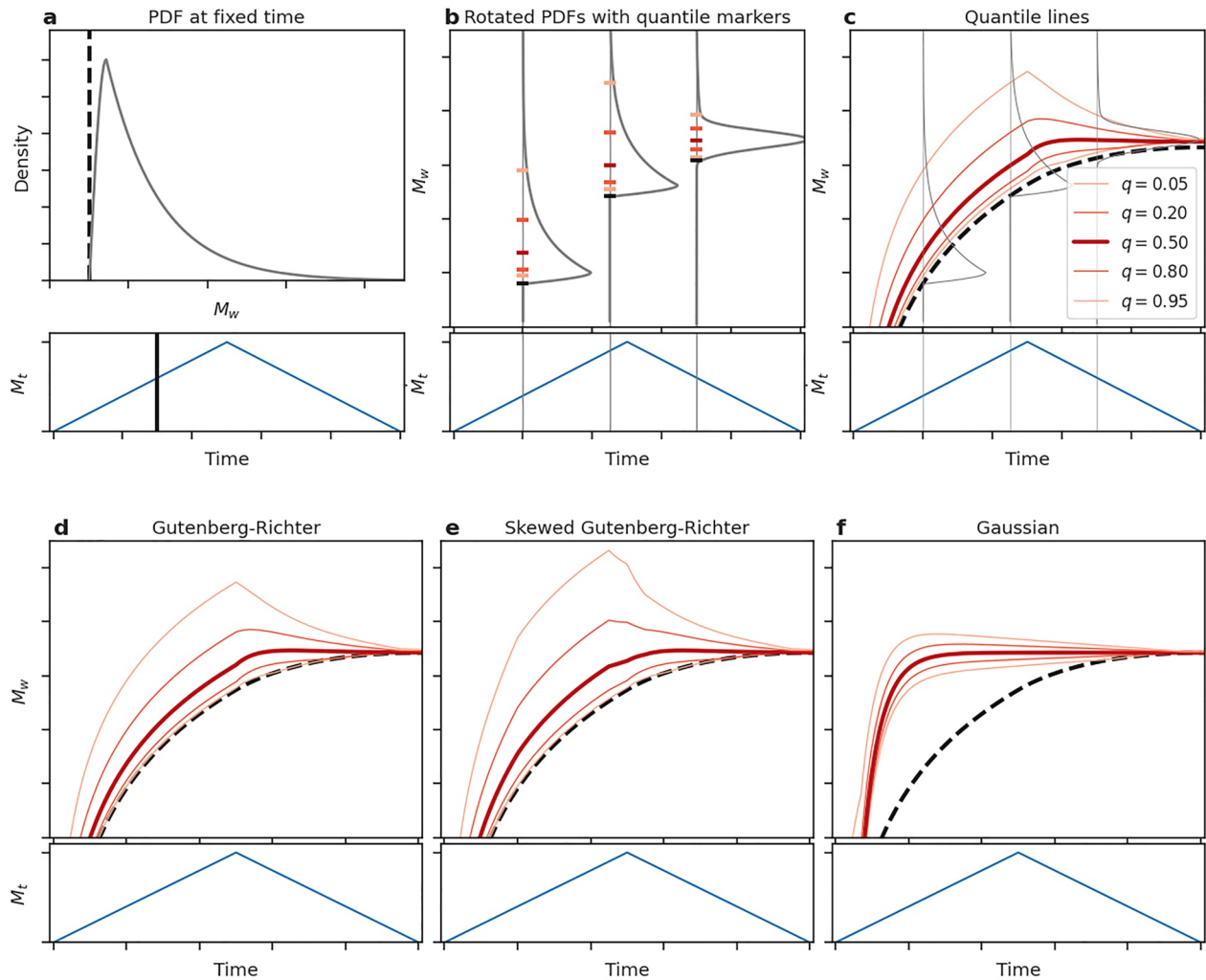


Figure 1. (a-c) Step-by-step explanation of quantile line plots (see subplot titles) (d-f) Magnitude estimate development for three different predictability models: Gutenberg-Richter (GR) (not predictable during growth phase), skewed GR (not point-predictable, but deviation from prior already during growth phase) and Gaussian (predictable). For each model, we use the same hypothetical event with prototypical triangular moment rate functions representing the first order moment release history; for predictable models to be viable, second order features of the moment rate function would differ between smaller and larger events. The dashed black line indicates the cumulative moment release. An additional visualization of the options at a fixed time is shown in Figure S3a in Supporting Information S1.

$\mathbb{P}_\theta(M|O_t) \approx \mathbb{P}(M|O_t)$, that is, a function $\mathbb{P}_\theta(M|O_t)$ mapping observables O_t to the corresponding conditional distribution of M . The parameters θ are learned to fit $\mathbb{P}(M|O_t)$ using $\{(M^i, O_t^i)\}_{i=1, \dots, n}$ (Gneiting & Raftery, 2007). Specifically, we parameterize $\mathbb{P}_\theta(M|O_t)$ using neural networks with Gaussian mixture outputs (Bishop, 1994), because both neural networks and Gaussian mixtures (Figure S3b in Supporting Information S1) have universal approximator properties, making them particularly well suited for our case (Bengio et al., 2017; Cybenko, 1989). Notably, this approach can be applied to any type of observables, simply by designing an appropriate neural network. We now use this approach for two choices of observables: moment rate functions and teleseismic waveforms.

