UNIVERSITY OF CALIFORNIA
SANTA CRUZ

FORECASTING LANDSCAPES AND EARTHQUAKES

A dissertation submitted in partial satisfaction of the
requirements for the degree of

DOCTOR OF PHILOSOPHY

in

EARTH AND PLANETARY SCIENCE

by

Kelian Dascher-Cousineau

September 2022

The Dissertation of Kelian Dascher-Cousineau
is approved:

Professor Emily E. Brodsky, Chair

Professor Thorne Lay

Professor Noah Finnegan

Peter Biehl
Vice Provost and Dean of Graduate Studies

Contents

List of Figures

List of Tables

Abstract

Forecasting landscapes and earthquakes

by

Kelian Dascher-Cousineau

Successive earthquakes can drive landscape evolution. However, the mechanism and pace with which landscapes respond remain poorly understood. Offset channels in the Carrizo Plain capture the fluvial response to lateral slip on the San Andreas Fault on millennial timescales. In Chapter 1, we developed and tested a model that quantifies competition between fault slip, which elongates channels, and aggradation, which causes channel infilling and, ultimately, abandonment. Validation of this model supports a transport-limited fluvial response and implies that measurements derived from present-day channel geometry are sufficient to quantify the rate of bedload transport relative to slip rate. Extension of the model identifies the threshold for which persistent change in transport capacity, obliquity in slip, or advected topography results in reorganization of the drainage network.

Chapters 2, 3 and 4, shift focus to earthquake statistics. Earthquakes follow well-known and remarkably robust empirical laws. The intensity of aftershocks after a large earthquake rapidly decreases with time according to Omori's law, wherein the total abundance of aftershocks depends strongly on the magnitude

of the mainshock. The relative abundance of large and small events is commonly modelled using the Gutenberg-Richter relationship. These empirical laws provide a first-order order description of seismicity and underlie current operational earthquake forecasts. However, very large fluctuations from these statistical models suggests non-stationarity in time and space. Are these random, or can these fluctuations be explained? Chapters 2, 3 and 4 aim to diagnose the causes of this variability and potential tools to better characterize these features in the context of earthquake forecasting.

In chapter 2, we examine aftershock productivity relative to the global average for all mainshocks ($M_W > 6.5$) from 1990 to 2019. A global map of earthquake productivity highlights the influence of tectonic regimes. Earthquake depth, lithosphere age and plate boundary type correspond well with earthquake productivity. We investigate the role of mainshock attributes by compiling source dimensions, radiated seismic energy, stress drop, and a measure of slip heterogeneity based on finite fault source inversions for the largest earthquakes from 1990 to 2017. On an individual basis, stress drop, normalized rupture width, and aspect ratio most strongly correlate with aftershock productivity. A multivariate analysis shows that a particular set of parameters (dip, lithospheric age and normalized rupture area) combines well to improve predictions of aftershock productivity on a cross-validated data set. Our overall analysis is consistent with a model in which the

volumetric abundance of nearby stressed faults controls the aftershock productivity rather than variations in source stress. Thus, we suggest a complementary approach to aftershock forecasts based on geological and rupture properties rather than local calibration alone.

Recognizing earthquakes as foreshocks in real-time would provide a valuable forecasting capability. In a recent study, Gulia and Wiemer (2019) proposed a traffic-light system that relies on abrupt changes in b-values relative to background values. The approach utilizes high-resolution earthquake catalogs to monitor localized regions around the largest events and distinguish foreshock sequences (reduced b-values) from aftershock sequences (increased b-values). In Chapter 3, we utilize the recent well-recorded earthquake foreshock sequences in Ridgecrest, California, and Maria Antonia, Puerto Rico, as an opportunity to test the procedure. For Ridgecrest, a b-value time series may have indicated an elevated risk of a larger impending earthquake during the M_W6.4 foreshock sequence and provided an ambiguous identification of the onset of the M_W7.1 aftershock sequence. The exact result depends strongly on expert judgment. Monte Carlo sampling across a range of reasonable decisions most often results in ambiguous warning levels. In the case of the Puerto Rico sequence, we record significant drops in b-value prior to and following the largest event (M_W6.4) in the sequence. The b-value has still not returned to background levels (12 February 2020). The Ridgecrest

sequence roughly conforms to expectations; the Puerto Rico sequence will only do so if a larger event occurs in the future with an ensuing b-value increase. Any real-time implementation of this approach will require dense instrumentation, consistent (versioned) low completeness catalogs, well-calibrated maps of regionalized background b-values, systematic real-time catalog production, and robust decision-making about the event source volumes to analyze.

Seismology is witnessing an explosive growth in the diversity and scale of earthquake catalogs owing to improved seismic networks and increasingly automated data augmentation techniques. A key assumption in this community effort is that more detailed observations should translate into improved earthquake forecasts. Current operational earthquake forecasts build on seminal work designed for sparse earthquake records. Advances in the past decades have mainly focused on the regionalization of the models, the recognition of catalog peculiarities and the extension to spatial forecasts; but have failed to leverage the wealth of new geophysical data. Here, we develop a neural-network based earthquake forecasting model that leverages the new data in an adaptable forecasting framework: the Recurrent Earthquake foreCAST (RECAST). We benchmark temporal forecasts generated by RECAST against the widely used Epidemic Type Aftershock Sequence (ETAS) model using both synthetic and observed earthquake catalogs. We consistently find improved model fit and forecast accuracy for Southern Cali-

fornia earthquake catalogs with more than 10,000 events. The approach provides a flexible and scalable path forward to incorporate additional data into earthquake forecasts.

Acknowledgments

This dissertation includes content from the following published materials:

Chapter 1: Dascher-Cousineau, K., Finnegan, N. J., and Brodsky, E. E. (2021). The life span of fault-crossing channels. *Science*, 373(6551), 204-207.

Chapter 2: Dascher-Cousineau, K., Brodsky, E. E., Lay, T., and Goebel, T. H. (2020). What controls variations in aftershock productivity? *Journal of Geophysical Research: Solid Earth*, 125(2), e2019JB018111.

Chapter 3: Dascher-Cousineau, K., Lay, T., and Brodsky, E. E. (2020). Two foreshock sequences post Gulia and Wiemer (2019). *Seismological Research Letters*, 91(5), 2843-2850.

Dascher-Cousineau, K., Lay, T., and Brodsky, E. E. (2021). Reply to "Comment on 'Two Foreshock Sequences Post Gulia and Wiemer (2019)' by Kelian Dascher-Cousineau, Thorne Lay, and Emily E. Brodsky" by Laura Gulia and Stefan Wiemer. *Seismological Society of America*, 92(5), 3259-3264.

Funding from NASA, NSERC, NSF and SCEC gave me the freedom to explore a range of topics without strict deliverables.

I would like to acknowledge the support of my advisors Emily Brodsky, Noah Finnegan and Thorne Lay.

Emily Brodsky deserves particular credit for allowing me to follow my scientific inquiries where ever they have led me, and being supportive throughout the process. I was encouraged to pursue science out of a genuine passion for the craft. I am quite confident that neither of us would have thought that three years into my Ph.D. I'd be working from my garden shed studying the intricacies of river bedload transport. I genuinely do not think that I could have found a better advisor. Its been a pleasure. I would like to thank Noah Finnegan for turning a short conversation about faults in the landscapes into a research product that I am truly proud of. Throughout this process, I was encouraged to let the science speak for itself. I hope I can hold on this clarity moving forward. Thank you to Thorne Lay for his seemingly endless knowledge of seismology and sharing it to the next generation of scientists. I admire his capacity to identify and amplify the scientific insight in his interactions.

Nearly every morning of my PhD, I have attended Seismo-Coffee, discussing the latest earthquake in Alaska, the seismic signals of nuclear testing in North Korea, or the subtle tilt detectable on a seismometer in the wake of 100m high waves propagating in the internal stratification of the ocean. Much of the discussion focuses on the fundamentals of earthquake science and the pressing research problems of the field. In this hour, I have tapped the pool of experience at the UCSC seismological laboratory. I am so thankful to everyone involved in Seismo-

coffee over the years. Grace Barcheck, Stephanie Taylor, Thomas Goebel, Shalev Siman Tov, Nadav Wetzler, Luca Maletesta, Travis Alongi, Heather Crume, Ricky Garza-Giron, Em Schnorr, Kristina Okamoto, Esteban Chaves, Huiyun Guo, Litong Huang, Alex Watson, Will Steinhardt, Chris Thom, Valere Lambert, Susan Schwartz, and Heather Savage... Spending time with this amazing group has been the highlight of my Ph.D.. Their enthusiasm and curiosity about science is contagious.

Thank you to my friends, sounding boards, housemates, science editors, adventure buddies, and second family at the Segre Place: Calvin Munson, Coby Abrahams, Colleen Murphy, and Zack Kaufmann. I have been lucky to live amongst an amazing group people in these last five years. I can't thank you enough.

A very special thanks goes to Alba Rodriguez Padilla for being my number one supporter through thick and thin. Finally, I would like to acknowledge my amazing family for endless support and care to get me up to, and through this experience.

Chapter 1

The lifespan of fault-crossing channels

1.1 Introduction

The recognition in 1908 that ephemeral streams in the Carrizo Plain, California, preserved a passive record of lateral fault offset transformed the study of active strike-slip faults (Lawson 1908; Sieh 1978; Sieh and Jahns 1984; Wallace 1968; Zielke et al. 2010). However, offset on these stream-channels generally does not exceed tens to hundreds of meters. Flow eventually overtops channels, typically spilling straight across the fault and resetting the recorded offset (Fig. 1.1A-C) (Sieh 1978; Wallace 1968; Zielke, Klinger, and Arrowsmith 2015). Variability in

channel offsets suggests a reset mechanism that depends primarily on local channel configuration rather than regional climate or earthquake history. Slip rates on the order of centimeters per year on the San Andreas Fault imply that channels reset on millennial timescales, thereby limiting their utility as recorders of fault slip (Wallace 1968). We instead leveraged this tectonic-geomorphic interaction to better understand how the drainage network responds to perturbation. We develop a quantitative prediction for the channel lifespan as a function of fault slip-rate, present-day channel geometry, and sediment transport capacity assuming transport-limited conditions. We validated our model by testing it against a lidar-derived record of fault-crossing channels in the Carrizo Plain and explore implications for the long-term evolution of strike-slip fault landscapes.

1.2 Theoretical foundation

Wallace (Wallace 1968) identified channel sedimentation as a control on the evolution of fault crossing streams (Ouchi 2005; Sims 1994). He noted that offset increases channel length without changing the drop in elevation (Fig. 1.1A), thus reducing channel gradients in fault crossing reaches. Where a channel enters a fault zone, the shallowing slope associated with a reduction in sediment transport causes deposition on the channel floor (Leopold and Bull 1979). Wherever the combined depth of aggrading sediment and water approaches the height of local

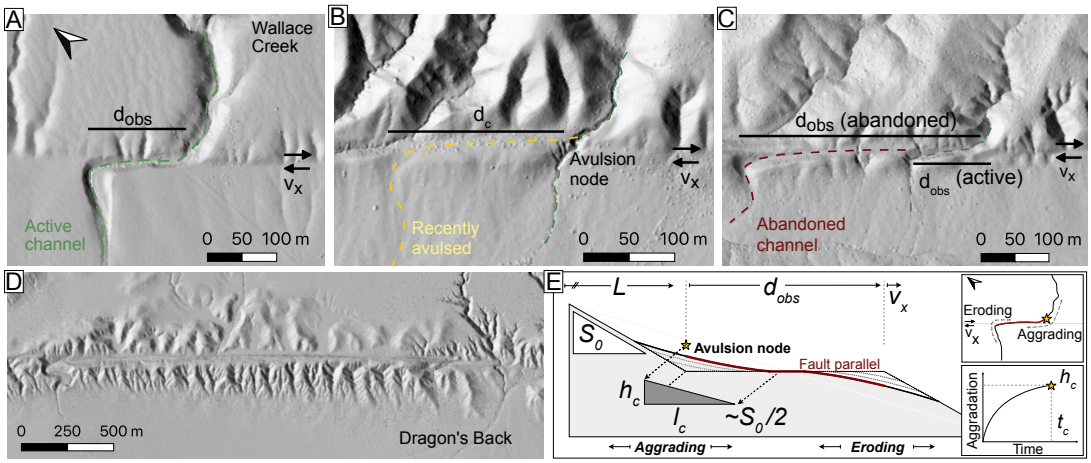

Figure 1.1: The lifespan of fault crossing channels. Hillshades (Bevis et al. 2017) illustrates, in progressive stages, (A) an offset channel, (B) a recently avulsed, or reset, channel at the critical offset, d_c, and (C) a subsequently offset channel. (D) Large fault parallel drainage along the advection path from the Dragon's Back. In this case, the penultimate avulsion caused a pulse of incision due to the resulting steeper gradient. Locations for A-D are shown in Fig. A.4. (E) Idealized elevation profile of an offset-channel (shown in plan-view in the upper inset) with a reach, L, an offset, d_{obs}, undergoing horizontal fault slip, v_x. Offset introduces a fault parallel segment with near-horizontal slope (dashed lines); the channel responds evolving with time up until its current geometry (bold line). Avulsion occurs at time t_c provided aggradation approaches the threshold height, h_c (see lower inset), which can be related to the upstream aggradation length scale, l_c, and the initial slope, S_0 (grey triangle).

topographic barriers, flow is susceptible to diversion in a sudden abandonment of the channel, known as an avulsion (Slingerland and Smith 2004). Flow then incises a new channel in the steepest descent direction, typically orthogonal to the fault trace. In this way, avulsion resets the offset and orientation of fault crossing channels (Fig. 1.1B) (Wallace 1968; Sims 1994). The upstream bend locally incurs the most pronounced change in channel slope and therefore marks an avulsion node, a persistent point for subsequent avulsions to occur (Fig. 1.1C), much in the same way that the backwater length scale in lowland coastal rivers controls the location of geologically persistent flow deceleration, sedimentation and therefore chronic avulsion in deltas (Ganti et al. 2016). Trenching and geochronology at Wallace and Phelan Creeks corroborate repeated cycles of aggradation, avulsion, and incision (Sieh and Jahns 1984; Sims 1994).

The importance of sedimentation in the lifespan of fault crossing channels suggests that bedload transport, rather than the stream's ability to erode the bed, limits their long-term evolution. In a transport-limited channel, the rate of aggradation, v_a [L/T] implicitly determines the time required for a channel to avulse, t_c [T] (Jerolmack and Mohrig 2007),

$$h_c = \int_0^{t_c} v_a(t)dt \tag{1.1}$$

where the avulsion threshold $h_c(t)$ is the net height of aggraded sediment required for the channel to overflow its container, which may be either a channel or a val-

ley. Eq. 1.1 evaluates the avulsion threshold at the time of avulsion, t_c; prior to this point in time, aggradation in active channels can never exceed the threshold. Throughout, the value of $h_c(t)$ may vary due to dip-slip along the fault, or advection of topographic features such as shutter ridges, depressions or abandoned channel heads. We first expand Eq. 1.1 with particular consideration to aggradation at the avulsion node and examine how the present-day geometry constrains and validates it, and then explore the influence of the avulsion threshold's time dependence.

If we assume bedload transport capacity to be linearly related to channel slope, then erosion or aggradation [L/T] is proportional to a change in slope in the channel profile (Paola, Heller, and Angevine 1992),

$$v_a = -\kappa \frac{\partial S}{\partial x}, \tag{1.2}$$

where x is the distance along the channel profile [L], S is the channel slope, and κ is a diffusivity coefficient [L^2/T]. The diffusivity coefficient represents the volumetric sediment transport capacity per unit channel width and slope. A channel conveying more sediment has a higher diffusivity coefficient and adjusts more rapidly to perturbations in the channel slope.

The channel elevation profile allows for useful approximations of the right-hand side of Eq. 1.2 (Fig. 1.1E). For a narrow fault zone, offset introduces a fault parallel segment with slope, S_f, much smaller than the original slope, S_0, such

that $\Delta S = S_f - S_0 \approx -S_0$ approximates the slope change without aggradation. In response, a wedge of sediment near the avulsion node grows as a symmetric diffusive pulse. We define its horizontal half-width at the time of avulsion as l_c and thus approximate the channel response leading up to the avulsion as $-\kappa\frac{\partial S}{\partial x} \approx \kappa\frac{S_0}{l_c}$

.

To relate Eq. 1.1 and 1.2, we cast l_c in terms of the initial channel slope, S_0, and the avulsion threshold height, h_c. Following aggradation to the threshold level, the rise, $S_0 l_c$ accommodates twice its original run, recovering half the original slope (Fig. 1.1E). This leads to the relationship $S_0/2 \approx h_c/l_c$ or, equivalently, $l_c \approx (2h_c)/S_0$, an approximation that is consistent with measurements of millennial-scale slope changes along alluvial rivers induced by check-dams (Leopold 1992) and holds when channel response is rate-limiting (Section A.31-2). Finally, by combining Eq. 1.1 and Eq. 1.2 and approximating the integral in Eq. 1.1 as a product, we obtained a characteristic avulsion time scale,

$$t_c \approx \frac{4h_c^2}{\kappa S_0^2}. \tag{1.3}$$

A full semi-analytical solution to Eq. 1.1 and 1.2 with discrete slip events (Section A.2) corroborates the approximate framework and suggests a response independent of whether slip is continuous or sequenced in individual earthquakes. Provided a long-term fault slip rate (v_x), a channel becomes unstable and susceptible to avulsion when the channel offset approaches a critical offset $d_c = v_x t_c$. In this

way, through a known slip-rate, the channel offset is an independent chronometer of the channel response. We can thus relate the advective and fluvial history of offset channels at the time of avulsion to determine a critical offset,

$$d_c \approx \frac{4h_c^2 v_x}{\kappa S_0^2}. \tag{1.4}$$

Eq. 1.4 is the first key result of the model. It defines the offset that resets fault-crossing channel and depends entirely on measurable quantities.

1.3 Analysis

We tested this relationship on a collection of active and abandoned fault-crossing channels in the Carrizo Plain. While it is rare to estimate d_c directly, Eq. 1.4 predicts observed offsets (d_{obs}) less than d_c on active channels and more than d_c on abandoned channels. We measured values of total offset (d_{obs}), avulsion threshold height (h_c), and the initial slope (S_0) interpreted from LiDAR data (e.g., Fig. 1.2A-C) for a set of 55 active and abandoned channels in the Carrizo Plain (Table A.1, Fig. A6-S64). We avoided channels where misalignment across the fault appears to result from the deflection of flow rather than progressive fault offset (e.g., downstream from an alluvial fan). We estimated the diffusivity coefficient using $\kappa \approx 0.1Lr$ where L is the corresponding catchment length (Fig. 1.2D) and r is the mean annual regional rainfall, recognizing that this approximation and

its constancy through time is the main source of error in our approach ((Paola, Heller, and Angevine 1992), Section A.4). We recorded partial overspilling on an active channel, or minimal offset on the active tributary of an abandoned channel as these respectively indicate an incipient or recent avulsion. In these cases, we assume that the measured offset approximates the critical offset, d_c.

We compared measured offsets to corresponding predictions of d_c (Fig. 1.2E). A logistic regression through the log-transformed data finds $d_{obs}/d_c = 0.6^{+0.5}_{-0.3}$ (99% confidence interval on 10000 bootstrap samples) best separates active and abandoned channels, indicating that the model prediction is within a factor of two of observed critical offsets (Fig. 1.3). If the timing of avulsions were independent of the diffusive channel response, the ratio d_{obs}/d_c would not separate active and abandoned channels (p = 0.0007, following a Wald test), and has no reason to approach unity. Instead, the scaling analysis predicts the lifespan of fault crossing channels well within model uncertainty, particularly in the parameterization of the diffusivity coefficient (Section A.4).

Under the assumptions of the model, any one parameter in Eq. 1.4 is constrained from the others. Posing Eq. 1.4 as an equality and isolating for the slip rate, the channel diffusivity coefficient (Section A.4) and channel geometry (Table A.1) on active and abandoned channels bracket a slip rate, respectively under- and over-estimating it. A logistic regression then yields a $2.1^{+1.7}_{-1.0}$ cm/yr slip rate in the

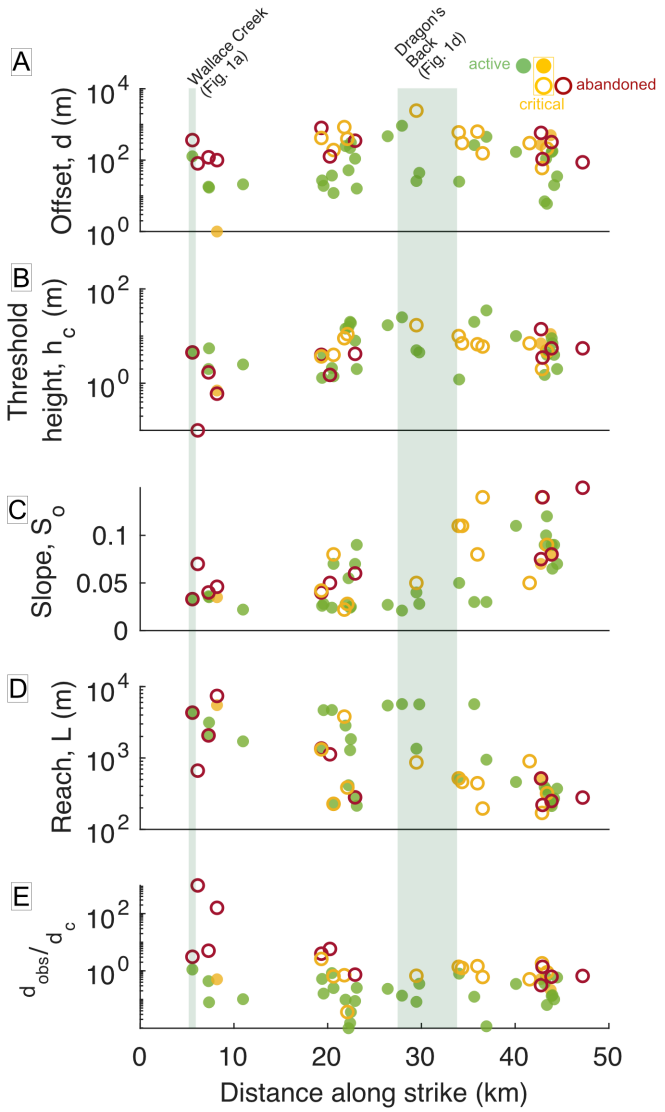

Figure 1.2: Channel geometry along the San Andreas Fault in the Carrizo Plain. Measurements of (A) offset (d_{obs}), (B) initial avulsion threshold height (h_c), (C) initial slope (S_0), (D) channel reach (L), and (E) normalized offset measurements (d_{obs}/d_c) (Table A.1). Solid markers indicated active channels; empty markers indicate abandoned channels. Yellow markers indicate channels that have evidence for incipient or recent avulsions. No individual set of measurements (A-D) appears sufficient to separate active and abandoned channels. Offsets in (A) and a slip rate of 3.5 cm/yr are consistent with a channel response over millennial timescales. The uncharacteristically large offsets in (A) collapse near unity in (E) when normalized by the critical offset.

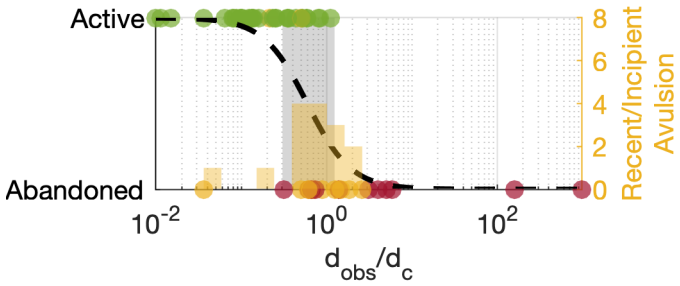

Figure 1.3: Channels classification. Logistic regression (dashed line) through active and abandoned channels (collapsing Fig. 1.2E along the x-axis) with respect to the normalized offset. Separation of the data, with a class boundary near unity, indicates consistency with the model framework. The grey box indicates the 99% confidence interval on the class boundary using 10000 bootstrap samples. Near critical channels (incipient or recent avulsion in yellow) are summarized in the histogram as an additional qualitative test

Carrizo Plain (99% confidence interval on 10000 bootstrap samples), compatible with geologic and geodetic estimates of 3.5 cm/yr (Schmalzle et al. 2006; Sieh and Jahns 1984). With a known fault slip rate, a similar approach calibrates the diffusivity (and hence sediment transport capacity) of channels. The critical offset encodes channel response best in arid environments where transport capacity is low relative to slip rate, perhaps a distinguishing feature of the Carrizo Plain (Section A.3). Where transport rates are high, the evolution of channels after individual earthquakes on centennial timescales may better capture the geomorphic process (Fig. A.3). Bedrock channels where sediment transport is not limiting may be better modeled as detachment-limited kinematic waves of erosion (Duvall and Tucker 2016; Gray et al. 2018; Reitman et al. 2019a) with fault damage promoting headward incision along the fault (Roy et al. 2016; Wallace 1968).

1.4 The role of topography and slip obliquity

Up to this point in our analysis, we utilized the avulsion threshold with respect to the present-day channel geometry, independent of consideration for its path and continued evolution. Expanding $h_c(t) = h_0 + \lambda v_x t + \cdots$. h_0 reflects initial channel height or incision immediately following an avulsion, with λ being the dimensionless ratio of apparent vertical to horizontal motion in the fault plane at the avulsion node.

Topographic gradients translated along strike and obliquity in slip (the ratio of strike-slip to dip-slip motion) contribute to λ. For the San Andreas Fault, obliquity in slip rarely exceeds 10% (Wallace 1990). For sharp topographic elements advected in front of the avulsion node, λ may be much higher (e.g. angle of repose $\tan 30 \approx 50\%$) over relatively short distances. The growth of the avulsion threshold, if sufficiently high, may outpace channel aggradation (Fig. 1.4). This results in persistent fault parallel drainage, analogous to a river diverted by active uplift (Humphrey and Konrad 2000). In such cases the penultimate avulsion may not occur at the avulsion node, where aggradation is highest, but rather downstream, where uplift wanes (e.g., Fig. A40-41). Unlike other fluvial environments susceptible to frequent avulsion such as deltas or fans (Slingerland and Smith 2004), aggradation in offset channels is effectively unbounded provided sustained growth of $h_c(t)$. Sedimentation can therefore far exceed flow depths.

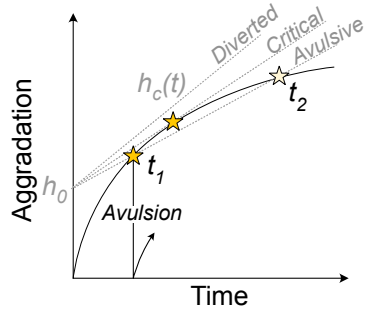

Figure 1.4: Channel aggradation and relative vertical motion along the fault. Aggradation at the avulsion node as a function of time for a channel with an avulsion threshold, $h_c(t)$, growing with time. Times t_1 and t_2 represent the two solutions to the quadratic form of the avulsion time scale, however, only time t_1 is realized as the system resets once avulsion occurs. In detail, aggradation exhibits jagged high frequency fluctuations representing cut-in-fill behavior but should be well approximated as a smooth diffusive curve on millennial timescales.

Introducing the time-dependent avulsion threshold into Eq. 1.3 yields a limit, λ_c, in which fluctuation in h_c outpaces aggradation,

$$\lambda_c \approx \frac{\kappa S_0^2}{16 v_x h_0}. \tag{1.5}$$

If $\lambda > \lambda_c$, channels are diverted along strike (Section A.5, Fig. 1.4). For the channels we studied, λ_c (estimated from Eq. 1.5) is typically on the order of a few percent and, therefore, sensitive to near-fault topography, which may contribute to the ubiquity of fault parallel drainage on the San Andreas Fault. Where near fault relief is high, the timing of avulsion is modulated by the valley spacing or abandoned channels that are advected along the fault (Harbert, Duvall, and Tucker 2018; Duvall and Tucker 2016; Reitman et al. 2019b). In the Carrizo Plain subdued near-fault topography and nearly horizontal slip implies that the ratio of

vertical to horizontal motion rarely exceeds the critical λ_c and results in abundant cross-fault drainage. Along the Dragon's Back, however, $\lambda \approx 6\%$ over a kilometer of the fault (Hilley and Arrowsmith 2008) is consistent with uncharacteristically long channel-offset and suppressed cross-fault drainage (Fig. 1.1D, A40-41). The case of $\lambda < 0$ (down dropping) results in an unstable channel geometry, which is likely to develop alluvial fans where channels reset in high flow events (Grant and Sieh 1994; Zielke et al. 2010). For large rivers, the numerator κS_0^2 of Eq. 1.5 is roughly two orders of magnitude larger than the ephemeral channels of the Carrizo Plain (Castelltort and Van Den Driessche 2003), suggesting that established fault-crossing rivers are less susceptible to diversion. Therefore, major fault parallel drainages along strike-slip faults may preferentially develop from the coalescence of smaller tributaries.

1.5 Conclusion

The geometry of offset channels is the product of an interactive fluvial-tectonic system. In the Carrizo Plain, channels offset by the San Andreas Fault provide an opportunity to characterize the long-term evolution of channels under a simple forcing history. In these reaches, the lifespan of fault-crossing channels is well described by a transport-limited response culminating in avulsions. The configuration of fault-crossing channels, readily obtained from topography data, quantifi-

ably constrains the relative pace of the tectonic and fluvial systems. Our analysis implies that the geomorphology of strike-slip fault zones exists in a careful balance where subtle changes to the system can toggle drainage from fault perpendicular to fault parallel valleys.

Chapter 2

What Control's Variations in Aftershock Productivity?

2.1 Introduction

Earthquakes cluster in time and space. In a typical sequence, the largest earthquake is the mainshock, those preceding are foreshocks, and those following are aftershocks (Omori 1895). The clustering behavior is well-described by three statistical relationships characterizing the temporal- and size-distributions of earthquakes (Ogata 1988). In this study, we focus on the aftershock productivity law:

$$N(M) = k10^{\alpha M} \tag{2.1}$$

where N is the number of aftershocks, M is the mainshock magnitude, and k is a constant of proportionality which depends on the number of aftershocks per mainshock above the catalog completeness level (Reasenberg and Jones 1989). In practice, the aftershocks are mingled with background events and a major challenge of any implementation is to accurately separate these signals in the face of regionally variable activity levels and detection thresholds. Previous efforts to isolate aftershocks have capitalized on the intrinsic clustering of earthquakes to suppress contamination and inferred that the productivity law fits a wide range of data with $\alpha \approx 1$ (Reasenberg and Jones 1989; Yamanaka and Shimazaki 1990; Arcangelis et al. 2016; Kisslinger 1996; Tahir and Grasso 2015; Page et al. 2016).

Eq. 2.1 empirically quantifies the basic phenomenon of an increasing number of aftershocks with magnitude. However, Fig. 2.1a shows 100-fold differences in the number of aftershocks for events of similar magnitude that far exceed 95% confidence bounds predicted from Poisson distribution given the rate predicted by Eq. 2.1. The particular counts reported in Fig. 1 are based on a time-space window counting scheme that will be discussed more fully below, but the variability of the aftershock rate is not unique to this work. Such excursions from the scaling relationship are well documented by prior work that used a variety of methods

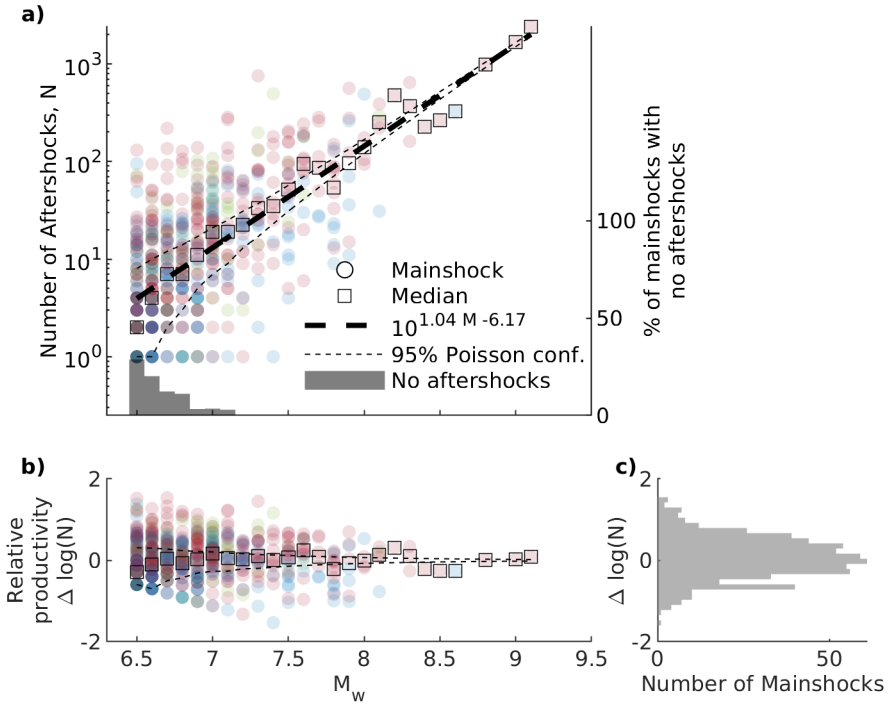

Figure 2.1: a) The number of aftershocks of $M_W \geq 4.5$ within three source dimensions and 60 days as a function of mainshock magnitude identified in the global ISC and NEIC catalogs from 1990 to 2019. Colors indicate faulting style of the mainshock; blue, green and red points correspond to earthquake sequences for which the mainshock was respectively strike-slip, normal or reverse. The global productivity law (dashed line) is fit using a least squares regression through the median log-number of aftershocks for each 0.1 magnitude bin (black squares). The median number includes mainshocks with no aftershocks which are not shown on the plot. Light dashed lines indicate 95% confidences bounds given a Poisson distribution. Individual earthquake sequences (circles) scatter significantly beyond these confidence limits. b) Relative productivity as a function of mainshock magnitude. The relative productivity distribution does not show events with no aftershocks and thus the lower left corner of the plot is underpopulated. c) Histogram of the relative productivity of mainshocks considered in this study.

(Marsan and Helmstetter 2017; Boettcher and Jordan 2004; Page et al. 2016; Tahir and Grasso 2014). In this work we investigate whether these variations are determined by features of particular sites or of the earthquakes themselves.

The answer has practical consequences. Aftershock forecasting is currently the only operational form of earthquake prediction and is routinely used in the wake of major earthquakes to advise on short-term hazard following an earthquake (Reasenberg and Jones 1989; Page et al. 2016; Hardebeck et al. 2018). Within a sequence, the probability of subsequent earthquakes exceeding the magnitude of the mainshock relates directly to the productivity (Reasenberg and Jones 1989; Reasenberg 1999). Currently, variability in aftershock productivity is calibrated regionally where data permit or extrapolated from nearby regions and updated over the course of the aftershock sequence (Reasenberg and Jones 1989; Reasenberg 1999; Ogata 2017). Many of these forecasts are empirical statistic models and, notably, rely heavily on high quality seismic data both before and during the mainshock (Gerstenberger et al. 2005; Omi et al. 2015). Other models embed physics-based rules that rely on local information about long-term seismicity, the state of stress and the geometry of the faults (Segou and Parsons 2016; Field et al. 2017). All of these approaches raise the critical issue of calibrating hazard in poorly instrumented areas.

Yet previous studies have found that features related to the setting of the

mainshock can inform aftershock abundance. Both global and regional studies indicate that tectonic regions have distinct aftershocks statistics (Chu et al. 2011; Page et al. 2016; Davidsen, Gu, and Baiesi 2015; Tahir, Grasso, and Amorse 2012; Ogata 2017). For instance, the Eastern Pacific has greater aftershock productivity than the Western Pacific (Singh and Suárez 1988; Wetzler, Brodsky, and Lay 2016). Seismotectonic subdivisions (plate boundaries, global geology, seismicity catalogs, and regional and local studies) yield 10-fold differences in aftershock productivity (Page et al. 2016). Active non-subduction continental regions have elevated earthquake productivity and, on average, larger aftershocks (Page et al. 2016; Mogi 1967; Davis and Frohlich 1991). In contrast, ridge transform faults are deficient in aftershocks (Davis and Frohlich 1991; Boettcher and Jordan 2004; McGuire, Boettcher, and Jordan 2005). The local geological structure is thought to underlie these distinctions (Boettcher and Jordan 2004; McCloskey et al. 2003). Yamanaka and Shimazaki 1990 report intraplate earthquakes as more productive than plate boundary earthquakes. Zaliapin and Ben-Zion 2016 find similar geographic patterns in clustering statistics and relate them to global heat flow. Case studies generally reinforce the importance of geological structures on the distribution and intensity of aftershocks (Das and Henry 2003; McCloskey et al. 2003). An extreme case comes from deep focus subduction zone earthquakes, which sometimes generate few or no observable aftershocks (Båth 1965; Frohlich

1989; Nyffenegger and Frohlich 2000; Wiens et al. 1997; Wu and Chen 1999; Persh and Houston 2004). Some work attributes the deficiency to the elevated temperature at depth (Nyffenegger and Frohlich 2000; Persh and Houston 2004).

Other studies focus on the importance of source effects. Theoretical arguments, supported by systematics in the variance of stress drop measurements and earthquake productivity in California, have suggested that increased stress drop should correspond to increased productivity (Marsan and Helmstetter 2017). The opposite relationship was documented for recent (1990-2016) major megathrust ruptures ($M_W \geq 7.0$) (Wetzler, Brodsky, and Lay 2016), suggesting that high stress drop corresponds to a smaller rupture area and therefore fewer aftershocks. This is supported by a tendency for megathrust aftershocks to occur on the periphery of large-slip zones (Wetzler, Brodsky, and Lay 2016; Van Der Elst and Shaw 2015). Earthquakes rupturing at supershear velocities also appear to have low aftershock productivity (Bouchon and Karabulut 2008). A relationship between the heterogeneity of a rupture and the number of aftershocks has been a long-standing contention (Mogi 1967; Yamanaka and Shimazaki 1990) with some support from rate and state models (Helmstetter and Shaw 2006; Marsan 2006), but few direct and quantitative measurements (Das and Henry 2003; Persh and Houston 2004). Finally, occurrence of dynamic triggering would suggest that radiated energy should influence the number of aftershocks (Felzer and Brodsky

2006).

Some observations cannot be distinctly associated with setting or source. There is a relationship between relative aftershock productivity and focal mechanism. Strike-slip earthquakes are proposed to be intrinsically less productive than dip-slip earthquakes (Tahir, Grasso, and Amorse 2012; Tahir and Grasso 2014; Tahir and Grasso 2015). However, given that the occurrence of strike-slip earthquakes are determined by regional tectonics, it is unclear whether the reduced productivity is a site or source effect.

These studies elucidate the use of insight gathered globally to calibrate forecasts locally. Yet the diversity in methods and data sets hinder an assessment of the relative importance of such inferences. The goal is to systematically assess both setting and source effects on aftershock productivity and distill them down to significant parameters. We then use these data to examine constraints on physical controls of aftershock productivity and provide a roadmap to calibrate productivity to physical factors.

In this study, we first present our measure of aftershock productivity. We then identify global patterns in earthquake productivity, considering setting effects such as plate boundary type, depth, and lithospheric age; and source effects such as radiated energy, stress drop, source geometry, and slip heterogeneity. In so doing, we explicitly address the risk of spurious correlations. We then establish

a parsimonious set of parameters that can help in aftershock prediction. Results indicate that the geometry of the source and the corresponding volumetric availability of stressed faults determine variations in earthquake productivity.

2.2 Metrics and Data

2.2.1 Earthquake Catalogs and Investigated Parameters

To examine variations in aftershock productivity, we utilize the National Earthquake Information Center (NEIC) for recent events and the International Seismological Center (ISC) locations and magnitudes as available for events from 1990 to 2019. We select earthquakes with moment magnitude exceeding global catalog completeness. We determine a global completeness utilizing the Kolmogorov-Smirnov test to evaluate the goodness of fit between a theoretical Gutenberg-Richter distribution and the data for a range of possible completeness magnitude. We identify catalog completeness as the lowest earthquake magnitude producing a local minimum in the Kolmogorov-Smirnov metric (Clauset, Shalizi, and Newman 2009; Goebel et al. 2017). We also test the sensitivity of our results to the magnitude of completeness, finding consistent results for $M_c = 4.5$ to 5.0 (main text versus Appendix B.1.2; Fig B.5-B.15).

We compare relative aftershock productivity to parameters that include site

and source effects. Table 2.1 outlines the selected attributes, coverage, and corresponding data sources. Locations, focal mechanism solutions, and radiated energy estimates were all obtained from the Incorporated Research Institutions for Seismology (IRIS) data management center. Finite fault inversions produced and cataloged by Hayes 2017 were used for other source parameters. Data and analytical limitations are such that rupture properties and source geometry are only available for 98 mainshocks with $M_W \geq 6.8$ (Hayes 2017). Supporting information Table S1 includes the source attributes for these mainshocks. We treat multi-segmented rupture separately.

For several attributes, we scale the parameters in an effort to remove magnitude dependence. For instance, radiated energy is normalized by moment. We normalize width and length by the standard empirical length-scaling, $10^{0.59M_W}$ Wells and Coppersmith 1994, Table 2A - Subsurface rupture dimensions. We normalize rupture area correspondingly, $10^{0.91M_W}$ Wells and Coppersmith 1994, Table 2A - Rupture area. Finally, we log-transform all of these values to linearize their distributions.

Following Noda, Lapusta, and Kanamori 2013, we compute stress drop measurements for all single plane finite fault inversions. Smoothing constraints on the finite fault inversions imply that the stress drop measurements are likely a lower, but consistent bound (Adams, Twardzik, and Ji 2017).

Finally, we introduce a novel quantification of slip heterogeneity (H) derived from finite fault inversions. The metric compares the observed slip to a smooth reference slip distribution,

$$H = \frac{\mu \int_{\Sigma} |u - u_{ref}| dS}{M_o}, \tag{2.2}$$

where μ is the shear modulus, u is the observed slip distribution, u_{ref} is a reference slip distribution, dS is an area element on the fault, Σ is the entire finite fault area, and M_o is the earthquake moment. The reference slip is prescribed as a positive ellipsoid fit to the slip distribution defined by free parameters u_c, x_0, y_0, a and b,

$$u_{ref} = \begin{cases} \sqrt{u_c^2 - \left(\frac{x-x_0}{a}\right)^2 + \left(\frac{y-y_0}{b}\right)^2}, & \text{if } \left(\frac{x-x_0}{a}\right)^2 + \left(\frac{y-y_0}{b}\right)^2 < 1 \\ 0, & otherwise \end{cases} \tag{2.3}$$

The heterogeneity measurement is designed to be most sensitive to large slip fluctuations such as asperities or barriers.

2.2.2 Measuring Aftershock Productivity

This study requires a consistent measure of aftershock productivity comparable on a global basis. To this end, we use a space-time windowing method to identify and count aftershocks. The event-level questions focusing on variations from the mean behavior examined here do not favor methods adaptively fitting different

Table 2.1: Attributes considered in this study.

Attributes	**Coverage**			**Comments**
	Spatial	M_c	Mainshocks	
Site effects				
Time-location[1]	Global	4.5	1011	
Plate boundaries[2]	Global	-	1011	Categorized by nearest digitized boundary provided the focal mechanism is congruent[7] for earthquakes < 400 km from plate boundary; others are intraplate.
Plate age and velocity[3]	Ocean basins	-	333	Determined by nearest crustal age measurement up to 30 km from the mainshock.
Source effects				
Radiated energy[5]	Global	6.0		
Source dimensions[6]	Global	7.0	98	Width (along-dip), length (along-strike) and aspect ratio of ruptures determined from autocorrelation width of finite fault inversions (Mai and Beroza 2000). See text for notes on scaling.
Rupture duration and velocity[7]	Global	7.0	98	
Material properties[7]	Global	7.0	98	V_p, V_s and density used for inversions
Stress drop[6]	Global	7.0	98	See text
Slip Heterogeneity[6]	Global	7.0	98	See text, Equation 2.3
Mixed effects				
Faulting style[8]	Global	5.5	1011	Categorized as strike-slip, normal, and reverse using the P and T axes
		Total	1011	

[1] National Earthquake Information Center (NEIC) and International Seismological Centre (ISC) catalogs downloaded from the Incorporated Research Institutions for Seismology (IRIS) data management service.

[2] Bird et al., 2003 (Bird 2003)

[3] Muller et al., 2008 (Müller et al. 2008)

[4] Produced by Convers et al., 2011 (Convers and Newman 2011) and downloaded from the Incorporated Research Institutions for Seismology (IRIS) data management service.

[5] Attributes were derived from finite fault inversions produced by Hayes, 2017 (Hayes 2017).

[6] Attributes directly measured by Hayes, 2017 (Hayes 2017).

[7] Harvard global Centroid Moment Tensor Solutions (gCMT) and National Earthquake Information Center (NEIC) focal mechanism solutions downloaded from the Incorporated Research Institutions for Seismology (IRIS) data management service.

time or space periods and other seismicity parameters (Ogata 2017). However, any aftershock counting method is subject to biases. We also utilize a clustering method (Zaliapin et al. 2008) as an alternative to test the robustness of our results to the chosen aftershock counting approach (see Appendix Section B.1.3; Fig. B.16-B.15).

For our primary windowing method, we classify earthquakes as foreshocks, mainshocks or aftershocks in a hierarchical sense (Felzer and Brodsky 2006; Brodsky 2011; Wetzler, Brodsky, and Lay 2016; Garza-Giron, Brodsky, and Prejean 2018). We define the largest earthquake in the catalog as a mainshock and then mark as foreshocks and aftershocks earthquakes within magnitude-dependent space and fixed time windows before and after the mainshock. The identified foreshocks and aftershocks are removed from further consideration to prevent double counting. A larger space and time window is used to ensure separation of sequences. Earthquakes within this larger window are also removed from consideration as potential mainshocks and do not count as aftershocks. We sequentially proceed to smaller mainshocks with this classification of foreshocks and aftershocks until we exhaust the catalog. The method is not designed to capture absolutely every aftershock, but rather to provide a consistent measure of aftershock productivity of each isolated mainshock. For the purpose of our study, we only consider mainshocks with $M_W \geq 6.5$ to ensure that the majority of analysed

sequences have observable aftershocks. The hierarchical approach captures both mainshocks that arise naturally from background seismicity, and mainshocks in more complex chains of seismicity in which later earthquakes in a sequence become the largest earthquake and thus the mainshock.

Specific trade-offs determine the choice of the time and space windows. Smaller windows increase the confidence that aftershocks are correctly attributed with few background events included; conversely, larger time windows include more aftershocks and limit the effect of censored statistics (mainshocks with no aftershocks), but may include more background events. We balance these trade-offs to find window selection criteria that yield as many aftershocks as possible without incurring a significant contribution from background activity (Figure 2.2). We assess the performance of each space-time window by comparing the results to the median of 100 time-shuffled catalogs that preserve the original spatial distribution but break the actual time sequence of the catalog (Garza-Giron, Brodsky, and Prejean 2018). The shuffled catalog effectively represents an upper bound on the contribution from background activity since it includes all events in the catalog (including potential aftershocks). The space-time window that includes the most aftershocks while separating the actual aftershock productivity relationship from the shuffled ones is the preferred choice.

Fig. 2.2 shows results for a suite of space and time windows. Space windows

are all measured in terms of source radius estimated following Table 2A of Wells and Coppersmith 1994, which provides the subsurface rupture length:

$$R_{source} \sim (2 \times 10^{0.59 M_W})\text{m} \tag{2.4}$$

For reference, M_W 6.0 and 9.0 earthquakes have dimensions on the order of ~ 7 km and ~ 400 km respectively. We also check the results by using the more recent geodetic-derived scaling relationship of Brengman et al. 2019. Supplemental Fig. B.2 demonstrates that the results are nearly the same, with 79% of the mainshocks having identical aftershock counts.

A spherical space window of three source dimensions in radius centered on the mainshock location and a time window of 60 days following the mainshock performs best. Smaller and shorter windows result in a clearer separation of the shuffled statistic. However, the number of mainshocks with no aftershocks increases dramatically (e.g., 30% of mainshocks in the case of the ten-day and one source dimension window, as opposed to 10% for our preferred window). Space-time windows that are too large may include a significant proportion of background events. However, for the preferred window, we show that in nearly all cases (97% of sequences) even a very conservative and absolute upper-bound on background seismicity (the shuffled catalog statistic) is within one standard deviation of Poisson counting error (Appendix Section B.1.4). The larger space window, used to eliminate events from further consideration, is four source dimensions and an ad-

ditional 40 days. We use the same combination of selection and screening in space and a 1 day time window to classify foreshocks. Using the space-time windowing approach we ensure that nearly all mainshocks (99%) are fully isolated in time and space from each other. The non-isolated cases occur when two mainshocks of different magnitudes have overlapping, but not coincident, aftershock sequences. These occurrences are rare (1%) and therefore we do not complicate the algorithm further to eliminate them.

As an alternative method, we use a clustering routine following Zaliapin et al. 2008 and Goebel et al. 2019. This approach seeks to build earthquake families by linking earthquakes to parent events based on a distance metric that combines magnitude, space and time. Pairs of parent and daughter events exhibit a statistical distribution with two modes: one that corresponds to clustered events and another that arises from a Poissonian background of seismicity. Separation of earthquake clusters is achieved by defining a decision boundary between these two modes and cutting all links that exceed this threshold. The largest event in each cluster is identified as a mainshock and aftershocks are counted as the number of events that follow it. See Zaliapin et al. 2008 for a detailed overview of the method, distance metrics, and theoretical connections to other schemes (e.g. ETAS). The specific parameters selected are consistent with previous implementations of Zaliapin et al. 2008 and are documented fully in the supplementary material (see

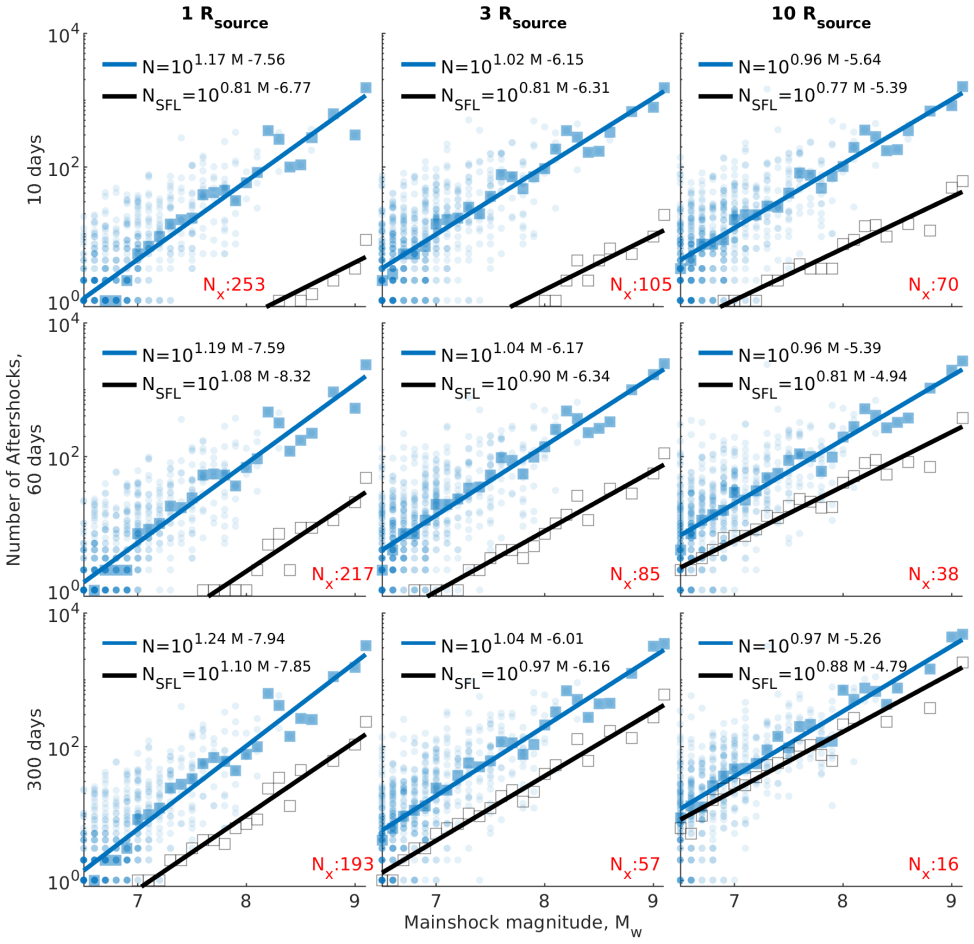

Figure 2.2: Sensitivity analysis of space-time windows. Time windows of 10, 60, and 100 days and spherical space windows with radii of 1, 3, and 10 source dimensions (R_{source}) are considered. Corresponding larger space and time windows are 4/3 and 5/3 of the selection windows. Blue circles are individual mainshocks identified with the hierarchical counting routine. The blue line indicates the outcome of a regression of the median log-number of aftershocks for each 0.1 magnitude bin (blue squares). For reference, we computed the same productivity relationship, N_{SFL} (black line), for the median of 100 time-shuffled catalogs (grey squares). For each space-time window, we indicate the number of mainshocks with no aftershocks in red (N_x). Note that as space and time windows increase, more mainshocks have measurable aftershock counts. However, as time and space windows increase, inflating aftershock counts and reducing α-values by overestimating the productivity of smaller events becomes increasingly prevalent.

Appendix Section B.1.3). Fig. B.16 shows that the aftershock productivity measured from both methods correlate well ($R^2 = 0.96$). This consistency stems from the relatively high completeness of our catalog which ensures that, for the relatively short time intervals of our aftershock windows, few, if any, background events are likely to be included in the event counts.

For both methods, we next compute the median number of aftershocks for each 0.1 magnitude bin in the mainshock catalog (including the counts of events with zero aftershocks) and the corresponding magnitude bin to perform a linear least squares inversion to determine k and α (Fig. 2.1a and B.17). Using these parameters, we define the relative productivity ($\Delta \log(N)$) for each mainshock as

$$\Delta \log(N) = log(N) - log(\hat{N}(M)) = log\left(\frac{N}{k10^{\alpha M}}\right) \tag{2.5}$$

where N is the observed number of aftershocks following a mainshock and $\hat{N}(M)$ is the number of aftershocks predicted for the mainshock magnitude, M, from Equation 2.1.

The relative productivity is closely related to variations in the Omori a-value or K value used in Page et al. 2016; Hardebeck et al. 2018; Utsu, Ogata, and Matsu 'ura 1995; Ogata 1988. The use of $\Delta \log(N)$ as a measure of relative productivity provides a means to assess whether scatter in aftershock production is related to specific mainshock parameters on an earthquake-by-earthquake basis.

It is important that relative productivity be independent of magnitude, as we run the risk of confounding variations in the average size of earthquakes with variations in productivity. We show that the median and interquartile range of relative productivity are not magnitude-dependent (Fig. 2.1b and B.1). Note that median and interquartile statistics which we use throughout ensure that we account for mainshocks with no aftershocks which have undefined $\Delta \log(N)$.

For each major result figure we provide three versions. We provide the preferred solution in the main text based on aftershock counts using the space-time windowing described above and a completeness threshold of $M_W 4.5$. In addition, we provide alternative solutions in Section B.1.2 and B.1.3 of the appendix. These are based on aftershock counts using the same space-time windowing but with a more conservative completeness threshold of $M_W 5.0$, and solutions based on the Zaliapin declustering method with a completeness threshold of $M_W 4.5$. All three Figures are referenced as each result is introduced for easy comparison. For the most part, the alternative methods are confirmatory and further commentary is only provided when conspicuous differences emerge. Given that the three treatments have varying intrinsic suppression of bias from background activity, consistency in the results strongly suggests that bias in N is negligible for our parameterizations.

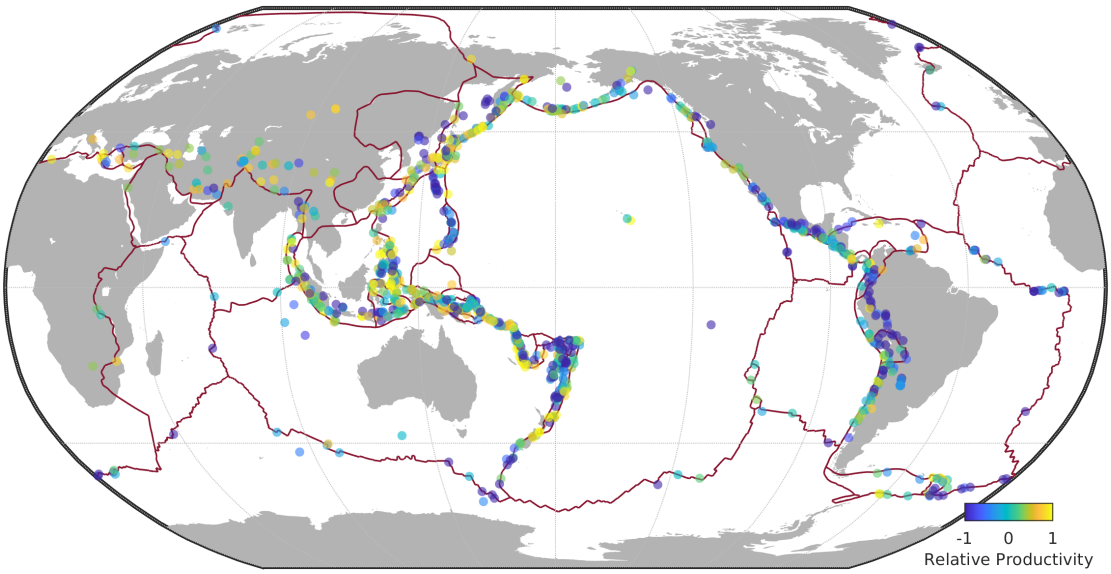

Figure 2.3: Global map of earthquake productivity. Red lines indicate the surface trace of the tectonic boundaries. Mainshocks with $M_W \geq 6.5$ color-coded according to their relative productivity (Equation 2.5).

2.3 Results

2.3.1 The Global Earthquake Productivity Map

Our analysis yields 1011 earthquake sequences with mainshocks exceeding $M_W > 6.5$. We map the global catalog of aftershock productivity (Fig. 2.3, B.7 and B.18). We highlight characteristic global patterns which we will examine in more detail in the following sections.

Intermediate- and deep-focus earthquakes stand out in the relative productivity map. Continental-scale bands of earthquakes with low relative productivity run along the Atacama, Japan, Izu-Ogasawara, Mariana and Tonga trenches.

