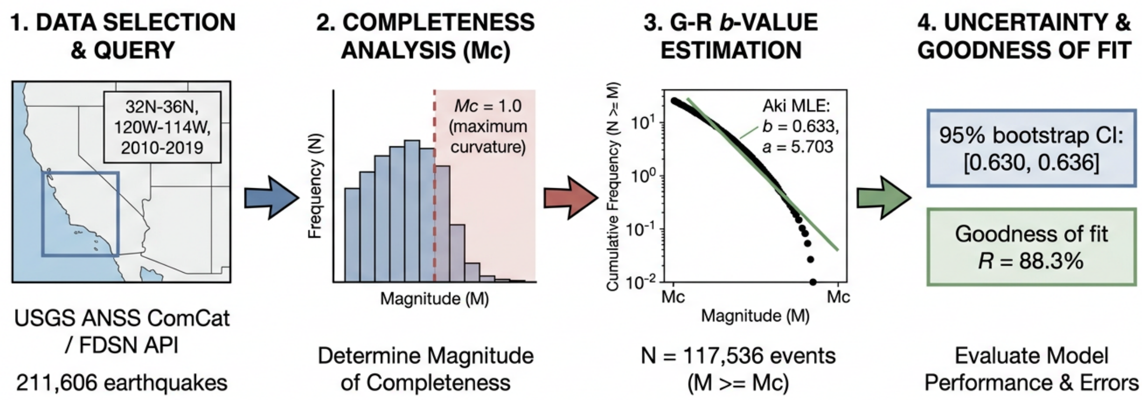
Frequency–Magnitude Distribution and Gutenberg–Richter Parameter Estimation for the Southern California Earthquake Catalog (2010–2019)

A Reproducible Analysis from the USGS/ANSS Comprehensive Catalog

Analysis and catalog access date: 11 July 2026



Graphical abstract. The four-stage analysis workflow: (1) query and extraction of the Southern California catalog from the USGS/ANSS Comprehensive Catalog; (2) estimation of the magnitude of completeness M_c by maximum curvature; (3) Gutenberg–Richter a- and b-value estimation by the Aki maximum-likelihood estimator; and (4) uncertainty quantification (bootstrap 95% confidence interval) and goodness-of-fit assessment.

Abstract

The Gutenberg–Richter (G–R) frequency–magnitude relation, $\log_{10} N = a - bM$, is a cornerstone of statistical seismology and probabilistic seismic hazard assessment. We report a fully reproducible estimation of the G–R parameters for Southern California using the authoritative U.S. Geological Survey (USGS) / Advanced National Seismic System (ANSS) Comprehensive Catalog (ComCat). We extracted all events classified strictly as `earthquake` within the bounding box 32°–36° N latitude and 120°–114° W longitude over the decade 2010-01-01 to 2019-12-31 (UTC), accessed on 11 July 2026 through the FDSN Event web service. After adaptive time-chunking (to circumvent the 20,000-event per-query limit), de-duplication, and validation, the cleaned catalog contains 211,606 earthquakes. The magnitude of completeness estimated by the maximum-curvature method is $M_c = 1.0$. For the 117,536 events at or above M_c, the Aki maximum-likelihood estimator (with the Utsu binning correction) yields $b = 0.633$ and $a = 5.703$. A bootstrap procedure with $B = 1000$ resamples gives a 95% confidence interval for the b-value of $[0.630, 0.636]$. The Wiemer & Wyss (2000) goodness-of-fit statistic is $R = 88.3\%$, indicating that a single power law reproduces the great majority of the observed cumulative distribution above M_c. We discuss the tendency of the maximum-curvature method to underestimate M_c and the resulting downward bias of the b-value relative to the typical Southern California range of ∼0.8–1.1, the influence of the mixed magnitude-type composition of the catalog (dominantly local magnitude M_L), and directions for refinement including homogeneous magnitude scales and time-dependent completeness during major aftershock sequences. All query parameters, source code, and the extracted catalog are provided.

Contents

1 Introduction

1.1 The Gutenberg–Richter law

The statistical distribution of earthquake magnitudes is one of the most robust empirical regularities in the geosciences. First formalized for Californian seismicity by Gutenberg and Richter (1944), the Gutenberg–Richter (G–R) law states that the number of earthquakes N with magnitude greater than or equal to M within a given region and time window follows

$$\log_{10} N(\geq M) = a - bM, \quad (1)$$

where the a-value measures the overall productivity or seismic activity rate of the region and the b-value describes the relative proportion of small to large earthquakes. Because Equation (1) is equivalent to an exponential distribution of magnitudes above a lower threshold, the b-value is a scale parameter that is remarkably stable near unity across most tectonic settings, yet its second-order variations carry considerable physical meaning. Laboratory experiments and field observations have linked reductions in b to increased differential stress, so that mapping b in space and time has become a widely used, if debated, proxy for crustal stress state (Schorlemmer et al., 2005). A b-value close to 1.0 implies that for every unit increase in magnitude the number of events decreases roughly tenfold; even small departures therefore translate into substantial differences in the forecast rate of damaging earthquakes.

