

A global classification and characterization of earthquake clusters

Ilya Zaliapin¹ and Yehuda Ben-Zion²

¹Department of Mathematics and Statistics, University of Nevada Reno, 1664 N Virginia St, Reno, NV 89557, USA. E-mail: zal@unr.edu

²Department of Earth Sciences, University of Southern California, Los Angeles, CA, USA

Accepted 2016 August 4. Received 2016 August 3; in original form 2016 March 27

SUMMARY

We document space-dependent clustering properties of earthquakes with $m \geq 4$ in the 1975–2015 worldwide seismic catalogue of the Northern California Earthquake Data Center. Earthquake clusters are identified using a nearest-neighbour distance in time–space–magnitude domain. Multiple cluster characteristics are compared with the heat flow level and type of deformation defined by parameters of the strain rate tensor. The analysis suggests that the dominant type of seismicity clusters in a region depends strongly on the heat flow, while the deformation style and intensity play a secondary role. The results show that there are two dominant types of global clustering: burst-like clusters that represent brittle fracture in relatively cold lithosphere (e.g. shallow events in subduction zones) and swarm-like clusters that represent brittle–ductile deformation in relatively hot lithosphere (e.g. mid-oceanic ridges). The global results are consistent with theoretical expectations and previous analyses of earthquake clustering in southern California based on higher quality catalogues. The observed region-specific deviations from average universal description of seismicity provide important constraints on the physics governing earthquakes and can be used to improve local seismic hazard assessments.

Key words: Earthquake dynamics; Earthquake interaction, forecasting, and prediction; Statistical seismology.

1 INTRODUCTION

Seismicity is often discussed as a prime natural example of universal self-similar behaviour (Bak & Tang 1989; Sornette & Sornette 1989; Keilis-Borok 1990; Rundle *et al.* 2003; Corral 2004; Turcotte & Malamud 2004). The term ‘universality’ implies validity of the same statistical laws in diverse geographic, geological, tectonic and physical settings; while ‘self-similarity’ refers to the abundance of earthquake characteristics described by power laws. A related term to ‘self-similarity’ is ‘scale-invariance’. We recall that the only function that is invariant with respect to changes of measurement units and/or scale of analysis is a power law. Table 2 in Ben-Zion (2008) lists various examples of power-law distributions of earthquake and fault quantities. The most established of those are the power-law distribution of seismic moments (Kagan 1999), which is an alternative form of the exponential distribution of earthquake magnitudes in the Gutenberg–Richter law (Gutenberg & Richter 1944), and the power decay rate of events following a large earthquake referred to as the Omori–Utsu law (Omori 1894; Utsu *et al.* 1995).

These laws were claimed to be universal on a global scale, at least at geologically long time intervals (e.g. Kagan 1999). In this view the documented discrepancies in observed forms and parameters of earthquake statistics are attributed to statistical fluctuations and artefacts of catalogue uncertainties (e.g. Kagan 1999, p. 569). An

alternative approach discussed by Ben-Zion (2008, Sections 2 and 3) considers the universal statistical descriptions to result in part from averaging data of large spatial domains having different event populations. If correct, clarifying the existence of different event populations in relation to physical properties of fault zones and the crust can increase the ability to extract detailed (region-specific) information from observed data and improve the understanding of earthquake physics.

With these goals in mind, Bailey *et al.* (2009, 2010) analysed patterns of earthquake focal mechanisms in southern California and found persisting differences in relation to geometrical properties of the major fault zones. Zaliapin & Ben-Zion (2011) analysed along-strike symmetry properties of aftershocks in catalogues of 25 fault zones in California, and established relations between deviations from generic symmetric distribution and contrasts of seismic velocities across the faults. See also Rubin & Gillard (2000) and Rubin (2002). The results are consistent with theoretical expectations on differences between ruptures on faults that do or do not separate different elastic solids (e.g. Weertman 1980; Ben-Zion 2001; Am-puero & Ben-Zion 2008). Yang & Ben-Zion (2009) and Enescu *et al.* (2009) showed that parameters of the Omori–Utsu aftershock decay law in southern California are correlated with the heat flow. Zaliapin & Ben-Zion (2013a,b) took this further by showing that there are distinctly different types of seismicity clusters in southern

Table 1. Statistics of singles, main shocks, aftershocks and foreshocks in the cluster analysis of events with $m \geq 4$ in the NCEDC catalogue during 1975–2015.

Magnitude range	Singles		Families					
			Main shocks (= no. of families)		Aftershocks		Foreshocks	
	No.	Per cent	No.	Per cent	No.	Per cent	No.	Per cent
All events: $m \geq 4$	116 228	45.2	19 612	7.6	105 790	41.2	15 363	6.0
$4 \leq m < 5$	98 385	47.9	8098	3.9	86 749	42.2	12 197	5.9
$5 \leq m < 6$	17 503	36.8	9141	19.2	17 999	37.8	2968	6.2
$6 \leq m < 7$	340	9.6	2049	57.7	973	27.4	187	5.3
$7 \leq m < 8$	0	0	298	79.5	66	17.6	11	2.9
$m \geq 8$	0	0	26	89.7	3	10.3	0	0

California with preferred locations correlated with the heat flow. These results are consistent with theoretical expectations on properties of earthquake sequences in regions with different effective viscosity (Ben-Zion & Lyakhovsky 2006). In the present paper we generalize the results of Zaliapin & Ben-Zion (2013a,b) to the global scale.

Our analysis focuses on earthquake clustering—partitioning of seismicity into groups closer in space and time than expected in a purely random distribution. Such groups reflect diverse triggering processes and prominently include traditional aftershock series, but also swarms and other types of clustering (Zaliapin & Ben-Zion 2013a; Vidale & Shearer 2006; Vidale *et al.* 2006; Zhang & Shearer 2016). Facilitated by high-quality catalogue and problem-specific statistical techniques, we demonstrated in an earlier study that the cluster style of seismicity in southern California is closely related to physical properties of the crust and is changing at the scale of tens of kilometres (Zaliapin & Ben-Zion 2013b). In particular, it was shown that there are two dominant types of clusters: (i) ‘Burst-like clusters’ with a prominently large main shock, small number of foreshocks and dominance of first-generation offspring. Such clusters reflect highly brittle rapid failure process in areas with cold crystalline rocks, decreased fluid content, and low heat flow production (increased effective viscosity). Burst-like cluster areas in southern California include the San Jacinto fault zone, Mojave, Ventura and San Gabriel regions. (ii) ‘Swarm-like clusters’ that lack a prominent main shock, have increased foreshock activity, and abundance of secondary, tertiary, etc. offspring. Such clusters reflect mixed brittle–ductile failure in areas with increased fluids and heat flow and/or soft sediments (decreased effective viscosity). Swarm-like cluster areas in southern California include the Salton Sea and Coso geothermal regions. The quality of data in southern California allowed us to validate the region-specific character of earthquake clustering by statistical differences in thirteen complementary cluster characteristics, including aftershock/foreshock intensity, magnitude difference between main shock and the largest aftershock/foreshocks, *b*-value, cluster area, duration, etc., all of which related to the effective viscosity of a region and hence to the cluster type (Zaliapin & Ben-Zion 2013b, Table 1 and Appendix C).

The above results from southern California demonstrate the existence of region-specific features that provide important information on earthquake dynamics and can contribute to improving seismic hazard assessments. However, extending the results to the global scale faces the problem of data quality. This is because higher magnitudes of completeness/reporting and earthquake location uncertainties impact cluster identification and lead to multiple artefacts (Zaliapin & Ben-Zion 2015). In particular, low catalogue quality blurs the underlying fine structure of earthquake clusters, artificially

making them more swarm-like, and moves some cluster events to the background mode. Accordingly, working with low-quality catalogues requires developing statistics tools that are robust to the catalogue uncertainties.

In the following sections we develop such a toolbox, and use it to reveal strong spatial dependence of global earthquake clustering that is mainly controlled by the local heat flow production. We confirm the existence of the two primary types of earthquake clusters—burst-like and swarm-like—and show that burst-like clusters are associated with cold regions (mainly shallow seismicity of subduction zones), while swarm-like clustering is typical for hot regions (mainly mid-oceanic ridges). The type of plate-boundary deformation is also examined and shown to play a secondary role in determining the cluster style of seismicity. The global results presented in this study are consistent with our previous regional findings in southern California based on higher-quality data. The analysis of possible sources of artefacts for each examined statistics provide results designed to be robust to the known catalogue uncertainties and deficiencies.

