Bimodal seismicity in the Himalaya controlled by fault friction and geometry

Luca Dal Zilio[1], Ylona van Dinther[2], Taras Gerya[1] & Jean-Philippe Avouac[3]

[1] *Geophysical Fluid Dynamics, Institute of Geophysics, ETH Zürich, Sonneggstrasse 5, 8092 Zürich, Switzerland*

[2] *Seismology and Wave Physics, Institute of Geophysics, ETH Zürich, Sonneggstrasse 5, 8092 Zürich, Switzerland*

[3] *Geological and Planetary Sciences, California Institute of Technology, Pasadena, California 91125, USA*

The 2015 magnitude 7.8 Gorkha earthquake is the latest large earthquake in the Himalaya. Ample observations document the bimodal distribution of Himalayan seismicity: large blind earthquakes (M7+) tend to cluster in the downdip part of the seismogenic zone, whereas infrequent great earthquakes (M8+) propagate up to the Himalayan frontal thrust. To explore the factors that regulate the bimodal seismicity, we developed a two-dimensional, seismic cycle model of the Nepal Himalaya. Our simulations successfully match the inter-seismic strain and produce a realistic earthquake cycle. Most importantly, we find that bimodal behaviour emerges as a result of relatively higher friction and a non-planar geometry of the Main Himalayan Thrust fault. Our findings strongly support the emerging view that the next large earthquake in Nepal may rupture an area similar to, or significantly larger than the 2015 Gorkha earthquake. This indicates the relevant seismic hazard for the millions of people living in the Indo–Gangetic Basin.

On 25 April 2015, an earthquake with moment magnitude M_W 7.8 struck the Nepal Himalaya[1–3], rupturing a 50-km-wide segment of the Main Himalayan Thrust (MHT) fault (Fig. 1a). The 2015 Gorkha earthquake has a similar location as the 1833 earthquake, with estimated magnitude M_W 7.6–7.7, which also caused significant damage in Kathmandu[4,5]. The geometry of the MHT is relatively well known in their hypocentral region from various geological and geophysical campaigns[6–8]. In particular, geodetic data (SAR, InSAR and GPS) and the detailed location of the Gorkha seismic sequence have provided new constraints on the geometry of the MHT[9,10]. This information allows us to investigate the relation between interseismic strain and seismicity—given the MHT geometry—and contribute to an ongoing debate on how the Himalayan wedge is deforming. Some authors claim that the location of the front of the high topography could be explained by a mid-crustal ramp along the MHT[11–13]. Conversely, others have argued for active out-of-sequence thrusting at the front of the high Himalaya[14,15]. Understanding how and where stresses build up in the Himalaya is important, because evaluating the balance between the interseismic strain accumulation and the elastic strain released during seismic events could potentially improve the seismic hazard assessment in central Nepal following the 2015 earthquake[16].

It has long be noticed that the seismicity in the Himalaya is bimodal[11,17,18]. Large blind earthquakes (M7+ or smaller) tend to cluster and repeatedly rupture the deeper portion of the MHT, whereas sporadic great earthquakes (M8+) completely unzip the entire width of the seismogenic zone (Fig. 1a). These blind ruptures are generally characterised by 10–15 km focal depths and clustered along the front of the Himalaya. They seem to occur in the vicinity of the mid-crustal ramp[11]. The M7.8 Gorkha earthquake is the largest known event in that category. On the other hand, paleoseismological field studies found evidence for surface ruptures at the Himalayan frontal fault (Main Frontal Thrust, MFT), probably associated with great (M8+) events[16,19–22]. The

1934 M8.4 Bihar Nepal[16,22] and the 1950 M8.7 Assam earthquake[23]—the largest intracontinental earthquake ever recorded—probably fall in that category. On the basis of these observations, the mechanism driving bimodal behaviour has been speculated for decades but remains poorly understood. One potential explanation is that the MHT consists of along-dip subsegments that rupture—either independently or jointly with neighbouring segments during larger earthquakes—with a non-periodic or even chaotic behaviour arising from stress transfers. This segmentation may partly be controlled by rheological[24] and geometrical complexities such as local non-planarity[5,25,26]. There is also growing evidence that fault frictional properties are also an influential and perhaps determining factor that affect the spatial extent, size, and timing of megathrust ruptures[27]. Dynamic simulations over multiple earthquake cycles with a linear slip-weakening friction law show that a large event that ruptures the entire fault is preceded by a number of small events with various rupture lengths[28]. These results are in keeping with dynamic modelling of the seismic cycle based on rate-and-state friction, which produce blind partial ruptures even in the case of a planar fault with uniform frictional properties[29]. However, how complete ruptures relate to partial ruptures and the geometry and mechanical properties of the MHT, has not yet been investigated quantitatively. ~~In this study, we show that the combined effect of specific frictional properties and geometry of the Himalayan megathrust can explain the observed bimodal distribution of earthquakes.~~

Here we report a self-consistent hypothesis that combines specific frictional properties and a ramp-flat-ramp geometry of the Himalayan megathrust to explain the observed bimodal distribution of earthquakes. We develop a novel two-dimensional (2D) numerical approach (Methods section) to explore the seismic rupture pattern on the MHT over many earthquake cycles (Fig. 1b). To model the India-Eurasia collisional system, we utilise geological structures constrained from geophysical campaigns[7] and new insights after the Gorkha sequence[1,5,9,10], whereas the

temperature distribution is based on a thermokinematic model derived from thermochronological and thermobarometric data[13] (Fig. S1). Also, the model is kinematically driven using a boundary condition that translates into a convergence rate of 38 mm yr^{-1}. The reference geometry of the MHT (Fig. 1b) is inferred from Elliott *et al.*[9] and denoted as Model EF. It is comprised of three segments to reflect the ramp-flat-ramp geometry: a shallow $\sim 30^{\circ}$ dipping ramp between the surface and 5-km-depth constrained by structural sections; a flat portion with a shallow angle reaching, finally, a steeper mid-crustal ramp[30]. Uncertainties regarding the geometry of the MHT still exist, and relatively gentle variation in geometry have also been proposed[10]. We therefore also perform numerical experiments considering this alternative, smoother fault model (Model DF; Fig. 1c and Supplementary Fig. S2). To test the sensitivity of the model to the fault geometry, we consider a simple planar fault as well (Model PF; Fig. 1c and Supplementary Fig. S2). For each of the three fault geometries adopted, we execute a parameter study of the fault frictional properties by testing values of effective static fault friction (that is, including pore-fluid pressure) between 0.06 and 0.2 (Table S1). This range is consistent with the results of a compilation of previously published data[31]. A detailed description of the numerical technique, model setup, modelling procedure and limitations is given in the Methods section.

