

Magnitude Compression at Two Completeness Thresholds and the Likelihood of Large Earthquakes: Pre-registered Tests of Blue-Dominance and Red-Persistence Rules

RamaKrishna Pasupuleti

workisfun415@gmail.com

Kakatiya University <https://orcid.org/0009-0008-8418-1430>

Research Article

Keywords: magnitude compression, completeness threshold, CSQ indicators, earthquake likelihood, pre-registration, seismicity state

Posted Date: September 30th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11175443/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: The authors declare no competing interests.

Abstract

Catalogue statistics have rarely shown reproducible short-term information about large earthquakes. We examine whether the relation between magnitude compression C computed at two completeness thresholds — $M \geq 1.5$ (blue) and $M \geq 3.0$ (red) — marks periods of different earthquake likelihood. C , together with quiescence Q and a skewness anomaly S , is computed from rolling 20-event windows with catalogue-wide baselines. Three rules, each specified and SHA-256 time-stamped before computation, were tested in the Noto (Japan), Iquique (Chile) and Kahramanmaras (Turkey) catalogues, which contain 15 large earthquakes. The outcome was a large earthquake within the following 90 days. (1) Blue dominance — no red-over-blue crossing and red above blue on at most 10% of the preceding 30 days — was followed by large earthquakes $\times 0.22$ as often as the average day (permutation $p = 0.005$), with lower rates in all three zones; this met the pre-registered criterion, and the association persisted after adjusting for recent activity (odds ratio 0.18, 95% CI 0.14–0.24). (2) Repeated crossings of red above blue (≥ 2 in 30 days) gave $\times 2.79$ ($p = 0.001$) but not in every zone, and the effect was explained by activity (odds ratio 1.12, $p = 0.33$). (3) Red above blue continuously for at least seven days gave ratios of $\times 1.32$ – $\times 1.44$ within zones but $\times 1.00$ pooled ($p = 0.17$), failing its criterion; only 4 of 14 evaluable large earthquakes followed such a phase. Blue dominance may therefore identify low-likelihood periods, while no rule identified the approach of a large earthquake. The findings rest on few earthquakes and are being tested, unchanged, on independent catalogues.

Key points

- Magnitude compression C was computed from events above $M 1.5$ (“blue”) and above $M 3.0$ (“red”) in three well-recorded zones, and three rules relating the two curves were tested with specifications fixed before computation.
- Persistent blue dominance was followed by large earthquakes about five times less often ($\times 0.22$), met its pre-registered criterion, and remained informative after accounting for seismic activity.
- Red rising above blue — repeatedly, or continuously for at least a week — did not meet its criteria; most large earthquakes were not preceded by a sustained red phase.

1. Introduction

Earthquake catalogues record the time, location and magnitude of every detected event, and many statistics derived from them have been proposed as indicators of approaching large earthquakes: changes in the frequency–magnitude distribution (b -value), seismic quiescence, accelerating moment release and others (Wyss & Habermann, 1988; Sobolev & Tyupkin, 1997; Gulia & Wiemer, 2019). Few have survived tests in which the method is fixed in advance and compared with appropriate reference forecasts on independent data (Zechar & Jordan, 2008; Jordan et al., 2011). The dominant, well-established source of short-term information in catalogues is earthquake clustering, described by the epidemic-type aftershock sequence (ETAS) model (Ogata, 1988).

In earlier work we developed three indicators — magnitude compression C , quiescence Q and a skewness anomaly S (the CSQ indicators) — and tested them extensively with pre-registered protocols. No reproducible short-term forecasting skill was found, and the sharp changes of C and S at large earthquakes were shown to arise from the mainshock magnitude entering the rolling window (Pasupuleti, 2026). Those tests used one magnitude threshold at a time.

C depends strongly on the threshold. Computed from all events above M 1.5, it reflects the abundant small-earthquake population and changes within hours to days; computed above M 3.0 it reflects the rarer, more energetic events and changes in steps. Inspection of the two curves suggested that which one dominates might matter: persistent dominance of the low-threshold curve seemed to accompany quiet periods, and sustained dominance of the high-threshold curve seemed to precede large earthquakes. Because such impressions are easily formed from a few striking cases, each version of the idea was written down and cryptographically time-stamped before any test was computed. This paper reports all versions, including those that failed.