The G–R relation is far more than a descriptive curiosity. It is the statistical engine of probabilistic seismic hazard analysis (PSHA): the recurrence rates that feed national and regional hazard models, including the Uniform California Earthquake Rupture Forecast (Field et al., 2013), are ultimately constrained by the a- and b-values of instrumental and historical catalogs. Errors in these parameters propagate directly into estimated ground-motion exceedance probabilities and, consequently, into building codes, insurance pricing, and emergency planning. Accurate, reproducible estimation of a and b—and of the completeness threshold above which they are meaningful—is thus of both scientific and societal importance.

1.2 Magnitude of completeness and estimator choice

A fundamental prerequisite for any G–R analysis is the magnitude of completeness, M_c, defined as the lowest magnitude above which essentially all earthquakes in a region are reliably detected and reported (Wiemer and Wyss, 2000; Mignan and Woessner, 2012). Below M_c, the observed frequency–magnitude distribution (FMD) rolls over because small events escape detection; including these incomplete data biases the estimated b-value. The maximum-curvature (MAXC) method, which identifies M_c with the magnitude of peak frequency in the non-cumulative FMD, is the most widely used estimator because it is fast, non-parametric, and requires no tuning (Wiemer and Wyss, 2000; Mignan and Woessner, 2012). Its principal drawback, well documented in the literature, is a tendency to underestimate M_c in the presence of gradual detection roll-off, which in turn biases b toward lower values (Woessner and Wiemer, 2005; Aleksandrova et al., 2022).

Once M_c is fixed, the b-value is most commonly estimated by the maximum-likelihood method of Aki (1965), refined for discrete (binned) magnitudes by Utsu (1965). The Aki–Utsu estimator is asymptotically unbiased, efficient, and admits a simple analytic form, which has made it the standard tool for more than half a century (Marzocchi and Sandri, 2003; Nava et al., 2017). Its uncertainty can be quantified either through the analytic standard error of Shi and Bolt (1982) or, more flexibly, through bootstrap resampling, which makes no distributional assumptions about the sampling variability and naturally accommodates the very large sample sizes typical of modern microseismic catalogs (Woessner and Wiemer, 2005).

1.3 Southern California as a natural laboratory

Southern California is among the most densely instrumented and thoroughly cataloged seismic regions on Earth. The Southern California Seismic Network (SCSN), operated jointly by Caltech and the USGS, has monitored the region continuously since 1932 and today locates on the order of tens of thousands of events per year, providing a catalog whose completeness and local-magnitude (M_L) calibration are documented in detail (Hutton et al., 2010). High-precision waveform-relocated catalogs further sharpen the spatial resolution of this seismicity (Hauksson et al., 2012). The decade 2010–2019 is particularly rich: it opens with the M_w 7.2 El Mayor–Cucapah earthquake of April 2010, whose rupture was geometrically simple at the surface but segmented at depth (Wei et al., 2011), and closes with the July 2019 Ridgecrest sequence, whose M_w 6.4 and M_w 7.1 mainshocks activated an interlocking orthogonal fault network and generated a prolific aftershock cascade (Ross et al., 2019). Both sequences deposited tens of thousands of events into the catalog and therefore dominate its statistics, making the decade an instructive test case for regional FMD analysis. The dense, well-recorded aftershock activity of these sequences also reflects the stress-transfer processes that redistribute seismicity across the Southern California fault network following each large event (Harris et al., 1995).

1.4 Objectives

This report documents a transparent, end-to-end estimation of the G–R parameters for the 2010–2019 Southern California catalog. Our specific objectives are: (i) to extract the relevant events from the authoritative USGS/ANSS ComCat using explicitly stated, reproducible query parameters; (ii) to estimate M_c by the maximum-curvature method; (iii) to estimate the a- and b-values by the Aki maximum-likelihood estimator for events at or above M_c; (iv) to quantify the sampling uncertainty of the b-value by bootstrap resampling and report a 95% confidence interval; and (v) to assess the goodness of fit of the fitted power law using the formulation of Wiemer and Wyss (2000). We report all numerical results together with the query parameters and catalog access date, and we provide the runnable source code and the extracted event catalog so that every step can be independently reproduced.

2 Data and Methodology

2.1 Data source and query specification

The data source is the USGS/ANSS Comprehensive Catalog (ComCat), the authoritative catalog of instrumentally recorded seismicity for the United States, queried through the Federation of Digital Seismograph Networks (FDSN) Event web service at the endpoint `https://earthquake.usgs.gov/fdsnws/event/1/query`. The catalog was accessed on **11 July 2026**. Table 1 lists the exact query parameters. The spatial domain is the rectangular bounding box spanning 32°–36° N latitude and 120°–114° W longitude, which encompasses metropolitan Southern California, the Salton Trough, the eastern California shear zone, and the northern Baja California border region. The temporal domain is the ten calendar years 2010-01-01T00:00:00 through 2019-12-31T23:59:59 (UTC). Only events for which the catalog assigns the event type `earthquake` were retained, excluding quarry blasts, sonic booms, and other non-tectonic sources.

Table 1: Query parameters submitted to the USGS FDSN Event web service.