Results

Consistency with interseismic deformation

An important goal in Himalayan studies over the past decades has been to refine the Himalayan convergence rate[32,33], because this is responsible for the productivity of Himalayan earthquakes[31,34]. We therefore emulate the observed velocity field by imposing a convergence rate of 38 mm yr^{-1}. The model produces about 19–20 mm/yr of convergence across the Himalaya, a value consistent with the long term geological rate, while the residual convergence rate is dissipated through aseis-

mic creep (that is, stable sliding) along the downdip continuation of the MHT at a temperatures larger than 350°C. Most importantly, the model fits the geodetic measurements of interseismic strain remarkably well (Fig. 2a). All three fault geometries yield predictions in good agreement with uplift rates measured from spirit-levelling[35], inSAR[36], and horizontal velocity measured from GPS[8] (Fig. S3). However, we note that model EF agrees particularly well with the data, in terms of both horizontal and vertical velocities.

The ~~mechanical consistency~~ elastic behaviour of the simulated tectonic regime is shown in Fig. 2b. The mid-crustal ramp operates as a geometric asperity during interseismic periods where elastic strain builds up and accounts for as much as two thirds of the convergence rate. Going deeper, the higher temperature favours the transition from frictionally unstable velocity-weakening behaviour to stable (velocity-strengthening) visco-plastic creep (Fig. S4). Visco-plastic strain rates show a sub-horizontal shear zone in the middle-lower crust, which corresponds to the aseismic creep along the MHT. Distributed viscous deformation also occurs in the vicinity of the kink along the MHT ramp-flat geometry. Another constraint on the simulated tectonic deformation comes from the off-megathrust events. The model shows that anelastic strain off the MHT tends to cluster beneath the topographic front of the Higher Himalaya (Fig. 2c,d). In fact, most of these events concentrate in a narrow zone near the edge of the mid-crustal ramp, ~~and~~ which correlates well with the microseismicity observed over the past decades[11]. This off-megathrust earthquake activity also shows a cut-off beneath the Higher Himalaya, which corresponds to the region where the viscous deformation is dominant and the axes of principal compressional stresses (σ_1) become (sub-)vertical.

Bimodal earthquake behaviour of the reference model

Despite the 2D limitations, the reference model produces a rich earthquake behaviour, similar to

that of natural faults. The spatiotemporal evolution of slip velocity of the reference model shows how coseismic slip events are released on the MHT-fault (Fig. 3a). Although the whole seismogenic zone is interseismically nearly fully locked, most of the simulated earthquakes nucleate and propagate only in the lower edge of the locked Main Himalayan Thrust, whereas only a few events unzip the whole flat-and-ramp system. The largest events tend to have similar size and recur quasi periodically every $\sim$1250 yr. Between them, a range of smaller events occurs, and releases only part of the accumulated strain energy. Using a rupture width–moment magnitude scaling law[37], it returns a moment magnitude of $M_W\sim$7.4–7.8 for partial rupture events (Fig. 3b). Such cluster of differently-sized partial ruptures leads up to a final complete failure of the MHT. These complete ruptures are the largest events with an estimated moment magnitude in the order of $M_W\sim$8.3–8.4 (Fig. 3b).

To understand the physical mechanism behind this behaviour, we investigate the spatiotemporal evolution of the stress and yield strength on the MHT. For example, event E9 (Fig. 3a) ruptures only the lower edge of the seismogenic zone and then event E18 is capable of propagating up to the surface. Our analysis indicates that the partial rupture event E9 nucleates close to the downdip limit of the seismogenic zone, before the mid-crustal ramp, where the stress build-up due to tectonic loading is fastest (Fig. 3c). The rupture propagation causes a local stress drop (Fig. 3d), unzipping only part of the seismogenic zone as it is stopped as a result of a large initial strength excess—that is, difference between stress and yield strength. For this event, we further estimate the slip resulting from the occurrence of such rupture. Our results indicate that event E9 produces $\sim$5–6 m of coseismic slip (Fig. 3e), mainly on the deeper flat portion of the MHT, between 10 and 15 km depth.

When only the downdip edge of the locked zone is unzipped, stress is transferred to the

neighbouring updip region by static stress change. Then a new downdip event nucleate sooner than expected from the average recurrence periods, with the next rupture being generally larger than the previous one. This occurs because the strength excess decreases in the frontal part of the MHT, as a result of the stress transfer and the ongoing tectonic loading. Consequently, partial ruptures contribute significantly to build up the stress state on the updip limit of the MHT to a critical level, as for example before event E18 (Fig. 3f). Thus, a complete event eventually propagates through the whole ramp-flat-ramp fault system and leads to a large stress drop (Fig. 3g). These complete ruptures results in a slip larger than 8 m (Fig. 3h), consistent with estimates from paleoseismic investigations[19,20,38]. Then a new cycle of partial ruptures begins, with an initial period of quiescence or small event activity. This is exactly what our model shows in Fig. 3: temporal evolution of the MHT displays a bimodal seismicity-dominated regime. Notably, rupture events are triggered by stress build-up near the downdip end of the locked fault zone, as is observed in nature[39]. Also, the model reproduces a realistic earthquake sequence of irregular moment magnitude main shocks, including events similar to the 2015 Gorkha earthquake. A simulation example is shown in Supplementary Movie S1.

For each numerical simulation, we determine the rupture styles of all events, which include pulse- and crack-like ruptures (Fig. S5). Pulse-like ruptures are characterised by the existence of significant healing during the rupture propagation[40,41]. In contrast, in a crack-like event slip is accommodated on the fault even when the rupture has reached the surface. Our results indicate that complete events are crack-like ruptures, and account for most of the stress transfer to the front of the thrust-fault system. Instead, a partial event usually begin as crack-like rupture and then turns into a bilateral (that is, both up- and downward propagation) pulse-like rupture until the end of the event (Fig. S5). The peak of slip occurs approximately when the rupture starts expanding along-

dip. It is noteworthy that these results are consistent with the 2015 Gorkha earthquake. The 2015 main-shock clearly indicate a pulse-like rupture with slip on any given portion of the fault occurring over a short fraction of the total $\sim$70 s duration of the earthquake source[2].

