Earthquake Shaking History and Probabilistic Seismic Hazard at the Agham Spur, Ladakh

Yash Singh

Developer and Mathematician

Iron.ws

iron.ws@proton.me

19 November 2026

Abstract

We assess the earthquake shaking that a building on the Agham spur, Ladakh (34°19′01″N, 77°47′09″E), should be able to withstand. An instrumental catalogue of 18,986 records (1902–2026) and the NOAA significant-earthquake database (211 events, 250 BCE to 2006) were corrected and merged into 19,045 distinct earthquakes. They are used to (i) reconstruct the shaking the site has experienced, (ii) select the 47 earthquakes relevant to design, and (iii) build a screening-level probabilistic seismic hazard analysis (PSHA) with a smoothed-seismicity background and a separate Karakoram Fault source. Among the events in the assembled dataset, the largest estimated median peak ground acceleration (PGA) at the site is about 0.05 g. That is a poor guide to the future, because the site is assumed to lie about 1.6 km from a strand of the Karakoram Fault, which has produced no large instrumental earthquake. If a strand lies there, a characteristic M7.3 rupture would give a median PGA of 0.46 g (84th percentile 0.84 g), about nine times the largest shaking in the dataset; the fault proximity is an assumption, and all fault-related results are conditional on it. Hazard at long return periods therefore depends mainly on whether, and how fast, the strand next to the site slips. We treat this with two explicit hypotheses: A, active at 2–4 mm/yr, and B, inactive or very slow (0–1 mm/yr). The blended mean rock PGA is 0.043, 0.17, 0.33 and 0.61 g at return periods of 72, 475, 975 and 2,475 years; under A the 475-, 975- and 2,475-year values are 0.20, 0.41 and 0.70 g, and under B 0.13, 0.20 and 0.36 g. The fault supplies 8% of the exceedance rate at 72 years, half at 475 years and 86% at 2,475 years. Catalogue magnitudes are the largest remaining statistical uncertainty: a recurrence fit restricted to M≥5.5, which is dominated by moment magnitudes, raises the 475-year PGA from 0.17 to 0.23 g. The results are provisional rock-site PGA values that rest on a single ground-motion model and an unmapped fault strand; they are not design parameters.

Contents

1 Introduction **4**

2 Data **4**

2.1 The instrumental catalogue 4

2.2 The NOAA significant-earthquake database 4

2.3 Merging and removing double counts 4

3 Seismicity and shaking history **5**

3.1 Seismicity around the site . . . 5

3.2 Shaking actually experienced . . . 5

3.3 Completeness of the record . . . 6

4 Ground-motion model **7**

5 Deterministic scenario: a Karakoram Fault earthquake **9**

5.1 Landslide screening (Newmark sliding block) . . . 9

6 Earthquakes relevant to design **10**

7 The return-period model **11**

7.1 Definitions and the Poisson assumption . . . 11

7.2 Source zone and declustering . . . 12

7.3 Magnitude-dependent completeness . . . 12

7.4 Gutenberg–Richter recurrence by maximum likelihood . . . 12

7.5 Hazard magnitude bins and maximum magnitude . . . 13

7.6 Spatial distribution: smoothed seismicity . . . 14

7.7 The Karakoram Fault source . . . 14

7.8 Evidence for fault slip at the site . . . 15

7.9 The hazard integral . . . 15

7.10 Logic tree . . . 16

7.11 Fault-activity hypotheses and the governing envelope . . . 16

7.12 Forecasting earthquakes and shaking in the building's life . . . 17

7.13 Disaggregation . . . 17

7.14 Back-test . . . 17

8 Results **18**

8.1 Return periods of ground shaking at the site . . . 18

8.2 Hazard by fault-activity hypothesis . . . 19

8.3 The shape of the hazard curve . . . 19

8.4 Disaggregation . . . 19

8.5 Probability of shaking in the building's life . . . 20

8.6 Probability of earthquakes in the region . . . 20

8.7 Sensitivity tests . . . 21

8.8 Slope screening with hazard-based shaking . . . 23

9 Discussion: implications for the building **23**

10 Limitations **24**

11 Conclusions **25**

A The 47 design-relevant earthquakes **28**

1 Introduction

Designing a long-lived building in a region with a short instrumental record requires three things: knowing what has shaken the site before, knowing which past earthquakes matter for design, and estimating how often future shaking of each severity will occur. This paper addresses all three for the Agham spur in the Nubra–Shyok sector of Ladakh, close to the Karakoram Fault.

The study proceeds in stages. The catalogues are cleaned and merged (Section 2); the shaking history is reconstructed with a ground-motion model (Sections 3 and 4); a deterministic fault scenario and a landslide screening follow (Section 5); design-relevant earthquakes are selected (Section 6); and a probabilistic model forecasts return periods of earthquakes and of ground shaking (Section 7), with results in Section 8. Implications, limitations and conclusions are given in Sections 9–11.

Status of the results. Fault-related results are conditional on the assumed proximity and activity of a fault strand next to the site, which has not been mapped or dated at site scale. This is a screening-level analysis of peak ground acceleration on reference rock. It is not a substitute for a site-specific hazard study, a geotechnical investigation, or the governing building code. Section 10 lists what is not modelled.

2 Data

2.1 The instrumental catalogue

The catalogue file contains 18,986 rows with unique identifiers, spanning 22 August 1902 to 8 April 2026, magnitudes M4.0–8.0 and epicentral distances from 25 to 1,694 km from the site. Twenty-three rows are typed "nuclear explosion" and were removed, leaving 18,963 earthquakes. Magnitude types are mixed (17,012 mb; about 1,850 moment-type magnitudes such as mw, mww, mwc, mwr and mwb; 102 Ms; a few others). All are treated as moment magnitude, an approximation whose consequences are examined in Section 8.7.

2.2 The NOAA significant-earthquake database

The NOAA file holds 211 events, with years from −250 to 2006, all with coordinates. Twenty-one have no `eqMagnitude`; these were filled from the Mw, Ms, Mb and unknown-magnitude fields where present. Three events within the ground-motion model's range still had no magnitude (1885 Kashmir, 1828 Kashmir, 1827 Lahore) and were scored at an assumed bounding magnitude of M7.5, so their results are upper bounds. The database lists only events that caused damage, deaths or other notable effects; it is not a complete catalogue and is not used to estimate recurrence rates.

2.3 Merging and removing double counts

Most NOAA events after 1902 are also in the catalogue. Each NOAA event after 1902 was matched to a catalogue event if it fell within a time window (2 days when the day was known, about ±17 days when only the month was known, about ±6 months when only the year was known), within 200 km, and within 1.5 magnitude units. Where several candidates qualified, the nearest was used. In total 129 NOAA events were matched; the 82 unmatched ones are NOAA-only (55 before 1902, 27 after). A matched event keeps the catalogue location and takes the larger of the two magnitudes, because mb saturates for large events. This rule changes the magnitude

of 22 of the 129 matched events; the two magnitudes may be on different scales, so "larger" is a heuristic for saturation, not a conversion, and keeping the catalogue magnitude instead changes no result (Section 8.7). The merged table contains **19,045 distinct earthquakes**: 18,834 catalogue-only, 129 in both and 82 NOAA-only. Of these, 1,648 have a model distance (R_{JB}, Section 4) of at most 400 km, the limit of the ground-motion model, and were scored.

3 Seismicity and shaking history

3.1 Seismicity around the site

Table 1 counts merged events by epicentral distance. Seismicity is sparse close to the site: one event within 25 km, 17 within 50 km (largest M5.7) and none of M≥7 within 150 km. Large earthquakes are mostly more than 150 km away, along the Himalayan arc and in the Kashmir–Tibet region.

Table 1: Merged events by epicentral distance from the site.

Radius (km)	All events	Max M	$M \geq 5$	$M \geq 6$	$M \geq 7$
25	1	4.2	0	0	0
50	17	5.7	3	0	0
100	38	6.4	7	1	0
150	138	6.9	22	3	0
200	297	7.5	46	4	1
300	882	7.9	152	28	8
500	2,550	7.9	435	61	15
1000	16,517	8.3	2,276	330	66

The nearest NOAA event is the 1885 Kashmir earthquake at 168 km; the 1905 Kangra earthquake (M7.9, intensity 10) is at 208 km, and 37 NOAA events lie within 500 km.

3.2 Shaking actually experienced

Each scored event was converted to a median PGA at the site with the ground-motion model of Section 4. The largest values are in Table 2.

Table 2: Events with the highest estimated median PGA at the site (rock).

Date	Source	M	Epicentral km	PGA (g)	Note
1941-01-10	USGS	5.7	34	0.049	40 km ESE of Leh
1905-04-04	USGS/NOAA	7.9	208	0.040	Kangra
1937-11-15	USGS	6.35	62	0.035	86 km NE of Leh
1885-06-06	NOAA	7.5*	168	0.029	Kashmir (M assumed)
1996-11-19	USGS	6.9	119	0.025	ENE of Thang
2020-09-25	USGS	5.3	39	0.024	58 km E of Leh
1555-09	NOAA	7.8	274	0.015	Srinagar

Only 6 events produced a median PGA of 0.02 g or more, 11 reached 0.01 g, and none reached 0.05 g. With the scatter of Section 4, the 84th-percentile PGA reaches 0.01 g for 31 events, 0.02 g for 12 and 0.05 g for 4. **Among the events in the assembled dataset, the strongest estimated shaking is therefore about 0.05 g at the median.** This is a statement about the dataset and the chosen ground-motion calculation, not a verified upper bound on the shaking the site has experienced: the NOAA database lists only damaging or notable earthquakes, and historical locations, magnitudes and source geometries are uncertain and unevenly covered. Figure 3 plots every scored event against time.

3.3 Completeness of the record

Figure 1 shows why a century of catalogue says little by itself. The number of M≥4 events per decade jumps from 206 in the 1960s to 1,449 in the 1970s (within 500 km: 25 to 246), when global networks improved. Cumulative curves for M≥4 and M≥5 bend in the 1970s; the M≥6 and M≥7 curves are smoother, indicating earlier completeness for large events. Before 1900 the only data are the 55 NOAA events, which are felt or damaging events and not a complete sample. The completeness years adopted for modelling (Section 7.3) come from this evidence.