2 DATA AND METHODS

2.1 Earthquakes

We work with the global catalogue produced by the Northern California Earthquake Data Center (NCEDC 2015). The examined catalogue covers the period 1/1/1975 to 6/9/2015 and contains 256 993 events. The minimal magnitude used in the analysis is $m_{\min} = 4$. This magnitude is higher than the completeness magnitude in many examined regions, in particular during the earlier times. We demonstrated (Zaliapin & Ben-Zion 2013a, Appendices D and E) that the cluster structure estimated by our technique (Section 3) is insensitive to the catalogue incompleteness as well as to the minimal reported magnitude. Accordingly, some cluster statistics, like the total number of clusters and partition of events into main shocks, foreshocks, and aftershocks (see Section 3.6 for definitions) are fairly robust with respect to the magnitude incompleteness. The incompleteness however does affect the cluster size distribution, as discussed in Section 4.2.

We only consider events with depth less than $z_c = 70$ km. The depth reporting in NCEDC catalogue is highly irresolute: 69 932 events (27.2 per cent) are assigned a depth of 33 km and 61 939 events (24.1 per cent) are assigned a depth of 10 km. Other popular (default) depth values are 35 km (14 421 events, 5.6 per cent), 30 km (4899 events, 1.9 per cent), and 5 km (2651 events, 1.0 per cent). In addition, there is a tendency, especially during earlier times, to assign depths divisible by 5 km (5, 10, 15, etc.). Our analysis is

based on earthquake epicentres and is not affected by the depth uncertainties.

Fig. 1(a) shows the spatial intensity $\Lambda(\mathbf{x})$ of events in the NCECD catalogue, in events with magnitude $m \geq 4$ per year per 10 000 km². Appendix B describes the process of producing smooth spatial maps of different seismic and physical characteristics used in this study. The intensity varies over several orders of magnitude, from 0.02 to 12.5, with the highest values associated with contracting subduction zones and lowest values associated with mid-oceanic spreading ridges. The global spatial distribution of the maximal observed earthquake magnitude m_{\max} of the examined seismicity is illustrated in Fig. 1(b). Naturally, the fluctuations of the maximal observed magnitude are closely related to the seismic intensity fluctuations. The distribution of hypocentral depth of the examined earthquakes is shown in Fig. A3(a). Its spatial variations resemble those for the earthquake intensity and maximal magnitude.

Fig. 1(c) presents the spatial variability of the magnitude completeness of the examined catalogue. It shows the proportion p_5 of earthquakes with magnitude $m \leq 5$, which serves as a proxy to completeness quality. Specifically, if the number $N(m)$ of events with magnitude above m is given by the Gutenberg–Richter law

$$\log_{10} N(m) = a - bm, \quad m \geq m_c \quad (1)$$

with $b = 1$ and $m_c = 4$, then $p_5 = 0.9$. Deficiency of events in small magnitude range results in lower values of p_5 . Mild fluctuations of the b -value also affect this completeness proxy albeit to a lesser degree. For instance, if we assume validity of eq. (1) with $m_c = 4$, then variation of b -values in the range [0.8, 1.2] results in p_5 in the range [0.84, 0.94], with lower b -values corresponding to lower p_5 values. Hence, the main fluctuations of p_5 (that goes as low as 0.5) observed in Fig. 1(c) are primarily due to incompleteness. The analysis suggests that catalogue quality is deteriorated for oceanic seismicity, relatively far from seismic networks, which mainly occur in the southern hemisphere.

Fig. 1 illustrates and summarizes the diversity of seismic regimes and parameters as well as variations in catalogue quality involved in a global study. The statistics used in our analysis are designed to be robust with respect to these obstacles and yet still reflect the essential characteristics of the regional cluster style.

2.2 Heat flow

The employed surface heat flow data is taken from Bird *et al.* (2008). The heat flow within the seismically active areas is mapped in Fig. 2. The distribution of the heat flow over the entire Earth surface is shown in Fig. A1. The heat flow production is prominently high along the oceanic spreading ridges, reaching 0.3 W m⁻².

2.3 Strain rate tensor

We use the global strain rate field modelled by Kreemer *et al.* (2014). Specifically, we consider the second invariant I_2 of the strain rate tensor $\dot{\epsilon}$:

$$I_2 = \sqrt{(\dot{\epsilon}_{\varphi\varphi})^2 + (\dot{\epsilon}_{\theta\theta})^2 + 2(\dot{\epsilon}_{\varphi\theta})^2} \quad (2)$$

and the tensor style S defined by Kreemer *et al.* (2014) as:

$$S = \frac{e_1 + e_2}{\max(|e_1|, |e_2|)}. \quad (3)$$

Here e_i are the eigenvalues of the strain rate tensor. The strain rate tensor style S can be used to roughly quantify the type of displacement into contraction ($S < -0.5$), strike-slip ($0.5 < S < -0.5$), and extension ($S > 0.5$). The maps of strain rate tensor second invariant and style are shown in Fig. A2.

2.4 Δ -Analysis

Any cluster analysis of earthquakes is affected by the existence of the catalogue lower cut-off magnitude m_{\min} (which may be smaller than the completeness magnitude m_c). For instance, if we analyse earthquakes with $m \geq m_{\min} = 4$, then an earthquake of $m = 4$ cannot have recorded aftershocks of a smaller magnitude, while an $m = 6$ event may have aftershocks with magnitudes $4 \leq m \leq 6$. To equalize the magnitude ranges for potential fore/aftershocks of main shocks with different magnitudes, we sometimes perform Δ -analysis that (i) only considers main shocks with magnitude $m \geq m_{\min} + \Delta$ and (ii) only considers fore/aftershocks with magnitude within Δ units below that of a main shock. The fore/aftershocks detected by this analysis are called Δ -fore/aftershocks. The conventional analysis that considers all events is referred to as *regular analysis*.

3 EARTHQUAKE CLUSTERS

3.1 Generalized earthquake distance

Consider a catalogue where each event i is characterized by its occurrence time t_i , hypocentre (ϕ_i, λ_i, d_i) , and magnitude m_i . We define the proximity η_{ij} of earthquake j to earthquake i following Baiesi & Paczuski (2004) as:

$$\eta_{ij} = \begin{cases} t_{ij}(r_{ij})^d 10^{-bm_i}, & t_{ij} > 0; \\ \infty, & t_{ij} \leq 0. \end{cases} \quad (4)$$

Here, $t_{ij} = t_j - t_i$ is the event intercurrence time, which is positive if event j occurred after event i ; $r_{ij} \geq 0$ is the spatial distance between the earthquake hypocentres; d is the (possibly fractal) dimension of the hypocentres or epicentres, and b is the parameter of the Gutenberg–Richter law (1). The motivation for and properties of this proximity measure are discussed in Zaliapin & Ben-Zion (2013a, 2015, 2016). In particular, the proximity to the nearest neighbour is inversely related to the seismic intensity. It is intuitive that the distance between events is smaller in a high-intensity process where a larger number of events occupy the same space–time volume; see Zaliapin *et al.* (2008) for formal derivations and Zaliapin & Ben-Zion (2013a) for simulation results.

3.2 Parent-offspring identification

For each event i we identify its unique nearest neighbour (*parent*) j with respect to the distance given by eq. (4), and denote for simplicity the nearest-neighbour distance by the same symbol η_{ij} . The event i is called an *offspring* of j . According to this definition, each event (except the first one in the catalogue) has a unique parent, and also might have multiple offspring.