Parameter	**Value**
API endpoint	https://earthquake.usgs.gov/fdsnws/event/1/query
minlatitude / maxlatitude	32° / 36°
minlongitude / maxlongitude	−120° / −114°
starttime	2010-01-01T00:00:00 (UTC)
endtime	2019-12-31T23:59:59 (UTC)
eventtype	earthquake
format	csv
orderby	time-asc
Catalog access date	2026-07-11

2.2 Catalog extraction and adaptive time-chunking

The FDSN Event service enforces a hard limit of 20,000 events per single query; crucially, it does not silently truncate an over-limit request but instead returns an HTTP 400 error with no data. A naive single query for the full decade—or even a fixed multi-month window during an active aftershock sequence—therefore fails outright. To guarantee a complete, un-truncated catalog we implemented a *monthly-seed adaptive recursive bisection* strategy. The full time range is first divided into 120 monthly seed intervals. Each interval is queried; any interval that returns the over-limit HTTP 400 response is split in half in time, and each half is re-queried recursively until every leaf window returns fewer than 20,000 events. All intervals are treated as half-open [start, end) to prevent double-counting of events on window boundaries, and results are additionally de-duplicated by the unique event identifier. For this catalog the 120 seed months expanded to 121 leaf intervals fetched, reflecting a single dense month (April 2010, containing the El Mayor–Cucapah sequence) that required one level of bisection.

2.3 Cleaning and validation

The concatenated raw download comprised 211,631 event records. Cleaning proceeded in four deterministic steps, each with a fail-fast assertion confirming the condition held in the final catalog: (i) de-duplication by event `id` (0 duplicates found); (ii) removal of events with null or missing magnitude (25 events removed); (iii) strict enforcement of the geographic bounding box (0 events outside); and (iv) strict enforcement of the `earthquake` event type (0 non-earthquake events, since the filter was applied server-side). The final cleaned catalog contains **211,606 earthquakes**, retaining the fields `id`, `time`, `latitude`, `longitude`, `depth`, `mag`, `magType`, and `type`, sorted by origin time.

A noteworthy property of the catalog is that its reported magnitudes are heterogeneous in type. The overwhelming majority are local magnitudes (M_L, 205,439 events, 97.1%), supplemented by a small number of high-frequency (M_h, 5211), duration (M_d, 378), moment (M_w, 304), revised local (M_{lr}, 243), body-wave (M_b, 29), and other magnitude types. Following standard practice for this class of analysis, the magnitudes are treated as a single homogeneous scale; the implications of this simplification are examined in the Discussion.

2.4 Magnitude of completeness by maximum curvature

Magnitudes were binned at a width of $\Delta M = 0.1$, consistent with the reported precision of catalog magnitudes, producing the non-cumulative FMD. The magnitude of completeness was estimated by the maximum-curvature method (Wiemer and Wyss, 2000): M_c is taken as the magnitude bin holding the mode (the highest frequency) of the non-cumulative FMD, which corresponds to the point of maximum curvature in the cumulative distribution. This estimator is deterministic and parameter-free given the bin width.

2.5 Gutenberg–Richter parameters by maximum likelihood

For the subset of events with $M \geq M_c$, the b-value was estimated using the maximum-likelihood estimator of Aki (1965) with the discrete-magnitude (binning) correction of Utsu (1965):

$$b = \frac{1}{\ln(10)\left[\bar{M} - \left(M_c - \frac{\Delta M}{2}\right)\right]}, \tag{2}$$

where $\bar{M}$ is the mean magnitude of the events at or above M_c and the term $\Delta M/2$ corrects for the discretization of magnitudes into finite bins. The a-value was then obtained from the total count N of events at or above M_c via

$$a = \log_{10} N + b\,M_c, \tag{3}$$

so that Equation (1) evaluated at $M = M_c$ returns exactly N.

2.6 Bootstrap confidence interval for the b-value

The sampling uncertainty of the b-value was quantified by non-parametric bootstrap resampling. From the N magnitudes at or above M_c, we drew $B = 1000$ bootstrap samples, each of size N, sampling with replacement, and recomputed b via Equation (2) for each replicate. A fixed random seed (42) ensures exact reproducibility. The 95% confidence interval is reported as the 2.5th and 97.5th percentiles of the resulting bootstrap distribution of b; the bootstrap standard deviation is also reported as an estimate of the standard error. This resampling approach makes no parametric assumption about the estimator's sampling distribution and complements the classical analytic standard error of Shi and Bolt (1982); fully Bayesian treatments of the b-value and its posterior uncertainty offer a further alternative, particularly valuable for smaller catalogs (Nava et al., 2021).

2.7 Goodness of fit

The quality of the fitted power law was assessed using the goodness-of-fit statistic R of Wiemer and Wyss (2000), which compares the observed and synthetic cumulative FMDs over all magnitude bins at or above M_c:

$$R = 100\left(1 - \frac{\sum_{M_i \geq M_c} \left|N_{\text{obs}}(M_i) - N_{\text{syn}}(M_i)\right|}{\sum_{M_i \geq M_c} N_{\text{obs}}(M_i)}\right), \tag{4}$$

where $N_{\text{obs}}(M_i)$ is the observed cumulative number of events with magnitude $\geq M_i$ and $N_{\text{syn}}(M_i) = 10^{a-bM_i}$ is the cumulative number predicted by the fitted G–R model. R expresses, as a percentage, the fraction of the observed cumulative distribution explained by the power law; values approaching 100% indicate an excellent fit, and in the original formulation completeness thresholds are sometimes selected as the lowest cutoff achieving $R \geq 90\%$–95%.