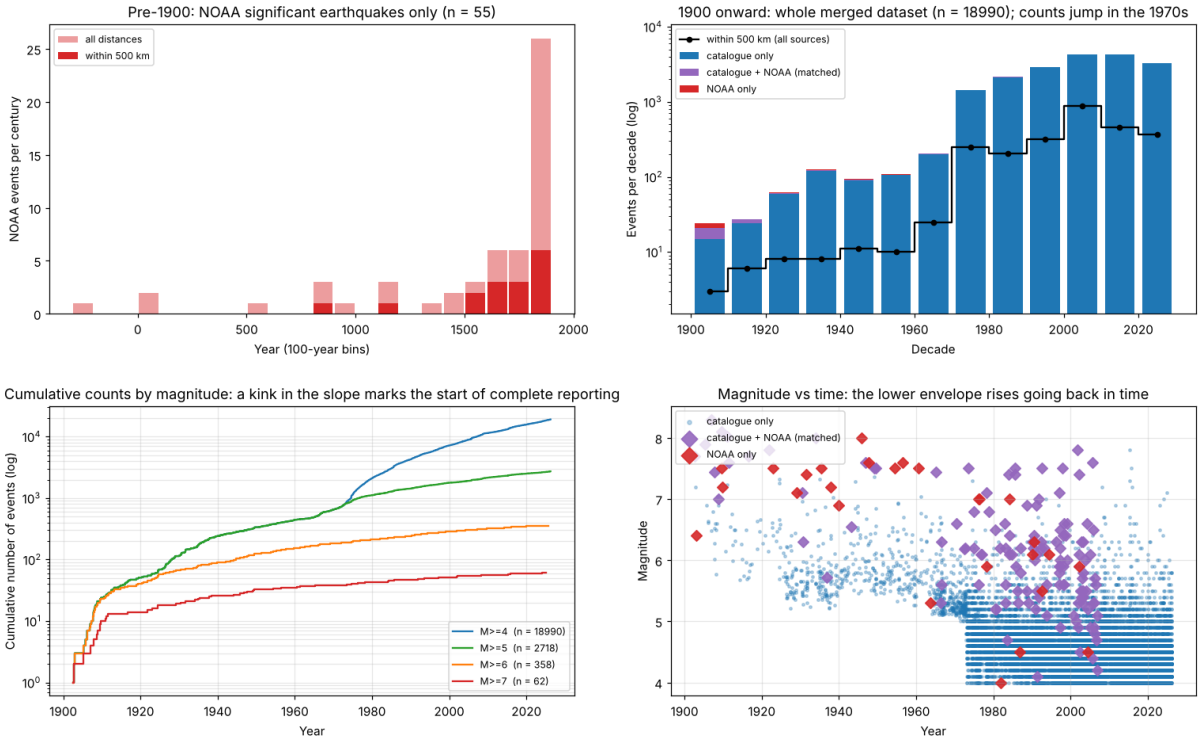


Figure 1: Completeness of the merged dataset. Top left: NOAA events per century before 1900. Top right: events per decade from 1900, stacked by source, with the within-500 km count. Bottom left: cumulative counts by magnitude threshold. Bottom right: magnitude against time; the lower envelope rises going back in time.

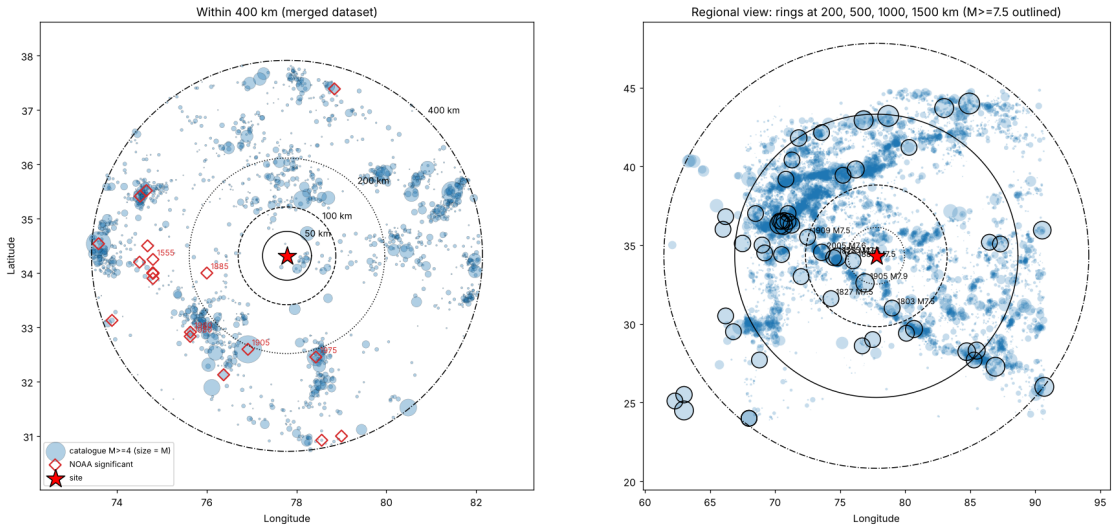


Figure 2: Merged events within 400 km (left) and a regional view with rings at 200–1500 km (right).

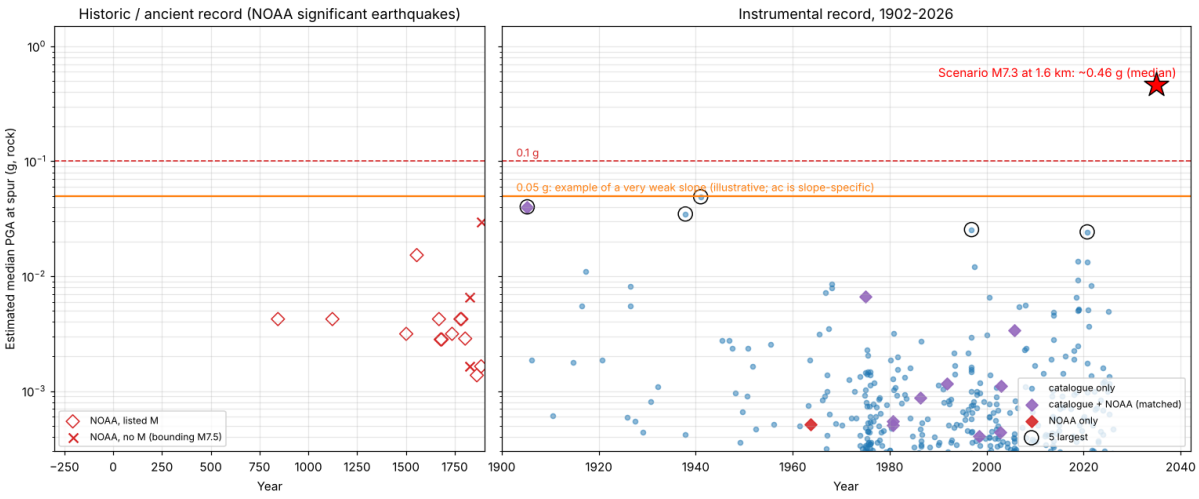


Figure 3: Estimated median PGA at the site from every scored event, with the scenario Karakoram Fault earthquake (red star). The orange line is 0.05 g, where loose debris starts to move.

4 Ground-motion model

Shaking is estimated with the BSSA14 ground-motion prediction equation (GMPE; Boore et al., 2014) for PGA, strike-slip faulting, on rock with $V_{S30} = 760\,\mathrm{m/s}$, the model's reference-rock condition, at which the site-amplification term is zero by construction. This is an assumption about the spur, not a finding: its actual V_{S30} and topography have not been measured. The median is

$$\ln \mathrm{PGA} = F_E(M) + F_P(R_{JB}, M), \tag{1}$$

$$F_E(M) = \begin{cases} e_1 + e_4(M - M_h) + e_5(M - M_h)^2, & M \le M_h \\ e_1 + e_6(M - M_h), & M > M_h \end{cases} \tag{2}$$

$$F_P = [c_1 + c_2(M - 4.5)]\ln(R/R_{\mathrm{ref}}) + c_3(R - R_{\mathrm{ref}}), \qquad R = \sqrt{R_{JB}^2 + h^2}, \quad R_{\mathrm{ref}} = 1\ \mathrm{km}. \tag{3}$$

The coefficients are $e_1 = 0.4856$, $e_4 = 1.431$, $e_5 = 0.05053$, $e_6 = -0.1662$, $M_h = 5.5$, $c_1 = -1.134$, $c_2 = 0.1917$, $c_3 = -0.008088$, $h = 4.5\,\mathrm{km}$.

Implementation check. The coefficients were checked against the independent `pygmm` implementation of BSSA14 at 49 combinations of magnitude (M4.5–7.9) and distance ($R_{JB} = 0.5$–$300\,\mathrm{km}$); the median PGA agreed to better than 10^{-5} relative error, including 0.459 g for M7.3 at $R_{JB} = 1.6\,\mathrm{km}$. The check was made in a separate environment; the production run used the embedded implementation, which the analysis script labels "unverified" when `pygmm` is absent, and the check can be repeated with the option `-gmpe both`. Agreement with `pygmm` shows that the equation is transcribed correctly; it does not show that BSSA14, a global model for shallow crustal earthquakes in active regions, is accurate for every Himalayan source and path. No alternative ground-motion model is used, so the absolute accelerations rest on a single model.

Other modelling choices.

- *Component.* BSSA14 predicts RotD50 horizontal PGA, the orientation-independent median of the two horizontal components; all accelerations here are in that sense.

- *Scatter.* Residuals are lognormal in $\ln \mathrm{PGA}$ with the BSSA14 total standard deviation $\sigma_{\ln} = \sqrt{\tau^2 + \phi^2}$, which depends on magnitude and distance. For PGA on rock ($V_{S30} \geq 300\,\mathrm{m/s}$, so the scatter model has no site-dependent term),

$$\begin{aligned} \tau(M) &= \tau_1 + (\tau_2 - \tau_1) f_M, \qquad f_M = \mathrm{clip}(M - 4.5, 0, 1), \\ \phi(M, R_{JB}) &= \phi_1 + (\phi_2 - \phi_1) f_M + \Delta\phi_R \, \frac{\ln\big(\mathrm{clip}(R_{JB}, R_1, R_2)/R_1\big)}{\ln(R_2/R_1)}, \end{aligned} \tag{4}$$

  with $\tau_1 = 0.398$, $\tau_2 = 0.348$, $\phi_1 = 0.695$, $\phi_2 = 0.495$, $\Delta\phi_R = 0.100$, $R_1 = 110\,\mathrm{km}$, $R_2 = 270\,\mathrm{km}$. This gives $\sigma_{\ln} = 0.61$ for the M7.3 fault scenario at 1.6 km, about 0.7 for M5 at 20 km and about 0.8 or more for M4–4.5 events. The 84th percentile is the median times $e^{\sigma_{\ln}}$ (a factor 1.83 for the scenario). The embedded formula matched `pygmm` to within 0.005 at 42 magnitude–distance pairs.

- *Truncation.* The scatter is truncated at $\pm 3\sigma$, two-sided (Section 7.9).

- *Finite faults.* For $M \geq 6.5$ the Joyner–Boore distance is $R_{JB} = \max(R_{\mathrm{epi}} - \frac{1}{2}L, 0)$ with surface rupture length $\log_{10} L = -3.22 + 0.69M$ (Wells and Coppersmith, 1994). This needs no rupture strike, so it is the *shortest* distance the rupture could have (it assumes the rupture points at the site): distances are biased short, and PGAs high, for large events whose rupture does not. For past events this errs on the conservative side and strengthens the conclusion that past shaking was small. Real distances need fault geometry.

- *Distance limit.* The model is calibrated to about 400 km; events with R_{JB} beyond that are not scored. Because the limit applies to R_{JB} after the finite-fault correction, a large event with an assumed magnitude can be scored although its epicentre is slightly farther (the 1827 Lahore event: epicentral distance 444 km, R_{JB} 399 km).