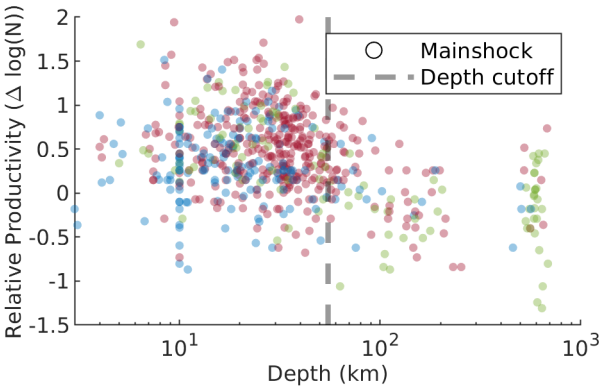

Figure 2.4: Relative aftershock productivity as a function of depth. Subsequent analysis will only consider earthquakes shallower than the 55 km cutoff (dashed line). Sequences are color-coded according to faulting style of the mainshock (blue: strike-slip, green: normal and red: reverse). Note: Discretization of depth is apparent in this plot as some events have default values. Depths of 33 km, 5 km, 10 km and 15 km are reported for 6%, 1%, 10% and 0.7%, respectively, of the catalog.

These are predominantly normal and reverse earthquakes rupturing intermediate- to deep-focus seismic zones. Over the entire catalog, earthquakes deeper than $\sim$ 55 km exhibit decreasing relative productivity with increasing depth (Fig. 2.4, B.8 and B.22). Similar observations are well documented (Båth 1965; Frohlich 1989; Nyffenegger and Frohlich 2000; Wiens et al. 1997; Wu and Chen 1999; Persh and Houston 2004). We, therefore, do not include mainshocks deeper than 55 km in the following analysis to avoid confounding depth with other influences.

Offshore earthquakes that are not in subduction zones have lower relative productivity. Oceanic transform and divergent boundaries usually host less productive aftershock seismicity than the global trend ($\Delta \log(N) < 0$). Consider for

example the 2018 M_W7.5 earthquake north of Honduras rupturing the Swan Islands oceanic transform fault. It had 2 aftershocks with $M_W > 4.5$ within 150 km of the epicentral location; for reference, the global median aftershocks count for mainshocks of M_W7.5 is 11 and therefore $\Delta log(N) = -0.7$. Earlier earthquakes along the transform have similarly low relative productivity. Very few of these offshore events are productive. With 170 aftershocks within 792 km, the M_W8.6 2012 intraplate strike-slip rupture offshore Sumatra followed by a M_W8.2 aftershock appears to be among the most productive earthquakes hosted in oceanic lithosphere not within a convergent boundary ($\Delta log(N) = -0.15$).

The western coastline of North America hosts spatially coherent patterns in relative aftershock productivity (Fig. 2.5a, B.9a and B.20a). The Aleutian Arc grades from earthquakes with generally high aftershock productivity in the West to low aftershock productivity in the East. Earthquakes on the Queen Charlotte Fault exhibit aftershock abundances similar to the global average. Offshore clusters of seismicity along the Blanco Fracture Zone and the Mendocino Triple Junction have low productivity. Continental earthquakes along the San Andreas Fault System are markedly more productive than the seismicity to the north and south. The pronounced decrease in productivity at the southern terminus corresponds to a shift from generally transpressional continental- to transtensional oceanic-tectonics. Subduction of the northernmost Section of the Cocos plate

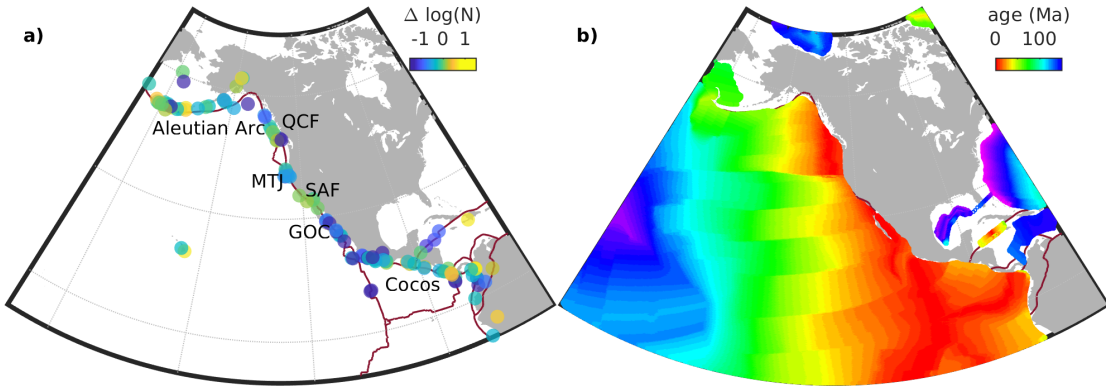

Figure 2.5: a) Aftershock productivity along the North American coastline. Individual mainshocks (circles) are color-coded according to their relative aftershock productivity ($\Delta \log(N)$, Equation 2.5). The Aleutian Arc, Queen Charlotte Fault (QCF), Mendocino Triple Junction (MTJ), San Andreas Fault (SAF), Gulf of California (GOC) and Cocos Plate Subduction Zone include areas with coherent productivity. Red line indicates major plate boundaries (Bird 2003). b) Seafloor crustal age estimates from Müller et al. 2008.

under Mexico is associated with low aftershock productivity. The number of aftershocks increases southward along central America. The gradational increases in productivity along the Aleutian Arc (westward) and Cocos Subduction zone (northward) correspond to increasing lithospheric age of the oceanic plate (Fig. 2.5b).

Continental seismicity has elevated aftershock productivity. Most notable is the India-Asia Collision Belt. Other examples include onshore seismicity in New-Zealand, the San Andreas Fault, and the Arabian Plate collision.

2.3.2 Tectonic Setting Effects

We investigate the effect of the seismo-tectonic setting on the productivity of earthquake sequences. Specifically, we subdivide the mainshock catalog by plate-boundary-type following the categorization detailed in Table 2.1. This comparison is shown in Fig. 2.6, B.10 and B.21. Both clustering algorithms single out oceanic transform faults as particularly deficient in aftershocks. Continental transform faults are more productive, but less so than the global average. Subduction zones, continental convergent boundaries, continental rift basins, and oceanic convergent boundaries have similar and generally elevated relative productivity. A two-sample Kolmogorov-Smirnov test comparing the cumulative distribution of each subset to the entire set suggests that the relative productivity of events on oceanic transform faults and continental convergent boundaries have a small probability, 10^{-9} and 3×10^{-5} respectively, of being randomly sampled from the overall distribution by chance; the remainder of the subsets are not significantly different from the overall distribution ($p = 0.05$). The result is not an effect of less coverage in the oceans because the same analysis with a higher catalog completeness ($M_W 5.0$) yields a similar separation of oceanic transform events.

Earthquake productivity is also related to the age of the lithosphere (Fig. 2.7, B.11 and B.22). Earthquake sequences in younger oceanic lithosphere tend to have fewer aftershocks. The increase in median productivity is strongest for

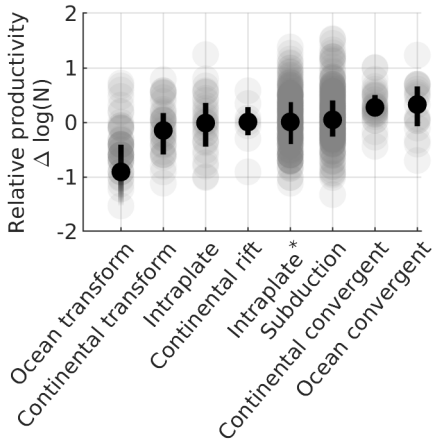

Figure 2.6: Earthquake productivity by tectonic boundary. Circles indicate the relative productivity of individual sequences. Solid markers and error bars indicate the median and the interquartile range. A faded lower error bar implies that mainshocks with no aftershocks are within the interquartile range. Intraplate* indicates earthquakes within 400 km of a plate boundary but with a faulting mechanism discordant with the plate boundary (e.g., outer rise events).

strike-slip mainshocks and those earthquake sequences in oceanic lithosphere with ages less $\sim$ 40 Ma. This same trend, although slightly weaker, is apparent for the higher catalog completeness threshold of M_W5.0 (Fig. B.11). With M_W5.0 as a completeness threshold, the fraction of mainshocks with no aftershocks in <40 Ma oceanic lithosphere is approximately twice the global average.

2.3.3 Source Effects

To identify mainshock source effects that may influence relative aftershock productivity, we linearly regress each available parameter with the relative productivity and compare the goodness of fit to random trials. The variance reduction

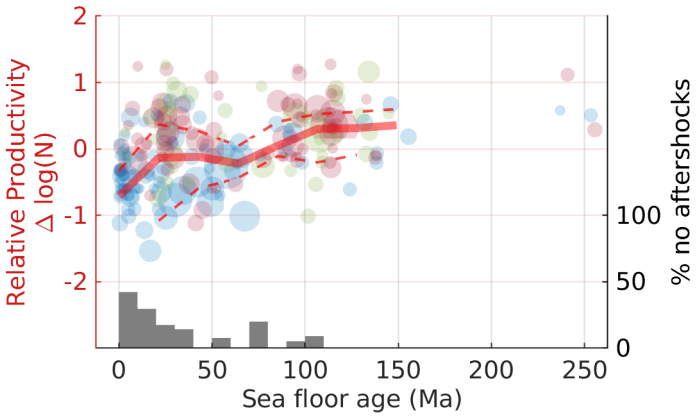

Figure 2.7: Relative productivity increases as a function of the age of the oceanic lithosphere. Each circle indicates an individual earthquake sequence. Sequences are color-coded by faulting style of the mainshock (blue: strike-slip, green: normal and red: reverse). The red line indicates the median average for 20 Ma crustal age bins. Dashed lines indicate the corresponding interquartile ranges. Vertical grey bars indicate the fraction of earthquakes with no aftershocks within each 10Ma crustal age bin and correspond to the right-hand axis.

for each regression is presented in Fig. 2.8a, B.12a and B.23a.

At face value, this analysis shows that aspect ratio, width, and stress drop are best correlated with the relative productivity. These correlations are seen in the raw data, albeit with significant scatter Figures 2.8b-d, B.12b-d and B.23b-d).

A potential problem with this analysis of 12 test parameters is that a spurious correlation could arise simply because of the numerous investigated regressions. We explicitly address this issue by investigating the probability of spurious correlations in randomized data. To do so, we first remove any causal relationship between our 12 predictors and relative productivity by shuffling the aftershock measurements and randomly reassigning each relative productivity measurement

to the parameters for a different mainshock. We then regress each parameter with the relative productivity and report the maximum variance reduction of any parameter in this shuffled set. This same routine is repeated 10000 times to generate a probability distribution function of the maximum variance reduction of 12 parameters should there be no causal relationship in the data. We refer to this evaluation of the extreme value for the full group, or family, of parameters as a family-wise test and it serves as a null hypothesis. We determine the percentile of our actual regression results within these random realizations, the family-wise p-value, for comparison (Fig. 2.8a - top axis). The rupture's normalized energy, normalized length, normalized area, Poisson's Ratio, log-heterogeneity, Young's Modulus, and velocity all yield a variance reduction with more than a 5% chance to arise by chance for the shuffled data and are thus not further considered. Refer to Fig. B.28 to see all correlations (Convers and Newman 2011; Hayes 2017).

Our analysis suggests that correlations with log-stress drop ($p = 0.07$) and normalized width ($p = 0.05$) are marginally significant. Aftershock productivity negatively correlates with the logarithm of stress drop and positively correlates with normalized rupture width (see Figures 2.8b-c, B.12b-c and B.23b-c).

Slip-zone aspect ratio ($p = 0.007$) is related to relative productivity in a statistically significant sense. Aftershock productivity negatively correlates with aspect ratio (see Fig. 2.8d).

The alternative counting algorithm yields estimates of relative productivity which best correlated with stress drop, width, area and aspect ratio (in decreasing order). The statistical significance and goodness of fit for these predictors is also much higher (Fig. B.23).

2.3.4 Focal Mechanism Dependence of Aftershock Productivity

Aftershock productivity exhibits a strong relationship with focal mechanism of the mainshock (Fig. 2.9a, B.13a and B.24a). Strike-slip mainshocks have relatively few aftershocks. The median number of aftershocks for strike-slip mainshock is three times fewer than dip-slip mainshocks of comparable magnitude. This separation by focal mechanism far exceeds 95% confidence intervals ($p \ll 0.05$).

Whether this is a site or source effect is ambiguous. We examined whether earthquakes with different focal mechanisms that share the same geographic location still exhibit statistically distinct earthquake productivity. If the tectonic setting is the only control on the relative productivity of earthquakes, then strike-slip earthquakes should not have fewer aftershocks when controlling for location.

We construct a catalog of co-located strike-slip and dip-slip earthquake pairs by iteratively cataloging the nearest pairs of strike-slip and dip-slip mainshocks. We explicitly avoid double counting throughout this process and ensure regular global

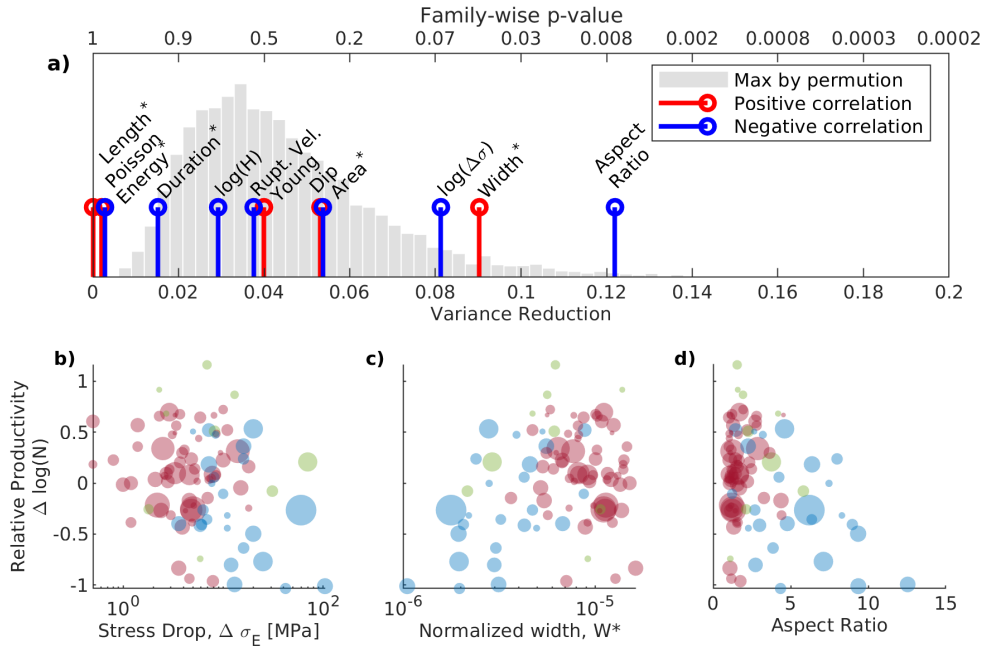

Figure 2.8: a) Goodness of fit of linear regressions for each source attribute in our combined catalog. Top and bottom axes respectively represent the p-value and goodness of fit of each attribute (stems). The probability distribution function in the backdrop indicates the maximum variance reduction outcome of 10000 permutation test of the entire data set. Asterisks indicate scaled and log-transformed variables. The scaled energy, length, duration and area, material properties, velocity, dip, and log-stress drop ($\Delta\sigma$) of the mainshock rupture all do not yield a statistically significant ($p = 0.05$) linear fit to the relative productivity. The normalized rupture width and aspect ratio of the rupture yield the best fitting linear regressions. Stems are color-coded to indicate whether the source attribute is positively (red) or negatively (blue) correlated with relative productivity. b-d): Relative earthquake productivity as a function of mainshock stress drop, normalized rupture width, and aspect ratio. Individual mainshocks are color-coded according to faulting style as in Fig. 2.1.

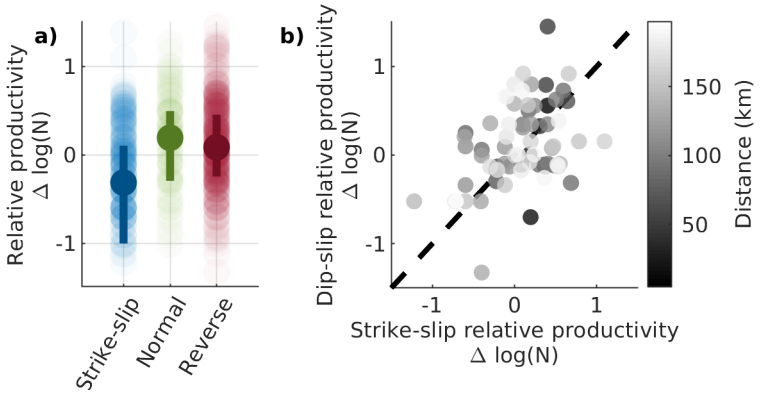

Figure 2.9: a) Relative aftershock productivity, $\Delta \log(N)$, by focal mechanism (Equation 2.5). b) Relative aftershock productivity for pairs of earthquake sequences with strike-slip and dip-slip mainshocks within 200 km from each other. Each pair is shaded according to its relative distance. Dashed line indicates a 1:1 relationship, the expectation for a purely site dominated control on relative productivity. Co-located mainshocks pairs generally follow this 1:1 trend, but exhibit considerable scatter.

coverage. The productivity of co-located strike-slip and dip-slip earthquakes generally follows a 1:1 trend (Fig. 2.9b, B.13b and B.24b). The distinction by faulting style is not statistically significant when comparing only co-located earthquakes ($p = 0.63$). This shift indicates that event-location alone partially explains why strike-slip earthquakes are deficient in aftershocks. Significant scatter implies that source effects (e.g., aspect ratio, stress drop, and dip) may contribute to the distinction.

2.3.5 A Multi-Attribute Prediction of Aftershock Productivity

We have considered variables on a case by case basis and have prescribed a functional form to the relationships. In this section, we allow for increased relational complexity to predict relative productivity given contextual information about the mainshocks setting and kinematics. We measure the quality of our predictions by computing the root mean squared error, $RMSE = \sqrt{\frac{\sum (\hat{f}_i - f_i)^2}{N}}$, where $\hat{f}_i$ is the prediction of relative productivity and f_i are one of N observed values. To prevent over-fitting the data, we perform leave-one-out cross-validation—individual predictions are calibrated on the remainder of the data (Witten, Frank, and Hall 2011).

As a point of reference, we produce predictions of relative productivity based on local seismicity. In this approach, aftershock productivity is predicted based on the median relative productivity of the 50 nearest mainshocks. Prediction accuracy is largely insensitive to the number of neighbors (20-100 yield comparable results). The chosen number of nearest neighbors produces the lower root mean squared error between predictions and observed relative productivity measurements. Model validation is shown in Figures 2.10a, B.14a and B.25a. This approach captures geographical effects, provides an 8% forecast improvement in the RMSE when compared to the global productivity law, and serves as a baseline

for the following.

We then tabulate earthquakes for which all parameters exist and systematically test predictions based on permutations of parameters on various machine learning algorithms. We find that Support Vector Machines (SVM) yield predictions with low root mean squared error (Figures 2.10b, B.14b and B.25b). An SVM finds hyperplanes that minimize the prediction-error by mapping data to higher dimensional spaces with coordinate transformations called kernels. Key differences between SVM regression and a linear regression are 1) a tolerance for a margin of error, 2) a simultaneous minimization of model complexity and 3) non-linearity which arises from the kernel-transformations. These features make SVMs particularly well suited to highly heterogeneous, relatively small, multivariate data sets such as ours (Witten, Frank, and Hall 2011). We manually calibrate the SVM by changing the kernel-transformation, the margin width, and the complexity trade-off. We present results from an SVM model trained using a quadratic kernel. The trained model is included in the supporting information files.

The following metrics yield the best predictions of relative productivity: scaled rupture area, fault dip, and plate age. Better predictions on training data, with adverse predictions on the validation data, indicate that additional parameters do not improve predictions and instead induce over-fitting.

The root mean squared error of the final model predictions of the SVM is 0.40, a 27% improvement on the global productivity law and a 20% improvement on the nearest neighbor algorithm, despite employing no direct geographical information. Notably, the SVM model better predicts extreme cases (highly productive or unproductive). The root mean squared error of the SVM is influenced by the following two outliers, the 2017 M_W7.0 Loyalty Islands earthquake and the 2013 M_W7.8 Scotia Sea earthquake. The productivity of these events is underestimated by the model. Interestingly, both these mainshocks were preceded by large foreshocks: two M_W6+ earthquakes for the former and a M_W6.8 earthquake for the latter. Removing these two outliers yields an additional 10% improvement to the predictions.

The preferred parameters of the SVM can be compared to those found in Section 2.3.3. For instance, dip appears to be more important in the SVM case. Aftershock productivity relates to dip non-monotonically with a maximum at intermediate dips ($\sim 25^o$) (See Fig. B.28, Panel l). The added non-linearity allowed by the SVM model kernels is useful in this situation. Normalized rupture area is also more important in the SVM than in the linear regressions. The cause of this difference is harder to discern, but is likely a combination of the added non-linearity allowed by the SVM model and the co-variance across the combined attributes. Such differences emphasize that the collapse of multivariate data into

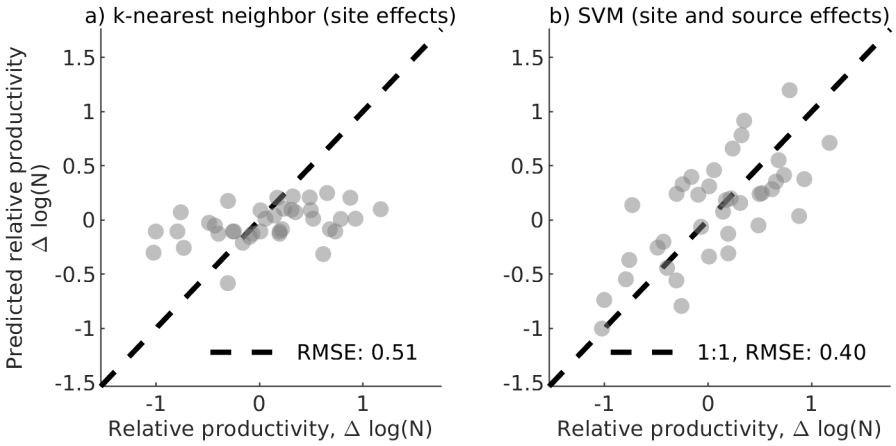

Figure 2.10: Response plots (prediction versus observation) for the k-nearest neighbor algorithm (a) and SVM models (b). Point indicates the predicted versus observed value of relative productivity for each earthquake sequences. A perfect prediction would place all values on the 1:1 line. The SVM model provides a 22% improvement in the root mean squared error when compared to k-nearest neighbor model. Combining both contextual information about the setting (crustal age) and the source (dip and normalized area) yields a root mean square value of 0.40.

single attribute linear relationships has its limitations.

2.4 Interpretation

Relative productivity is sensitive to both setting and source effects. From setting, dominant factors are lithospheric age, lithosphere type (oceanic vs. continental) and depth. Important parameters related to mainshock source effects are are area, width, aspect ratio and stress drop. Focal mechanism is an additional factor that can be construed as stemming from both tectonic setting and source. Fig. 2.11 (B.15 and B.26) synthesizes these results and demonstrates their relative importance. Similar results for large ($M_W > 7.5$) and small ($6.5 < M_W < 7.5$) mainshocks indicate no significant regional scaling differences confounds our results (Figures B.3-B.4).

As demonstrated in Sections B.1.2 and B.1.3, our results are robust to aftershock selection method and catalog completeness, which are the primary two factors that could influence our measurements. A higher catalog completeness of $M_W 5.0$ reproduces results similar to those we presented using a global completeness of $M_W 4.5$, though far more mainshocks have no observed aftershocks (Section B.1.2) – when more than half of mainshocks considered in a group are censored, the estimation of a median productivity is obscured. Using an alternative clustering routine, the relationship between relative productivity and crust age is slightly

less pronounced but still apparent (Section B.1.2). We find better relationships between aftershock productivity and normalized rupture area (p=0.004), stress drop (p=0.0008) and normalized width (p=0.002). Note that these changes occurred among variables that are strongly co-varied (as shown in Fig. 2.12). The correlation between source properties and aftershock productivity are otherwise unchanged using this alternative definition of aftershock clusters. The consistent results across these sensitivity tests highlight that our major results are robust.

2.4.1 Relative Importance of Setting and Source Effects

Our parameterization of slip heterogeneity does not correlate well with productivity. This observation contradicts some modelling predictions (Helmstetter and Shaw 2006; Marsan 2006) and long-standing interpretations (Mogi 1967). The observation suggests that the number of observable aftershocks is dominated by surrounding volume outside the main slip zone and less strongly modulated by residual stress on the fault plane. This particular aspect of source does not appear to exert a strong influence on relative productivity.

Measurements of relative productivity show no correlation with the scaled radiated energy of the mainshock. Within three source dimensions of the mainshock, static and quasi-static effects dominate earthquake triggering.

Our observations (Fig. 2.8a-b), particularly underscored by Fig. B.12a-b),

indicate that productivity decreases, albeit weakly, with increasing stress drop. High stress drop ruptures drive more elevated stresses at the periphery of the fault which may contribute to higher aftershock productivity; however, this effect might be reversed by the size of the rupture which is smaller for a higher stress drop. Our observations indicate that the geometric compactness that results from high stress drop dominates: a smaller volumetric activation from high stress drop ruptures results in fewer triggered aftershocks (Wetzler, Brodsky, and Lay 2016).

As discussed above, some source attributes strongly co-vary with each other (Fig. 2.12). Width, length, rupture heterogeneity, aspect ratio, stress drop and dip are particularly correlated and separating their relative importance is complicated. However, we can distinguish the categories of parameters that appear to be related to relative productivity versus those that do not. For instance, in all analyses rupture length is uncorrelated with productivity in spite of co-varying with important parameters. Other trends get mapped into different parameters by the two methods. For instance, normalized area strongly anti-correlates with stress drop as both are measures of the compactness of the source. Which one of these two similar parameters dominates depends on the exact data selection criteria. Fig. 2.12 shows that for the subset of earthquakes that have all the variables available that the SVM requires, productivity correlates slightly more strongly with normalized area than stress drop. Therefore, it is not surprising that

the SVM mapped the trend into normalized area rather than stress drop and the results are compatible with the linear regression (Fig. 2.8). The important result here is that a combination of a measure of compactness, lithospheric age and dip together are a better predictor of productivity than any individual parameter.

Comparing source and setting influences (Figures 2.11, B.15, and B.26) emphasizes the role of the setting. The source parameters we consider tend to have a more subdued influence on the relative productivity than do some of the tectonic controls. In particular, crustal age, lithosphere type, and depth separate populations of earthquakes with significantly different aftershock productivity.

2.4.2 Importance of Fault Availability

The dependence on crustal age, lithosphere type, depth, aspect ratio, stress drop, and the underlying magnitude scaling are all congruent with a primary physical control: the volume of nearby rock susceptible to brittle failure influences aftershock productivity.

Fig. 2.11, B.15 and B.26 places plate age as a pronounced control on aftershock statistics following depth effects. With increasing age, the oceanic lithosphere becomes colder, thicker, more brittle and less buoyant. Subduction zone earthquakes hosted along younger lithosphere tend to generate larger earthquakes (lower b-values), particularly so in the first $\sim$ 70 Ma. Nishikawa and Ide 2014

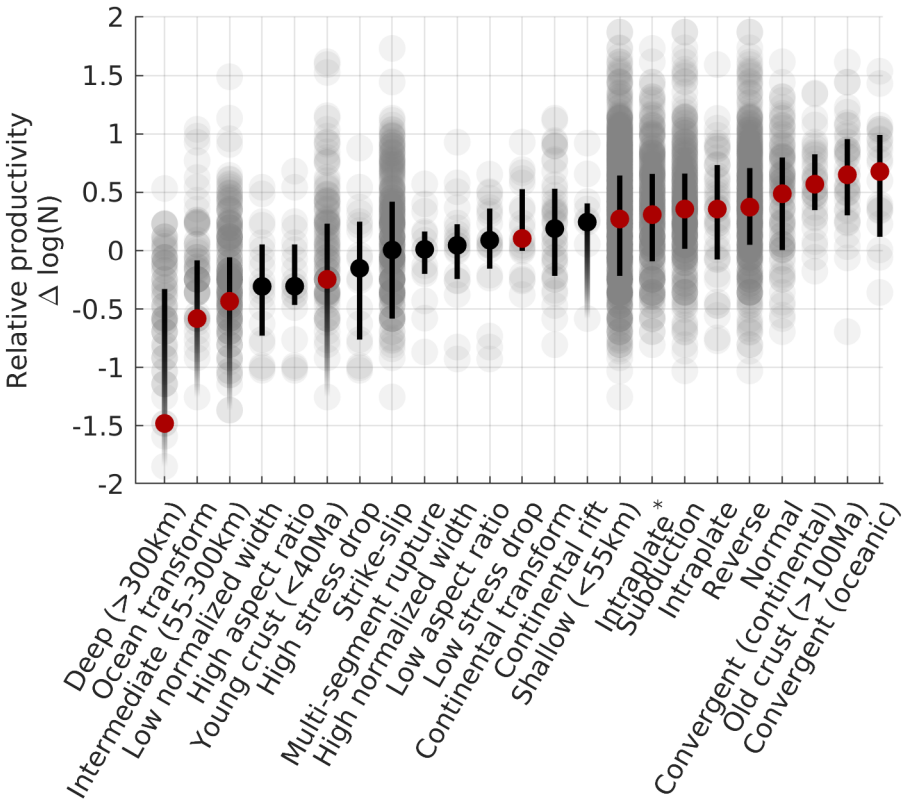

Figure 2.11: Synthesis of relative productivity according to catalog subsets. The group considered here are the short list which best distinguished relative productivity based on our different lines of investigation (Sections 2.3.1-2.3.3). 'High' and 'low' subsets respectively refer to $>$ 80th and $<$ 20th percentile ranges of the data. Grey circles are individual mainshocks. Opaque points and error bars respectively indicate the median and interaquartile range of the subset. Fading error bars imply that mainshock sequences with no aftershocks are within the interquartile range of the data. Attributes with red markers are consistent with the hypothesis that they are sampled from a different continuous distribution than the overall population of earthquakes using a 2-sample Kolmogorov-Smirnov test at a 5% significance threshold.

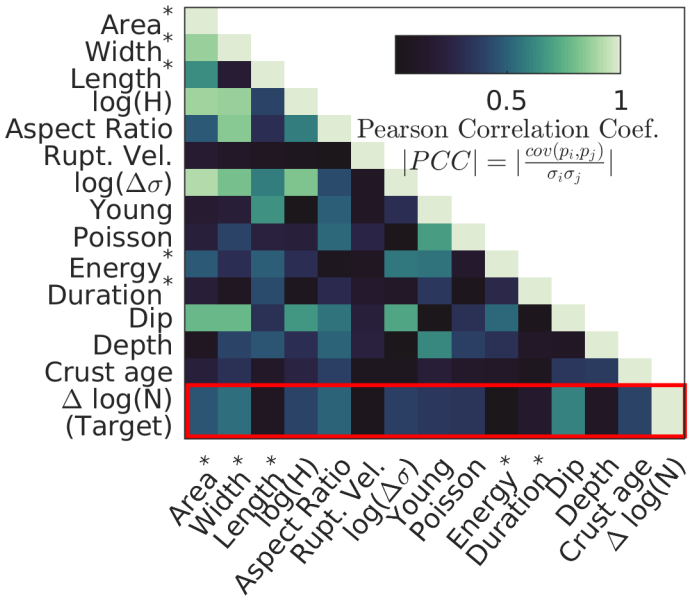

Figure 2.12: Pearson Correlation Coefficient matrix for quantitative predictors of relative earthquake productivity. Correlation coefficients comprises earthquake shallower than 55 km for which all source attributes were calculated. Brighter colors indicate higher absolute values in correlation. Note that examining the correlation coefficient does not fully capture more non-linear interactions.

attributed this trend to changes in the buoyancy of the subducting slab. While variations in b-value statistics may covary with aftershock productivity, we find it difficult to reconcile their physical interpretation with the increasing aftershock productivity along transform faults that we observe.

Old oceanic lithosphere is also thicker which, in turn, increases the volume in the brittle failing regime. The change in susceptibility of the surrounding volume to earthquakes stressing is a natural explanation for productivity increase with plate age. Though the effect was more subdued, a careful analysis of subduction zones also revealed the same pattern (Wetzler, Brodsky, and Lay 2016). Old ($>$ 100Ma) oceanic lithosphere hosts seismicity along Japan and in the western Mediterranean Sea, which are highly productive in aftershocks (see Fig. 2.3, B.7, and B.18). Generally, the contrasting aftershock productivity of the Eastern and Western Pacific can be related to lithospheric age.

We may cast continental earthquakes, which we observe to have elevated aftershock productivity, as a highly thickened (old) lithosphere end member. Equivalently, convergent boundaries which effectively double the thickness of the lithosphere are associated with high aftershock productivity. The deeper the zone of low-temperature brittle material, the higher the aftershock production.

Intermediate- and deep-focus earthquakes are the opposite end-member, instead exhibiting very low aftershock productivity. Deep earthquakes within sub-

ducting oceanic lithosphere are confined above and below by viscous mantle. Moreover the lithosphere is thinned by thermal conduction from the warmer mantle. Subducting lithosphere becomes rejuvenated and hence aftershock poor.

Long aspect ratio ruptures reflect the saturation of the brittle crust which is limited in this case by the surface of the Earth above and by the ductile zone below (Scholz 2019). Spatial confines limit the volume that may host aftershocks. It is telling to observe that the down-dip width of the rupture correlates far better with aftershock productivity than its length (Fig. 2.8, B.12, and B.23). In other words, vertical confines (the surface and the ductile base of the seismogenic zone) are better controls on the productivity of earthquakes than the lateral limits.

The final evidence for the significance of volume availability arises from the scaling of aftershock productivity with mainshock size. The scaling captures the mainshocks ability to activate a larger volume with increasing earthquake size. Our results are a natural extension of this basic premise. Fluctuations around the magnitude scaling, i.e. relative productivity, result from fluctuations in the mainshock's ability to brittlely deform its surrounding volume.

Previous work has found that aftershock clustering statistics negatively correlate with heat flow, and interpret their findings in the context of a temperature dependent rheology (Ben-Zion and Lyakhovsky 2006; Zaliapin and Ben-Zion 2016). The observations are consistent with fault availability playing an important role.

Regions with thin lithosphere have high heat flow and so are predicted to be aftershock poor in both frameworks. However, the correlation of focal mechanism (Figures 2.9a, B.13 and B.24) with aftershock productivity is more challenging to explain by rheological changes alone. The direct correlation between aftershock productivity and heat flow may be a special case of the influence of fault availability since high heat flow can reduce the number of available faults for aftershocks by thinning the seismogenic crust.

2.4.3 Improving Aftershock Forecasts

Short-term hazard assessment following a large earthquake relies on regional catalogs to calibrate the statistical behavior of the local seismicity. Unfortunately, seismic records are temporally limited and generally do not span large-earthquake-cycle timescales, particularly for the determination of aftershock parameters such as aftershock productivity. For large events where local data are limited, expert judgement will often rely on past ruptures to serve as analogs for the upcoming hazard. We extend and formalize this approach with a statistical treatment using flexible prediction tools. Using local analogs to determine the aftershock hazard is often impossible. Our analysis provides a complimentary approach. Indeed, we show that contextual attributes are strongly indicative of upcoming aftershock productivity. Thus, teleseismic data and a coarse knowledge of the local tec-

tonic context can significantly help constrain short-term hazard following large earthquakes in poorly instrumented or quiescent areas. In particular, Fig. 2.10b suggests an algorithm to use in future aftershock prediction, indicating 10-fold improvements in the forecast of aftershock counts.

2.5 Conclusion

We synthesize multiple possible relationships between aftershock productivity and the effects of earthquake setting and rupture properties from 1990 to 2019. In addition to the decrease in productivity with depth, global patterns indicate that earthquake productivity is particularly low along oceanic transform faults and tracks lithosphere ages. The relationship suggests that productivity increases because of the cooling and thickening of oceanic lithosphere with time. In assessing source properties, we found that the rupture's aspect ratio, the down-dip width and, to a lesser degree, stress drop, correlate most strongly with aftershock productivity; other parameters including rupture duration, and length, scaled radiated energy and material parameters did not correlate significantly. The short-term seismicity forecast is improved ten-fold from a combination of plate age, dip, and normalized rupture area. These improvements do not require regional calibration based on historical seismicity and therefore lend themselves well to large remote earthquakes where teleseismic data are available, but long-term monitoring is not.

Source geometry and the availability of stressed faults are inferred to provide a primary control on the number of aftershocks triggered.

Chapter 3

Two Foreshock Sequences

Post-Gulia and Wiemer (2019)

3.1 Introduction

Global statistical analyses of earthquake sequences and the advent of high-resolution earthquake catalogs demonstrate that foreshock sequences often precede large earthquakes (Bouchon et al. 2013; Marsan et al. 2014; Mignan 2015; Trugman and Ross 2019). Recognizing that an event is a foreshock would provide useful forecasting capability. However, decades of work have failed to establish a robust feature of individual foreshocks that distinguishes them from mainshocks or their aftershocks. An alternative would be to find a distinct statistical at-

tribute of a sequence of events that occurs prior to a mainshock versus a sequence of events that follow a mainshock (Helmstetter and Sornette 2003). Foreshocks may be distinctive due to some precursory loading process or influence from the locked zone of the subsequent mainshock, neither of which will exist for the aftershock sequence (Brodsky and van der Elst, 2014). Recently, Gulia and Wiemer 2019 proposed that abrupt changes in magnitude-frequency distribution relative to background levels in localized regions around the largest events provide such a distinguishing attribute. They evaluate their method for 27 earthquake sequences not followed by a second mainshock, plus two followed by a second mainshock (Amatrice, Italy, and Kumamoto, Japan). Eighteen correct negatives, two true positives, ten neutral alarms, counted neither as successes or failures, and one false alarm yielded a 95% success rate in discriminating foreshock sequences from aftershock sequences. Their approach relies on time-varying estimates of the slope of the cumulative magnitude-frequency distribution, or b-value of the Gutenberg-Richter relationship,

$$b = \frac{N}{\log(10) \sum_{i=1}^{N} M_i - M_c + \delta_M/2}. \tag{3.1}$$

The b-value, here estimated by the maximum likelihood method (Aki, 1965), is calculated in a sliding time window, where N is the number of events in the selected window of space and time, M_i is the magnitude of each event, M_c is the catalog

completeness magnitude (lowest magnitude for which all events are thought to be detected), and $\Delta_M/2$ is the Utsu (1965) correction for magnitude value round-off. Gulia and Wiemer (2019) find that a positive change in the b-value relative to the background value following a sizable earthquake indicates that an aftershock sequence is underway; conversely, a negative change indicates that a foreshock sequence is ongoing. Part of the approach's apparent success is attributed to the narrow space window, measurable with high-resolution catalogs, used to probe the earthquake source process (Gulia et al. 2018). The authors proposed the following stop-light classification. Green: a 10% increase in the b-value indicates that the largest event in the region defined by the catalog has occurred and an aftershock sequence has begun; red: a 10% drop in the b-value indicates that a foreshock sequence is ongoing and a larger earthquake is yet to come; orange: a change in b-value less than 10% is ambiguous.

Consideration of spatial and temporal b-values as indicators of imminent failure is not new (Molchan, Kronrod, and Nekrasova 1999). Since its first introduction as a parametric model by Gutenberg and Richter (Gutenberg and Richter 1944), the frequency-magnitude distribution of earthquakes has been used broadly to characterize diverse seismogenic environments and probe the state of stress. Studies attribute variations in b-values to fault zone heterogeneity (Mogi 1963), stressing rate (Scholz 1968; Wyss 1973), variations in pore pressure (Wyss 1973) or ther-

moelastic stress relief in areas subject to high thermal gradients (Warren-Smith et al. 2019). The use of b-value as a predictive tool gained some traction in the 1970s with observations of b-value anomalies preceding large mainshocks (Fiedler 1974; Smith 1981). Similar observations in laboratory experiments suggested that b-value is a diagnostic of impending brittle failure (Main, Meredith, and Jones 1989). However, a low b-value preceding a large earthquake is, to some degree, self-fulfilling. A low b-value indicates a relatively high number of larger magnitude events in a sequence and, thus, experiencing subsequent large events should not be surprising (Helmstetter, Kagan, and Jackson 2005). In the absence of high-resolution seismic catalogs with low catalog completeness, closure on this topic has been elusive.

Yet Gulia and Wiemer (2019) suggest that earthquake sequences recorded with dense networks provide an informative window into a critical time of the earthquake cycle. Parametric earthquake forecasting models, some of which include time-varying b-values in a probabilistic sense, typically yield a $<15\%$ probability of a larger aftershock (Hardebeck et al. 2018; Shcherbakov et al. 2019). The 95% success rate of Gulia and Wiemer seemingly outperforms the forecasting models by a large margin and implies that a remarkable degree of measurable determinism underlies earthquake sequences.

In predictive studies of this nature, hypotheses are inevitably developed and

tested on the same dataset. Confirmation bias and overfitting are a concern. Moreover, further testing in different geological environments is necessary. Recent earthquake foreshock sequences in Ridgecrest, California, and Maria Antonia, Puerto Rico, provide an opportunity to put the method, as specified, to the test. We apply the analysis to these two sequences. The details of the implementation are in Appendix C.

3.2 Results

On 4 July 2019, an M_W6.4 earthquake ruptured previously unmapped orthogonal faults near Ridgecrest (Liu et al. 2019; Ross et al. 2019a). A very productive sequence of earthquakes followed this event (Ross et al. 2019a; Shelly 2020). Two days later, an M_W 7.1 earthquake ruptured a northwest-trending fault system (Fig. 3.1a). This sequence was well-recorded by the local network. Relocated catalogs enhanced by template matching were available promptly after the bulk of the sequence. High-resolution seismicity catalogs reveal complex faulting with multiple secondary structures of the orthogonal set on a range of scales (Fig. 3.1a, (Ross et al. 2019a; Shelly 2020)).

On 28 December 2019, M_W 4.7 and, a few hours later, M_W 5.0 earthquakes marked the start of a vigorous sequence of small earthquakes just south of Maria Antonia, Puerto Rico, with over 400 recorded events in the next ten days (Fig.

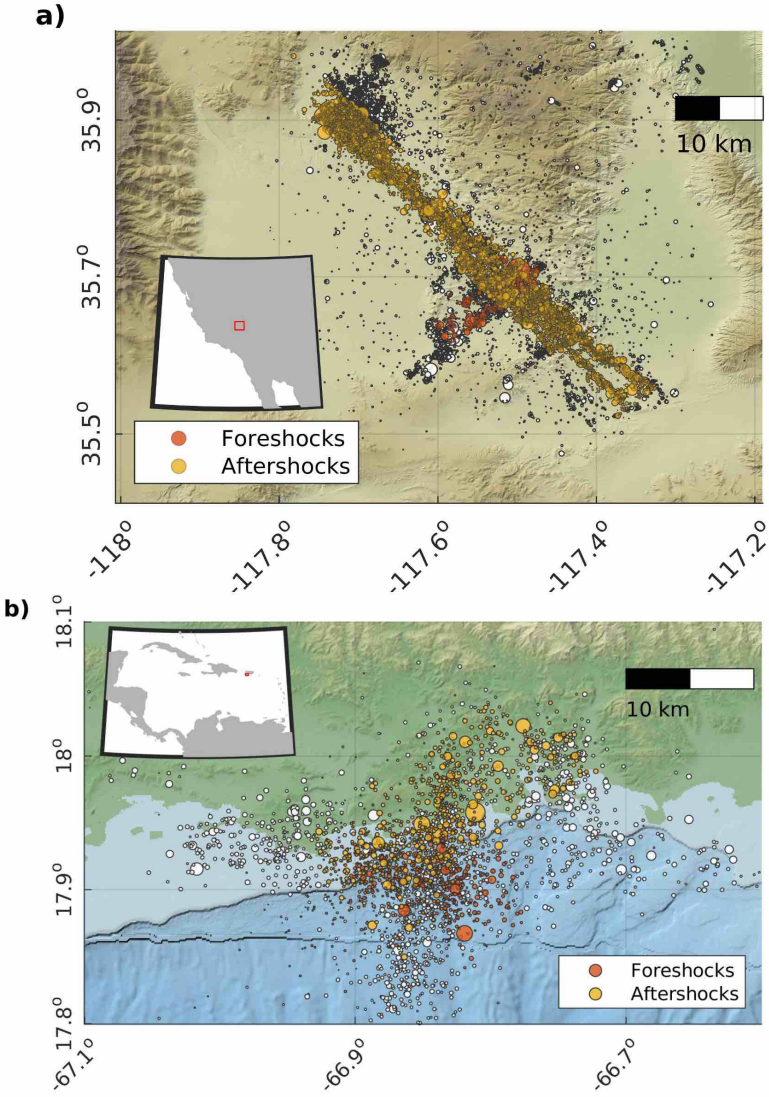

Figure 3.1: Overview of the seismic activity. a) Ridgecrest earthquake activity (Shelly 2020). b) Earthquakes from the Puerto Rico Seismic Network. Orange and Yellow events are selected for the *b*-value time series pre- and post-mainshock respectively; note this selection is established in 3 dimensions as specified in the Appendix.

3.1b). On 6 January 2020, an M_W 5.8 earthquake struck, followed the next day by an M_W 6.4 event with a normal faulting mechanism. At present, the latter event appears to be a mainshock. In the month following the mainshock, the productivity of this sequence was remarkably high (96th percentile globally when accounting for magnitude, following Dascher-Cousineau, Lay, and Brodsky 2020). The high productivity and occurrence of numerous large events before and after the mainshock suggest features approaching those of an earthquake swarm. The largest aftershock (M_W 5.9 on 11 January 2020) was half a magnitude unit smaller than the mainshock, still within the typical range for largest aftershocks (Båth 1965; Helmstetter and Sornette 2003). The sequence occurred shoreward of the Muertos Trough, a convergent zone with little recorded seismic activity (Mann et al. 2005). Shallow hypocenters, diverse focal mechanisms and diffuse seismicity indicate intraplate faulting on a complex network of faults (Liu et al. 2019).

3.3 *b*-value time series

We follow the procedure documented in Gulia and Wiemer (2019) by adapting their script to the new catalogs. This procedure is described in detail in the Appendix. We specify any deviation from the original method in the main text.

During the Ridgecrest foreshock sequence *b*-value in the selected source volume around the M_W6.4 foreshock is variable but lower than the estimated background

level of 0.9 (measured from the nearest 250 events since 2000). A stepwise increase in b-value to a level generally around background in a new source volume follows the mainshock (Fig. 3.2). The b-value then continues to oscillate above and below background for the remainder of the documented period (7-16 July 2019). Gulia and Wiemer (2019) address time-varying completeness after large events by screening out affected time windows. The calibration of this blind time relies on expert judgment. If we apply a screening after the 6.4 and 7.1 of 0.05 and 2 days respectively, the stop-light procedure outlined by Gulia and Wiemer would have indicated a red warning during the foreshock sequence changing to orange during the aftershock sequence.

For the Puerto Rico earthquake sequence, the 29 December 2019 foreshock is followed by a red warning level (Fig. 3.3). For the source region surrounding this event used for computing a b-value, we relax the nominal spatial window of 3 km from the source to 10 km in order to determine stable b-values. For this reason, the time series produced for the $M_W5.0$ foreshock is not a strict test of the method proposed by Gulia and Wiemer (2019), but is nonetheless interesting to consider. In the short time window between the $M_W5.8$ foreshock and the apparent mainshock, the b-value in the initial foreshock source volume drops further relative to background, sustaining a red warning, indicating that the activity is still a foreshock sequence. The time window between the $M_W5.8$ event and the

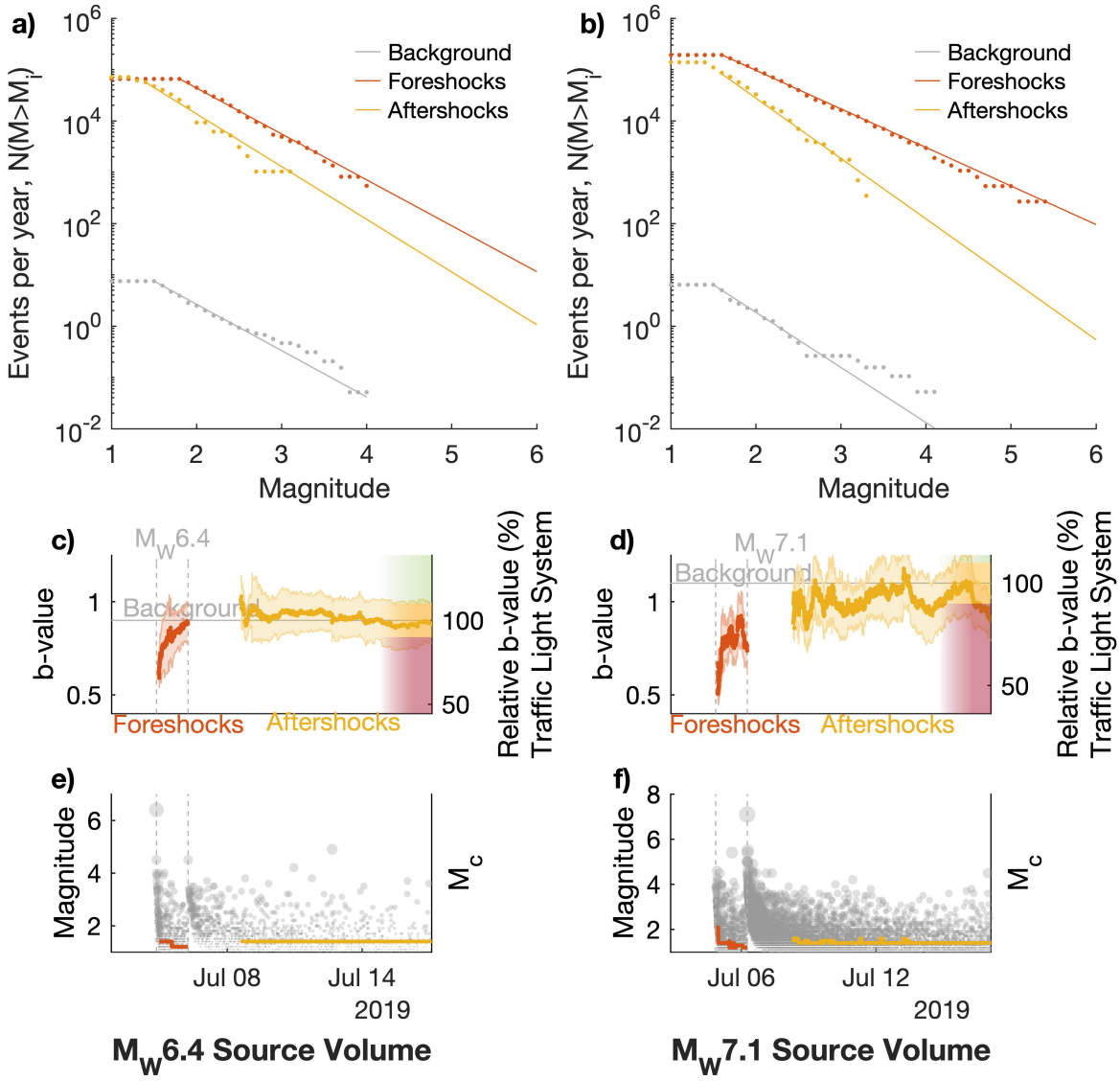

Figure 3.2: a)-b) Comparison of Gutenberg-Richter distributions for the background to and aftershock seismicity of the Ridgecrest sequence in the M_W6.4 foreshock and M_W7.1 mainshock source volumes, respectively for the time period shown in e) and f). c-d) Time series of b-value estimates during the sequence with 1σ error bars for the corresponding source volumes. Dashed lines indicate the timing of the 4 July 2019 M_W6.4 foreshock and the 5 July 2019 M_W7.1 mainshock. The traffic light criteria relative to the background level is indicated on the right. e-f) Time series of event magnitudes during the sequence in the corresponding volumes. Colored curves indicated the time-varying catalog completeness, M_c, during the intervals of the foreshock and aftershock sequences used for b-value computation.

mainshock is too short to measure an independent b-value time series in a new box around this foreshock. For its part, the time series produced for the mainshock exactly follows the proposed procedure. The sequence of events following the M_W6.4 "mainshock"exhibits a further decrease in b-value. The post-mainshock b-value has slowly drifted to larger values yet is still far below background level for the period we considered (Fig. 3.3). We estimate a high background b-value level ($b \sim 1.3$) for the M_W6.4 event source volume, which may be an unstable reference and low relative values during the sequence. However, the 50% decrease in b-value after the M_W 6.4 event relative to the earlier activity (independent of background level uncertainty) conflicts with the model expectations if this is a mainshock with ensuing aftershocks.

3.4 Discussion

For the Ridgecrest sequence, the procedure proposed by Gulia and Wiemer (2019), appears to characterize distinct behavior in the reference source volumes of the foreshock sequence ($b = 0.8 \pm 0.1$) relative to the aftershock sequence ($b = 1.0 \pm 0.1$). The aftershocks, however, are not distinctly above the background level. For the Puerto Rico sequence, the foreshock sequence is difficult to diagnose conclusively due to catalog irregularities and sparse background seismicity in a relatively small M_W5.0 source volume. Though we deviate from the

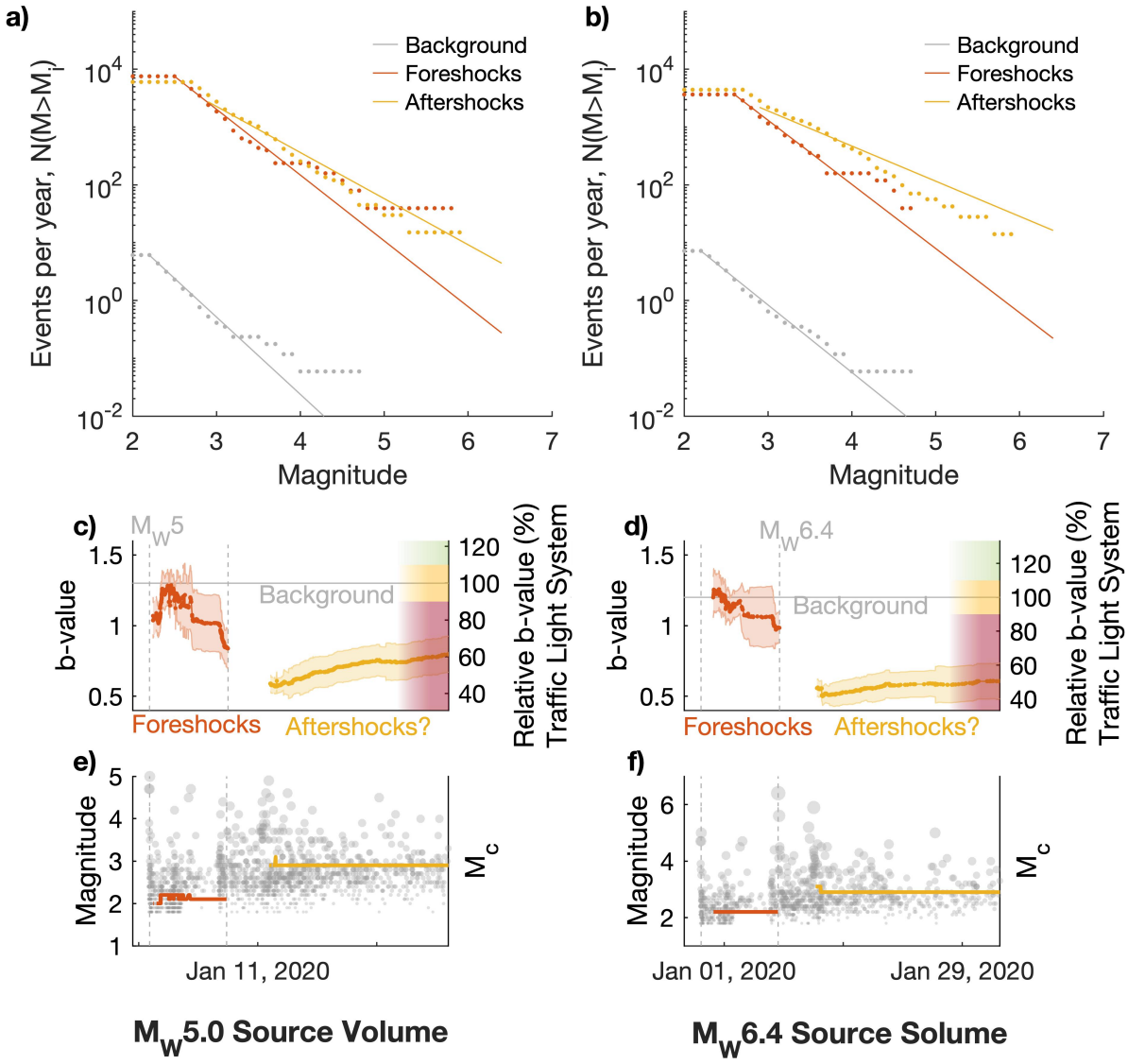

Figure 3.3: a-b) Comparison of Gutenberg-Richter distributions for the background, foreshock and aftershock seismicity of the Puerto Rico sequence in the M_W5.0 foreshock and M_W6.4 mainshock source volumes, respectively for the time period shown in e) and f). c-d) Corresponding time series of b-value estimates during the sequence with 1σ error bars. Dashed lines indicate the timing of the 29 December 2019 M_W5.0 foreshock and the 7 January 2020 M_W6.4 mainshock. The traffic light criteria relative to the background level is indicated on the right. e-f) Time series of event magnitudes during the sequence in the corresponding volumes. Lines indicated the time-varying catalog completeness, M_c, during the intervals of the foreshock and aftershock sequences used for b-value computation.

original method by considering a foreshock event with $M_W < 6.0$ and expanding the source volume, we do find a b-value drop and a red warning leading up to the mainshock agrees with the prediction. Events following the $M_W 6.4$ mainshock were distinctly vigorous and, on average, large, consistent with b-values far below background levels. As 12 February 2020, no b-value increase would indicate the transition to an aftershock sequence, and the stop-light procedure predicts that a larger event is yet to come. Time will tell; however, if no larger mainshock occurs, the b-value procedure for identifying a foreshock sequence will fail in this instance[1].

Both sequences feature volatile b-value time series, raising an issue of real-time warning levels. A single representative b-value for the foreshock sequence is not a straightforward product of the method, as stated. Real-time measurements from the 400-event moving-window would be highly erratic; conversely, the time-window used to produce the aggregated measurements shown in panels 3.2a and 3.2a can only be established retrospectively once the mainshock has occurred. How to robustly bridge this gap remains an open question.

The stop-light warnings are critically sensitive to 1) reliable estimation of background b-value in the specified source regions around the largest events, 2) time-varying catalog completeness and quality, and 3) parameterization of the

[1]Update: As of 23 June, 2021, the largest aftershock to occur remains the $M_W 5.9$ event on 11 January, 2020. With 36 $M_W 4.5+$ aftershocks the sequence is nonetheless a remarkably vigorous and persistent aftershock sequence.

b-value measurement. For instance, during the review of this work, we learned that Gulia and Wiemer obtained results on Ridgecrest using the same catalogs and an adaptation of the published code (Gulia, Wiemer, and Vannucci 2020). Discrepancies in the resulting time-series arise from the decisions left to expert judgment that must be attuned to the individual circumstances of each earthquake. The need to pick one of two possible fault planes, when both appeared to have ruptured, further complicates the implementation.