A particular feature of the Himalayan wedge is the seismic–aseismic transition zone, which seems to coincide with the mid-crustal ramp beneath the front of the high Himalaya[8,11]. However, the feedback between the geometry and the rheological behaviour of the mid-crustal ramp are difficult to ascertain on the basis of natural data alone. When a rupture occurs in our simulations, it generally expands upward from the locked edge, but not much downwards. This occurs because the zone of aseismic slip acts as an efficient barrier to downdip propagation of ruptures. This self-consistent feature of our models is interpreted as an effect of the temperature increase with depth, which in turn decreases the viscosity of rocks. Also, our models show that all hypocentre locations fall in a narrow zone near the edge of the mid-crustal ramp (Fig. 3a), indicating a pivotal role of this crustal asperity in localising the strain both on and off the megathrust (Fig. 2b). Thus, our results suggest that both the geometric-structural and the thermal-rheological strength of the mid-crustal ramp control the downdip rupture width on the MHT.

Quantifying the effects of fault friction and geometry on seismic ruptures

To explore the conditions that explain this bimodal seismicity, we next analyse the effect that frictional properties and geometry of the MHT have on the resulting pattern of large earthquakes. We first explore the parameter space in terms of static fault friction (μ_s), and the maximum friction drop from static to dynamic friction coefficient ($\gamma = 1 - \mu_d/\mu_s$). To elucidate the observed bimodal seismicity, we further analyse the median S parameter, which represents the ratio between the average strength excess before an event and the average coseismic static stress drop[41]. Our model produces distinctly different rupture patterns within a narrow range of frictional parameters (Fig.

4). In fact, an increase of both the static fault friction and friction drop leads to an increase of (i) the number of events per cycle (Fig. 4a), (ii) the average recurrence interval between the largest events (Fig. 4b), and (iii) the median S values (Fig. 4c). As illustrated in Fig. 4, this corresponds to a transition from ordinary (unimodal) cycles to irregular cycles, which display a bimodal seismicity.

The bimodal seismicity-dominated regime model clearly shows the evolution from low stress drop, high strength excess pulse-like events into a few high stress drop, low strength excess crack-like large events, which leads to a high median S value. In contrast, the spatiotemporal evolution of the model with a lower static fault friction (μ_s=0.1) shows a more ordinary recurrence pattern of quasi-periodic large events (Fig. S6). These events mostly nucleate near the edge of the mid-crustal ramp, grow large and propagate both up- and downward, activating the whole flat-and-ramp system. Evolving ruptures readily break the entire locked zone of the MHT in a crack-like style, and lead to significant stress drops. Consequently, this model is related to a low median S value (Fig. 4c).

For each of the three fault geometries adopted, we further analyse the relation between the S parameter and rupture width of all events when a bimodal seismicity pattern is observed (μ_s=0.16 and γ=0.7; Fig. 5a–c). Results from the reference model EF (Fig. 5a) indicate that the S parameter decreases with increasing rupture width. Most importantly, we find that this ramp-flat-ramp geometry results in a rupture-width gap between 60–65 km and 90–95 km. A very similar trend is also observed in Model DF (Fig. 5a). Pulse-like partial ruptures are confined to a critical width of 60–65 km, whereas large crack-like events propagate through the whole seismogenic zone. Consequently, models EF and DF result in a bimodal distribution of rupture widths. On the other hand, results from the simple planar fault (Fig. 5c) indicate that the S parameter decreases linearly with increasing rupture width. This means that the larger the event, the higher the stress

197 released and the resulting S parameter is lower. Although this model displays a wide spectrum of
198 rupture widths, the general pattern does not indicate any bimodal distribution.

199 To examine the role of the MHT geometry, we further analyse the average downdip stress
200 vs. strength distribution for the three fault geometries adopted (Fig 4g-i). In general, these profiles
201 suggest that, the steeper the fault dips in the updip region of the MHT, the higher would be the
202 pressure-dependent fault strength. This, together with a relatively higher fault friction, increases
203 the fault strength even further. Consequently, the strength excess also increases, and a higher
204 pre-stress is thereby necessary to reach a critical level at which eventually a crack-like event
205 ruptures the entire megathrust (Fig. S7). As in the case of model EF (Fig. 5d), and even more
206 clearly on model DF (Fig. 5e), the strength excess in the shallower region of the MHT is notably
207 high. This behaviour arises because when the model accounts for a ramp-and-flat fault geometry,
208 the far-field tectonic loading is not fast enough to bring the pre-stress up to a critical state in the
209 upper edge of the MHT. Most of the simulated earthquakes are thus capable of rupturing only a
210 fraction of the seismogenic zone. Then, the static stress distribution left over from these previous
211 partial ruptures contribute significantly to increase the stress state in the updip limit of the MHT.
212 On the other hand, the planar fault geometry (Model PF) maintains a relatively low strength excess
213 throughout the seismogenic zone (Fig. 5f), thereby allowing the propagation of frequent complete
214 ruptures.

215 ## Discussion

216 ### ~~Seismic gaps versus failures: implications for seismic hazard~~

217 Our simulations show that it is probably incorrect to assume that relatively large earthquakes (M7+)
218 known to have occurred along the Himalayan front over history[42] are representative of the greatest
219 possible earthquakes. In our model, the same segment of the MHT can in principle produce

220 a sequence of partial ruptures similar to the Gorkha earthquake and occasionally much larger
221 events, similar to the 1934 M8.4 event or even larger. This is confirmed by moment conservation
222 calculations at the scale of the Himalayan arc, which require M$\sim$9 earthquakes with a 1000 yr
223 return period[43]. Our models indicate that a great earthquake (M8+) can occur at the same location
224 as a M7+ earthquake, and that it may strike sooner than would be anticipated from considerations
225 of renewal time from plate convergence rates. While we cannot rule out the plausible presence
226 of along-strike heterogeneities given the lack of the third-dimension, our models show that the
227 combined effects of fault geometry and frictional properties in controlling the along-dip bimodal
228 behaviour of the MHT could potentially hold for the entire Himalayan arc. In support of this claim,
229 recent pattern of interseismic coupling on the MHT along the entire Himalayan arc do not indicate
230 any aseismic barrier that could affect the seismic segmentation of the arc and limit the along-strike
231 propagation of seismic ruptures[44] (Fig 1a).