3. Predictions From Moment Rate Functions

We first apply this framework to source time functions (STFs), also known as moment rate functions, a commonly used observable in predictability studies (Danré et al., 2019; Meier et al., 2017). We use three STF databases: SCARDEC (3,514 events, $5.4 \leq M_w \leq 9.1$) (Vallée & Douet, 2016) and the ones from USGS (Hayes, 2017) (190 events, $6.8 \leq M_w \leq 9.1$) and Ye et al. (2016) (119 events, $6.8 \leq M_w \leq 9.1$). Figure 2 shows characteristics

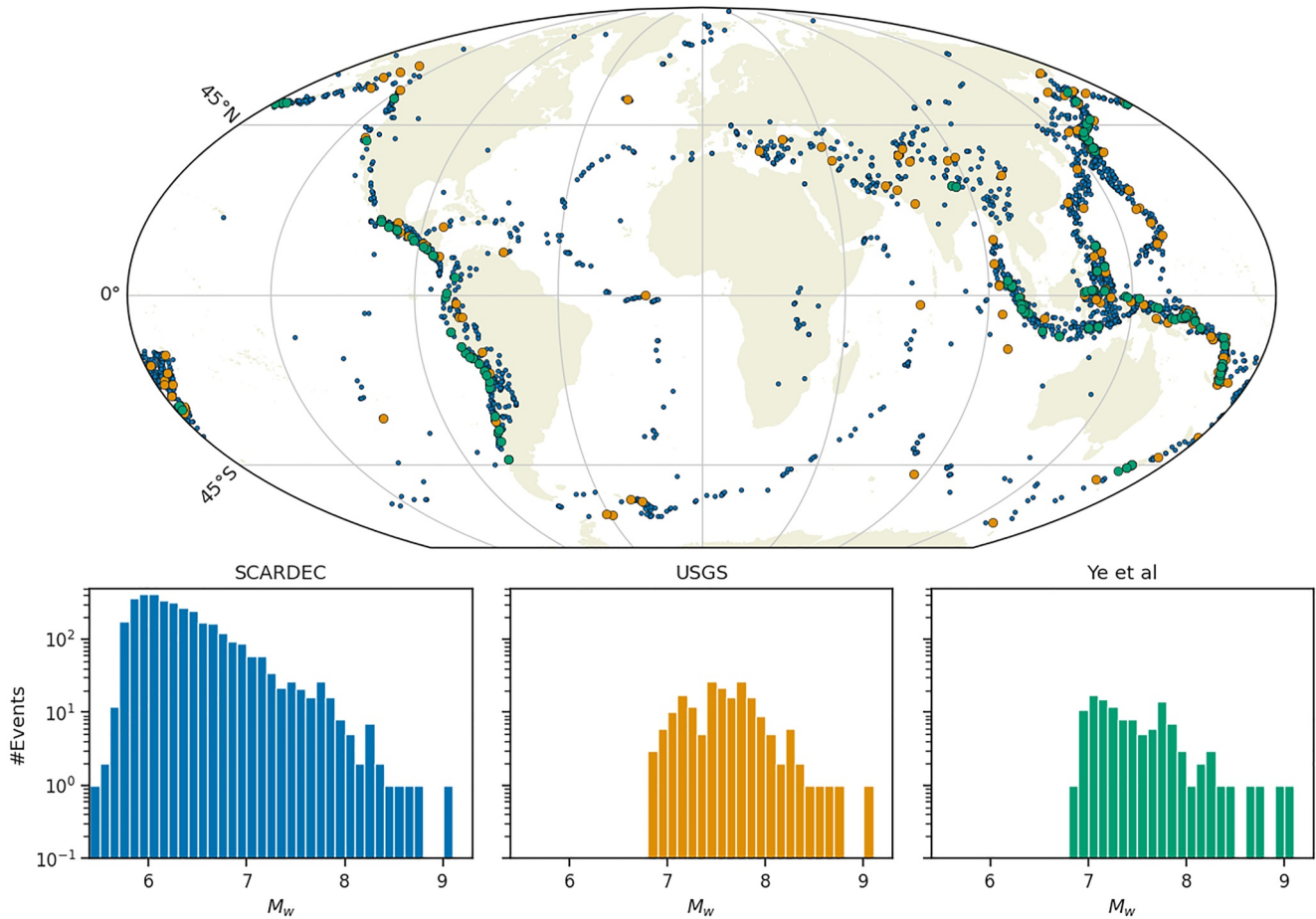


Figure 2. Distribution of events and histograms for magnitude distribution for the three STF datasets. The events are color coded by their data set. Ye et al. is plotted on top of USGS, on top of SCARDEC. This might lead to few events not being visible due to overlaps.

of the datasets. The three STF databases were generated using two different methodologies. SCARDEC uses a point source approximation and conducts a constrained deconvolution of body waves. In contrast, USGS and Ye et al. (2016) calculate finite fault solutions from body and surface waves using prior knowledge of the fault plane orientation. As the spatial extent of the source is modeled, these STF generally represent more high frequency details than the SCARDEC ones. On the other hand, the SCARDEC method can provide STF for smaller events than the finite-fault inversion schemes.