2. Data

Table 1

Catalogues. Yearly completeness (maximum-curvature magnitude + 0.2; Wiemer & Wyss, 2000) is between M 1.2 and 1.8 in every year of every zone, so both thresholds are above completeness. A large earthquake is the first event of any 30-day sequence above the zone threshold, which removes aftershocks of earlier large events from the target list.

Zone	Catalogue and zone definition	Events M ≥ 1.5	Events M ≥ 3.0	Large earthquakes
Noto, Japan	Zenodo 10.5281/zenodo.12799164; circle 40 km around 37.53°N, 137.31°E; 2003–2024	10 842	656	7 (M ≥ 4.5)
Iquique, Chile	GFZ Data Services 10.5880/GFZ.4.1.2023.004; circle 100 km around 19.90°S, 70.50°W; 2007–2021	7 892	1 235	5 (M ≥ 6.0)
Kahramanmaras, Turkey	github.com/sadeghkarimpouli/Family_based_features; box 36.82–37.82°N, 36.41–37.61°E; 2017–2023	1 280	59	3 (M ≥ 4.0)

3. Methods

3.1 CSQ indicators at a given threshold

For a catalogue containing all events above a threshold (1.5 or 3.0), sorted by time, the baselines are the population standard deviation of all magnitudes σ_0 , the mean inter-event time $r_0 = (t_N - t_1)/(N - 1)$, and the mean skew₀ and three times the standard deviation s of the skewness of all 20-event windows. After each event i , using that event and the 19 before it:

$$C = \max(0, 1 - \sigma_w / \sigma_0)$$

$$Q = \max(0, 1 - r_0 / r_w), r_w = (t_i - t_{\{i - 19\}}) / 19$$

$$S = \max(0, (\text{skew}_w - \text{skew}_0) / s)$$

C is high when recent magnitudes are unusually similar, Q when events arrive unusually slowly, and S when the recent distribution is unusually skewed toward larger magnitudes. The daily value is the value after the last event before 24:00 UTC. The calculation is provided as csq_calc.py.

3.2 Two-threshold quantities

$$\text{blue}(t) = C \text{ at } M \geq 1.5 \quad \text{red}(t) = C \text{ at } M \geq 3.0 \quad D(t) = \text{red}(t) - \text{blue}(t)$$

A crossing occurs on day t when $D(t - 1) \leq 0$ and $D(t) > 0$. The run length $L(t)$ is the number of consecutive days up to t with $D > 0$; any day with blue above red resets it to zero.

3.3 Rules tested

Table 2
The four pre-registered rules, in the order they were specified.

Rule (specification, SHA-256)	Definition	Claim
Red persistence (RedBlue_Spec, 7c017f63...)	share of days with $D > 0$ in the last 15/20/30 days; area of $D > 0$	top-10% values precede large earthquakes within 21 days
Blue dominance (Crossing_Spec, dfcf1e47...)	no crossing in the last 30 days and $D > 0$ on $\leq 10\%$ of them	fewer large earthquakes within 90 days
Red activity (Crossing_Spec, dfcf1e47...)	≥ 2 crossings in the last 30 days	more large earthquakes within 90 days
Red sustained (RedRun_Spec, 8b88946c...)	$L(t) \geq 7$ days	more large earthquakes within 90 days

3.4 Statistical evaluation

For the 90-day rules, the probability of a large earthquake within 90 days after days in a state was divided by the probability after all valid days, giving a ratio R per zone and pooled over zones. Only days with a fully observed 90-day window were used. Significance was assessed with 1 000 replicates in which each zone’s large-earthquake dates were replaced by the same number of random dates; this preserves the long, autocorrelated runs of each state, which ordinary tests would treat as independent. Pre-registered success required a pooled ratio of at least 1.5 (or at most 0.67 for blue dominance), a permutation p -value below 0.05, and a ratio in the claimed direction in all three zones. The 21-day persistence rule was scored with the Hanssen–Kuipers skill and compared with a “busy zone” baseline (the 7-day event count).