232 For a finite range of static fault friction (μ_s=0.06–0.2), our model exhibits a large spectrum
233 (250–1500 yr) of recurrence time of great earthquakes. It also shows that an indication for the
234 temporal proximity of such a M8+ earthquake can come from the maximum updip limit of the prior,
235 partial earthquake, which provides an indication for a likely critically stressed MHT. Our results
236 indicate that an average recurrence time of $\sim$600 yr leads to coseismic slip of 8–10 m in order to
237 release the elastic strain accumulated during such interseismic periods. However, partial ruptures
238 account only for an average slip of 4–6 m, in agreement with the average slip of moderate (M7+)
239 Himalayan earthquakes such as the Gorkha earthquake[9].

240 Finally, it appears that the static stress change due to partial ruptures is the major factor
241 introducing irregularity in the seismic cycle. This is the main reason that could explain why the
242 model obeys neither the slip- nor time-predictable behaviour at any given point on the fault (Fig.

S8), since it does not incorporate a fixed threshold shear stress for slip to occur. This is because, after each earthquake, the stress on the ruptured area drops to a low level, approximately determined by the rate-dependent friction formulation evaluated at the coseismic slip rate.

~~Conclusion~~

~~We developed a new numerical model that incorporates numerous geological and geophysical constraints on the Nepal Himalaya and the MHT and reconciles a suite of independent interseismic and coseismic observations.~~

To summarise, this seismo-thermo-mechanical model constrained by observations provides physical explanations to understand the behaviour of the seismic cycle in the Himalaya. It shows that frictional properties and non-planar geometry of the MHT control a variety of observations of the MHT behaviour, such as the along-dip stress conditions, the bimodal seismicity, the relative persistence of along-dip variations of seismic ruptures and the variable recurrence time of large (M7+) and great (M8+) earthquakes. Based on our numerical experiments, we postulate that large crack-like earthquakes on the MHT may incorporate and release a heterogeneous historical reservoir of mid-décollement stress inherited from former pulse-like partial ruptures. These very large events account for the bulk of the deformation that is transferred to the most frontal structures in the Sub-Himalaya. If this mechanism is in fact correct, it has consequences for the assessment of seismic moment where only rupture length and surface slip are known, as is the case for all palaeoseismic ruptures inferred from slip on the MFT[16,19,20,22]. Because a heterogeneous along-dip stress condition is likely to prevail throughout the Himalayan arc, our results may provide an answer to the long-standing difficulties in explaining the source of the stored stresses needed to drive large ($>$8–10 m) paleoseismic surface ruptures recorded on the MFT.

In light of our modelling results, the updip arrest of the 2015 Gorkha earthquake calls for special attention, as the nearly 800-km-long stretch between the 1833/2015 ruptures and the 1905 M_W 7.8 Kangra earthquake is a well-identified seismic gap with no large earthquake for over 500 years[1]. The MHT is clearly locked there[8,44] and its deficit of slip may exceed $\sim$10 m. The last large earthquake in that area occurred in 1505, and could have exceeded M_W 8.5 (ref. 45). Our results suggest that the next great earthquake in southern Tibet may rupture an area significantly greater than the section from the Gorkha earthquake, and that it is prudent to consider that the region to the west of Kathmandu is well along in a elastic strain accumulation cycle prior to a great earthquake, most likely much greater than the one occurred in 2015. Continued geodetic monitoring of the Himalayan arc in the coming years will help to provide new constraints and to ascertain these speculations.

References

1. Avouac, J.-P., Meng, L., Wei, S., Wang, T. & Ampuero, J.-P. Lower edge of locked main himalayan thrust unzipped by the 2015 gorkha earthquake. *Nature Geoscience* (2015).

2. Galetzka, J. *et al.* Slip pulse and resonance of the kathmandu basin during the 2015 gorkha earthquake, nepal. *Science* **349**, 1091–1095 (2015).

3. Wang, K. & Fialko, Y. Slip model of the 2015 mw 7.8 gorkha (nepal) earthquake from inversions of alos-2 and gps data. *Geophysical Research Letters* **42**, 7452–7458 (2015).

4. Bilham, R. Location and magnitude of the 1833 nepal earthquake and its relation to the rupture zones of contiguous great himalayan earthquakes. *Current Science* **69**, 101–128 (1995).

5. Mugnier, J.-L. *et al.* Segmentation of the himalayan megathrust around the gorkha earthquake (25 april 2015) in nepal. *Journal of Asian Earth Sciences* (2017).

6. Lavé, J. & Avouac, J.-P. Active folding of fluvial terraces across the siwaliks hills, himalayas of central nepal. *Journal of Geophysical Research* **105**, 5735–5770 (2000).

7. Nábělek, J. *et al.* Underplating in the himalaya-tibet collision zone revealed by the hi-climb experiment. *Science* **325**, 1371–1374 (2009).

8. Ader, T. *et al.* Convergence rate across the nepal himalaya and interseismic coupling on the main himalayan thrust: Implications for seismic hazard. *Journal of Geophysical Research: Solid Earth (1978–2012)* **117** (2012).

9. Elliott, J. *et al.* Himalayan megathrust geometry and relation to topography revealed by the gorkha earthquake. *Nature Geoscience* (2016).

10. Duputel, Z. *et al.* The 2015 gorkha earthquake: A large event illuminating the main himalayan thrust fault. *Geophysical Research Letters* **43**, 2517–2525 (2016).

11. Pandey, M., Tandukar, R., Avouac, J., Lave, J. & Massot, J. Interseismic strain accumulation on the himalayan crustal ramp (nepal). *Geophysical Research Letters* **22**, 751–754 (1995).

12. Lavé, J. & Avouac, J. Fluvial incision and tectonic uplift across the himalayas of central nepal. *Journal of Geophysical Research: Solid Earth* **106**, 26561–26591 (2001).

13. Herman, F. *et al.* Exhumation, crustal deformation, and thermal structure of the nepal himalaya derived from the inversion of thermochronological and thermobarometric data and modeling of the topography. *Journal of Geophysical Research: Solid Earth* **115** (2010).

14. Wobus, C., Heimsath, A., Whipple, K. & Hodges, K. Active out-of-sequence thrust faulting in the central nepalese himalaya. *Nature* **434**, 1008–1011 (2005).