- *Distance definition.* R_{JB} is used exactly as BSSA14 defines it, the closest horizontal distance to the surface projection of the rupture; no source depth is added, because the model already contains its own pseudo-depth term h. In the hazard grid, the source distance of a cell is $\sqrt{d^2 + (0.38 \times 10\ \mathrm{km})^2}$; the added 3.8 km represents the spread of sources inside a 10 km cell and is not a source depth.

- *Source depth and type.* BSSA14 is a model for shallow crustal earthquakes, and the catalogue is not: of 1,631 events inside the zone, 362 have a reported depth below 35 km, 202 below 50 km and 100 below 70 km, while 457 carry the default depth of 33 km and 538 the

default of 10 km, so many depths are not measured. Events deeper than 35 km are scored with R_{JB} only and are flagged as indicative. None of the events that control the historical results is deeper than 35 km, and removing the deeper events from the recurrence model changes the hazard little (Section 8.7).

- *Fault style.* The strike-slip equation is used for every source, although thrust earthquakes dominate the Himalayan arc, strike-slip faulting characterises the Karakoram Fault and the region also contains extensional and intermediate-depth seismicity. BSSA14 includes style-of-faulting terms, but they were not varied here, and one equation across these source types is a simplification.

5 Deterministic scenario: a Karakoram Fault earthquake

The site is assumed to lie about 1.6 km (roughly one mile) from the nearest strand of the Karakoram Fault; the strand has not been mapped at site scale for this study, and its existence, continuity and activity next to the site are not established (Section 7.8). Every result in this section is therefore a conditional calculation: if such a strand exists and ruptures, a rupture on it would put the site essentially on the fault. Table 3 gives median PGA for a range of magnitudes and distances around the central scenario of M7.3 at 1.6 km, the central branch of the characteristic magnitudes used in Section 7. It yields **0.46 g median and about 0.84 g at the 84th percentile**, which is 9.4× the largest shaking in the entire record.

Table 3: Median PGA (g, rock) for Karakoram Fault scenarios.

R_{JB} (km)	M7.0	M7.3	M7.6	M7.7
0.5	0.46	0.48	0.49	0.50
1.6	0.44	0.46	0.48	0.49
5	0.35	0.37	0.39	0.40
10	0.24	0.27	0.29	0.30

Why the past may understate the future is shown by a slip budget, which holds only if the fault is active at the site. At 2–4 mm/yr, accumulating the 1.8 m of average slip of an M7.3 rupture takes 445 to 890 years (2.0 to 4.1 m and 1,020 to 2,040 years for M7.7). The 124 years of instrumental data cover only 0.25–0.50 m of that accumulation, so the absence of large shaking in the catalogue is what is expected *if* the fault is active. It is equally what is expected if the fault is inactive at the site; the catalogue cannot distinguish the two (Section 7.8). The M7.7 column is a potential upper magnitude, not a measured rupture.

5.1 Landslide screening (Newmark sliding block)

Debris on slopes at the spur was screened with the empirical sliding-block regression of Jibson (2007),

$$\log_{10} D_n = -2.710 + \log_{10}\left[\left(1 - \frac{a_c}{a_{\max}}\right)^{2.335}\left(\frac{a_c}{a_{\max}}\right)^{-1.478}\right] + 0.424\,M, \tag{5}$$

with D_n in cm, a_c the slope's critical acceleration and $a_{\max}$ the PGA. The regression has a standard deviation of about 0.5 log units, so the mean plus or minus one standard deviation spans about an order of magnitude, and its author recommends it for regional-scale mapping and rapid preliminary screening, not for site-specific design. Three further cautions apply. First, critical

acceleration is specific to each slope (geometry, material strength, water content), so no value is a universal threshold; the values 0.02–0.30 g are used here only to bracket very weak to moderately strong slopes until slope-specific values are measured. Second, the regression was calibrated over a finite range of critical-acceleration ratios; table entries with $a_c/a_{\max}$ outside roughly 0.05–0.40 are marked † as possible extrapolations (the exact bounds should be confirmed against Jibson, 2007). Third, a predicted displacement is not a probability of failure; it is conditional on the assumed shaking and a_c, and no universal displacement threshold separates stable from failed slopes. The magnitude used is that of the event producing the shaking considered (the M5.7 earthquake of 1941 for the largest past shaking). For the largest estimated shaking in the dataset (0.049 g) displacement is 0.55 cm at $a_c = 0.02$ g (an extrapolated entry) and zero for $a_c \geq 0.05$ g. For the scenario median it is 225 cm at $a_c = 0.02\,\text{g}^{\dagger}$, 49 cm at 0.05 g, 13 cm at 0.10 g, 2.2 cm at $0.20\,\text{g}^{\dagger}$ and 0.4 cm at $0.30\,\text{g}^{\dagger}$ (Table 13). On this screening, slopes with low critical accelerations would be candidates for field investigation if the scenario occurred; the calculation does not show that any particular slope will fail.

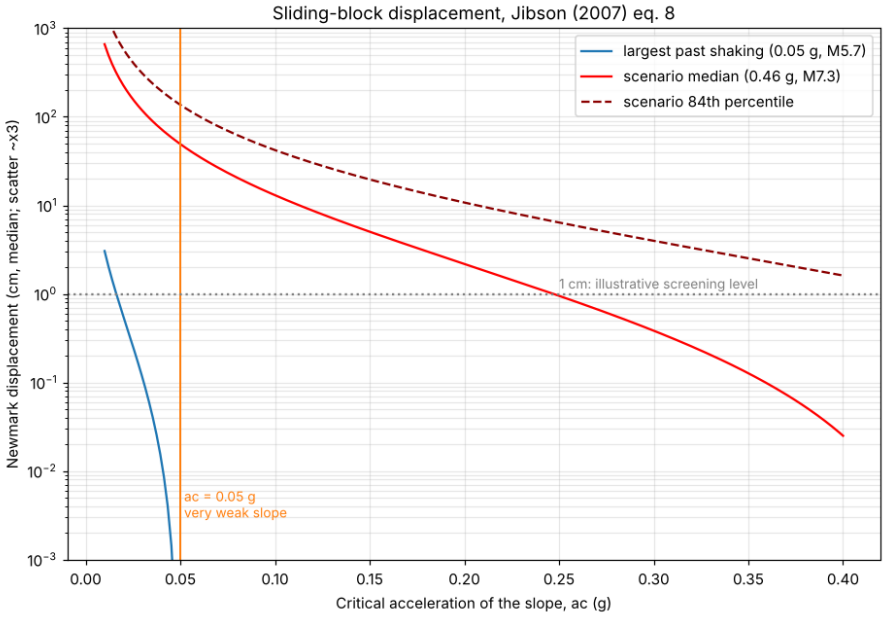


Figure 4: Newmark displacement against slope critical acceleration for the largest estimated shaking in the dataset and the scenario.

6 Earthquakes relevant to design

A machine-readable list, `design_relevant_earthquakes.json`, selects every merged event that meets *any* of these rules (thresholds are command-line options):

A median PGA at the site ≥ 0.02 g;

B 84th-percentile PGA ≥ 0.05 g;

C $M \geq 6$ within 200 km;

D $M \geq 7$ within 500 km;

E $M \geq 5$ within 100 km;

F listed by NOAA as a significant (damaging or notable) event within 500 km.

Forty-seven events qualify; rule memberships (an event may satisfy several) are F: 37, D: 15, E: 7, A: 6, B: 4, C: 4. Each entry records its date, location, depth, magnitude and type, distances, median and 84th-percentile PGA, the rules that selected it and notes (assumed magnitude, mb saturation, deep event, beyond model range). The file also contains the scenario event and 32 very large events (M≥7.5 within 1,000 km) that lie beyond the model's range and are listed for context only. The events are listed in Appendix A and mapped in Figure 5; a GeoJSON copy supports GIS use.

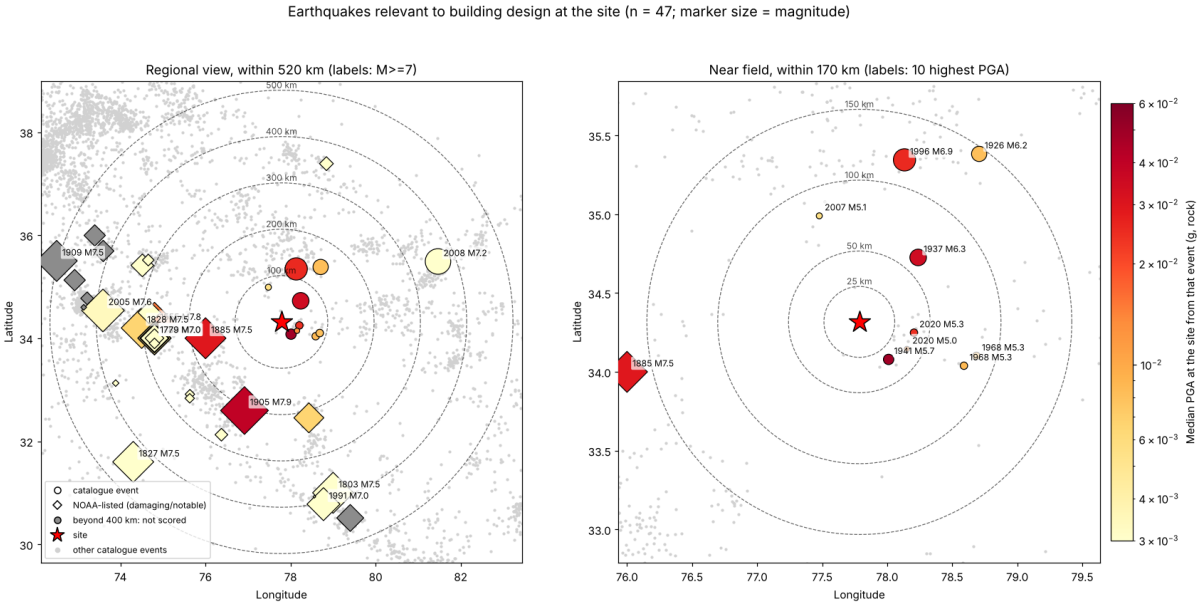


Figure 5: The 47 design-relevant earthquakes. Colour is the median PGA each produced at the site, marker size is magnitude, diamonds are NOAA-listed events, grey markers lie beyond the 400 km scoring limit.

7 The return-period model

This section documents the model that predicts how often earthquakes, and shaking at the site, recur. It is a screening-level PSHA in the tradition of Cornell (1968), with a gridded background source and a separate fault source, implemented in the function `run_forecast` of the analysis script (Section 11).

7.1 Definitions and the Poisson assumption

Earthquakes are assumed to occur as a Poisson process: independent, with constant average rate λ per year. The probability of at least one event in t years is then

$$P(t) = 1 - e^{-\lambda t}, \tag{6}$$

and the *return period* is $T = 1/\lambda$. For a ground-motion level x, $\lambda(x)$ is the annual rate of exceeding x at the site, and the PGA with return period T is the x_T satisfying $\lambda(x_T) = 1/T$. A 475-year return period corresponds to a 10% chance of exceedance in 50 years and 2,475 years to 2%. Because earthquakes cluster, the Poisson model is applied only to declustered mainshocks. A return period of ground motion is the reciprocal of an exceedance *rate*; it is not the recurrence interval of any particular earthquake.