To illustrate the importance of these factors, we performed a Monte Carlo simulation that randomly sampled a reasonable range of possibilities within the method's scope. The time series shown in Fig. 3.4 illustrate how minor differences impact warning for the Ridgecrest sequence. For this analysis, we uniformly sampled the range of reasonable source volume (orthogonal planes ruptured in the foreshock), assumed completeness threshold correction (0.1 to 0.3), background catalog start date (2000-2012) and blind times following the foreshock (0.01-0.5 days) and mainshock (0.5-5 days). The variability introduced by these decisions exceeds the standard deviation of the measured b-values (Shi and Bolt 1982). The relative b-value of the Ridgecrest mainshock generally rises after the mainshock with a modal value of a 10% increase (Fig. 3.4c). Yet from the set of 1000 Monte Carlo simulations, 7.4% of the time series correctly identify the foreshock and aftershock sequence, 76.7% provide neutral assessments that do not contradict the

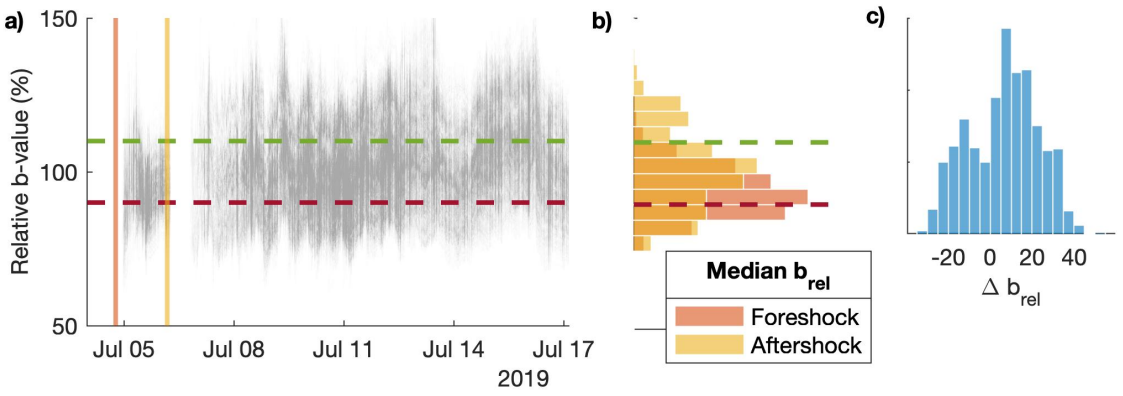

Figure 3.4: a) Monte Carlo simulation of 1000 summary time series of the relative b-value. Each individual time series samples uniform priors for the background catalog start date (2000-2012), foreshock source volume choice (choosing one or the other of the two orthogonal planes ruptured in the foreshock), M_c maximum curvature correction (0.1-0.3) and blind times after the foreshock (0.01-0.5 days) and mainshock (0.5-5) days. Warning thresholds are indicated by the horizontal dashed lines. For all calculations, the source volume and corresponding background value is updated at the mainshock. b) Corresponding histogram of the median relative b-value of the foreshock and aftershock periods. c) Distribution of the change in the median value of the relative b-value after the M_W7.1 mainshock across all the realizations.

observed outcome. The correct identifications are possible, perhaps even preferable, given careful expert judgment in the specification of parameters, but are not a representative outcome of allowable decisions.

Both sequences did not rupture plate boundary faults. Correspondingly low background seismicity causes unstable estimates of background b-values and, ultimately, warning levels. Should there not be enough events in the 3 km wide source volume (< 250), Gulia and Wiemer (2019) measure the background b-values using the nearest 250 events leading up to the sequence. Background b-values can

thus be sensitive to anomalous seismicity clusters far from the source volume. In the Ridgecrest region, for example, background activity established far from the source volume is influenced by anomalous seismicity in the Coso volcanic field; background activity considering events too far back in time is influenced by changes in the detection capabilities of the seismic network. Production of stable regionalized b-value maps to serve as references for temporal changes during future large event sequences is a requirement for this general approach.

During highly productive earthquake sequences, a combination of technical and logistical factors results in unstable and time-varying catalog completeness. Extraneous factors, including analyst overwhelm (person-power), and damaged infrastructure, may cause b-values to fluctuate as an artefact of varying catalog completeness. Gulia and Wiemer (2019) partially addressed this issue by reevaluating the catalog completeness at each windowing step and removing events in the immediate wake of larger foreshocks and mainshocks based on expert judgement. The impact of large events on catalog completeness varies with time and regional network operations. The maximum likelihood estimation of the b-value is most sensitive to small events, which are likely incomplete beyond the maximum curvature threshold used to infer completeness.

The limitations outlined above are practical. Method development and the dedication of resources during ongoing sequences would mitigate these problems.

The Puerto Rico sequence highlights other potential limitations. The prediction that the sequence is a foreshock sequence is not yet fulfilled. Is this an indication that the method is not as deterministic as originally envisioned? Or is it a peculiarity of a complex tectonic environment with unusually swarm-like seismicity? How the method applies to different environments and types of earthquake sequences must still be established.

3.5 Conclusion

The procedure of Gulia and Wiemer (2019) offers the potential to recognize foreshock sequences in real-time. The Ridgecrest sequence roughly conforms to expectations but is not a definitive success. The Puerto Rico sequence appears to not follow the predicted behavior but could potentially do so in the future if a larger event occurs with an ensuing b-value increase. We find that variable warning levels result from subtle differences in catalog production and model parameterization left to expert judgment. Future implementation of this approach will require robust decision-making, well-calibrated maps of regionalized background b-values for a large distribution of possible source volumes, consistent and version-controlled low completeness catalogs during sequences, dense instrumentation and systematic real-time catalog production.

3.6 Comment and Reply

The publication of this chapter was concurrent with a pseudo prospective evaluation of the foreshock traffic-light system in Ridgecrest by Gulia, Wiemer and Vannucci (2020). In contrast to our analysis and our preferred results, the authors obtained a red light in the foreshock sequences, and a green light in the aftershock sequence. The results were shown to be robust to the magnitude cutoff and blind time. A subsequent comment on our publication and our reply discussed how the analyses of the same sequence deviated from each other, and, more importantly, the original method (Gulia and Wiemer 2019, as stated in). Our reply is included in the Appendix C). Gulia and Wiemer (2021) revealed six potential ways in which the replication of their analysis may have deviated from the proposed method. We showed that for four of the six claimed deviations, our analysis conformed to the criteria outlined in Gulia and Wiemer 2019), and that, in some cases, the criticisms are in direct contradiction with the guidelines in Gulia and Wiemer 2019. There are two true deviations from the defined procedure that we should have better articulated. We explain and discuss the rationale for these deviations. One attempts to reconcile the code distributed by Gulia and Wiemer 2019 with the documentation. The other stems from a decision to encompass a volume that was robust to uncertainty in early hypocenter depths, as would be required for a real time application. Accounting for these true deviations yield a

red warning during both the foreshock and aftershock sequences Fig. C.2. The range in results obtained by the various analyses underscores the relatively large epistemic uncertainty involved the production of these results (e.g. Fig. 3.4. We again emphasize the influence of expert judgment in the analysis.

3.7 Data and Resources

Figures and analysis were all produced using MATLAB ver. R2019b. Maps figures were specifically produced using Topotoolbox and used topographic data from the Global Multi-Resolution Topography (GMRT) Data Synthesis and ALOS Global Digital Surface Model (AW3D30) hosted on OpenTopography. Earthquake event catalogs for Ridgecrest were from the ANSS Comprehensive Earthquake Catalog (ComCat) and augmented by Shelly (2019) during the sequence itself. Earthquake event catalogs for Puerto Rico were downloaded from the Puerto Rico Seismic Network. All the data and code used for this analysis are provided here:

https://github.com/keliankaz/DC2020_reply/tree/main/DC2020

The analysis for our subsequent reply is available here:

https://github.com/keliankaz/DC2020_reply/tree/main/COMMENT

Chapter 4

Flexible and Scalable Earthquake Forecasting

4.1 Introduction

Seismology is witnessing an explosive growth in the diversity and scale of earthquake catalogs owing to improved seismic networks and increasingly automated workflows (Obara 2003; White, Ben-Zion, and Vernon 2019; Ross et al. 2019b; Shelly 2017; Tan et al. 2021). A key assumption in this community effort is that more detailed observations should translate into improved earthquake forecasts. Current operational earthquake forecasts build on seminal work designed for sparse earthquake records (Reasenberg and Jones 1989; Ogata 1988)

that combines the canonical statistical laws of seismology (Ogata 1988; Gutenberg and Richter 1944; Utsu 1970). This parsimonious approach is remarkably robust and stubbornly difficult to improve upon (Beroza, Segou, and Mostafa Mousavi 2021). Advances in the past decades have mainly focused on the regionalization of the models (Page et al. 2016; Ogata 2017; Field et al. 2017), the recognition of catalog peculiarities (Ross 2021; Mizrahi, Nandan, and Wiemer 2021) and the extension to spatial forecasts (Ogata 1998); but have failed to leverage the wealth of new geophysical data. Here, we develop a neural-network based earthquake forecasting model that leverages the new data in an adaptable forecasting framework: the Recurrent Earthquake foreCAST (RECAST). We benchmark temporal forecasts generated by RECAST against the widely used Epidemic Type Aftershock Sequence (ETAS) model using both synthetic and observed earthquake catalogs. We consistently find improved model fit and forecast accuracy for Southern California earthquake catalogs with more than 10^4 events. The approach provides a flexible and scalable path forward to incorporate additional data into the earthquake forecast.

4.2 RECAST model description

RECAST builds on recent developments in machine learning known as neural temporal point processes (Du et al. 2016a; Omi, and Aihara 2019; Shchur, Biloš,

and Günnemann 2020; Shchur et al. 2021; Zhou et al. 2021). The model uses a general-purpose encoder-decoder neural network architecture that predicts the origin time of the next event given the history of past events (Fig. 1.1A) (Du et al. 2016a; Omi, and Aihara 2019; Shchur, Biloš, and Günnemann 2020; Zhou et al. 2021; Upadhyay, De, and Gomez Rodriguez 2018; Mei and Eisner 2017). RECAST is based on the Gated Recurrent Unit neural network architecture (Cho et al. 2014) that encodes the variable-length history of past earthquakes, $\mathcal{H}(t_i) = \{(t_1, M_1), ..., (t_{i-1}, M_{i-1})\}$, into a fixed-dimensional hidden state vector $\boldsymbol{h}_i$. The hidden state ensures that the influence of previous events can be carried forward in the model. Next, the RECAST decoder uses $\boldsymbol{h}_i$ to parametrize the probability density function of the next origin time t_i,

$$f_{RECAST}(t_i|\mathcal{H}(t_i)) = f(t_i|\boldsymbol{h}_i). \tag{4.1}$$

We model the probability density function of the next event origin time with a Weibull Mixture Distribution. This choice results in a model where training and simulation can be done efficiently and exactly. Moreover, a mixture distribution with a sufficiently large number of components is able to approximate any other probability distribution arbitrarily well (Shchur, Biloš, and Günnemann 2020). Note that the RECAST output is equivalently expressed continuously in time in terms of the conditional intensity function $f(t_i|\mathcal{H}(t))$ commonly used in point process literature (see Methods for further implementation details). The model'

s architecture and hyperparameters follow the more general application in Eq. (Shchur, Biloš, and Günnemann 2020), without explicit tuning for the earthquake forecasting task (see Methods for the full model description).

4.3 Results

As a benchmark, we model the recurrence of earthquakes with an Epidemic Type Aftershock Sequence (ETAS) model (Fig. 1.1B), which is a standard statistical model that calculates the time-dependent earthquake intensity $\lambda(t_i|\mathcal{H}(t))$ as a combination of a steady background rate and parameterized aftershock sequences (Ogata 1988; Aalen, Borgan, and Gjessing 2008). All earthquakes produce aftershocks and thus the current earthquake rate is an explicit function of the entire event history $\mathcal{H}$. Bearing similarity to Eq. 1.1, ETAS yields a probability density function for the next event origin t_i,

$$f_{ETAS}(t_i|\mathcal{H}(t_i)) = \lambda(t_i|\mathcal{H}(t_i))e^{-\int_{t_{i-1}}^{t_i}\lambda(t_i|\mathcal{H}(s))ds}. \tag{4.2}$$

RECAST and ETAS are both temporal point process models. They are trained by maximizing the joint log-likelihood of the cataloged origin times using the same optimization procedure (See Methods). However, RECAST introduces practical benefits. In this comparative study, we only consider origin times and magnitudes as model inputs, but additional event features, such as locations, source proper-

ties, and geophysical data, can be seamlessly introduced into the model. More generally, the model's components are modular and easily adapted using standard machine learning frameworks (Paszke et al. 2019). Another key distinction is that RECAST processes events sequentially, summarizing the event history into a fixed-dimensional vector (Fig. 4.1A). In contrast, ETAS references all previous events to determine the intensity function and, in turn, the event likelihood (Fig. 4.1B). As a result, the computational time and space complexity for evaluating the likelihood of a catalog is linear for RECAST and quadratic for ETAS (Fig. 4.1C). Quadratic growth is computationally cumbersome or potentially prohibitive for the real-time evaluation of earthquake forecasts given earthquake catalogs with more than 10,000 events. This limitation is mitigated in the RECAST model that can be trained on catalogs with more than 1,000,000 events on a single consumer-grade GPU (Fig. D.1).

As a first step, we assess whether the chosen model architecture is well suited to learn known features within earthquake catalogs. We evaluate the performance of RECAST on a synthetic catalog generated by a prescribed ETAS model. In this case an ETAS model achieves the highest possible expected joint log-likelihood on the out-of-sample sequences since it is the generative model. If RECAST's performance approaches this limit, it would demonstrate that the model can be trained to capture realistic earthquake temporal clustering from the event data

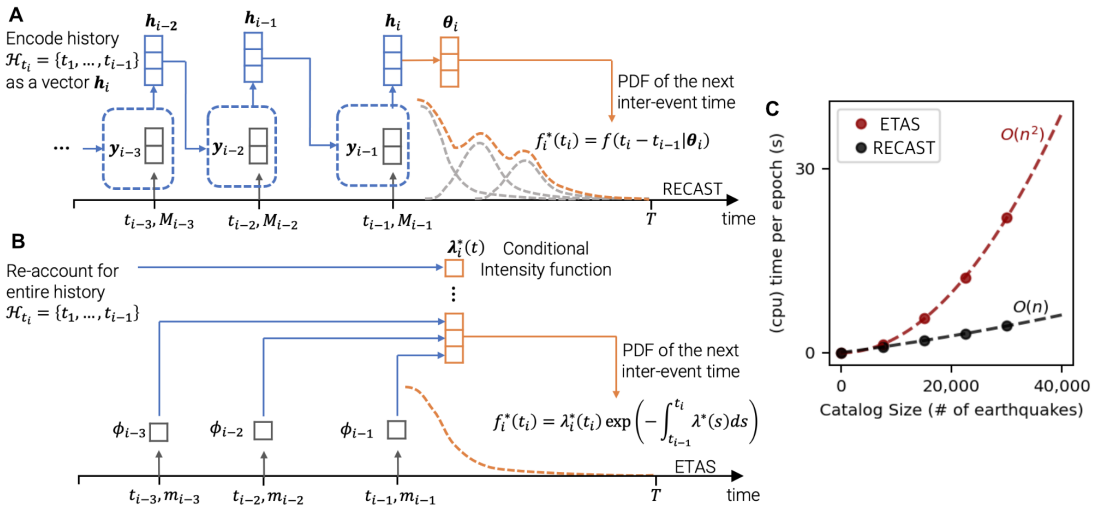

Figure 4.1: Structurally alike, conceptually distinct earthquake forecasting models. (A) Model architecture for the RECAST model. For each likelihood evaluation, the network takes as input the previous hidden state h_{i-1} with features h_{i-1} of the previous event (t_{i-1}, M_{i-1}) and outputs an updated hidden state h_i given the gated recurrent unit update equations. An affine layer connects the hidden state to the three parameters of each Weibull Mixture component which in turn determine the model likelihood given an event t_i (B) Model architecture for the ETAS model. ETAS parametrizes the probability density function of the next origin time given the history of previous events. See methods for extended model descriptions (C) The computational time and space complexity of evaluating the log-likelihood of an entire earthquake catalog for RECAST is linear, $O(N)$. In contrast, the complexity for the ETAS model is quadratic, $O(N^2)$.

alone. Specifically, we evaluate the model performance on a set of synthetic ETAS-generated catalogs split into training, validation, and test sets. Each catalog is a random realization that spans 10,000 days and contains around 1,000 events on average. We train and evaluate both models on the same set of synthetic ETAS-generated sequences. RECAST successfully captures sharp fluctuations in the underlying conditional intensity function associated with Omori's law (Fig. 4.2A-B). The goodness of fit, over both the training and validation set, approaches the performance of ETAS (Fig. 4.2C). Without narrowly tailoring the RECAST architecture for the task, the model independently recovers the classical statistical laws of seismology that compose ETAS, and does so from event data alone using magnitudes as an additional event feature. The architecture for the marks is general and thus the successful recovery of standard magnitude-based behavior is demonstration of the effectiveness of the approach to incorporating additional inputs into the model.

We next consider an earthquake catalog from 2008 to 2021 (White, Ben-Zion, and Vernon 2019) bounding the San Jacinto fault (Fig. 4.3A). This catalog makes a good test case because of its dense station-coverage of a particularly seismically active area (See methods for more detail on data pre-processing). We find that RECAST outperforms ETAS in terms of goodness-of-fit in the reserved test period (Fig. 4.3B). We gain a sense for the volume of data required to outperform ETAS

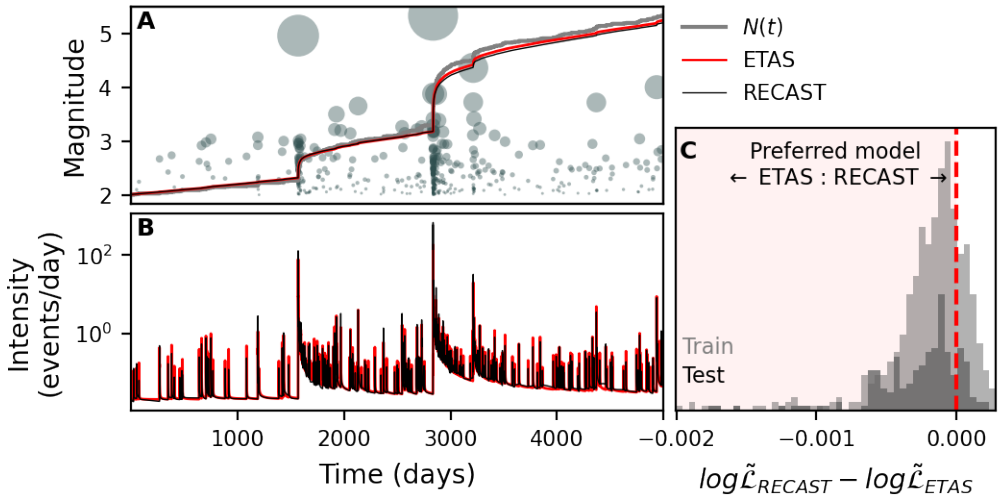

Figure 4.2: RECAST performance on synthetic earthquake catalogs. (A) Example of a synthetic earthquake catalog generated from an ETAS model in the test set. Events magnitudes and origin times are indicated by the gray markers with the gray line tracking the cumulative number of events, $N(t)$. The red (ETAS) and black (RECAST) lines indicate the integral of the conditional intensity function in (B), which corresponds to the expected number of events for each model. (C) For any particular catalog $\log \widetilde{\mathcal{L}}_{RECAST} - \log \widetilde{\mathcal{L}}_{ETAS}$ greater than zero implies that RECAST performed better than ETAS (and vice versa). In most cases, $\log \widetilde{\mathcal{L}}_{RECAST}$ approaches the $\log \widetilde{\mathcal{L}}_{ETAS}$ model, an upper bound in expectation since ETAS is the generative model.

Table 4.1: Summary statistics of the catalogs. The specified dates indicate the intervals used for training, validation, and testing (Fig. 4.4).

Dataset name	**Catalog start**	**Catalog end**	**Number of events**	**Train NLL start**	**Validation Start** NLL	**Test NLL start**	M_c
White et al.	2008-01	2021-01	134,975	2009-01	2014-01	2017-01	0.6
SCEDC	1981-01	2020-01	125,421	1985-01	2005-01	2014-01	2
QTM Salton Sea	2008-01	2018-01	44,133	2009-01	2014-01	2016-01	1
QTM San Jacinto	2008-01	2018-01	20,790	2009-01	2014-01	2016-01	1

by incrementally extending the training period backward in time while maintaining the same testing period. ETAS is the preferred model when trained on fewer than 10,000 events; extending beyond 10,000 events, RECAST outperforms ETAS. Comparable improvements arise when considering the entire Southern California Earthquake Data Center catalog from 1981 to 2021 (Hutton, Woessner, and Hauksson 2010), and subregions of the Quake Template Matching Catalog (Ross et al. 2019b) (Fig. 4.4, D.3-D.5, Table 4.1). Consistent out-of-sample improvements on these disparate catalogs suggest that RECAST is robust to multiple methods of data production and regional variations in seismicity. Importantly, there is signal discovered by RECAST implying a greater degree of predictability than is currently captured in aftershock forecasting methods.

We finally compare earthquake forecasts on a fixed interval in time. Once trained, RECAST provides a straightforward way to simulate potential catalog continuations. We can generate a sample earthquake by drawing an origin time from the mixture distribution in Eq. 4.1 and a magnitude from the Gutenberg–Richter distribution (Gutenberg and Richter 1944). Repeatedly adding events

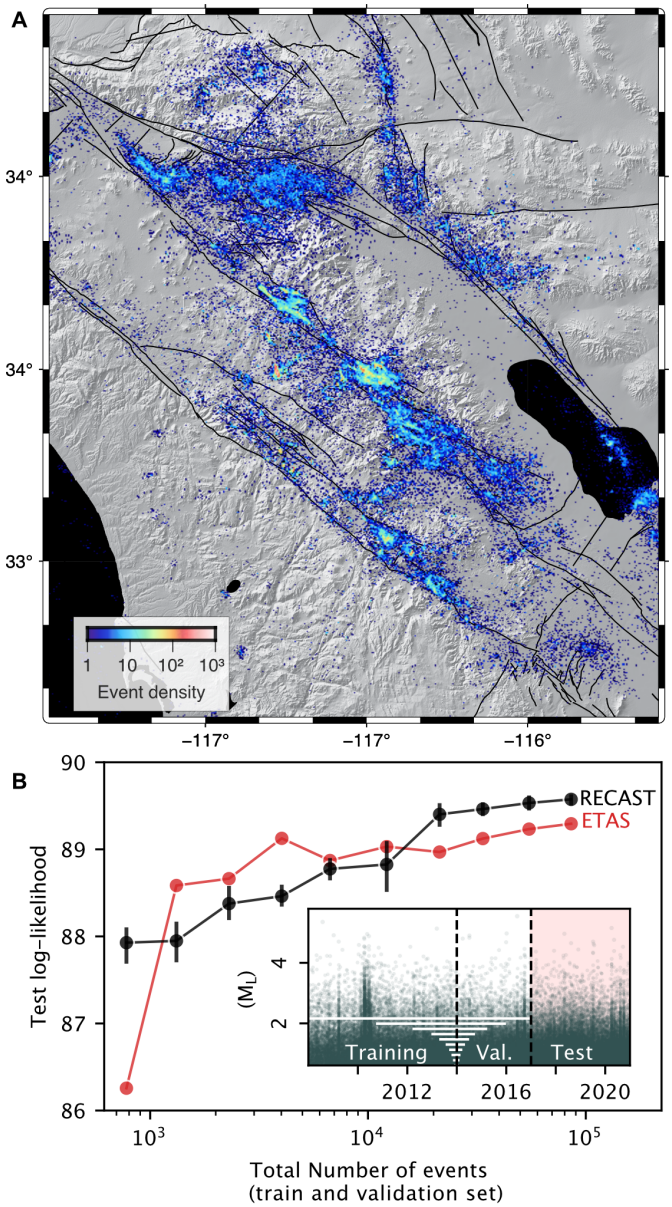

Figure 4.3: Performance on modern earthquake catalogs. (A) Seismicity around the San Jacinto Fault zone in Southern California (White, Ben-Zion, and Vernon 2019) (B) Model relative goodness of fit on the out-of-sample test period as measured by the time-averaged joint log-likelihood for RECAST and ETAS model. Error bars indicate the 95% confidence interval of 1000 bootstrap samples for five random initializations of RECAST. The models were trained with incrementally longer training and validation sets. The inset shows the time series for the seismicity considered. Each white bar shows the corresponding periods used to train and validate ETAS and RECAST. Given a training and validation catalog in excess of $\sim 10,000$ earthquakes (fourth white bar from the top in the inset), the test period is best modeled by RECAST.

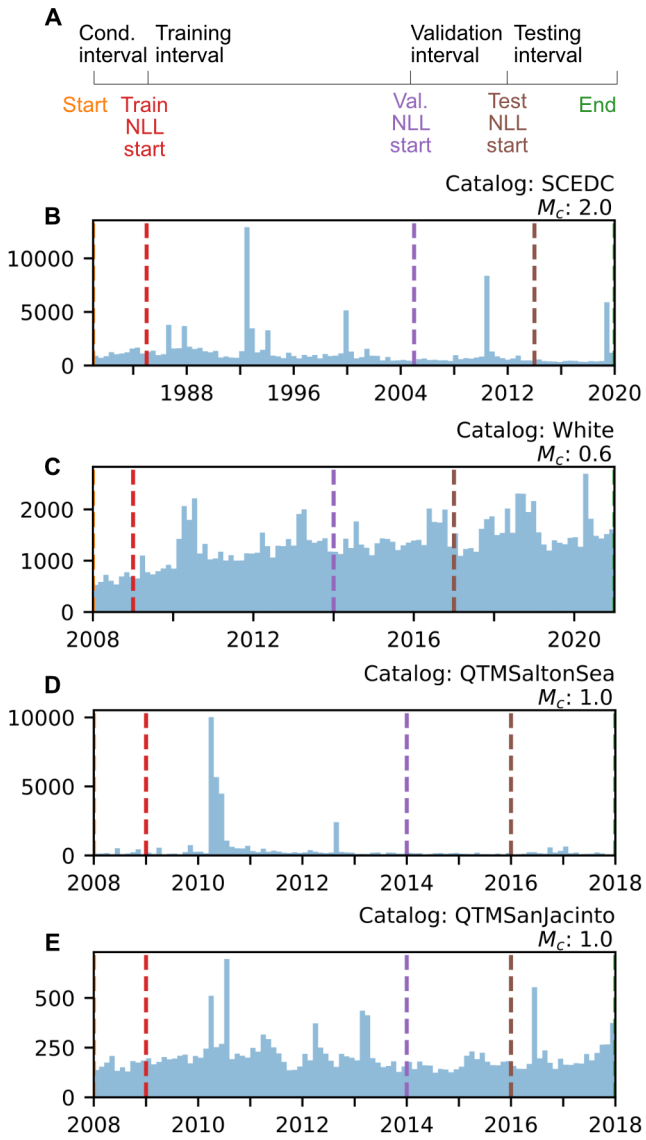

Figure 4.4: Earthquake catalogs considered. (A) Catalogs are split into a conditioning, training, validation and test intervals. The Negative Log-likelihood is evaluated on the training, validation and test intervals. NLL on the training interval [Train NLL start, Val. NLL start] is used to optimize model parameters; NLL on the validation interval [Val. NLL start, Test NLL start] is used to perform early stopping; and NLL on the test interval [Test NLL start, End] is used for comparing the models. In all cases, the NLL is conditioned on all the past events. For example, we condition the NLL of the training interval on events from [Start, Train NLL start] to avoid orphaned events (Elst 2017). We consider seismicity rate for (B) the entire SCEDC29, (B) the San Jacinto fault zone, (C) the QTM Salton Sea area, and (D) the QTM San Jacinto fault zone catalogs. Note that in the SCEDC, the conditioning and training period feature a higher background rate of earthquakes. See Table 4.1 for further detail.

to the catalog yields continuations spanning a range of potential outcomes. We refer to this set as an earthquake forecast. Specific aspects of a forecast such as the abundance of events or the probability of a particularly large earthquake can then derived from statistics of the set of catalog continuations. We consider two-week earthquake forecasts with the Southern California earthquake catalog (Hutton, Woessner, and Hauksson 2010) with an additional set of results for the San Jacinto fault catalog in the Extended Data. Each two-week forecast comprises 50,000 simulated catalog continuations during the testing period. Individual continuations exhibit realistic features, including aftershock decay, secondary bursts of aftershocks and stochasticity (Fig. 4.5A). To quantitatively compare the RECAST and ETAS forecast, we compute the fraction of simulated continuations, r, that reproduce the total observed number events in the interval (Fig. 4.5B and Fig. 4.7B). A forecast is said to be more accurate if r is greater (see Methods for detailed definition). The vast majority of two-week intervals in the test set were best forecast by RECAST (138 out of 155 14-day intervals, Fig. 4.5C-D, 4.7C-D). Periods best fit by RECAST tend to be background periods. Nonetheless, RECAST outperforms ETAS following the largest earthquakes as well. Another metric of success is the ability to accurately capture the full range of potential outcomes. Outcomes outside the 95% confidence interval of the forecasts are far less frequent for RECAST relative to the ETAS model (14% vs. 33%, respectively,

Fig. 4.6, 4.8). Outcomes more extreme than the entire set of 50,000 simulations occurred only once for the RECAST forecasts and 10 times for the ETAS forecasts (Fig. 4.5C). Full misses for ETAS cluster during the Ridgecrest earthquake aftershock sequence. While RECAST certainly does not foresee the Ridgecrest sequence, it quickly recovers relatively accurate forecasts after the onset of the sequence.

4.4 Conclusion

Results consistently indicate improved performance in terms of goodness of fit and extended forecasts. RECAST therefore captures signals inaccessible to ETAS (Fig. 4.3, 4.5). Given the real datasets we consider, we interpret ETAS underfits event sequences (e.g. Fig. 4.8) and is inflexible to long-term trends in seismicity (e.g. 4.6). In addition, its forecasts inherit the characteristics of the training set and are biased accordingly (Fig. 4.5, 4.4). Though RECAST uses the same training periods, updates to the hidden state allow the model to carry information forward within the validation and test sequences. The model's memory plays a critical role allowing RECAST to account for time-varying trends in the validation and test sets and, as a result, perform better.

Dense earthquake catalogs reveal a richer window into the seismic cycle. It has been a challenge to translate finer detail and increased data volume into im-

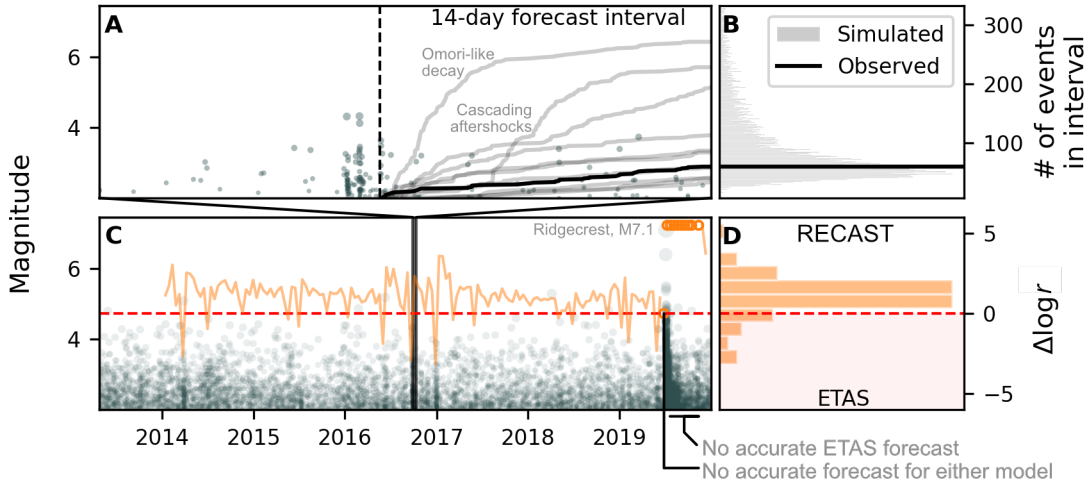

Figure 4.5: Evaluation of earthquake forecasts with RECAST and ETAS models. (A) Example two-week earthquake RECAST forecast issued at the vertical dashed line. The observed cumulative number of earthquakes is shown in black with sample RECAST simulations in gray. Punctuated increases in the cumulative number of events in these samples indicate spontaneous event clusters which gradually decay in time. (B) Full distribution of the cumulative number of events from the sampled trajectories compared to the observation (black). (C) Test catalog with the evolution of the log-score for the tested 14-day forecast intervals. Empty orange markers and the corresponding annotation indicate the intervals where both model forecasts (1 instances), or just the ETAS model forecast (9 instances) yielded no accurate forecast ($r = 0$ and $\log r$ is undefined). These all cluster during the Ridgecrest earthquake sequence. (D) Comparison of the relative accuracy (forecast log-likelihood) of RECAST and ETAS models. Positive log-score indicates a more accurate RECAST forecast. In most intervals, the RECAST is more accurate.

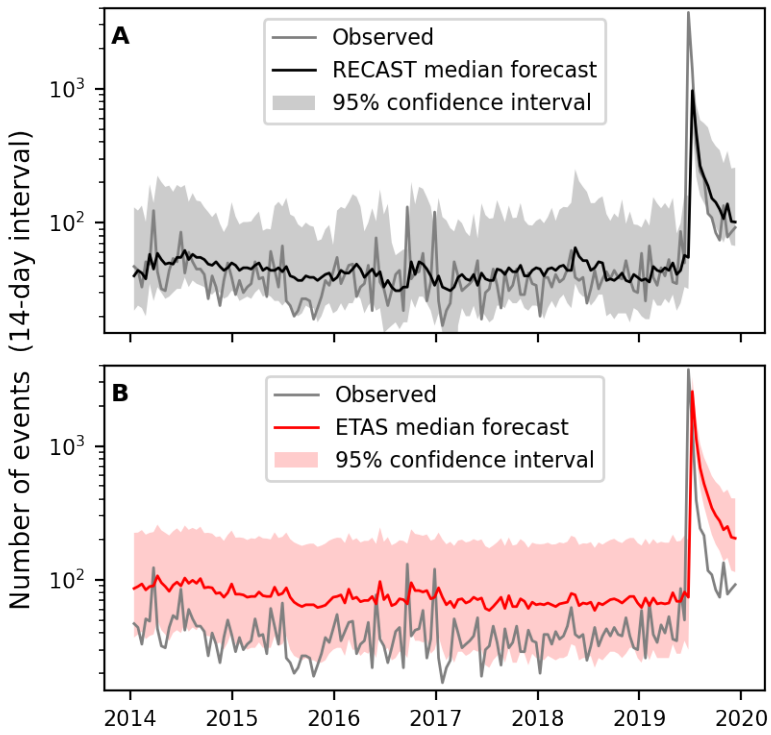

Figure 4.6: Comparison of 14-day forecasts during the test period for the SCEDC catalog given RECAST (A) and ETAS (B) models. Note that ETAS inherits the abnormally elevated rate of events that characterize the training set (Fig. 4.4).

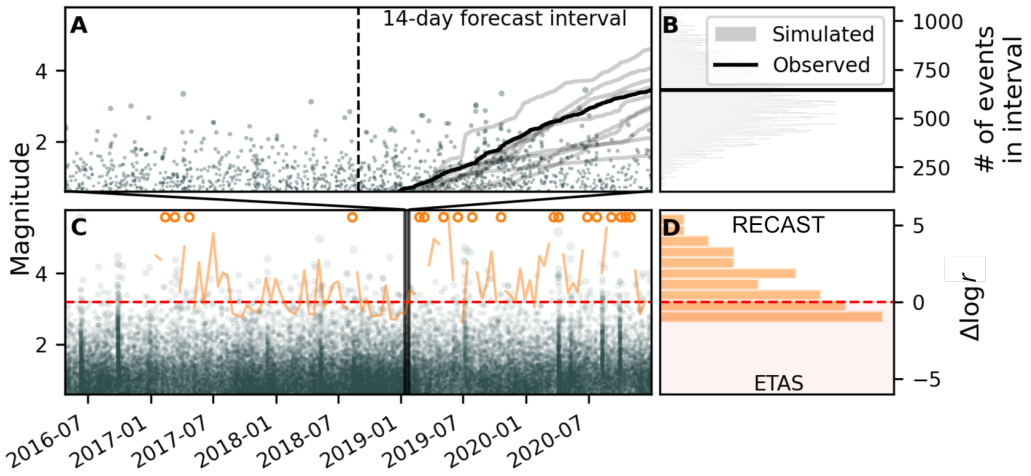

Figure 4.7: Simulated trajectories over forecast intervals and evaluation of earthquake forecasts for the White et al., 2019, earthquake catalog. (A) Example two-week earthquake RECAST forecast issued at the vertical black line. The observed cumulative number of earthquakes is shown in black with example RECAST simulations in gray. (B) Full distribution of the cumulative number of events from the sampled trajectories compared to the observation (black). (C) Test catalog with the evolution of the log-score for the tested 14-day forecast intervals. Circles indicate one or both models yielded no simulated trajectories consistent with the observed seismicity; top (18 instances), bottom (0 instances) and middle (0 instances) respectively correspond to complete ETAS, RECAST and collective misses. (D) Comparison of the relative accuracy of RECAST and ETAS. Log-score greater than zero indicates a more accurate RECAST forecast. In most intervals, RECAST is more accurate.

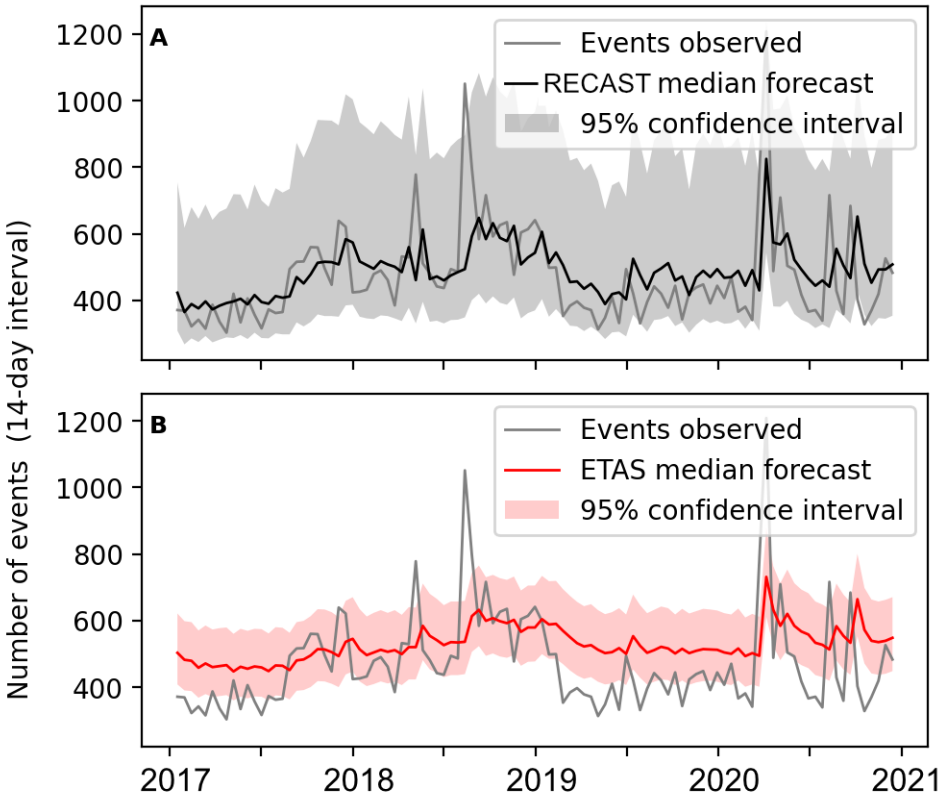

Figure 4.8: Comparison of 14-day forecasts during the test period for the San Jacinto (White, Ben-Zion, and Vernon 2019) catalog given RECAST (A) and ETAS (B) models.Observations fall outside the 95th percent confidence interval of RECAST and ETAS respectively 3 (6%) and 56 (54%) of the 103 intervals considered. Note that the ETAS mode is both overconfident and inflexible to the fluctuations in seismicity.

proved earthquake forecasts. A neural network architecture such as RECAST is sufficiently general to independently recover well-known statistical patterns on synthetic catalogs. The model outperforms ETAS on large-scale real-world catalogs indicating the existence of signal elusive to standard forecasts but discoverable by more flexible forecasting frameworks. RECAST, and more generally neural temporal point process models complement and extend current capabilities in earthquake forecasting. It remains to be seen whether the increase in performance is due to more accurately capturing the observational limitations of catalogs or reflects a physical process that controls long-term seismicity evolution. In either case, forecasters can now capitalize on the production of augmented catalogs by leveraging the flexibility and scalability of neural networks.

Bibliography

Aalen, Odd, Ornulf Borgan, and Hakon Gjessing (Sept. 2008). *Survival and Event History Analysis: A Process Point of View*. Springer.

Adams, Mareike, Cedric Twardzik, and Chen Ji (Jan. 2017). "Exploring the uncertainty range of coseismic stress drop estimations of large earthquakes using finite fault inversions." *Geophysical Journal International* 208.1, pp. 86–100.

Arcangelis, Lucilla de et al. (Apr. 2016). "Statistical physics approach to earthquake occurrence and forecasting." *Physics Reports* 628, pp. 1–91.

Arrowsmith, J. Ramón, Dallas D. Rhodes, and David D. Pollard (May 1998). "Morphologic dating of scarps formed by repeated slip events along the San Andreas Fault, Carrizo Plain, California." *Journal of Geophysical Research: Solid Earth* 103.B5, pp. 10141–10160.

Båth, Markus (Dec. 1965). "Lateral inhomogeneities of the upper mantle." *Tectonophysics* 2.6, pp. 483–514.

Ben-Zion, Yehuda and Vladimir Lyakhovsky (2006). "Analysis of aftershocks in a lithospheric model with seismogenic zone governed by damage rheology." *Geophys. J. Int* 165.

Beroza, Gregory C., Margarita Segou, and S. Mostafa Mousavi (Aug. 2021). "Machine learning and earthquake forecasting—next steps." *Nature Communications* 12.1, p. 4761.

Bevis, M.G. et al. (Dec. 2017). *Classified point cloud and gridded elevation data from the 2005 B4 Lidar Project, southern California, USA - ScienceBase-Catalog.*

Bird, Peter (Mar. 2003). "An updated digital model of plate boundaries." *Geochemistry, Geophysics, Geosystems* 4.3.

Boettcher, Margaret S and T. H. Jordan (2004). "Earthquake scaling relations for mid-ocean ridge transform faults." *Journal of Geophysical Research: Solid Earth* 109.12, pp. 1–21.

Bouchon, Michel and Hayrullah Karabulut (2008). "The Aftershock Signature of Supershear Earthquakes." *Science* 320, pp. 1323–1325.

Bouchon, Michel et al. (Apr. 2013). "The long precursory phase of most large interplate earthquakes." *Nature Geoscience* 6.4, pp. 299–302.

Brengman, Clayton M. J. et al. (Oct. 2019). "Earthquake-Scaling Relationships from Geodetically Derived Slip Distributions." *Bulletin of the Seismological Society of America* 109.5, pp. 1701–1715.

Brodsky, Emily E (2011). "The spatial density of foreshocks." *Geophysical Research Letters* 38.10.

Castelltort, Sebastien and Jean Van Den Driessche (Mar. 2003). "How plausible are high-frequency sediment supply-driven cycles in the stratigraphic record?" *Sedimentary Geology* 157.1-2, pp. 3–13.

Cho, Kyunghyun et al. (Sept. 2014). "Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation." *arXiv:1406.1078 [cs, stat]*.

Chu, Annie et al. (2011). "Comparison of ETAS Parameter Estimates across Different Global Tectonic Zones." *Bulletin of the Seismological Society of America* 101.5, pp. 2323–2339.

Clauset, Aaron, Cosma Rohilla Shalizi, and M. E.J. J Newman (2009). "Power-law distributions in empirical data." *SIAM Review* 51.4, pp. 661–703.

Convers, J A and A V Newman (2011). "Global evaluation of large earthquake energy from 1997 through mid-2010." *J. Geophys. Res* 116.

Das, Shamita and C. Henry (Sept. 2003). "Spatial relation between main earthquake slip and its aftershock distribution." *Reviews of Geophysics* 41.3, p. 1013.

Dascher-Cousineau, Kelian, Thorne Lay, and Emily E. Brodsky (2020). "Two foreshock sequences post gulia and wiemer (2019)." *Seismological Research Letters* 91.5, pp. 2843–2850.

Davidsen, J., C. Gu, and M. Baiesi (May 2015). "Generalized Omori–Utsu law for aftershock sequences in southern California." *Geophysical Journal International* 201.2, pp. 965–978.

Davis, Scott D. and Cliff Frohlich (Apr. 1991). "Single-link cluster analysis of earthquake aftershocks: Decay laws and regional variations." *Journal of Geophysical Research: Solid Earth* 96.B4, pp. 6335–6350.

Du, Nan et al. (Aug. 2016a). "Recurrent Marked Temporal Point Processes: Embedding Event History to Vector." *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Francisco California USA: ACM, pp. 1555–1564.

— (Aug. 2016b). "Recurrent Marked Temporal Point Processes: Embedding Event History to Vector." *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '16. New York, NY, USA: Association for Computing Machinery, pp. 1555–1564.

Duvall, Alison R. and Gregory E. Tucker (2016). "Dynamic Ridges and Valleys in a Strike-Slip Environment." *Journal of Geophysical Research: Earth Surface* 120, pp. 2016–2026.

Elst, Nicholas J. van der (Nov. 2017). "Accounting for orphaned aftershocks in the earthquake background rate." *Geophysical Journal International* 211.2, pp. 1108–1118.

Felzer, Karen R. and Emily E. Brodsky (June 2006). "Decay of aftershock density with distance indicates triggering by dynamic stress." *Nature* 441.8, pp. 735–738.

Fiedler, G. B. (Aug. 1974). "Local b-values related to seismicity." *Tectonophysics. Focal processes and the prediction of earthquakes* 23.3, pp. 277–282.

Field, Edward H. et al. (2017). "A spatiotemporal clustering model for the third uniform California earthquake rupture forecast (UCERF3-ETAS): Toward an operational earthquake forecast." *Bulletin of the Seismological Society of America* 107.3, pp. 1049–1081.

Frohlich, Cliff (May 1989). "The Nature of Deep-Focus Earthquakes." *Annual Review of Earth and Planetary Sciences* 17.1, pp. 227–254.

Ganti, Vamsi et al. (2016). "Experimental river delta size set by multiple floods and backwater hydrodynamics."

Garza-Giron, Ricardo, Emily E. Brodsky, and Stephanie G Prejean (2018). "Mainshock-Aftershock Clustering in Volcanic Regions." *Geophysical Research Letters* 45.3, pp. 1370–1378.

Gerstenberger, Matthew C. et al. (May 2005). "Real-time forecasts of tomorrow's earthquakes in California." *Nature* 435.7040, pp. 328–331.

Goebel, T. H.W. et al. (Sept. 2019). "Aftershock deficiency of induced earthquake sequences during rapid mitigation efforts in Oklahoma." *Earth and Planetary Science Letters* 522, pp. 135–143.

Goebel, Thomas H.W. et al. (Sept. 2017). "What allows seismic events to grow big?: Insights from b-value and fault roughness analysis in laboratory stick-slip experiments." *Geology* 45.9, pp. 815–818.

Grant, Lisa B and Kerry Sieh (1994). "Paleoseismic evidence of clustered earthquakes on the San Andreas fault in the Carrizo Plain, California." 99.94, pp. 6819–6841.

Gray, Harrison J. et al. (Jan. 2018). "Off-fault deformation rate along the southern San Andreas fault at Mecca Hills, southern California, inferred from landscape modeling of curved drainages." *Geology* 46.1, pp. 59–62.

Gulia, L. et al. (2018). "The Effect of a Mainshock on the Size Distribution of the Aftershocks." *Geophysical Research Letters* 45.24, pp. 13, 277–13, 287.

Gulia, Laura and Stefan Wiemer (Oct. 2019). "Real-time discrimination of earthquake foreshocks and aftershocks." *Nature* 574.7777, pp. 193–199.

Gulia, Laura, Stefan Wiemer, and Gianfranco Vannucci (July 2020). "Pseudo-prospective Evaluation of the Foreshock Traffic-Light System in Ridgecrest

and Implications for Aftershock Hazard Assessment." *Seismological Research Letters* 91.5, pp. 2828–2842.

Gutenberg, B. and C. F. Richter (Oct. 1944). "Frequency of earthquakes in California." *Bulletin of the Seismological Society of America* 34.4, pp. 185–188.

Harbert, Sarah A., Alison R. Duvall, and Gregory E. Tucker (Nov. 2018). "The Role of Near-Fault Relief Elements in Creating and Maintaining a Strike-Slip Landscape." *Geophysical Research Letters* 45.21, pp. 611–683.

Hardebeck, Jeanne L. et al. (Jan. 2018). "Updated California aftershock parameters." *Seismological Research Letters* 90.1, pp. 262–270.

Hayes, Gavin P. (June 2017). "The finite, kinematic rupture properties of great-sized earthquakes since 1990." *Earth and Planetary Science Letters* 468, pp. 94–100.

Helmstetter, Agnès, Yan Y. Kagan, and David D. Jackson (2005). "Importance of small earthquakes for stress transfers and earthquake triggering." *Journal of Geophysical Research: Solid Earth* 110.B5.

Helmstetter, Agnès and Bruce E. Shaw (July 2006). "Relation between stress heterogeneity and aftershock rate in the rate-and-state model." *Journal of Geophysical Research* 111.B7, B07304.

Helmstetter, Agnès and Didier Sornette (Oct. 2003). "Båth's law derived from the Gutenberg-Richter law and from aftershock properties." *Geophysical Research Letters* 30.20, n/a–n/a.

Hilley, George E. and J. Ramón Arrowsmith (May 2008). "Geomorphic response to uplift along the Dragon's Back pressure ridge, Carrizo Plain, California." *Geology* 36.5, pp. 367–370.

Humphrey, Neil F. and Sarah K. Konrad (Jan. 2000). "River incision or diversion in response to bedrock uplift." *Geology* 28.1, p. 46.

Hutton, Kate, Jochen Woessner, and Egill Hauksson (Apr. 2010). "Earthquake Monitoring in Southern California for Seventy-Seven Years (1932–2008)." *Bulletin of the Seismological Society of America* 100.2, pp. 423–446.

Jaeger, John C. and Horatio Scott Carslaw (1959). *Conduction of heat in solids*. Oxford: Oxford University Press.

Jerolmack, Douglas J. and David Mohrig (May 2007). "Conditions for branching in depositional rives." *Geology* 35.5, pp. 463–466.

Kisslinger, Carl (1996). "Aftershocks and Fault-Zone Properties." *Adavances in geophysics Volume 38*. London: Academic Press, pp. 1–36.

Lawson, Andrew C. (1908). *The California Earthquake of April 18, 1906*. Carnegie Institution of Washington.

Leopold, Luna B. and William B. Bull (June 1979). "Base Level, Aggradation, and Grade." *Proceedings of the American Philosophical Society* 123.3, pp. 168–202.

Liu, Chengli et al. (2019). "Coseismic Rupture Process of the Large 2019 Ridgecrest Earthquakes From Joint Inversion of Geodetic and Seismological Observations." *Geophysical Research Letters* 46.21, pp. 11820–11829.

Mai, P. Martin and Gregory C. Beroza (2000). "Source Scaling Properties from Finite-Fault-Rupture Models." *Bull. Seism. Soc. Am.* 90.3, pp. 604–615.

Main, Ian G., Philip G. Meredith, and Colin Jones (1989). "A reinterpretation of the precursory seismic b-value anomaly from fracture mechanics." *Geophysical Journal International* 96.1, pp. 131–138.

Mann, Paul et al. (Jan. 2005). "Neotectonics of southern Puerto Rico and its offshore margin."

Marsan, David (June 2006). "Can coseismic stress variability suppress seismicity shadows? Insights from a rate-and-state friction model." *Journal of Geophysical Research: Solid Earth* 111.6.

Marsan, David and Agnès Helmstetter (July 2017). "How variable is the number of triggered aftershocks?" *Journal of Geophysical Research: Solid Earth* 122.7, pp. 5544–5560.

Marsan, David et al. (Oct. 2014). "Foreshock activity related to enhanced aftershock production." *Geophysical Research Letters* 41.19, pp. 6652–6658.

McCloskey, John et al. (June 2003). "Structural constraints on the spatial distribution of aftershocks." *Geophysical Research Letters* 30.12.

McGuire, Jeffrey J., Margaret S Boettcher, and Thomas H Jordan (2005). "Foreshock sequences and short-term earthquake predictability on East Pacific Rise transform faults." *Nature* 434, pp. 457–461.

Mei, Hongyuan and Jason Eisner (Nov. 2017). "The Neural Hawkes Process: A Neurally Self-Modulating Multivariate Point Process." *arXiv:1612.09328 [cs, stat]*.

Mignan, Arnaud (May 2015). "The debate on the prognostic value of earthquake foreshocks: A meta-analysis." *Scientific Reports* 4.1, p. 4099.

Mizrahi, Leila, Shyam Nandan, and Stefan Wiemer (2021). "Embracing Data Incompleteness for Better Earthquake Forecasting." *Journal of Geophysical Research: Solid Earth* 126.12, e2021JB022379.

Mogi, K (Mar. 1963). "Magnitude-Frequency Relation for Elastic Shocks Accompanying Fractures of Various Materials and Some Related problems in Earthquakes (2nd Paper)." = *Bulletin of the Earthquake Research Institute, University of Tokyo* 40.4, pp. 831–853.

Mogi, Kiyoo (Dec. 1967). "Earthquakes and fractures." *Tectonophysics* 5.1, pp. 35–55.

Molchan, G. M., T. L. Kronrod, and A. K. Nekrasova (Mar. 1999). "Immediate foreshocks: time variation of the b-value1In memoriam A.G. Peozorov.1." *Physics of the Earth and Planetary Interiors* 111.3, pp. 229–240.

Müller, R. Dietmar et al. (2008). "Age, spreading rates, and spreading asymmetry of the world's ocean crust." *Geochemistry, Geophysics, Geosystems* 9.4, pp. 1–19.

Nishikawa, Tomoaki and Satoshi Ide (Dec. 2014). "Earthquake size distribution in subduction zones linked to slab buoyancy." *Nature Geoscience* 7.12, pp. 904–908.

Noda, Hiroyuki, Nadia Lapusta, and Hiroo Kanamori (June 2013). "Comparison of average stress drop measures for ruptures with heterogeneous stress change and implications for earthquake physics." *Geophysical Journal International* 193.3, pp. 1691–1712.

Nyffenegger, Paul and Cliff Frohlich (Apr. 2000). "Aftershock occurrence rate decay properties for intermediate and deep earthquake sequences." *Geophysical Research Letters* 27.8, pp. 1215–1218.

Obara, Kazushige (2003). "Hi-net: High sensitivity seismograph network, Japan." *Methods and Applications of Signal Processing in Seismic Network Operations*. Vol. 98. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 79–88.

Ogata, Yosihiko (1988). "Statistical models for earthquake occurrences and residual analysis for point processes." *Journal of the American Statistical Association* 83.401, pp. 9–27.

— (June 1998). "Space-Time Point-Process Models for Earthquake Occurrences." *Annals of the Institute of Statistical Mathematics* 50.2, pp. 379–402.

— (Aug. 2017). "Statistics of Earthquake Activity: Models and Methods for Earthquake Predictability Studies." *Annual Review of Earth and Planetary Sciences* 45.1, pp. 497–527.

Omi, Takahiro, naonori ueda naonori, and Kazuyuki Aihara (2019). "Fully Neural Network based Model for General Temporal Point Processes." *Advances in Neural Information Processing Systems*. Vol. 32. Curran Associates, Inc.

Omi, Takahiro et al. (2015). "Intermediate-term forecasting of aftershocks from an early aftershock sequence: Bayesian and ensemble forecasting approaches." *Journal of Geophysical Research: Solid Earth* 120.4, pp. 2561–2578.

Omori, F. (1895). *On the aftershocks of earthquakes*.

Ouchi, Shunji (2005). "Development of offset channels across the San Andreas fault." *Geomorphology* 70.1-2, pp. 112–128.

Page, Morgan T. et al. (Oct. 2016). "Three Ingredients for Improved Global Aftershock Forecasts: Tectonic Region, Time-Dependent Catalog Incompleteness,

and Intersequence Variability." *Bulletin of the Seismological Society of America* 106.5, pp. 2290–2301.

Paola, Chris, Paul L. Heller, and Charles L. Angevine (June 1992). "The large-scale dynamics of grain-size variation in alluvial basins, 1: Theory." *Basin Research* 4.2, pp. 73–90.

Paszke, Adam et al. (2019). "PyTorch: An Imperative Style, High-Performance Deep Learning Library." *Advances in Neural Information Processing Systems*. Vol. 32. Curran Associates, Inc.

Perron, J. Taylor, James W. Kirchner, and William E. Dietrich (2009). "Formation of evenly spaced ridges and valleys." *Nature* 460.7254, pp. 502–505.

Persh, Steven E and Heidi Houston (Oct. 2004). "Strongly Depth-Dependent Aftershock Production in Deep Earthquakes." *Bulletin of the Seismological Society of America* 94.5, pp. 1808–1816.

Reasenberg, Paul A. (Mar. 1999). "Foreshock occurrence before large earthquakes." *Journal of Geophysical Research: Solid Earth* 104.B3, pp. 4755–4768.

Reasenberg, Paul A. and Lucile M. Jones (1989). "Earthquake Hazard after a Mainshock in." *Science* 243.4895, pp. 1173–1176.

Reitman, Nadine G. et al. (2019a). "Offset Channels May Not Accurately Record Strike-Slip Fault Displacement: Evidence From Landscape Evolution Models." *Journal of Geophysical Research: Solid Earth* 124.12, pp. 13427–13451.

Reitman, Nadine G. et al. (Dec. 2019b). "Offset Channels May Not Accurately Record Strike-Slip Fault Displacement: Evidence From Landscape Evolution Models." *Journal of Geophysical Research: Solid Earth* 124.12, pp. 13427–13451.

Ross, Gordon J. (May 2021). "Bayesian Estimation of the ETAS Model for Earthquake Occurrences." *Bulletin of the Seismological Society of America* 111.3, pp. 1473–1480.

Ross, Zachary E. et al. (Oct. 2019a). "Hierarchical interlocked orthogonal faulting in the 2019 Ridgecrest earthquake sequence." *Science* 366.6463, pp. 346–351.