To infer the magnitude distributions from (partial) source time functions, we use a simple multi-layer perceptron. As input we use five observables derived from the source time function at time t : (a) cumulative moment M_t ; (b) current moment rate \dot{M}_t ; (c) average moment rate $\frac{1}{t}M_t$; (d) peak moment rate $\max_{\tau \leq t} \dot{M}_\tau$; (e) current moment acceleration \ddot{M}_t . We use features instead of full STF to avoid the danger of overfitting due to the high dimensionality of time series but low numbers of training examples. The selected features represent both current (b, e) and past (a, c, d) aspects of the STF. Furthermore, they are sufficient to approximate the observables in previous STF-based studies, for example, average moment acceleration (Melgar & Hayes, 2019), or characteristic time and early moment acceleration (Meier et al., 2017). We conducted experiments with an extended set of features (Figures S4 and S5 in Supporting Information S1) and obtained identical conclusions with slightly degraded performance. Therefore, in the following we only use the set of five features. We train the model on SCARDEC, as it is the largest of the datasets, using 10 fold cross validation. We use the continuous ranked probability score as loss, as its optimization behavior is more favorable in face of highly skewed underlying distributions than for log-likelihood. Further details on the model and training procedure are provided in Text S1 in Supporting Information S1.

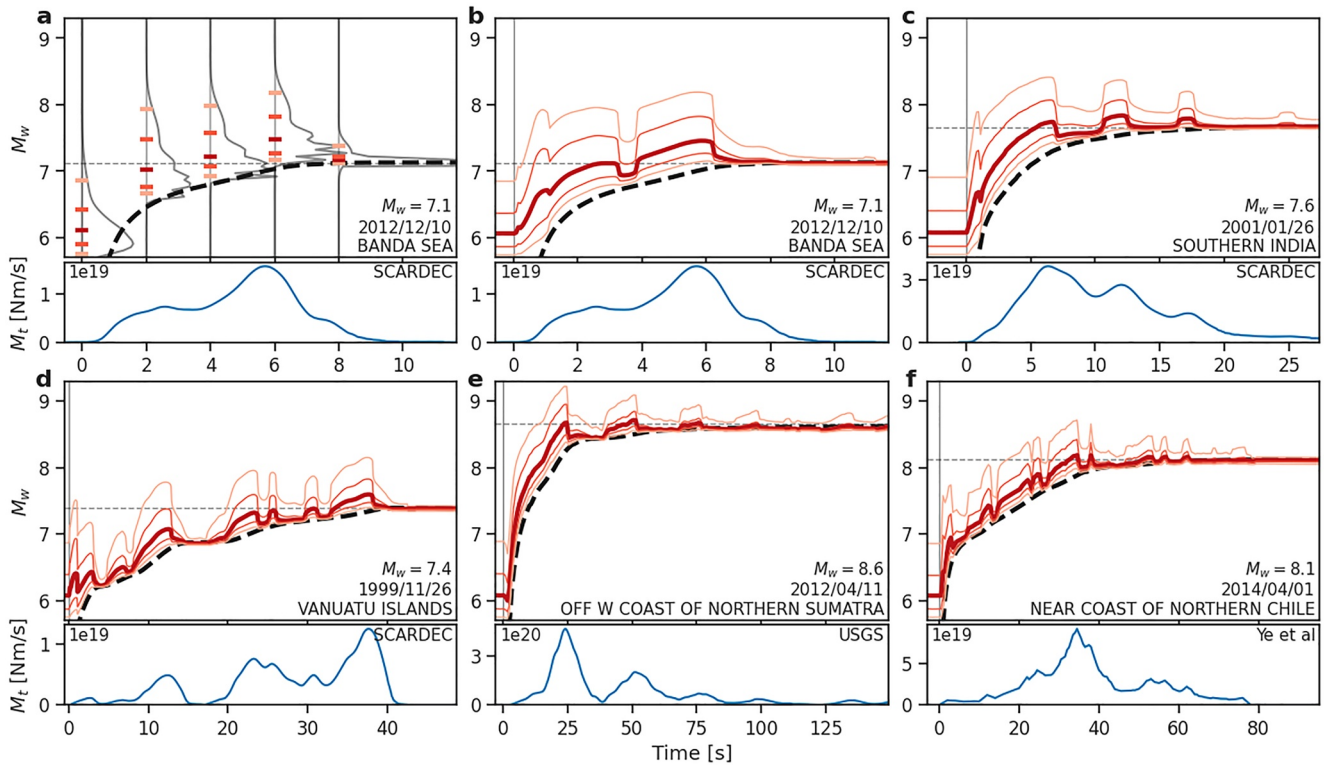


Figure 3. (a) Probability density functions (PDFs) calculated from the source time function (STF) model just before onset, and at 2, 4, 6 and 8 s after onset. Colored ticks on the PDFs indicate 0.05, 0.2, 0.5, 0.8, 0.95 quantiles. (b–f) Example predictions from the STF model visualized by the 0.05, 0.2, 0.5, 0.8, 0.95 quantiles over time. (b) shows the same event as (a) The lower right gives information on the event. The black dashed line shows the moment released so far, that is, the trivial lower bound. The bottom plots show the STFs. The upper right indicates the STF database used. For a step-by-step explanation on the quantile plots, see Figures 1a–1c.