7.2 Source zone and declustering

The background source zone is the disc of radius 400 km around the site (the limit of the ground-motion model). To avoid edge effects, declustering and smoothing use events out to 550 km. Since 1902 there are 1,631 events inside the zone. Foreshocks and aftershocks were removed with the windows of Gardner and Knopoff (1974): a smaller event is dropped if it lies within the distance d and time t of a larger one,

$$d = 10^{0.1238M+0.983}\ \text{km}, \qquad t = \begin{cases} 10^{0.032M+2.7389}\ \text{days}, & M \geq 6.5 \\ 10^{0.5409M-0.547}\ \text{days}, & M < 6.5, \end{cases} \tag{7}$$

applied $\pm t$ in time and processed from the largest event downward. Of the 1,631 events, 814 (50%) remain as mainshocks. For M≥4.5, 483 of 890 are kept (Figure 6, right). Rate estimation, the back-test and the spatial smoothing use catalogue-sourced events only: the NOAA records are a selected subset. One NOAA-only mainshock lies inside the zone (1963, M5.3, 279 km); it falls before its magnitude bin's completeness year, so excluding it changes nothing.

7.3 Magnitude-dependent completeness

Completeness years by 0.5-unit magnitude bin were read from the decade counts and cumulative curves of the merged data:

Bin lower edge	4.0	4.5	5.0	5.5	6.0	≥6.5
Complete from	excluded	1973	1964	1930	1904	1902

The M4.0–4.5 bin is excluded because its observed rate is lower than that of the M4.5–5.0 bin, which can only mean incomplete reporting in this remote region. Including it flattens the fitted b-value to 0.78 and produces an implausible M≥6 return period of 3 years. The completeness table can be overridden with `-completeness`.

7.4 Gutenberg–Richter recurrence by maximum likelihood

Magnitudes are assumed to follow the Gutenberg–Richter law (Gutenberg and Richter, 1944), $\log_{10} N(\geq M) = a - bM$, equivalently an exponential distribution with $\beta = b \ln 10$. Because each bin has its own observation period, the parameters are estimated by the maximum-likelihood method of Weichert (1980). In bin i with edges m_i^{lo}, m_i^{hi}, observed for T_i years with n_i mainshocks, the expected rate is

$$\lambda_i = A\, f_i(\beta), \qquad f_i(\beta) = e^{-\beta(m_i^{lo}-M_0)} - e^{-\beta(m_i^{hi}-M_0)}, \qquad M_0 = 4.0, \tag{8}$$

and the Poisson log-likelihood is $\sum_i [n_i \ln(T_i\lambda_i) - T_i\lambda_i]$. Maximizing over A gives $\hat{A} = N / \sum_i T_i f_i(\beta)$ with $N = \sum n_i$, leaving a profile likelihood in β only,

$$\ell(\beta) = \sum_i n_i \ln f_i(\beta) - N \ln \sum_i T_i f_i(\beta), \tag{9}$$

which is maximized on a fine grid. The standard error is $\sigma_\beta = (-\ell''(\hat{\beta}))^{-1/2}$ from the numerical curvature.

Result (476 mainshocks): $b = 1.14 \pm 0.05$. The annual rates inside the 400 km zone are 2.18 events/yr for $M \geq 5$, 0.158/yr for $M \geq 6$ (return period 6.3 years) and 0.0115/yr for $M \geq 7$ (return period 87 years). The standard error is statistical only, conditional on the catalogue,

the completeness table and the magnitudes as given; it excludes uncertainty from magnitude type (in the fit window 310 of 319 events in M4.5–5.0 and 58 of 105 in M5.0–5.5 are mb), event selection and completeness (Section 8.7). Table 4 compares fitted and observed counts.

Table 4: Recurrence fit: observed and expected mainshock counts within 400 km in each bin's completeness window, with two-sided Poisson p-values (computed from unrounded expected counts).

Bin	Years observed	Observed	Expected	p
M4.5–5.0	53	319	315.9	0.87
M5.0–5.5	62	105	99.4	0.60
M5.5–6.0	96	26	41.4	0.01
M6.0–6.5	122	19	14.2	0.25
M6.5–7.0	124	4	3.9	1.00
M7.0–7.5	124	1	1.0	1.00
M7.5–8.0	124	2	0.3	0.07

Goodness of fit. The fit is in-sample, so these tests are generous. The largest bin stands out: two M7.5–8.0 mainshocks (presumably the 1905 Kangra and 2005 Kashmir earthquakes) were observed against 0.3 expected, about seven times the expectation; the probability of two or more under a Poisson model with mean 0.28 is 3.3% (one-sided), or 0.07 on the two-sided scale used in the table. With seven bins, one such value is not surprising, and magnitude scales in the catalogue are heterogeneous and the sample is tiny, so it is a warning and not proof of misfit. Pooling all M≥6.5 gives 7 observed against about 5.2 expected ($P(N \geq 7) = 0.27$), no evidence of misfit for the pooled tail. The M5.5–6.0 bin (26 against 41, $p = 0.01$) is the clearest individual misfit and probably reflects completeness or magnitude-conversion problems rather than physics. The magnitude types support the second reading: the M5.5–6.0 bin is 23 of 26 moment-type, whereas the bins that dominate the b-value are mostly mb, so the deficit appears on a different magnitude scale from the one that sets the slope. Together with the clustering of the largest events, this is a reason to investigate magnitude homogenisation and source zonation (Section 10), not a minor statistical detail; Section 8.7 shows how much the result moves. The practical effect on the site is small: an M7.9 event 150 km away gives only about 0.03 g at the site, and above a few hundred years the hazard is controlled by the fault, not the background.

7.5 Hazard magnitude bins and maximum magnitude

For the hazard integral the recurrence law is evaluated in 0.1-unit bins from M4.5 (smaller events are not damaging) up to a maximum magnitude $M_{\max}$:

$$\lambda_j = A\left[e^{-\beta(m_j^{lo}-M_0)} - e^{-\beta(\min(m_j^{hi},M_{\max})-M_0)}\right] \quad (m_j^{lo} < M_{\max}), \qquad \lambda_j = 0 \text{ otherwise.} \tag{10}$$

Rates in the observed magnitude range are therefore unchanged by the choice of $M_{\max}$, which only removes the tail. $M_{\max}$ is a logic-tree branch (7.5, 7.9, 8.3 with weights 0.3, 0.5, 0.2). The largest catalogue magnitude within 400 km is the M7.9 Kangra earthquake of 1905, so any branch below it contradicts the catalogue: every $M_{\max}$ branch below the largest *catalogue* magnitude within 400 km is raised to it and the weights of merged branches are added (events with an assumed magnitude, such as the 1885 Kashmir M7.5, do not set the floor). The branches become 7.9 (weight 0.8) and 8.3 (weight 0.2). The floor has little effect on shaking at the site, because it only adds rate to M7.5–7.9 events that are typically more than 150 km away. Kangra is M7.9 in

both data files; some compilations give M_w7.8, which would move the floor by 0.1 unit and, by the same argument, matter little.

7.6 Spatial distribution: smoothed seismicity

The zone is a grid of 5,025 cells of 10×10 km. The share w_i of the zone's seismicity assigned to cell i is obtained by Gaussian smoothing (as in Frankel, 1995) of the declustered M≥4.5 mainshocks,

$$w_i \propto \sum_e a_e \exp\left(-\frac{d_{ie}^2}{2\sigma_s^2}\right), \qquad a_e = \frac{1}{t_{\text{end}} - t_c(m_e)}, \qquad \sum_i w_i = 1, \tag{11}$$

where d_{ie} is the distance from cell i to event e, t_c is the completeness year for the event's magnitude (so each event counts as the annual rate it represents) and $\sigma_s = 50$ km (option `-smooth-km`). The rate of magnitude bin j in cell i is $\lambda_j w_i$. Quiet surroundings mean that only 0.8% of the zone's rate falls within 50 km of the site, 3.5% within 100 km and 22.6% within 200 km. Each cell's source distance is $\sqrt{d_i^2 + (0.38 \times 10\text{ km})^2}$, so a cell on top of the site is not at zero distance; this term accounts for the spread of sources within a cell and is not a source depth (Section 4).

7.7 The Karakoram Fault source

The background catalogue contains no large events on the fault, so the fault is modelled separately as a characteristic-earthquake source at $R_{JB} = 1.6$ km. The rate is set by the slip budget, not the catalogue. A point on the fault must slip on average $\bar{D}(M)$ in every characteristic rupture, with (Wells and Coppersmith, 1994) $\log_{10}\bar{D} = -6.32 + 0.90M$ metres, giving 0.95 m for M7.0, 1.78 m for M7.3, 3.31 m for M7.6 and 4.07 m for M7.7. The annual rate is

$$\lambda_f = \frac{c\,s}{\bar{D}(M_c)}, \tag{12}$$

where s is the slip rate and c the coupling, the fraction of slip released seismically (default $c = 1$, the conservative choice; option `-coupling`). Slip-budget recurrence intervals $1/\lambda_f$ are (years):

	$s = 1$	$s = 2$	$s = 3$	$s = 4$ mm/yr
M7.0	955	477	318	239
M7.3	1,778	889	593	445
M7.6	3,311	1,656	1,104	828
M7.7	4,074	2,037	1,358	1,018

These are conditional on the fault being active, and are slip-budget estimates, not observed recurrence intervals. The characteristic magnitudes are *alternative* descriptions of the same fault, not events to be added together. Weighting the branches below, the mean characteristic-earthquake rate is about 0.0011 per year (mean return period about 900 years): about 0.0015 per year (650 years) under hypothesis A and 0.0003 per year (3,100 years) under hypothesis B. These derived figures depend entirely on the assumed slip rates and coupling, and on the assumption that every characteristic rupture releases the average slip at the site-adjacent reach; they are the model's largest single uncertainty.

7.8 Evidence for fault slip at the site

The slip-rate range is taken from published geodetic and geological studies of the Karakoram Fault, summarised in Table 5.

Table 5: Published constraints on Karakoram Fault slip relevant to the Agham reach.

Source	Finding	Caveat
Mohanty et al. (2024)	Continuous GPS over more than 300 km: Nubra valley (K2 segment of the northern fault) inactive; Bangong–Chaxikang (southern fault) 1.6–3.2 mm/yr, on the Pangong strand; the Tangtse strand takes no slip.	Nine-year GPS snapshot; applies to the strands crossed by its profiles, not shown to include the trace 1.6 km from the site.
Wang et al. (2026)	GNSS and InSAR block model: northern segment (about 36.2–37°N) can be below 1 mm/yr; the middle and southern segments (about 30.5–35°N) are faster.	Regional model rate. The authors warn that strong coupling on the Main Frontal Thrust may bias the inferred rate and that it may not equal the long-term rate.
Brown et al. (2002)	4 ± 1 mm/yr from cosmogenic dating of offset debris flows and moraines in Ladakh.	A geological rate for another reach (the Tangtse Valley study area); the basis of the 4 mm/yr upper branch, not a measurement at the site and not transferable to the assumed strand without evidence.
Bohon et al. (2010)	A Quaternary strand in the Tangyar valley with offset alluvial surfaces; slip rate undetermined at the time.	Shows that a possibly active parallel strand exists in the sector; not shown to be the trace at the site.
Ojha (2026)	Segmented, "stop-and-go" behaviour; last kinematic readjustment about 0.5 Ma.	A structural age, not a dated last rupture and not a recurrence interval.
Wang and Shen (2020)	Continental GPS model: northern segment about 0, southern about 3 mm/yr.	Not independent of the GPS data used in other analyses; not a second confirmation.