Ross, Zachary E. et al. (2019b). "Searching for hidden earthquakes in Southern California." *Science*.

Roy, S. G. et al. (Oct. 2016). "A fault runs through it: Modeling the influence of rock strength and grain-size distribution in a fault-damaged landscape." *Journal of Geophysical Research: Earth Surface* 121.10, pp. 1911–1930.

Schmalzle, Gina et al. (2006). "Strain accumulation across the Carrizo segment of the San Andreas Fault, California: Impact of laterally varying crustal properties." *J. Geophys. Res* 111, p. 5403.

Scholz, Christopher H. (May 1968). "Mechanism of creep in brittle rock." *Journal of Geophysical Research* 73.10, pp. 3295–3302.

Scholz, Christopher H. (2019). *The Mechanics of Earthquakes and Faulting*. Cambridge, United Kingdom: Cambridge University Press.

Segou, Margarita and Tom Parsons (2016). "Prospective earthquake forecasts at the himalayan front after the 25 April 2015 M 7.8 Gorkha Mainshock." *Seismological Research Letters* 87.4, pp. 816–825.

Shcherbakov, Robert et al. (Sept. 2019). "Forecasting the magnitude of the largest expected earthquake." *Nature Communications* 10.1, p. 4051.

Shchur, Oleksandr, Marin Biloš, and Stephan Günnemann (Jan. 2020). "Intensity-Free Learning of Temporal Point Processes." *International Conference on Learning Representations*.

Shchur, Oleksandr et al. (Aug. 2021). "Neural Temporal Point Processes: A Review." *International Joint Conference on Artificial Intelligence*.

Shelly, David R. (2017). "A 15 year catalog of more than 1 million low-frequency earthquakes: Tracking tremor and slip along the deep San Andreas Fault." *Journal of Geophysical Research: Solid Earth* 122.5, pp. 3739–3753.

— (Jan. 2020). "A High-Resolution Seismic Catalog for the Initial 2019 Ridgecrest Earthquake Sequence: Foreshocks, Aftershocks, and Faulting Complexity." *Seismological Research Letters* 91.4, pp. 1971–1978.

Shi, Yaolin and Bruce A. Bolt (Oct. 1982). "The standard error of the magnitude-frequency b value." *Bulletin of the Seismological Society of America* 72.5, pp. 1677–1687.

Sieh, Kerry E (1978). "Slip along the San Andreas fault associated with the earthquake." *Bulletin of the Seismological Society of America* 68.5, pp. 155–159.

Sieh, Kerry E and Richard H Jahns (1984). "Holocene activity of the San Andreas fault at Wallace Creek, California." *Geological Society of America Bulletin* 95, pp. 883–896.

Sims, J D (1994). "Stream channel offset and abandonment and a 200-year average recurrence interval of earthquakes on the San Andreas fault at Phelan Creeks, Carrizo Plain, California." *Proceedings of the Workshop on Paleoseismology.* Vol. 94, pp. 170–172.

Singh, S. K. and Gerardo Suárez (Feb. 1988). "Regional variation in the number of aftershocks (mb greater than 5) of large, subduction-zone earthquakes (Mw greater than 7.0)." *Bulletin of the Geological Society of America* 78.1, pp. 230–242.

Slingerland, Rudy and Norman D Smith (Dec. 2004). "River Avulsions and their Deposits." *Annu. Rev. Earth Planet. Sci* 32, pp. 257–285.

Smith, Warwick D. (Jan. 1981). "The b-value as an earthquake precursor." *Nature* 289.5794, pp. 136–139.

Tahir, Mohammad and Jean Robert Grasso (Feb. 2014). "Aftershock patterns of Ms $>$ 7 earthquakes in the India-Asia collision belt: Anomalous results from the Muzaffarabad earthquake sequence, Kashmir, 2005." *Bulletin of the Seismological Society of America* 104, pp. 1–23.

Tahir, Mohammad and Jean-Robert Grasso (Oct. 2015). "Faulting style controls for the space–time aftershock patterns." *Bulletin of the Seismological Society of America* 105.5, pp. 2480–2497.

Tahir, Mohammad, Jean-Robert Grasso, and D. Amorse (2012). "The largest aftershock: How strong, how far away, how delayed?" *Geophysical Research Letters* 39.4, pp. 1–5.

Tan, Yen Joe et al. (Apr. 2021). "Machine-Learning-Based High-Resolution Earthquake Catalog Reveals How Complex Fault Structures Were Activated during the 2016–2017 Central Italy Sequence." *The Seismic Record* 1.1, pp. 11–19.

Trugman, Daniel T. and Zachary E. Ross (Aug. 2019). "Pervasive Foreshock Activity Across Southern California." *Geophysical Research Letters* 46.15, pp. 8772–8781.

U.S. Geological Survey and New Mexico Bureau of Mines and Mineral Resources (2020). *Quaternary fault and fold database for the United States.*

Upadhyay, Utkarsh, Abir De, and Manuel Gomez Rodriguez (2018). "Deep Reinforcement Learning of Marked Temporal Point Processes." *Advances in Neural Information Processing Systems*. Vol. 31. Curran Associates, Inc.

Utsu, Tokuji (Mar. 1970). "Aftershocks and Earthquake Statistics(1) : Some Parameters Which Characterize an Aftershock Sequence and Their Interrelations." *Journal of the Faculty of Science, Hokkaido University*. 7th ser. 3.3, pp. 129–195.

Utsu, Tokuji, Yosihiko Ogata, and Ritsuko S Matsu 'ura (1995). "The Centenary of the Omori Formula for a Decay Law of Aftershock Activity." *Journal of Physics of the Earth* 43.1, pp. 1–33.

Van Der Elst, Nicholas J. and Bruce E. Shaw (July 2015). "Larger aftershocks happen farther away: Nonseparability of magnitude and spatial distributions of aftershocks." *Geophysical Research Letters* 42.14, pp. 5771–5778.

Wallace, Robert E (1968). *Notes on stream channels offset by the San Andreas fault, southern Coast Ranges, California*. Tech. rep., pp. 6–21.

Warren-Smith, E et al. (2019). "Episodic stress and fluid pressure cycling in subducting oceanic crust during slow slip." *Nature Geoscience* 12.June, pp. 475–481.

Wells, D.L. L Donald L. and Kevin J. K.J. J Coppersmith (1994). "Empirical relationships among magnitude, rupture length, rupture width, rupture area

and surface displacements." *Bulletin of the Seismological Society of America* 84.4, pp. 974–1002.

Wetzler, Nadav, Emily E. Brodsky, and Thorne Lay (Dec. 2016). "Regional and stress drop effects on aftershock productivity of large megathrust earthquakes." *Geophysical Research Letters* 43.23, pp. 12, 012–12, 020.

White, Malcolm C. A., Yehuda Ben-Zion, and Frank L. Vernon (2019). "A Detailed Earthquake Catalog for the San Jacinto Fault-Zone Region in Southern California." *Journal of Geophysical Research: Solid Earth* 124.7, pp. 6908–6930.

Wiens, Douglas A. et al. (Aug. 1997). "Aftershock sequences of moderate-sized intermediate and deep earthquakes in the Tonga Subduction Zone." *Geophysical Research Letters* 24.16, pp. 2059–2062.

Witten, Ian H., Eibe Frank, and Mark A. Hall (2011). *Data Mining Practical Machine Learning Tools and Techniques*. Burlington: Elsevier.

Wu, Li-Ru Ru and Wang-Ping Ping Chen (July 1999). "Anomalous aftershocks of deep earthquakes in Mariana." *Geophysical Research Letters* 26.13, pp. 1977–1980.

Wyss, Max (1973). "Towards a Physical Understanding of the Earthquake Frequency Distribution." *Geophysical Journal of the Royal Astronomical Society* 31.4, pp. 341–359.

Yamanaka, Yoshiko and Kunihiko Shimazaki (1990). "Scaling Relationship between the Number of Aftershocks and the Size of the Main Shock." *Journal of Physics of the Earth* 38.4, pp. 305–324.

Zaliapin, Ilya and Yehuda Ben-Zion (Oct. 2016). "A global classification and characterization of earthquake clusters." *Geophysical Journal International* 207.1, pp. 608–634.

Zaliapin, Ilyar et al. (2008). "Clustering analysis of seismicity and aftershock identification." *Physical Review Letters* 101.1.

Zhou, Zihao et al. (Dec. 2021). "Neural Point Process for Learning Spatiotemporal Event Dynamics." *arXiv:2112.06351 [cs]*.

Zielke, Olaf, Yann Klinger, and J Ramón Arrowsmith (Jan. 2015). "Fault slip and earthquake recurrence along strike-slip faults —Contributions of high-resolution geomorphic data." *Tectonophysics* 638, pp. 43–62.

Zielke, Olaf et al. (Feb. 2010). "Slip in the 1857 and earlier large earthquakes along the Carrizo Plain, San Andreas Fault." *Science* 327.5969, pp. 1119–1122.

Appendix A

Supplementary Material for chapter 1

A.1 Materials and Methods

We consider channels in the Carrizo Plain along the San Andreas. We use the USGS Quaternary Fault database to identify active faults (U.S. Geological Survey and New Mexico Bureau of Mines and Mineral Resources 2020). 1m resolution altimetry from the B4 Lidar Project is used to quantify relevant morphological features (Bevis et al. 2017). We manually digitize a diverse sampling of channels within the lidar swath using a shaded relief map (Fig. A6-64). The sampling strategy emphasized collecting a diverse set of channel geometries in order to

bracket the critical offset and test end-member cases of the model. The validation exercise we present (Eq. 1.3-1.4 and Fig. 1.3) expands the rate of aggradation at the avulsion node. We therefore exclude channels where the avulsion occurs far away from the avulsion node relative to the fault offset (e.g. Fig. A40-41). This might occur if changes in fault kinematics along strike suppress avulsion near the avulsion node (by increasing h_c) and promote avulsion downstream (decreasing h_c). We further exclude channels for which offset is too small to confidently measure (e.g. Fig. A20, A24). The data include abandoned channels with small offsets and active channels with kilometers of offset (Fig. 1.2a). For each channel, we infer the initial slope (S_0), the avulsion threshold (h_c), the catchment length (L) and offset (d_{obs}) (Table A.1, Fig. A6-64). More detailed examples are provided further below (Fig. A.5).

To approximate the initial slope, we collect a representative value of the slope upstream of the fault, ensuring that the channel profile is roughly linear over the sampling interval. The avulsion threshold (h_c), the vertical height of aggradation required to avulse the channel, is an inferred quantity which relies on geomorphic interpretation. Aggraded sediment obscures the initial channel floor elevation datum. For recent avulsions, rapid incision rapidly recovers this original datum. In these cases, we estimate h_c, likely as a lower bound, in a channel cross section at the avulsion node. In the absence of this direct measurement, we measure the

difference between the elevation of the channel wall near the avulsion node and elevation of the channel floor midway along the fault parallel segment, where we infer the channel should neither aggrade nor erode. Catchment length is measured as the distance from the avulsion node to the start of the channel. Offset is measured as the fault parallel distance required to backslip channels to a straight fault crossing geometry. Our measurements presume flow initially crosses the fault undisturbed when the offset was last reset. This initial condition is often the end-result of a previous avulsion. This assumption fails, for example, if an avulsion at or upstream of the fault causes flow to reroute or deflect along pre-existing topographic features resulting in deflected but not offset initial channel geometries. Another case incurs from a channel that establishes its geometry by eroding headward along the fault. In neither of these cases is channel misalignment simply related to channel offset. Our sampling excludes channels for which the assumptions of initial conditions are unclear and stands to significantly impact the offset measurements.

An abandoned channel is said to have a recent avulsion when there is little to no offset on the active tributary of the channel relative to the total offset. An active channel is recorded as featuring incipient avulsion if aggradation approaches the avulsion threshold or evidence of overspilling near the avulsion node is present (e.g., Fig. 1.1b). In our sampling, we avoid channels that appear to be cut by

multiple strands of the San Andreas, thus ensuring that a single site-wide slip rate of 3.5 cm/yr is appropriate for the entire dataset (Sieh and Jahns 1984).

A.2 Semi-analytical forward solutions

As a validation of the analytical approximation of the channel aggradation rate, we include a semi-analytical model that approximates the diffusive response to offset between punctuated increments of slip, akin to earthquake forcing.

e consider a steady state long-channel profile z(x) with length L, discretized in elements of length dx, with fixed starting and finishing elevations H and 0 respectively (e.g. Fig. A.1a). In this profile, a zone of width w_0 is subject to fault offset. An increment of slip is introduced by stretching this zone and re-interpolating the profile to maintain approximately regular point spacing. The post-slip profile, $z_0(x)$, is then allowed to respond diffusively with a diffusivity κ during the earthquake recurrence interval, T. This initial value problem is well posed and has the analytical solution (Jaeger and Carslaw 1959):

$$z(x,T) = H - \frac{Hx}{L} + \sum_{n=1}^{\infty} B_n \sin\left(\frac{n\pi x}{L}\right) e^{k\frac{n\pi}{L}^2 T} \tag{A.1}$$

where the coefficients are given by,

$$B_n = \frac{2}{L}\int_0^L \left(z_0(x) - H + \frac{Hx}{L}\right) \sin\left(\frac{n\pi x}{L}\right) dx \tag{A.2}$$

For repeated earthquakes it is convenient to numerically evaluate the integral solution for the coefficients and expand the solution to the Nth degree, ensuring that the solution remains stable. This approach allows for a robust and efficient computation of the channel response over a large densely discretized domain over a long period of time. An implementation of this model included here: https://doi.org/10.5281/zenodo.4766502

The avulsion time scale can therefore be obtained by incrementally cycling through slip and inter-seismic diffusive responses until the aggradation along the profile exceeds the avulsion threshold, which, for its part can be prescribed to vary arbitrarily with time.

Measurements in this study, initial slope, avulsion threshold, diffusivity, fault zone width, domain size and slip, can all be encoded to produce initial profiles. Recurrence intervals, earthquake slip, and channel diffusivity then dictate the evolution of the channel profile given realistic parameters. As such, we can determine unique estimates of t_c for each channel (e.g. Fig. A.1). Though we expect solutions derived in this approach to be more accurate than the order of magnitude treatment presented in the main text, these results do not afford the same degree of interpretability.

A.3 Slip- and diffusion-limited channel response

The relative length-scales of the diffusive pulse of aggrading sediment, the slip, and the channel size determine different physical regimes for the channel response. In this study, we primarily discuss a diffusion-limited response. We explore the spatio-temporal limits of this approximation, demonstrate its applicability to our data, and briefly describe system- and slip-limited responses that may arise.

Aggradation emplaces a triangular wedge of sediment at the avulsion node (Fig. A.2). With time, the basal dimensions of the triangular wedge grow according to the rate limiting dimension. If the diffusive length scale, $l_{diff} \sim \sqrt{\kappa t}$, exceeds the advective length scale $l_{slip} \sim vt/2$, the channel response is slip-limited. If $l_{diff} < l_{slip}$, the channel response is diffusion-limited. Upstream aggradation can also encounter the system length scale (L), the channel reach. In each case, similar triangles relate the relevant length scale of the offset channel (Fig. A.2):

$$\frac{S_0 L_u}{L_u + L_d} \approx \frac{h_c}{L_d} \tag{A.3}$$

where S_0 is the initial slope, h_c is the critical avulsion threshold, and L_{u-d} are the limiting up- and downstream horizontal length scales.

A.3.1 Diffusion-limited

In the diffusion-limited case, $vt_c/2 < \sqrt{\kappa t_c}$ (presented in the main manuscript), the diffusive pulse near the avulsion node propagates up- and downstream sym-

metrically and is not influenced by the converse erosive pulse propagating upstream. In this case, $L_{u-d} \approx \sqrt{\kappa t_c}$ and,

$$\frac{S_0\sqrt{\kappa t_c}}{2\sqrt{\kappa t_c}} \approx \frac{h_c}{\sqrt{\kappa t_c}} \tag{A.4}$$

Thus, isolating for t_c, we obtain the same solution as presented in the main section of the manuscript, $t_c \approx (4h_c^2)/(\kappa S_0^2)$.

A.3.2 Slip-limited

In the slip-limited case $vt_c/2 < \sqrt{\kappa t_c}$, the downstream diffusive pulse is balanced by the erosive pulse propagating upstream. A resulting neutral point midway along the offset pins the geometry of the aggrading wedge of sediment such that $L_u \approx \sqrt{\kappa t_c}$, $L_d \approx vt/2$ and,

$$\frac{S_0\sqrt{\kappa t_c}}{\sqrt{\kappa t_c} + vt/2} \approx \frac{h_c}{vt_c/2} \tag{A.5}$$

In the case that $vt_c/2 \ll \sqrt{\kappa t_c}$, the critical avulsion time scale is insensitive to the channel diffusivity and $t_c \approx 2h_c/vS_0$.

A.3.3 System- and diffusion-limited

If the channel is allowed to equilibrate fully over its entire reach but is not slip-limited downstream, $L \ll \sqrt{\kappa t_c}$, but slip has eventually outpaced the channel

response, $\sqrt{\kappa t_c} < vt_c/2$, $L_u = L$, $L_d \approx \sqrt{\kappa t_c}$, and,

$$\frac{S_0 L}{L + \sqrt{\kappa t_c}} \approx \frac{h_c}{\sqrt{\kappa t_c}} \tag{A.6}$$

Then a solution for the avulsion time exists, $t_c \approx \left(\frac{L}{S_0 L/h_c - 1}\right)^2 / \kappa$, but is highly sensitive to initial conditions since $(S_0 L)/h_c - 1$ is close to 0.

A.3.4 Phase space

The diffusion-, slip- and system-limited channel responses delineate a phase space which is most strongly modulated by the channel diffusivity and the avulsion time scale (Fig. 1.3). In the extremes of the temporal scales we consider, the validity of a time-independent diffusivity coefficient should be carefully considered. Importantly, we show that given output from our semi-analytical model, which should be accurate across the entire phase space, most of our data are well described by a diffusion-limited response.

A.4 Diffusivity validation

The diffusion-, slip- and system-limited channel responses delineate a phase space which is most strongly modulated by the channel diffusivity and the avulsion time scale (Fig. 1.3). In the extremes of the temporal scales we consider, the validity of a time-independent diffusivity coefficient should be carefully con-

sidered. Importantly, we show that given output from our semi-analytical model, which should be accurate across the entire phase space, most of our data are well described by a diffusion-limited response.

A.4.1 Empirical measurements

Channel diffusivity is most directly measured from an equilibrium channel slope (S_{eq}) of a channel for which the sediment transport (Q_{sed}) per channel width (W) is modulated by the diffusivity (κ):

$$S_{eq} = \frac{Q_{sed}}{W\kappa} \tag{A.7}$$

Thus, provided direct measurements of sediment transport, slope and channel width, diffusion can be constrained (Castelltort and Van Den Driessche 2003). However, long-term average values of the sediment flux are difficult to obtain directly. The sediment flux can, however, be estimated indirectly from the erosion rate (E) integrated over the catchment area (A).

$$Q_{sed} = \int_A Eda \sim EL^2 \tag{A.8}$$

At the Carrizo Plain, representative values of channel slope (0.1), reach (500 m), erosion (0.1mm/yr) and width (3 m) yield a diffusivity on the order of

$$\kappa \sim \frac{EL^2}{WS_{eq}} \sim 10^{-6} m^2/s \tag{A.9}$$

A.4.2 Stream power and bedload mobility

The approach presented in the manuscript calculates diffusivity by estimating the water flux in the channel and relating it to a bedload transport capacity,

$$\kappa = \frac{-8qA\sqrt{c_f}}{(C_0(s-1)} \approx 0.1q \tag{A.10}$$

where q is the long-term volumetric water flux normalized to the width of the basin, A (0.1-1) is a non-dimensional constant related to the channel cross-sectional geometry, c_f is a dimensionless drag coefficient typically on the order of 0.01, C_0 is the sediment concentration (~ 0.7), and s is the specific gravity (2.7) (Paola, Heller, and Angevine 1992). The major uncertainties in this approach are estimating the water discharge and the channel geometry (A). We use the estimate that $q = rL$ and $A = 0.1$, where r is the annual rainfall (5 cm/yr for the Carrizo Plain). However, it is difficult to assess whether modern rainfall rates are representative of the long-term average. For a typical channel draining the Temblor range, this yields diffusivity on the order of $10^{-7}m^2/s$.

A.4.3 Scarp diffusivity

An absolute lower bound on the channel diffusivity may be derived from scarp dating techniques in the Carrizo Plain. Since sediment transport in the landscape is mediated by much less effective geomorphic processes (e.g. rain splash, soil creep and burrowing). Here, morphologic dating of fault scarps (Arrowsmith,

Rhodes, and Pollard 1998) suggests landscape diffusivity of $8.6 \pm 0.8 m^2/kyr$ or $\sim 3x10^{-10} m^2/s$, which is consistent with our estimates. The sources including hillslope-derived sediment and bedrock incision of the fault crossing reach contribute to the aggradation rate in the channel, but are neither sufficiently large nor sufficiently perturbed to contribute significantly to the channel morphology (Perron, Kirchner, and Dietrich 2009).

A.5 Diffusive and advective timescales

We consider a time dependent avulsion threshold which includes first order time dependence in a Taylor series expansion, $h_c(t) = h_0 + \lambda v_x t + \cdots$, where h_0 reflects initial channel height or incision immediately following an avulsion and λ is the ratio of local vertical to horizontal motion at the avulsion node. Introducing the first order term of the time dependent avulsion threshold into Eq. 1.3, we obtain

$$\frac{h_0 + \lambda v_x t_c}{t_c} \approx \frac{\kappa S_0^2}{4(h_0 + v_x t_c)}. \tag{A.11}$$

Again, isolating for the critical avulsion timescale, we obtain the quadratic equation,

$$t_c \approx \frac{1}{4\lambda^2 v_x^2}(-4\lambda v_x h_0 + \kappa S_o^2 \pm S_0\sqrt{(\kappa^2 S_0^2 - 16 v_x h_0 \kappa)}). \tag{A.12}$$

This solution represents the two intercepts of the square root growth of sediment aggradation of the channel floor and the linear growth of the avulsion threshold (Fig. 4.5). Since the avulsion resets the channel geometry, we can rule out the second solution. If the radicand is negative, aggradation is outpaced by the growth of h_c such that the offset of the channel is boundless. Well established fault parallel drainage will therefore develop provided these conditions are maintained. For $\lambda > 0$, setting the radicand to be zero yields the ratio

$$R_z = \frac{16h_0}{\kappa S_0^2}, \tag{A.13}$$

which compares vertical advection and the diffusive response. When $R_z < 1$ channels can avulse; when R_z>1, channels cannot avulse in response to the tectonic forcing (Fig. 4.5). Setting R_z to 1, we can determine the critical value of λ for which this transition occurs,

$$\lambda_c = \frac{\kappa S_0^2}{16 v_x h_0} \tag{A.14}$$

Note that if the initial avulsion threshold is small or when λ is negative, the avulsion frequency will be high ($t_c \sim 0$) and the channel geometry will be highly unstable.

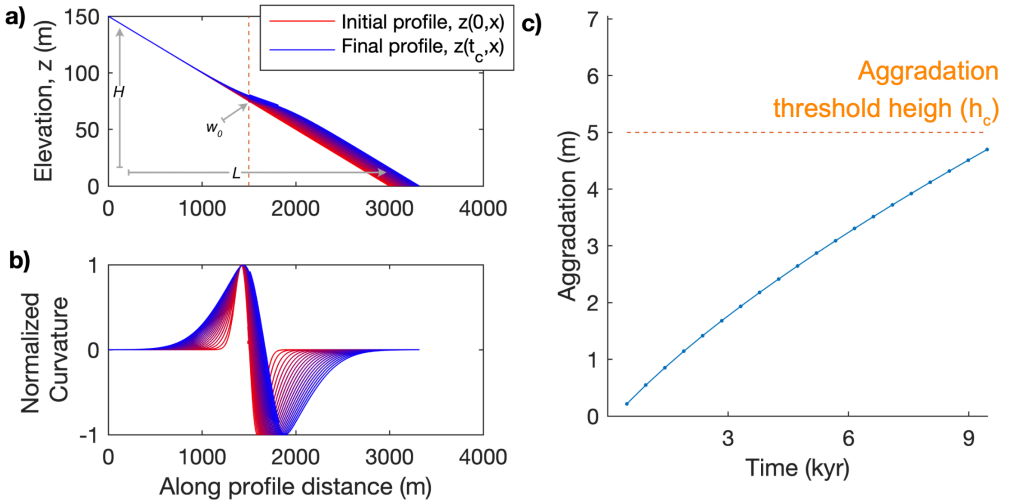

Figure A.1: Semi-analytical evolution of offsetting channel profile. a) Modelled channel long profile. b) Corresponding curvature along the profile. c) Aggradation at the avulsion node. Each point corresponds to an earthquake. Parameterization for this model run was selected to be analogous to Wallace Creek.

A.6 Data

We include the measurements of channel geometry we used in the manuscript below. The extent of the area of interest is shown in the inset of Fig. A.4. We first provide examples of how these measurements are collected. The data are then tabulated in Table A.1.

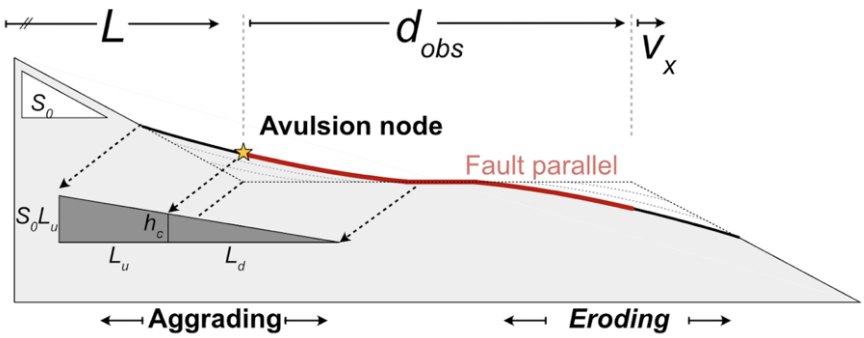

Figure A.2: Schematic channels profile with the aggrading wedge of sediment.

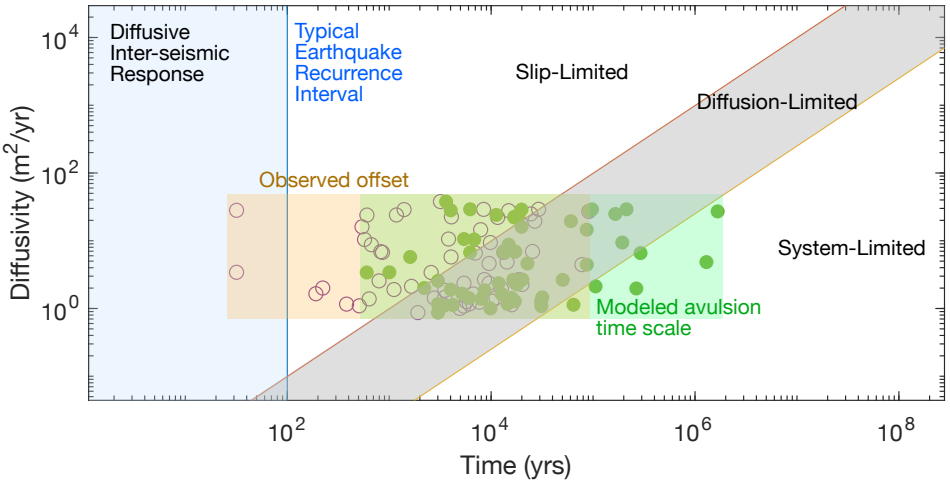

Figure A.3: Phase space of the channel response visualized along a section of constant slip rate with the system size implicitly related to the diffusivity. Most of the data are well approximated by a diffusion-limited response. Slip-limited and system-limited responses bracket the estimated avulsion timescale derived from forward models of our semi-analytical model solutions provided initial conditions from our data. Here, we use diffusivity estimates representative of the Carrizo Plain. The zone bracketing the diffusion-limited response would widen in a locale where channel response is slower or where fault slip is faster.

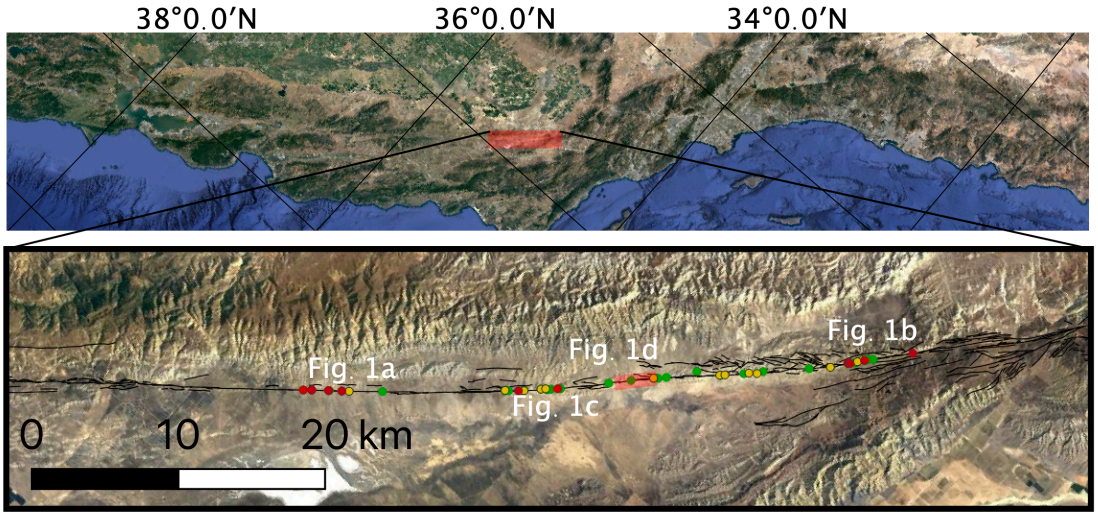

Figure A.4: Satellite imagery of the study area, the Carrizo Plain, California, USA. Locations featured in Fig. 1.1 and 1.2 are indicated in the inset. Quaternary faults are indicated in black.

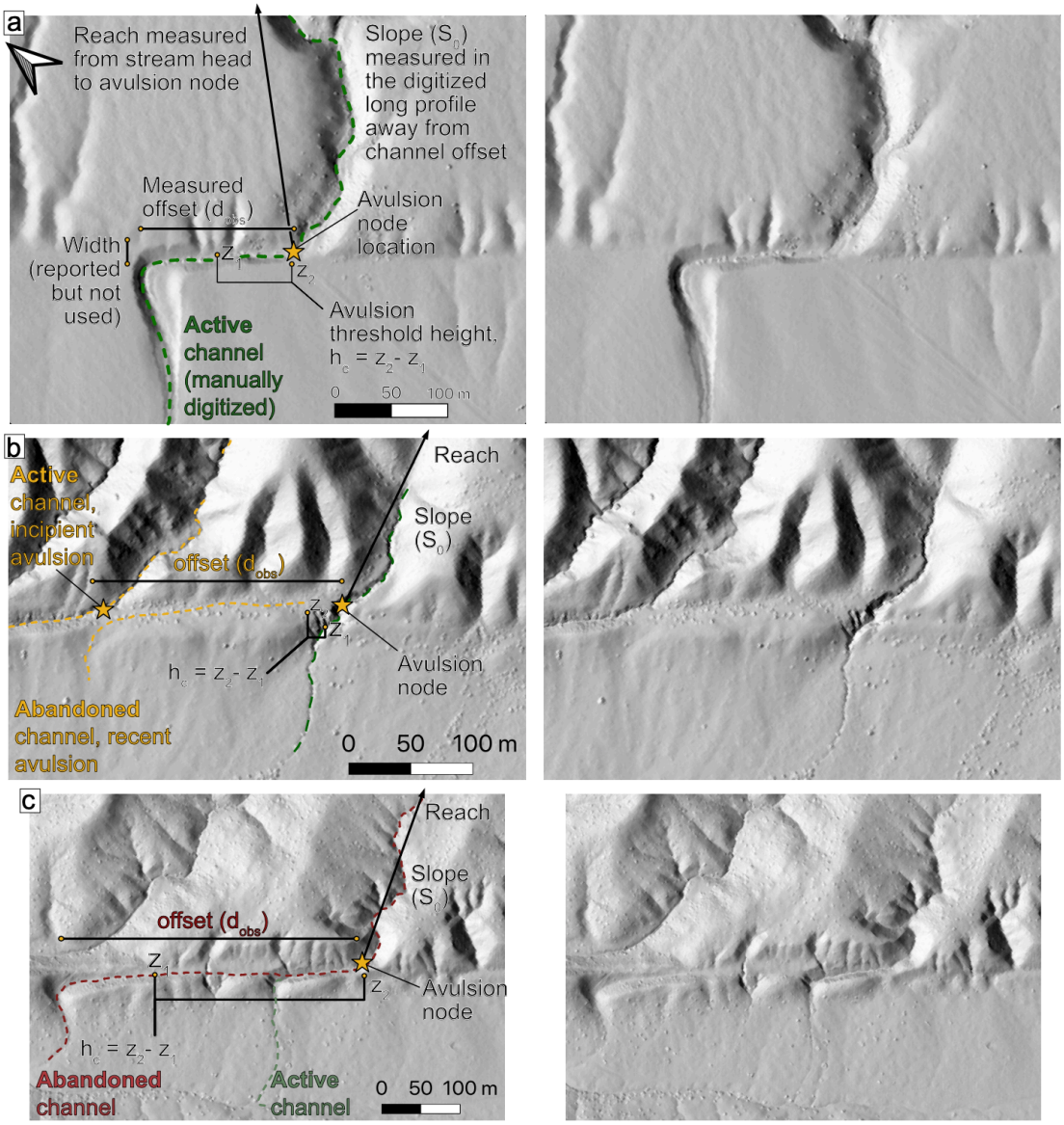

Figure A.5: Annotated examples of channel geometry measurements. a-c) correspond to Fig. 1.1a-c and can be located according to Fig. A.4. These are examples of a channel that is a) active, b) recently avulsed and c) abandoned. In c), the avulsion threshold could be equivalently obtained by deriving z_1 midway along the offset of the active branch.