Because crossings and runs can only change when earthquakes occur, every state may partly reflect how active a zone is — and activity itself raises the chance of a large earthquake through clustering. We therefore fitted logistic models of the outcome with zone, the logarithms of the 30-day counts of $M \geq 1.5$ and $M \geq 3.0$ events, and the state. This activity adjustment was added after the first results and is reported as supporting evidence, not as a pre-registered test; its day-level confidence intervals are too narrow because neighbouring days are dependent.

4. Results

4.1 The two curves

Blue varies rapidly because twenty $M \geq 1.5$ events span hours to days during the Noto swarm; red changes in steps as the rarer $M \geq 3.0$ events arrive, and drops abruptly when a large earthquake enters its window (Figure 1). Blue-dominant periods occurred between large earthquakes; red-sustained phases were short and scattered. Figure 2 shows all three indicators at both thresholds through the 2014 Iquique sequence: Q is near zero at both thresholds throughout the active period, and S rises at both thresholds when the largest events enter the window.

4.2 Pre-registered results

Table 3. Probability of a large earthquake within 90 days after days in each state, and ratio to all days.

Zone	All days	Blue-dominant	Red-active	Red ≥ 7 days
Noto, Japan	48.1%	4.1% ($\times 0.09$)	62.1% ($\times 1.29$)	63.6% ($\times 1.32$)
Iquique, Chile	7.3%	1.5% ($\times 0.20$)	16.6% ($\times 2.29$)	9.8% ($\times 1.35$)
Kahramanmaras, Turkey	19.2%	11.2% ($\times 0.58$)	0.0% ($\times 0.00$)	27.8% ($\times 1.44$)
Pooled ratio (p)	—	$\times 0.22$ (0.005)	$\times 2.79$ (0.001)	$\times 1.00$ (0.17)
Pre-registered criterion	—	MET	not met	not met

Blue dominance. After blue-dominant days the probability fell to $\times 0.09$ of the average in Noto, $\times 0.20$ in Iquique and $\times 0.58$ in Kahramanmaras; pooled $\times 0.22$ with permutation $p = 0.005$. All criteria were met.

Red activity and red persistence. Repeated crossings raised the pooled ratio to $\times 2.79$ ($p = 0.001$), but Kahramanmaras had only 25 red-active days and no following earthquake, so the all-zone criterion failed. The 21-day persistence rule reached at best a mean skill of 0.103 (PERSIST_15), below the busy-zone baseline (0.349), and failed.

Red sustained for at least seven days. Within each zone the probability after such phases was modestly higher ($\times 1.32$, $\times 1.35$, $\times 1.44$), but the pooled ratio was $\times 1.00$ ($p = 0.17$) and the criterion failed. The pooled ratio is diluted because Iquique, with a low base rate, contributes most red-sustained days while Noto, with a high base rate, contributes few — an instance of Simpson’s paradox that the pre-registered pooled criterion did not anticipate. The probability did not rise steadily with run length (Figure 4).

4.3 State before each large earthquake

Table 4. Only 4 of 14 evaluable large earthquakes followed a red run of seven days or more; one (Kahramanmaras, 20 October 2022) occurred during a blue-dominant period.

Zone	Large earthquake	State on the day before	Red-over-blue run (days)
Noto	2021-09-16	red-active	0
Noto	2022-03-08	red-active	1
Noto	2022-06-19	red-active	9
Noto	2023-01-06	red-active	0
Noto	2023-05-05	red-active	4
Noto	2024-01-01	red-active	1
Noto	2024-02-11	red-active	5
Iquique	2008-02-04	mixed	0
Iquique	2009-04-17	mixed	74
Iquique	2009-11-13	mixed	83
Iquique	2014-03-16	mixed	1
Iquique	2014-06-19	red-active	4
Kahramanmaras	2017-11-29	not evaluable	—
Kahramanmaras	2022-07-15	mixed	169
Kahramanmaras	2022-10-20	blue-dominant	0

4.4 Is it only activity?

Table 5. Logistic models with zone and the 30-day counts of $M \geq 1.5$ and $M \geq 3.0$ events. Intervals and p -values treat days as independent and are therefore too narrow.