15. Whipple, K. X., Shirzaei, M., Hodges, K. V. & Arrowsmith, J. R. Active shortening within the himalayan orogenic wedge implied by the 2015 gorkha earthquake. *Nature Geoscience* **9**, 711–716 (2016).

16. Bollinger, L., Tapponnier, P., Sapkota, S. & Klinger, Y. Slip deficit in central nepal: omen for a repeat of the 1344 ad earthquake? *Earth, Planets and Space* **68**, 12 (2016).

17. Ni, J. & Barazangi, M. Seismotectonics of the himalayan collision zone: Geometry of the underthrusting indian plate beneath the himalaya. *Journal of Geophysical Research: Solid Earth* **89**, 1147–1163 (1984).

18. Baranowski, J., Armbruster, J., Seeber, L. & Molnar, P. Focal depths and fault plane solutions of earthquakes and active tectonics of the himalaya. *Journal of Geophysical Research: Solid Earth* **89**, 6918–6928 (1984).

19. Lavé, J. *et al.* Evidence for a great medieval earthquake ($\sim$1100 ad) in the central himalayas, nepal. *Science* **307**, 1302–1305 (2005).

20. Kumar, S. *et al.* Paleoseismological evidence of surface faulting along the northeastern himalayan front, india: Timing, size, and spatial extent of great earthquakes. *Journal of Geophysical Research: Solid Earth* **115** (2010).

21. Hetényi, G. *et al.* Joint approach combining damage and paleoseismology observations constrains the 1714 ad bhutan earthquake at magnitude 8$\pm$0.5. *Geophysical Research Letters* **43** (2016).

22. Sapkota, S. *et al.* Primary surface ruptures of the great himalayan earthquakes in 1934 and 1255. *Nature Geoscience* **6**, 71–76 (2013).

23. Chen, W.-P. & Molnar, P. Seismic moments of major earthquakes and the average rate of slip in central asia. *Journal of Geophysical Research* **82**, 2945–2969 (1977).

24. Huc, M., Hassani, R. & Chéry, J. Large earthquake nucleation associated with stress exchange between middle and upper crust. *Geophysical research letters* **25**, 551–554 (1998).

25. Hubbard, J. *et al.* Structural segmentation controlled the 2015 mw 7.8 gorkha earthquake rupture in nepal. *Geology* **44**, 639–642 (2016).

26. Qiu, Q. *et al.* The mechanism of partial rupture of a locked megathrust: The role of fault morphology. *Geology* **44**, 875–878 (2016).

27. Kaneko, Y., Avouac, J.-P. & Lapusta, N. Towards inferring earthquake patterns from geodetic observations of interseismic coupling. *Nature Geoscience* **3**, 363–369 (2010).

28. Duan, B. & Oglesby, D. D. The dynamics of thrust and normal faults over multiple earthquake cycles: Effects of dipping fault geometry. *Bulletin of the Seismological Society of America* **95**, 1623–1636 (2005).

29. Michel, S., Avouac, J.-P., Lapusta, N. & Jiang, J. Pulse-like partial ruptures and high-frequency radiation at creeping-locked transition during megathrust earthquakes. *Geophysical Research Letters* **44**, 8345–8351 (2017).

30. Cattin, R. & Avouac, J. Modeling mountain building and the seismic cycle in the himalaya of nepal. *Journal of Geophysical Research* **105**, 13389–13407 (2000).

31. Avouac, J. Mountain building: From earthquakes to geologic deformation. *Treatise on Geophysics,* **6**, 381–432 (2015).

32. Bilham, R. *et al.* Gps measurements of present-day convergence across the nepal himalaya. *Nature* **386**, 61–64 (1997).

33. Bettinelli, P. *et al.* Plate motion of india and interseismic strain in the nepal himalaya from gps and doris measurements. *Journal of Geodesy* **80**, 567–589 (2006).

34. Bilham, R., Gaur, V. K. & Molnar, P. Himalayan seismic hazard. *Science* **293**, 1442–1444 (2001).

35. Jackson, M. & Bilham, R. Constraints on himalayan deformation inferred from vertical velocity fields in nepal and tibet. *Journal of Geophysical Research* **99**, 13–897 (1994).

36. Grandin, R. *et al.* Long-term growth of the himalaya inferred from interseismic insar measurement. *Geology* **40**, 1059–1062 (2012).

37. Wells, D. L. & Coppersmith, K. J. New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement. *Bulletin of the seismological Society of America* **84**, 974–1002 (1994).

38. Bollinger, L., Henry, P. & Avouac, J. Mountain building in the nepal himalaya: Thermal and kinematic model. *Earth and Planetary Science Letters* **244**, 58–71 (2006).

39. Bollinger, L., Avouac, J., Cattin, R. & Pandey, M. Stress buildup in the himalaya. *Journal of Geophysical Research: Solid Earth (1978–2012)* **109** (2004).

40. Heaton, T. H. Evidence for and implications of self-healing pulses of slip in earthquake rupture. *Physics of the Earth and Planetary Interiors* **64**, 1–20 (1990).

41. Herrendörfer, R., Van Dinther, Y., Gerya, T. & Dalguer, L. A. Earthquake supercycle in subduction zones controlled by the width of the seismogenic zone. *Nature Geoscience* **8**, 471–474 (2015).

42. Bilham, R. Earthquakes in india and the himalaya: tectonics, geodesy and history. *Annals of Geophysics* **47** (2004).

43. Stevens, V. & Avouac, J.-P. Millenary mw>9.0 earthquakes required by geodetic strain in the himalaya. *Geophysical Research Letters* **43**, 1118–1123 (2016).

44. Stevens, V. & Avouac, J. Interseismic coupling on the main himalayan thrust. *Geophysical Research Letters* **42**, 5828–5837 (2015).

45. Bilham, R. & Wallace, K. Future mw 8 earthquakes in the himalaya: implications from the 26 dec 2004 mw= 9.0 earthquake on india's eastern plate margin. *Geol. Surv. India Spec. Publ* **85**, 1–14 (2005).

Acknowledgements

We thank to J.-P. Ampuero and the STM-group for comments. V.L. Stevens kindly provided the coupling data. We are grateful to J. Singer for providing us with a basic GMT script to plot Fig. 1a. Numerical simulations were performed on ETH cluster Euler. This study was funded by the SNF 2-77090-14 project Swiss-AlpArray SINERGIA.