3.3 Bimodal distribution of the nearest-neighbour distance

Consider the space and time distances between event i and its parent j normalized by the magnitude of the parent event (Zaliapin

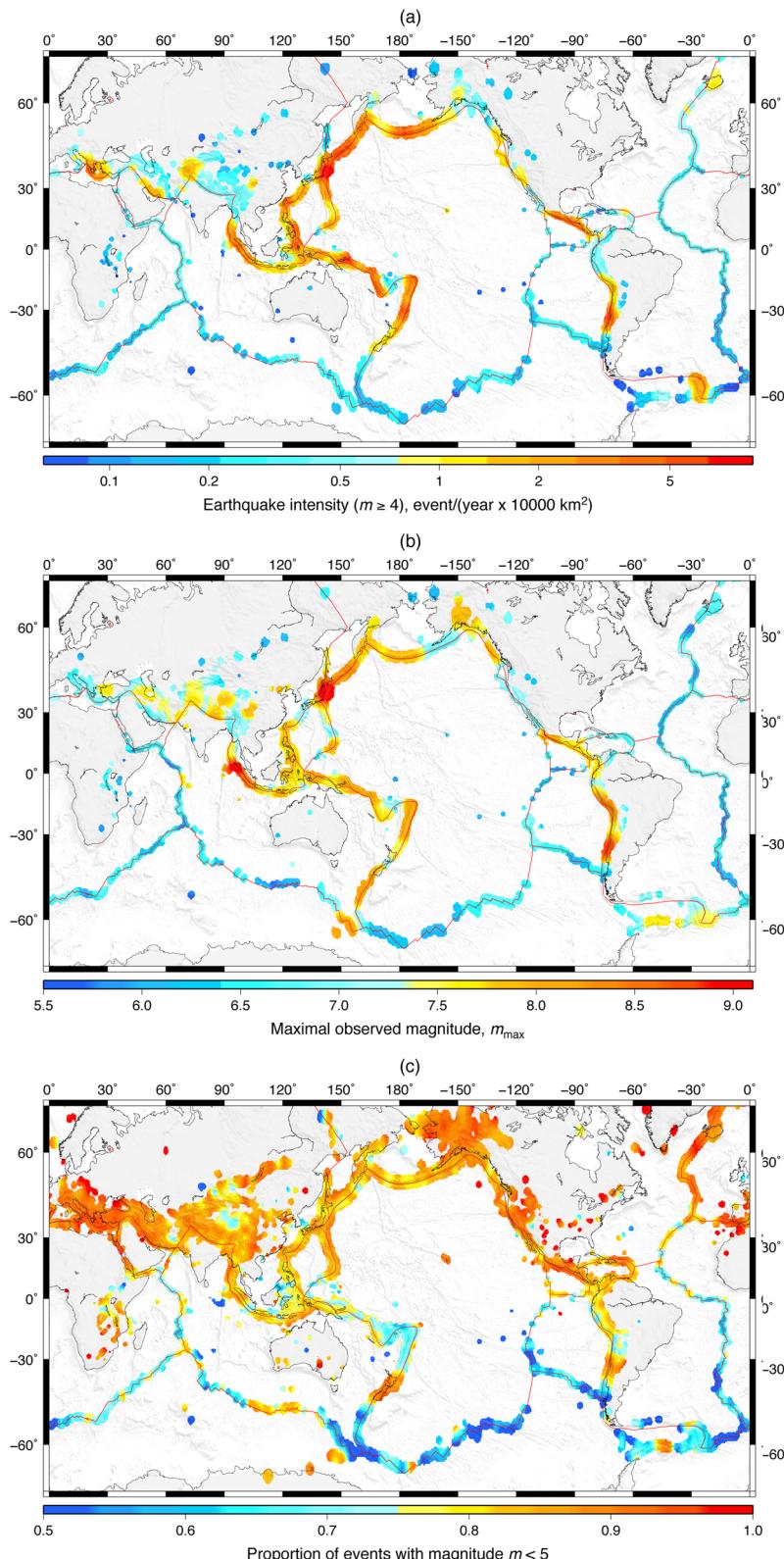


Figure 1. Spatial distribution of selected statistical characteristics of earthquakes with magnitude $m \geq 4$ according to the global NCEDC catalogue during 1975–2015. A point is included in this graph if the circle of radius 100 km centred at the point contains 5 or more earthquakes of magnitude $m \geq 5$. Red lines depict major tectonic faults. Shadings correspond to bathymetry and topography. Continents are depicted by grey colour. (a) Earthquake intensity Λ in events $\times \text{yr}^{-1} \times 10000 \text{ km}^{-2}$ (b) Maximal observed magnitude m_{\max} . (c) Proportion of events with $m \leq 5$.

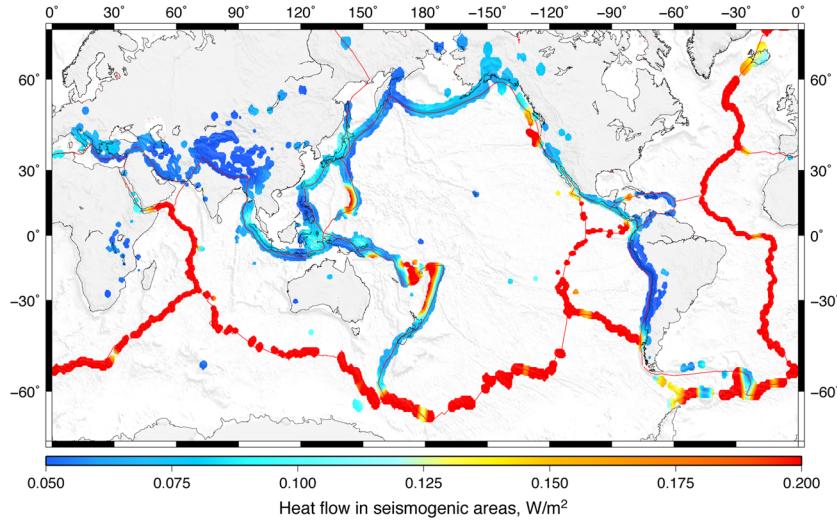


Figure 2. Heat flow in the seismically active regions. A point is included in this graph if the circle of radius 100 km centred at the point contains 5 or more earthquakes of magnitude $m \geq 5$. The heat flow values are clipped at $H = 0.2$ (the maximal reported value is 0.3).

et al. 2008):

$$T_{ij} = t_{ij} 10^{-q b m_i}; R_{ij} = (r_{ij})^d 10^{-p b m_i}; q + p = 1. \quad (5)$$

This is convenient because now $\log T_{ij} = \log T_{ij} + \log R_{ij}$. Zaliapin et al. (2008) and Zaliapin & Ben-Zion (2013a) demonstrated that a time-stationary space-inhomogeneous Poisson flow of events with Gutenberg–Richter magnitudes corresponds to a unimodal distribution of $(\log T, \log R)$ that is concentrated along a line $\log_{10} T + \log_{10} R = \text{const}$. Observed seismicity, however, shows a bimodal joint distribution of $(\log_{10} T, \log_{10} R)$, as has been documented in multiple studies of various regions (e.g. Zaliapin et al. 2008; Zaliapin & Ben-Zion 2011, 2013a,b, 2015, 2016; Gu et al. 2013; Davidsen et al. 2015; Reverso et al. 2015; Schoenball et al. 2015). One of the modes is similar to that observed in a Poisson process and corresponds to *background* events, while the other consists of *clustered* events located considerably closer in time and space to their parents than expected in a Poisson process (see Fig. 4).

3.4 Separating the background and cluster modes: a Gaussian mixture model approach

Analysis of statistical properties of the background and cluster modes is one of the main tools of this study (see Section 4.2). The bimodality of the earthquake distance distribution (Figs 4b and d) allows one to use a suitably chosen nearest-neighbour threshold η_0 to formally attribute each event to either the background (if $\eta_{ij} > \eta_0$) or cluster (if $\eta_{ij} < \eta_0$) mode. The threshold selection is done here according to a Gaussian mixture model with two modes.

A two-mode Gaussian mixture model assumes that sample $x_i \in R^m$, $i = 1, \dots, n$ comes from the distribution

$$F(x) = wN(x; \mu_1, \Sigma_1) + (1 - w)N(x; \mu_2, \Sigma_2), \quad (6)$$

where w is the mixture weight of the first mode and $N(x; \mu, \Sigma)$ denotes the Gaussian (Normal) distribution, with mean μ that is a vector with m components and variance Σ that is a positive-definite $m \times m$ matrix. The estimation of such model can be done using the Expectation-Maximization algorithm (Dempster et al. 1977).

In our setting, we can either apply a 1-D Gaussian mixture model to the log-distance $\log_{10} \eta$ or a 2-D Gaussian mixture model to the logarithmic components $(\log_{10} T, \log_{10} R)$. Both approaches

give very close results. The details of numerical implementation are discussed in Hicks (2011). A model assigns to each event the probabilities w and $(1-w)$ of being attributed to one or the other mode. We make the final mode assignment according to the maximal probability $v = \max(w, 1-w)$. This corresponds to choosing the mode separation threshold η_0 that equalizes the densities of the two estimated Gaussian modes:

$$N(\eta_0; \mu_1, \Sigma_1) = N(\eta_0; \mu_2, \Sigma_2). \quad (7)$$

The background (cluster) events now can be equivalently defined by the condition $\eta_{ij} > \eta_0$ ($\eta_{ij} \leq \eta_0$). In 1-D case, the position η_{bg} of the background is defined as the mean value of the estimated rightmost Gaussian mode: $\eta_{bg} = \max(\mu_1, \mu_2)$. In 2-D case we define $\eta_{bg} = \max(\mu_1[1] + \mu_1[2], \mu_2[1] + \mu_2[2])$, where the notation $[i]$ refers to the i th component of a vector. Alternatively, one can define η_{bg} as (a) the mean generalized earthquake distance η_{ij} of the background events, or (b) the mean generalized distance of events that happened at large spatial distance from their parent (say, $R > 5, 50$). The last approach is motivated by the observation that the majority of events at large spatial distances from their parent belong to the background mode (see Fig. 4). These alternate approaches give results (not shown) that are very close to those obtained with our main method.