State	Odds ratio (activity-adjusted)	95% interval	p
Blue-dominant	0.18	0.14–0.24	< 0.001
Red-active	1.12	—	0.33
Red sustained ≥ 7 days	3.43	2.84–4.14	< 0.001

Blue dominance remained associated with lower likelihood at every activity level (Figure 5). The red-active effect disappeared once activity was included. The red-sustained state retained an association, but it met neither its pre-registered criterion nor a consistent dose–response, and preceded a minority of large earthquakes.

5. Discussion

5.1 What the results support

One rule — persistent blue dominance — met a pre-registered criterion: the probability of a large earthquake within 90 days was about five times lower than average, in the same direction in an intraplate swarm, a subduction

megathrust and a continental strike-slip fault, and the association persisted when periods of equal activity were compared. It is the first rule derived from the CSQ indicators to meet a pre-registered criterion in our programme, in which earlier single-threshold versions and several published indicators did not (Pasupuleti, 2026). A reliable marker of reduced likelihood could be useful in practice, for example to identify periods in which routine rather than heightened monitoring is appropriate.

5.2 What the results do not support

No rule identified the approach of a large earthquake. Repeated red crossings reflected activity; red dominance lasting a week or more was associated with a modest within-zone increase but did not meet its criterion, showed no dose–response and preceded only 4 of 14 large earthquakes. The idea that sustained high-threshold compression marks energy accumulation is therefore not supported by these data. Interpreting any C state as a measure of stress would require independent physical evidence; C describes the recent magnitude distribution, not stress.

5.3 Possible interpretation of blue dominance

Blue dominance means that small earthquakes are, relative to their long-term behaviour, more uniform in size than larger ones, with no episode in which the larger events became the more uniform set. One hypothesis is that it marks periods of steady background activity without the clustered sequences – foreshock–mainshock–aftershock activity with a broad range of magnitudes – in which most large earthquakes occur. The observation that the association survives adjustment for event counts suggests that the magnitude structure, not only the number of events, carries the information. This interpretation is untested; the state could also be influenced by detection changes near the lower threshold, and it should be examined with ETAS-based simulations and independent catalogues.

5.4 Multiple versions of one idea

Four related rules were tested on the same 15 earthquakes, each refined after the previous result. Even with pre-registration of each version, testing several versions increases the chance that one appears successful by accident. The blue-dominance result is the strongest ($p = 0.005$ and consistent across zones), but it should be regarded as a hypothesis generated and first-tested in these catalogues, to be confirmed on data that played no part in its development.

5.5 Limitations

- Only 15 large earthquakes in three zones; seven are from one swarm (Noto).
- The activity adjustment was added after the results, and its day-level uncertainties are too narrow.
- The pooled ratio criterion is sensitive to differences in base rate between zones (Simpson’s paradox), which affected the red-sustained test.
- Kahramanmaras has only 59 events above M 3.0, so its red curve changes rarely.
- Thresholds, look-back, run length and horizon follow the author’s hypotheses and were not optimised; other choices are untested.
- No comparison with a fitted ETAS forecast has been made.

6. Conclusions

Comparing magnitude compression at $M \geq 1.5$ and $M \geq 3.0$ identified one pattern of practical interest: persistent dominance of the low-threshold curve marked periods in which large earthquakes were about five times less likely, a result that met its pre-registered criterion and persisted after accounting for activity. Rules in which the high-threshold curve rises above the low-threshold curve — repeatedly or for a week or more — did not reliably precede large earthquakes. The blue-dominance rule and all code have been frozen and will be tested unchanged on independent Japanese and Californian catalogues.