The site (34.32°N) falls in the latitude band of the faster middle and southern segments of Wang et al., but lies within the Nubra sector, where the inactive result of Mohanty et al. applies. Which strand lies 1.6 km from the site, and whether it is the inactive Nubra/K2 strand, an active parallel Quaternary strand, or a distinct reach, is not settled by the published literature at coordinate level. For that reason the model keeps both outcomes open instead of selecting one slip rate (Sections 7.10 and 7.11). The weights placed on the alternatives are analyst judgement, not quantities established by the literature.

7.9 The hazard integral

Given a median PGA $\mathrm{PGA}_{med}(m, r)$ from Section 4, the probability that a magnitude-m event at distance r exceeds x at the site, with scatter truncated at $\pm 3\sigma$, is

$$P(x \mid m, r) = \frac{\Phi(3) - \Phi(\max(-3, \min(z, 3)))}{\Phi(3) - \Phi(-3)}, \qquad z = \frac{\ln x - \ln \mathrm{PGA}_{med}(m, r)}{\sigma_{\ln}(m, r)}, \tag{13}$$

with Φ the standard normal distribution function and $\sigma_{\ln}(m, r)$ the BSSA14 standard deviation of Section 4. The two-sided truncation requires the normalising constant $\Phi(3) - \Phi(-3)$ *and*

clamping z at both ends, so that $P = 1$ when x is more than 3σ below the median and $P = 0$ when it is more than 3σ above. The annual rate of exceeding x for one logic-tree branch is the sum of the background and fault contributions,

$$\lambda(x) = \underbrace{\sum_j \sum_i \lambda_j \, w_i \, P(x \mid m_j, r_i)}_{\text{background}} + \underbrace{\lambda_f \, P(x \mid M_c, 1.6 \text{ km})}_{\text{Karakoram Fault}} . \tag{14}$$

Median PGAs are tabulated on a distance grid and interpolated, which is exact for the embedded model and fast enough for `pygmm`. The level x runs over 70 values from 0.005 to 2 g.

7.10 Logic tree

Uncertain parameters are combined as independent branches with weights:

Parameter	Values	Weights
β (b-value)	$\hat{\beta} - \sigma_\beta$, $\hat{\beta}$, $\hat{\beta} + \sigma_\beta$	0.25, 0.5, 0.25
Background $M_{\max}$ (floored at the largest observed magnitude)	7.9, 8.3	0.8, 0.2
Fault slip rate (mm/yr)	0, 1, 2, 3, 4	0.15, 0.20, 0.30, 0.25, 0.10
Characteristic magnitude	7.0, 7.3, 7.6, 7.7	0.25, 0.35, 0.25, 0.15

This gives $3 \times 2 \times 5 \times 4 = 120$ branches. The mean hazard is the weighted mean of the branch rates, $\bar{\lambda}(x) = \sum_b W_b \lambda_b(x)$. The 16th and 84th percentiles are weighted quantiles of the branch exceedance rates $\lambda_b(x)$ at each x (percentile *curves*), and the tabulated 16% and 84% PGAs come from inverting those curves at $1/T$. These bands are *epistemic*: they show spread among logic-tree branches, not the scatter of one earthquake's ground motion, which is the aleatory $\sigma_{\ln}$ already inside every branch curve. Where branch curves cross, this is not the same as inverting each branch first and taking weighted percentiles of the branch PGAs; the script computes that version as a cross-check, and here the two agree to within 0.00014 g at every tabulated return period. The return-period PGA x_T is found by log–log interpolation of $\bar{\lambda}(x)$ at $1/T$. The slip-rate and magnitude weights are judgements, not measurements; the weight of 0.15 on zero slip and the weights on the M7.6 and M7.7 branches in particular are analyst choices. The next subsection shows how results are reported so that they do not depend on them.

7.11 Fault-activity hypotheses and the governing envelope

The slip-rate branches fall into two explicit hypotheses: **A**, the fault slips at a rate consistent with the geodetic and geological studies of the active segments (2, 3, 4 mm/yr; tree weight 0.65) and **B**, inactive or very low (0, 1 mm/yr; tree weight 0.35). Within each hypothesis the branch weights are renormalised and the mean hazard $\bar{\lambda}_A(x)$, $\bar{\lambda}_B(x)$ is computed with the background and magnitude branches unchanged. The *governing envelope* is $\lambda_{\text{env}}(x) = \max\big(\bar{\lambda}_A(x), \bar{\lambda}_B(x)\big)$ at each x. The 0.65/0.35 split is a judgement; the conditional curves and the envelope do not depend on it. These are conditional curves, not predictions: the model does not say which hypothesis is true. The envelope is conservative only relative to A and B; it does not cover the ground-motion-model, site-response, rupture-geometry or catalogue uncertainties listed in Section 10.

7.12 Forecasting earthquakes and shaking in the building's life

Using the mean rates, the rate of earthquakes of magnitude at least m within radius R is

$$\Lambda(M \geq m, R) = \Big(\sum_{d_i \leq R} w_i \Big) \sum_{j:\, m_j^{lo} \geq m} \bar{\lambda}_j + \sum_{M_c \geq m} \bar{\lambda}_f(M_c), \tag{15}$$

where the last term adds characteristic fault events when they qualify. The probability of at least one in t years is then $1 - e^{-\Lambda t}$. The probability that PGA at the site exceeds x in t years is $1 - e^{-\bar{\lambda}(x)t}$. A 475-year return period is a statistical measure of how often shaking of that size is exceeded under a Poisson model; it does not mean that an earthquake occurs every 475 years or that the next one is 475 years away.

7.13 Disaggregation

To identify the controlling earthquakes, the mean-model contribution to $\lambda(x_T)$ was split into magnitude bins (0.5 wide) and distance bins (0–10, 10–25, 25–50, 50–100, 100–200, 200–300, 300–400 km). Fault events are placed in the 0–10 km bin, and the fault source is also tracked separately so that it can be told apart from background earthquakes that happen to lie within 10 km.

7.14 Back-test

A split-sample back-test fitted the recurrence model to mainshocks before 2000 ($b = 1.11$) and compared predicted with observed counts for 2000–2026 (Table 6). Two-sided Poisson p-values are given. The sparse M6.5–7.0 and M7.0–7.5 bins are statistically uninformative.

Table 6: Back-test of the background recurrence model, 2000–2026.

Bin	Expected	Observed	p (two-sided)
M4.5–5.0	144.4	167	0.07
M5.0–5.5	40.0	51	0.11
M5.5–6.0	11.1	5	0.07
M6.0–6.5	3.1	3	1.00
M6.5–7.0	0.9	0	0.85
M7.0–7.5	0.2	1	0.42

The model predicts counts of the right order in every bin and no bin is rejected at the 5% level. There is a pattern: more M4.5–5.5 events than predicted and fewer M5.5–6.0, consistent with a shift in the catalogue's magnitude scale after about 2000, as moment magnitudes replaced mb for mid-sized events. The back-test tests the background source only; the fault source cannot be tested against a record that contains no large fault earthquake.

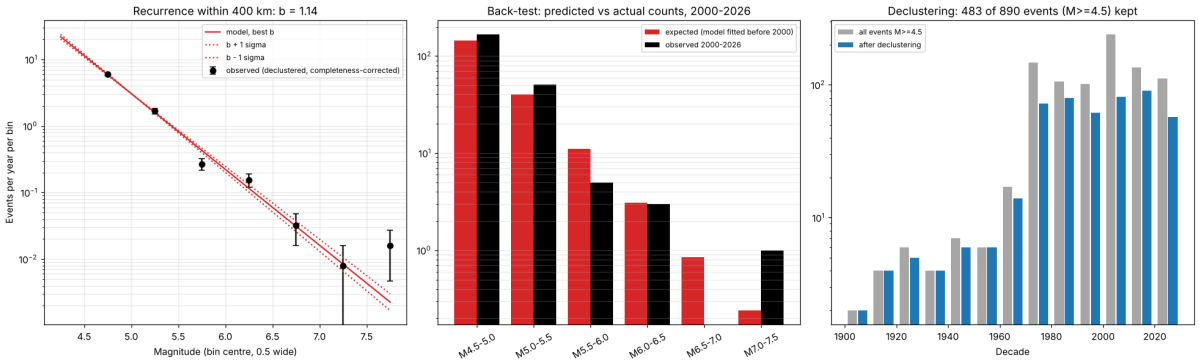


Figure 6: Left: observed and fitted recurrence (with $\pm 1\sigma$ in b). Middle: back-test. Right: effect of declustering.

8 Results

All values are for rock ($V_{S30} = 760\,\mathrm{m/s}$) and are mean PGA unless stated.

8.1 Return periods of ground shaking at the site

Table 7 gives the uniform-hazard PGA of the blended logic tree. The 72-year level (0.043 g) is set by background seismicity (the fault supplies 8% of the rate); by 475 years the fault supplies half of the exceedance rate and by 2,475 years 86%. The 16–84% epistemic band across branches widens from 0.040–0.045 g at 72 years to 0.34–0.77 g at 2,475 years, because the fault slip rate is the dominant uncertainty.

Table 7: Mean hazard (blended tree): PGA with a given return period; 16% and 84% are epistemic logic-tree quantiles.

Return period (yr)	P in 50 yr	P in 100 yr	Mean PGA (g)	16%	84%	Fault share
72	50.1%	75.1%	0.043	0.040	0.045	8%
475	10.0%	19.0%	0.168	0.123	0.259	50%
975	5.0%	9.7%	0.330	0.187	0.480	76%
2475	2.0%	4.0%	0.610	0.335	0.766	86%
4975	1.0%	2.0%	0.830	0.504	0.988	88%

8.2 Hazard by fault-activity hypothesis

Table 8: PGA (g, rock) by fault-activity hypothesis, the governing envelope and the blended tree.

Return period (yr)	A (2–4 mm/yr)	B (0–1 mm/yr)	Envelope	Blended tree
72	0.044	0.041	0.044	0.043
475	0.203	0.129	0.203	0.168
975	0.409	0.200	0.409	0.330
2475	0.700	0.365	0.700	0.610
4975	0.924	0.543	0.924	0.830

The envelope equals hypothesis A at every level, as expected, because an active fault always raises the hazard. The two hypotheses differ little at 72 years (0.044 against 0.041 g) and by a factor of 1.6 at 475 years, 2.0 at 975 years and 1.9 at 2,475 years. The return periods are ground-motion exceedance periods, not recurrence intervals of the characteristic earthquake.