The regional mode separation quality Q is defined as the average value of the mode assignment probability $v = \max(w, 1-w)$ over all events in a region. According to this definition, the quality is constrained by $0.5 \leq Q \leq 1$, where $Q = 1$ corresponds to a perfect separation (each event is attributed to one of the modes with probability 1) and $Q = 0.5$ corresponds to an indeterminate separation (each event is attributed to either mode with the same probability of 0.5).

3.5 Cluster identification

Connecting each earthquake in the catalogue to its nearest neighbour (parent) according to the nearest-neighbour distance η of eq. (4) produces a single cluster (spanning network) that contains all examined events. From a graph-theoretical perspective, this cluster is a *tree graph*, which means that it does not have loops (Zaliapin & Ben-Zion 2013a; Baiesi & Paczuski 2004). Removing all links

that correspond to large parent-offspring distances, defined by the condition $\eta \geq \eta_0$, creates a spanning forest—a collection of trees each representing a separate earthquake cluster. The forest contains many single-event trees, which we call *singles*. The other clusters contain multiple events and are called *families*.

3.6 Event classification

In each family, the earthquake with the largest magnitude is called *main shock*. If there are several earthquakes with same largest magnitude within a family, the first one is considered to be the main shock, so each family has a single main shock. All events in a family that occurred after the main shock are called *aftershocks*. All events that occurred prior to the main shock are called *foreshocks*. Each single is also considered to be a main shock (that has no foreshocks and aftershocks).

3.7 Parameters

In this study we use event epicentres rather than hypocentres because the depth coordinates are often less accurate than those of the epicentres, and location errors lead to various analysis artefacts as discussed in detail by Zaliapin & Ben-Zion (2015). In addition, we fix $b = 1$, $d = 1.3$ and $p = 0.5$ in eqs (4) and (5). Zaliapin & Ben-Zion (2013a) showed that the estimated cluster structure is fairly robust with respect to the values of these parameters. Accordingly, the main conclusions of this study are not sensitive to the precise parameter values.

We refer to Zaliapin & Ben-Zion (2013a) for further details on and examples of performance of our cluster technique, as well as detailed analysis on its stability. The statistical artefacts of catalogue uncertainties that affect cluster analysis based on parent-offspring identification are examined in Zaliapin & Ben-Zion (2015).

4 RESULTS

4.1 Basic characteristics of seismicity

For the purpose of this study, a point on the Earth surface is denoted seismically active if there are 5 or more events with magnitude $m \geq 5$ within 100 km of this point according to the NCEDEC catalogue during 1975–2015. The total Earth surface area is $510.1 \times 10^6 \text{ km}^2$; about 9 per cent of this area, or $44.9 \times 10^6 \text{ km}^2$, is seismically active. Fig. 1 shows basic characteristics of seismicity in the active areas: (panel a) Intensity, in event/(year $\times 10^3 \text{ km}^2$), of earthquakes with magnitude $m \geq 4$, (panel b) maximal observed magnitude and (panel c) proportion of events with magnitude $m \leq 5$ which serves as a proxy for catalogue completeness (see discussion in Section 2.1).

Next we relate the three characteristics illustrated in Fig. 1 to the type of lithospheric deformation. Fig. 3(a) displays the partitioning of seismically active area with respect to values of the strain rate tensor style S and second invariant I_2 . The analysis is done within $0.5^\circ \times 0.5^\circ$ cells that tile the Earth surface. About half of the active area corresponds to lowest values of the second invariant ($I_2 < 100$) attributed to subduction zones and other contraction environments. The largest 10 per cent of the values of the second invariant ($I_2 > 1000$) are associated with mid-ocean ridges and other extension environments. The partition of seismic area into different tensor styles is fairly independent of the values of I_2 —approximately 60 per cent in contraction, 20 per cent in strike-slip and 20 per cent in extension.

Fig. 3(b) shows the average value of seismic intensity for events with $m \geq 4$ in the same coordinates (S, I_2). The highest average intensity of $\Lambda \approx 3$ events per year per 10^3 km^2 is exclusively observed within contracting environments: $S < -0.75$, $I_2 > 100$. The lowest average intensity, $0.2 < \Lambda < 1$, is observed within extending environments. The intensity in strike-slip zones is intermediate around and slightly below $\Lambda = 1$. The highest seismic activity typically occurred in subduction zones, which explain a close resemblance in the patterns of seismic intensity (Fig. 3b) and that of the average hypocentral depth (Fig. A4a). The average heat flow has the highest values ($H > 0.25$) exclusively within extension environments—along the mid-oceanic spreading ridges (Fig. 3c). The results in Figs 3(b) and (c) emphasize that the spatial distribution of seismic intensity is inversely related to that of the heat flow. We also perform Spearman's rank correlation analysis (see Appendix C for definition and discussion) and generalized linear model (GLM) analysis (see Appendix D) for earthquake intensity versus heat flow H and strain rate tensor parameters S and I_2 . The results, summarized in Tables 5, 6 and Figs D1(a), (b) and (c), corroborate the observations of Fig. 3. We show below that the heat flow production also governs the space-dependent style of earthquake clustering.

Fig. 4 presents the distribution of the generalized earthquake distance η of eq. (4) and the joint distribution of its normalized space and time components (R, T) of eq. (5) for the earthquakes from areas with low ($H < 0.1$) and high ($H > 0.25$) heat flow value. Panels (a, c) show the joint two-dimensional density of (R, T) estimated for all earthquakes in areas with the indicated heat flow level. Colours represent the relative number of points with given values of (R, T) , as indicated in the colourbar on the right. The integral over the entire distribution is 1. Panels (b, d) show the estimated density for the nearest-neighbour distance η on a logarithmic scale—the histogram values for $\log_{10}\eta$ divided by the total number of examined events—so that the integral over the distribution is 1. The background and cluster modes are seen clearly in both cases. However, there exist several notable differences in the cluster style of earthquakes in high versus low heat flow regions: (i) The typical position η_{bg} of the background mode in low heat flow regions ($\log_{10}\eta_{bg} \approx -4.7$) is lower than that in high heat flow regions ($\log_{10}\eta_{bg} \approx -4.0$). This reflects higher earthquake intensity in low heat flow regions that have predominantly contracting and transform deformation style, and corroborates our earlier observations in Figs 3(b) and (c). (ii) The proportion of background events in low heat flow regions is lower than that in high heat flow regions. Accordingly, the proportion of clustered events is higher in low heat flow regions. (iii) The time decay of cluster events is faster in high heat flow regions, leading to stronger time separation between the background and cluster modes. This can be seen by comparing how the cluster mode is blending with the background mode in panels (a) and (c) of Fig. 4. The offspring duration is longer in low heat flow regions, which is reflected by a horizontally elongated shape of the cluster mode in Fig. 4(a), as opposed to a more confined location of the cluster mode in high heat flow regions in Fig. 4(c). (iv) Proportion of repeaters—events that happen at short spatial and large temporal distances from the parent and hence occupy the lower right corner of the (T, R) plots in Figs 4(a) and (c)—is larger in high heat flow regions. This observation is further illustrated in Fig. 5 that shows the distribution of rescaled time to parent T for offspring within two parent rupture lengths to the parent. In cold regions (panel a) the cluster and background modes are largely overlapping at these short distances to parent. The background mode is centred at about $\log_{10}T = -3$ and has lower intensity and spread