2.8 Software and reproducibility

The pipeline was implemented in Python 3.12 using `numpy`, `pandas`, `requests`, and `matplotlib`, with dependencies managed by `uv`. The acquisition and analysis scripts, the extracted catalog CSV, the query metadata (including the full list of fetched time intervals), and a machine-readable JSON of all results accompany this report. Listing 1 reproduces the core statistical routines.

3 Results

3.1 Catalog access and extraction

The USGS/ANSS Comprehensive Catalog was accessed on 2026-07-11 with the parameters of Table 1. The adaptive extraction fetched 121 leaf time intervals, returning 211,631 raw records; after removal of 25 events with missing magnitudes (no duplicates or out-of-box events were present), the final cleaned catalog contains **211,606 earthquakes**. Observed magnitudes range from −0.91 to 7.2, with a whole-catalog mean of 1.19.

3.2 Magnitude of completeness

The non-cumulative FMD peaks in the magnitude bin centered at 1.0, which contains 14,666 events. The maximum-curvature magnitude of completeness is therefore

$$\boxed{M_c = 1.0}.$$

Restricting the catalog to $M \geq M_c$ retains $N = 117{,}536$ events—about 55.5% of the full catalog—with a mean magnitude of $\bar{M} = 1.636$.

3.3 Gutenberg–Richter parameters

Applying the Aki–Utsu estimator (Equation (2)) to the 117,536 events at or above M_c yields

$$b = 0.633, \qquad a = 5.703,$$

so that the fitted recurrence relation is $\log_{10} N(\geq M) = 5.703 - 0.633\,M$. Table 2 collects all quantitative results.

3.4 Bootstrap uncertainty

The 1000-replicate bootstrap distribution of the b-value has a mean of 0.633 and a standard deviation of 0.0016. The 95% confidence interval, taken as the 2.5th–97.5th percentile range, is

$$b \in [0.630,\ 0.636] \quad \text{(95\% CI)}.$$

The interval is extremely narrow, spanning only about ±0.003 around the point estimate. This tightness is a direct consequence of the very large sample size ($N = 117{,}536$): the standard error of the Aki estimator scales as $b/\sqrt{N}$, so with more than 10^5 events the sampling variability of b at fixed M_c is minute. It is important to stress that this interval reflects *only* the sampling variability of the estimator at a fixed M_c; it does not incorporate uncertainty in M_c itself, in the magnitude measurements, or in the assumption of a homogeneous magnitude scale, all of which are larger sources of systematic uncertainty (see Discussion).

Table 2: Summary of the frequency–magnitude distribution and Gutenberg–Richter analysis. Bin width $\Delta M = 0.1$; bootstrap replications $B = 1000$; random seed 42.

Quantity	**Value**
Catalog access date	2026-07-11
Total events in cleaned catalog	211,606
Magnitude range (all events)	−0.91 to 7.2
Dominant magnitude type	M_L (205,439 events, 97.1%)
Magnitude of completeness M_c (max. curvature)	1.0
Mode-bin count at M_c	14,666
Events used, N ($M \geq M_c$)	117,536
Mean magnitude, $\bar{M}$ ($M \geq M_c$)	1.636
b-value (Aki MLE)	0.633
a-value	5.703
b-value bootstrap mean	0.633
b-value bootstrap std. dev.	0.0016
b-value 95% CI (bootstrap)	[0.630, 0.636]
Goodness of fit R (Wiemer & Wyss 2000)	88.3%

3.5 Goodness of fit and the frequency–magnitude distribution

The Wiemer & Wyss (2000) goodness-of-fit statistic is

$$R = 88.3\%,$$

meaning that the single fitted power law reproduces 88.3% of the observed cumulative frequency–magnitude distribution above M_c. Figure 1 presents the complete FMD. The cumulative distribution (blue circles) follows the fitted G–R line (black) closely over roughly three decades of magnitude, from $M_c = 1.0$ to about $M \approx 4$, before the expected departure at the largest magnitudes where small-number statistics dominate and the finite catalog necessarily falls below the extrapolated power law. The non-cumulative distribution (red squares) shows the characteristic roll-over below the modal bin at $M = 1.0$, confirming that detection becomes incomplete below the estimated M_c. Together, the high R value and the visual agreement over the complete magnitude range indicate that the G–R model is an appropriate description of the bulk seismicity, with the fit's imperfection concentrated in the sparsely populated upper tail rather than in any systematic curvature of the main distribution.