Declarations

Data and code availability

All catalogues are public (Table 1). The code package accompanying this paper contains `csq_calc.py` (C, Q and S at both thresholds, crossings, runs and states for any catalogue), `make_figures.py` (all figures), `run_crossing.py` and `run_redrun.py` (the pre-registered tests) and the specifications with their SHA-256 hashes: RedBlue_Spec 7c017f633447e29747518302fddc361b512b0a265802b5b351663a178fb1b37f; Crossing_Spec dfcf1e4758142076a2e8ec1641accb107b4286880128551a0202ec645a4fcc98; RedRun_Spec 8b88946c3d947201d04064b6a67da1ae2b822eea6cea3f108fad63e64df52596. A full reproducibility archive is deposited on DOI 10.5281/zenodo.22973854.

Acknowledgements

The author thanks the data providers of the Noto, Iquique and Kahramanmaras catalogues for making their data openly available.

Competing interests

The author declares no competing interests.

References

1. Gulia, L., & Wiemer, S. (2019). Real-time discrimination of earthquake foreshocks and aftershocks. *Nature*, 574, 193–199.
2. Gutenberg, B., & Richter, C. F. (1944). Frequency of earthquakes in California. *Bulletin of the Seismological Society of America*, 34, 185–188.
3. Jordan, T. H., Chen, Y.-T., Gasparini, P., Madariaga, R., Main, I., Marzocchi, W., et al. (2011). Operational earthquake forecasting: State of knowledge and guidelines for utilization. *Annals of Geophysics*, 54, 315–391.
4. Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, 83, 9–27.
5. Pasupuleti, R. (2026). Catalogue-derived earthquake precursor indicators: development, pre-registered validation, mechanism experiments and corrections. Manuscript and Zenodo archive.
6. Reasenber, P. A., & Jones, L. M. (1989). Earthquake hazard after a mainshock in California. *Science*, 243, 1173–1176.
7. Sobolev, G. A., & Tyupkin, Y. S. (1997). Low-seismicity precursors of large earthquakes in Kamchatka. *Volcanology and Seismology*, 18, 433–446.

8. Wiemer, S., & Wyss, M. (2000). Minimum magnitude of completeness in earthquake catalogs. *Bulletin of the Seismological Society of America*, 90, 859–869.
9. Wyss, M., & Habermann, R. E. (1988). Precursory seismic quiescence. *Pure and Applied Geophysics*, 126, 319–332.
10. Zechar, J. D., & Jordan, T. H. (2008). Testing alarm-based earthquake predictions. *Geophysical Journal International*, 172, 715–724.