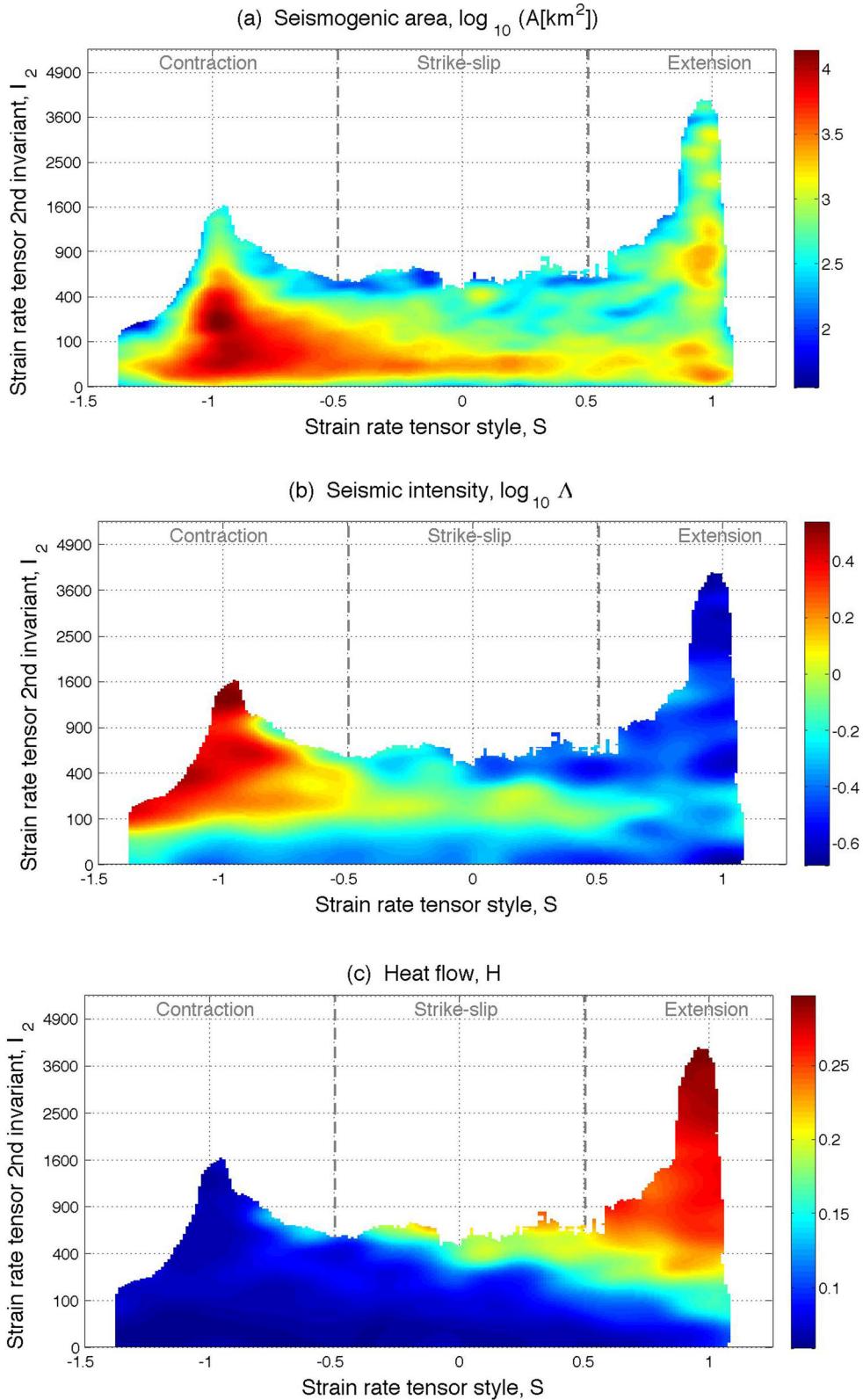


Figure 3. Average values of selected statistics as a function of strain rate tensor style S and second invariant I_2 . (a) Seismogenic area A [km^2] (the values are reported on a logarithmic scale). (b) Seismic intensity Λ [$\text{event yr}^{-1} 10 000 \text{ km}^{-2}$] (the values are reported on a logarithmic scale). (c) Heat flow H [W m^{-2}].

compared to that of the clustered mode. The latter is centred at $\log_{10} T = -6$ and has much larger spread, interpreted as slow decay of intensity of offspring earthquakes. In hot regions (panel b), on the contrary, the two modes are well separated. The background mode

is centred at about $\log_{10} T = -2$ (lower intensity of background events compared to those in cold regions), and has notably higher intensity than the cluster mode. The cluster mode is centred at about $\log_{10} T = -6.5$ and has smaller spread than that of the cluster mode

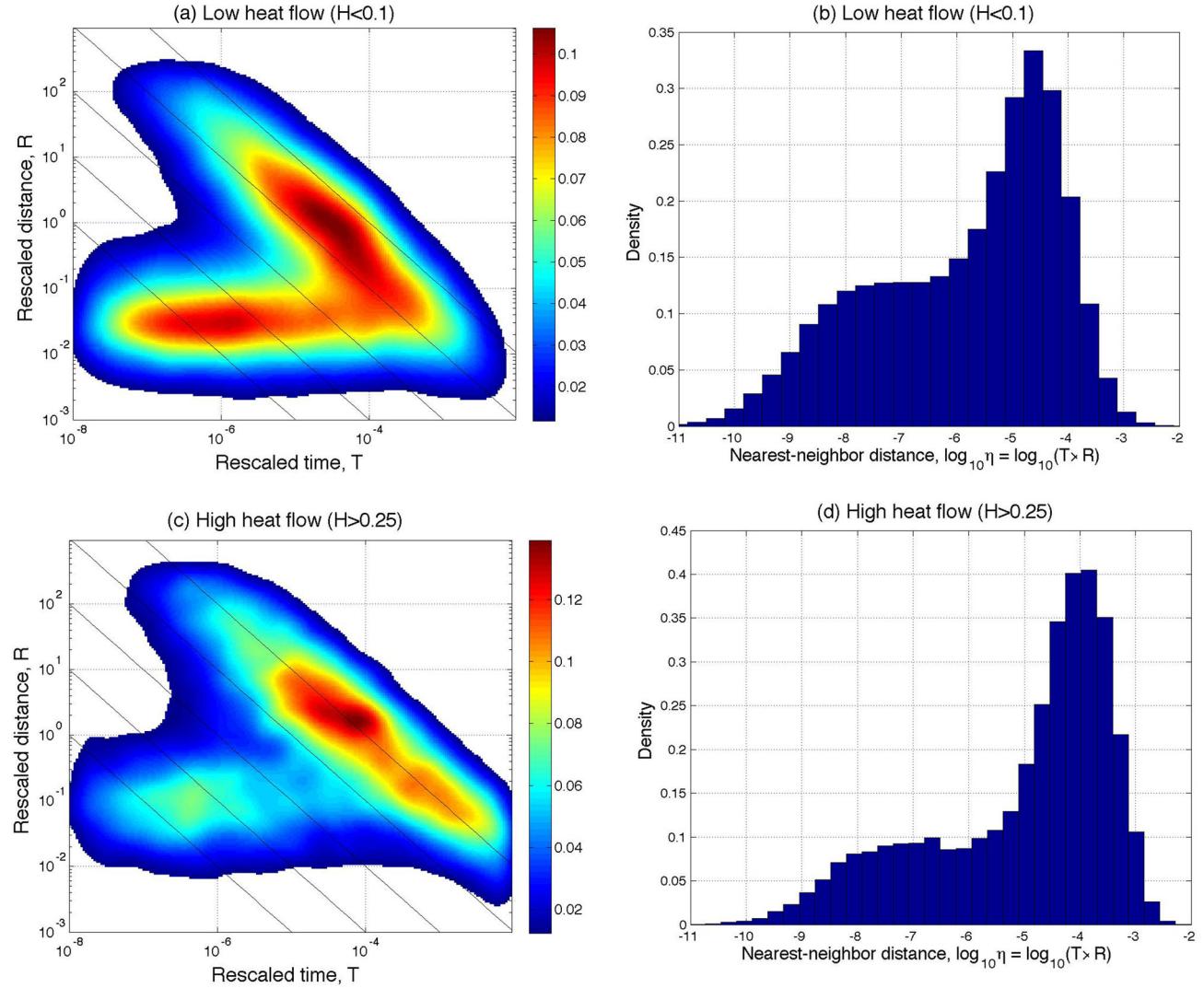


Figure 4. Generalized earthquake distance η of eq. (4) and its normalized space and time components (T, R) of eq. (5) in regions with different level of heat flow H . (a, b) Earthquakes in regions with low heat flow, $H < 0.1$. (c, d) Earthquake in regions with high heat flow, $H > 0.25$. (a, c) Joint distribution of the rescaled components (T, R) of the earthquake nearest-neighbour distance. (b, d) Distribution of the values of the nearest-neighbour distance η . Black diagonal lines in panels (a, c) depict levels of constant distance η (from top to bottom): $-\log_{10}\eta = 4, 5, 6, 7, 8$.

in cold regions, suggesting faster decay of intensity of offspring earthquakes.