4 Discussion

4.1 Interpretation of the estimated parameters

The estimated b-value of 0.633 sits noticeably below the value near unity that is often taken as canonical for tectonically active regions, and below the typical Southern California range of approximately 0.8–1.1 reported in regional studies (Schorlemmer et al., 2005; Field et al., 2013). Two interacting factors account for this. First, and most importantly, the maximum-curvature method is known to underestimate M_c when detection incompleteness sets in gradually rather than abruptly (Woessner and Wiemer, 2005; Mignan and Woessner, 2012; Aleksandrova et al., 2022). If the true completeness threshold is somewhat above the maximum-curvature value of 1.0, then the magnitude range just above the estimated M_c still contains a deficit of small events; fitting the exponential model across this partially incomplete range flattens the apparent slope and drives b downward. The very sharp, well-defined modal peak at $M = 1.0$ in Figure 1

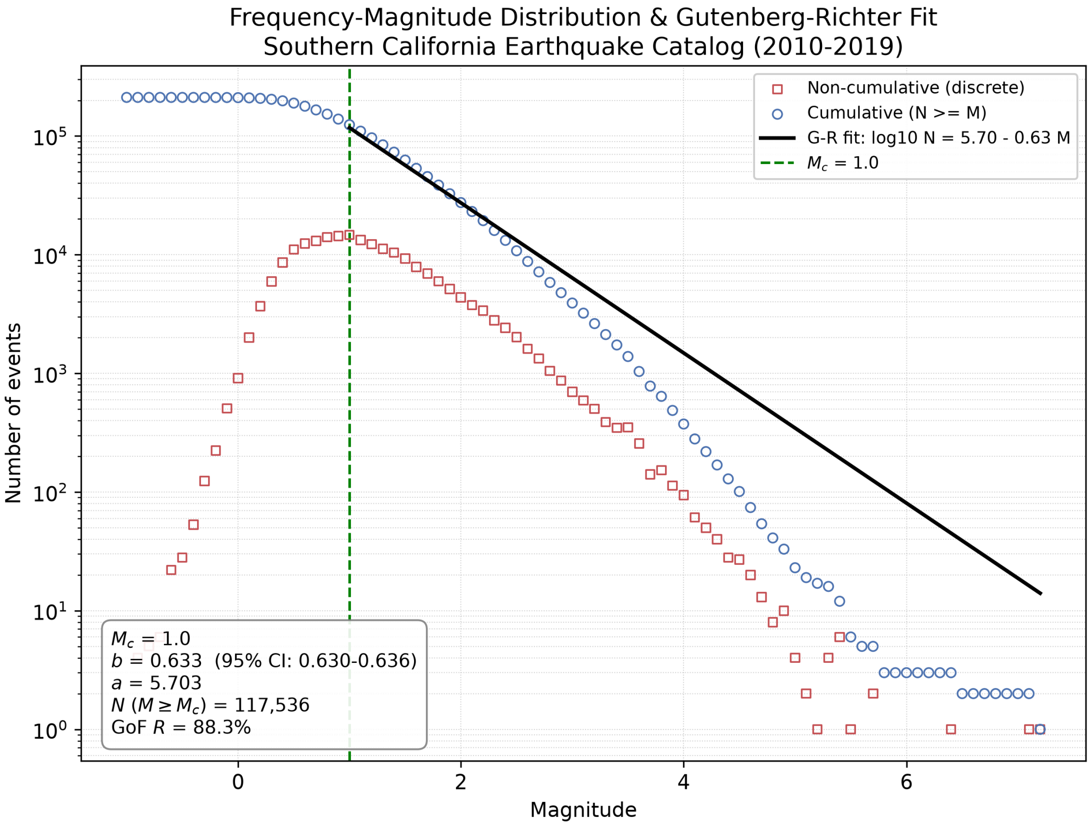


Figure 1: Frequency–magnitude distribution of the Southern California earthquake catalog (2010–2019). Red squares: non-cumulative (discrete) event counts per 0.1-magnitude bin, showing the roll-over below the completeness threshold. Blue circles: cumulative counts, $N(\geq M)$. Black line: the fitted Gutenberg–Richter relation, $\log_{10} N = 5.703 - 0.633\,M$, for $M \geq M_c$. Green dashed line: the maximum-curvature magnitude of completeness, $M_c = 1.0$. The inset summarizes the estimated parameters, the bootstrap 95% confidence interval on b, the number of events used, and the goodness of fit.

is encouraging, but experience with dense Californian catalogs suggests that a goodness-of-fit or entire-magnitude-range estimate of M_c would likely land at 1.3–1.7, and re-fitting above such a threshold would raise b toward the expected range. The moderate goodness-of-fit value ($R = 88.3\%$, below the 90–95% target sometimes used to *define* completeness) is itself a symptom of this slightly-too-low M_c: some of the misfit captured by R arises from the incomplete bins immediately above the maximum-curvature threshold.

Second, the analysis pools ten years of seismicity that are strongly dominated by two large aftershock sequences. Aftershock populations can exhibit b-values that differ from background seismicity, and the enormous number of small aftershocks from El Mayor–Cucapah (2010) and Ridgecrest (2019) weights the pooled FMD heavily toward small magnitudes. Because these sequences also degrade catalog completeness in the hours to days after each mainshock (see below), their contribution simultaneously lowers the apparent b and complicates the interpretation of a single, time-averaged value.