8.3 The shape of the hazard curve

Figure 7 shows the hazard curve and the disaggregation. Below about 0.05 g the curve follows the background seismicity (an annual rate of 0.011 at 0.05 g, a 91-year return period). From about 0.1 g up, the curve is held up by the Karakoram Fault, whose component is almost flat at about 1×10^{-3} per year up to roughly 0.2 g and then falls as the scatter tail is exhausted. The jump from 0.043 g at 72 years to 0.17 g at 475 years and 0.61 g at 2,475 years reflects the transition from quiet background seismicity to the single nearby fault.

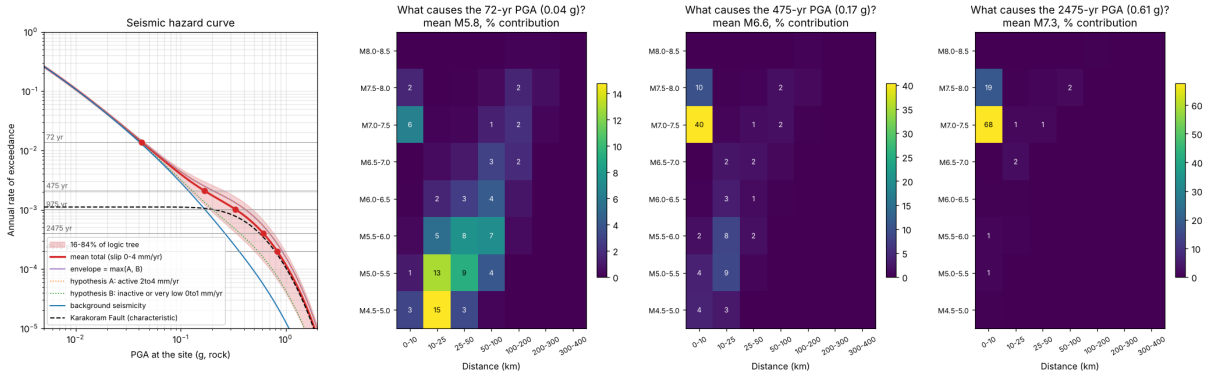


Figure 7: Hazard curve (mean, 16–84% epistemic band, background and fault components, hypotheses A and B and their envelope) and disaggregation of all sources at the 72-, 475- and 2475-year PGA. Numbers are % contributions; the 0–10 km column includes both the fault and background earthquakes within 10 km.

8.4 Disaggregation

- 72-year PGA (0.043 g): mean M5.8 at about 44 km; modal bin M4.5–5.0 at 10–25 km. The fault source supplies 8% and all sources inside 10 km 13%. This level is controlled by small and moderate background events.

- 475-year PGA (0.17 g): mean M6.6 at about 16 km. The Karakoram Fault source supplies 50% of the exceedance rate (M7.0–7.5: 40%; M7.5–8.0: 10%), matching the fault share in Table 7. A further 10% comes from background earthquakes inside 10 km, so the whole 0–10 km bin is 60%; the rest comes from moderate background events at 10–50 km. The 0–10 km bin must not be read as the fault alone.
- 2475-year PGA (0.61 g): mean M7.3 at about 9 km (bin resolution). The fault source supplies 86% (M7.0–7.5: 68%; M7.5–8.0: 19%); the 0–10 km bin totals 90%.

The hazard-consistent scenario at 2,475 years is therefore an M7.3 earthquake on the Karakoram Fault at about 1.6 km. At 475 years the fault and background contribute comparably, so no single scenario describes that level well. The distances reported by the disaggregation are bin-resolution.

8.5 Probability of shaking in the building's life

Table 9: Probability that PGA at the site exceeds a level at least once (mean model).

PGA (g)	Return period (yr)	10 yr	25 yr	50 yr	100 yr	200 yr
0.05	91	10.4%	24.0%	42.2%	66.6%	88.8%
0.10	250	3.9%	9.5%	18.1%	33.0%	55.1%
0.20	574	1.7%	4.3%	8.3%	16.0%	29.4%
0.30	876	1.1%	2.8%	5.5%	10.8%	20.4%
0.50	1,724	0.6%	1.4%	2.9%	5.6%	11.0%

For a 100-year life, a 33% chance of exceeding 0.10 g (16–84% range 27–39%) and a 5.6% chance of exceeding 0.5 g (2–9%) are the model's mean estimates. The wide range at 0.5 g reflects the fault slip rate.

8.6 Probability of earthquakes in the region

Table 10: Probability of at least one earthquake of the given size and distance (mean model, including characteristic fault events).

Event	Return period	10 yr	25 yr	50 yr	100 yr	200 yr
$M \geq 5$ within 50 km	54 yr	16.8%	36.8%	60.1%	84.0%	97.5%
$M \geq 5$ within 100 km	13 yr	53.9%	85.6%	97.9%	100%	100%
$M \geq 6$ within 100 km	150 yr	6.4%	15.3%	28.3%	48.5%	73.5%
$M \geq 6$ within 200 km	27 yr	30.9%	60.3%	84.2%	97.5%	99.9%
$M \geq 7$ within 200 km	283 yr	3.5%	8.5%	16.2%	29.8%	50.7%
$M \geq 7$ within 400 km	85 yr	11.1%	25.6%	44.6%	69.3%	90.6%

These are comparable with the record: since 1902 there are three M≥5 events within 50 km (a return period of about 41 years against the model's 54) and four M≥6 events within 200 km (about 31 years against 27). For M≥7 within 400 km the model's 85-year return period is longer than the record's three events in 124 years (about 41 years: the 1905 Kangra, 2005 Kashmir

and 2008 Hotan earthquakes). The model, whose fault rate is anchored to the slip budget, is therefore not inflated by the fault relative to the short record; if anything the record is more active than the model, though three events are too few to conclude that.

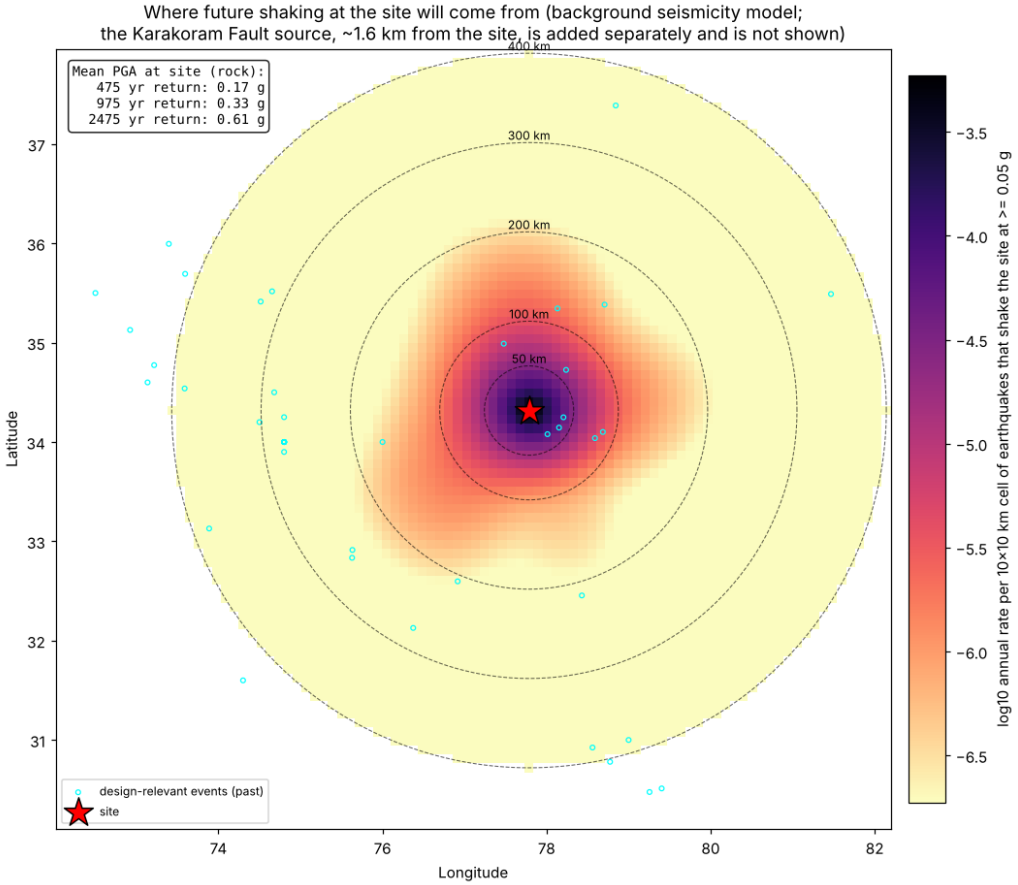


Figure 8: Where future shaking at the site will come from in the background model: $\log_{10}$ annual rate per cell of earthquakes that shake the site at 0.05 g or more. The Karakoram Fault source is added separately and is not shown.

8.7 Sensitivity tests

Fault coupling. The fault's contribution scales linearly with coupling c, so the hazard can be recomputed cheaply (Table 11).

Table 11: Mean PGA (g) against the fraction of Karakoram Fault slip released in large earthquakes.

Coupling	475 yr	975 yr	2475 yr
0.25	0.127	0.194	0.348
0.50	0.138	0.230	0.448
1.00	0.168	0.330	0.610

Across this range the 475-year PGA changes by about 30% (0.13 to 0.17 g), the 975-year value by 70% and the 2,475-year value by 75%. Coupling matters less than the choice between hypotheses A and B (Table 8), which changes the 2,475-year PGA by a factor of 1.9.

Ground-motion scatter. With a fixed $\sigma_{\ln} = 0.65$ instead of the BSSA14 model, the blended tree gives 0.042, 0.167, 0.328 and 0.626 g at 72, 475, 975 and 2,475 years, so the choice changes the 2,475-year PGA by about −2.5% and the other levels by 1% or less.

Catalogue magnitudes and source depth. Table 12 shows the effect of four choices about the earthquake catalogue. Keeping the catalogue magnitude, instead of the larger of the catalogue and NOAA values, for matched events changes nothing. Shifting all mb magnitudes by ±0.2 is a stress test, not a calibrated mb–Mw conversion. Restricting the recurrence fit and the spatial smoothing to M≥5.5, where the catalogue magnitudes are almost all moment magnitudes, removes most of the mb-dominated bins; it leaves only 52 mainshocks, so it is a sensitivity test and not a replacement for the base case. Removing events deeper than 50 or 70 km tests the effect of deeper sources on the background rates.

Table 12: Sensitivity of the blended mean rock PGA (g) to catalogue treatment.