For qualitative insights into the predictions and as a basis for interpreting the average results, we visualize representative examples (Figure 3). In all cases, the sign of the moment acceleration largely defines the anticipated potential for growth: positive acceleration, that is, the growth phase, indicates high growth potential, negative acceleration low potential. Furthermore, the higher the current moment release is, the higher the growth potential. This results from the STF's smoothness: at high moment rates it will likely take longer to arrest than at low rates. Notably, the model does not predict future asperities within a multiple asperities rupture (Figures 3d and 3e); for times after the peak of the moment rate function has been passed, the model expects a steady decay. Once the moment rate approaches zero, the estimated further growth is low (e.g., Figure 3d at 15 s, 3e at 40 s). If moment release accelerates again, the model immediately expects another asperity to break and higher growth potential is inferred yet again. These effects lead to sudden changes of the PDF at local maxima and minima of the STF (e.g., Figure 3d at 20 s, 3e at 25 s).

For a systematic analysis, we average $\mathbb{P}(M|O_t)$ by magnitude buckets (Figures 4a, 4c and 4e). For the datasets using finite fault solutions (Figures 4c and 4e), during the first 2 s of the STF, the predicted distributions are mostly identical across buckets. Afterward, the buckets split up over time: $M_w = 6.5$ –7.0 at ~ 2 s, $M_w = 7.0$ –7.5 at ~ 8 s, $M_w = 7.5$ –8.0 at ~ 16 s, $M_w = 8.0$ –8.5 at 25–40 s. These times match typical half-durations of events in these magnitude ranges (Gomberg et al., 2016).

SCARDEC (Figure 4a) shows similar splitting over time, but exhibits an apparent skew in the early predictions: higher magnitude buckets exhibit higher likelihood for becoming large. Similarly, the SCARDEC examples in Figures 3b–3d show high predictions within the first 2 s and abruptly fall afterward. We attribute this apparent predictability to artifacts of the SCARDEC processing, in particular a limited dynamic range. Due to the point source approach and separate deconvolutions at each station, the minimum resolvable moment release depends on the peak and cumulative moment release of an event (Vallée & Douet, 2016). This leads to a systematic bias in the first seconds of the STF (Figure S6 in Supporting Information S1), as previously reported by Meier et al. (2020). We confirmed that a similar apparent predictability can be introduced in the other two STF datasets by artificially