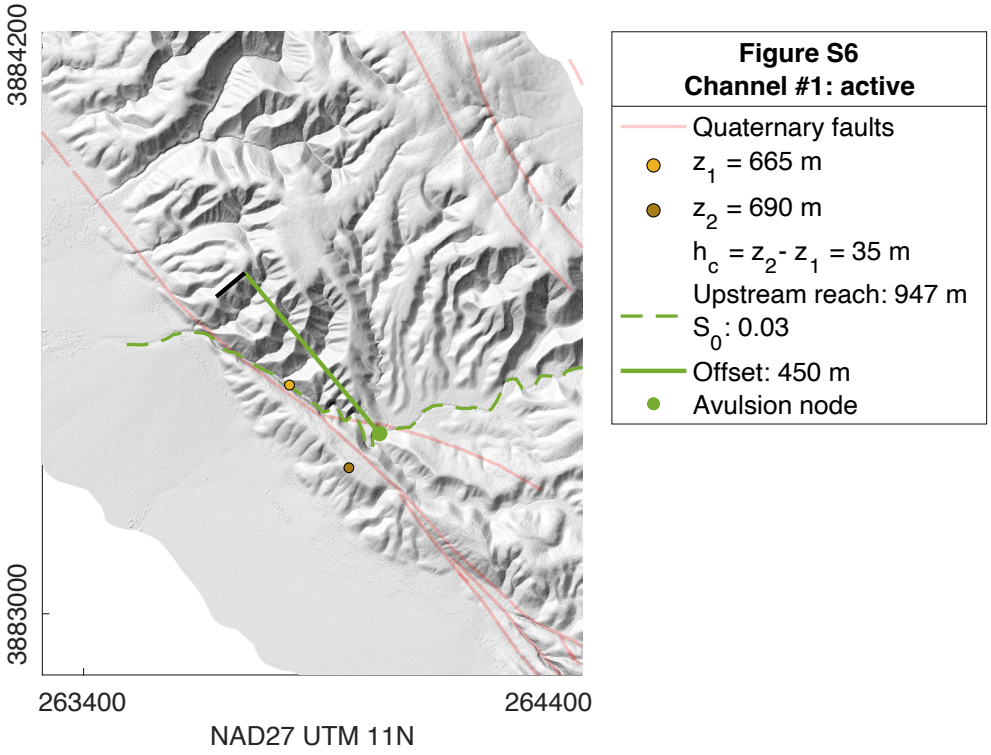

Figure A.6: Measurements for channel 1

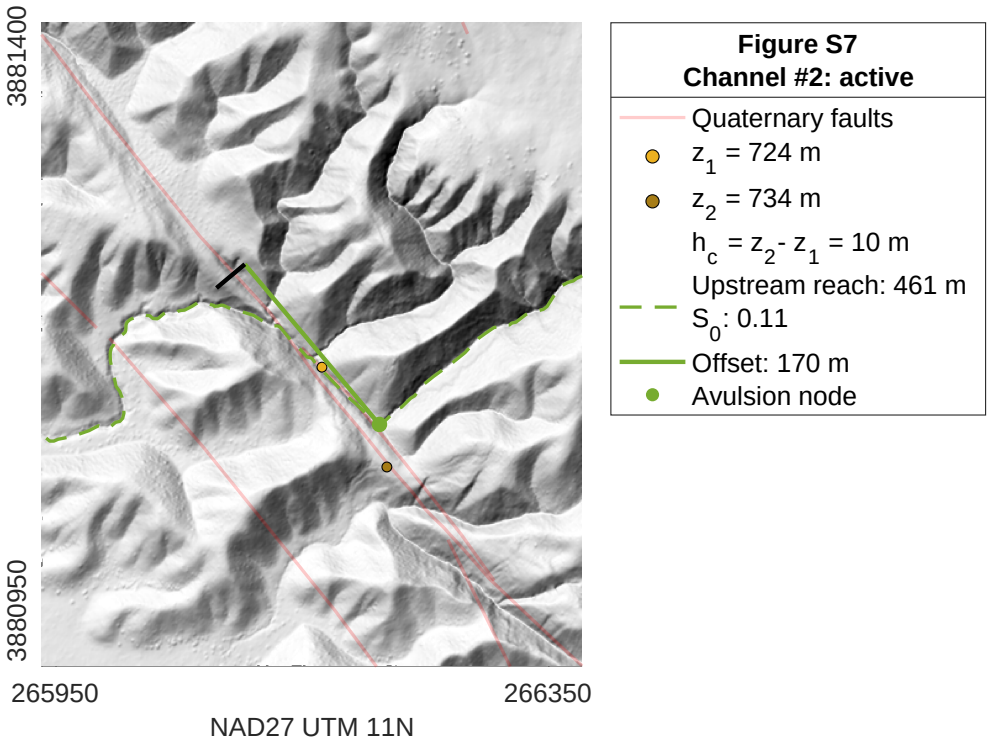

Figure A.7: Measurements for channel 2

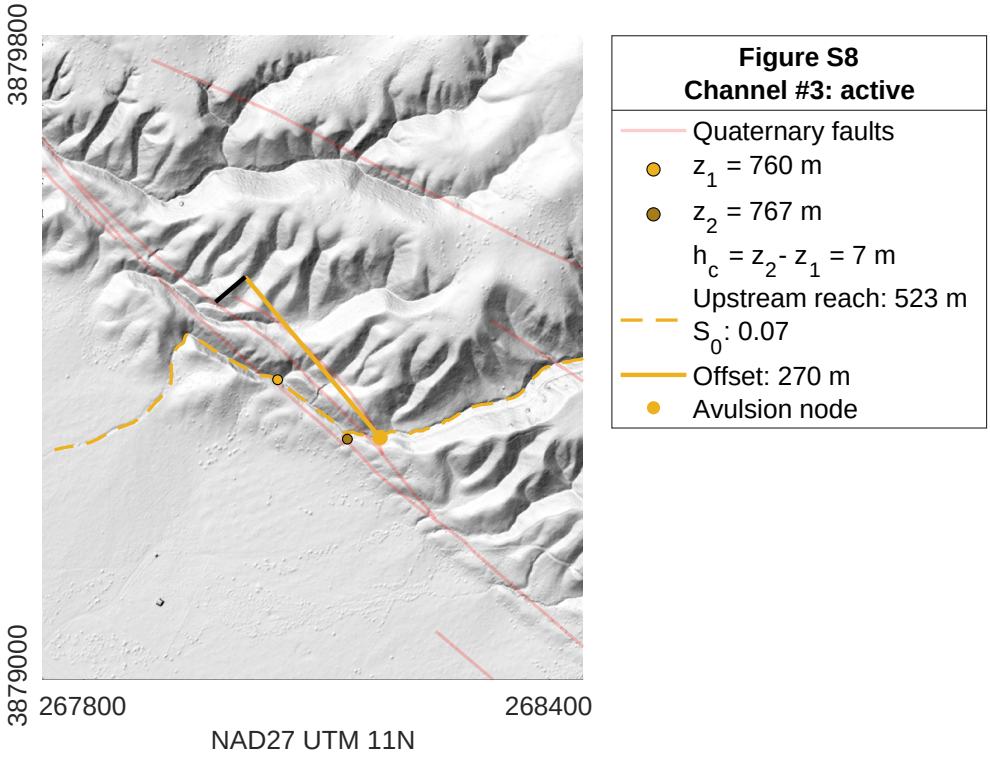

Figure A.8: Measurements for channel 3

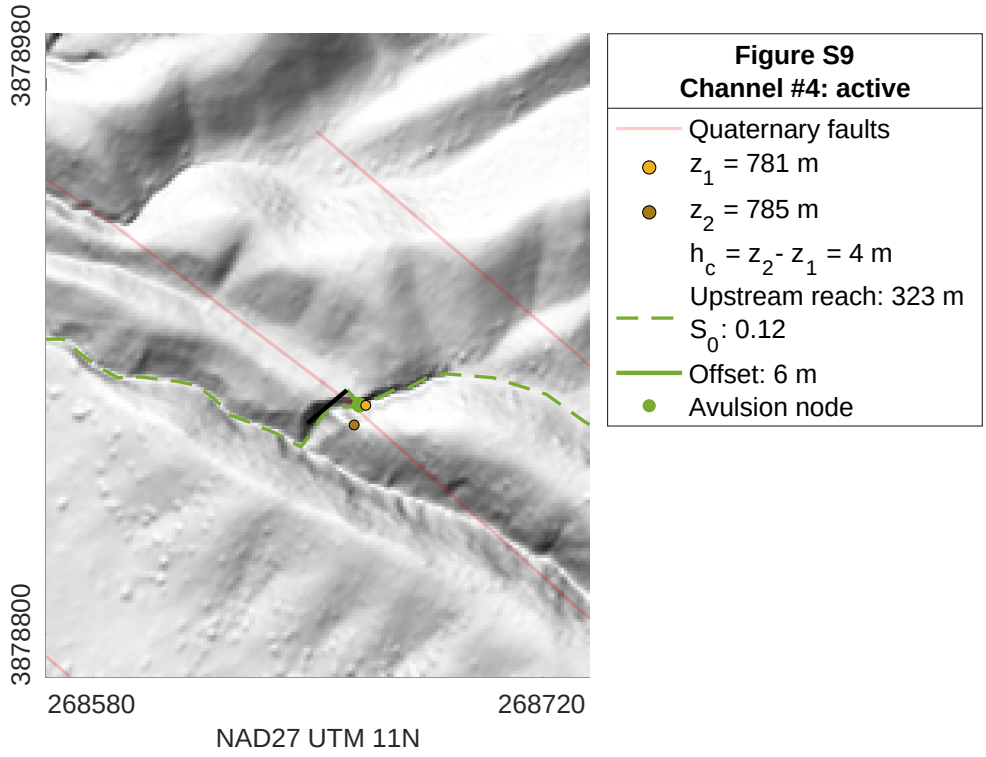

Figure A.9: Measurements for channel 4

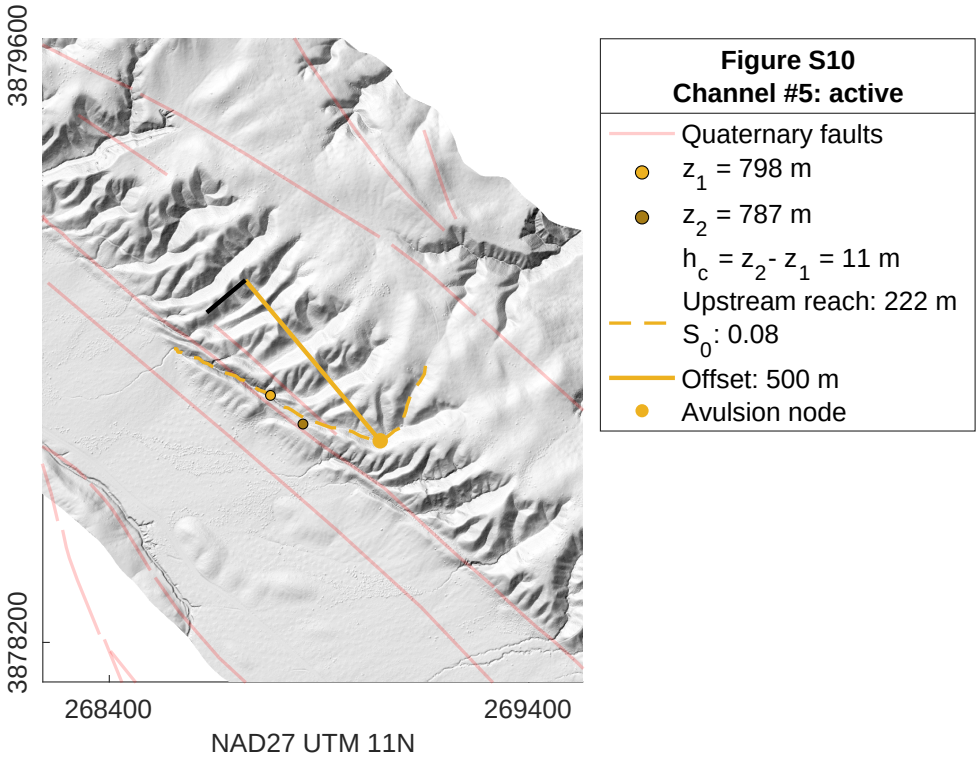

Figure A.10: Measurements for channel 5

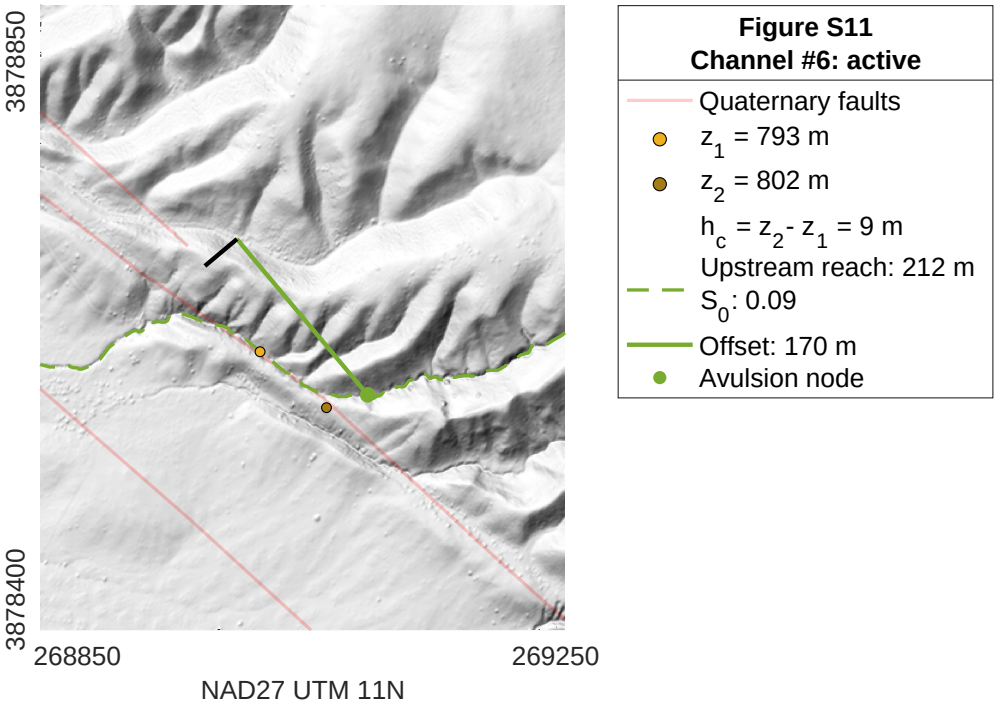

Figure A.11: Measurements for channel 6

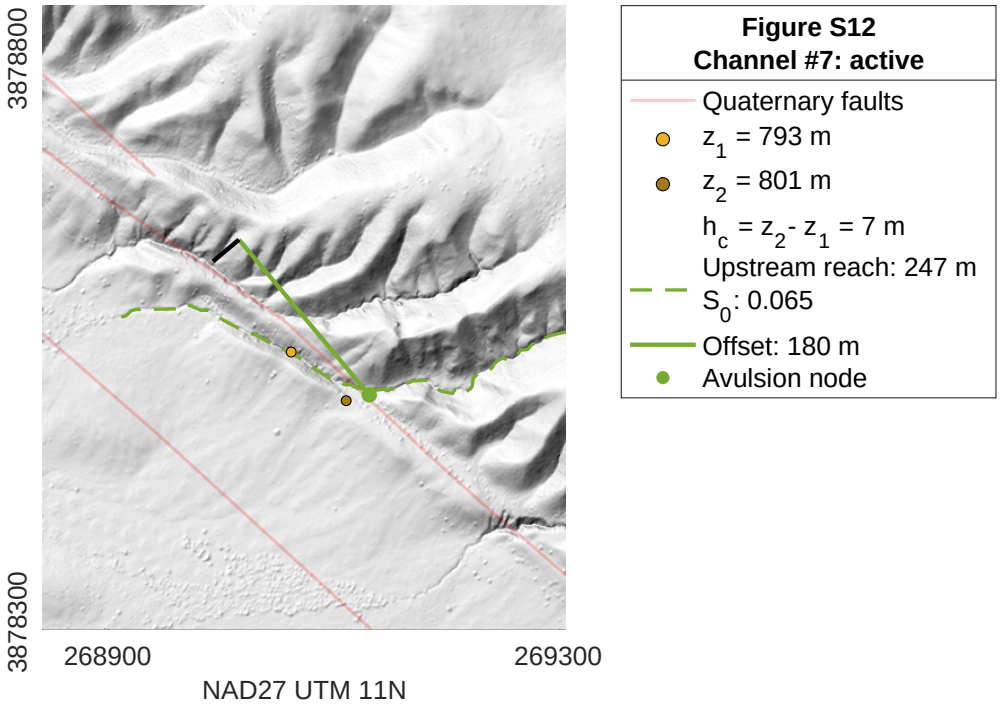

Figure A.12: Measurements for channel 7

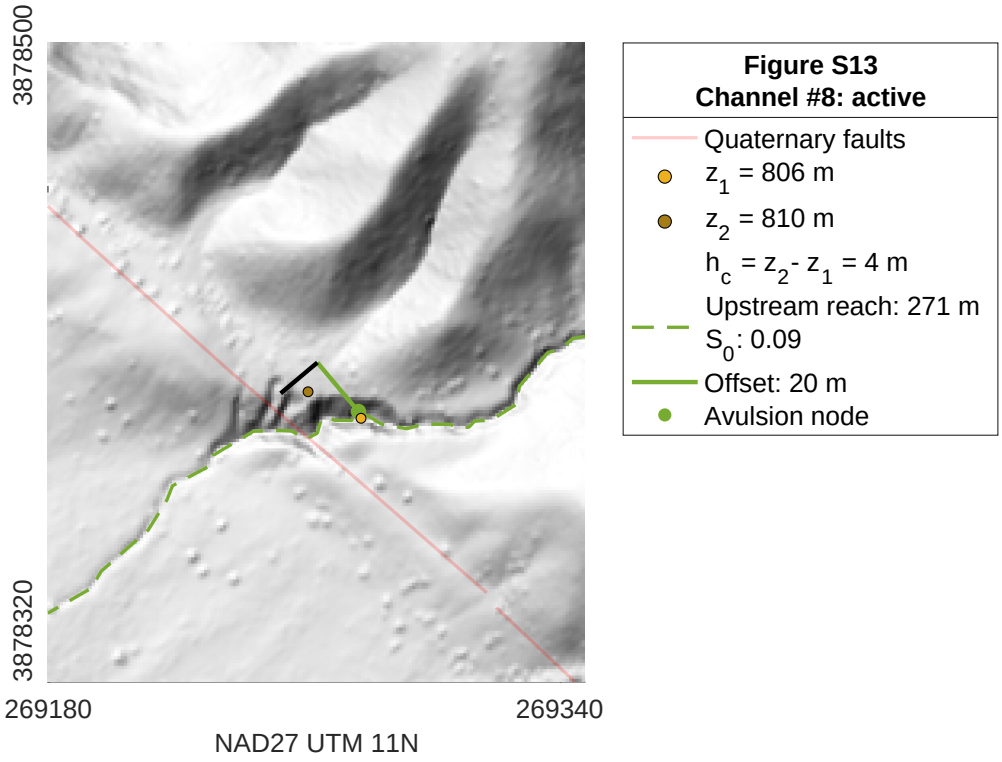

Figure A.13: Measurements for channel 8

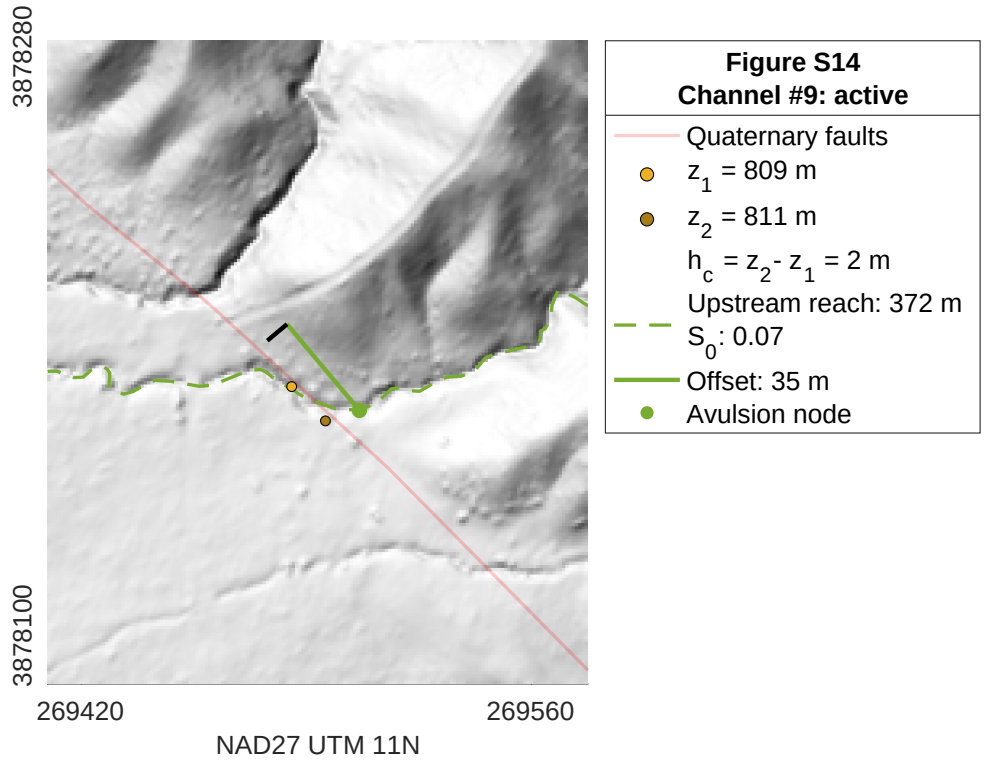

Figure A.14: Measurements for channel 9

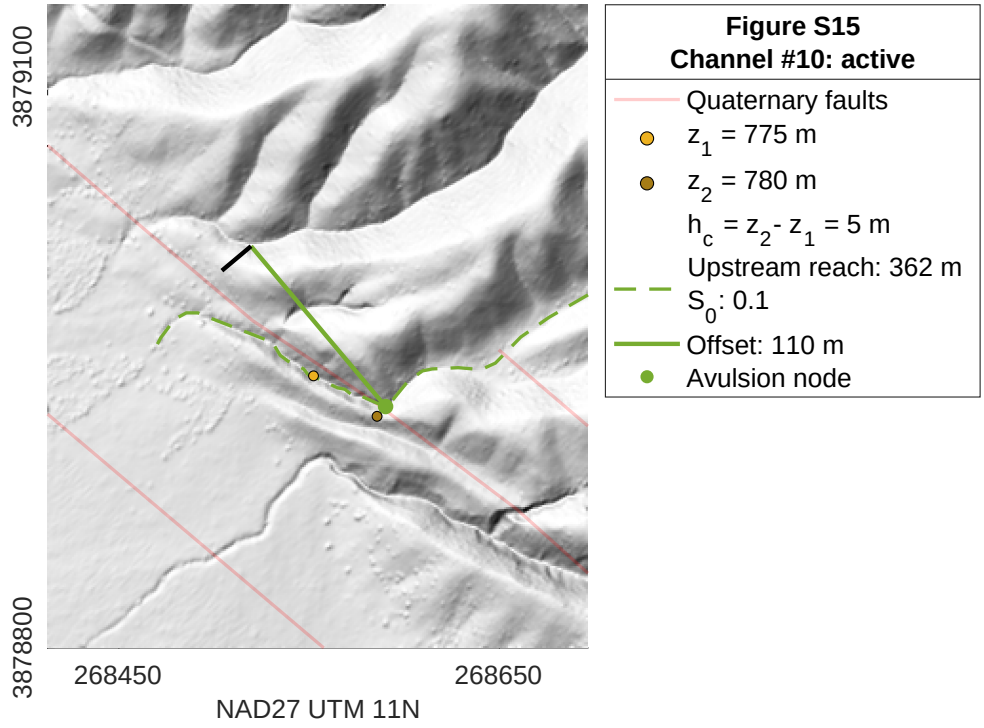

Figure A.15: Measurements for channel 10

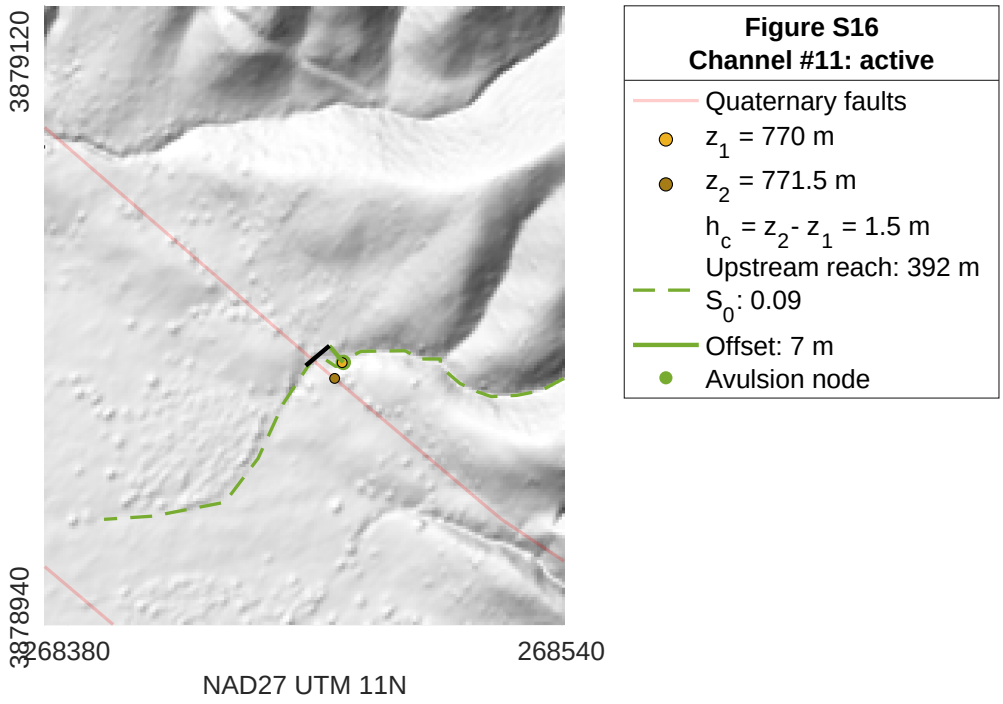

Figure A.16: Measurements for channel 11

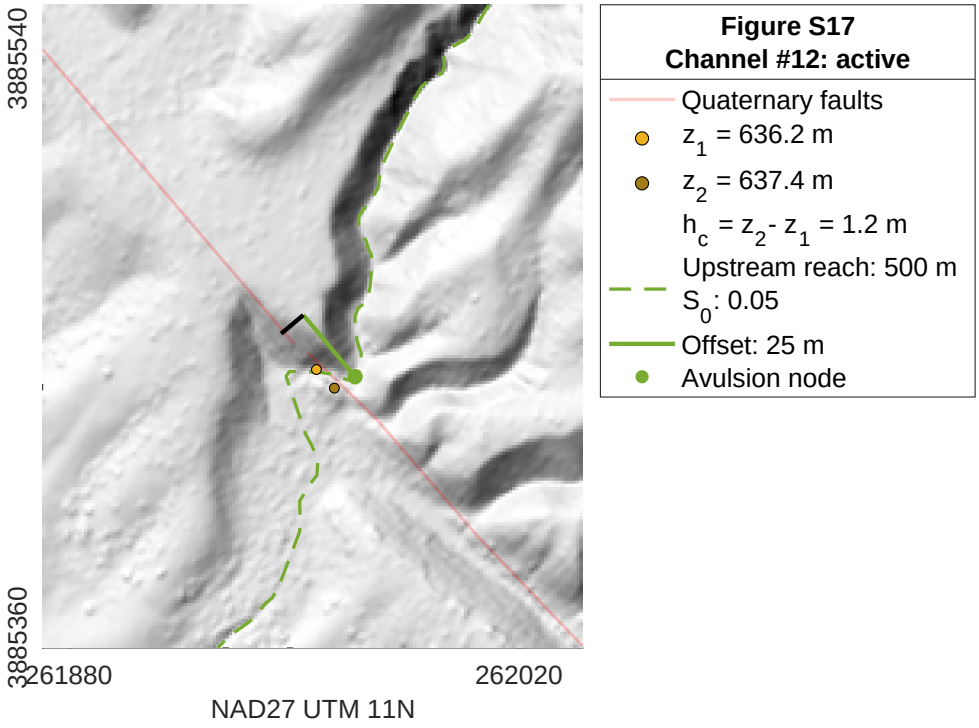

Figure A.17: Measurements for channel 12

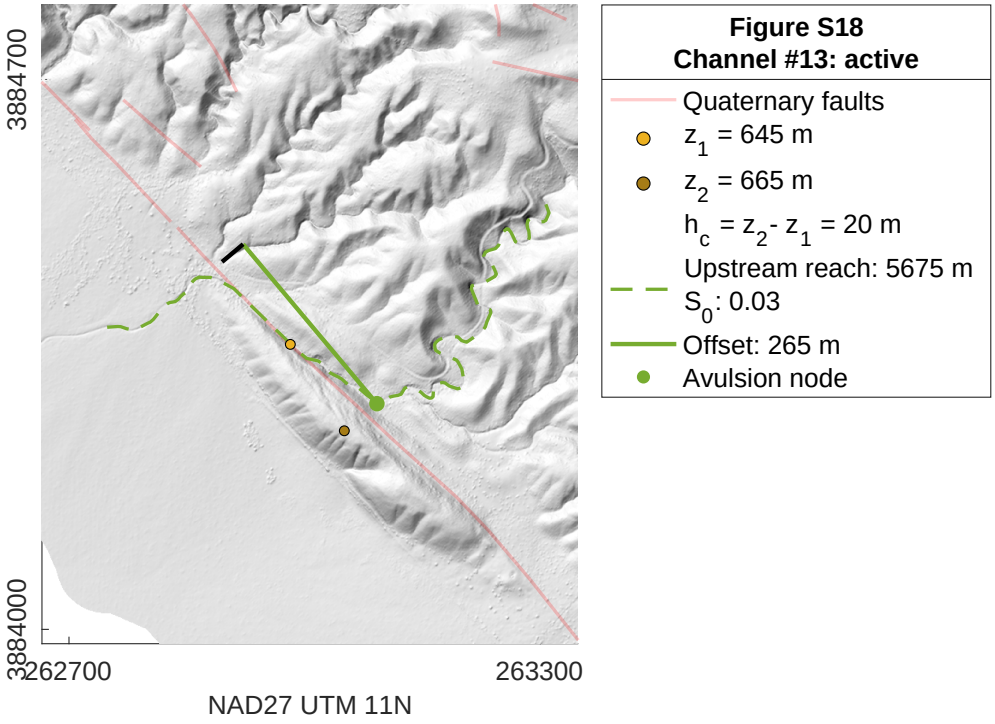

Figure A.18: Measurements for channel 13

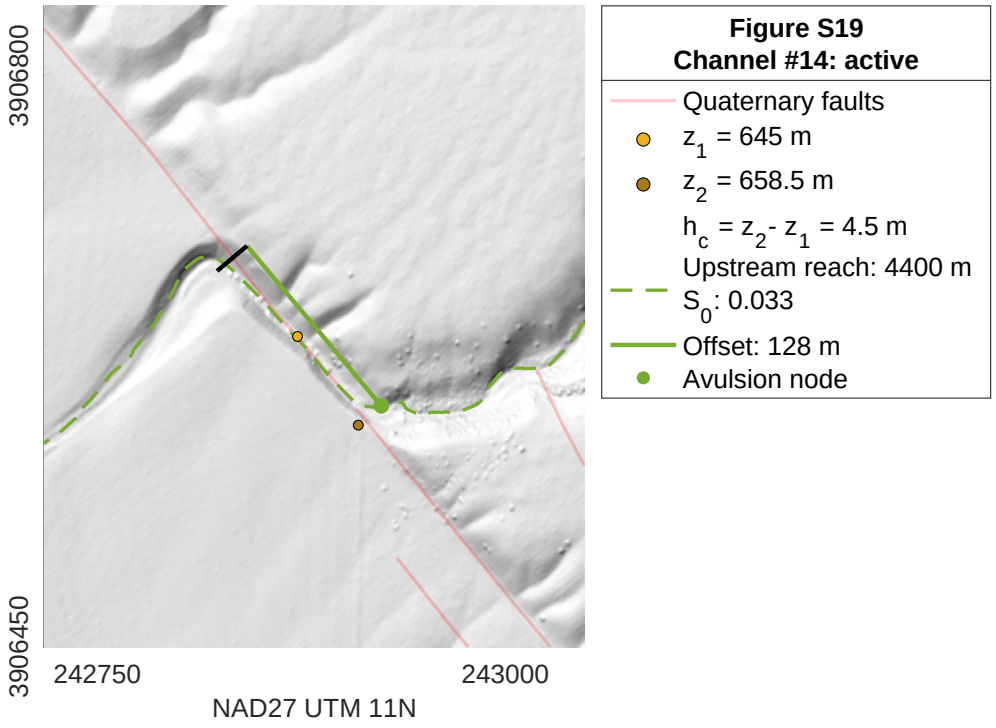

Figure A.19: Measurements for channel 14

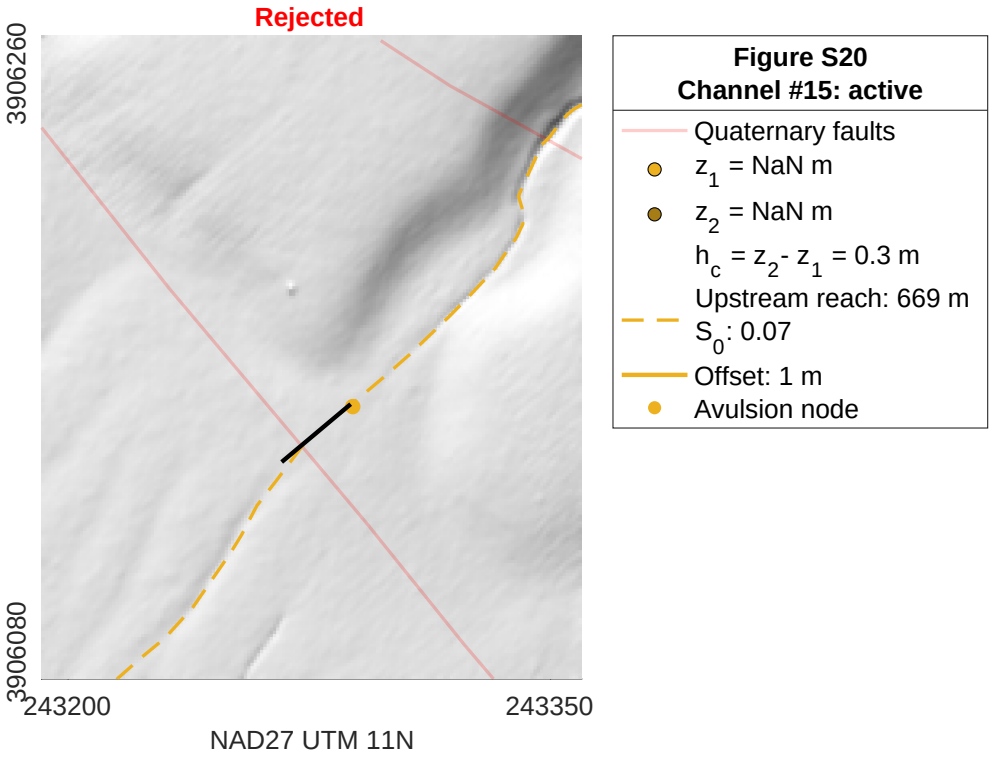

Figure A.20: Measurements for channel 15

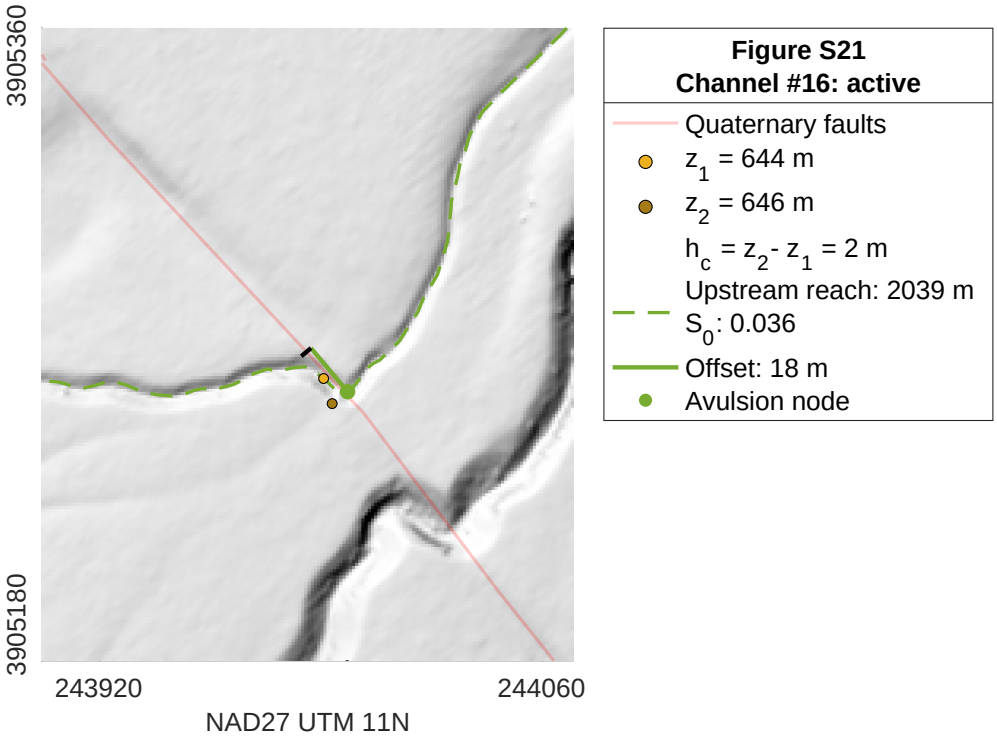

Figure A.21: Measurements for channel 16

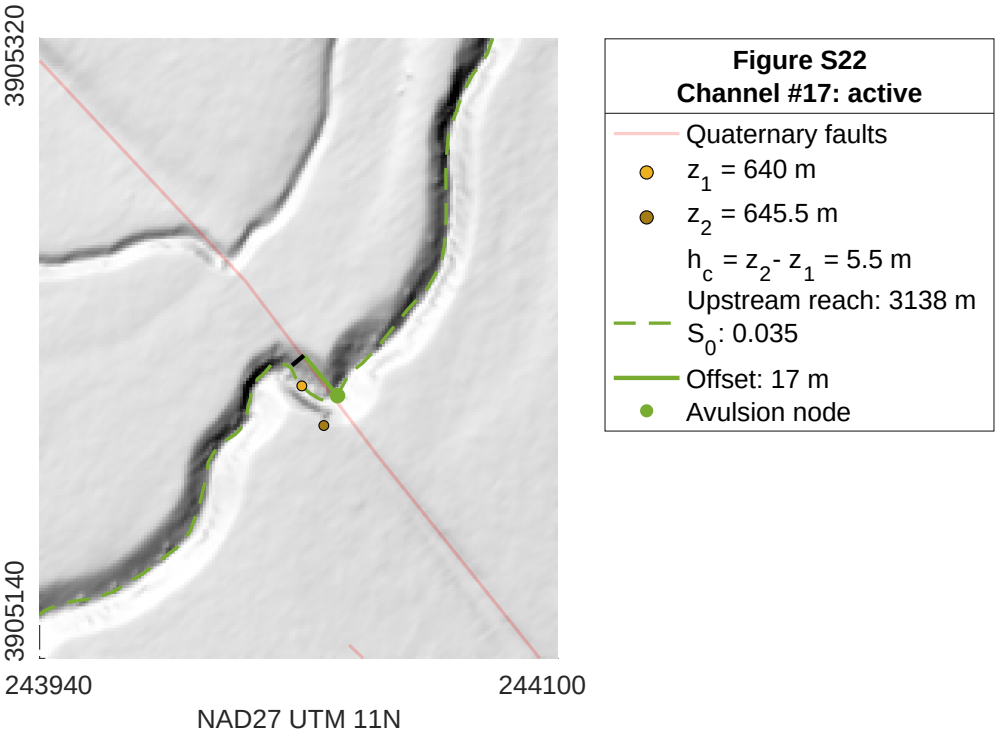

Figure A.22: Measurements for channel 17

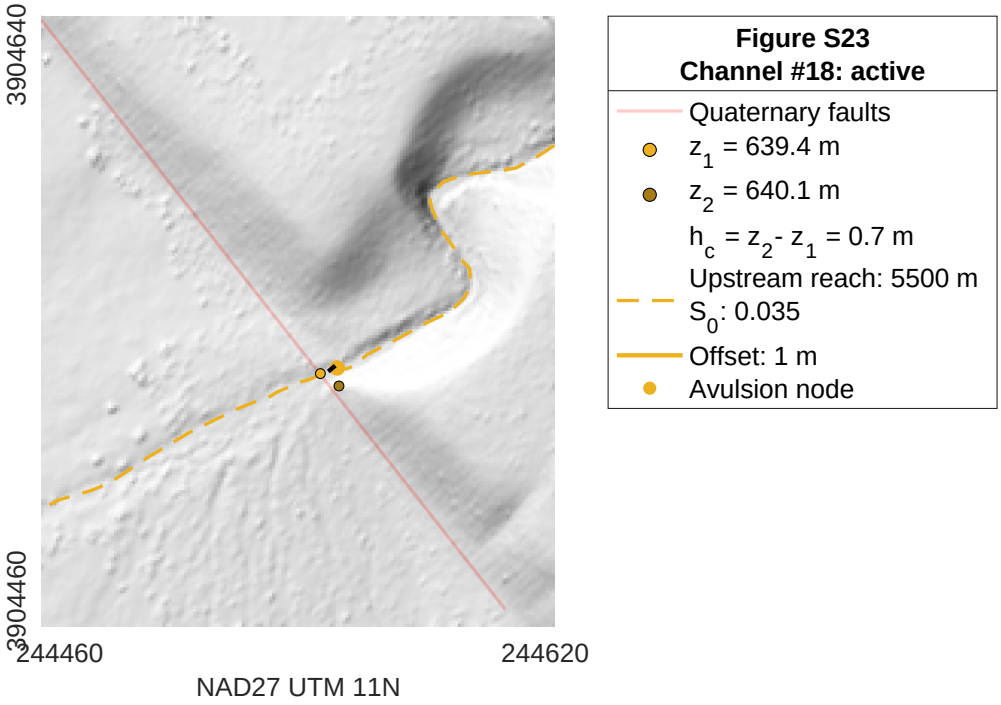

Figure A.23: Measurements for channel 18

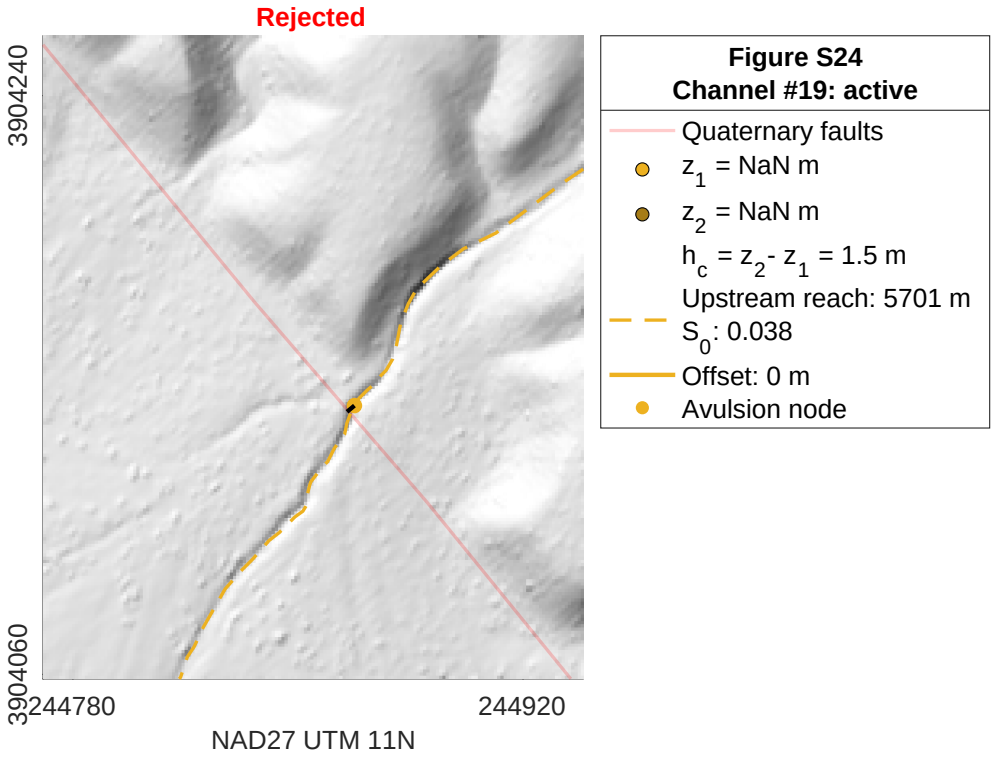

Figure A.24: Measurements for channel 19

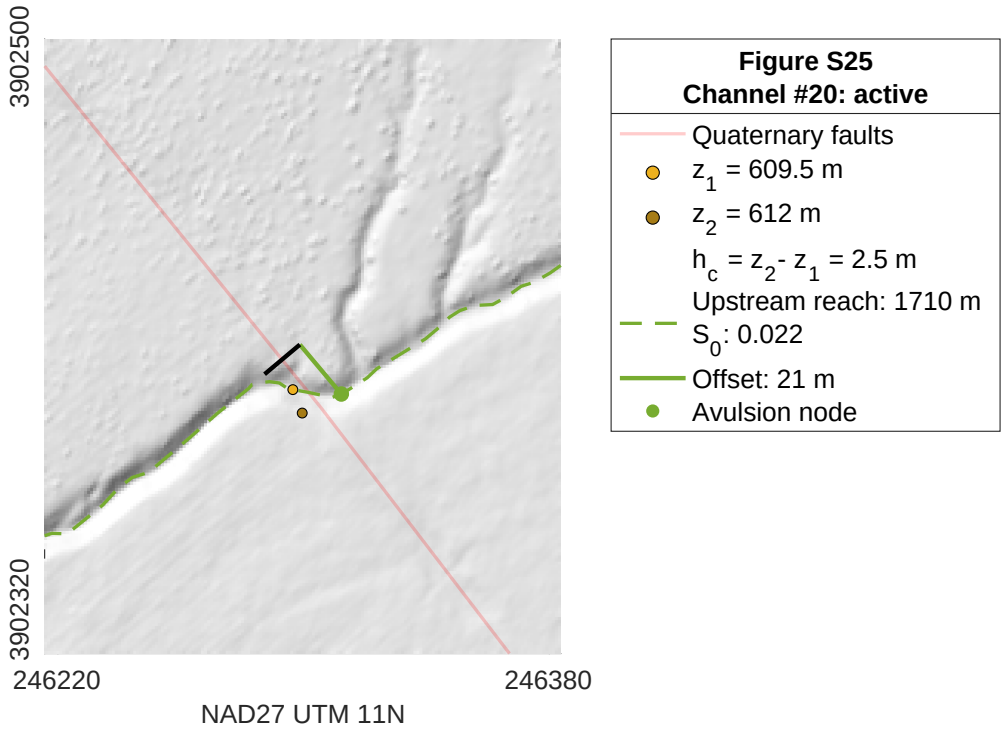

Figure A.25: Measurements for channel 20

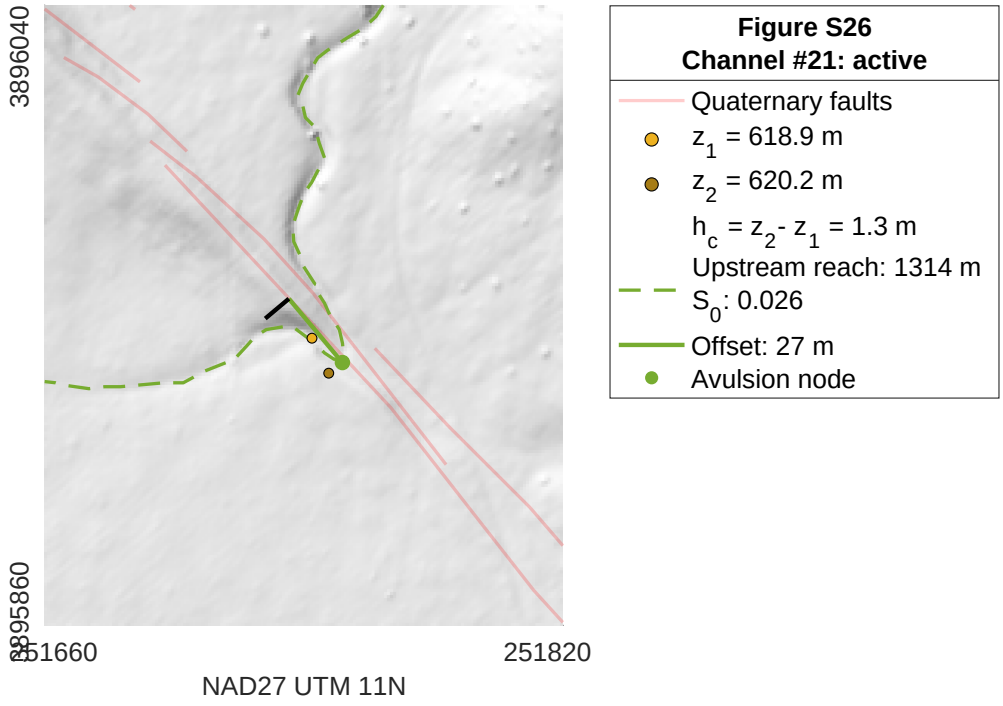

Figure A.26: Measurements for channel 21

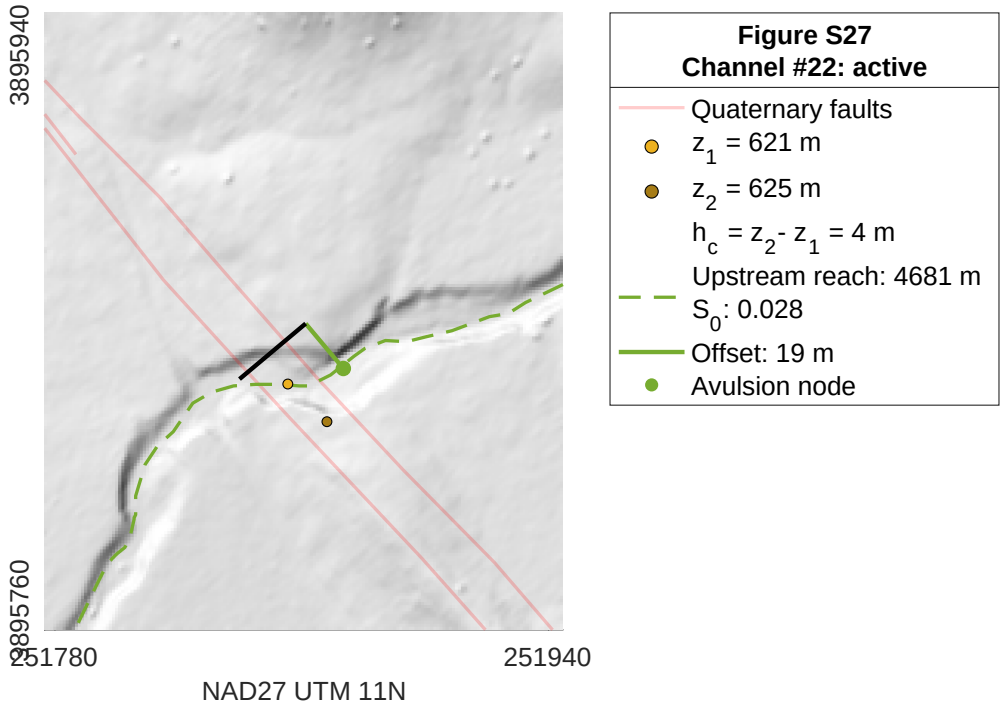

Figure A.27: Measurements for channel 22

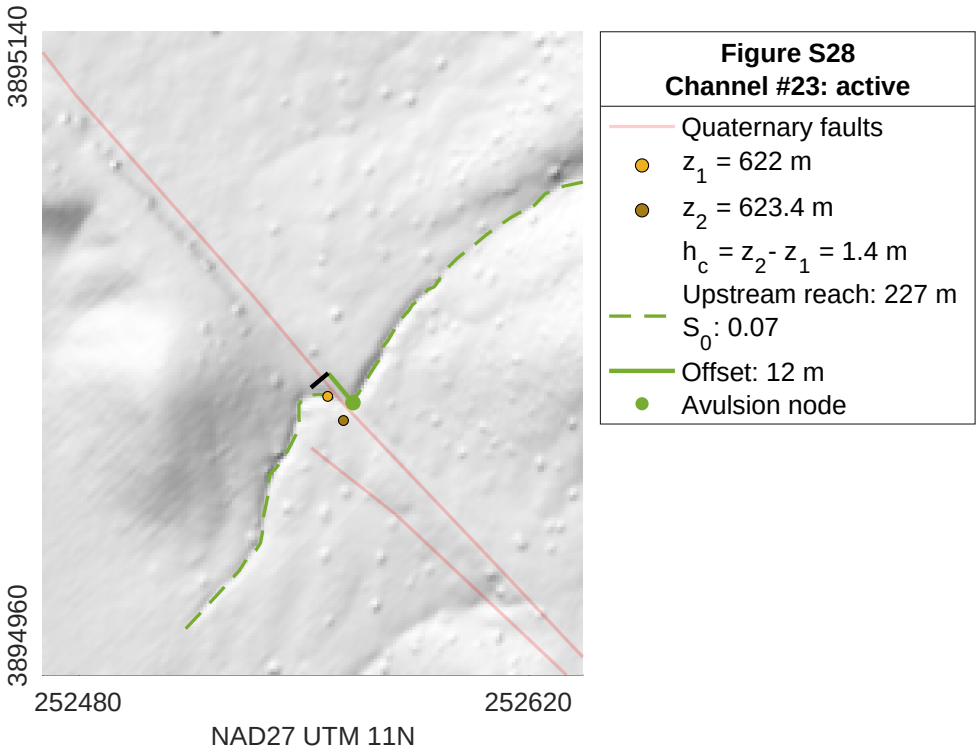

Figure A.28: Measurements for channel 23

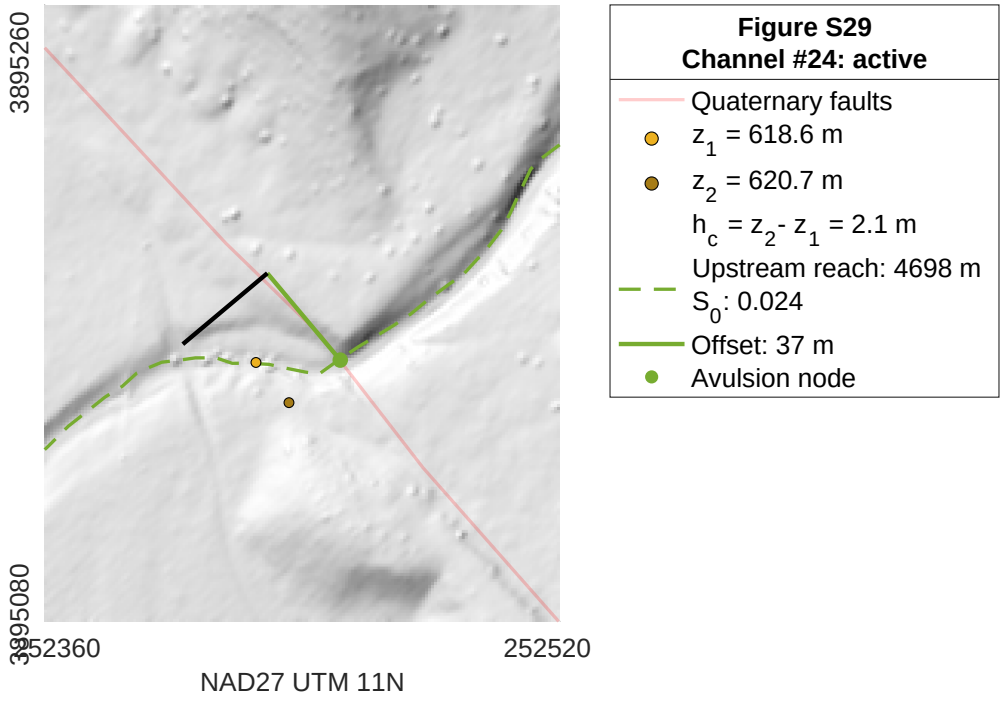

Figure A.29: Measurements for channel 24

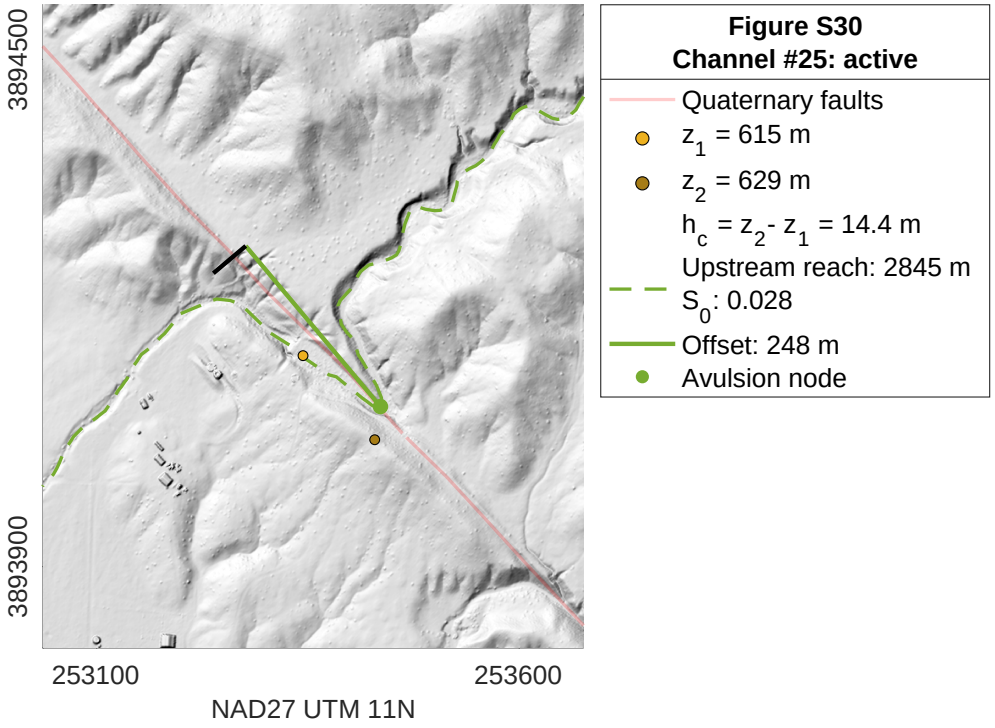

Figure A.30: Measurements for channel 25

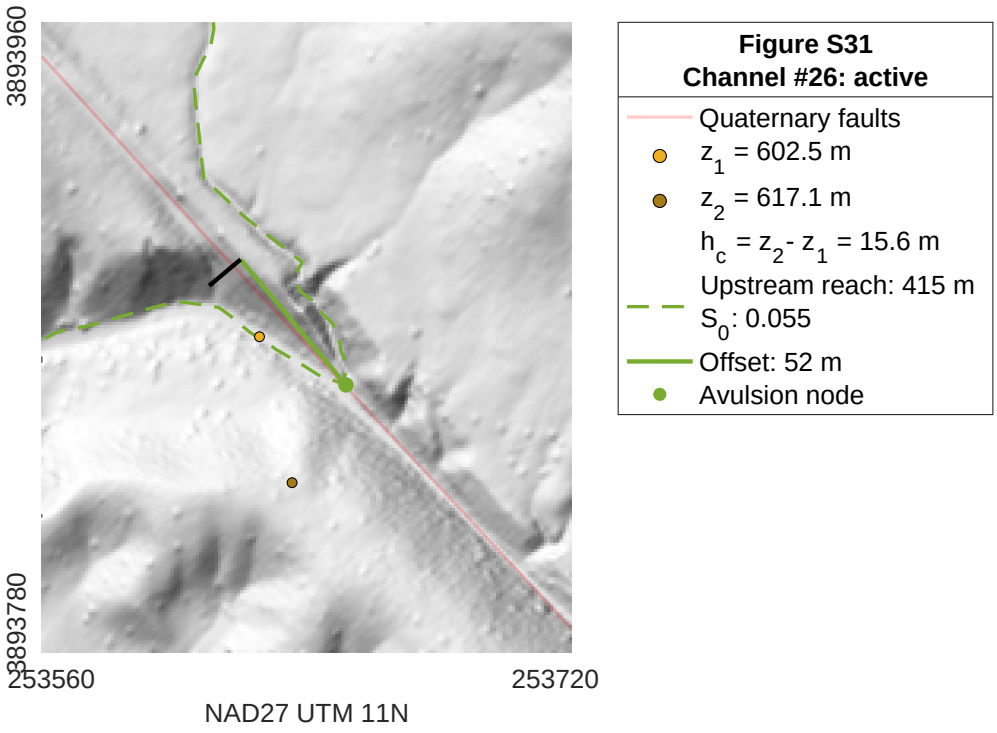

Figure A.31: Measurements for channel 26

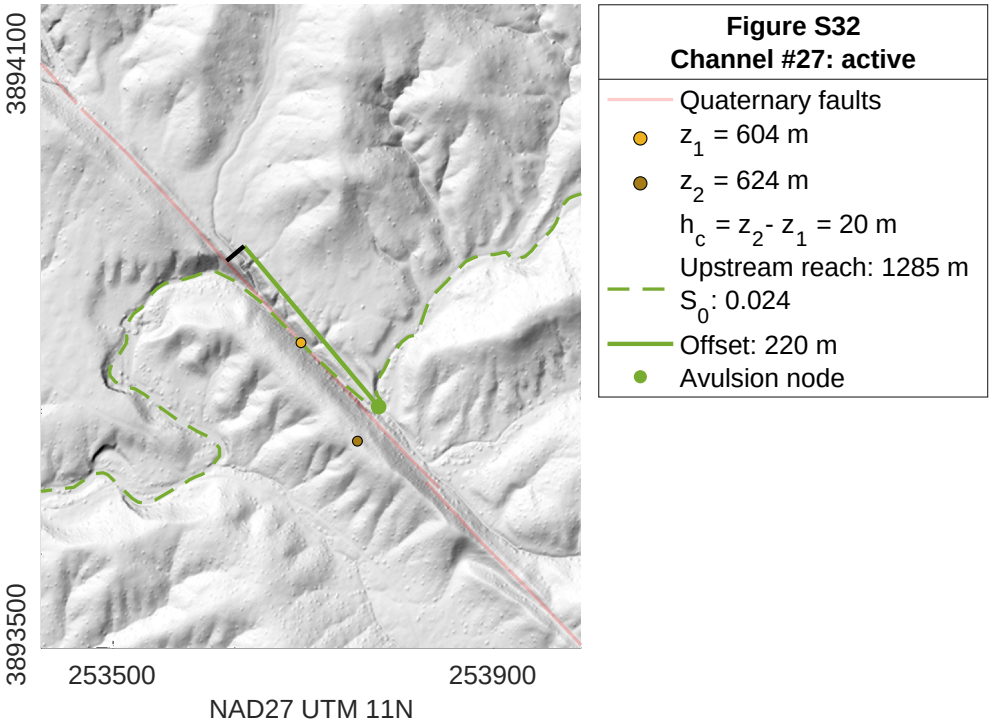

Figure A.32: Measurements for channel 27

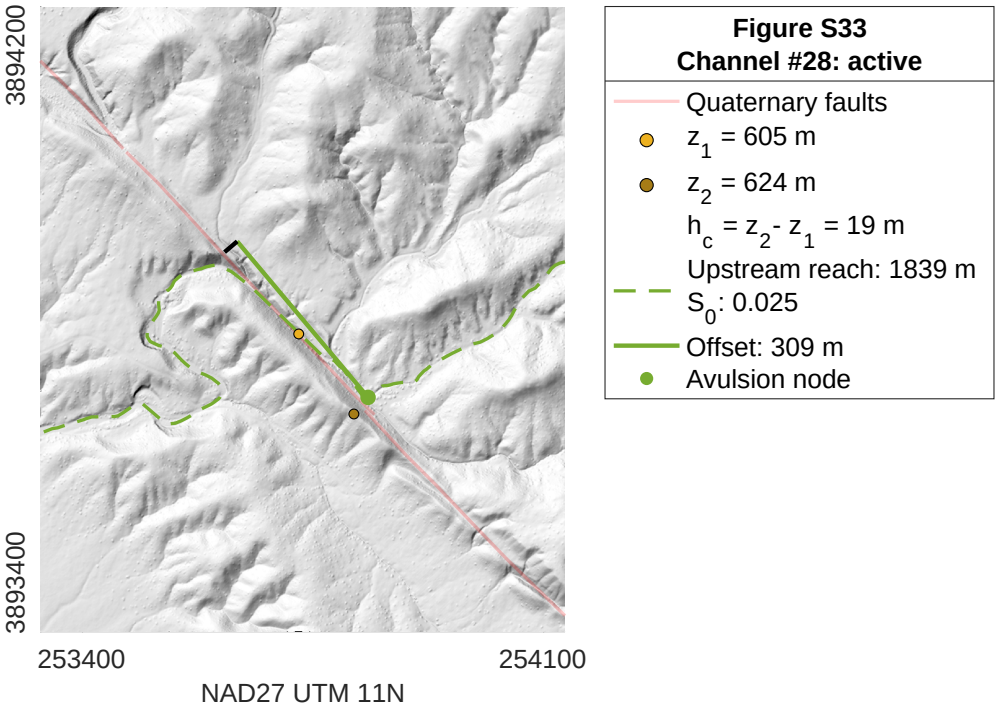

Figure A.33: Measurements for channel 28

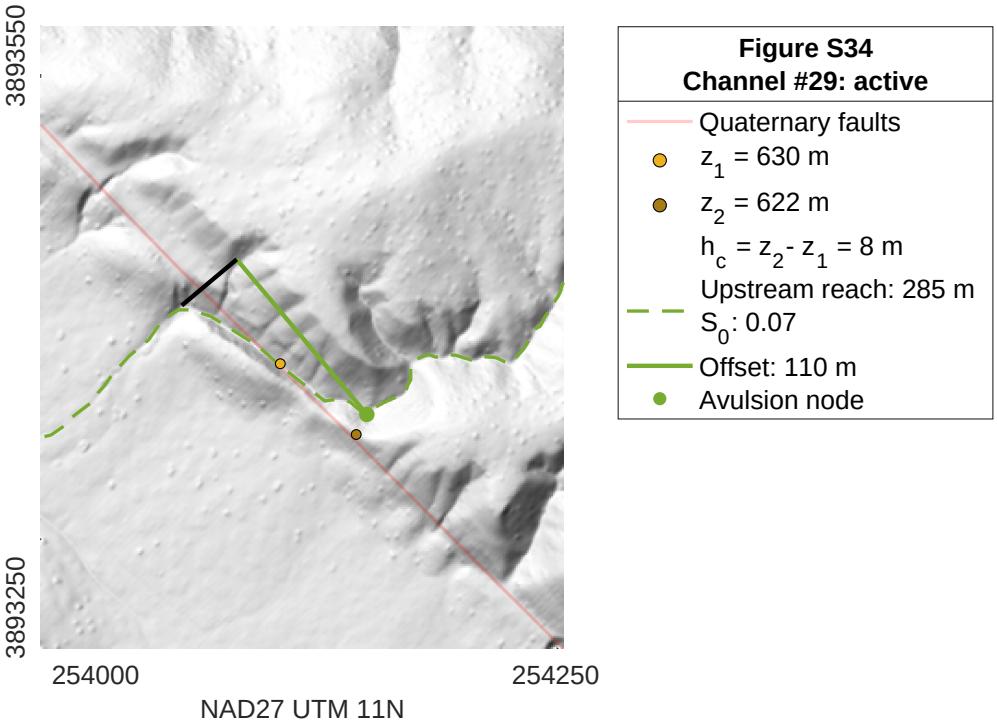

Figure A.34: Measurements for channel 29

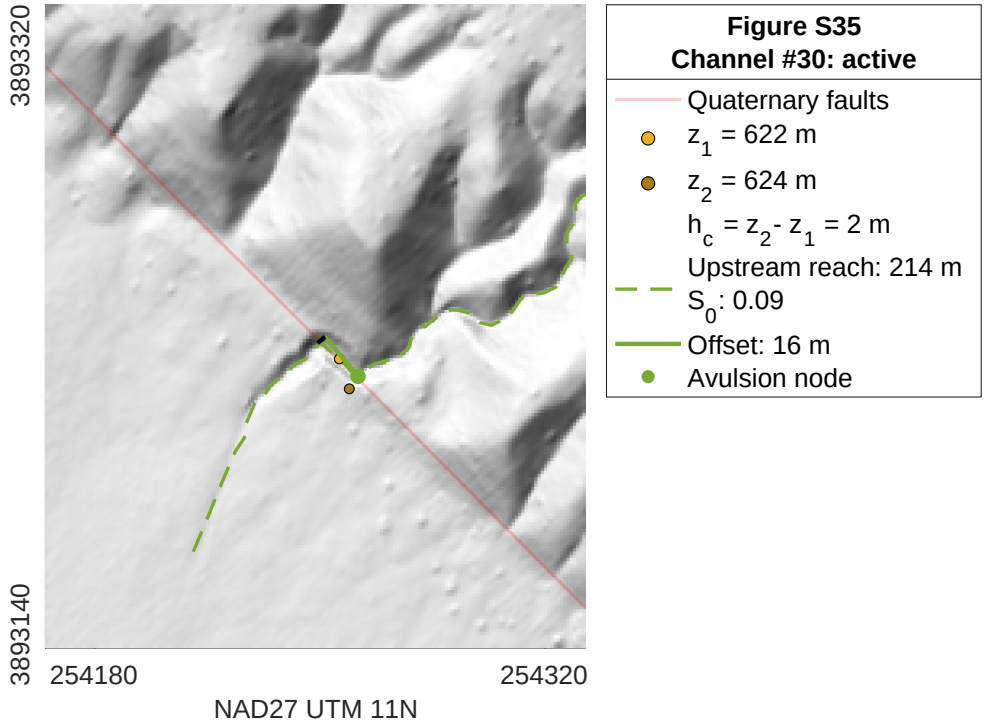

Figure A.35: Measurements for channel 30

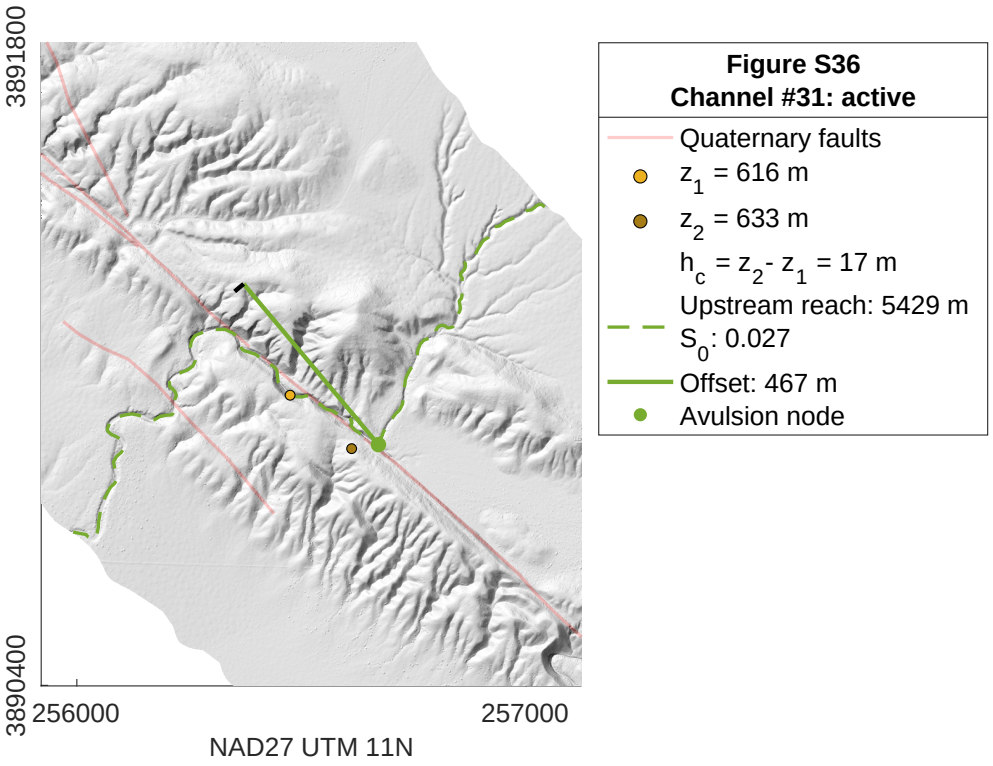

Figure A.36: Measurements for channel 31

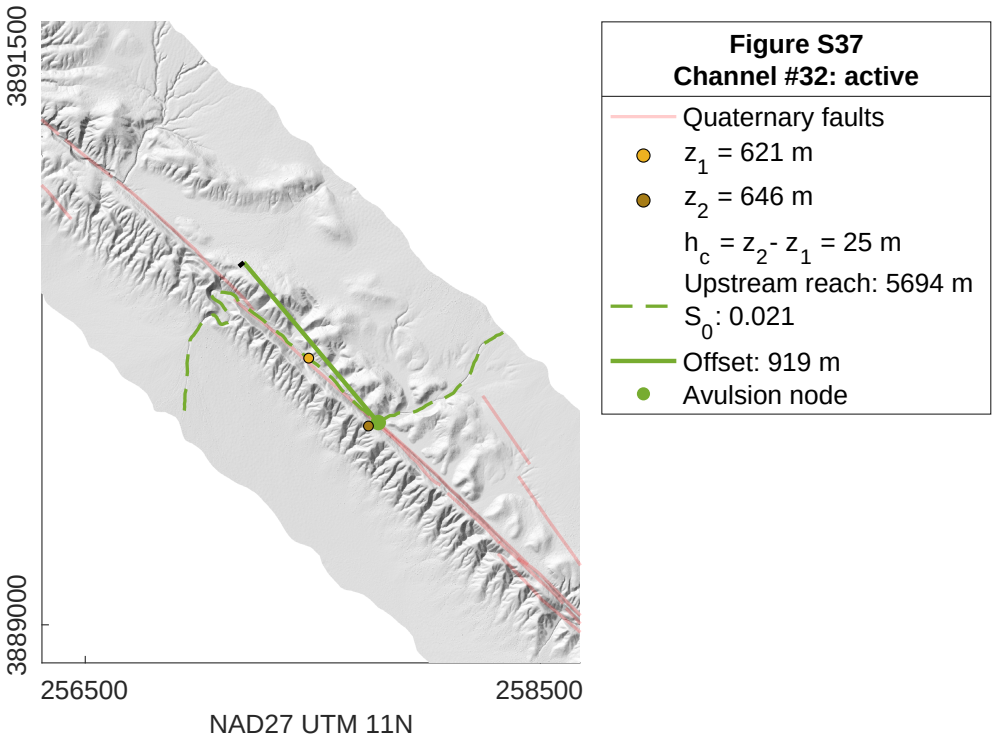

Figure A.37: Measurements for channel 32

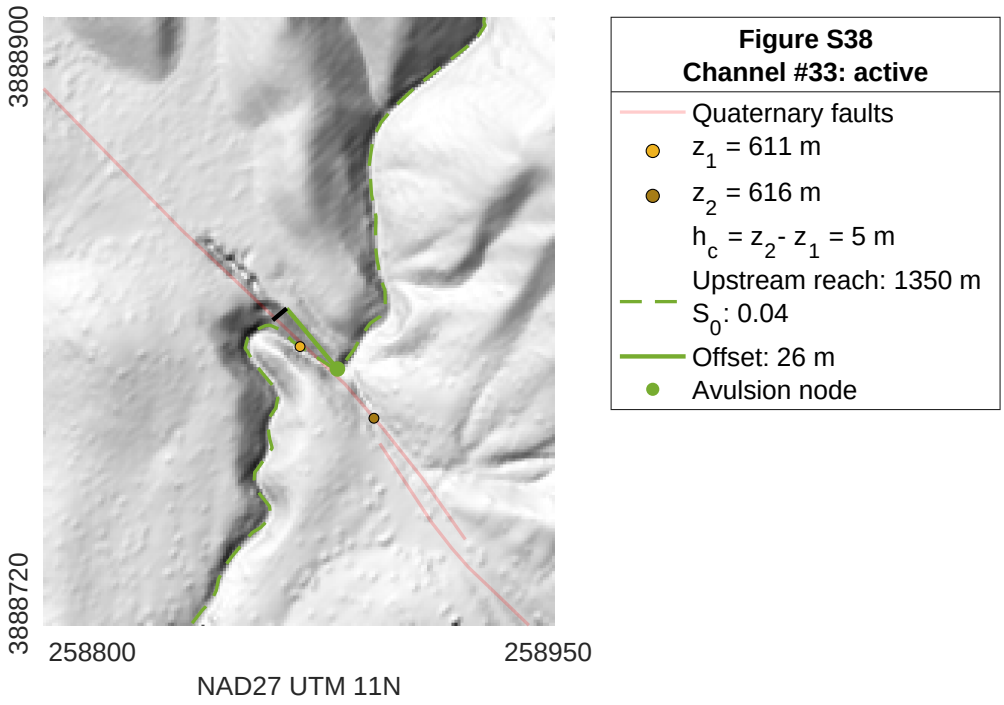

Figure A.38: Measurements for channel 33

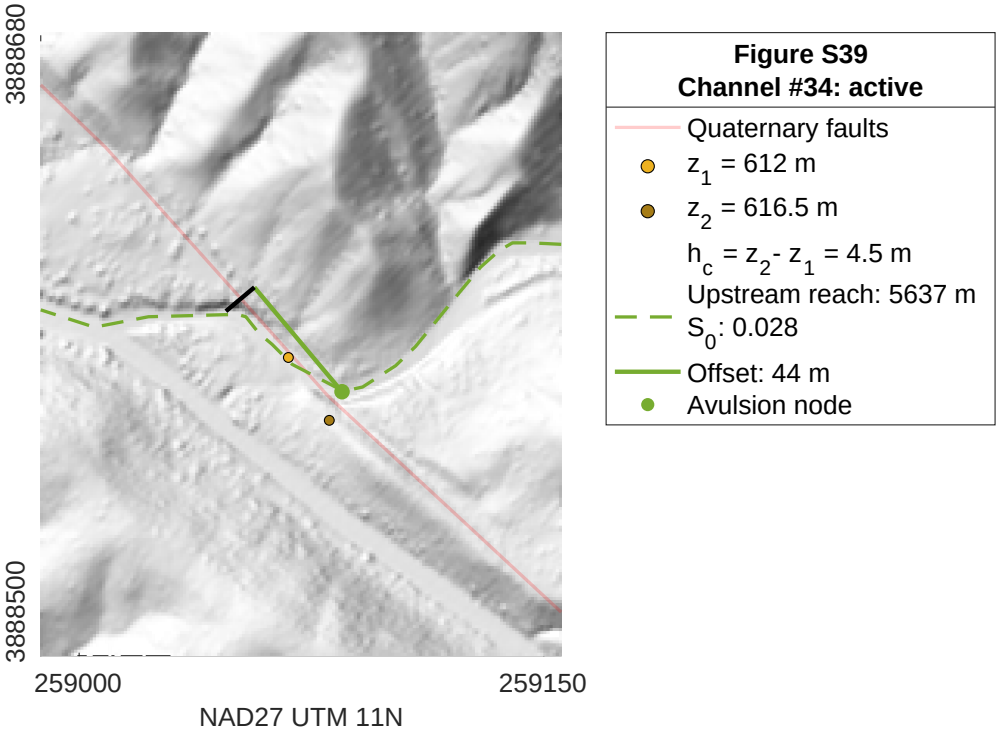

Figure A.39: Measurements for channel 34

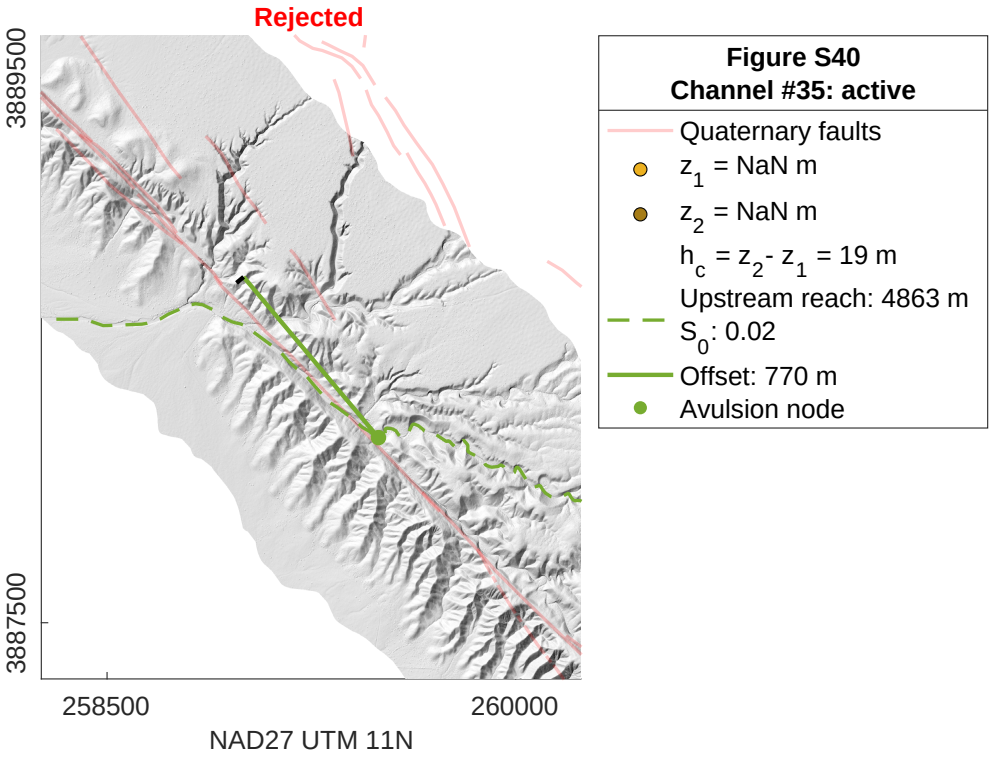

Figure A.40: Measurements for channel 35

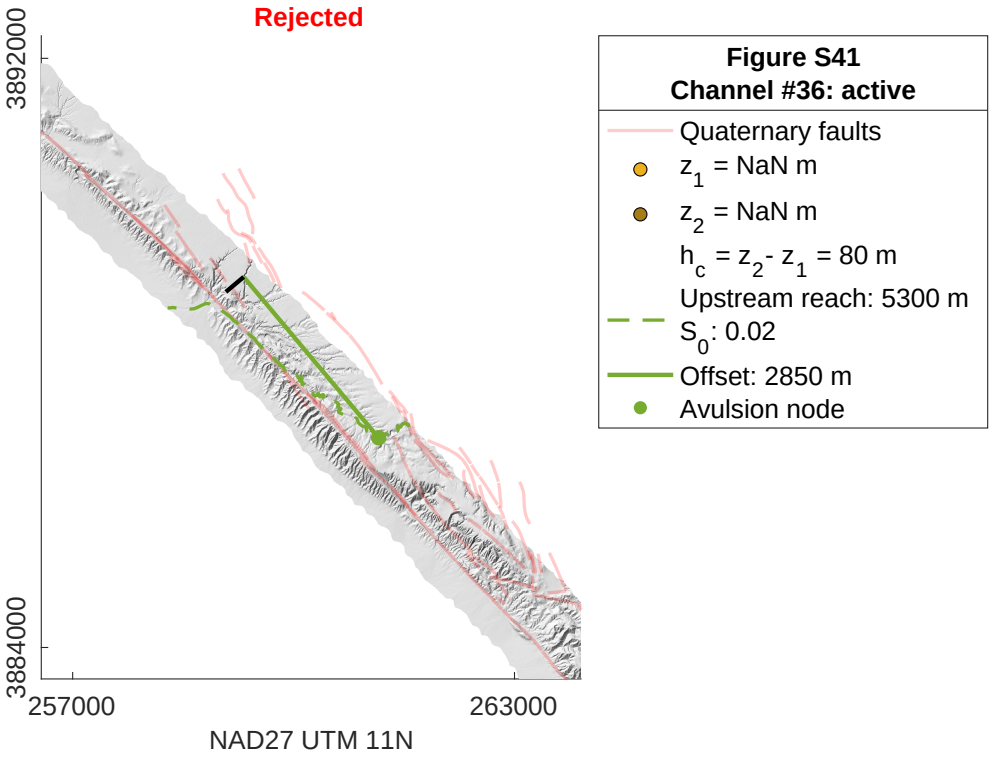

Figure A.41: Measurements for channel 36

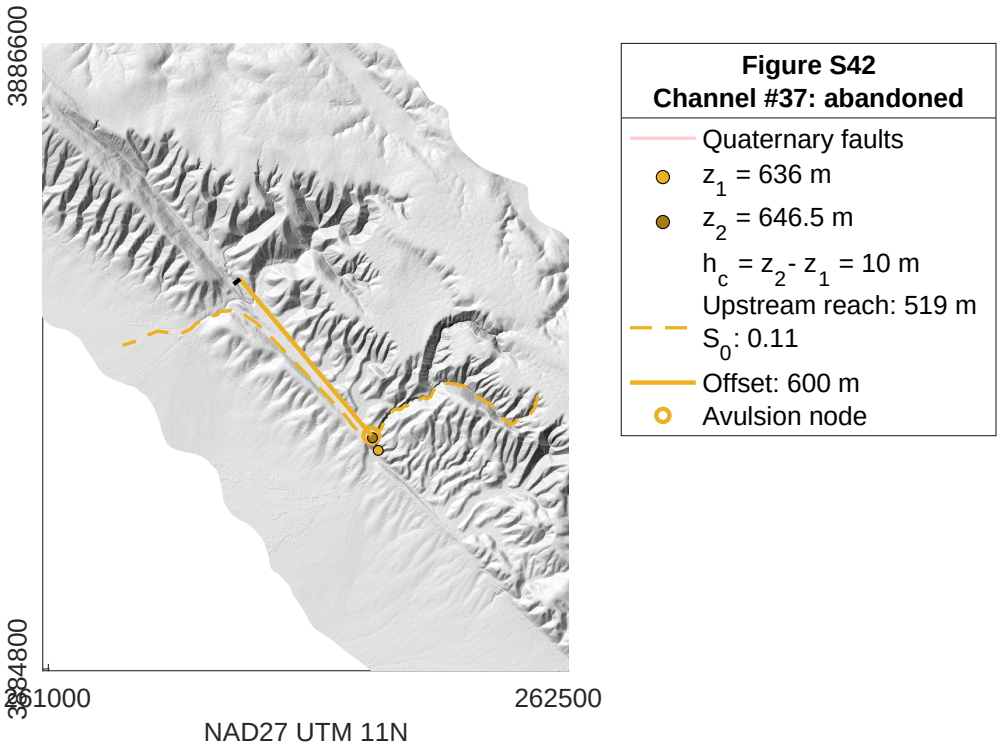

Figure A.42: Measurements for channel 37

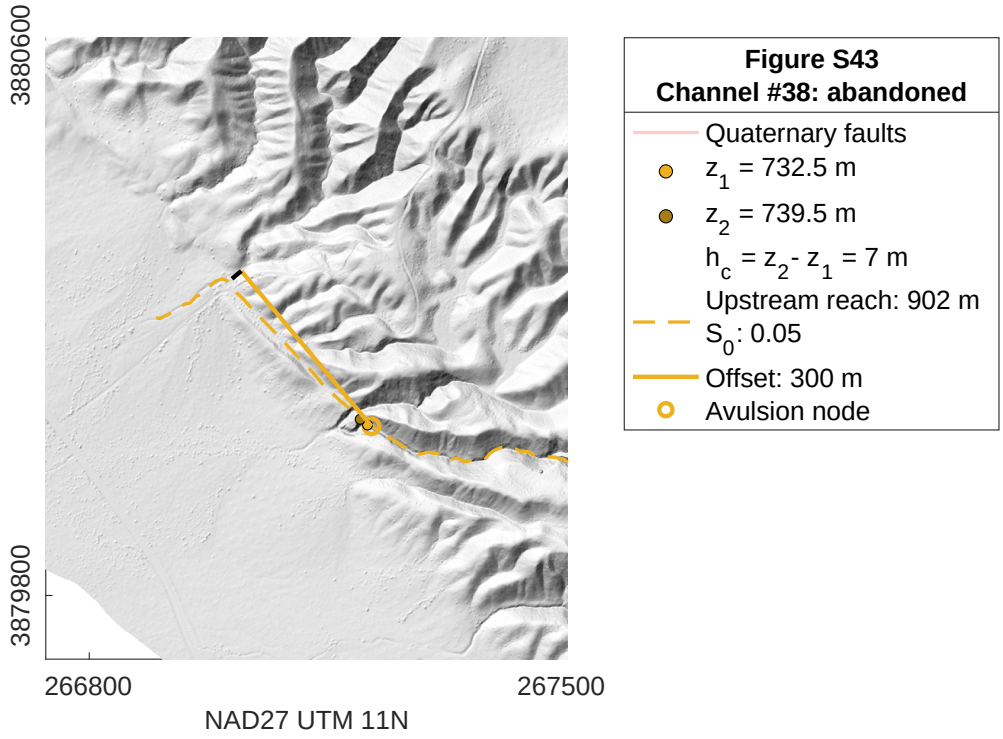

Figure A.43: Measurements for channel 38

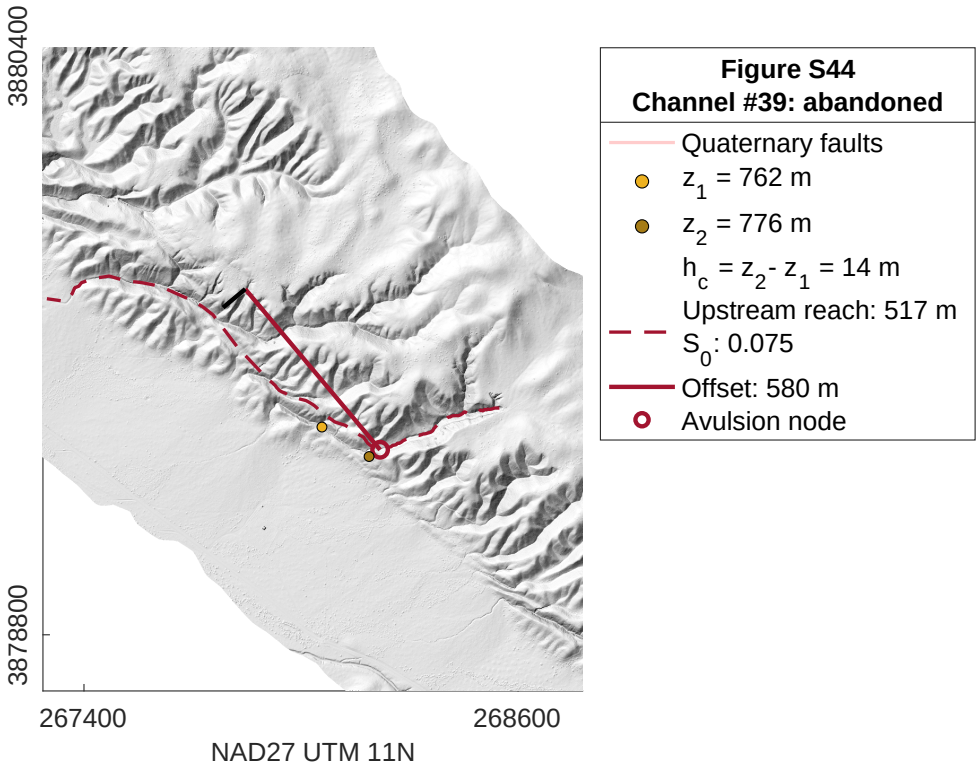

Figure A.44: Measurements for channel 39

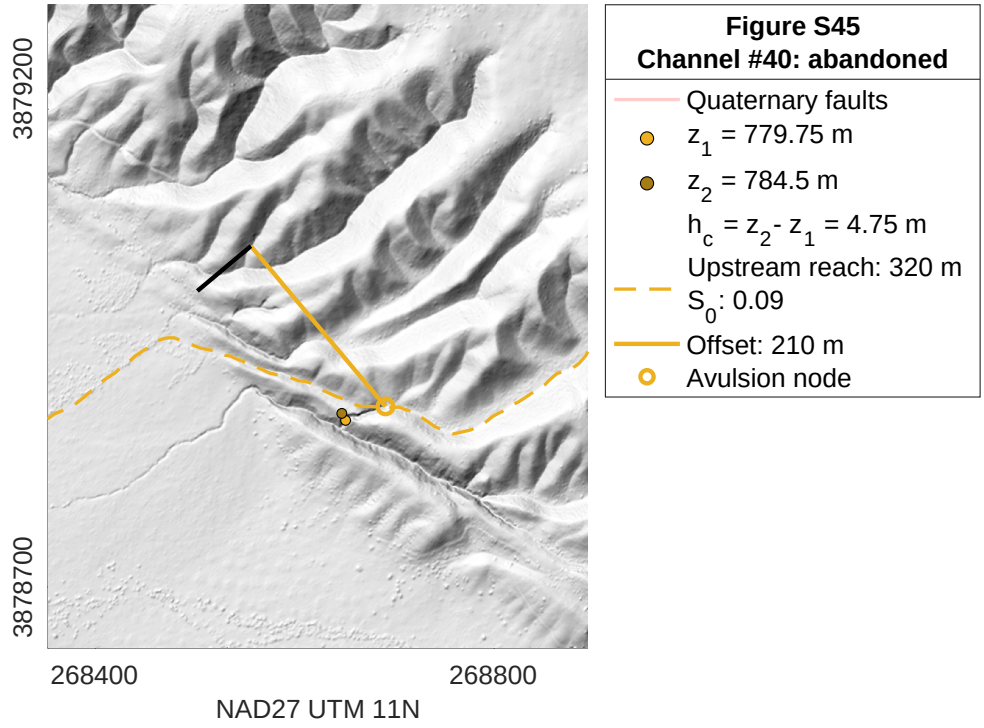

Figure A.45: Measurements for channel 40

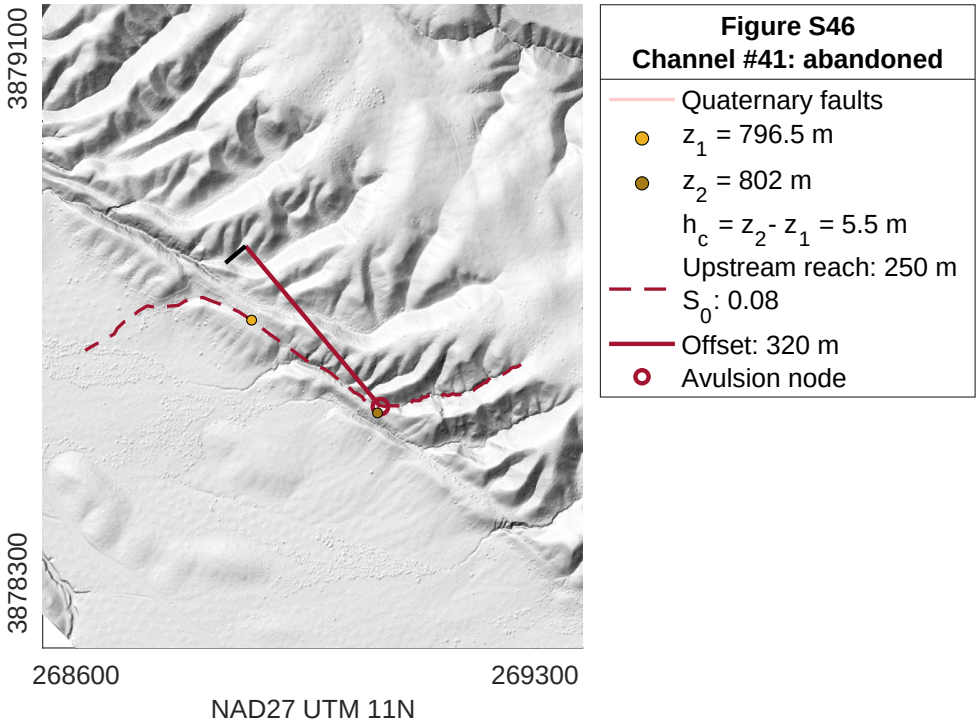

Figure A.46: Measurements for channel 41

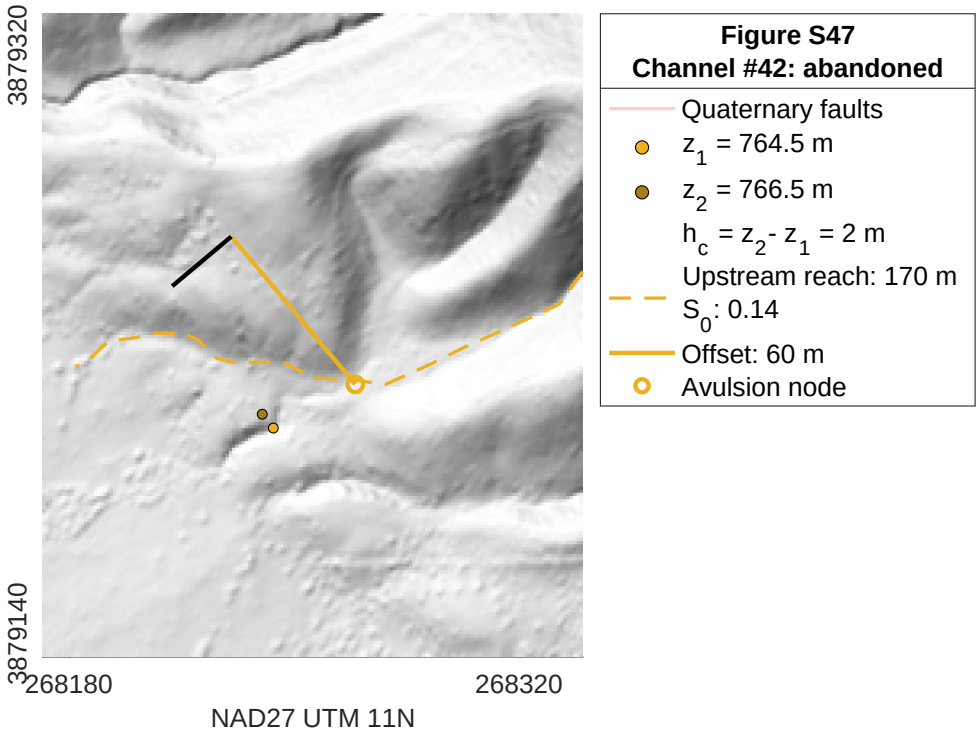

Figure A.47: Measurements for channel 42

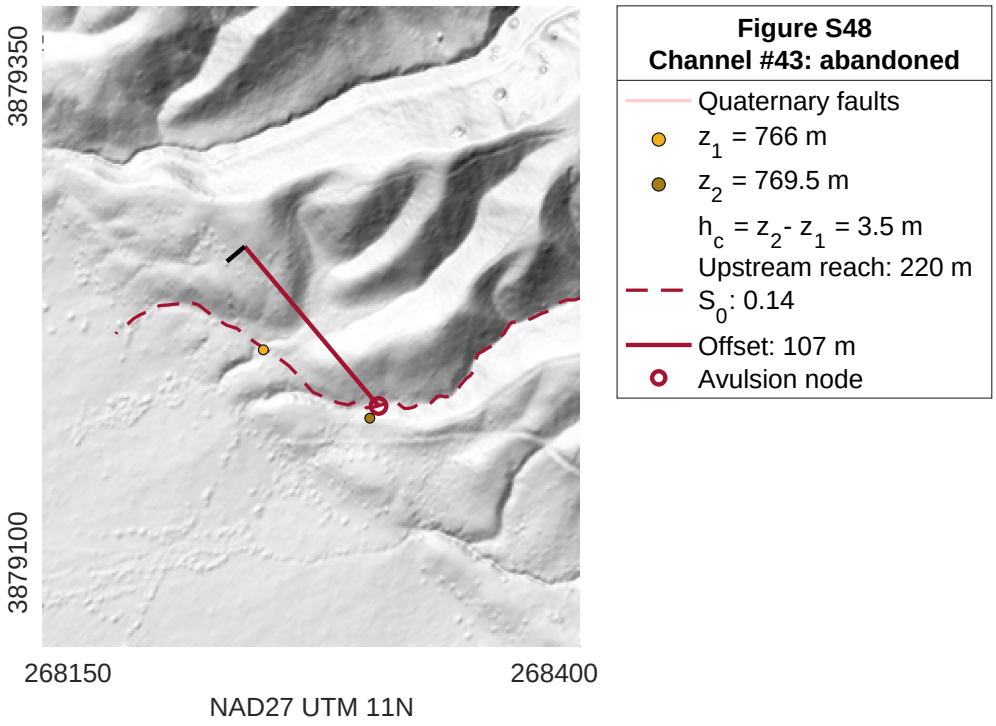

Figure A.48: Measurements for channel 43

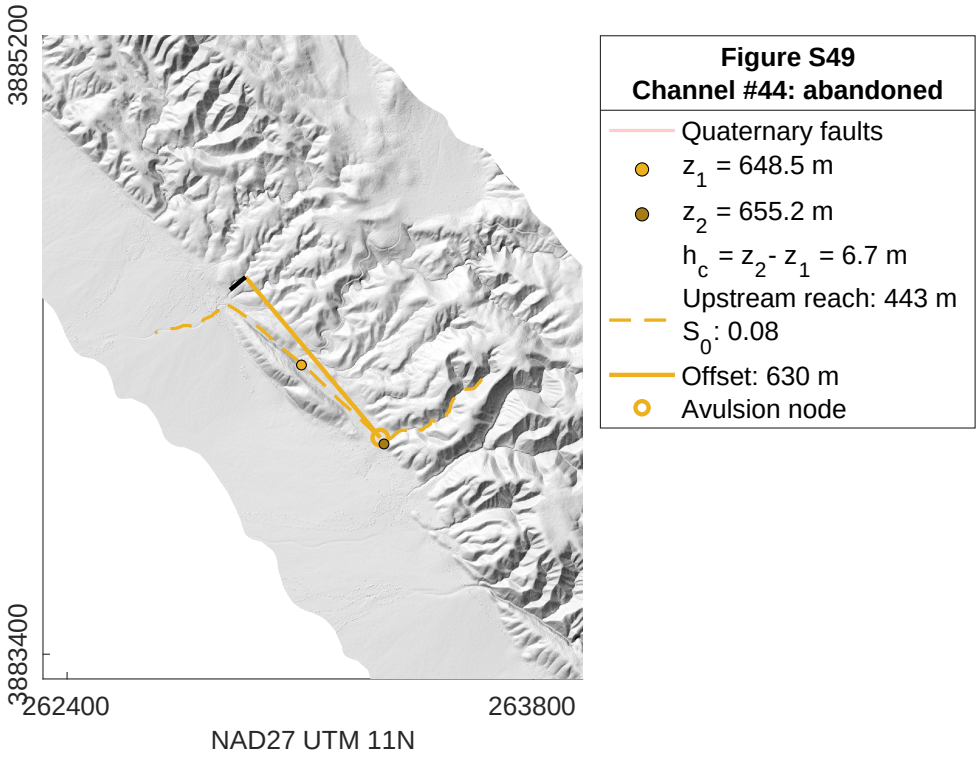

Figure A.49: Measurements for channel 44

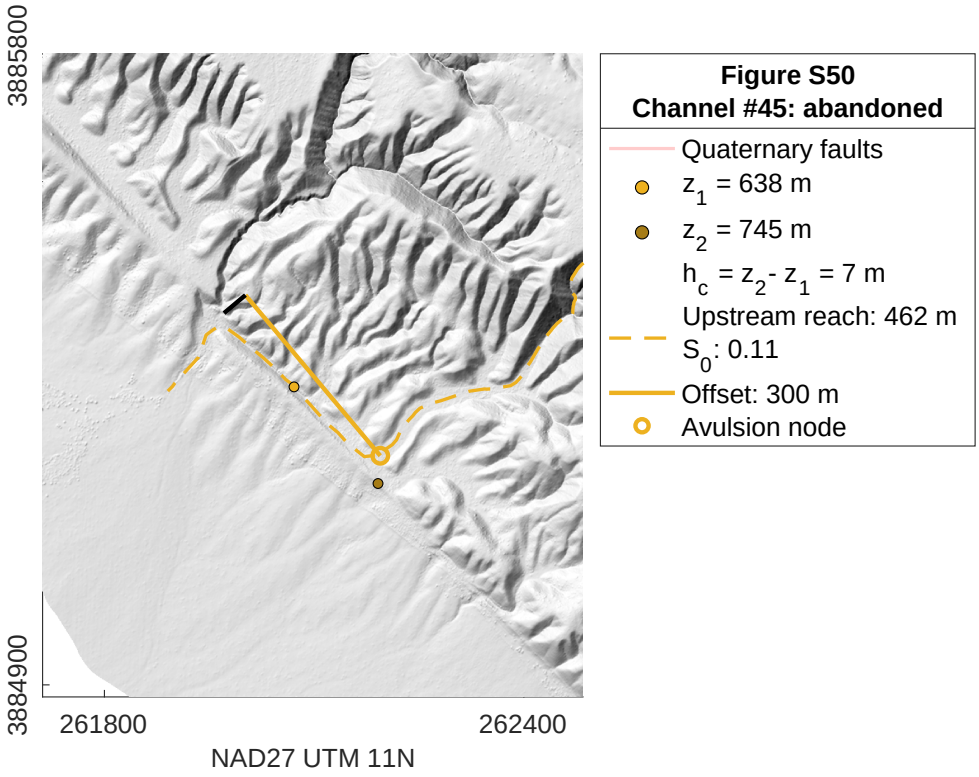

Figure A.50: Measurements for channel 45

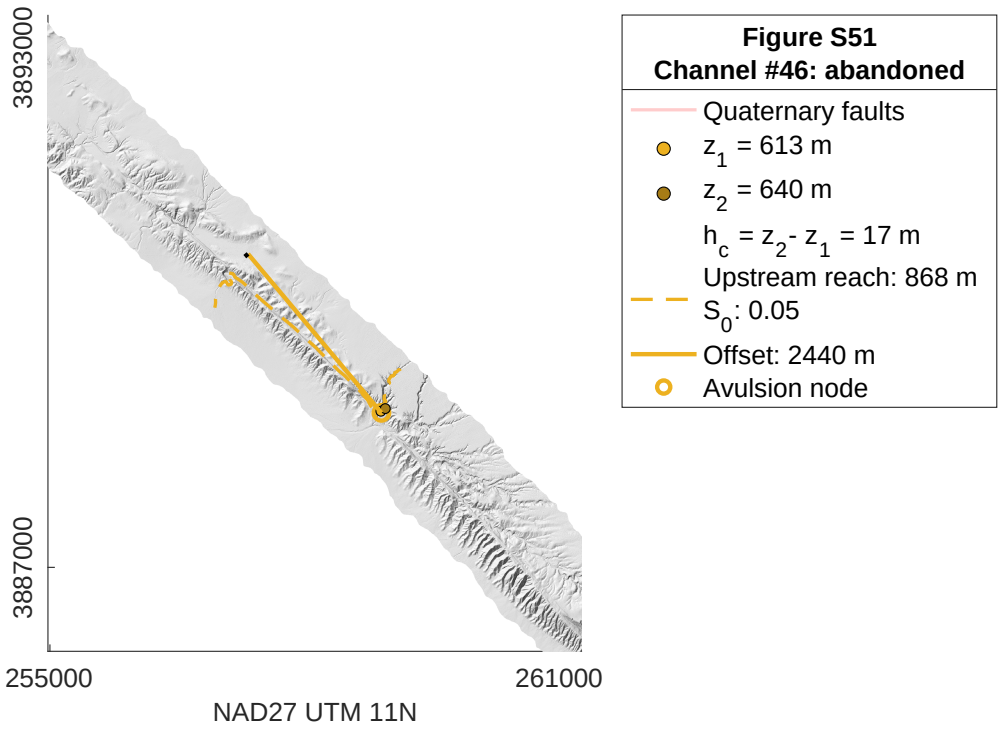

Figure A.51: Measurements for channel 46

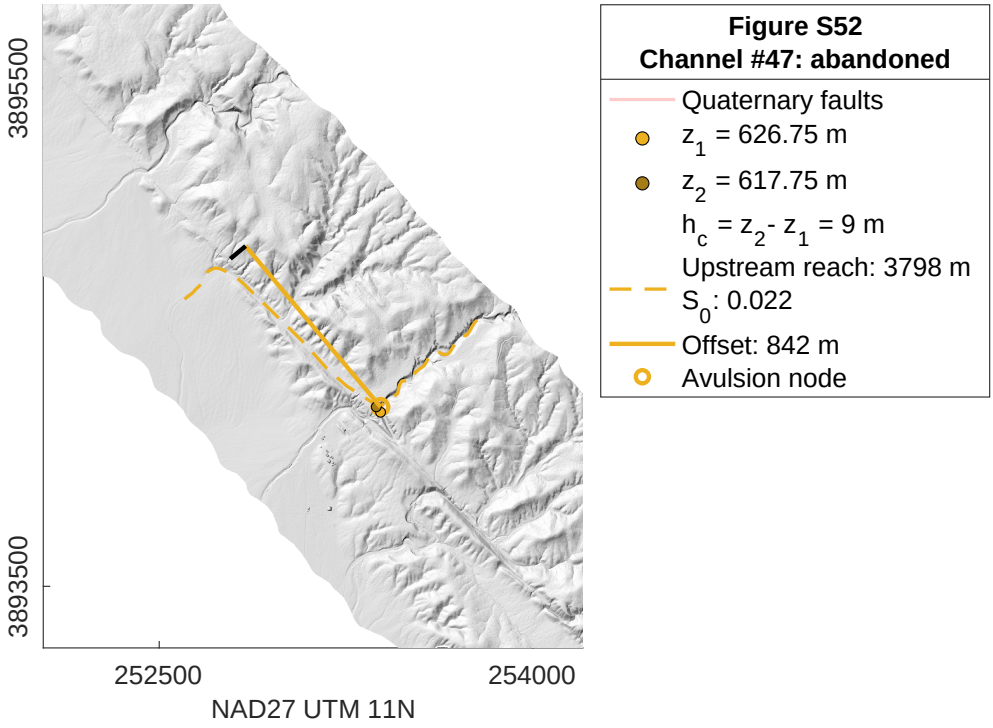

Figure A.52: Measurements for channel 47

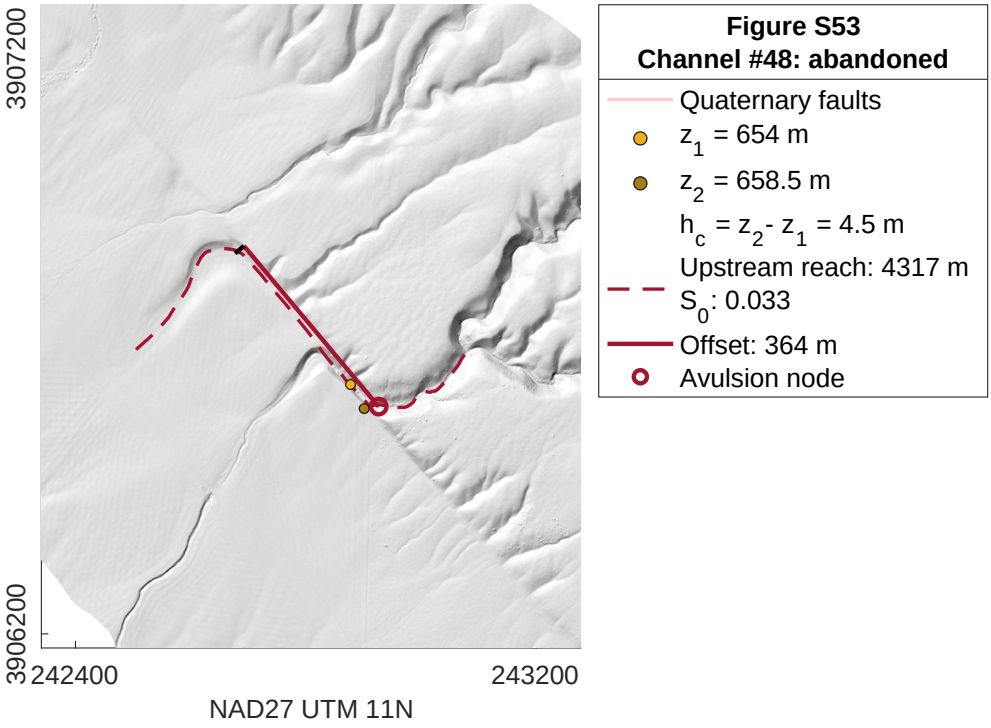

Figure A.53: Measurements for channel 48

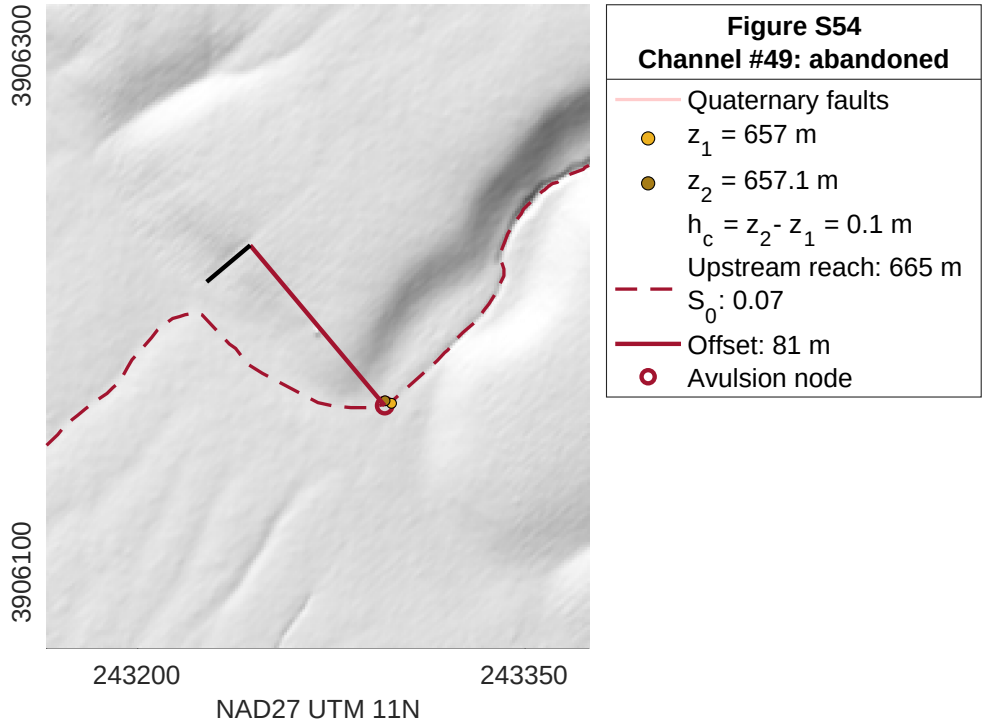

Figure A.54: Measurements for channel 49