Case	b	72 yr	475 yr	975 yr	2,475 yr
Base case	1.14	0.043	0.168	0.330	0.610
Catalogue magnitude kept when matched	1.14	0.043	0.168	0.330	0.610
mb shifted by −0.2	0.99	0.041	0.175	0.347	0.629
mb shifted by +0.2	1.10	0.056	0.206	0.369	0.640
Events deeper than 70 km removed	1.11	0.043	0.170	0.334	0.614
Events deeper than 50 km removed	1.10	0.043	0.173	0.337	0.617
Fit restricted to M≥5.5 (52 mainshocks)	0.78	0.047	0.228	0.420	0.703

The merge rule has no effect, and the removal of deeper events changes the hazard by less than 3%. The ±0.2 shifts raise the 475-year PGA by 4–23% and the 2,475-year PGA by under 5%; the 72-year PGA moves by −5 to +30%. The M≥5.5 fit has the largest effect: it gives $b = 0.78 \pm 0.12$ and an M≥7 return period of 30 years inside the zone against 87 years in the base case, and raises the 475-, 975- and 2,475-year PGAs by 36%, 27% and 15%. The fault share falls to 46%, 60% and 65% at those levels because the background is higher. With 52 mainshocks and about three M≥7 events the estimate is itself uncertain, but it shows that the base-case recurrence, which is anchored by mb-dominated small-magnitude bins, may understate the rate of larger events. The base case should therefore not be read as a conservative value for the background. This is smaller than the factor of 1.6–2 between fault hypotheses A and B, but it is not negligible, and real conversions depend on magnitude and region; a regional magnitude homogenisation with its uncertainty is needed before the recurrence model can be considered settled.

8.8 Slope screening with hazard-based shaking

Table 13: Newmark displacement (cm) for different critical accelerations, using the mean 475- and 2475-year PGAs and the scenario median (0.46 g) and 84th percentile (0.84 g). †: $a_c/a_{\max}$ outside roughly 0.05–0.40, a possible extrapolation of the regression. Regression scatter is about 0.5 log units.

a_c (g)	475-yr PGA	2475-yr PGA	Scenario median	Scenario 84th
0.02	22	345†	225†	577†
0.05	3.3	79	49	136
0.10	0.33†	23	13	42
0.20	0.0†	4.9	2.2†	11
0.30	0.0†	1.4†	0.38†	4.0

At the 475-year level only very weak slopes move by more than a few centimetres. At the 2,475-year level and in the scenario, slopes with a_c up to about 0.2 g move by a few centimetres or more. Because of the regression scatter, the extrapolated entries and the slope-specific nature of a_c, these figures identify where field investigation is needed; they are not predictions of failure.

9 Discussion: implications for the building

These points follow from the results; they are engineering considerations to confirm with a seismic engineer and a geotechnical specialist, not design values.

1. **Do not design from the record.** The largest estimated shaking among the events in the dataset (124 years of instrumental data and the incomplete NOAA historical record) is about 0.05 g; this is a property of the dataset, not a demonstrated bound on past shaking. The scenario gives 0.46 g at the median; the probabilistic 475-year value is 0.17 g (0.13 g if the fault is inactive, 0.20 g if active) and the 2,475-year value 0.61 g (0.36 to 0.70 g), all depending on the fault hypothesis (Table 8).

2. **The fault governs the long return periods.** If the fault is active (hypothesis A), an M7.0–7.7 rupture on the Karakoram Fault supplies about half of the exceedance rate at the 475-year level (blended tree) and the large majority from about 975 years upward. A mapped fault trace, an assessment of surface rupture and off-fault deformation, and a setback decision are essential at 1.6 km.

3. **Probabilistic and deterministic checks agree in order of magnitude.** The 2,475-year mean PGA (0.61 g blended, 0.70 g under A) lies between the scenario median (0.46 g) and the scenario 84th percentile (0.84 g). The 4,975-year value (0.83 g blended, 0.92 g under A) approaches the scenario 84th percentile; it is not a true upper bound, because the logic tree omits ground-motion-model, site, rupture-geometry and catalogue uncertainty. Comparing a mean hazard value with the scenario 84th percentile is only a plausibility check: that percentile is aleatory scatter for one assumed event, whereas the hazard bands are epistemic.

4. **Site effects are missing.** Values are for rock. The actual V_{S30}, and topographic amplification of a spur, can increase shaking and are not modelled.

5. **Spectral demand.** Only PGA is modelled. Structural design needs spectral accelerations at the building's periods, which require the full BSSA14 coefficient set (for example through `pygmm`) run through the same logic tree.

6. **Slope stability.** If the scenario occurred, the screening suggests that slopes with low critical accelerations could move by centimetres to tens of centimetres (Table 13); critical accelerations are slope-specific and need field measurement, and the regression is a screening tool.

7. **Reduce the biggest uncertainty.** Decide hypothesis A or B for the strand 1.6 km from the site by mapping it and assigning it to a published segment, with a slip-rate or paleoseismic constraint (trenching, offset dating). Until then, the envelope (0.20, 0.41 and 0.70 g at 475, 975 and 2,475 years) is a conditional screening case; it exceeds the blended tree but is not a complete upper bound. The choice between hypotheses moves the 475-year PGA between 0.13 and 0.20 g and the 2,475-year PGA between 0.36 and 0.70 g.

10 Limitations

- **Ground-motion model.** The embedded BSSA14 median and standard deviation reproduce `pygmm` (Section 4), but BSSA14 is one model for shallow crustal events in active regions. It has not been compared with a ground-motion model calibrated to the Himalaya, and the absolute accelerations, especially at long return periods, depend on that single choice. Himalayan models (for example Sharma et al., 2009; Anbazhagan et al., 2013; Bajaj and Anbazhagan, 2019) and the other NGA-West2 models (Abrahamson et al., 2014; Campbell and Bozorgnia, 2014; Chiou and Youngs, 2014) should form a ground-motion logic tree in a site-specific study. Because the 475-year and longer levels are dominated by earthquakes within 10 km, near-source behaviour (finite-fault geometry, directivity, saturation) matters more than agreement with other models at larger distances.

- **Magnitude heterogeneity.** Catalogue magnitudes (mostly mb) are treated as Mw. mb saturates for large events, and its relation to Mw depends on magnitude range, region and data source, so no single conversion applies; changes of magnitude type over time probably contribute to the back-test's M5.5–6.0 shortfall. In the fit window the small-magnitude bins are mostly mb and the large bins moment-type. Section 8.7 shows that a fit restricted to M$\geq$5.5 raises the 475-year PGA by 36%; a uniform $\pm$0.2 shift is a stress test, not a conversion. A full study needs a regionally appropriate magnitude homogenisation with its uncertainty carried through the logic tree.

- **Source zonation.** One Gutenberg–Richter relation and one smoothing kernel are applied across a 400 km zone that contains Himalayan thrust, Karakoram strike-slip, Tibetan extensional and deeper seismicity, which need not share recurrence behaviour, depth distribution or ground-motion characteristics. Removing the deeper events changes the hazard little, but separate source zones were not defined, and the M5.5–6.0 misfit and the clustering of the largest events (Section 7) are reasons to examine zonation rather than minor statistical details.

- **Historical record.** The historical component comes from a database of significant earthquakes, which is not a complete catalogue; older locations, magnitudes and source geometries are uncertain, so the largest estimated shaking in the dataset is not a verified bound on shaking at the site.

- **Completeness and sampling.** Rates for M$\geq$7 rest on three events since 1902 (M7.6 in 2005, M7.2 in 2008 and M7.9 in 1905 within 400 km). The NOAA record is incomplete and

uneven in time.

- **Poisson, time-independent.** The model ignores the time since the Karakoram Fault last ruptured and any partial creep; the characteristic-earthquake rate follows from slip-rate assumptions with no seismic confirmation, and the M7.7 branch is a potential magnitude rather than a measured one.
- **Slip rates and fault geometry.** The 0–4 mm/yr range rests on regional geodetic and geomorphic studies with the caveats in Section 7.8. Which strand lies at the site is unresolved, and the branch weights and the A/B split are judgements. The slip-budget rate assumes that every characteristic rupture reaches the site-adjacent reach and releases the average displacement there; segmentation and the probability that a rupture reaches the site are not modelled.
- **Simplifications.** Strike-slip mechanism everywhere; shallow depths assumed in the hazard grid; finite-fault distances approximated; smoothing length (50 km) and grid (10 km) chosen by judgement; the 84th percentile uses the BSSA14 $\sigma_{\ln}(M, R_{JB})$ evaluated at the scenario geometry.
- **Uncertainty bands.** The 16–84% bands are epistemic logic-tree spread only; they do not include ground-motion-model, site, geometry or catalogue uncertainty.
- **Regional comparison.** Regional PSHA studies of the north-western Himalaya (for example Bind et al., 2026) use area sources, many ground-motion models and different site definitions; no quantitative comparison with them is made here, and the 475-year value (0.17 g) should be reconciled with such studies in a site-specific analysis.
- **Not modelled.** Near-fault directivity and fling, surface rupture, liquefaction, site and topographic amplification, aftershock hazard, and spectral accelerations.
- **Unscored events.** Events beyond 400 km, including the very large events listed for context, are not in the hazard; long-period shaking from them is possible.

11 Conclusions

1. Among the events in the assembled dataset, the largest estimated shaking at the Agham spur is about 0.05 g (median PGA). This is a statement about the dataset, and it says little about the future, because the dataset contains no large earthquake on the nearby Karakoram Fault.
2. Conditional on a Karakoram Fault strand lying 1.6 km from the site, a characteristic M7.3 rupture would give a median rock PGA of 0.46 g (0.84 g at the 84th percentile), and screening suggests that low-strength slopes on the spur could move substantially. The fault proximity is an assumption, not a mapped result.
3. The probabilistic rock PGA is 0.043 g at 72 years and rises to 0.17, 0.33 and 0.61 g at 475, 975 and 2,475 years. Background seismicity controls the first, the fault supplies half of the rate at 475 years and 86% at 2,475 years.
4. Whether the strand next to the site is active, and at what rate, is the dominant uncertainty: the 2,475-year PGA is 0.70 g if it is active (2–4 mm/yr) and 0.36 g if it is inactive or very slow. The published literature does not settle which applies at the site; a fault study can.

5. These are provisional, rock-site PGA estimates from a single ground-motion model. The treatment of catalogue magnitudes is the largest remaining statistical uncertainty (a fit restricted to M≥5.5 raises the 475-year PGA to 0.23 g). A site-specific study should map the fault strand, add a ground-motion logic tree and spectral accelerations, homogenise magnitudes, separate the tectonic source zones, and include site response before any design value is adopted.

Data and code availability

The analysis is a single Python script, `quake_analysis_merged.py`, which reproduces every number in this paper:

```
python quake_analysis_merged.py \
    --catalog 4-to-10-further-away.csv \
    --historic earthquakes.json --outdir out
```

The fixed-scatter sensitivity uses the same command with `-sigma-model fixed`, and the catalogue sensitivities use `-merge-mag catalogue`, `-mb-offset` (±0.2), `-max-depth-km` (50 and 70) and `-completeness 4.5:none,5.0:none` (the M≥5.5 fit). Other options: `-gmpe {auto,pygmm,embedded,both}`, `-scenario-mag`, `-assume-hist-mag`, `-min-pga`, `-min-pga84`, `-smooth-km`, `-coupling`, `-completeness`, `-fault-trace`. Outputs in `out/`: figures 1–8; `pga_per_event.csv`; `historic_by_distance.csv`; `design_relevant_earthquakes.json` and `.geojson`; `future_earthquake_forecast.json` (hazard table, per-hypothesis PGA table, probabilities, disaggregation matrices, scenarios, slope table, back-test, hazard curves, design considerations); and `summary.txt`.