Figures

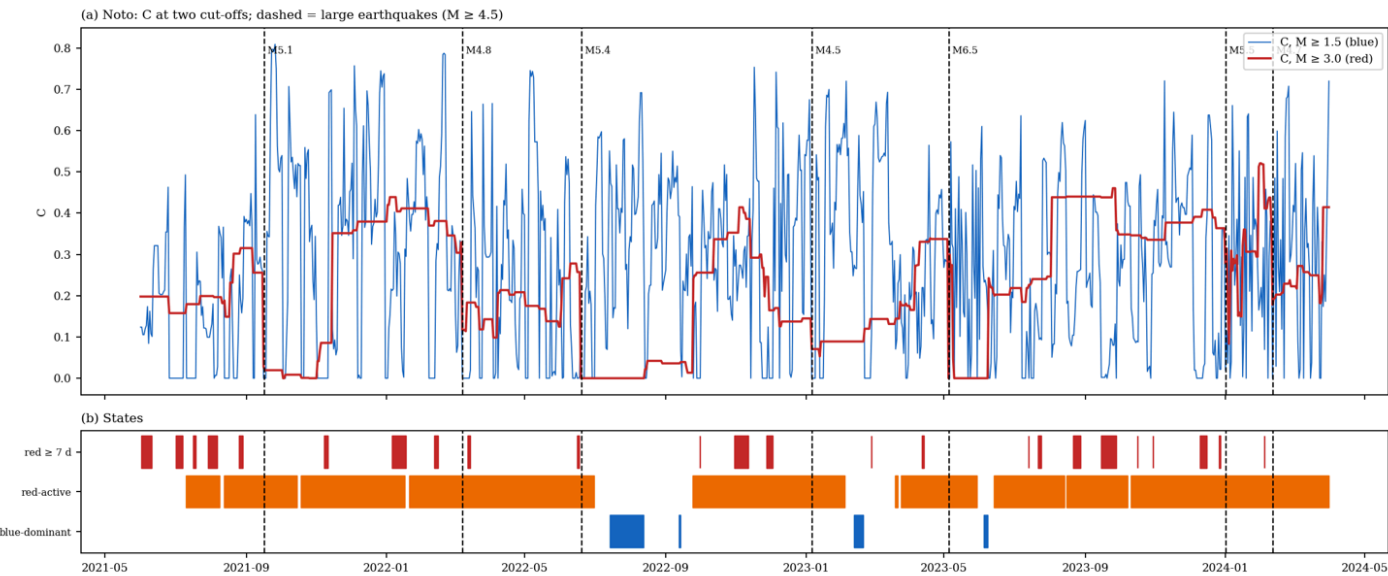


Figure 1

(a) C from $M \geq 1.5$ (blue) and $M \geq 3.0$ (red) in the Noto zone, June 2021 – March 2024; dashed lines mark the seven large earthquakes. (b) Days in each state: blue-dominant, red-active (≥ 2 crossings in 30 days) and red sustained (red above blue for ≥ 7 consecutive days).