Author Contributions

L.D.Z. designed the study and model setup, carried out and analysed the numerical experiments and wrote the paper. Y.v.D. and T.V.G. developed the STM methodology and analysed the results. J.P.A. contributed to the concept development. All authors discussed the results and commented on the paper.

Competing Interests

The authors declare that they have no competing financial interests.

Correspondence

Correspondence and requests for materials should be addressed to L.D.Z.

(*luca.dalzilio@erdw.ethz.ch*).

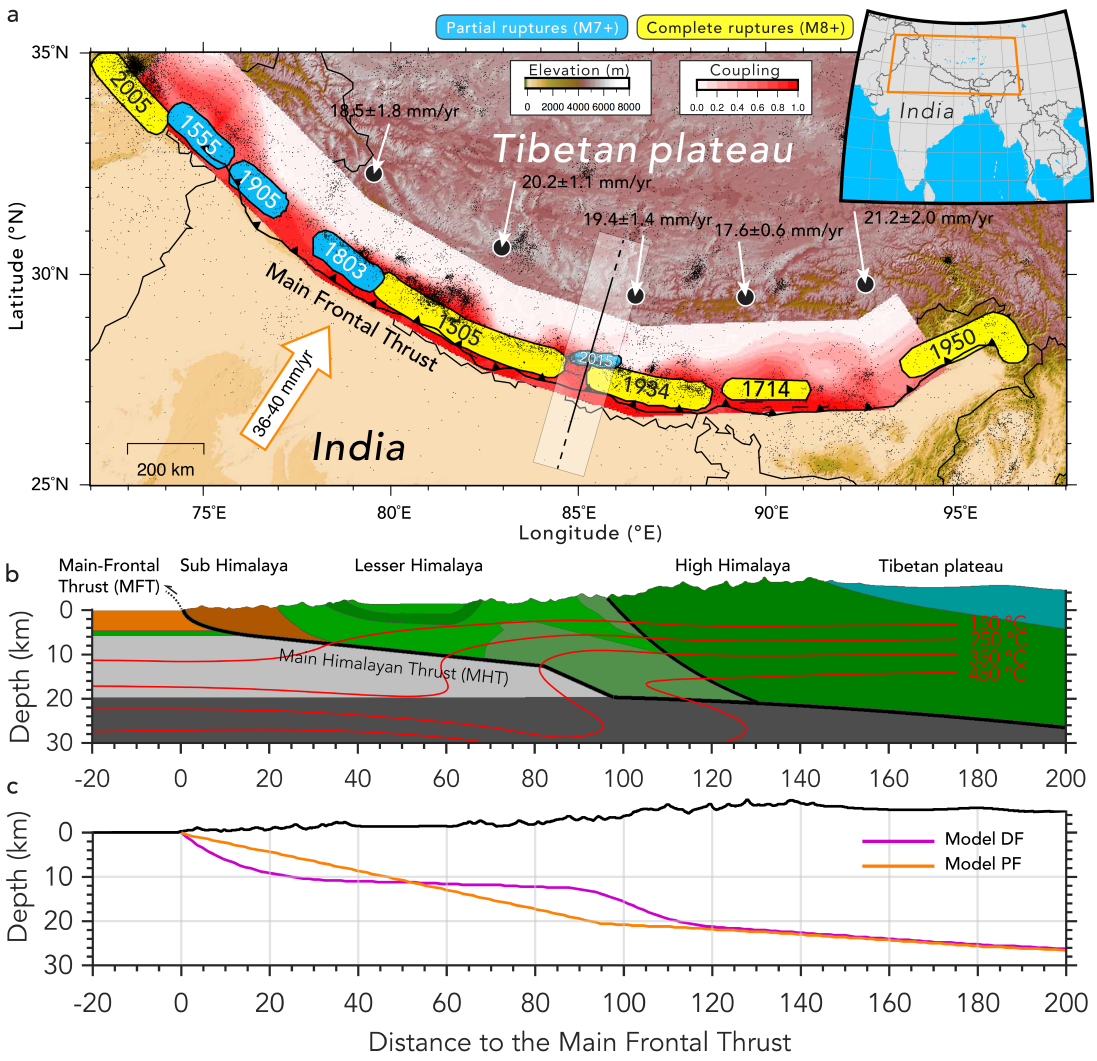

Figure 1: Seismotectonic context, model setup, and fault geometries. **a**, Topographic relief, coupling mode, and historical seismicity. The white arrows show the long-term shortening across the arc. The interseismic coupling is shown as shades of red (ref. 44). A coupling value of 1 means the area is fully locked, while a value of 0 means fully creeping. Coloured patches indicate the supposed rupture zones since 1505 (refs 4, 21, 22): blue patches display blind ruptures of large (M7+) earthquakes, whereas yellow patches indicate surface ruptures of great (M8+) events. Black line indicates the cross-section utilised for the numerical model setup. **b**, Zoom of the initial reference setup (model EF) and temperature. The numerical setup represents the geological cross section of the Nepal Himalaya constrained from the mainshock and aftershocks of the Gorkha sequence (ref. 9). **c**, Additional MHT geometries employed in the numerical experiments: model DF, from Duputel *et al.* (ref. 10), and a planar fault geometry (Model PF).