References

Abrahamson, N.A., Silva, W.J., Kamai, R. (2014). Summary of the ASK14 ground motion relation for active crustal regions. *Earthquake Spectra* 30(3), 1025–1055.

Anbazhagan, P., Kumar, A., Sitharam, T.G. (2013). Ground motion prediction equation considering combined dataset of recorded and simulated ground motions. *Soil Dynamics and Earthquake Engineering* 53, 92–108.

Bajaj, K., Anbazhagan, P. (2019). Regional stochastic GMPE with available recorded data for active region: application to the Himalayan region. *Soil Dynamics and Earthquake Engineering* 126.

Bind, A.P. et al. (2026). Surface-consistent probabilistic seismic hazard microzonation and damage potential modelling based on 1D/2D site characterization from Kashmir to North-Western Indian Himalaya. *Discover Hazards* 2, 35.

Bohon, W. et al. (2010). A new Quaternary strand of the Karakoram fault system, Ladakh Himalayas.

Boore, D.M., Stewart, J.P., Seyhan, E., Atkinson, G.M. (2014). NGA-West2 equations for predicting PGA, PGV and 5%-damped PSA for shallow crustal earthquakes. *Earthquake Spectra* 30(3), 1057–1085.

Brown, E.T. et al. (2002). Slip rates of the Karakorum Fault, Ladakh, India, determined using cosmic ray exposure dating of debris flows and moraines. *Journal of Geophysical Research* 107(B9).

Campbell, K.W., Bozorgnia, Y. (2014). NGA-West2 ground motion model for the average horizontal components of PGA, PGV and 5%-damped linear acceleration response spectra. *Earthquake Spectra* 30(3), 1087–1115.

Chiou, B.S.-J., Youngs, R.R. (2014). Update of the Chiou and Youngs NGA model for the average horizontal component of peak ground motion and response spectra. *Earthquake Spectra* 30(3), 1117–1153.

Cornell, C.A. (1968). Engineering seismic risk analysis. *Bulletin of the Seismological Society of America* 58(5), 1583–1606.

Frankel, A. (1995). Mapping seismic hazard in the central and eastern United States. *Seismological Research Letters* 66(4), 8–21.

Gardner, J.K., Knopoff, L. (1974). Is the sequence of earthquakes in southern California, with aftershocks removed, Poissonian? *Bulletin of the Seismological Society of America* 64(5), 1363–1367.

Gutenberg, B., Richter, C.F. (1944). Frequency of earthquakes in California. *Bulletin of the Seismological Society of America* 34(4), 185–188.

Jibson, R.W. (2007). Regression models for estimating coseismic landslide displacement. *Engineering Geology* 91(2–4), 209–218.

Mohanty, A., Gahalaut, V.K., Chowdhury, S., Bansal, A.K., Gautam, P., Catherine, J. (2024). Geodetic constraints on slip rate on the Karakoram fault and its role in the Himalayan arc deformation. *Earth and Planetary Science Letters* 626, 118512.

Ojha, A.K. (2026). Stop-and-go tectonics of the Karakoram Fault segments induced pulsating localised crustal extension. *Journal of Structural Geology* 207, 105695.

Sharma, M.L., Douglas, J., Bungum, H., Kotadia, J. (2009). Ground-motion prediction equations based on data from the Himalayan and Zagros regions. *Journal of Earthquake Engineering* 13(8), 1191–1210.

Wang, D., Elliott, J.R., et al. (2026). Constraining fault kinematics and India–Tarim convergence partitioning in southwestern Tibet with geodetic block modeling. *Geophysical Research Letters* 53, e2026GL123891.

Wang, M., Shen, Z.-K. (2020). Present-day crustal deformation of continental China derived from GPS and its tectonic implications. *Journal of Geophysical Research: Solid Earth* 125.

Weichert, D.H. (1980). Estimation of the earthquake recurrence parameters for unequal observation periods for different magnitudes. *Bulletin of the Seismological Society of America* 70(4), 1337–1346.

Wells, D.L., Coppersmith, K.J. (1994). New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement. *Bulletin of the Seismological Society of America* 84(4), 974–1002.

Wilson, R.C., Keefer, D.K. (1985). Predicting areal limits of earthquake-induced landsliding. In Ziony, J.I. (ed.), *Evaluating Earthquake Hazards in the Los Angeles Region*, USGS Professional Paper 1360, 317–345.

A The 47 design-relevant earthquakes

Ranked by median PGA at the site (rock). Rules: A (median PGA ≥ 0.02 g), B (84th percentile ≥ 0.05 g), C ($M \geq 6$ within 200 km), D ($M \geq 7$ within 500 km), E ($M \geq 5$ within 100 km), F (NOAA significant event within 500 km). Dates include the origin time (UTC, to the minute) for instrumental records. Depth is in km (blank for NOAA-only records); Epi. is the epicentral distance and R_{JB} the model distance used for scoring, which includes the finite-fault correction for $M \geq 6.5$. The 400 km scoring limit applies to R_{JB}, so the 1827 Lahore event (assumed magnitude, epicentre 444 km, R_{JB} 399 km) is scored; a dash in the R_{JB} column means the event is beyond the limit and not scored. An asterisk marks an assumed magnitude. The two M5.2 records of 23 August 1980 (ranks 33 and 35) are distinct events: they have different USGS identifiers (usp00019h8 and usp00019h9), origin times 14 minutes apart and different depths, and the NOAA database also lists two events on that day (M4.8 and M4.9).

#	Date (UTC)	Source	M	Depth	Epi.	R_{JB}	PGA (g)	Rules	Place
1	1941-01-10 07:38	USGS	5.7	15	34	34	0.049	ABE	40 km ESE of Leh, India
2	1905-04-04 00:49	USGS/NOAA	7.9	20	208	122	0.040	ABDF	The 1905 Kangra, India Earthquake
3	1937-11-15 21:37	USGS	6.35	30	62	62	0.035	ABCE	86 km NE of Leh, India
4	1885-06-06	NOAA	7.5*		168	123	0.029	ABCDF	India: Kashmir
5	1996-11-19 10:44	USGS	6.9	33	119	101	0.025	AC	130 km ENE of Thang, India
6	2020-09-25 10:57	USGS	5.3	10	39	39	0.024	AE	58 km E of Leh, India
7	1555-09	NOAA	7.8		274	202	0.015	DF	INDIA: KASHMIR: SRINAGAR (1555 Kashmir e...
8	2020-09-27 17:27	USGS	5.0	10	39	39	0.013	E	52 km E of Leh, India
9	1968-02-10 17:03	USGS	5.3	15	80	80	0.009	E	93 km E of Leh, India
10	1926-08-06 22:45	USGS	6.2	10	145	145	0.008	C	169 km NE of Leh, India
11	1968-02-11 20:38	USGS	5.3	15	86	86	0.008	E	101 km E of Leh, India
12	1975-01-19 08:02	USGS/NOAA	6.8	33	216	201	0.007	F	120 km E of Manali, India
13	1828-06-26	NOAA	7.5*		302	257	0.007	DF	India: Kashmir
14	2007-12-11 15:56	USGS	5.1	16	80	80	0.006	E	62 km E of Thang, India
15	844	NOAA	7.0		277	257	0.004	DF	India: Kashmir (Srinagar area)
16	1123	NOAA	7.0		277	257	0.004	DF	India: Kashmir (Srinagar area)
17	1669-06-23	NOAA	7.0		277	257	0.004	DF	India: Kashmir: Srinagar
18	1779	NOAA	7.0		277	257	0.004	DF	India: Kashmir (Srinagar area)
19	1784	NOAA	7.0		277	257	0.004	DF	India: Kashmir (Srinagar area)
20	2005-10-08 03:50	USGS/NOAA	7.6	26	386	333	0.003	DF	21 km NNE of Muzaffarabad, Pakistan
21	1501-09-24	NOAA	6.75		277	263	0.003	F	India: Kashmir (Srinagar area)
22	1736	NOAA	6.75		277	263	0.003	F	India: Kashmir (Srinagar area)
23	1803-09-01	NOAA	7.5		386	341	0.003	DF	India: Uttarakhand, Mathura (Muttra)
24	1678	NOAA	6.65		277	265	0.003	F	India: Kashmir (Srinagar area)
25	1683	NOAA	6.65		277	265	0.003	F	India: Kashmir (Srinagar area)
26	2008-03-20 22:32	USGS	7.2	10	360	332	0.002	D	226 km SE of Hotan, China
27	1827-09-26	NOAA	7.5*		444	399	0.002	DF	Pakistan: Lahore, Punjab
28	1885-05-30	NOAA	6.3		286	286	0.002	F	India: Kashmir: Srinagar
29	1863	NOAA	6.0		277	277	0.001	F	India: Kashmir (Srinagar area)

#	Date (UTC)	Source	M	Depth	Epi.	R_{JB}	PGA (g)	Rules	Place
30	1991-10-19 21:23	USGS/NOAA	7.0	10	404	384	0.001	DF	32 km E of Uttarkashi, India
31	2002-11-20 21:32	USGS/NOAA	6.3	33	322	322	0.001	F	58 km SSE of Gilgit, Pakistan
32	1986-04-26 07:35	USGS/NOAA	5.5	33	277	277	0.001	F	11 km ENE of Kangar, India
33	1980-08-23 21:36	USGS/NOAA	5.2	25	253	253	0.001	F	10 km SW of Bhadarwah, India
34	1963-09-02	NOAA	5.3	25	279	279	0.001	F	India: Kashmir: Sw
35	1980-08-23 21:50	USGS/NOAA	5.2	33	259	259	0.001	F	18 km SSW of Bhadarwah, India
36	2002-11-01 22:09	USGS/NOAA	5.4	33	315	315	0.000	F	54 km SE of Gilgit, Pakistan
37	1998-05-28 21:11	USGS/NOAA	5.6	33	355	355	0.000	F	13 km NNE of Zangguy, China
38	2006-03-10 07:50	USGS/NOAA	4.9	10	384	384	0.000	F	12 km E of New Mirpur, Pakistan
39	2005-08-16 03:45	USGS/NOAA	4.4	43	384	384	0.000	F	24 km NNE of Uttarkashi, India
40	1981-09-12 07:15	USGS/NOAA	6.2	33	411	–	–	F	69 km WSW of Gilgit, Pakistan
41	2004-02-14 10:30	USGS/NOAA	5.5	11	422	–	–	F	20 km ENE of Battagram, Pakistan
42	2006-04-04 09:12	USGS/NOAA	4.8	10	427	–	–	F	13 km SE of Battagram, Pakistan
43	1972-09-04 13:42	USGS/NOAA	6.2	15	441	–	–	F	38 km WSW of Gahkuch, Pakistan
44	2005-12-14 07:09	USGS/NOAA	5.3	44	449	–	–	F	17 km WNW of Papalkoti, India
45	1999-03-28 19:05	USGS/NOAA	6.6	15	449	–	–	F	9 km NNW of Papalkoti, India
46	1974-12-28 12:46	USGS/NOAA	6.2	33	453	–	–	F	50 km N of Shingli Bala, Pakistan
47	1909-07-08	NOAA	7.5*		500	–	–	DF	Pakistan