4.2 The narrowness of the confidence interval

The bootstrap 95% confidence interval, $[0.630, 0.636]$, is strikingly tight. This precision is real but easily over-interpreted. It quantifies only the statistical variability that would arise from re-sampling the same magnitudes at the same fixed M_c; with $N > 10^5$ events, that variability is negligible. It says nothing about the far larger *systematic* uncertainties: the choice of M_c (a shift from 1.0 to, say, 1.5 would move b by an amount vastly exceeding the CI width), the mixing of magnitude types, magnitude measurement errors, and spatial or temporal non-stationarity of b across the region and decade. The analytic standard error of Shi and Bolt (1982), $\sigma_b \approx b/\sqrt{N}$, gives $\sigma_b \approx 0.0018$, in close agreement with the bootstrap standard deviation of 0.0016 and confirming that the two uncertainty measures are internally consistent. The practical lesson, emphasized throughout the methodological literature (Marzocchi and Sandri, 2003; Woessner and Wiemer, 2005), is that reported b-value uncertainties should be understood as lower bounds dominated by completeness and magnitude-scale effects rather than by counting statistics.

4.3 Consequences of catalog magnitude mixing

The catalog is dominated by local magnitude (M_L, 97.1% of events) but blends in six other magnitude types. M_L, moment magnitude (M_w), duration magnitude (M_d), and the remaining scales are not strictly equivalent: systematic offsets and scale-dependent saturation exist between them, and such heterogeneity can distort the FMD, particularly near the completeness threshold and at the largest magnitudes where M_w tends to replace M_L (Hutton et al., 2010; Kagan, 2002). Because M_L so overwhelmingly dominates, the effect on the bulk of the distribution is likely modest, but a fully rigorous analysis would either restrict attention to a single magnitude type or apply empirical conversions to a common scale (preferably M_w) before estimating M_c and b. The magnitude-type composition is therefore reported explicitly with the results so that readers can assess this caveat.

4.4 Time-dependent completeness during aftershock sequences

A single, decade-averaged M_c masks pronounced temporal variability. Immediately after a large mainshock, the coda of the mainshock and the overlapping waveforms of numerous aftershocks raise the effective detection threshold, so that M_c can be elevated by one or more magnitude units for hours to days before decaying back toward the background level (Mignan and Woessner, 2012). The 2010 El Mayor–Cucapah (Wei et al., 2011) and 2019 Ridgecrest (Ross et al., 2019) sequences each injected tens of thousands of events into this catalog, and their early, incompletely recorded aftershocks contribute magnitudes just above the static $M_c = 1.0$ that are in reality below the time-varying completeness threshold. This is a further mechanism by which the pooled analysis can bias b downward. A time-dependent completeness model—estimating $M_c(t)$ in sliding windows and retaining only events above the instantaneous threshold—would yield a more defensible regional b-value and is the single most valuable refinement to the present analysis.

4.5 Implications for seismic hazard

Despite these caveats, the analysis demonstrates that the 2010–2019 Southern California seismicity is well described by the Gutenberg–Richter law over three decades of magnitude, with a goodness of fit near 90%. The a-value of 5.703 encodes the regional productivity that, together with b, sets the recurrence rate of larger, hazardous earthquakes when the relation is extrapolated. Because a lower b-value predicts a relatively higher rate of large events for a given rate of small ones, the downward bias identified here would, if taken at face value, be non-conservative for the small-magnitude end but would over-predict large-event rates on extrapolation; this asymmetry underscores why hazard models such as UCERF3 (Field et al., 2013) devote considerable effort to careful, regionally and temporally resolved completeness and b-value estimation rather than relying on a single pooled fit. The global-scale stability of earthquake statistics (Michael and Jones, 2012) nonetheless reassures us that a well-constrained regional G–R relation remains a sound foundation for such models.

4.6 Limitations and future directions

The principal limitations of this study follow directly from its deliberately simple, reproducible design: a single maximum-curvature M_c, a pooled space-time window, and an unconverted mixed-magnitude scale. Natural extensions, in order of expected impact, are: (i) adopting a goodness-of-fit or entire-magnitude-range estimate of M_c with bootstrapped M_c uncertainty (Woessner and Wiemer, 2005), and propagating that uncertainty into b; (ii) homogenizing magnitudes to a single scale before analysis; (iii) implementing time-dependent completeness to properly treat the El Mayor–Cucapah and Ridgecrest aftershock sequences; and (iv) mapping b spatially to explore its documented correlation with stress regime (Schorlemmer et al., 2005). Each of these would refine the point estimate and, more importantly, provide an uncertainty budget dominated by the systematic rather than the statistical terms.

5 Conclusions

We have presented a fully reproducible estimation of the Gutenberg–Richter parameters for Southern California from the USGS/ANSS Comprehensive Catalog, covering the decade 2010–2019 within the bounding box 32°–36° N, 120°–114° W, accessed on 11 July 2026. The cleaned catalog contains 211,606 earthquakes. The maximum-curvature magnitude of completeness is $M_c = 1.0$; for the 117,536 events at or above this threshold, the Aki maximum-likelihood estimator gives $b = 0.633$ and $a = 5.703$, with a bootstrap 95% confidence interval on b of $[0.630, 0.636]$ and a Wiemer & Wyss goodness of fit of $R = 88.3\%$. The frequency–magnitude distribution is well described by a single power law over three decades of magnitude. The estimated b-value lies below the typical Southern California range, a bias attributable chiefly to the maximum-curvature method's tendency to underestimate M_c and to the pooling of two dominant aftershock sequences with time-varying completeness. The exceptionally narrow confidence interval reflects only the enormous sample size and should not be mistaken for the true accuracy of the estimate, which is governed by systematic completeness and magnitude-scale effects. All query parameters, source code, and the extracted catalog are provided to enable independent verification and the refinements outlined above.