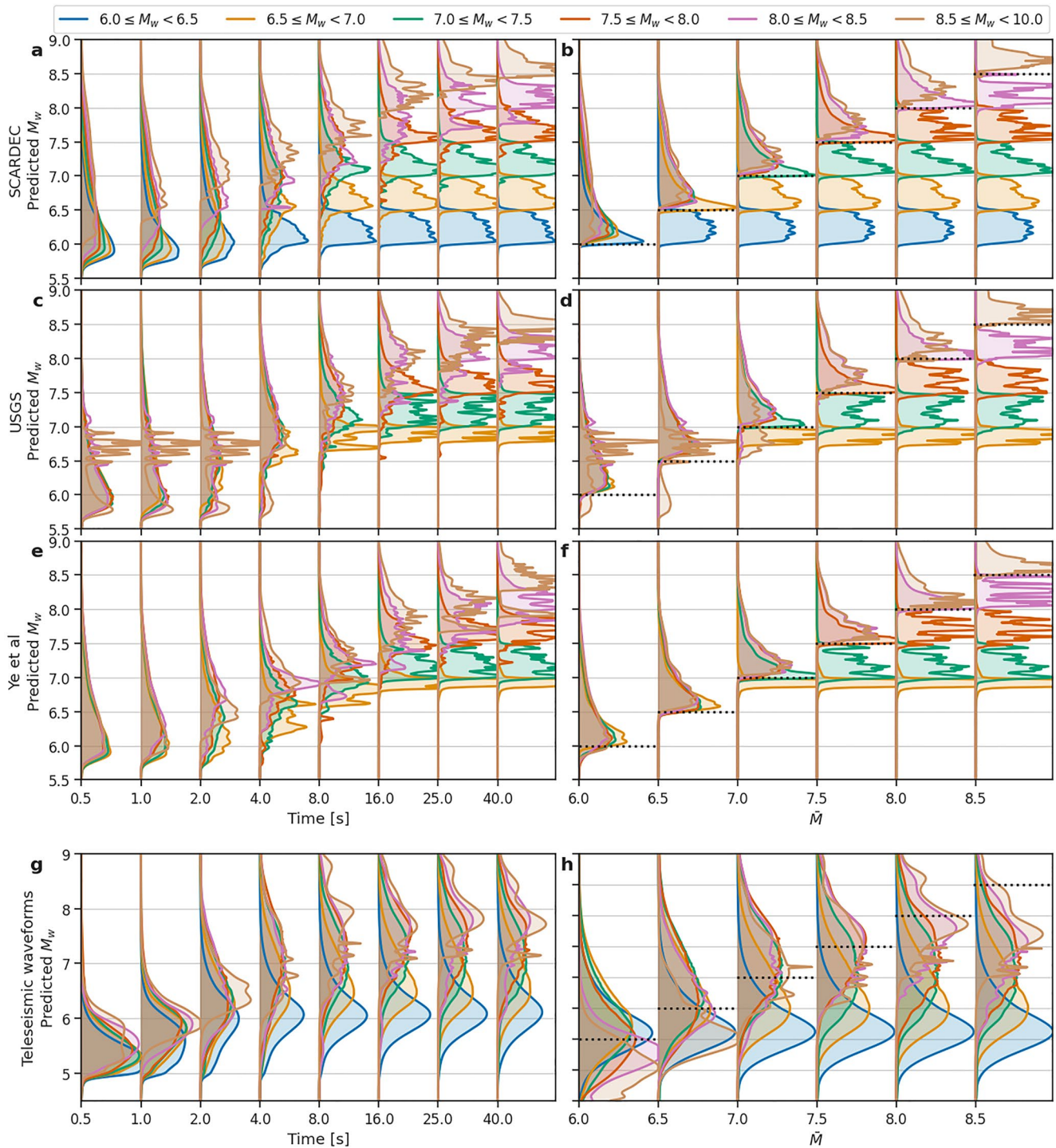


Figure 4. Average predicted Probability density functions (PDFs) based on source time function (STFs) (a)-(f) and teleseismic waveforms (g)-(h) grouped by magnitude bin. Left column shows results at time t after onset ($P(M|O_t)$), right column after cumulative moment equals magnitude \bar{M} ($P(M|O_{\bar{M}})$). The STF model has been trained on the SCARDEC data set and evaluated on each STF data set. See Figure S8 in Supporting Information S1 for STF results from a neural network trained with the USGS data set. PDFs were truncated in visualization to avoid overlap between different times/base magnitudes. Black dotted lines in (b, d, f, g) indicate the current base magnitude. The apparent skew between buckets in panel (b) and prediction difference in (g) for $\bar{M} = 6.0$ likely results from SCARDEC processing artifacts. For determining $t_{\bar{M}}$ in (h) we used the SCARDEC data set. See Figure S12 in Supporting Information S1 for plots with the other STF datasets. Events differ between panels (g) and (h): (h) only includes those events present in both the teleseismic data set and SCARDEC ($\sim 3,500$ events) and (g) all of the former ($\sim 38,000$ events).

limiting their dynamic range (Figure S7 in Supporting Information S1). Additionally, we trained models with the much smaller USGS data set without limiting dynamic range. These results lack the apparent early predictability observed for SCARDEC, supporting the explanation above (Figure S8 in Supporting Information S1).

Predictions at a fixed time t after onset describe both moment release until t and future development, with only the latter being relevant for predictability. To isolate this aspect, we define $\mathbb{P}(M|O_{\bar{M}}) = \mathbb{P}(M|O_t, \bar{M})$, where $t_{\bar{M}} = \sup_t \{M(t) \leq \bar{M}\}$ is the time when the cumulative moment release equals \bar{M} . When analyzing $\mathbb{P}(M|O_{\bar{M}})$, all three datasets exhibit the same trends (Figures 4b, 4d and 4f). All magnitude buckets with lower bounds at least $\bar{M} + 0.5$ show nearly identical predictions: a sharp increase in likelihood from \bar{M} to $\sim \bar{M} + 0.2$ and an exponential tail. $\bar{M} + 0.2$ represents roughly twice the seismic moment of \bar{M} and, due to the symmetry of STFs (Meier et al., 2017), half the event duration. For buckets with lower bound equal to \bar{M} , peak likelihood occurs around \bar{M} , again with exponential tails. The decay is steeper for these buckets, as most events are already past the peak and substantial future growth can therefore only result from future asperities, but not from further growth of the current one. The results are independent of the faulting mechanism (Figures S9, S10, S11 in Supporting Information S1). The systematic analysis therefore confirms the hypothesis that the final magnitude can only be assessed after the peak of the STF has been passed and that the rupture of further asperities cannot be anticipated.

4. Predictions From Teleseismic P Arrivals

STFs have limited temporal resolution, giving only a low-pass filtered view of the source process. Consequently, potential higher frequency details indicative of future rupture development might be hidden. To resolve this issue, we apply our approach to teleseismic P arrival waveforms. In contrast to STFs, these contain spectral information in a wider frequency band up to ~ 1 Hz, above which it will be largely attenuated. The lower limit of the spectral information is defined by the high-pass filter (in our case at 0.025 Hz) that is required due to instrumental limitations of seismic recordings (Goldberg et al., 2019; McGuire et al., 2021) and the limited time window used. We collated a data set of $\sim 35,000$ event with $\sim 750,000$ manually labeled first P arrivals. Our neural network model is an adaptation of TEAM-LM (Münchmeyer et al., 2021). TEAM-LM consists of a combination of convolutional layers, a transformer network, and a mixture density output and predicts the event magnitude directly from the seismic waveforms at a flexible set of input stations. Further details on the data set and model are provided in appendix 2.