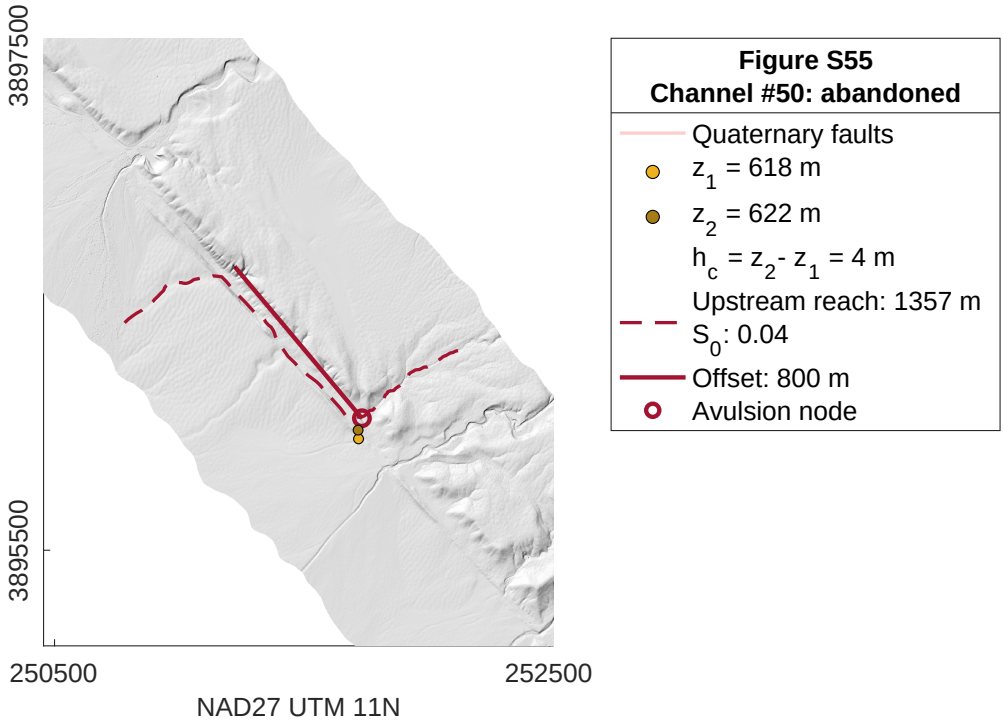

Figure A.55: Measurements for channel 50

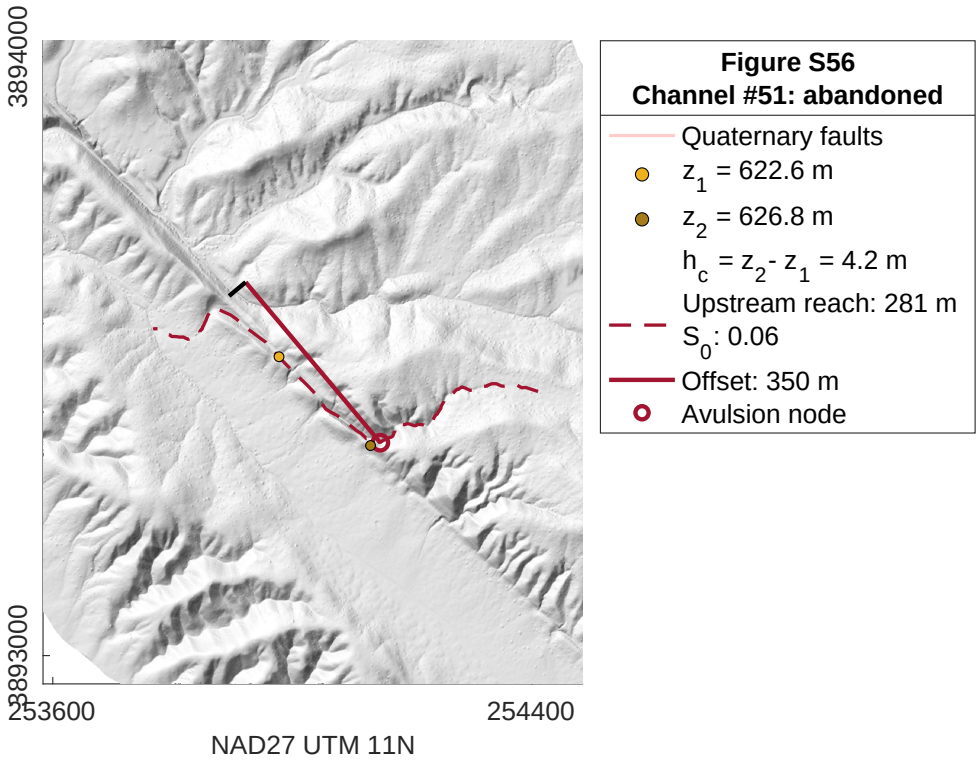

Figure A.56: Measurements for channel 51

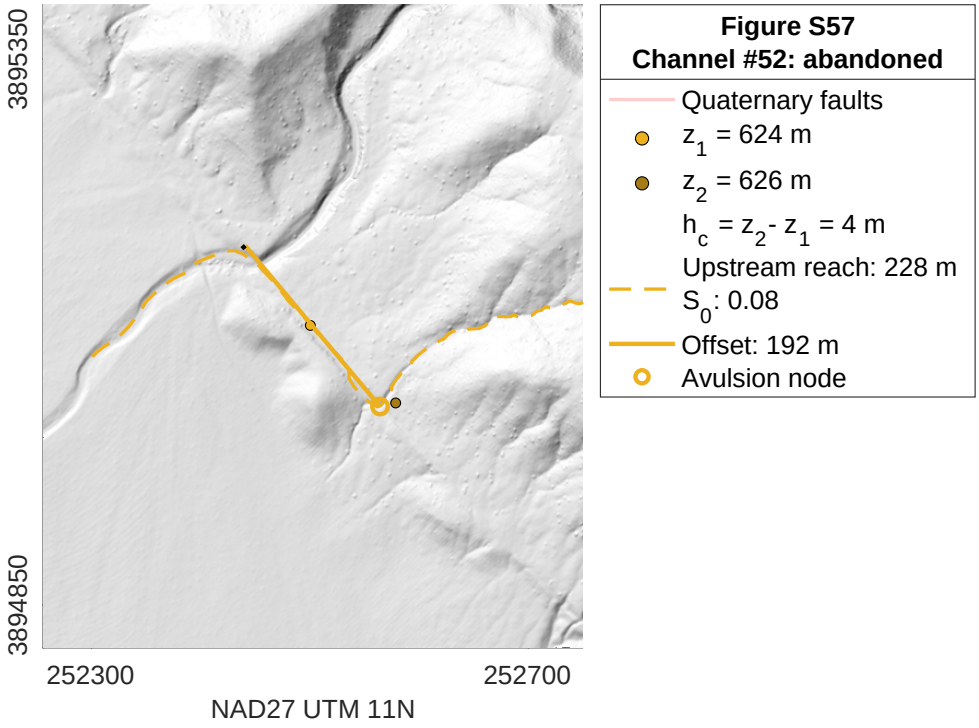

Figure A.57: Measurements for channel 52

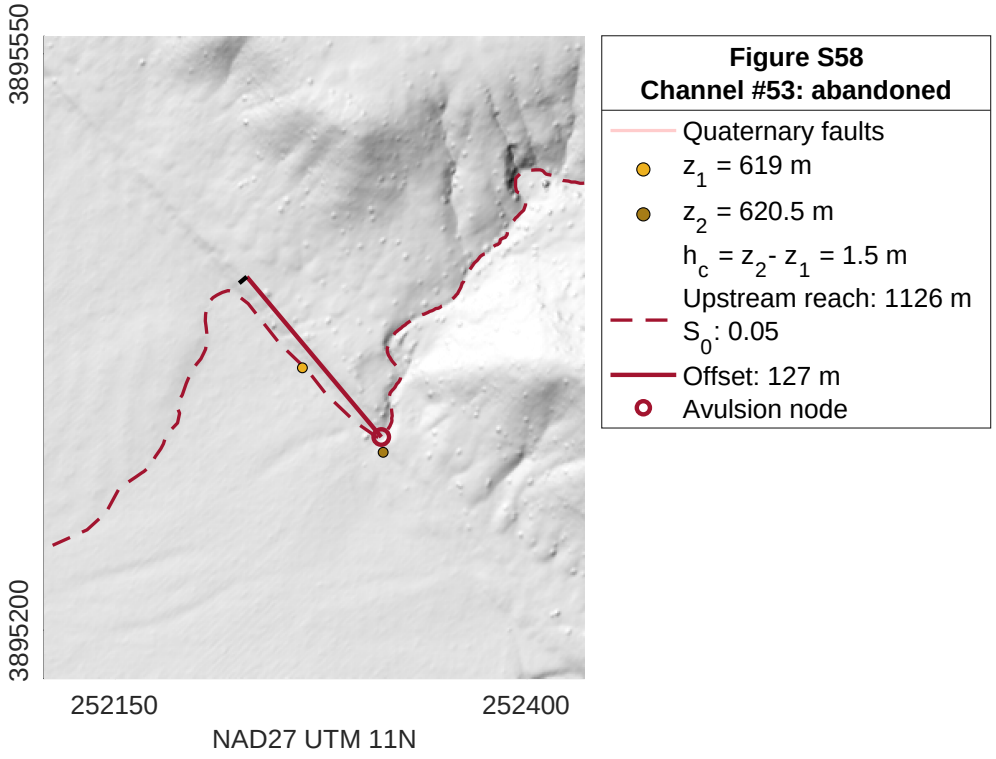

Figure A.58: Measurements for channel 53

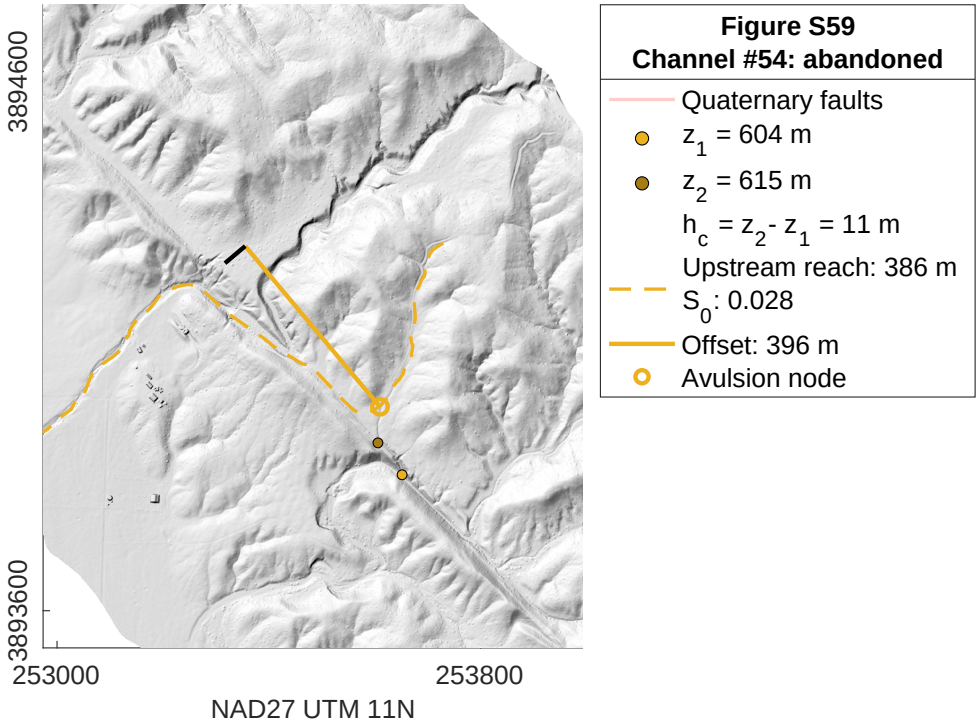

Figure A.59: Measurements for channel 54

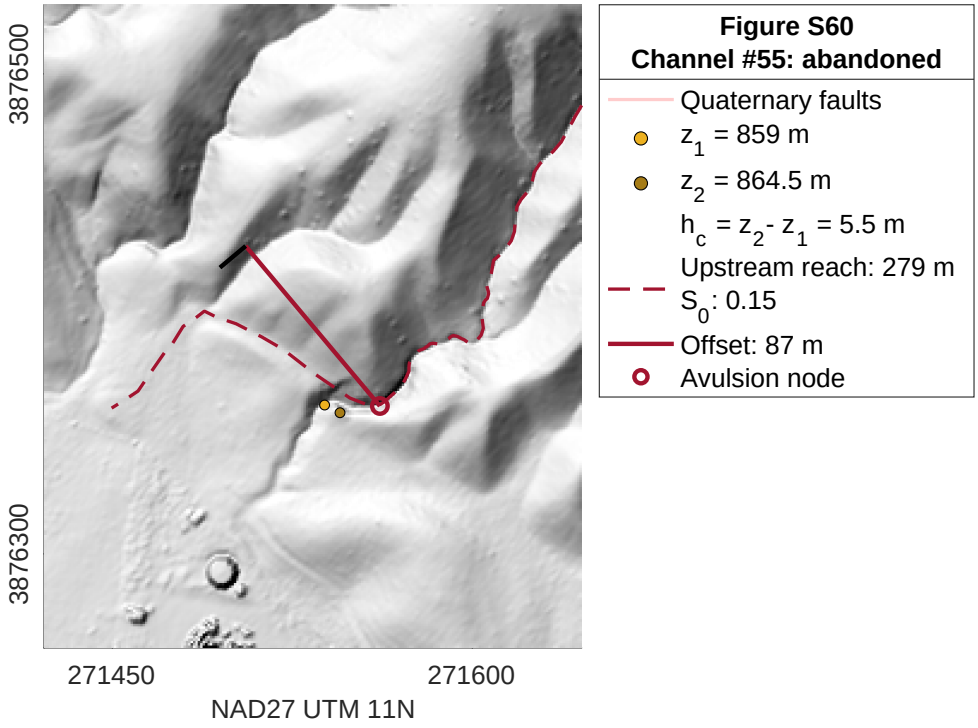

Figure A.60: Measurements for channel 55

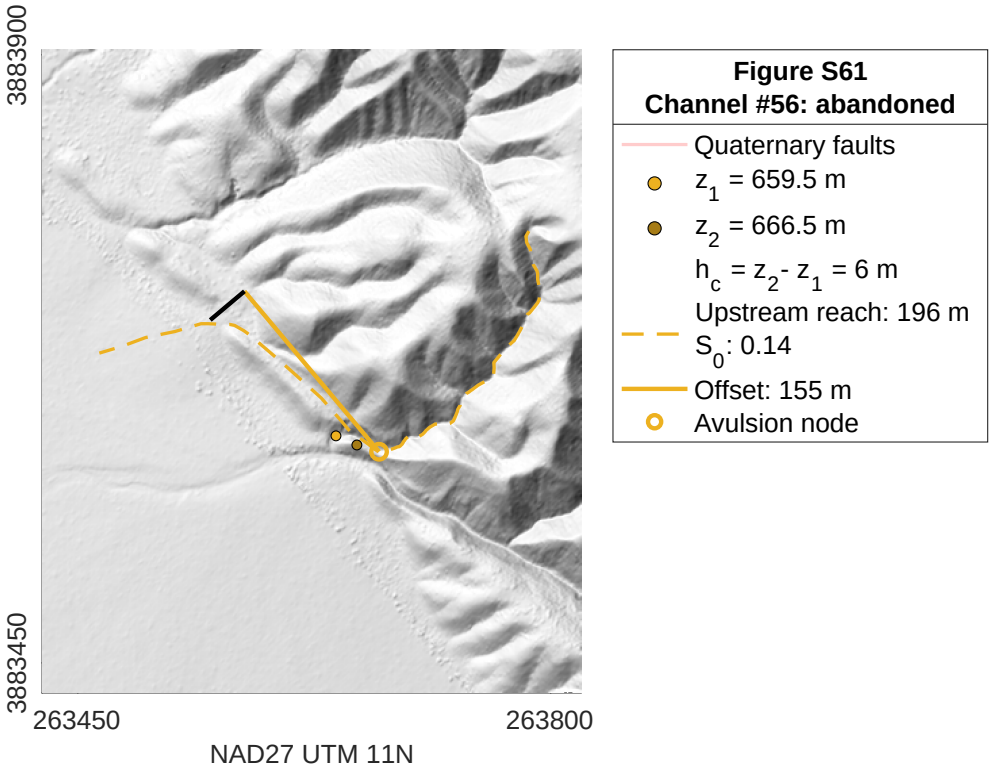

Figure A.61: Measurements for channel 56

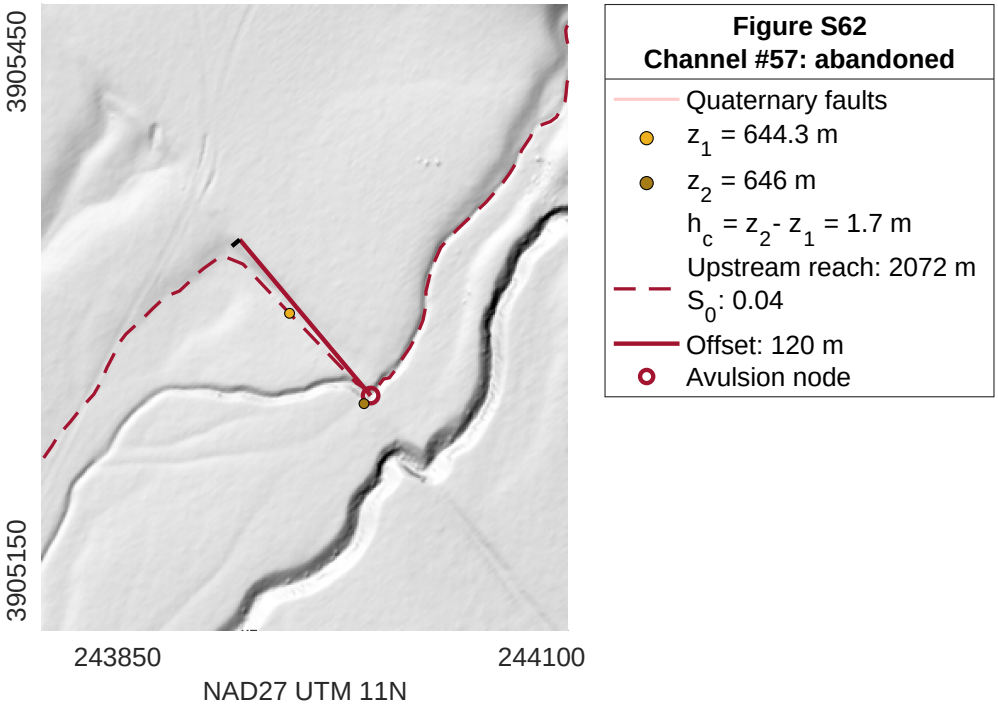

Figure A.62: Measurements for channel 57

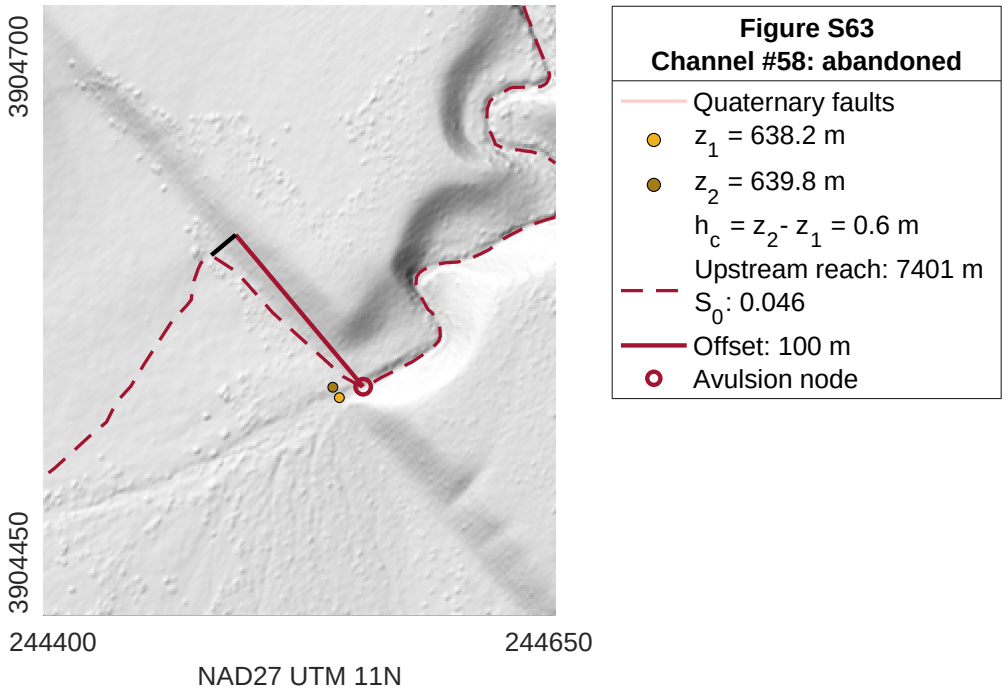

Figure A.63: Measurements for channel 58

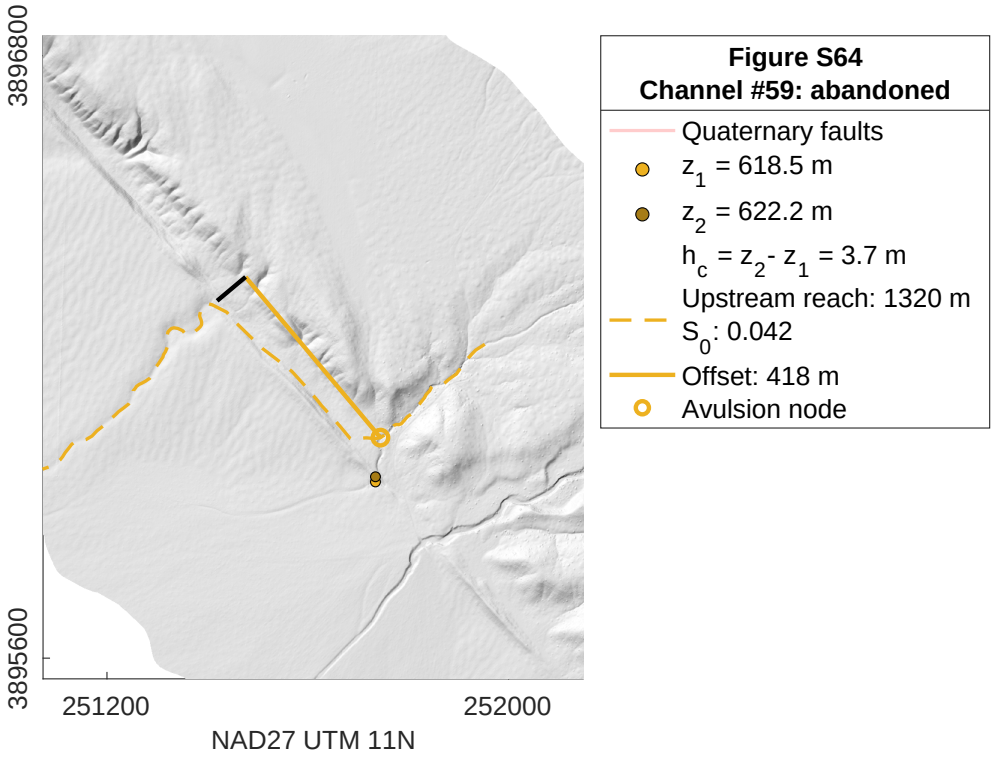

Figure A.64: Measurements for channel 59

Table A.1: Channel geometry measurements. See text for the measurement protocol. Easting and Northing coordinates correspond to the avulsion node.

ID	Easting (UTM)	Northing (UTM)	Offset (m)	Width (m)	Slope (unitless)	hc (m)	Reach (m)	Stage	Active (True/False)
1	264035.236	3883386.63	450	80	0.03	35	947	green	1
2	266212.496	3881100.63	170	30	0.11	10	461	green	1
3	268181.69	3879306.41	270	50	0.07	7	523	yellow	1
4	268662.946	3878871.36	6	16	0.12	4	323	green	1
5	269045.714	3878680.48	500	120	0.08	11	222	yellow	1
6	269091.9	3878571.07	170	35	0.09	9	212	green	1
7	269131.441	3878501.39	180	30	0.065	7	247	green	1
8	269276.698	3878388.4	20	15	0.09	4	271	green	1
9	269506.465	3878169.85	35	8	0.07	2	372	green	1
10	268590.168	3878926.64	110	20	0.1	5	362	green	1
11	268469.139	3879022.56	7	10	0.09	1.5	392	green	1
12	261968.713	3885441.08	25	9	0.05	1.2	500	green	1
13	263091.981	3884287.23	265	35	0.03	20	5675	green	1
14	242923.374	3906597.34	128	24	0.033	4.5	4400	green	1
16	243998.384	3905253.62	18	4	0.036	2	2039	green	1
17	244033.281	3905209.74	17	5	0.035	5.5	3138	green	1
18	244551.126	3904536.73	1	3	0.035	0.7	5500	yellow	1
20	246312.281	3902390.09	21	15	0.022	2.5	1710	green	1
21	251753.186	3895940.86	27	10	0.026	1.3	1314	green	1
22	251873.483	3895838.38	19	28	0.028	4	4681	green	1
23	252565.163	3895032.3	12	7	0.07	1.4	227	green	1
24	252452.09	3895162.91	37	36	0.024	2.1	4698	green	1
25	253435.839	3894089.27	248	50	0.028	14.4	2845	green	1
26	253653.177	3893855.57	52	13	0.055	15.6	415	green	1

27	253779.115	3893739.37	220	24	0.024	20	1285	green	1
28	253833.471	3893680.19	309	25	0.025	19	1839	green	1
29	254146.844	3893354.79	110	39	0.07	8	285	green	1
30	254261.828	3893220.14	16	3	0.09	2	214	green	1
31	256671.514	3890936	467	26	0.027	17	5429	green	1
32	257787.387	3889888.8	919	28	0.021	25	5694	green	1
33	258880.311	3888797.03	26	6	0.04	5	1350	green	1
34	259084.857	3888570.92	44	12	0.028	4.5	5637	green	1
37	261948.287	3885485.16	600	24	0.11	10	519	yellow	0
38	267217.817	3880050.52	300	18	0.05	7	902	yellow	0
39	268218.617	3879311.57	580	80	0.075	14	517	red	0
40	268690.891	3878880.99	210	70	0.09	4.75	320	yellow	0
41	269062.675	3878568.7	320	40	0.08	5.5	250	red	0
42	268268.769	3879217.54	60	24	0.14	2	170	yellow	0
43	268309.064	3879180.35	107	12	0.14	3.5	220	red	0
44	263340.806	3884050.49	630	60	0.08	6.7	443	yellow	0
45	262194.584	3885227.68	300	40	0.11	7	462	yellow	0
46	258898.384	3888816.83	2440	50	0.05	17	868	yellow	0
47	253387.976	3894220.3	842	80	0.022	9	3798	yellow	0
48	242926.523	3906594.79	364	20	0.033	4.5	4317	red	0
49	243295.633	3906164.4	81	22	0.07	0.1	665	red	0
50	251741.627	3896033.04	800	2	0.04	4	1357	red	0
51	254146.958	3893355.42	350	36	0.06	4.2	281	red	0
52	252564.239	3895031.83	192	4	0.08	4	228	yellow	0
53	252311.425	3895317.29	127	6	0.05	1.5	1126	red	0
54	253610.081	3893985.28	396	50	0.028	11	386	yellow	0
55	271561.472	3876361.62	87	14	0.15	5.5	279	red	0
56	263672.945	3883606.21	155	33	0.14	6	196	yellow	0
57	243998.511	3905252.99	120	6	0.04	1.7	2072	red	0
58	244555.862	3904538.6	100	16	0.046	0.6	7401	red	0
59	251744.771	3896039.13	418	74	0.042	3.7	1320	yellow	0

A.6.1 Example measurements

In Fig. A.5, we feature annotated examples of measurements of channel geometry. We first digitized all channels by hand within the lidar swath.