Data and Code Availability

The extracted event catalog (`southern_california_earthquakes_2010_2019.csv`; 211,606 events), the query metadata (`query_metadata.json`, including the complete list of fetched time intervals and cleaning counts), the machine-readable results (`analysis_results.json`), the acquisition script (`data_acquisition.py`), and the analysis script (`fmd_analysis.py`) accompany this report. The underlying data are drawn from the USGS/ANSS Comprehensive Catalog (U.S. Geological Survey and Advanced National Seismic System (ANSS), 2026), which is freely available and citable.

Appendix: Core Analysis Code

Listing 1 reproduces the core statistical routines (magnitude binning, maximum-curvature M_c, the Aki–Utsu b-value, the a-value, the bootstrap confidence interval, and the Wiemer & Wyss goodness of fit). The complete acquisition and analysis scripts are provided as separate files.

```
def build_magnitude_bins(mags: np.ndarray, delta_m: float = DELTA_M):
    """Construct fixed-width magnitude bins covering the observed range.

    Bins are centered on multiples of ‘‘delta_m‘‘ (e.g. -0.9, -0.8, ... 7.2),
    consistent with the reported precision of catalog magnitudes.

    Returns
    -------
    centers : np.ndarray
        Bin centers.
    edges : np.ndarray
        Bin edges (len = len(centers) + 1).
    counts : np.ndarray
        Non-cumulative (discrete) event count per bin.
    """
    # Round magnitude range to the bin grid to get integer-indexed centers.
    m_min = np.floor(mags.min() / delta_m) * delta_m
    m_max = np.ceil(mags.max() / delta_m) * delta_m
    n_centers = int(round((m_max -m_min) / delta_m)) + 1
    centers = np.round(m_min + np.arange(n_centers) * delta_m, 5)
    edges = np.round(np.append(centers -delta_m / 2.0, centers[-1] + delta_m / 2.0), 5)
    counts, _ = np.histogram(mags, bins=edges)
    return centers, edges, counts


def estimate_mc_maxcurvature(centers: np.ndarray, counts: np.ndarray) -> float:
    """Magnitude of completeness via the maximum-curvature method.

    Mc is defined as the center of the bin holding the mode (highest frequency)
    of the non-cumulative frequency-magnitude distribution.
    """
    mode_idx = int(np.argmax(counts))
    return float(centers[mode_idx])


def aki_b_value(mags_ge_mc: np.ndarray, mc: float, delta_m: float = DELTA_M) -> float:
    """Aki (1965) maximum-likelihood b-value with Utsu binning correction.

        b = 1 / ( ln(10) * ( <M> -(Mc -dM/2) ) )
    """
    mean_mag = float(np.mean(mags_ge_mc))
    denom = LN10 * (mean_mag -(mc -delta_m / 2.0))
    return 1.0 / denom

```

```python

def gr_a_value(n_events: int, b_value: float, mc: float) -> float:
    """Gutenberg-Richter a-value: a = log10(N) + b * Mc."""
    return float(np.log10(n_events) + b_value * mc)


def bootstrap_b_ci(
    mags_ge_mc: np.ndarray,
    mc: float,
    n_boot: int = N_BOOTSTRAP,
    delta_m: float = DELTA_M,
):
    """Bootstrap the b-value and return (mean, std, ci_low, ci_high, samples).

    The 95% CI is taken as the 2.5th and 97.5th percentiles of the bootstrap
    distribution of Aki-MLE b-values.
    """
    n = mags_ge_mc.size
    b_samples = np.empty(n_boot, dtype=float)
    rng = np.random.default_rng(RANDOM_SEED)
    t0 = time.time()
    for i in range(n_boot):
        sample = rng.choice(mags_ge_mc, size=n, replace=True)
        b_samples[i] = aki_b_value(sample, mc, delta_m)
        if (i + 1) % 100 == 0:
            print(
                f" bootstrap {i + 1}/{n_boot} "
                f"(elapsed {time.time() -t0:.1f}s)",
                flush=True,
            )
    ci_low, ci_high = np.percentile(b_samples, [2.5, 97.5])
    return (
        float(np.mean(b_samples)),
        float(np.std(b_samples, ddof=1)),
        float(ci_low),
        float(ci_high),
        b_samples,
    )


def cumulative_counts(centers: np.ndarray, counts: np.ndarray) -> np.ndarray:
    """Cumulative number of events with magnitude >= each bin center."""
    return np.cumsum(counts[::-1])[::-1]


def goodness_of_fit_R(
    centers: np.ndarray,
    counts: np.ndarray,
    mc: float,
    a_value: float,
    b_value: float,
) -> float:
    """Wiemer & Wyss (2000) goodness-of-fit statistic R (percent).

        R = 100 * ( 1 -sum|Nobs -Nsyn| / sum Nobs )

    evaluated over cumulative counts for bins with center >= Mc.
    """
    cum_obs = cumulative_counts(centers, counts)
    mask = centers >= (mc -1e-9)
    m_i = centers[mask]
    n_obs = cum_obs[mask].astype(float)
    n_syn = 10.0 ** (a_value -b_value * m_i)
    R = 100.0 * (1.0 -np.sum(np.abs(n_obs -n_syn)) / np.sum(n_obs))
    return float(R)
```

Listing 1: Core FMD and Gutenberg–Richter routines (excerpt from `fmd_analysis.py`).