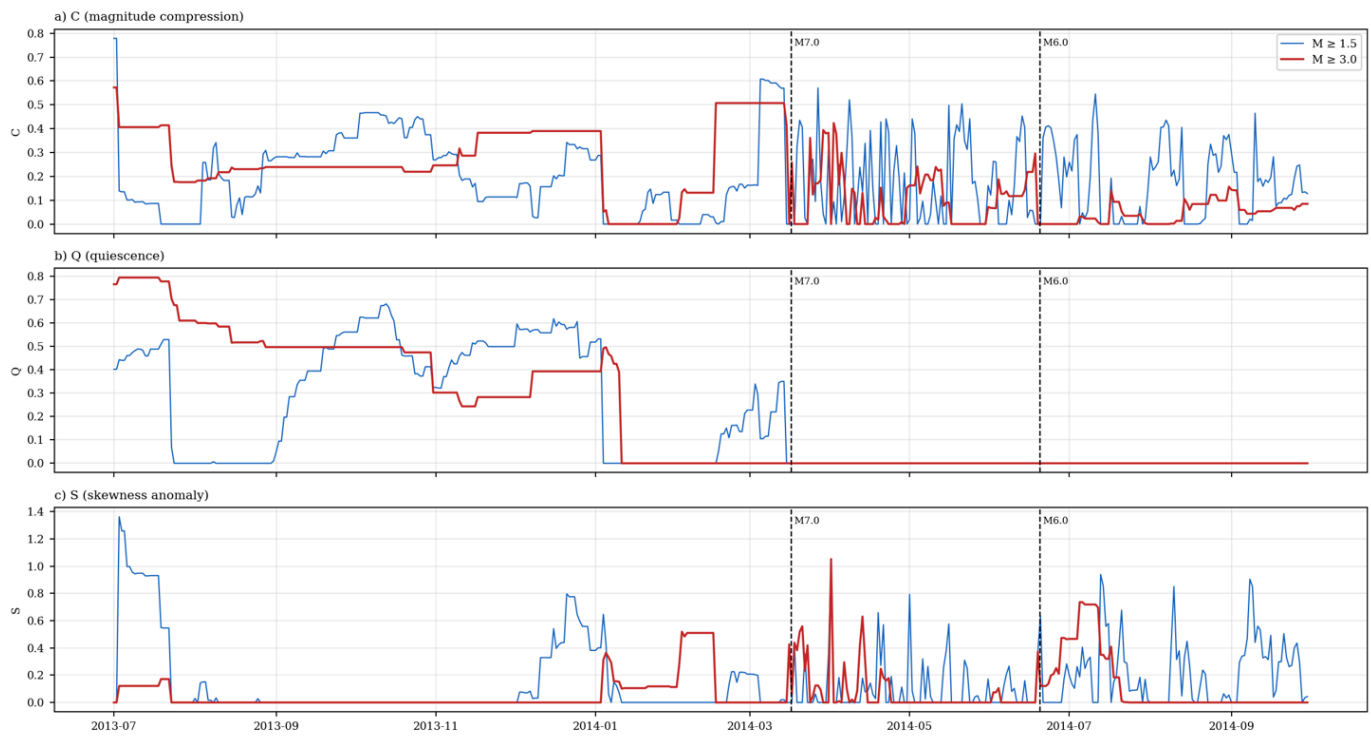


Figure 2

C, Q and S at $M \geq 1.5$ (blue) and $M \geq 3.0$ (red) in the Iquique zone, July 2013 – September 2014, including the 16 March 2014 $M 7.0$ and 19 June 2014 $M 6.0$ earthquakes (dashed).

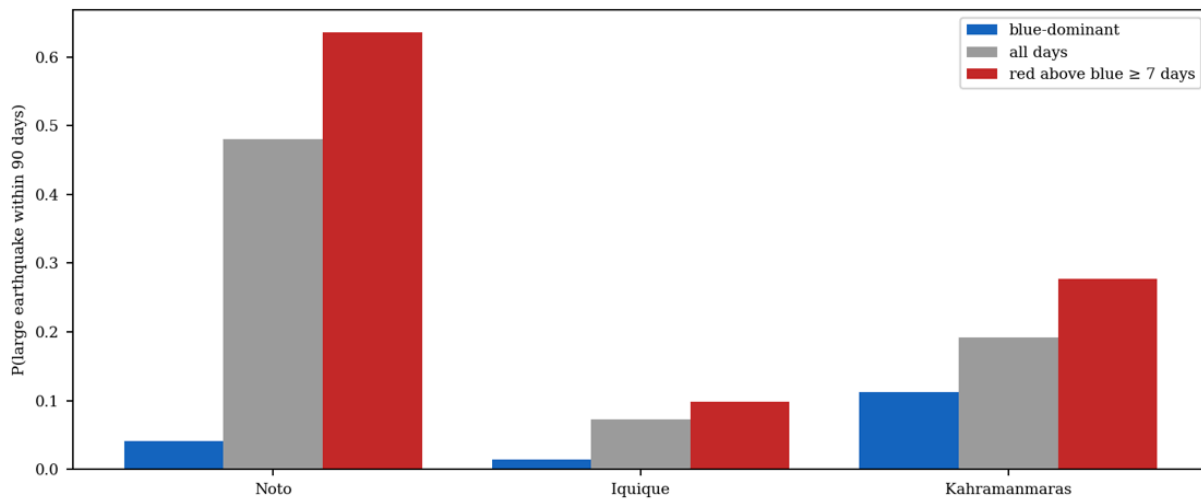


Figure 3

Probability of a large earthquake within 90 days after blue-dominant days, all days and red-sustained days.