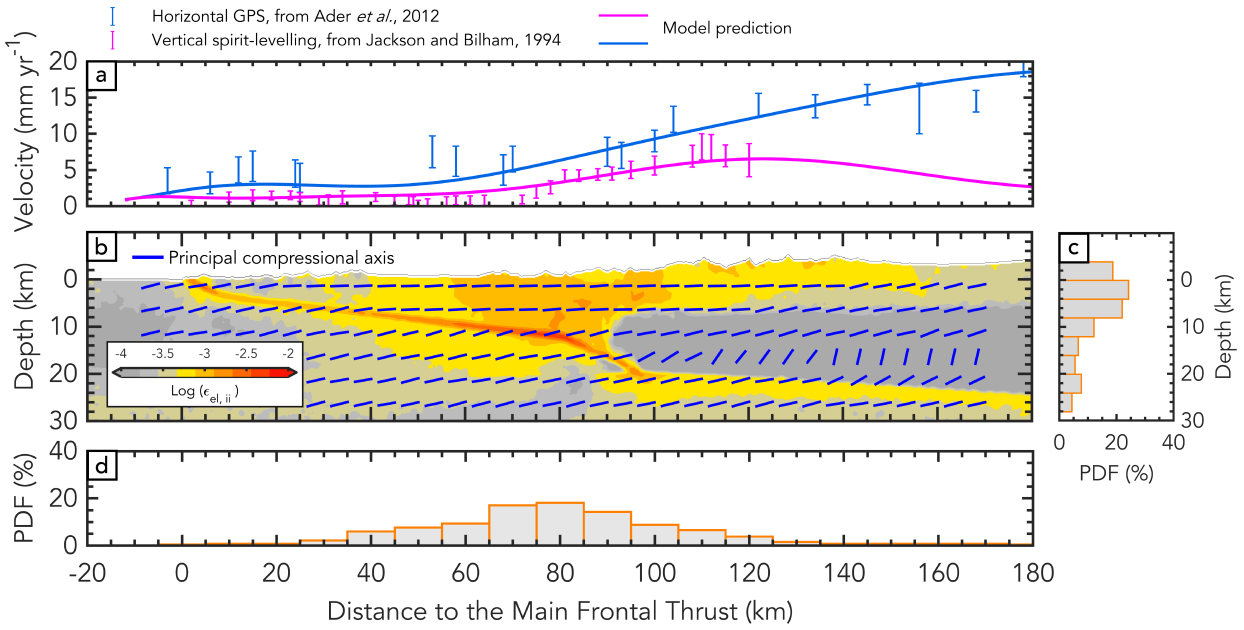

Figure 2: Interseismic behaviour computed in the 2D model. **a**, Observed vs. synthetic present-day velocity fields. Observed field is shown in blue (horizontal GPS, ref. 8) and violet (spirit-levelling, ref. 35) bars, respectively. Solid lines show the corresponding horizontal and vertical modelling prediction. **b**, Elastic strain regime across the Himalaya inferred over an interseismic period of 350 yr and orientation of principal compressional axes (blue bars). Histograms in (**c**) and (**d**) show the vertical and horizontal off-megathrust faulting distributions, respectively.

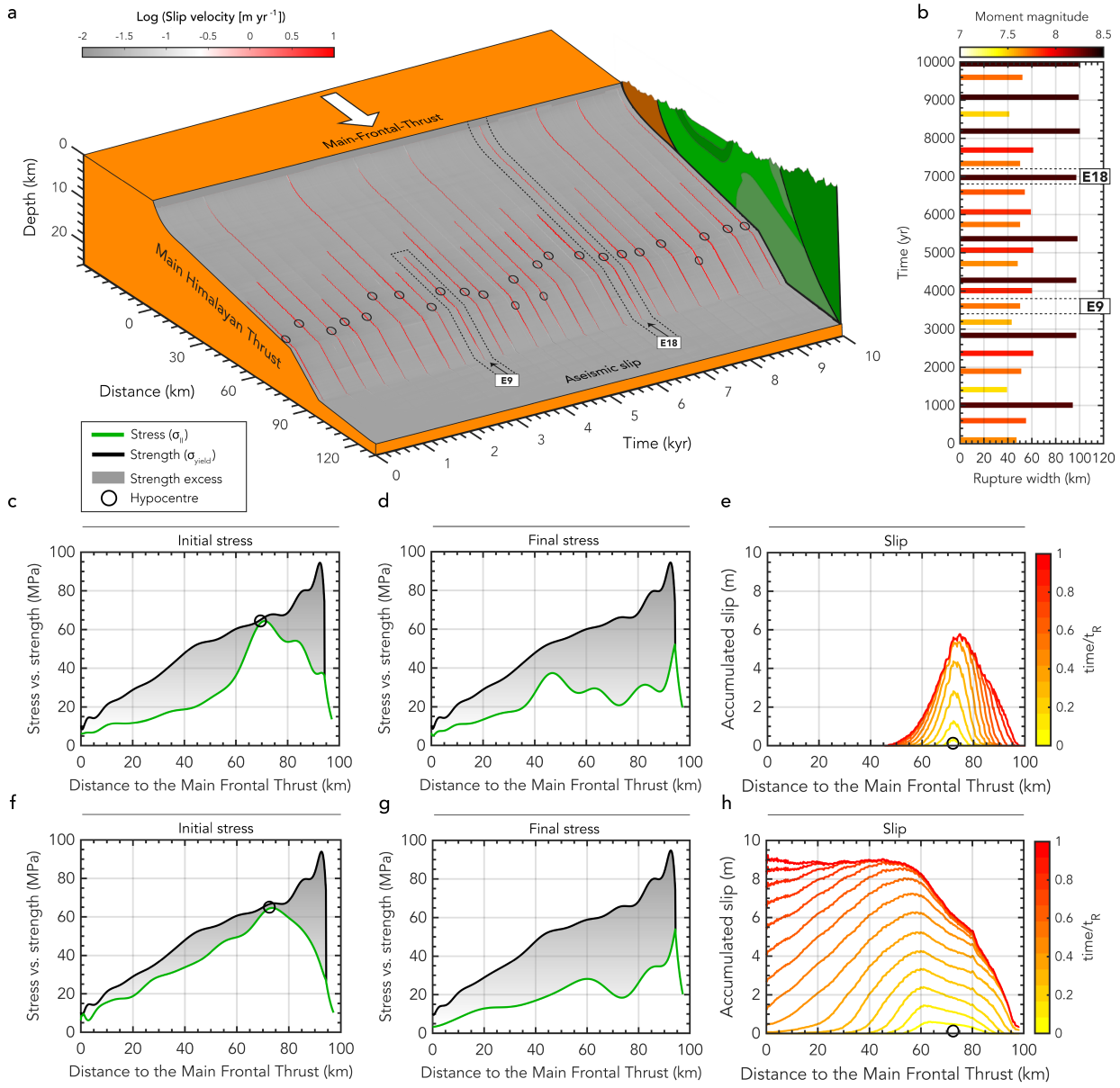

Figure 3: Megathrust behaviour computed in the 2D model (EF) over 10,000 yr. **a**, Spatiotemporal evolution of slip on the MHT for the reference model. Red lines show slip during the simulated earthquakes. Note that hypocenters (black circles) are typically located in the lower edge of the flat segment, just before the mid-crustal ramp. **b**, Time evolution of downdip rupture width. Colorbar indicates the corresponding moment magnitude. **c-d**, Along megathrust profiles of initial (**c**) and final (**d**) stress vs. strength for the partial rupture event E9. **e**, Contours of accumulated coseismic slip throughout event E9. **f-g**, Along megathrust profiles initial (**f**) and final (**g**) stress vs. strength for the complete rupture event E18. **h**, Contours of accumulated coseismic slip throughout event E18.