4.2 Cluster and background modes

The results of Figs 3–5 demonstrate that earthquake clustering style is space-dependent and related to the heat flow production. We now complement these analyses by additional statistics involving the earthquake nearest-neighbour distances. Specifically, we apply a 1-D Gaussian mixture model (Section 3.3) to the nearest-neighbour distances $\log_{10}\eta$ of events within circles of radius $r = 200$ km centred at the epicentres of all examined earthquakes. The model is used to estimate the space-dependent threshold η_0 that separates the cluster and background modes, partition the events into cluster and background populations, estimate the characteristic position η_{bg} of the background mode, and assess the quality Q of mode separation. Fig. 6 shows the spatial maps of the position η_{bg} of the background mode (panel a) and the proportion of events in the background mode

(panel b). The map of the quality Q of the mode separation is shown in Fig. A3(c).

The location of the background mode is primarily controlled by the absolute intensity of the background events (Zaliapin *et al.* 2008; Zaliapin & Ben-Zion 2013a). This explains the inverse relation between the background location (Fig. 6a) and earthquake intensity (Fig. 1a). Furthermore, the values of η_{bg} follow a three-modal distribution, clearly outlining the major tectonic environments in agreement with Fig. 3(b). The highest earthquake intensity and lowest values of $\eta_{\text{bg}} < -4.5$ are observed within convergent environments. The lowest earthquake intensity and largest values of $\eta_{\text{bg}} > -3.75$ are observed along divergent boundaries. Intermediate values of earthquake intensity and background position $-4.5 < \eta_{\text{bg}} < -3.55$ are observed along transform boundaries. Fig. A3(b) shows the worldwide spatial distribution of a related feature—the threshold η_0 that separates the background and cluster mode, according to a Gaussian mixture model.

The other examined cluster characteristics exhibit similar spatial variations. In particular, divergent environments have uniformly increased background proportions $p_{\text{bg}} > 0.7$ (Fig. 6b) and high

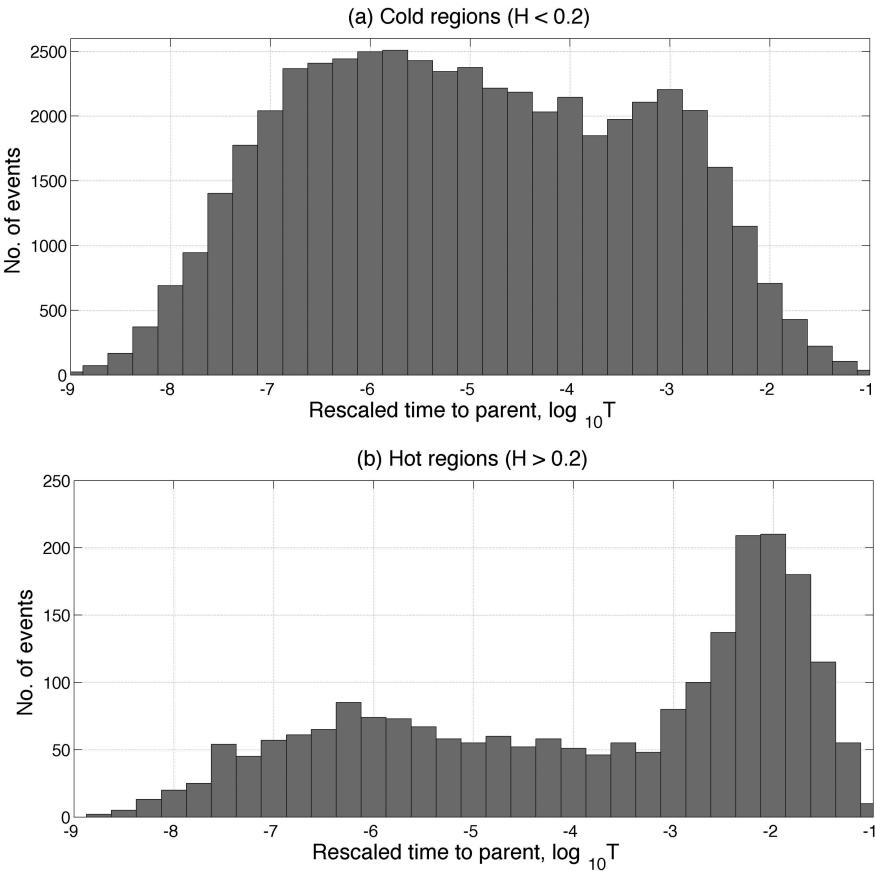


Figure 5. Distribution of the rescaled time to parent T of eq. (5) for offspring that occurred within two parent rupture distances from the parent. (a) Cold regions, $H < 0.2$. (b) Hot regions, $H > 0.2$.

mode separation quality $Q \approx 0.95$ (Fig. A3c). Convergent environments exhibit much larger spatial variability and intermittence in the values of background proportions p_{bg} and separation quality Q . For instance, Fig. 6(b) shows that in the Northwestern part of the Pacific plate, along the Kuril-Kamchatka and Japan trenches, the background proportion varies widely in the range 0.2–0.7 over a scale of hundreds of kilometers that coincides with the spatial resolution of our analysis. Another example of highly intermittent spatial behaviour is the Persia-Tibet-Burma orogeny in the Eurasian plate. Overall, however, the average proportion of the background events in transform and convergent environments is lower than in divergent environments, as illustrated in Fig. 7(b). Similarly, the mode separation quality Q shows high intermittency within transform and convergent boundaries. It changes in the range between 0.9 and 0.95 with rather sharp spatial gradients (Fig. A3c), and has overall lower values compared to divergent zones (Fig. 7c). The observed clear spatial variations in the cluster parameters are not spurious but governed by local tectonic and physical settings. This was shown in a local study for southern California (Zaliapin & Ben-Zion 2013b); a detailed demonstration of such correlations in the global setting is outside the resolution of this study. Fig. 7 compares the three examined parameters of seismic clustering as functions of strain rate tensor's style S and second invariant I_2 . The comparison of earthquake cluster statistics with heat flow and strain rate tensor parameters using Spearman's correlation and GLM approach is illustrated in Tables 5, 6 and Figs D1(d)–(o). This further documents the coupling between the examined cluster characteristics and their correlation with the heat flow (*cf.* Fig. 3c).

4.3 Properties of earthquake clusters

The 256 993 events of the examined catalogue have been partitioned into 135 840 clusters according to the procedure of Section 3. Of those clusters, 116 228 (85.6 per cent) are singles and 19 612 (14.4 per cent) are families with sizes ranging from 2 to 6584. Tables 1 and 2 summarize the individual event classification (into singles, main shocks, foreshocks, and aftershocks) in the regular and Δ -analysis, respectively.

Fig. 8(a) shows the distribution of cluster size N for clusters in areas with high ($H > 0.2$) and low ($H < 0.2$) heat flow levels. The distribution tail in both cases can be approximated by a power law

$$S(N) = \text{Prob} [\text{cluster size} > N] \propto N^{-\alpha} \quad (8)$$

with power index $\alpha \approx 2$ in hot areas and $\alpha \approx 1$ in cold areas. The value of $\alpha \approx 1$ was previously reported for the cluster size distribution in southern California (Zaliapin & Ben-Zion 2013a). The observed difference in the cluster size distributions implies that (i) cold areas have much larger clusters—indeed, the maximal cluster size in cold areas is $\max(N | H < 0.2) = 6584$ while the maximal cluster size in hot areas is 35 times smaller, $\max(N | H \geq 0.2) = 186$; and (ii) the proportion of clusters with size $N > 10$ is larger in cold areas. Recall that the cluster size statistically increases with the maximal observed magnitude, since larger events have more offspring (e.g. Utsu 1970); it also increases as the magnitude of completeness decreases. Accordingly, the dominance of large clusters in cold regions observed in Fig. 8(a) is explained by statistically higher maximal magnitude (Fig. 1b) and better quality of catalogues (Fig. 1c) in cold regions compared to hot ones.