Compared to the STF model, predictions $\mathbb{P}(M|O_t)$ and $\mathbb{P}(M|O_{\bar{M}})$ show higher uncertainties and systematic underestimation of the largest magnitudes at all times (Figure 4g). Higher uncertainties result from the fact that assessing magnitude from waveforms is harder than from STFs, whereas underestimation can be attributed to data sparsity (Münchmeyer et al., 2021). Due to the higher model uncertainties, all tails look rather like exponentially modified Gaussians than exponential distributions, as for the STF case. We note that the apparent lower uncertainty for the highest magnitude buckets compared to the lower magnitude buckets results from the number of samples in each bucket: with fewer samples available, the result gets less smooth but also less wide. Nonetheless, the general trends are highly similar to the results from the STF analysis before. Early predictions ($t \leq 2$ s) are indistinguishable, except for the bin $M_w = 6.0-6.5$, where event durations are often < 4 s. Bins split over time, similar to the STF model, although with higher overlap in predictions between bins. Splits occur around 4 s for $M_w = 6.5-7.0$, 8 s $M_w = 7.0-7.5$, and 16–25 s for both $M_w = 7.5-8.0$ and $M_w = 8.0-8.5$, again representing typical event half-durations.

As for $\mathbb{P}(M|O_t)$, predictions for $\mathbb{P}(M|O_{\bar{M}})$ exhibit similar behavior to the ones from the STF model (Figure 4h). While \bar{M} is considerably below the final magnitude, the predictions are indistinguishable between the buckets. Splitting of buckets occurs slightly later than for the STF model, that is, clear differences only become apparent once \bar{M} exceeds the upper bound of the bucket. This likely results from the higher uncertainties. We therefore argue, that assessment is likely still possible from the moment peak onward, again up to potential further asperities.

5. Comparison With Previous Results

Our results find no predictability from both STF and teleseismic waveforms. These observations seem contradictory to several previous results. We discuss potential reasons for some studies below. Melgar and Hayes (2019) found differences in moment acceleration for earthquakes of different size. As our STF model has access to the

acceleration parameter investigated in this study, we would expect to be able to reproduce this effect. However, later analysis demonstrated that these results were caused by a sampling bias, which our study confirms (Meier et al., 2020).

Danré et al. (2019) analyzed STFs as well, in this case by decomposing them into subevents, and also found predictability. Large events exhibited higher moment in early subevents and in addition showed higher complexity, that is, had more subevents. We suspect that their different conclusion might also result from the SCARDEC processing, which hides small subevents within large earthquakes and thereby makes the first identifiable subevent within a large event comparably larger.

Melgar and Hayes (2017) demonstrated from near-field geodetic observations of several great subduction earthquakes a pulse-like behavior, that is, the static displacement was seen to accumulate over a time period much shorter than the total rupture duration. They further analyzed slip pulse behavior and found a correlation between mean rise time and moment magnitude, making magnitude assessment possible after ~ 15 s. While this conclusion would contradict our results, the significance of the findings for events with $M_w > 7.5$ is unclear, given the low number of very large events and several intermediate events with long rise time. On the other hand, ~ 15 s does not imply any further predictability than found in our study for events with $M_w \leq 7.5$, due to their comparatively short duration. Furthermore, such a scaling in mean-rise time might as well be explained with a self-similar pulse model, for example, if the pulse grows over time (Nielsen & Madariaga, 2003). In this case, the mean rise time of a large event might be higher than of a small event solely because of later breaking segments giving no indication for predictability. We note that geodetic observations are candidates for studying potential pulse behavior of very large earthquakes, as they can resolve low frequency detail and static displacement (Goldberg et al., 2019; McGuire et al., 2021).

Colombelli et al. (2020) found differences in the slope of early peak ground motion parameters at local distances between earthquakes of magnitude 4–9. In our analysis of teleseismic waves, this effect could be hidden by the inherent attenuation of high frequencies in teleseismic waveforms. Therefore, our results do neither confirm nor contradict Colombelli et al. (2020). Similarly, while our study practically rules out predictability given STFs and teleseismic waveforms, it still leaves the option of approaches, where the tell-tale signals might only be observable in local waveforms, or require geodetic observations.

6. Conclusion

We conclude that there are no signs of early rupture predictability in either STFs or broadband teleseismic P waveforms. Instead, our analysis indicates that the total moment of an event based on such data can only be estimated after the peak moment release. However, even then it is not possible to anticipate future asperities.

While our analysis finds no early predictability, it highlights the feasibility of real time rupture tracking, at least using STFs and teleseismic waveforms. Transferring the methods to regional waveforms might therefore still significantly benefit early warning. Care has to be taken to counteract potentially biased estimations for the largest events, for which undersampling is hard to avoid (Münchmeyer et al., 2020).