Table A.1 features the following columns:

- ID [integer]: Identification number. ID numbers are in correspondence to those indicated in Fig.A6-64. Channels excluded from the analysis (see Measurement protocol, Fig. A20, A24, A40-41), are not included in the table.

- Avulsion node location: reported in NAD27 UTM Zone 11N easting and northing coordinates (meters, columns 1 and 2 of Table A.1) correspond to the upstream intersection of the active channel branch and the fault (yellow star in Fig. A.5). Uncertainty on this measurement is likely on the order of the channel width (not to be confused fault zone width).

- Offset [meters]: The fault parallel distance required to backslip channels to a straight fault crossing geometry determines the channel offset (see Fig. A.5). This distance is always measured relative to the avulsion node location.

- Width [meters]: The zone fault-perpendicular extent over which we observe significant deflection of the channel. We report this measurement but do not utilize it in our analysis.

- Slope [unitless]: The initial slope, S_0, is measured away from the fault to avoid confounding the channel response, but not so far as to confound the concavity of the channel profile, typically a third to two thirds of the upstream reach.

- hc [meters]: The critical avulsion threshold height is an inferred quantity. We defined this quantity to reflect the height of sediment relative to the original channel floor required for the channel to avulsed. However, sediment obscures the channel floor datum and complicates this measurement. In some channels (e.g., Fig. A.5a), we leverage the elevation of a neutral point which we interpret to neither aggrades or erodes midway along the deflection of an active channel (z_1). We then obtain h_c by comparing this original datum to the lowest point in the immediate proximity of the avulsion node that would enable overspilling of the channel (z_2). In other channels (e.g., Fig. A.5b), we can obtain $h_c$directly from the channel walls of a recent avulsion. Cases in which the channel is abandoned obtaining h_c can be less straightforward (e.g., Fig. A.5c). In such cases the original channel floor datum is obtained midway along the active strand of the channel or equivalently midway along the offset leading to the avulsion of the original now abandoned strand. These are then compared again to the height of topographic barriers as the present-day avulsion node (z_2). Note that we

find agreement between these different approaches (e.g., Fig. A.5c).

- Reach [meters]: Channel reach is measured as the crow flies from the avulsion node to the start of the channel.

- Stage [categorical]: The channel stage can take on the following categorical stages: active with no evidence of incipient avulsion (green), incipient or recent avulsion (yellow), abandoned with no evidence of recent avulsion (red). The determination of the incipient or recent avulsions is documented in Fig. A.5. Panel b) shows two examples of critical channels which feature incipient (to the left) and recent (to the right) avulsions. We categorized channels to feature recent avulsion if offset on the active (avulsed) channel is much smaller than the total offset on the abandoned channel. We categorize channels to feature incipient avulsions if the elevation (original datum plus aggraded sediment) at the avulsion node approaches the height of the topography confining the channel. This differential height should not be confused with the critical avulsion threshold height that is measured from a neutral datum along the offset (i.e. $h_c = z_2 - z_1$, as shown in each panel).

- Active: [categorical] We document whether the channel we digitized is active or abandoned based one whether or not there is a contiguous flow path connecting a tributary to the outlet. Abandoned channel outlets typically feature a more muted morphology from colluvial fill post-dating the loss of

catchment area upstream of the avulsion node.

Appendix B

Supplementary Material for Chapter 2

B.1 Examination of relative productivity

In this section we examine potential biases to relative productivity that may arise from our counting method, its parameterization and catalog artifacts.

B.1.1 Scaling effects

We first assess whether any biases are introduced by the definition of relative productivity as the number of aftershocks referenced to the number of aftershocks expected for a given magnitude. If the global trend with magnitude is incorrectly

modelled, a systematic trend of relative productivity with magnitude would be observed. Fig. B.1 shows that no such trend exists and thus definition of relative productivity appears reasonable.

Next, we examined whether there may be issues related to the definition of the scaling relationship we used in space windowing. Fig. B.2 compares the measurement of relative productivity utilizing two different inferred scalings between rupture length and magnitude (Brengman et al. 2019; Wells and Coppersmith 1994). We find that 79% of sequences have the same aftershock counts regardless of the scaling used and that the excursions tend to be small.

Using a the group of earthquakes with $M_W < 7.5$ reveals more statistically significant subgroups for attributes which do not require finite fault inversion (Fig. B.4). Conversely, those attributes derived from finite fault inversion become less stable and, as a result, their relative productivity statistics are more varied (Fig. B.3). The attributes defining the largest and smallest relative productivity are the same in both cases. Note that within for the set of smaller mainshocks, the median value of relative productivity can be censored, implying that the median mainshock of the subgroup had no measured aftershock.

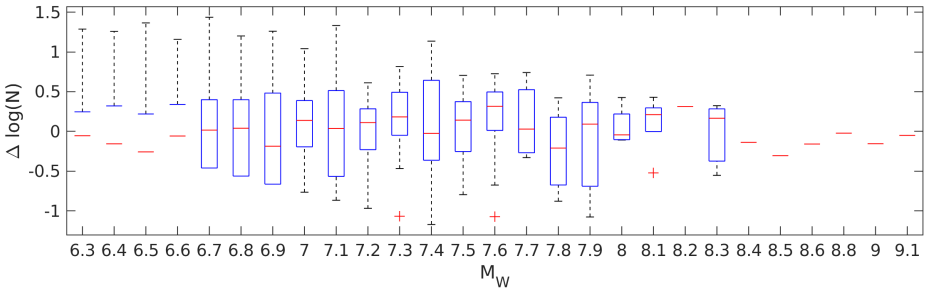

Figure B.1: Box and whiskers plot of relative productivity as a function of magnitude. Boxes outline the interquartile range. Whiskers outline the range of the data and +'s indicate outliers. Censoring, $\Delta \log(N) = -inf$, at the lower magnitudes ($M_W < 6.7$) and limited data at highest magnitudes ($M_W > 8.1$) prevent proper assessments of the interquartile range. Nonetheless, we do not observe a systematic decrease in variance that may introduce bias in subsequent analysis.

B.1.2 Catalog completeness

We test the robustness of our primary findings by testing their validity with a more conservative magnitude of completeness of $M_W 5$ instead of $M_W 4.5$ as done in the main text. In these results, there are fewer aftershocks detected and statistical significance is therefore decreased while the chance of inclusion of any background events in the counts is lowered because of the reduced rate for the larger magnitude cut-off, but the results do not violate any of the major findings of the main text. We reproduce Fig. 2.1-2.11 of the main text with the alternative completeness in Fig. B.5–B.15, which preserve the same order as the main text.

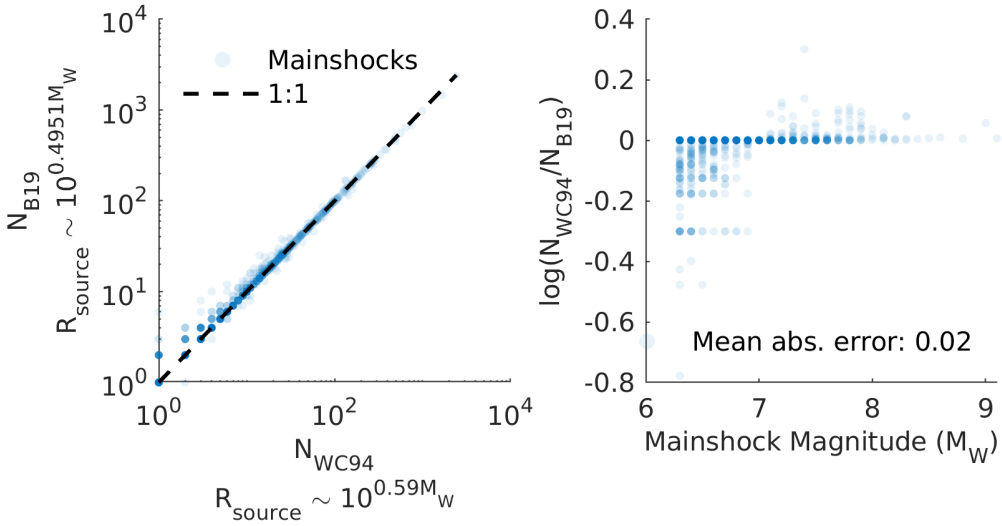

Figure B.2: Comparison between relative productivity statistic as obtained from two different scaling relationships: 1) WC94 following (Wells and Coppersmith 1994) and 2) B19 following (Brengman et al. 2019). Left: Direct comparison of inferred number of aftershocks. Right: Discrepancy in measurements as a function of mainshock magnitude. We find that 79% of sequences have that exact same aftershock counts. Larger ruptures yield slightly smaller a aftershock counts the B19 scaling, whereas smaller ruptures have the converse relationship. Generally, the effect is subtle with a mean absolute error of 0.02. Note that the discrepancies, $log(N_{WC}/N_{B19})$, are directly equal to differences in measurements of relative productivity.

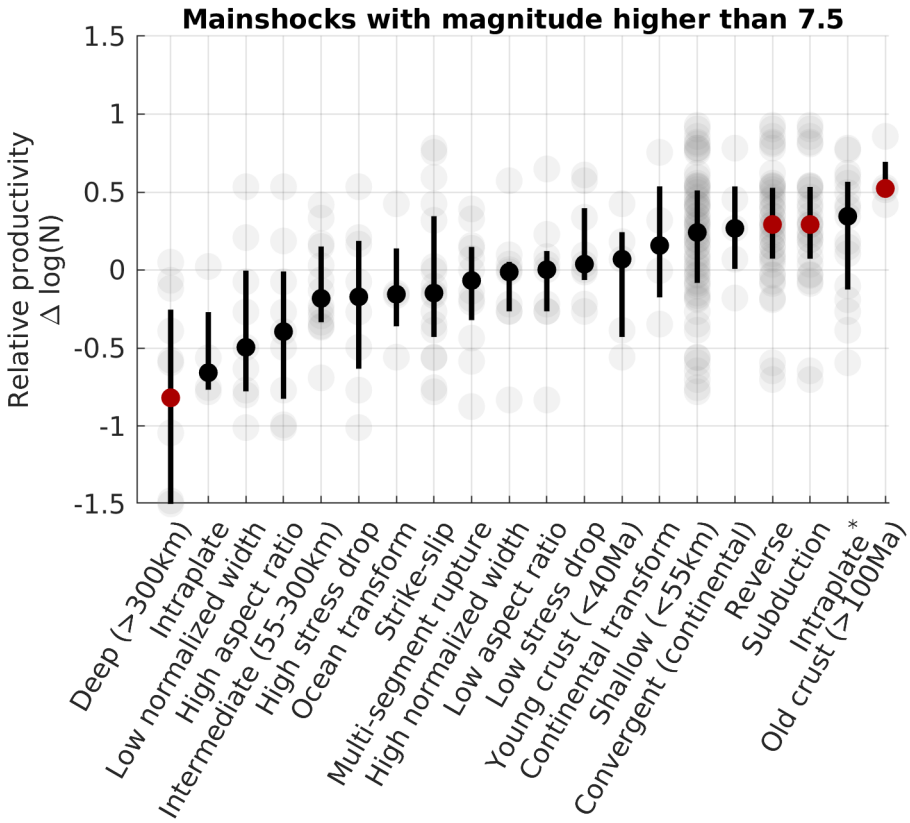

Figure B.3: Major results of this study presented for mainshocks with M_W magnitude greater than 7.5. Attributes with red markers are more consistent with the hypothesis that they are sampled from a different continuous distribution than the overall population of earthquakes using a 2-sampled Kolmogorov-Smirnov test at a 5% significance threshold.

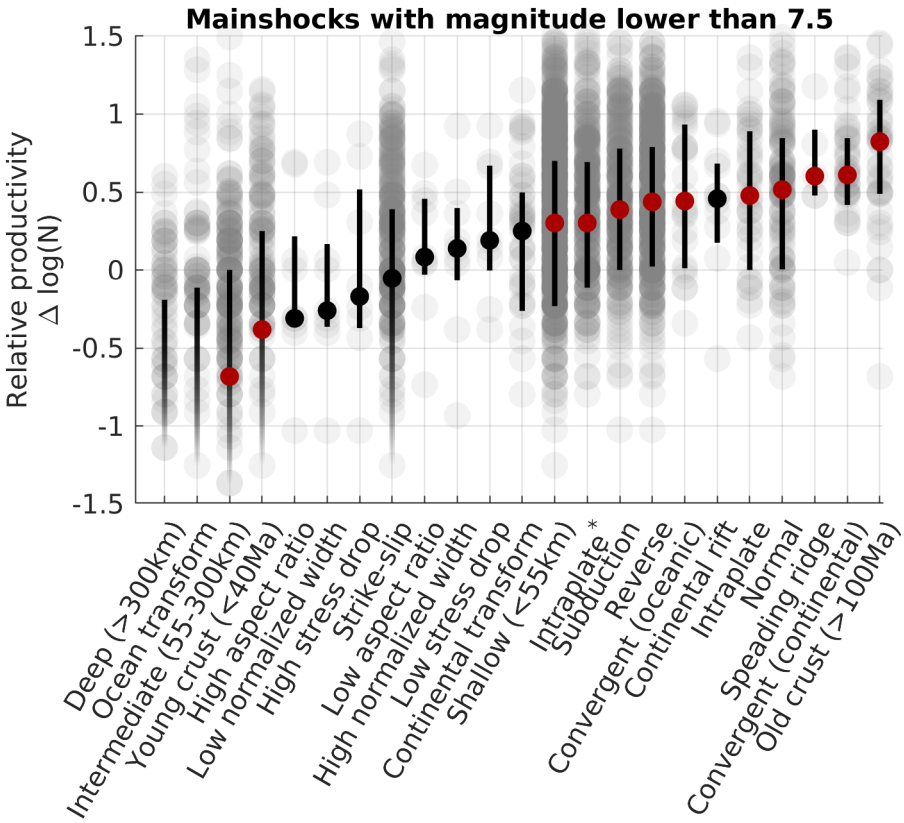

Figure B.4: Major results of this study presented for mainshocks with M_W magnitude less than 7.5. Attributes with red markers are consistent with the hypothesis that they are sampled from a different continuous distribution than the overall population of earthquakes using a 2-sample Kolmogorov-Smirnov test at a 5% threshold.

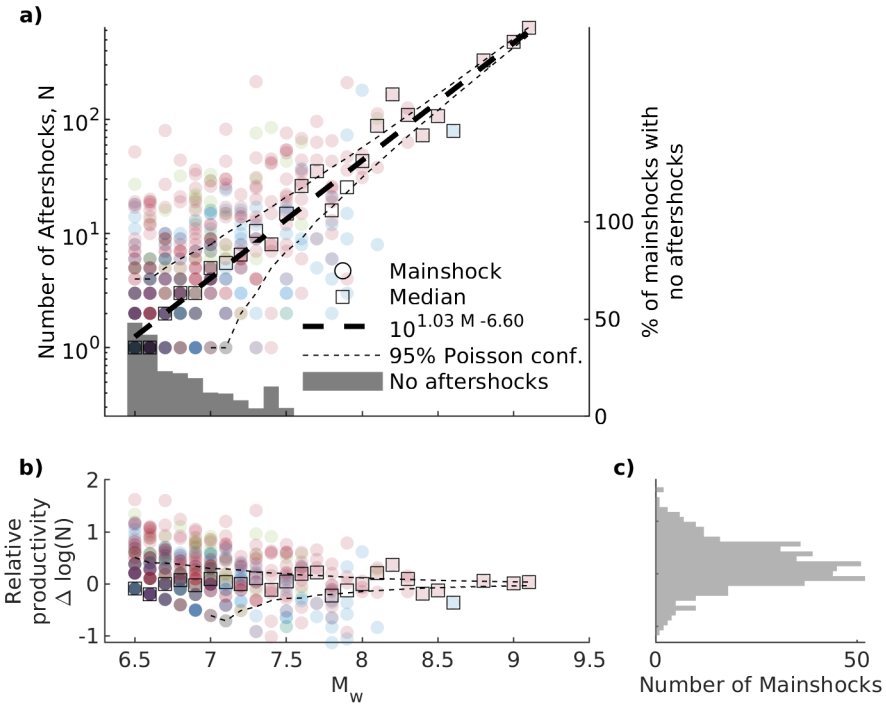

Figure B.5: The number of aftershocks of $M_W \geq 5$ within three source dimensions and 60 days as a function of mainshock magnitude identified in the global ISC and NEIC catalogs from 1990 to 2019. Colors indicate faulting style of the mainshock; blue, green and red points correspond to earthquake sequences for which the mainshock was respectively strike-slip, normal or reverse. The global productivity law (dashed line) is fit using a least squares regression through the median log-number of aftershocks for each 0.1 magnitude bin (black squares). The median number includes mainshocks with no aftershocks which are not shown on the plot. Note the individual earthquake sequences (circles) exhibit significant scatter around the productivity law. b) Relative productivity as a function of mainshock magnitude. The relative productivity distribution does not show events with no aftershocks and thus the lower left corner of the plot is underpopulated. c) Histogram of the relative productivity of mainshocks considered in this study.

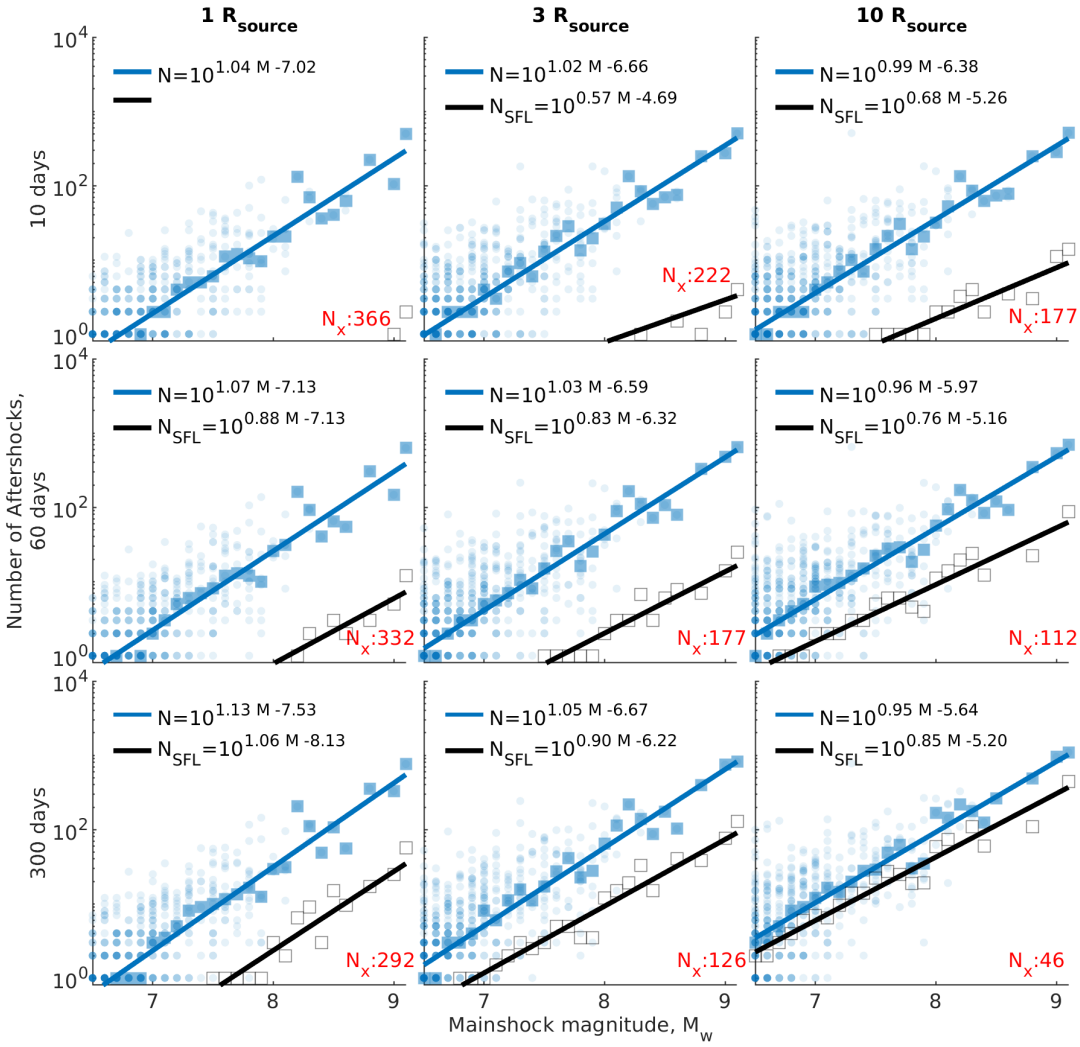

Figure B.6: Sensitivity analysis of space-time windows. Time windows of 10, 60, and 100 days and spherical space with radii of 1, 3, and 10 source dimensions (R_{source}) are considered. Blue data are mainshocks identified through our hierarchical declustering routine. Circles are individual mainshocks. Squares are median values for each 0.1 magnitude bins. Regressions are computed using least squares through the median log-number of aftershocks for each 0.1 magnitude bin. For reference, we computed the median productivity relationship (grey squares) for 100 time-shuffled catalogs and the corresponding scaling relationship (black line). For each space-time window, we indicate the number of mainshocks with no aftershocks in red (N_x). Note that as space and time windows increase, more mainshocks have measurable aftershock counts. However, the likelihood of counting background productivity and significantly affecting subsequent parameterization becomes an increasing concern.

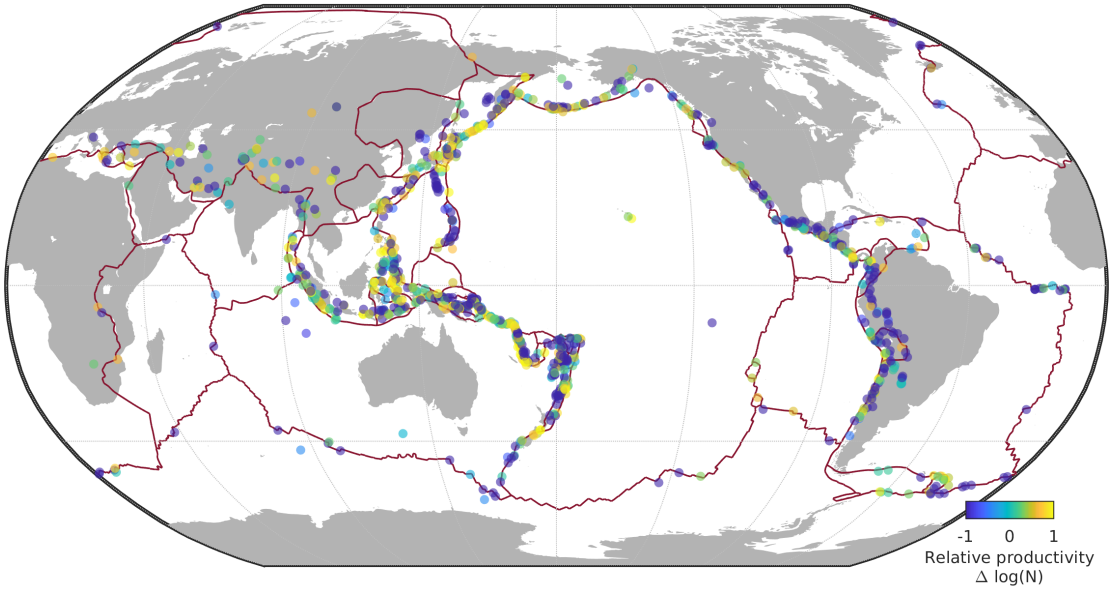

Figure B.7: Global map of earthquake productivity as measured with a catalog completeness threshold of M_W5.0. Red lines indicate the surface trace of the tectonic boundaries. Mainshocks with $M_W \geq 6.5$ color-coded according to their relative productivity.

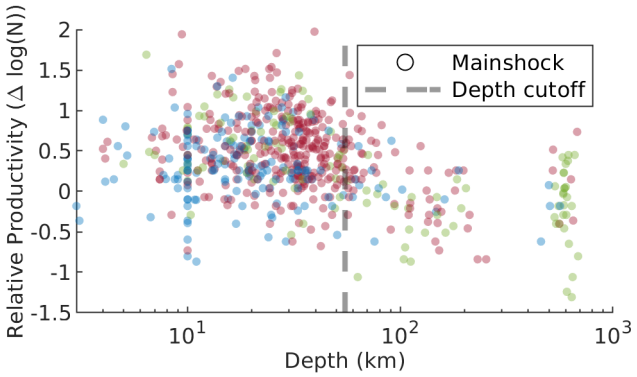

Figure B.8: Relative aftershock productivity measured with a catalog completeness threshold of M_W5.0 as a function of depth. Subsequent analysis will only consider earthquakes shallower than the 55 km cutoff (dashed line). Sequences are color-coded according to faulting style of the mainshock (blue: strike-slip, green: normal and red: reverse).

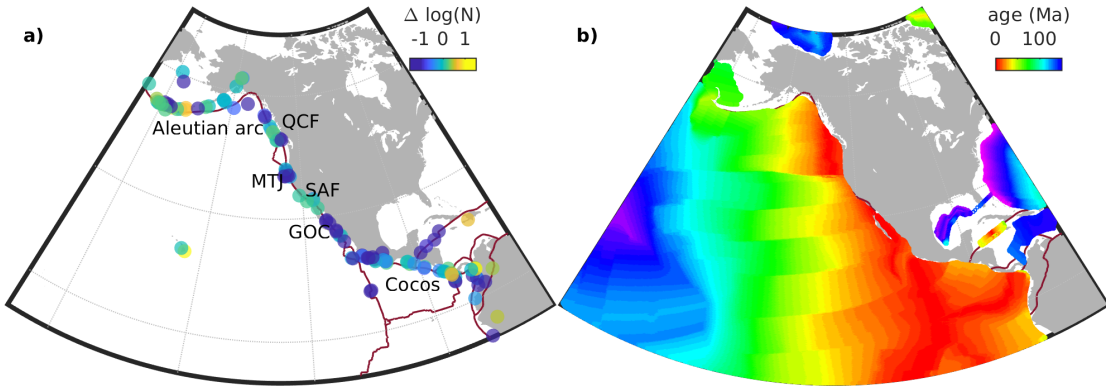

Figure B.9: a) Aftershock productivity measured with a catalog completeness threshold of $M_W 5.0$ along the North American coastline. Individual mainshocks (points) are color-coded according to their relative aftershock productivity ($\Delta \log(N)$). The Aleutian arc, the Queen Charlotte Fault (QCF), the Mendocino Triple Junction (MTJ), San Andreas Fault (SAF), the Gulf of California (GOC) and the Cocos plate subduction include areas with coherent productivity. Red line indicates major plate boundaries (Bird 2003). b) Seafloor crustal age estimates from Müller et al. 2008.

B.1.3 Alternative clustering method

In this section we present the salient results of our study using an alternate clustering approach. The main manuscript presents results obtained using a hierarchical space-time windowing scheme. We preferred this method because it is more readily reproducible on an event by event basis and therefore provides transparency. A reader is readily capable of inspecting individual past or future earthquakes to see where they fall with respect to the global average, and whether they corroborate or challenge our conclusions without having to perform extensive additional calculations. However, it is important to also check the result

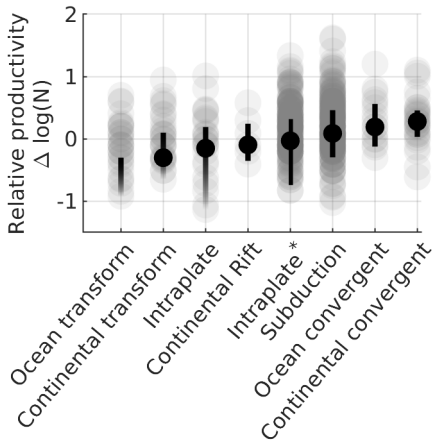

Figure B.10: Earthquake productivity measured with a catalog completeness threshold of M_W5.0 by tectonic boundary. Points indicate the relative productivity of individual sequences. Solid markers and error bars indicate the median and the interquartile range. A faded lower error bar implies that mainshocks with no aftershocks are within the interquartile range. Intaplate* indicates earthquakes within 400km from a plate boundary but with a faulting mechanism discordant with the plate boundary (e.g., outer rise events).

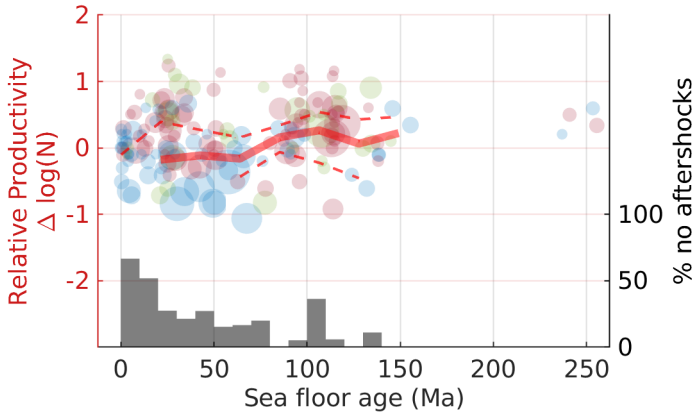

Figure B.11: Relative productivity measured with a catalog completeness threshold of M_W5.0 increases as a function of the age of the oceanic lithosphere. Each point indicates an individual earthquake sequence. Sequences are color-coded by faulting style of the mainshock (blue: strike-slip, green: normal and red: reverse). The red line indicates the median average for 20Ma crustal age bins. Dashed lines indicate the corresponding interquartile ranges. Bars indicate the fraction of earthquakes with no aftershocks within each 10 Ma crustal age bin.

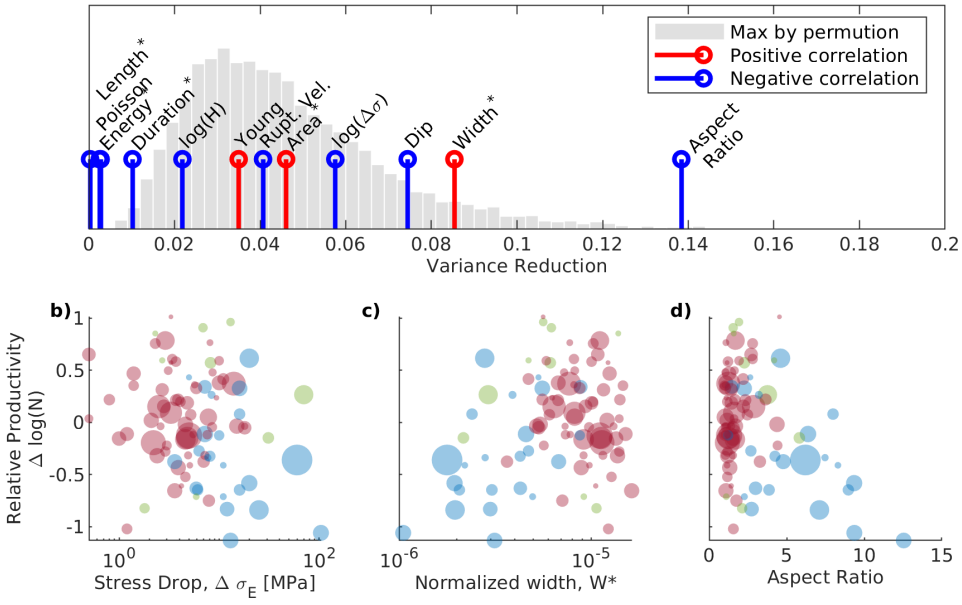

Figure B.12: Inspection of the relationship between source attributes and relative productivity measure with a catalog completeness of M_W5. a) Goodness of fit of linear regressions for each source attribute in our combined catalog. Top and bottom axes respectively represent the p-value and goodness of fit of each attribute (stems). The probability distribution function in the backdrop indicates the maximal variance reduction outcome of 10000 permutation test of the entire data set. The probability of obtaining a spurious correlation by chance over the whole family of attributes we tested for is derived from the number of random shuffles exceeding the measured variance reduction and normalized to the overall sample (10000). Asterisks indicate scaled and log-transformed variables. The scaled energy, length, duration and area, material properties, velocity, dip, and stress drop ($\Delta\sigma$) of the mainshock rupture all do not yield a statistically significant ($p = 0.05$) linear fit to the relative productivity; the normalized rupture width and aspect ratio of the rupture yield the best fitting linear regressions. Stems are color-coded to indicate whether the source attribute is positively (red) or negatively (blue) correlated with relative productivity. b-d): Relative earthquake productivity as a function of mainshock stress drop, normalized rupture width, and aspect ratio. Individual mainshocks are color-coded according to faulting style as in Fig. 2.1.

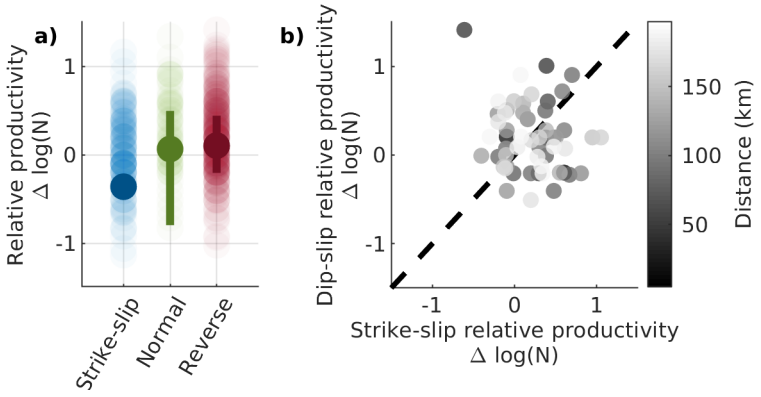

Figure B.13: Inspection of the relationship between focal mechanism and relative productivity as measured with a catalog completeness threshold of M_W5. a) Relative aftershock productivity ($\Delta \log(N)$) by focal mechanism. b) Relative aftershock productivity for pairs of earthquake sequences with strike-slip and dip-slip mainshocks within 200 km from each other. Each pair is shaded according to its relative distance. Dashed line indicates a 1:1 relationship, the expectation for a purely site dominated control on relative productivity. Co-located mainshocks pairs generally follow this 1:1 trend, but exhibit considerable scatter.

robustness with more mathematically rigorous aftershock detection methods.

Here we present our major results using aftershock counts obtained following Zaliapin et al. 2008. This approach seeks to build earthquake families by linking earthquakes to parent events. Separation of clusters is achieved by finding a decision boundary between background events and clustered events. See Zaliapin et al. 2008 for a detailed overview of the method, distance metrics, and theoretical connections to other schemes (ETAS). Though there are discrepancies, aftershock counts obtained using this method correspond well to counts obtained using space-time windowing (see Fig. B.16). Additionally, we find few differences to the major results of the study (see Fig. B.17–B.26). The workflow and detailed explanations

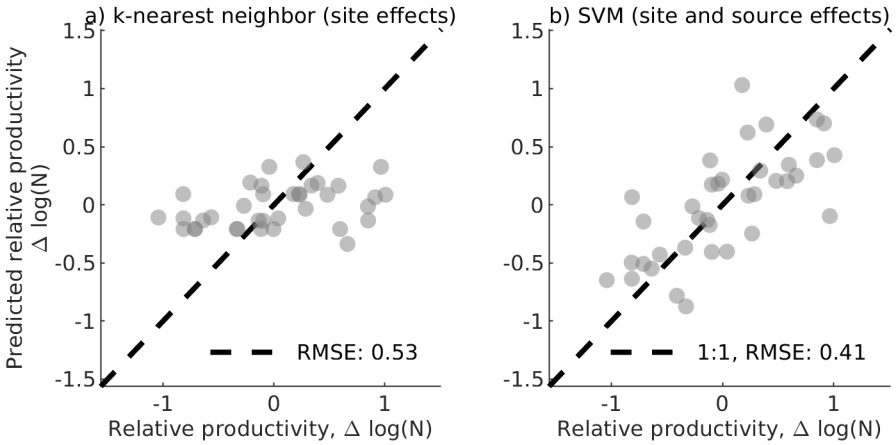

Figure B.14: Sensitivity of two forecasting approaches to catalog completeness. Measurements here of relative productivity are computed using a catalog completeness of $M_W 5.0$. Response plots (prediction versus observation) for the k-nearest neighbor algorithm and SVM models. Each point is an individual earthquake sequence. A perfect prediction would place all values on the 1:1 line. The SVM model outperforms the k-nearest neighbor model. Hold-one-out cross-validation ensures that the data for model calibration is separate from the prediction data. Combining both contextual information about the setting (crustal age) and the source (dip and normalized area) yields a root mean square value of 0.39. In particularly, the SVM model better predicts extreme cases (highly productive or unproductive).

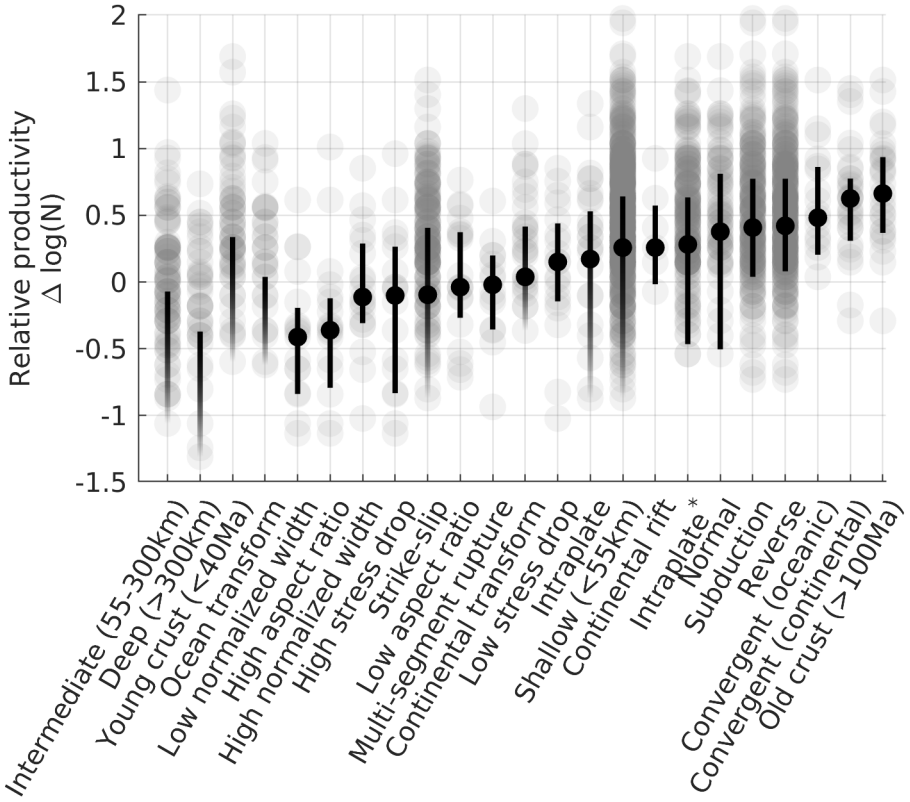

Figure B.15: Synthesis of relative productivity measured with a catalog completeness threshold of M_W5.0 according to catalog subsets. The group considered here are the short list which best distinguished relative productivity based on our different lines of investigation. 'High' and 'low' subsets respectively refer to >80th and <20th percentile ranges of the data. Grey circles are individual mainshocks. Black points and error bars respectively indicate the median and interaquartile range of the subset. Fading error bars imply that mainshock sequences with no aftershocks are within the interquartile range of the data.

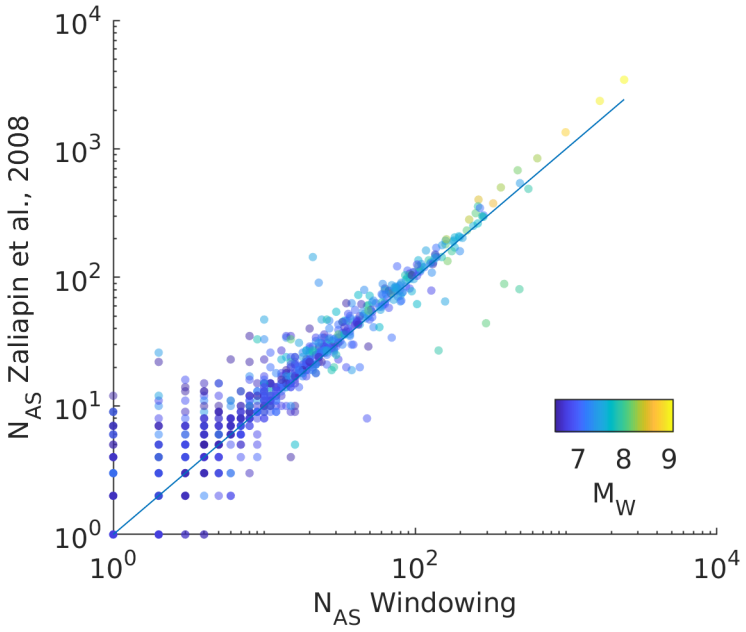

Figure B.16: Comparison of aftershock counts obtained using space time windowing and those obtained using nearest neighbor clustering (Zaliapin et al. 2008). Each point represents an individual mainshock. Mainshocks are colored according to moment magnitude. Blue line indicates a 1 to 1 correspondence.

of the clustering routine are available here:

https://github.com/tgoebel/clustering-analysis.

B.1.4 Background Seismicity

Since our selection criteria for aftershocks is magnitude dependent, there is a concern that aftershock counts will be biased by background seismicity. In this section, we show that the background rate of seismicity is in fact negligible and within counting error of a Poisson process.

Measuring background seismicity is challenging and an ongoing topic of re-

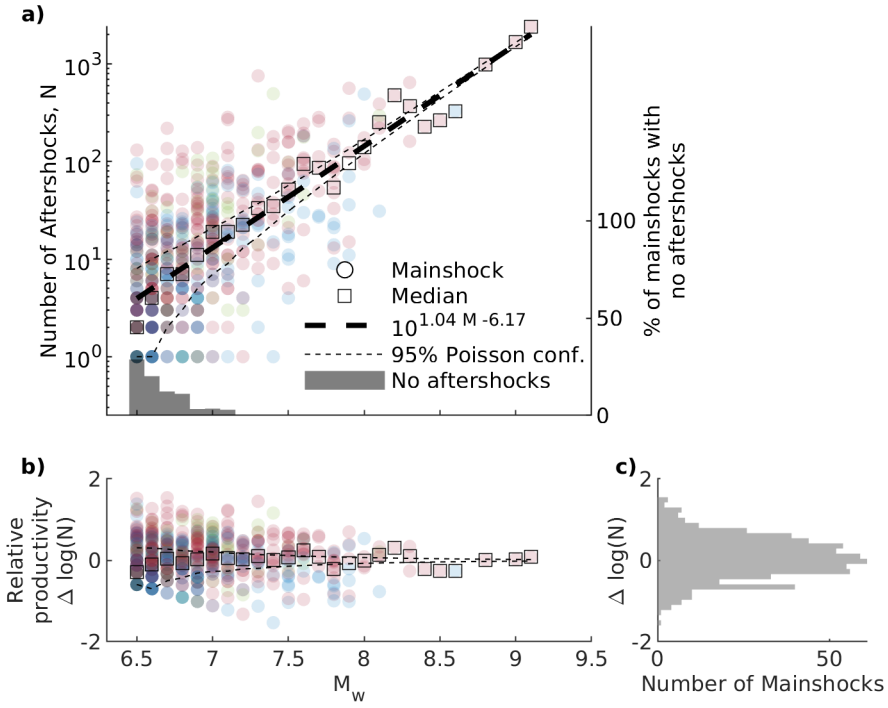

Figure B.17: Results using Zaliapin et al. 2008 clustering. a) The number of aftershocks of $M_W \geq 4.5$ within three source dimensions and 60 days as a function of mainshock magnitude identified in the global ISC and NEIC catalogs from 1990 to 2019. Colors indicate faulting style of the mainshock; blue, green and red points correspond to earthquake sequences for which the mainshock was respectively strike-slip, normal or reverse. The global productivity law (dashed line) is fit using a least squares regression through the median log-number of aftershocks for each 0.1 magnitude bin (black squares). The median number includes mainshocks with no aftershocks which are not shown on the plot. Individual earthquake sequences (circles) scatter significantly above and below the productivity law. b) Relative productivity as a function of mainshock magnitude. The relative productivity distribution does not show events with no aftershocks and thus the lower left corner of the plot is underpopulated. c) Histogram of the relative productivity of mainshocks considered in this study.

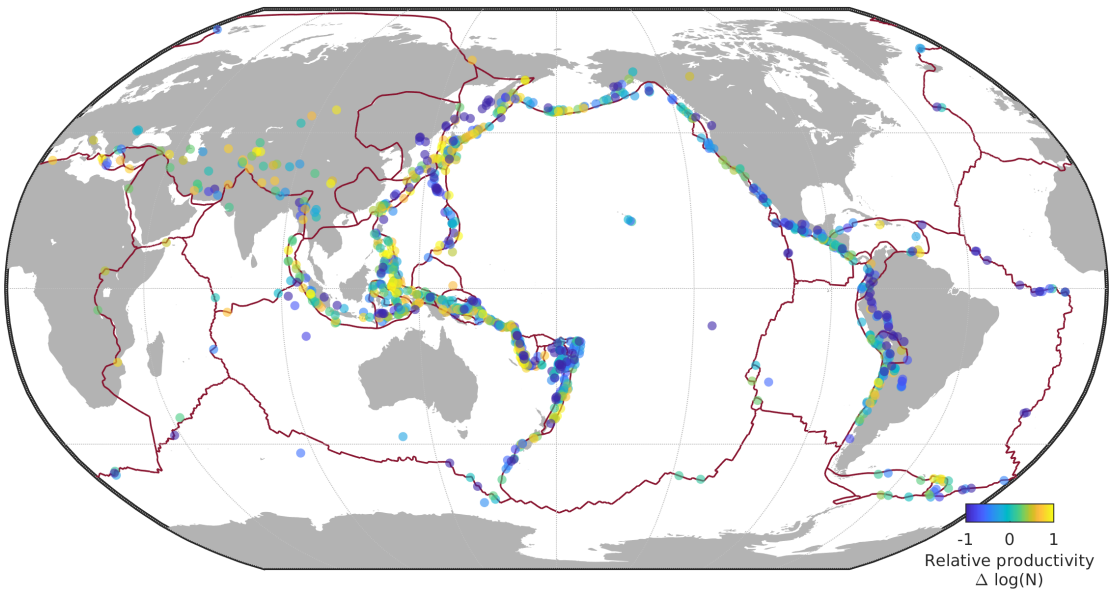

Figure B.18: Results using Zaliapin et al. 2008 clustering. Global map of earthquake productivity. Red lines indicate the surface trace of the tectonic boundaries. Mainshocks with $M_W \geq 6.5$ color-coded according to their relative productivity

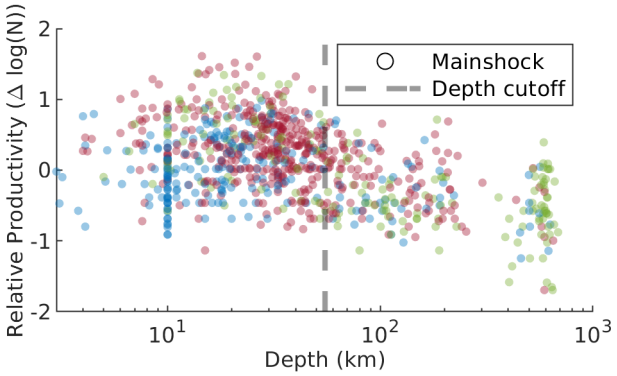

Figure B.19: Results using Zaliapin et al. 2008 clustering. Relative aftershock productivity as a function of depth. Subsequent analysis will only consider earthquakes shallower than the 55 km cutoff (dashed line). Sequences are color-coded according to faulting style of the mainshock (blue: strike-slip, green: normal and red: reverse). Note: Discretization of depth is apparent in this plot as some events have default values. Depths of 33 km, 5 km, 10 km and 15 km are reported for 6%, 1%, 10% and 0.7%, respectively, of the catalog.

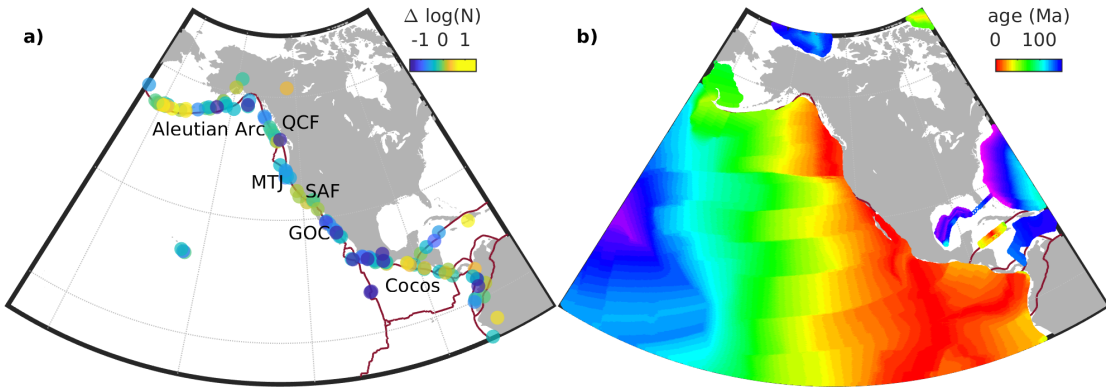

Figure B.20: Results using Zaliapin et al. 2008 clustering. a) Aftershock productivity along the North American coastline. Individual mainshocks (circles) are color-coded according to their relative aftershock productivity. The Aleutian Arc, Queen Charlotte Fault (QCF), Mendocino Triple Junction (MTJ), San Andreas Fault (SAF), Gulf of California (GOC) and Cocos Plate Subduction Zone include areas with coherent productivity. Red line indicates major plate boundaries (Bird 2003). b) Seafloor crustal age estimates from Müller et al. 2008.

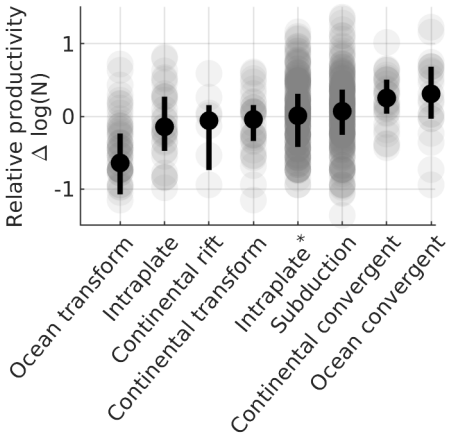

Figure B.21: Results using Zaliapin et al. 2008 clustering. Earthquake productivity by tectonic boundary. Circles indicate the relative productivity of individual sequences. Solid markers and error bars indicate the median and the interquartile range. A faded lower error bar implies that mainshocks with no aftershocks are within the interquartile range. Intraplate* indicates earthquakes within 400 km of a plate boundary but with a faulting mechanism discordant with the plate boundary (e.g., outer rise events).

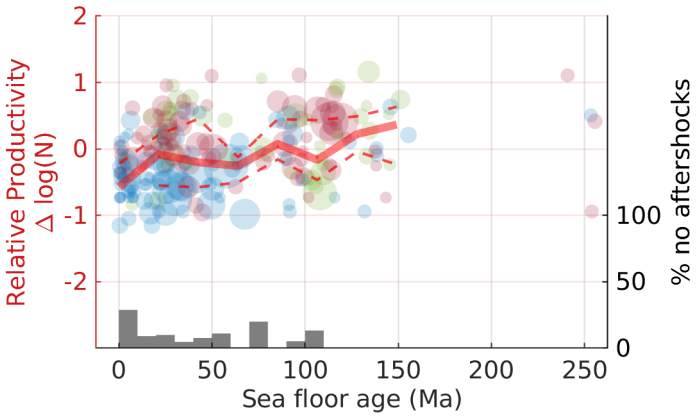

Figure B.22: Results using Zaliapin et al. 2008 clustering. Relative productivity increases as a function of the age of the oceanic lithosphere. Each circle indicates an individual earthquake sequence. Sequences are color-coded by faulting style of the mainshock (blue: strike-slip, green: normal and red: reverse). The red line indicates the median average for 20 Ma crustal age bins. Dashed lines indicate the corresponding interquartile ranges. Bars indicate the fraction of earthquakes with no aftershocks within each 10Ma crustal age bin and correspond to the right-hand axis.

search. Approaches to measure this quantity include modelling the earthquake process as an epidemic type sequence with a stationary background rate and various declustering algorithms. However, these methods suffer from instability, subjectivity and are subject to strong trade-off between precision (granularity) and accuracy. In the extreme, event-wise calibration of background seismicity is particularly under constrained. Background seismicity is also critically dependent on the duration of an aftershock sequence (if we let aftershock sequences continue to infinity, background vanishes). Fundamentally, background estimation is sensitive to the definition of a declustering method.

Here, we circumvent these problems by considering all events (including after-

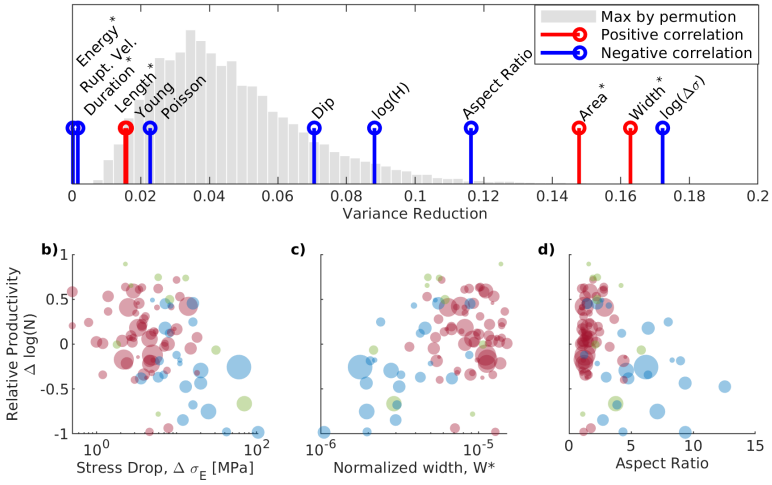

Figure B.23: Results using Zaliapin et al. 2008 clustering. a) Goodness of fit of linear regressions for each source attribute in our combined catalog. Top and bottom axes respectively represent the p-value and goodness of fit of each attribute (stems). The probability distribution function in the backdrop indicates the maximal variance reduction outcome of 10000 permutation test of the entire data set we tested. Asterisks indicate scaled and log-transformed variables. The scaled energy, length, duration and area, material properties, velocity, dip, and log-stress drop ($\Delta\sigma$) of the mainshock rupture all do not yield a statistically significant ($p = 0.05$) linear fit to the relative productivity —the normalized rupture width and aspect ratio of the rupture yield the best fitting linear regressions. Stems are color-coded to indicate whether the source attribute is positively (red) or negatively (blue) correlated with relative productivity. b-d): Relative earthquake productivity as a function of mainshock stress drop, normalized rupture width, and aspect ratio.

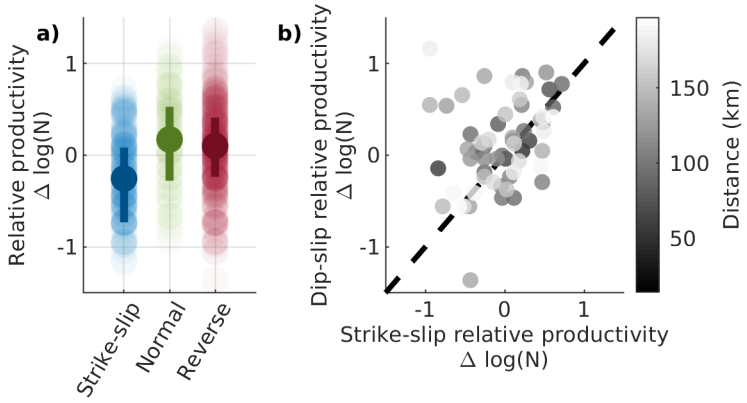

Figure B.24: Results using Zaliapin et al. 2008 clustering. a) Relative aftershock productivity ($\Delta \log(N)$ by focal mechanism (Eq. 2.5). b) Relative aftershock productivity for pairs of earthquake sequences with strike-slip and dip-slip mainshocks within 200 km from each other. Each pair is shaded according to its relative distance. Dashed line indicates a 1:1 relationship, the expectation for a purely site dominated control on relative productivity. Co-located mainshocks pairs generally follow this 1:1 trend, but exhibit considerable scatter.

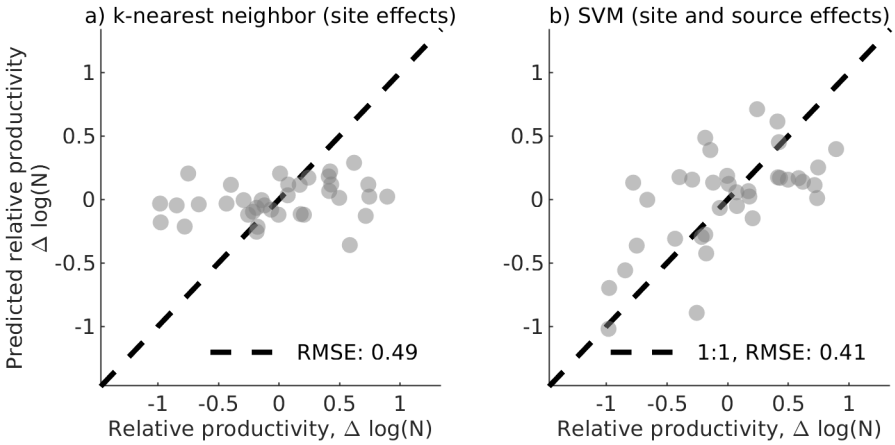

Figure B.25: Results using Zaliapin et al. 2008 clustering. Response plots (prediction versus observation) for the k-nearest neighbor algorithm (a) and SVM models (b). Each point indicates prediction of relative productivity relative to that which was observed for individual earthquake sequences. A perfect prediction would place all values on the 1:1 line.