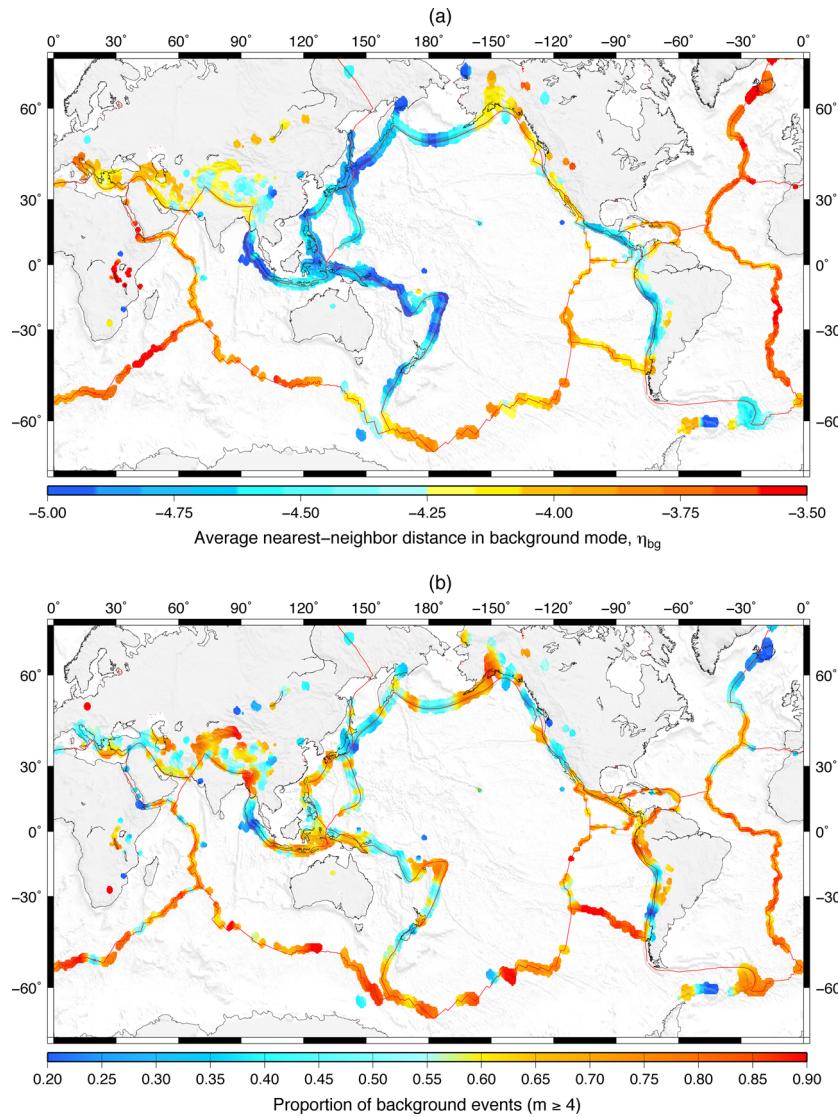


Figure 6. Global maps of selected parameters of seismic clustering. (a) Average nearest-neighbour distance $\log_{10}\eta_{bg}$ in the background mode. (b) Proportion of events in the background mode.

To eliminate effects related to differences in the largest regional magnitude, we compare the cluster size distributions in relatively hot and cold areas using Δ -analysis with $\Delta = 2$ (Fig. 8b). This approach equalizes the cluster sizes with main shocks of different magnitudes (Zaliapin & Ben-Zion 2013a), and hence should eliminate the discrepancy caused by different levels of seismic activity in cold and hot regions. The results indicate that the cluster size is *stochastically larger* in cold area. Recall that a random variable X is said to be stochastically larger than Y if the survival function of X is larger than that of Y for all arguments. Finally, we compare the cluster size distribution for clusters with intermediate-magnitude main shocks. As shown in Fig. 8(c), the size of clusters with main shock magnitude $m < 5$ is stochastically larger in hot regions. Similarly, the cluster size is stochastically larger in hot regions for clusters in all magnitude ranges below $m = 6$ (not shown). In addition, the number of foreshocks and aftershocks per cluster with main shock magnitude $m < 6$ is significantly higher in hot regions (not shown).

In summary, stochastically larger cluster size in cold regions is related to the presence of large-magnitude clusters with main shock

magnitude $m > 6$. At the same time, the size of intermediate-magnitude clusters (with main shock magnitude $m < 6$) is stochastically larger in hot regions. These two observations are consistent with the finding of Zaliapin & Ben-Zion (2013b) in southern California, who also pointed out the difference in clustering styles of the largest regional events and the rest of earthquakes, and reported larger cluster size of intermediate-magnitude clusters in hot regions.

As a particular case of small-cluster size analysis, we notice that the proportion p_S of smallest clusters—singles—among all detected clusters is higher in cold areas: $p_S(H < 0.2) = 0.86$ vs $p_S(H \geq 0.2) = 0.80$. The observed difference in proportions is highly significant, with p -value being essentially zero ($p < 10^{-16}$) according to the Fisher test (Agresti 2007). This effect is noteworthy, since the higher maximal magnitude, better quality of catalogues, and lower completeness magnitude in cold regions should decrease the number of singles (e.g. Zaliapin & Ben-Zion 2015). On the other hand, the probability of being a single is higher for small-magnitude events (e.g. it is more probable for $m = 4$ event to have no offspring than for $m = 7$). Accordingly, an increased detected proportion of small-magnitude events in cold regions (Fig. 1c) might inflate

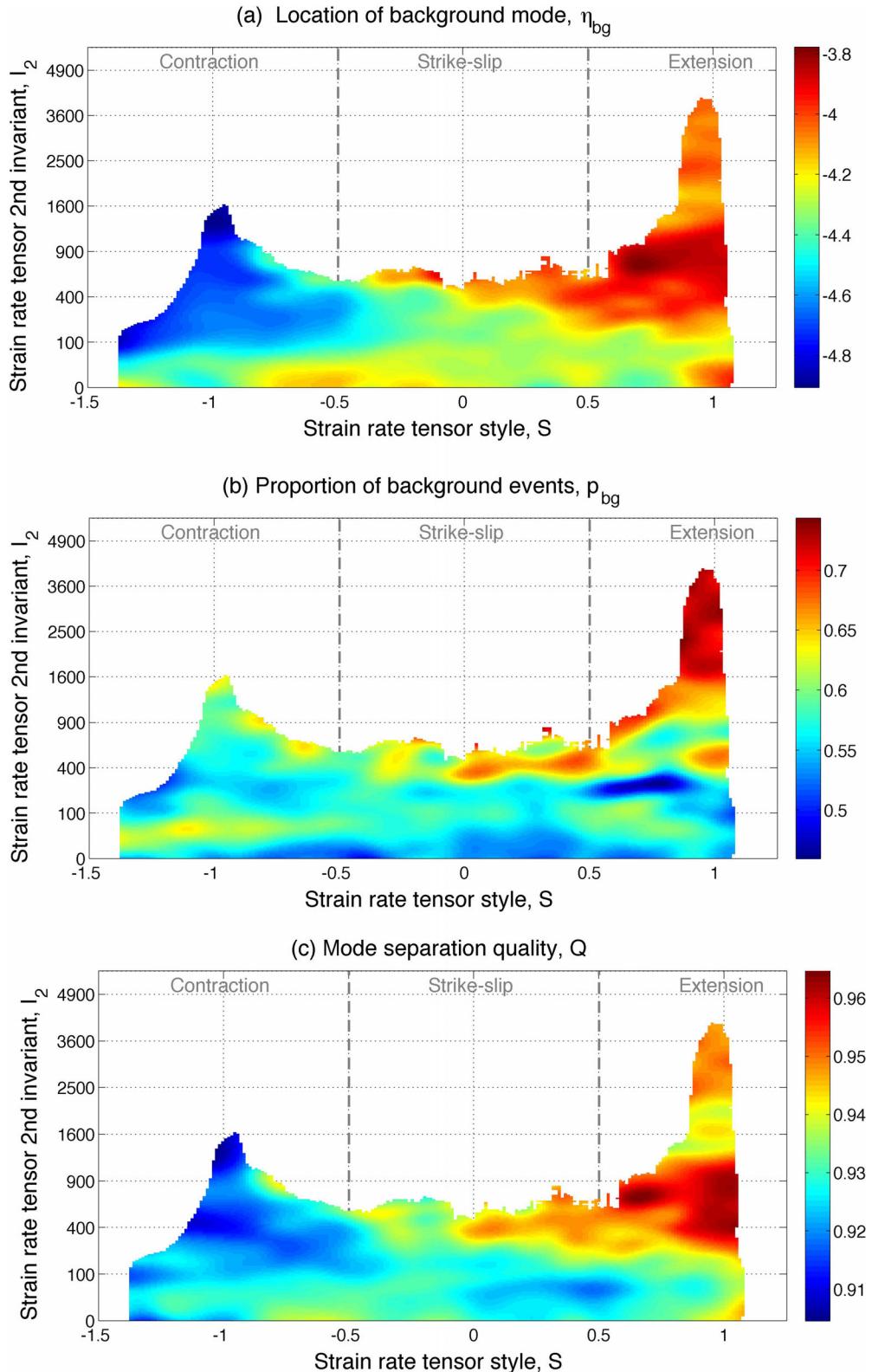


Figure 7. Statistics of background and cluster modes as functions of strain rate tensor's style S and second invariant I_2 . (a) Average nearest-neighbour distance $\log_{10} \eta_{bg}$ in the background mode. (b) Proportion of background events p_B . (c) Quality Q of background/cluster mode separation.