Data Availability Statement

Moment rate datasets were obtained from the US Geological Survey (<https://earthquake.usgs.gov/data/finite-fault/>), Linling Ye (slip models are supplement to Ye et al. (2016) and linked in the acknowledgements therein), and the SCARDEC project (<http://scardec.projects.sismo.ipgp.fr/>). We downloaded manual phase picks from the ISC (International Seismological Centre, 2021) and USGS (U.S. Geological Survey, 2017). Seismic waveforms were downloaded from the IRIS and GEOFON data centers. We use waveforms from the GE (GEOFON Data Centre, 1993), G (Institut De Physique Du Globe De Paris (IPGP) & Ecole Et Observatoire Des Sciences De La Terre De Strasbourg (EOST), 1982), GT (Albuquerque Seismological Laboratory (ASL)/USGS, 1993), IC (Albuquerque Seismological Laboratory (ASL)/USGS, 1992), II (Scripps Institution Of Oceanography, 1986), and IU (Albuquerque Seismological Laboratory (ASL)/USGS, 1988) seismic networks.

Acknowledgments

Jannes Münchmeyer acknowledges the support of the Helmholtz Einstein International Berlin Research School in Data Science. We thank Martin Vallée for the insightful discussion regarding the apparent early predictability in the SCARDEC data set. We thank Jean-Paul Ampuero and an anonymous reviewer for their feedback that helped to improve the manuscript.