References

Aki, K. (1965). Maximum likelihood estimate of b in the formula $\log n = a - bm$ and its confidence limits. *Bulletin of the Earthquake Research Institute, University of Tokyo*, 43:237–239.

Aleksandrova, N. K., Nazaretyan, S. N., and Sergeeva, N. A. (2022). Comparative analysis of the methods for estimating the magnitude of completeness in earthquake catalogues. *Izvestiya, Physics of the Solid Earth*, 58(1):58–72.

Field, E. H., Biasi, G. P., Bird, P., Dawson, T. E., Felzer, K. R., Jackson, D. D., Johnson, K. M., Jordan, T. H., Madden, C., Michael, A. J., Milner, K. R., Page, M. T., Parsons, T., Powers, P. M., Shaw, B. E., Thatcher, W. R., Weldon II, R. J., and Zeng, Y. (2013). Uniform california earthquake rupture forecast, version 3 (ucerf3): The time-independent model. Open-File Report 2013-1165, U.S. Geological Survey.

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

Harris, R. A., Simpson, R. W., and Reasenberg, P. A. (1995). Influence of static stress changes on earthquake locations in southern california. *Nature*, 375(6528):221–224.

Hauksson, E., Yang, W., and Shearer, P. M. (2012). Waveform relocated earthquake catalog for southern california (1981 to june 2011). *Bulletin of the Seismological Society of America*, 102(5):2239–2244.

Hutton, K., Woessner, J., and Hauksson, E. (2010). Earthquake monitoring in southern california for seventy-seven years (1932–2008). *Bulletin of the Seismological Society of America*, 100(2):423–446.

Kagan, Y. Y. (2002). Seismic moment distribution revisited: I. statistical results. *Geophysical Journal International*, 148(3):520–541.

Marzocchi, W. and Sandri, L. (2003). A review and new insights on the estimation of the b-value and its uncertainty. *Annals of Geophysics*, 46(6):1271–1282.

Michael, A. J. and Jones, L. M. (2012). Determining the statistical significance of global earthquake rate changes. *Proceedings of the National Academy of Sciences*, 109(9):2949–2954.

Mignan, A. and Woessner, J. (2012). Estimating the magnitude of completeness for earthquake catalogs. *Community Online Resource for Statistical Seismicity Analysis (CORSSA)*.

Nava, F. A., Bello-Segura, D. I., and Márquez-Ramírez, V. H. (2021). Bayesian estimation of b-value in the gutenberg–richter relationship: A sample size reduction approach. *Natural Hazards*, 108(3):2709–2729.

Nava, F. A., Márquez-Ramírez, V. H., Zúñiga, F. R., Lomnitz, C., and Reyes-Dávila, G. A. (2017). Gutenberg–richter b-value maximum likelihood estimation and sample size. *Journal of Seismology*, 21(1):127–135.

Ross, Z. E., Idini, B., Jia, Z., Stephenson, O. L., Zhong, M., Wang, X., Zhan, Z., Simons, M., Fielding, E. J., Yun, S.-H., Hauksson, E., Moore, A. W., Liu, Z., and Jung, J. (2019). Hierarchical interlocked orthogonal faulting in the 2019 ridgecrest earthquake sequence. *Science*, 366(6463):346–351.

Schorlemmer, D., Wiemer, S., and Wyss, M. (2005). Variations in earthquake-size distribution across different stress regimes. *Nature*, 437(7058):539–542.

Shi, Y. and Bolt, B. A. (1982). The standard error of the magnitude–frequency b value. *Bulletin of the Seismological Society of America*, 72(5):1677–1687.

U.S. Geological Survey and Advanced National Seismic System (ANSS) (2026). ANSS comprehensive earthquake catalog (ComCat), accessed via the FDSN event web service. https://earthquake.usgs.gov/fdsnws/event/1/. Accessed 2026-07-11.

Utsu, T. (1965). A method for determining the value of b in a formula $\log n = a - bm$ showing the magnitude–frequency relation for earthquakes. *Geophysical Bulletin of Hokkaido University*, 13:99–103.

Wei, S., Fielding, E., Leprince, S., Sladen, A., Avouac, J.-P., Helmberger, D., Hauksson, E., Chu, R., Simons, M., Hudnut, K., Herring, T., and Briggs, R. (2011). Superficial simplicity of the 2010 el mayor–cucapah earthquake of baja california in mexico. *Nature Geoscience*, 4(9):615–618.

Wiemer, S. and Wyss, M. (2000). Minimum magnitude of completeness in earthquake catalogs: Examples from alaska, the western united states, and japan. *Bulletin of the Seismological Society of America*, 90(4):859–869.

Woessner, J. and Wiemer, S. (2005). Assessing the quality of earthquake catalogues: Estimating the magnitude of completeness and its uncertainty. *Bulletin of the Seismological Society of America*, 95(2):684–698.