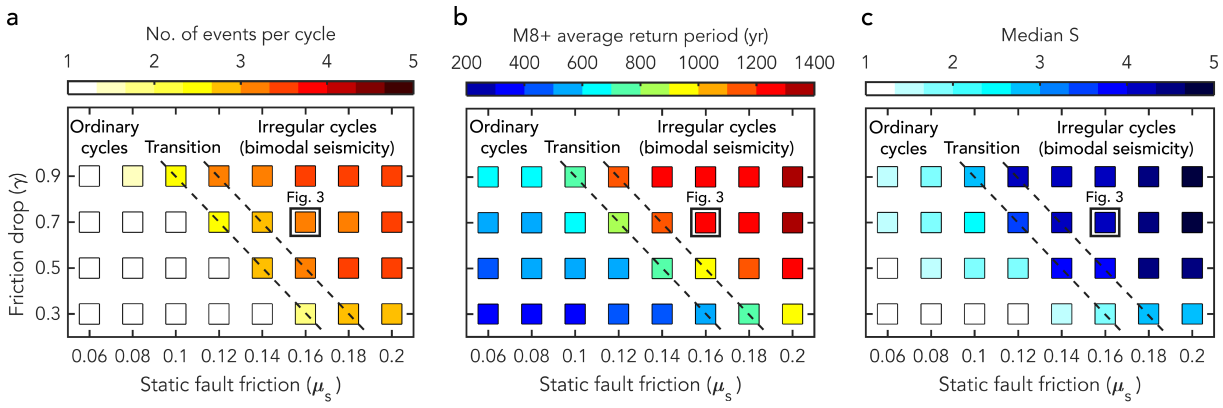

Figure 4: Effect of frictional properties on the seismic behaviour of model EF. Average (**a**) number of events per cycle, (**b**) recurrence time of complete ruptures (M8+ events), and (**c**) median of the S parameter. Dashed black lines indicate the transition from ordinary cycles to irregular cycles (bimodal seismicity).

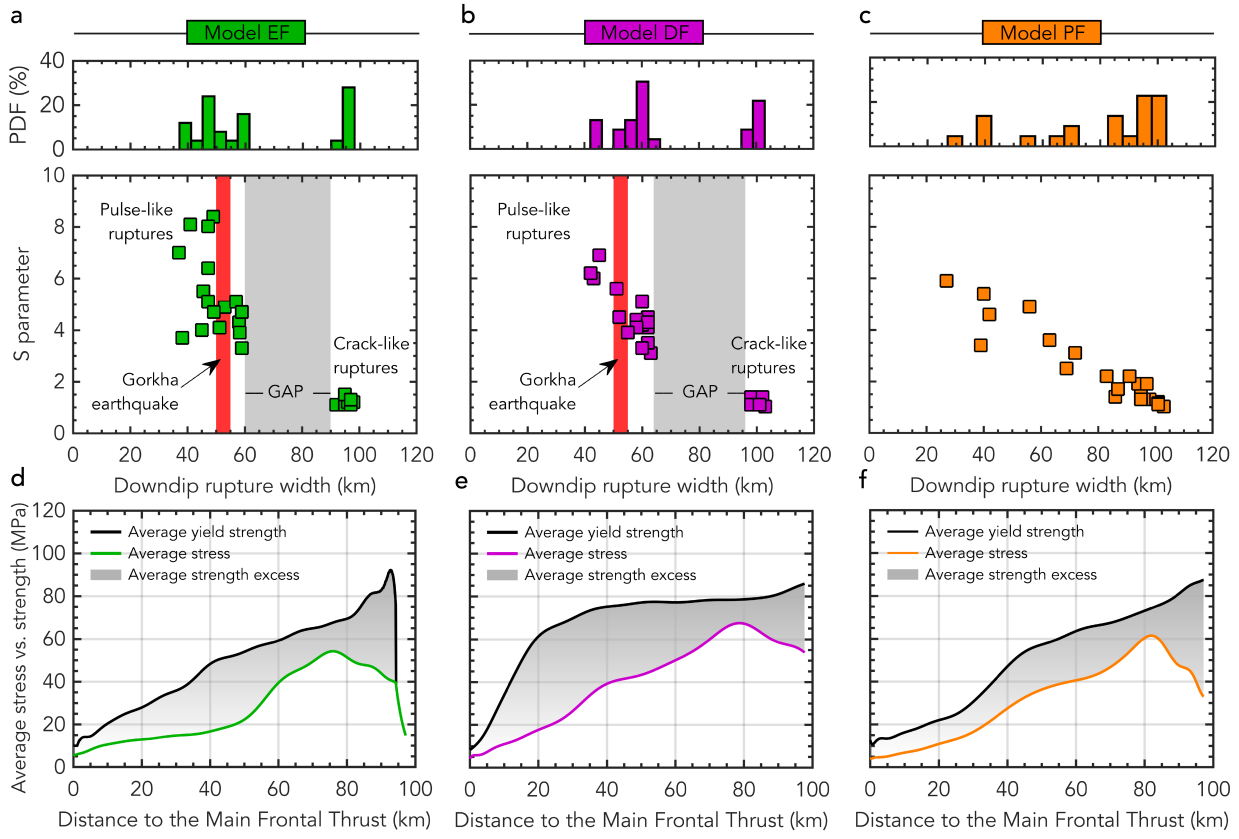

Figure 5: Impact of the three fault geometries on the rupture patterns. Relationship between S parameter and rupture width for models adopting a realistic ramp-flat-ramp fault geometry inferred from Elliott *et al.* (ref. 9) (**a**), Duputel *et al.* (ref. 10) (**b**), which also indicate the dominance of different rupture styles (pulse- vs. crack-like ruptures), and a planar fault geometry (**c**). **d-f**, Along megathrust profiles of the average stress vs. strength for the three fault geometries adopted: model EF (**d**), DF (**e**), and PF (**f**).