References

- Abercrombie, R. E., & Mori, J. (1994). Local observations of the onset of a large earthquake: 28 June 1992 Landers, California. *Bulletin of the Seismological Society of America*, 84(3), 725–734.
- Albuquerque Seismological Laboratory (ASL)/USGS. (1988). *Global seismograph network (gsn - iris/usgs)*. International Federation of Digital Seismograph Networks. Retrieved from <http://www.fdsn.org/doi/10.7914/SN/IU>
- Albuquerque Seismological Laboratory (ASL)/USGS. (1992). *New China digital seismograph network*. International Federation of Digital Seismograph Networks. Retrieved from <http://www.fdsn.org/doi/10.7914/SN/IC>
- Albuquerque Seismological Laboratory (ASL)/USGS. (1993). *Global telemetered seismograph network (usaf/usgs)*. International Federation of Digital Seismograph Networks. Retrieved from <http://www.fdsn.org/doi/10.7914/SN/GT>
- Allen, R. M., & Melgar, D. (2019). Earthquake early warning: Advances, scientific challenges, and societal needs. *Annual Review of Earth and Planetary Sciences*, 47(1), 361–388. <https://doi.org/10.1146/annurev-earth-053018-060457>
- Bengio, Y., Goodfellow, I., & Courville, A. (2017). *Deep learning* (Vol. 1). MIT press Massachusetts.
- Bishop, C. M. (1994). *Mixture density networks (Tech. Rep.)*. Aston University.
- Böse, M., & Heaton, T. H. (2010). Probabilistic prediction of rupture length, slip and seismic ground motions for an ongoing rupture: Implications for early warning for large earthquakes. *Geophysical Journal International*, 183(2), 1014–1030. <https://doi.org/10.1111/j.1365-246x.2010.04774.x>
- Colombelli, S., Festa, G., & Zollo, A. (2020). Early rupture signals predict the final earthquake size. *Geophysical Journal International*, 223(1), 692–706. <https://doi.org/10.1093/gji/ggaa343>
- Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, 2(4), 303–314. <https://doi.org/10.1007/bf02551274>
- Danrè, P., Yin, J., Lipovsky, B. P., & Denolle, M. A. (2019). Earthquakes within earthquakes: Patterns in rupture complexity. *Geophysical Research Letters*, 46(13), 7352–7360. <https://doi.org/10.1029/2019GL083093>
- Ekström, G., Nettles, M., & Dziewoński, A. (2012). The global CMT project 2004–2010: Centroid-moment tensors for 13, 017 earthquakes. *Physics of the Earth and Planetary Interiors*, 200–201, 1–9. <https://doi.org/10.1016/j.pepi.2012.04.002>
- Ellsworth, W. L., & Beroza, G. C. (1995). Seismic evidence for an earthquake nucleation phase. *Science*, 268(5212), 851–855. <https://doi.org/10.1126/science.268.5212.851>
- GEOFON Data Centre. (1993). *Geofon seismic network*. Deutsches GeoForschungsZentrum GFZ. <https://doi.org/10.14470/TR560404>
- Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477), 359–378. <https://doi.org/10.1198/01621450600001437>
- Goldberg, D., Melgar, D., & Bock, Y. (2019). Seismogeodetic p-wave amplitude: No evidence for strong determinism. *Geophysical Research Letters*, 46(20), 11118–11126. <https://doi.org/10.1029/2019gl083624>
- Gomberg, J., Wech, A., Creager, K., Obara, K., & Agnew, D. (2016). Reconsidering earthquake scaling. *Geophysical Research Letters*, 43(12), 6243–6251. <https://doi.org/10.1002/2016GL069967>
- Hayes, G. P. (2017). The finite, kinematic rupture properties of great-sized earthquakes since 1990. *Earth and Planetary Science Letters*, 468, 94–100. <https://doi.org/10.1016/j.epsl.2017.04.003>
- Ide, S. (2019). Frequent observations of identical onsets of large and small earthquakes. *Nature*, 573(7772), 112–116. <https://doi.org/10.1038/s41586-019-1508-5>
- Institut De Physique Du Globe De Paris (IPGP) Ecole Et Observatoire Des Sciences De La Terre De Strasbourg (EOST). (1982). *Geoscope, French global network of broad band seismic stations*. Institut de physique du globe de Paris (IPGP). <https://doi.org/10.18715/GEOSCOPE.G>
- International Seismological Centre. (2021). ISC bulletin. Author. <https://doi.org/10.31905/d808b830>
- Kilb, D., & Gomberg, J. (1999). The initial subevent of the 1994 Northridge, California, earthquake: Is earthquake size predictable? *Journal of Seismology*, 3(4), 409–420. <https://doi.org/10.1023/A:1009890329925>
- McGuire, J. J., Minson, S. E., Murray, J. R., & Brooks, B. A. (2021). Commentary: The role of geodetic algorithms for earthquake early warning in cascadia. *Geophysical Research Letters*, 48(6), e2020GL092324. <https://doi.org/10.1029/2020gl092324>
- Meier, M.-A., Ampuero, J.-P., Cochran, E., & Page, M. (2020). Apparent earthquake rupture predictability. *Geophysical Journal International*, 225(1), 657–663. <https://doi.org/10.1093/gji/ggaa610>
- Meier, M.-A., Ampuero, J. P., & Heaton, T. H. (2017). The hidden simplicity of subduction megathrust earthquakes. *Science*, 357(6357), 1277–1281. <https://doi.org/10.1126/science.aan5643>
- Melgar, D., & Hayes, G. P. (2017). Systematic observations of the slip pulse properties of large earthquake ruptures. *Geophysical Research Letters*, 44(19), 9691–9698. <https://doi.org/10.1002/2017GL074916>
- Melgar, D., & Hayes, G. P. (2019). Characterizing large earthquakes before rupture is complete. *Science Advances*, 5(5), eaav2032. <https://doi.org/10.1126/sciadv.aav2032>
- Mori, J., & Kanamori, H. (1996). Initial rupture of earthquakes in the 1995 Ridgecrest, California sequence. *Geophysical Research Letters*, 23(18), 2437–2440. <https://doi.org/10.1029/96GL02491>
- Münchmeyer, J., Bindi, D., Leser, U., & Tilmann, F. (2020). The transformer earthquake alerting model: A new versatile approach to earthquake early warning. *Geophysical Journal International*, 225(1), 646–656. <https://doi.org/10.1093/gji/ggaa609>
- Münchmeyer, J., Bindi, D., Leser, U., & Tilmann, F. (2021). Earthquake magnitude and location estimation from real time seismic waveforms with a transformer network. *Geophysical Journal International*, 226(2), 1086–1104. <https://doi.org/10.1093/gji/ggab139>
- Nielsen, S., & Madariaga, R. (2003). On the self-healing fracture mode. *Bulletin of the Seismological Society of America*, 93(6), 2375–2388. <https://doi.org/10.1785/0120020090>
- Noda, S., & Ellsworth, W. L. (2016). Scaling relation between earthquake magnitude and the departure time from P wave similar growth. *Geophysical Research Letters*, 43(17), 9053–9060. <https://doi.org/10.1002/2016GL070069>
- Olson, E. L., & Allen, R. M. (2005). The deterministic nature of earthquake rupture. *Nature*, 438(7065), 212–215. <https://doi.org/10.1038/nature04214>
- Scripps Institution of Oceanography. (1986). *Iris/ida seismic network*. International Federation of Digital Seismograph Networks. Retrieved from <http://www.fdsn.org/doi/10.7914/SN/II>
- Trugman, D. T., Page, M. T., Minson, S. E., & Cochran, E. S. (2019). Peak ground displacement saturates exactly when expected: Implications for Earthquake Early Warning. *Journal of Geophysical Research: Solid Earth*, 124(5), 4642–4653. <https://doi.org/10.1029/2018JB017093>
- U. S. Geological Survey. (2017). Advanced national seismic system (anss) comprehensive catalog of earthquake events and products. <https://doi.org/10.5066/F7MS3QZH>
- Vallée, M., & Douet, V. (2016). A new database of source time functions (STFs) extracted from the SCARDEC method. *Physics of the Earth and Planetary Interiors*, 257, 149–157. <https://doi.org/10.1016/j.pepi.2016.05.012>

- Ye, L., Lay, T., Kanamori, H., & Rivera, L. (2016). Rupture characteristics of major and great ($M_w \geq 7.0$) megathrust earthquakes from 1990 to 2015: 1. Source parameter scaling relationships. *Journal of Geophysical Research: Solid Earth*, *121*(2), 826–844. <https://doi.org/10.1002/2015JB012426>
- Zollo, A., Lancieri, M., & Nielsen, S. (2006). Earthquake magnitude estimation from peak amplitudes of very early seismic signals on strong motion records. *Geophysical Research Letters*, *33*(23), L23312. <https://doi.org/10.1029/2006GL027795>