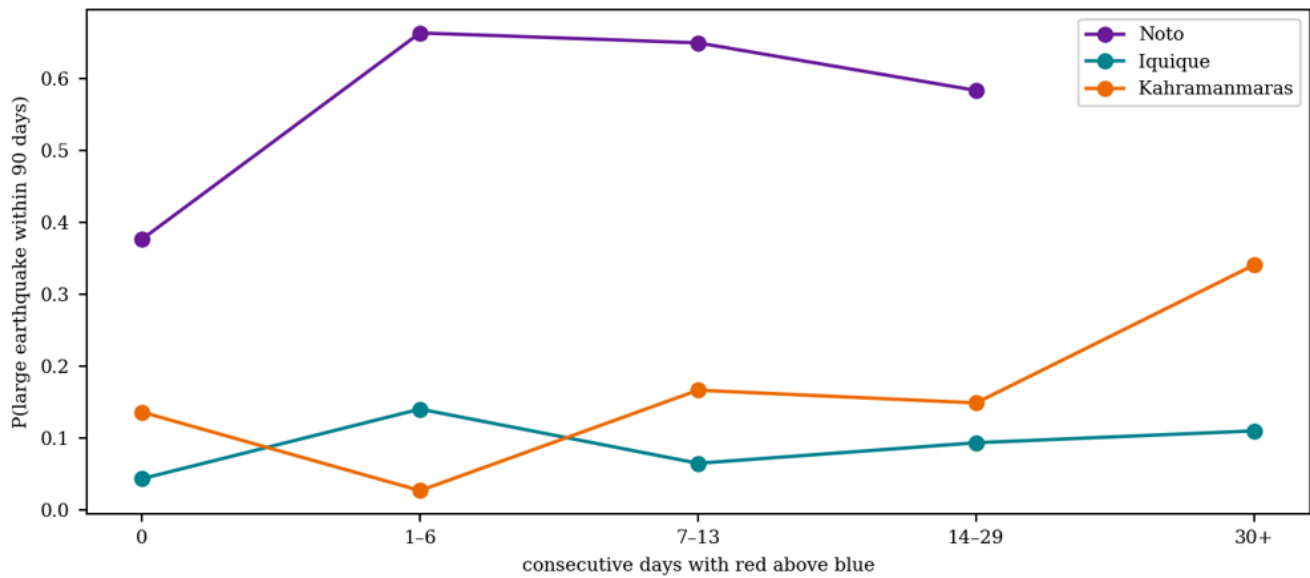


Figure 4

Probability of a large earthquake within 90 days by the number of consecutive days with red above blue.

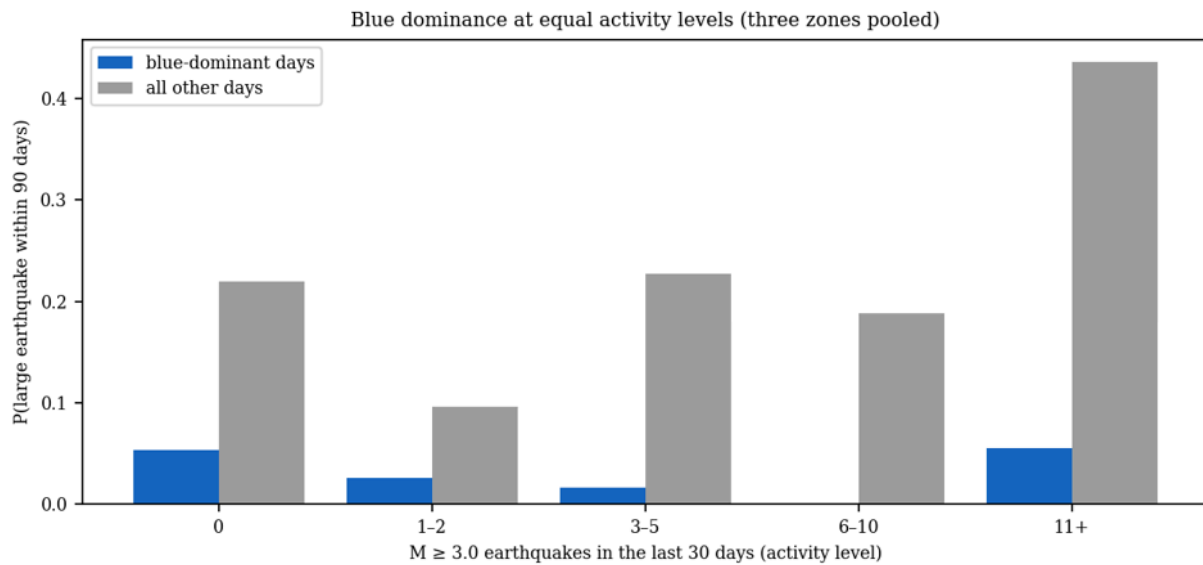


Figure 5

Probability of a large earthquake within 90 days on blue-dominant and other days, at equal levels of recent $M \geq 3.0$ activity (three zones pooled).