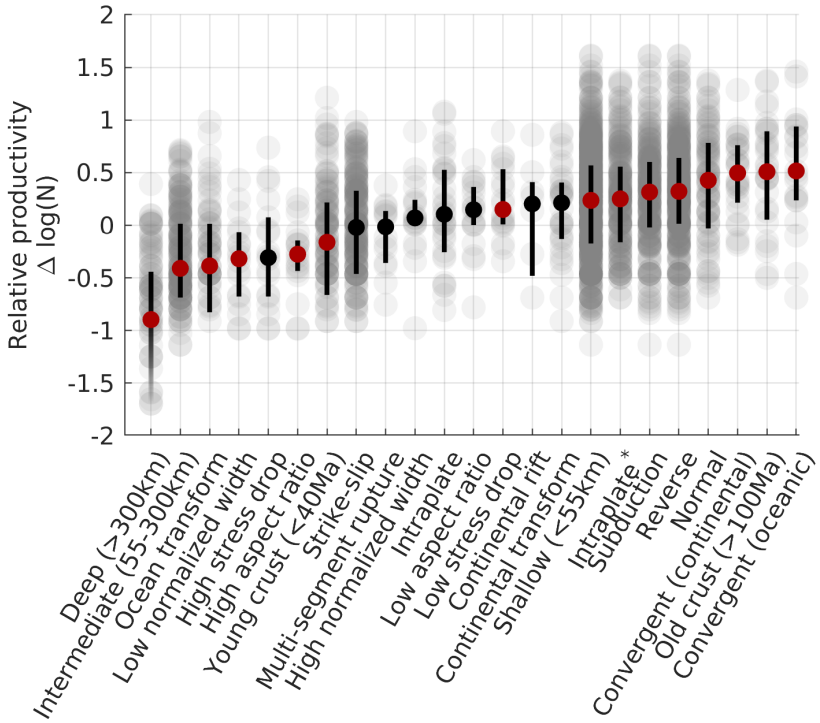

Figure B.26: Results using Zaliapin et al. 2008 clustering. Synthesis of relative productivity according to catalog subsets. The group considered here are the short list which best distinguished relative productivity based on our different lines of investigation. 'High' and 'low' subsets respectively refer to >80th and <20th percentile ranges of the data. Grey circles are individual mainshocks. Opaque points and error bars respectively indicate the median and interaquartile range of the subset. Fading error bars imply that mainshock sequences with no aftershocks are within the interquartile range of the data. Attributes with red markers are consistent with the hypothesis that they are sample from a different continuous distribution than the overall population of earthquakes using a 2-sample Kolmogorov-Smirnov test at a 5% significance threshold.

shocks) to produce an apparent background rate. We then compare this apparent background rate of events to aftershock counts. This is a conservative approach since the apparent background rate may include aftershocks that would not abide to a strict definition of background seismicity.

Fig. 2.2 in the manuscript shows how the aftershock-counts compare to a catalog where temporal information has been shuffled. The spatial structure, however, is conserved. This test is both a means to optimize the choice of space-time window but also an absolute upper-bound of the effect background seismicity as a function of magnitude. The effect is indeed scale dependent. However, comparing this effect to the aftershock counts reveals that the bias is small, typically 1 to 2 orders of magnitude below the aftershock counts. Fig. B.27 demonstrates that the bias from background seismicity lies within one standard deviation of a Poisson counting error for 97% of mainshocks. Again, this calculated bias is an absolute upper-bound which includes aftershocks and likely a significant overestimate of background activity. The occurrence of $M_W 4.5$ events in a specific small space and time window is very rare.

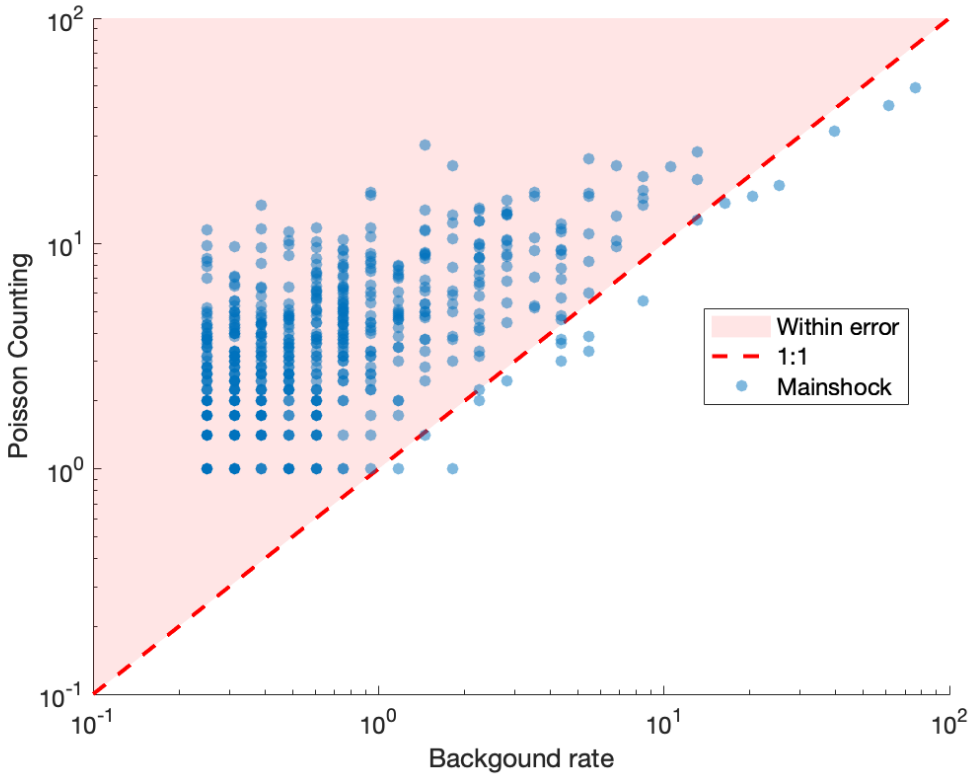

Figure B.27: Comparison of the apparent background seismicity to one standard deviation of a Poisson counting error. Dashed line indicates a one-to-one relationship. 97% of events fall in the upper half of the plot highlighted in red indicating that they are within the error of the aftershock count.

B.2 Relative productivity as a function of miscellaneous parameters

Fig. B.28 presents how the relative productivity changes as a function of each variable we tested for in this study.

B.3 Additional Supporting Information

B.3.1 Trained SVM model

We provide as a separate file, *train_SVM.m*, the function used to train our SVM mode using leave-one-out cross-validation. The function includes hyperparamaters used for our analysis. Given 1) a 2-D array of predictors where rows represent individual earthquakes and columns represent predictors (in our case dip, normalized area and plate age); and 2) a response vector (in our case the relative productivity), the function returns the predictions, *validationPredictions*, the root means squared error, *validationRMSE*, and the trained model, *regressionSVM*.

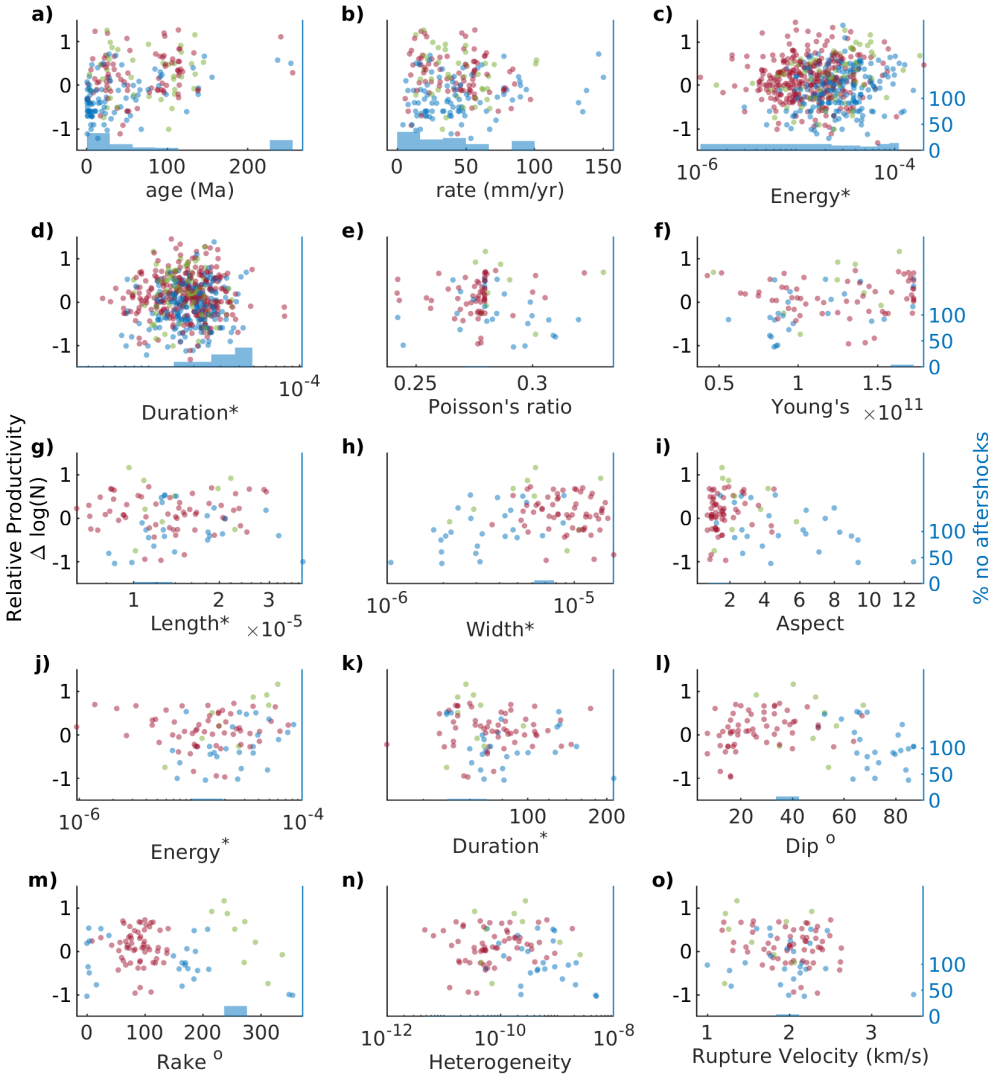

Figure B.28: Miscellaneous setting and source attributes as a function of relative productivity (N^*). Source attributes that scale with magnitude are normalized (norm.) to eliminate any such dependence. Note that for the normalized rupture durations and the normalized radiated energy, we show their relationship to relative productivity for both the Convers and Newman 2011 and Hayes 2017 estimates. Mainshocks are color-coded by focal mechanism (strike-slip: blue, normal: green, and reverse: red). Note that a linear relationship has its limitations for more non-linear relationships (e.g. dip and rake).

B.3.2 Catalogued Source Parameters

We provide separately a table (Table A.1), *FSPcat.csv*, for all the source attributes we consider. The table includes earthquake date (in *matlab datenum* format), latitude, longitude, depth (km), moment magnitude, moment (Nm), dip, rake, rupture velocity (km/s), rupture duration (s), Young's modulus (Pa), and Poisson's ratio as reported by Hayes 2017. We also include the focal mechanism, width (km), length (km), aspect ratio (length over width), slip heterogeneity and stress drop (MPa) that we derived.

B.3.3 Alternative clustering method code

The full description of the alternative aftershock detection routine is available in a Python Jupyter Notebook at the following GitHub page:

https://github.com/tgoebel/clustering-analysis

In addition, this repository contains all the functions and utilities used in the process in addition to the earthquake catalog used in this study.

Appendix C

Supplementary Material for Chapter 3

C.1 Methods

We follow the procedure outlined by Gulia and Wiemer (2019, Gulia and Wiemer 2019). For the Ridgecrest sequence, we used events starting on January 2000 from the Advanced National Seismic System Comprehensive Earthquake Catalog (see Data and Resources) up until the first M_W6.4 July 4th earthquake, after which, we use the augmented catalog produced by Shelly 2020. For the Puerto Rico sequence, we use events from the Puerto Rico Seismic Network starting on January 2003. Though the sequence is offshore, the shallow seismicity is

proximal enough to be well within the instrumental coverage of the local seismic network. We select the starting dates in both catalogs to maximize the number of events while ensuring generally consistent catalog practices.

The analysis is restricted to the local source regions of the large events of interest. The source region is established as follows. First, two candidate planes are obtained from the global centroid moment tensor catalog. The strike and dip of the plane that best fits the aftershock distribution is chosen. In the case of Ridgecrest, this determination is ambiguous and depends on the early time period we select to determine the source volume. We present results for the E-W volume in Fig. 3.2. Next, the lateral dimensions of the fault are derived from standard scaling relationships, $SRL = 10^{0.62M-2.57}$ [km] (Wells and Coppersmith 1994). Finally, the width of the source region is set to 3km. When a new large event occurs, the source location, attitude, and lateral dimensions are updated accordingly. For the Puerto Rico sequence, these strict criteria result in too few events during the foreshock sequence. We relax these spatial criteria to a 10 km radius sphere for this specific foreshock sequence. We opted to use a sphere since the lateral dimensions, which follow Wells and Coppersmith 1994 are less than the 10 km radius. We only deviate from the original method for this event. The analysis of the mainshock conforms to the original method. In a preliminary screening of the catalog, all events below catalog completeness as determined by maximum

curvature with a correction factor of 0.2 are removed. In addition, we remove all events within the 0.05-day window of the foreshock events in the sequence when catalog completeness is highly inhomogeneous. For mainshocks, we remove all events within two days. Pre- and post- sequence events and corresponding b-value time-series are treated separately. The time series of the pre-sequence background level is constructed by marching forward with a 250-event window. Within this window, events are again screened for completeness. In so doing, we corrected an apparent error in the published code and ensured that this second completeness screening is performed as described in the original method. The b-value for the group of earthquakes is estimated using maximum likelihood (Equation 1) if there are more than 50 remaining events after screening. Corresponding 1σ error is established in accordance with Shi and Bolt (1982). If there are too few earthquakes to establish background, the nearest 250 events to the first event are used to measure the background b-value. This alternative approach was used for both Ridgecrest and Puerto Rico. Following the onset of the sequence, b-value estimates proceed with a 400-event window and the completeness same screening. For each event, we ensure the time series is acausal (does not include information from future events) such that a b-value is attributed to the time-stamp of the last earthquake in the group.

C.2 Replying to Gulia and Wiemer comment on "Two foreshock sequences post Gulia and Wiemer (2019)"

In a comment on our paper, Gulia and Wiemer (2021; hereinafter GW2021) asserted that our replication of their analysis deviated in at least six different ways from the proposed method (as given in GW2019, Gulia and Wiemer 2019). These purported deviations lead to different pseudo-prospective warning levels, particularly for the Ridgecrest aftershock sequence.

Here we show that for four of the six claimed deviations, the analysis in Chapter 3 (hereonafter, DC2020) conformed to the criteria outlined in GW2019, and that, in some cases, the criticisms are in direct contradiction with the guidelines in GW2019. There are two true deviations from the defined procedure that we should have better articulated. We explain and discuss the rationale for these deviations. One attempts to reconcile the code distributed by GW2019 with the published documentation in GW2019. The other stems from a decision to encompass a volume that was robust to uncertainty in early hypocenter depths, as would be required for a real time application. We conclude, as we did originally in DC2020, by emphasizing the influence of expert judgment in the analysis. Considering the importance and public impact of real-time earthquake warnings and

the scarce opportunities for validating proposed methodologies, a next generation of the FTLS needs to be robust to differences in expert judgment, propagating model uncertainty into warning levels. To facilitate comparison with GW2021, we follow the same labeling. We group deviations 1.3 and 3.1 together as they reflect the same issue. Where it is relevant, we directly quote GW2019 to avoid misrepresentation.

C.2.1 Discussion

C.2.1.1 Deviation 1.1: DC2020 uses a catalog from the year 2000, while we would advise (and do so in Gulia et al., 2018, 2020) to use data from 1981 (GW2021).

We did use the year 2000 as a starting point for our background catalog for the Ridgecrest earthquake sequence. The start date is important for defining the b-value of background activity within a target event source volume and is a decision left to expert judgment. Our choice followed the guidelines in GW2019:

> *The start time of the pre-event catalogue depends on the quality and completeness of the local network and sometimes on avoiding overlap with past sequences (in our case, we choose 1 January 2012 for both Japan and Italy; in Italy, to avoid overlap with the L'Aquila aftershocks and in Kumamoto to avoid the influence of the 2011 Tohoku M_W9 megathrust event). The pre-event period should ideally contain several years of seismicity for a robust estimate. (GW2019)*

Figure C.1 shows local seismicity from 1980 up to the Ridgecrest sequence.

The 1980-2000 period is seismically active, with apparent triggering from large regional earthquakes including 1992 Landers and, to a lesser extent, 1999 Hector Mine, along with the strong local 1995 Ridgecrest sequence. In the accompanying comment, the authors do not mention the local 1995 Ridgecrest sequence and, despite a notable spike in local seismicity in 1992, they argue that the Landers (and Hector Mine) events are too far away to influence seismicity within the source volumes of the Ridgecrest sequence. It was our judgment that the background should be established after 2000 when seismicity is not perturbed by past local sequences or triggering from large events, as prescribed in GW2019. The background catalog interval of 18 years more than doubled the intervals used to establish a background for the two foreshock sequences analyzed in GW2019. Our choice for the catalog start date is a decision left to expert judgment by GW2019, not a deviation of the method as asserted. Some discrepancies discussed below are a direct consequence of this choice, which we view as more reasonable than extending the catalog back to 1981.

The sensitivity to the catalog start date is the first point we emphasized in our original discussion and something we explored more fully using start dates ranging from 2000 to 2012 in the Monte Carlo sampling.

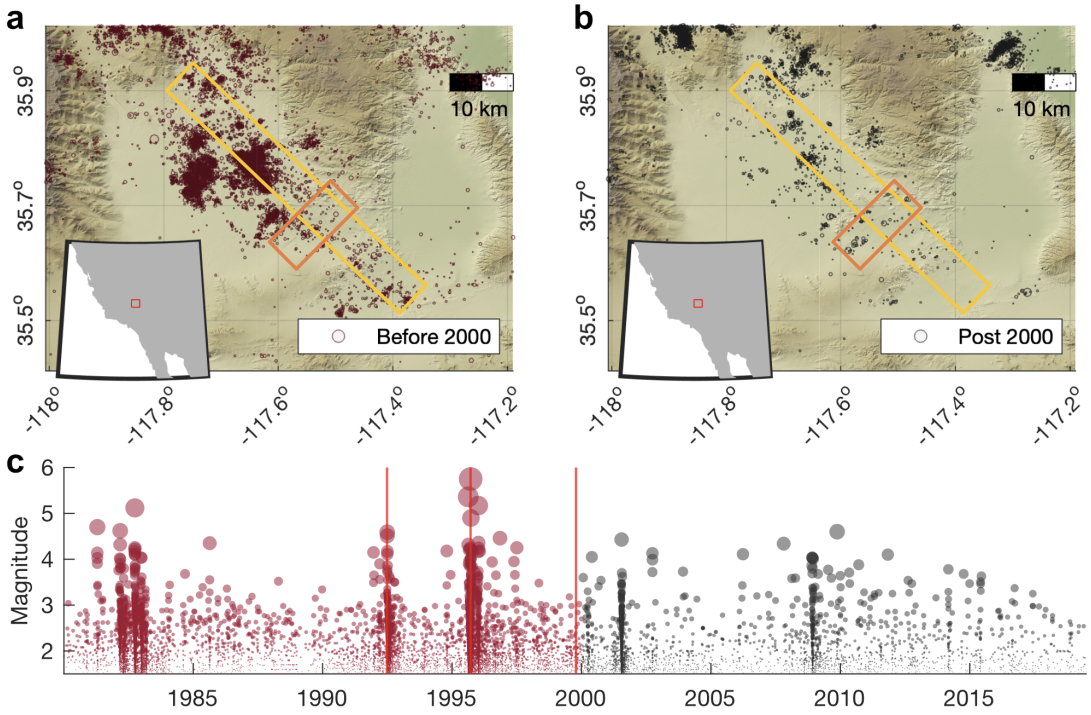

Figure C.1: Local seismicity along with approximate source volumes used for the 2018 Ridgecrest foreshock (orange box) and the mainshock (yellow box) for the period prior to 2000 (a) and after 2000 (b). The true source volume is three dimensional with lateral dimensions scaling with magnitude and orientation prescribed by the gCMT solution. (c) Time-sequence of regional seismicity in the Ridgecrest area from 1980 to 2019 prior to the Ridgecrest sequence. The area features triggered seismicity in 1992 following the Landers earthquake and intense activity in 1995 associated with the regional M_W5.7 earthquake along the 2018 Ridgecrest mainshock rupture.

C.2.1.2 Deviation 1.2: DC2020 apply erroneously an additional Mc increment 0.4 rather than 0.2 (GW2021).

DC2020 does indeed apply the magnitude of completeness correction twice. We used the code distributed by GW2019 which necessitated adaptation for individual mainshocks because of hard-wired parameters which we struggled to reconcile with the procedure as described in the Methods section of GW2019. In GW2019's primary code (Run_TLS_Gulia_Wiemer.m), the multi-step correction procedure is inconsistent across the background, foreshock, and aftershock intervals. The background events are screened for events below the completeness thresholds of MMAXC which is implemented with a +0.2 increment at line 162 and a -0.2 increment at line 172. The foreshocks are screened with MMAXC+0.2 (lines 168-176). The aftershocks are screened with Mc=1.8 (line 237). After these initial screenings, the data are further screened for completeness with an additional +0.2 correction (function calls on lines 192 or 214 for the background period; lines 226-227 for the foreshock period and lines 244-245 for the aftershock period). Thus, aggregate corrections of 0.2, 0.4, and 0.2 were applied respectively. Faced with a non-transferrable processing scheme that both did not describe how an initial $M_c = 1.8$ was selected for the aftershock interval and differed in our reading from the documentation provided in the published paper, we opted to uniformly use the conservative and consistent correction of 0.4 to the maximum curvature

estimates (+0.2 globally, with additional +0.2 in every event window for our primary result) that is consistent with the coded method in GW2019 for the most important period, i.e., the foreshocks.

Estimating catalog completeness is susceptible to the peculiarities of local catalogs. We therefore also included in DC2020 a range of corrections ranging from 0.1 to 0.3 in our Monte Carlo simulation with the corrections applied at both screening steps, resulting in aggregate corrections of 0.2-0.6. Using more conservative completeness thresholds trades off precision (reported with the standard deviation at every time-step), for accuracy. For the sake of completeness, we have now additionally performed the primary calculation of DC2020 with a single +0.2 correction (Fig. C.2). Lower relative b-values particularly during the aftershock sequence toggle the aftershock period from a yellow (ambiguous) to red warning indicating an impending larger earthquake that has not yet been observed. Since the original code provided with GW2019 hard-wired specific mainshocks into the code, we provide at the end of this chapter our version that produce the figures in DC2020 as well as Fig. C.2 below.

C.2.1.3 Deviation 1.3 and 3.1: To establish the reference b-value, DC2020 sample events in a circular region of about 10 km around the epicenter, while Gulia et al., (2020) use events in a box within 3km of the rupture plane (GW2021).

The background seismicity in the Ridgecrest source volumes from 2000 to 2018 is very sparse (Fig. C.1b), presenting difficulty estimating a stable b-value. For sparse background seismicity, GW2019 indicate the following procedure:

> *(b) If fewer than N_{pre} [250] events are available, we use the N_{pre} [250] events that are nearest to the epicentre and then compute a single regional background b value as reference, following the computational approach defined in (a). This procedure was used for the $M_W = 6.5$ Kumamoto event (Fig. 2a), sampling a distance of up to 17 km from the epicenter.(GW2019)*

N_{pre} (250 events) refers to the minimum number of background events remaining after the first completeness correction in the target source volume (a box within 3 km of the estimated rupture surface inferred from scaling relationships for the event magnitude). Since we used a later catalog start date (2000 instead of 1981, as discussed above) in a region with low true background activity, we have far fewer background events in the two choices of foreshock and the mainshock source volumes for the Ridgecrest sequence (Fig. C.1). This situation triggers the above-stated processing step in the original code. Thus, instead of using the deficient number of events in the box within 3 km from the target faults in Ridgecrest, we use the nearest 250 events to the epicenter in the regional catalog spanning

the catalog start time to the occurrence of the MW6.4 foreshock, as specified by GW2019. Low background activity levels trigger this condition for both the foreshock ("deviation 1.3") and mainshock ("deviation 3.1") for the Ridgecrest and Puerto Rico events.

C.2.1.4 Deviation 2.1: DC2020 selected aftershock of the first hour, rather than the first 24 hours to define the active fault. They thus selected the alternate fault plane for estimating the b-values of the aftershocks following the first mainshock (GW2021).

We note the guidance from GW2019 on this point:

> *Typically, one to several hours of aftershocks are sufficient to select the right plane, and rapid source-inversion approaches can also deliver a finite fault model within 1–2 h.*

GW2019 offer no further guidelines. DC2020 indeed selected the SW-NE fault plane based on the first hour of aftershocks for the main solution. No point in the original paper specifies that the first 24 hours should be used to define the fault. We note that early finite fault models also assumed rupture on the SW-NE plane because the early aftershock distribution has greater numbers and length on that plane than on the orthogonal plane of the "L-shaped" rupture. The 2018 Ridgecrest foreshock is truly ambiguous in terms of defining the "right"source volume in the procedure of GW2019, given that both planes ruptured simultaneously

(Liu et al. 2019). There is no guidance in GW2019 for a case involving simultaneous rupture of two orthogonal planes. We therefore included either E-W or N-S choices of source volume in the Monte Carlo suite of calculations. The deficiency of background events from 2000 to 2018 triggers the 'nearest 250'event option for estimating a reference *b*-value for either choice of the foreshock target source volume, and with the epicenter being near the intersection of the two planes, this gives the same reference *b*-value estimate for both cases.

C.2.1.5 Deviations 2.2: DC2020 do not limit the analysis to events with 3 km depth below and above the fault plane, but extend the sampling down to 20 km.

We thank Gulia and Wiemer for raising the issue of the depth-extent of the source volume. This is a challenge for near-real-time applications due to time-varying estimates of hypocentral position. In DC2020, we established source volume boxes centered upon the hypocentral locations and rotated according to the strike and dip of the respective foreshock and mainshock faulting mechanisms. These source volumes extend 3 km on either side of the rupture plane, as prescribed in GW2019. Estimated horizontal dimensions of the source-volume boxes were defined by the relationship in Wells and Coppersmith 1994 for strike-slip faulting subsurface rupture length (RLD in their Table 2A). This gave lengths of 25 km for the foreshock and 68 km for the mainshock. The relationships in Wells

and Coppersmith 1994 also estimate downdip subsurface width (RW) for strike-slip faulting, which gives 9.3 km (5.25-16.44 km range) for the foreshock and 14.4 km (7.9 to 26 km range) for the mainshock. One can then center the box on the hypocenter (as prescribed by GS2019). We deviated from this depth windowing procedure by including source volumes extending across the $\sim$15 km seismogenic layer depth; this is a true deviation from the defined procedure that we should have articulated, with a justification as given below. Early catalog locations gave hypocentral depths of 10.7 km for the foreshock (available by 9:21 am on July 15, 2019) and 17 km for the mainshock (available on July 6, 2019), but the latter was quickly revised to 8 km. Our post-processing of the sequences used the catalog of (Shelly 2020) which gives hypocentral depth estimates of 15 km and 3 km for the foreshock and mainshock, respectively. This substantial variability in hypocentral depths combined with a choice of centering the source volume box on the uncertain hypocentral depth significantly impacts the placement of the source volume used for fault plane preference and b-value time series calculations. The vast majority of seismicity in this region is from 1 to 15 km depth. Immediate aftershock hypocenters of both events spanned this full depth range. Using the final catalog depth of 8 km for the mainshock, with central value for RW gives a depth range of 8 ± 7.2km, which spans the main seismogenic zone. By allowing a greater depth extent, only a few additional events are included in our analysis

of the mainshock sequence. There would be a significant truncation of activity if we centered the source volume on hypocenters at either 17 km or 3 km for the same RW value. For the foreshock hypocenter at 10.7 km, using the central value for RW gives a depth range of 10.7 ± 4.7km or 10.7 ± 8.2km for the maximum range of RW. Using the Shelly catalog depth of 15 km, centering the source volume leads to up-dip truncation of the seismicity. Our choice of extending the source volume across the seismogenic zone has a greater impact on the foreshock sequence than on the mainshock sequence. Still, with early seismicity extending across the depth range, and considering that Wells and Coppersmith 1994 themselves calibrate rupture dimensions on early aftershocks, it is not clear that our decision is any less or more correct regarding the down-dip width of the source volume. Uncertainty in actual hypocentral depth and ambiguity as to whether the hypocenter is at the top, bottom or center of the actual rupture area, warranted extending the source volumes throughout the depth range of the activity. In Southern California, where most seismicity is less than 15 km depth, this would be our recommended way of dealing with the uncertainty in hypocentral depth estimates and down-dip width of the source volume for large events. For other regions, this may not be appropriate.

C.2.2 Puerto Rico Case study

As the GW2020 notes, DC2020 explicitly stated that the analysis of the Puerto Rico Sequence did not strictly comply with the criteria of GW2019, in part due to the low magnitude of the largest foreshocks and in part due to paucity of events within the putative M 5.0 foreshock source volume, which was widened from 3 km to 10 km from the fault to capture sufficient activity. In the same way that GW2019 included the Tohoku sequence without giving weight to the findings for it, we included application to a second clear foreshock sequence despite the catalog limitations. The near-coast sequence is unusually well monitored for an offshore event (in contrast to the Tohoku case); and its aftershock sequence was vigorous and ongoing during our submission process, making it a fully prospective test (albeit of a slightly adapted hypothesis).

C.2.3 Conclusion

Readers that have studied GW2019, DC2020, GWV2020, and the comment and reply may find this discussion somewhat unsatisfactory. The prognostic value of the traffic light system and the physical implications thereof remain ambiguous. GW2019, GWV2020, and to some degree DC2020, all suggest that there may indeed be a precursory signal in the b-value time series of foreshocks. Following our best judgment, DC2020 found that the Ridgecrest foreshock sequence did have

reduced b-value along the E-W plane, that dipped below background level during the foreshock sequence (red warning). This is consistent with GWV2020 who used an earlier catalog start time and the N-S plane for estimating the background b-value. The magnitude of this dip and its recovery with time, however, are found to be sensitive to seemingly innocuous decisions associated with expert judgement which determine whether options in the FTLS method of GW2019 are exercised. For the Ridgecrest mainshock, the main difference in DC2020 and GWV2020 is that the latter paper estimates a lower background b-value apparently because their time window includes a prior earthquake sequence and triggered activity. This enhances the b-value increase of the aftershock sequence relative to background, giving a green light in GWV2020, whereas the higher background b-value estimate in DC2020 yields a yellow light, or ambiguous definition of aftershock activity. The goal of DC2020 was not to stunt progress with this approach but rather to contribute to its advancement. Independent research groups replicating scientific methods is a desirable feature that we hope to continue fostering in our community. The outcome of this exercise illustrates the concerns we raised in our original submission. The reliance on expert judgement suggests that the method, as stated, lacks the specificity to robustly monitor the underlying physical process accurately. The method, as defined in GW2019, needs the following improvements: 1) explicit and physically motivated criteria for determining background

b-values free of contamination by prior sequences and remotely triggered activity, particularly for low background level regions, 2) a more robust method for the determination of catalog completeness for b-value estimation and 3) a provision for multi-fault ruptures and uncertain mainshock rupture source dimensions. Foregoing empirical calibration in favor of early aftershock distributions to directly estimate the source volume may help address this last point. Improving robustness of the b-value estimation with respect to time-varying catalog completeness may greatly help reduce short term fluctuations and increase usable data during an ongoing sequence. Other concerns arise with respect to using sufficient numbers of observations for any b-value estimate to be meaningful. Addressing these issues will ensure that prospective tests are comparable across multiple research groups without ambiguity.

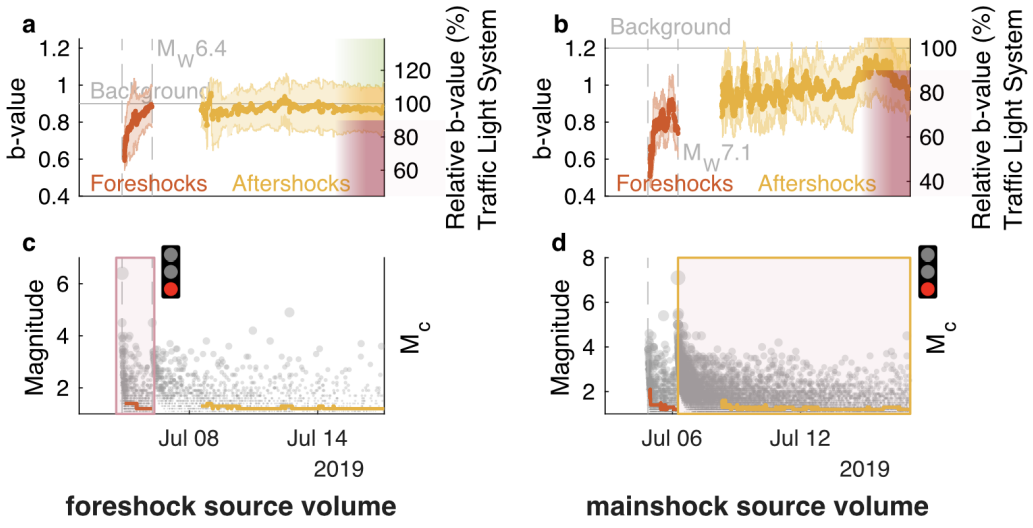

Figure C.2: Primary result for Ridgecrest with a single completeness correction. (a-b) Time series of b-value estimates during the sequence with 1σ error bars for the corresponding source volumes indicated below (c-d) and shown in Fig. C.1b. Dashed lines indicate the timing of the 4 July 2019 M_W6.4 foreshock and the 5 July 2019 M_W7.1 mainshock. The traffic-light criteria relative to the background level are indicated on the right. (c-d) Time series of event magnitudes during the sequence in the corresponding volumes (Fig. C.1b). Colored curves indicated the time-varying catalog completeness, M_c, during the intervals of the foreshock and aftershock sequences used for b-value computation. Note that with a single correction the traffic light warning for both the foreshock (c - red box) and aftershock (d - yellow box) periods are red, indicating that there should be an impending larger earthquake that has not yet occurred.

Appendix D

Supplementary Material for Chapter 4

D.1 RECAST implementation

RECAST encodes the event history into a fixed-dimensional vector $\boldsymbol{h}_i$ using a recurrent neural network (RNN). Then, the history encoding $\boldsymbol{h}_i$ is used to parametrize the conditional PDF $f^*(t_i)$ of the next origin time t_i, where the * indicates that the probability distribution function is conditioned on the past events history $\mathcal{H}(t_i)$.

The RNN starts with the initial hidden state $\boldsymbol{h}_1$ represented by an H-dimensional vector. Every time an event (t_i, M_i) occurs, a new hidden state

$\boldsymbol{h}_{i+1}$ is computed from the previous hidden state $\boldsymbol{h}_i$ and the observed values of (t_i, M_i) using the RNN update equation. In this study, we use a Gated Recurrent Unit (GRU) architecture for the RNN, as defined in the original paper (Shchur, Biloš, and Günnemann 2020). We represent each event (t_i, M_i) with a 2-dimensional vector $\boldsymbol{y}_i = (\log(t_i - t_{i-1}) - \log \bar{\tau}, M_i - M)$ that we provide as input to the RNN. The feature vector $\boldsymbol{y}_i$ consists of the logarithm of the inter-event time and the earthquake magnitude, both standardized by respectively subtracting the log of the mean inter-event time and mean magnitude M in the training set. If other features (e.g., locations) are available, they can be concatenated with y_i. Once the hidden state $\boldsymbol{h}_i$ is computed by the RNN, it is used to obtain the probability density function (PDF) $f^*(t_i)$ of the next origin time t_i.

More specifically, we model the time until the next event $\tau_i = t_i - t_{i-1}$ with a Weibull mixture distribution $f(t_i|\boldsymbol{\theta}_i)$ with R components, where the parameters $\boldsymbol{\theta}_i$ are computed from the hidden state $\boldsymbol{h}_i$. Our choice of the Weibull mixture distribution has several important advantages. The distribution is flexible, as it is able to capture multimodality, as well as increasing, decreasing or constant hazard rates. Moreover, the survival function of the Weibull distribution can be computed and inverted analytically, which allows us to efficiently compute likelihood and generate samples from the TPP model. Many popular existing neural TPP models fail to satisfy all these conditions simultaneously and rely on

slow approximate iterative methods during training and sampling (Du et al. 2016b; Mei and Eisner 2017; Omi, and Aihara 2019). The PDF of the Weibull mixture distribution is defined by 3R strictly positive parameters $\boldsymbol{\theta}_i = \{\pi_r, b_r, k_r\}_{r=1,\dots,R}$ where r are the mixing weights, and (b_r, k_r) define the rate and shape of each component,

$$f(\tau; \boldsymbol{\theta}_i) = \sum_{r=1}^{R} \pi_r f(\tau | b_r, k_r) = \sum_{r=1}^{R} \pi_r b_r k_r \tau^{k_r - 1} \exp\left(-b_r \tau^{k_r}\right) \tag{D.1}$$

We obtain the parameters $\boldsymbol{\theta}_i$ by passing the history encoding $\boldsymbol{h}_i$ through an affine layer, $\boldsymbol{\theta}_i = A\boldsymbol{h}_i + \boldsymbol{\beta}$, and applying suitable nonlinearities to enforce the constraints on the parameters (softmax for r's and softplus for b_r's and k_r's). We clamp the minimum values for the scale and shape parameters to be no less than -5 to prevent numerical instabilities in PDF computation. To summarize, the learnable parameters of the RECAST model consist of weights of the GRU encoder and weights of the affine decoder.

For full reproducibility, we refer readers to the PyTorch implementation of RECAST:

https://gitlab.com/shchur/recast

For an introductory walkthrough of an (unmarked) neural temporal point process we refer readers here:

https://shchur.github.io/blog/2021/tpp2-neural-tpps/

The selection of hyperparameters and components of the model architecture

provide some additional degrees of freedom and introduce the risk of over-fitting. We adopted fixed values $H = 32$, $R = 32$, as suggested in the previous work. We tried some modifications (larger H and R, deeper networks). This, however, did not noticeably change the loss on the validation set, so we chose the simplest model and used it in all the experiments. We have not used the test set to tune any hyperparameters or model architecture to ensure a fair comparison with the baseline.

D.2 ETAS implementation

The ETAS model parametrizes the conditional intensity function $\lambda^*(t)$ as

$$\lambda^*(t) = \mu + \sum_{i:t_i<t} \Phi(t - t_i, M_i) \tag{D.2}$$

where μ is the background rate of events and $\Phi(\tau, M) = k10^{\alpha(M-M_c)}(\tau+c)^{-p}$ is the excitation function. The excitation function consists of the aftershock productivity law $k10^{\alpha(M-M_c)}$ that characterizes the expected number of aftershocks given the magnitude, and Omori's law, $(\tau + c)^{-p}$ that characterizes the temporal decay of the aftershock rate. Thus, the contribution of past events depends on the catalog completeness M_c and the positive model parameters $\{\mu, k, \alpha, c, p\}$. To ensure positivity of the parameters, we perform optimization for ETAS in the log space. We implemented ETAS in PyTorch using the same framework as RECAST. Our

implementation of ETAS therefore benefits from automatic differentiation, GPU acceleration and integration with standard optimization algorithms.

D.3 Model Training

The parameters of both models are learned using maximum likelihood estimation. Given a sequence of earthquakes $((t_1, M_1), \ldots, (t_N, M_N))$ in the time interval $[T_{start}, T_{end})$, the time-averaged log-likelihood function $\log \widetilde{\mathcal{L}}$ describes how well a given model predicts the origin times of the events

$$\log \widetilde{\mathcal{L}} = \frac{1}{T_{end} - T_{start}} \left(\sum_{i=1}^{N} \log f^*(t_i) + \log S^* \left(T_{end} - t_N \right) \right), \tag{D.3}$$

where the additional term $S^*(T_{end} - t_N)$ is the probability of observing no event between the last event, t_N, and the end of the interval. $\log \widetilde{\mathcal{L}}$ is equivalently expressed in terms of the conditional intensity function $\lambda^*(t)$ as

$$\log \widetilde{\mathcal{L}} = \frac{1}{T_{end} - T_{start}} \left(\sum_{i=1}^{N} \log \lambda^*(t_i) - \int_{T_{start}}^{T_{end}} \lambda^*(s) ds \right) \tag{D.4}$$

Both definitions of $\log \widetilde{\mathcal{L}}$ correspond to the exact same quantity. Time normalization ensures that experiments optimized over multiple catalogs (e.g. synthetic catalogs), catalogs with more events contribute more heavily to the overall log-likelihood score.

We maximize the log-likelihood of both models on the training set by per-

forming gradient ascent with Adam30 on the model parameters with a learning rate of 0.01. To avoid overfitting, after each training step over the entire event catalog (hereinafter, one epoch), we evaluate the log-likelihood of the model on a validation set and stop training once the validation objective stops improving. We allow for a maximum of $1,500$ training epochs and perform early stopping if the validation loss does not improve in the last 200 epochs. After training, we recover the parameters of the model with the best performance on the validation set.

In our experiments we train both RECAST and ETAS models using 5 random initializations of the model parameters. This is done to mitigate the potential optimization problems caused by presence of local optima in the loss landscape16. We initialize ETAS parameters with an initial set of parameters $\{\mu, k, \alpha, c\}_0 = \{0.02, 0.0073, 1.0, 1.08\}$ and randomly perturb them by adding Uniform(-0.3, 0.3) noise in the log-space. For both model types, we report the final results for the model that performed best on the validation set. We found that in practice both RECAST and ETAS are not sensitive to different random initialization when trained using Adam optimizer on the catalogs we consider ($> 10,000$ events). Moreover, ETAS consistently converges to nearly identical sets of parameters and loss for all random seeds and catalogs we considered.

For synthetic datasets, we generate many individual catalogs. The training,

validation and testing is conducted on a set of $1,000$ synthetic earthquake catalogs (available here: https://gitlab.com/shchur/recast). The training, validation and testing sets respectively comprise 600, 200, and 200 catalogs from the set. For real earthquake catalogs, we split each individual catalog into conditioning, training, validation and test periods (Fig. 4.4), the splits for each catalog considered are specified in Table 4.1.

D.4 Forecast generation

Given sequence of earthquakes leading to a certain time T, we would like to predict seismicity for the following W days (i.e., the time period $[T, T + W]$).

For ETAS we use standard sampling approaches for a branching process31 over the forecast interval. Event magnitudes are drawn from the Gutenberg–Richter Distribution with the b-value estimated using the maximum likelihood estimation on the training set.

For RECAST, we can generate a collection of catalog continuations conditioned on the observed events before T. Each such continuation is generated iteratively one event at a time.

1. The next inter-event time τ_i is sampled from the PDF produced by the model;

(a) The component of the mixture distribution is chosen according to the categorical distribution;

(b) Inter-event time τ_i is generated from the chosen Weibull component with inverse transform sampling;

2. The magnitude M_i for the next event is drawn from the Gutenberg–Richter Distribution with the same b-value estimated on the training set;

3. The event is appended to event history;

4. The RNN hidden state is updated given the new event;

5. The PDF of the next inter-event time is generated by the model.

This sequential procedure is repeated until the end of the forecasting interval, thus producing a single sample continuation. As is the case during training, sequential processing implies that RECAST is more efficient during sampling as well, $O(C \times L)$ for RECAST vs. $O(N \times C \times L)$ for ETAS, where C is the number of catalog continuations and L is the expected number of events in the forecasting interval. Moreover, RECAST can generate thousands of sample continuations in parallel, leading to even higher efficiency.

D.5 Model comparison

D.5.1 The conditional intensity function

The ETAS model is usually presented in terms of the conditional intensity function $\lambda^*(t)$, while we defined RECAST using the PDF of the origin times $f^*(t_i)$. Both representations, however, are equivalent, and we can compute the conditional intensity of the RECAST model, continuously in time, as

$\lambda^*(t_i) = \frac{f^*(t_i)}{S^*(t_i)}$,

in each event interval where $S^*(t_i) = \int_{t_i}^{\infty} f^*(t)dt$ is the survival function of the ith event that can be computed in closed form for the Weibull mixture distribution.

D.5.2 Model likelihood and goodness of fit

Since the log-likelihood describes the goodness of fit for both ETAS and RECAST, it also provides a principled way to compare the two models. For this, we compute the logarithm of the likelihood ratio (LLR) defined as the difference of the log-likelihoods normalized by the duration of the evaluation.

$$LLR = \log \widetilde{\mathcal{L}}_{RECAST} - \log \widetilde{\mathcal{L}}_{ETAS} \tag{D.5}$$

Positive LLR means that RECAST fits the data better than ETAS, and $LLR < 0$ implies the converse. Note that when we evaluate the log-likelihood on the test

interval (e.g Fig. 4.2B), both models are conditioned on the previous events that occurred in the train and validation intervals.

D.5.3 Forecast accuracy

In the case of an extended forecast, to evaluate the quality of a forecast, we compare the cumulative number of events during the forecasting interval across all sample continuations against the observed value. More formally, the accuracy is measured by the log-score:

$$\log r = \log \left(\sum_{k=1}^{N_{tot}} I(n_k = n_{obs})/N_{tot} \right) \tag{D.6}$$

where $I(x)$ is the indicator function that is 1 where x is true and 0 otherwise, n_k is the number of events in the kth simulation, nobs is the number of events actually observed during the forecasting interval and N_{tot} is the total number of simulations that we set to 50,000. r is a proper scoring rule and so is a theoretically sound way to compare model forecasts32. Note that we explicitly choose not to prescribe a probability distribution function to the forecast since forecasts (particularly RECAST forecasts) may not follow standard distributions. This metric of forecast performance can then be compared between the models considered such that

$$\Delta \log r = \log r_{RECAST} - \log r_{ETAS} \tag{D.7}$$

is an estimate of the relative forecasting accuracy, where positive values indicate superior RECAST performance.

D.6 Data

To supplement the analysis in the main text, we conducted our analysis at various scales and sub-regions in Southern California.

D.6.1 Southern California Seismic Network

The Southern California Seismic Network has developed and maintained the standard earthquake catalog for Southern California as documented in Hutton, Woessner, and Hauksson 2010. The catalog features events from 1932 to present (2022). Over this period, the catalog completeness has improved significantly from M_c3.25 to 1.8. Significant improvements to the network occurred from 1970-1975, punctuated improvements in the following 1981 and 1985 years, and steady improvement from 1995 to present. The catalog provides relatively consistent onshore coverage from the United-States Mexico Border up to the edge of the historical southern California network north of the San Andreas Big Bend (Fig. D.3). For our analysis, we set an M_c2.0 consistent with a conservative interpretation of on shore catalog completeness using the entire magnitude range from 1980 to present, per Fig. 5 in Hutton, Woessner, and Hauksson 2010. The most

notable events in this catalog include the 1952 MW 7.5 Kern County, 1992 M_W7.3 Landers, 1999 M_W7.1 Hector Mine and M_W7.1 Ridgecrest sequences. For model evaluation the catalog was split into training (1981-01-01 to 2004-12-31), validation (2005-01-01 to 2014-12-31) and test (2015-01-01 to 2021-12-31) periods (Fig. 4.4). The catalog was downloaded from the Southern California Earthquake Data Center (https://service.scedc.caltech.edu/ftp/catalogs/SCSN/) on January 31st, 2022.

D.6.2 Detailed Earthquake Catalog for the San Jacinto Fault-Zone Region in Southern California

White, Ben-Zion, and Vernon 2019 have augmented the seismicity catalog with a particular focus on the San Jacinto region in Southern California by leveraging a dense seismic network. Automated phase detection is combined with relocation procedures. The San Jacinto features persistent elevated rates of seismicity and is sufficiently well monitored to capture numerous moderate sized earthquake sequences (Fig. 4.4). The catalog features events from 2008 to 2021. The authors report 99% event detection rate in the focus region for events with magnitude greater than 0.6. A secular increase in the number of events detected every year along with a gradual increase in the number of stations (stabilizing around 2014) suggests that the catalog completeness increases with time. The focus area fea-

tures the following notable events: 2010 M_w5.4 Borrego Springs, 2013 M_L4.7 Anza Borrego, and 2016 M_w5.2 Borrego Springs. The catalog has marked edge artifacts with events clustered on the edge of the domain. In preparing this catalog, we trimmed the edge (5% of the width and height respectively) of the domain to limit the influence of this artifact. For model evaluation the catalog was split into training (2008-01-01 to 2013-12-31), validation (2014-01-01 to 2017-12-31) and test (2018-01-01 to 2020-12-31) periods. The catalog was downloaded January 2022 (https://data.mendeley.com/datasets/7ywkdx7c62/1).

D.6.3 Quake Template Matching (QTM) catalog for Southern California

Using the SCSN seismic network, Ross et al. 2019b applied a region-wide template matching analysis to greatly increase the density of earthquake observations relative to the regional catalog (Hutton, Woessner, and Hauksson 2010). A nominal magnitude of completeness is reported to be around 0.3. However, given the template matching approach is most effective at revealing events within close proximity of previously cataloged events, the catalog completeness is susceptible to significant spatial variability. The catalog is sufficiently dense in time that the minimum inter-event duration, around 2s, is susceptible to influence the catalog, particularly during active earthquake sequences. While we do not expect this to

Table D.1: Time-averaged NLL scores for the catalogs examined in the study. Positive differences imply improved model fit.

Dataset Name	**Model name**	**Training NLL**	**Validation NLL**	**Test NLL**	**Test NLL Difference**
White et al.	RECAST	-63.633	-76.981	-89.636	
	ETAS	-63.178	-76.661	-89.309	+0.327
SCEDC	RECAST	-22.876	-13.637	-14.261	
	ETAS	-22.636	-13.491	-14.255	+0.006
QTM Salton Sea	RECAST	-72.691	-3.836	-13.109	
	ETAS	-73.819	-3.758	-13.022	+0.087
QTM San Jacinto	RECAST	-7.613	-2.995	-6.745	
	ETAS	-7.443	-2.939	-6.673	+0.072

have a strong influence on a temporal forecast, it is worth noting the catalog is not completely causal. That is, templates can be used to identify events that precede their occurrence. As such, a pseudo prospective forecast will unfairly benefit from a catalog that is spatially more dense where earthquakes will happen later in the catalog. The catalog includes events from 2008 to 2017 and contains more than 1.81 million earthquakes (Fig. 4.4, D.5). During this period, the El Mayor Cucapah earthquake (outside the study area) resulted in a strong increase in regional seismicity. We used this catalog to support results obtained using catalogs for the San Jacinto area (lat: [33.0, 34.0] lon: [-117.0, -116.0], Fig. D.5) and also tested whether improvements documented persist for the Salton Sea area (lat: [32.5, 33.3] lon: [-116.0, -115.0]). The catalog was downloaded January 2022 (https://service.scedc.caltech.edu/ftp/QTMcatalog/qtm_final_12dev.hypo).

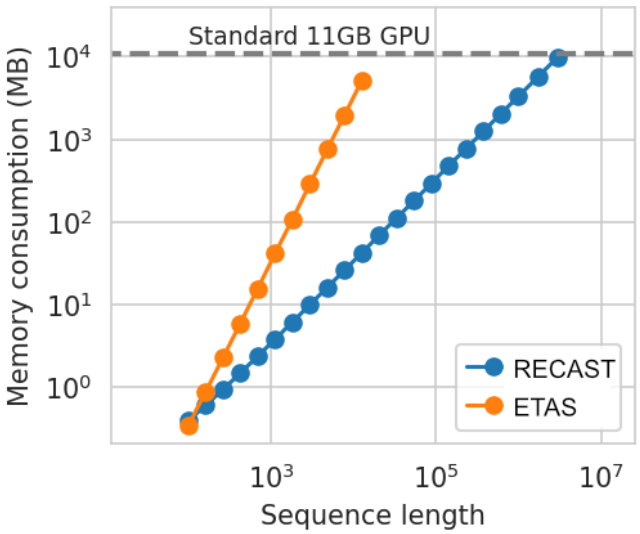

Figure D.1: Memory Consumption as a function of sequence length. Since we leverage a recurrent neural network to update a hidden state we drastically reduce the memory consumption both during training (in the evaluation of the model log-likelihood) and during sampling to generate earthquake forecasts.

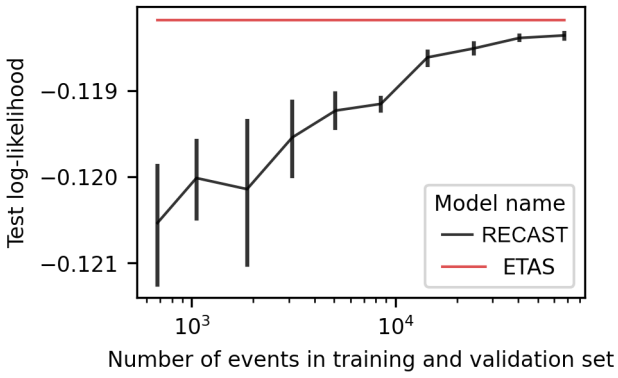

Figure D.2: Model performance on synthetic data. On synthetic catalogs generated by ETAS, RECAST approaches the theoretically optimal performance of the ETAS model as we increase the training set size. In contrast, on real-world datasets, the RECAST outperforms ETAS given a large enough training set size (Fig. 4.3).

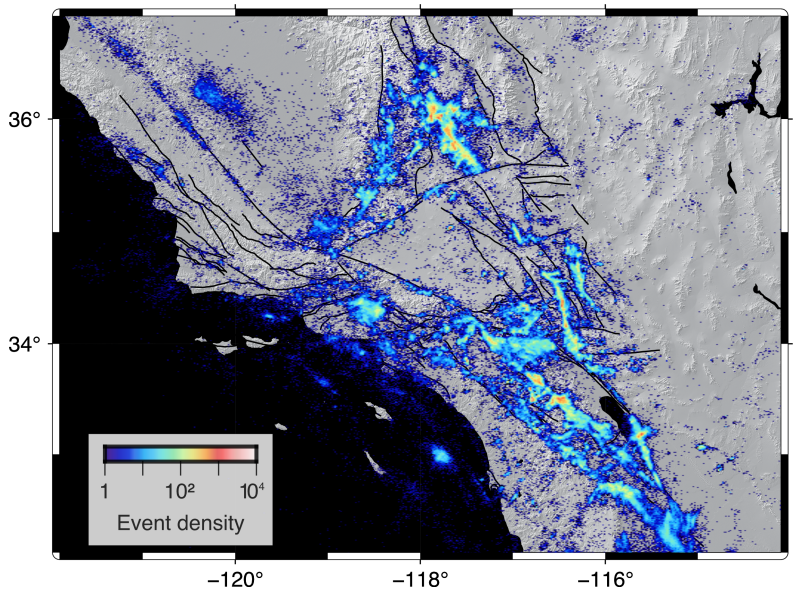

Figure D.3: Standard Southern California Catalog developed by Hutton et al., 2008 and maintained by the SCEDC.

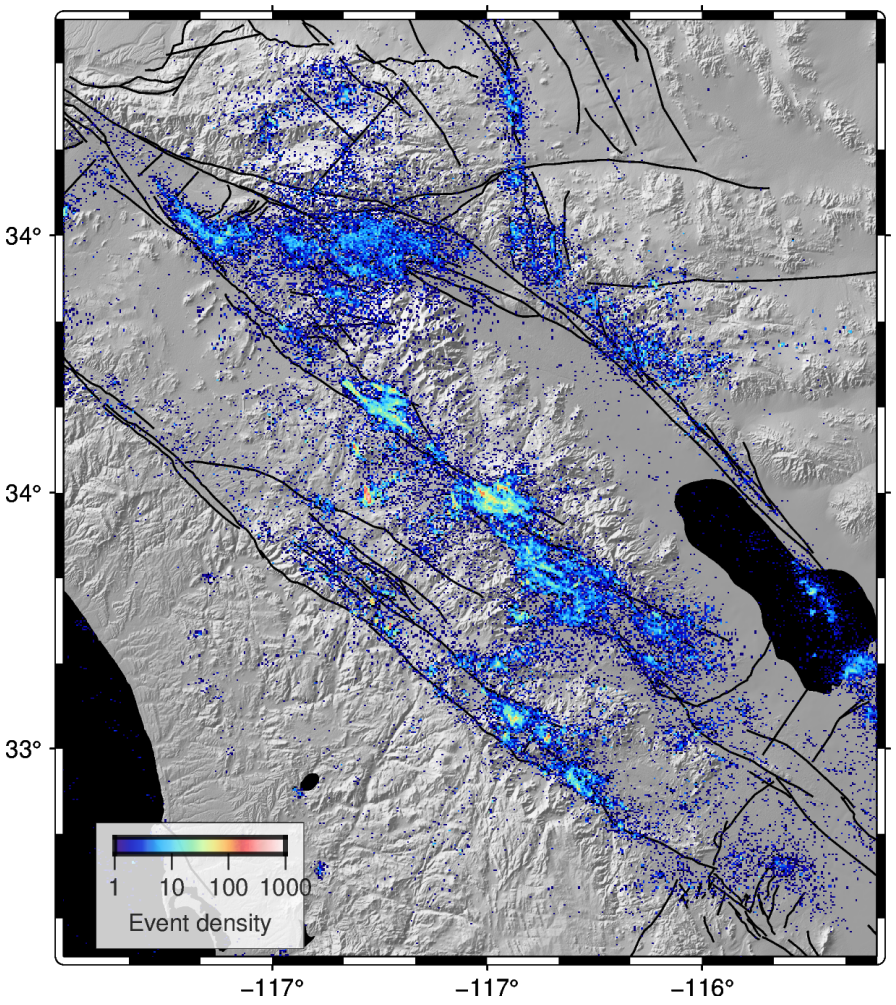

Figure D.4: Augmented catalog developed by White et al., 2019, for the San Jacinto Region.

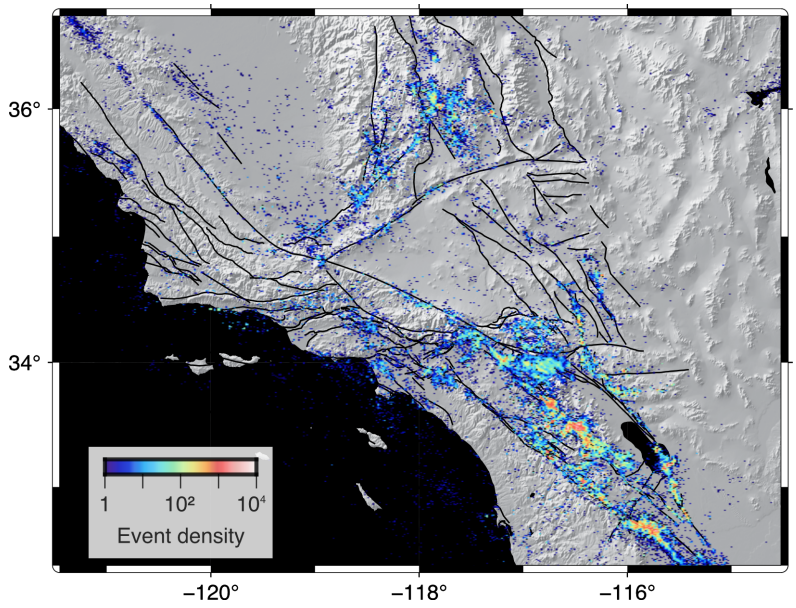

Figure D.5: Quake template matching (QTM) catalog developed by Ross et al., 2019.