the proportion of singles. We demonstrate that this effect is not ultimately responsible for an increased proportion of singles in cold regions by repeating the analysis within each magnitude interval of length 0.1: [4.0, 4.1), [4.1, 4.2), etc. The proportion of

singles is higher in cold regions (not shown) for each magnitude interval from [4.0, 4.1), to [5.5, 5.6). Table 3 summarizes the results of a formal statistical testing that uses magnitude intervals of length 0.5 and confirms the statistical significance of the observed

Table 2. Statistics of singles, main shocks, aftershocks and foreshocks in the cluster Δ -analysis of events with $m \geq 2$ in the NCECD catalogue during 1975–2015.

Magnitude range	Singles				Families			
			Main shocks (= no. of families)				Aftershocks	
	No.	Per cent	No.	Per cent	No.	Per cent	No.	Per cent
All events: $m \geq 4$	404	1.9	2309	10.9	16 165	76.3	2314	10.9
$4 \leq m < 5$	–	–	–	–	9280	88.4	1,216	11.6
$5 \leq m < 6$	–	–	–	–	6067	86.9	913	13.1
$6 \leq m < 7$	390	11.8	1999	60.3	750	22.6	174	5.2
$7 \leq m < 8$	14	3.7	284	75.9	65	17.4	11	2.9
$m \geq 8$	0	0	26	89.7	3	10.3	0	0

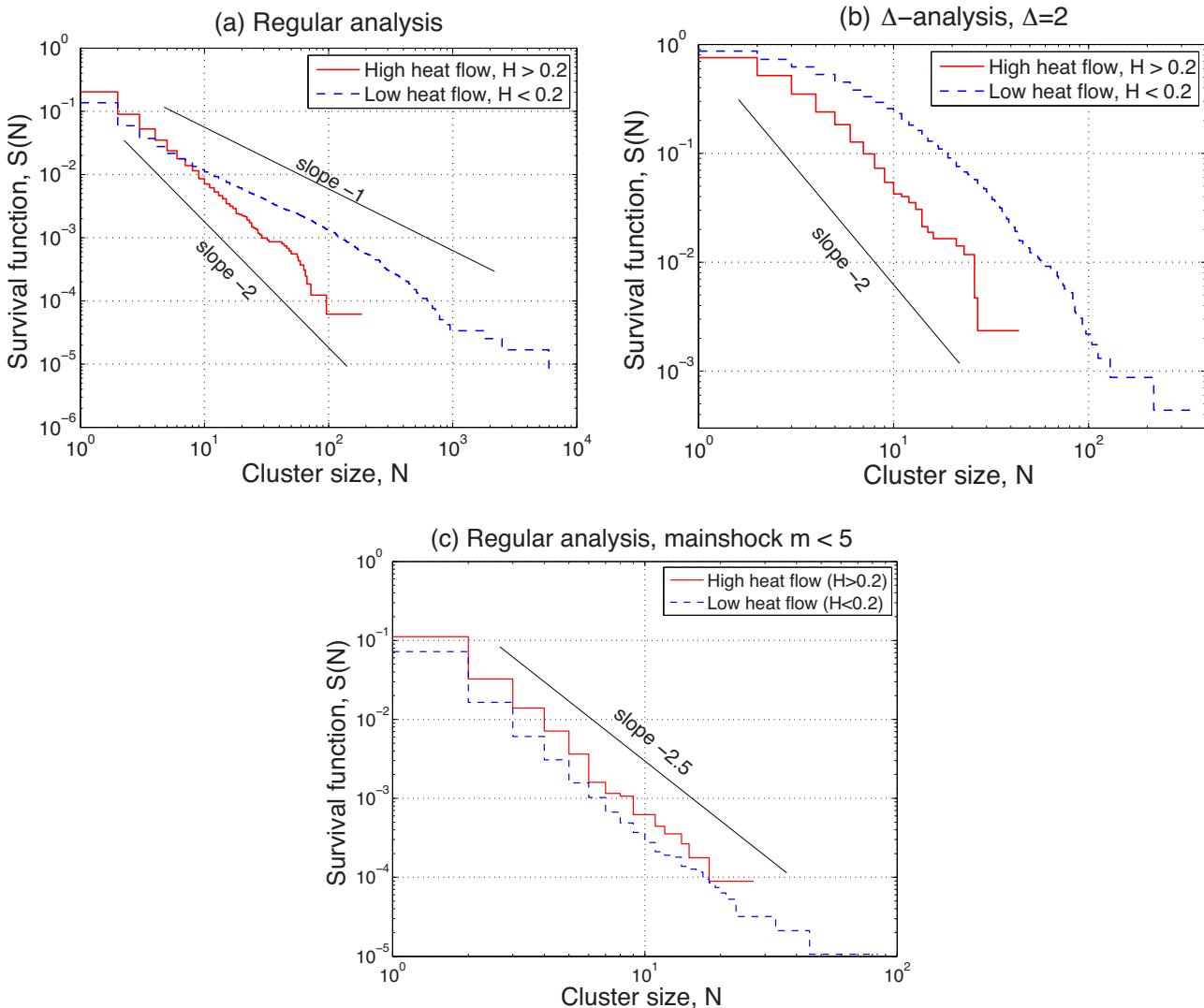


Figure 8. Distribution of cluster size N in regions with high ($H > 0.2$, red solid line) and low ($H < 0.2$, blue dashed line) values of heat flow. For families, the heat flow value is estimated at the main shock epicentre. The y-axis shows the survival function $S(N) = \text{Prob.}[\text{cluster size} > N]$. The lines that correspond to power laws $S(N) \propto N^{-\alpha}$ with indices $\alpha = 1$ and $\alpha = 2$ are shown for visual convenience. (a) Regular analysis, all clusters. (b) Delta analysis with $\Delta = 2$. (c) Regular analysis, clusters with main shock magnitude $m < 5$.

differences for events with magnitudes from 4.0 to 5.5. This analysis also demonstrates that events with $m > 5.5$ in cold regions become singles less often than those in hot regions (not shown). It is difficult to conclude with the existing data whether this effect is related to the inferior catalogue quality in hot regions or is a real physical property.

Fig. 9(a) shows the proportion p_S of singles among families in different regions. The proportion varies between 0.7 and 0.95. The highest values ($p_S > 0.9$) are typically observed within cold regions, while lowest values ($p_S < 0.75$) almost exclusively belong to hot areas. A closer examination reveals that the proportion of singles exhibits abrupt spatial fluctuations in some areas (e.g. mid-ocean

Table 3. Testing the hypothesis H_0 : Proportion of singles among the clusters is the same in cold and hot regions.

Magnitude	Cold regions, $H < 0.2$		Hot regions, $H \geq 0.2$		Fisher test p -value	Decision at 0.01 level
	No. singles/clusters	Prop. singles	No. singles/clusters	Prop. singles		
$4.0 \leq m < 4.5$	48 085/49 906	0.963	4216/4422	0.953	0.001	Reject H_0
$4.5 \leq m < 5.0$	39 654/44 655	0.888	6430/7500	0.857	7×10^{-14}	Reject H_0
$5.0 \leq m < 5.5$	12 238/16 872	0.725	2477/3638	0.681	9×10^{-8}	Reject H_0

ridges) over hundreds of kilometres. These fluctuations are caused by local tectonic and physical settings, such as transition from transform to extension faulting, but are not the focus of this study. The relation between the proportion of singles and heat flow is further illustrated in Fig. 10(a) that shows p_S as a function of strain tensor parameters (S, I_2). The large-scale averaging used in this analysis clearly demonstrates an increased proportion of singles within cold areas. The comparison of p_S with heat flow and strain rate tensor parameters using Spearman's correlation and GLM approach is summarized in Tables 5, 6 and Figs D1(p)–(r).

4.4 Foreshocks and aftershocks

The global spatial distribution of the proportion p_F of foreshocks among foreshocks and aftershocks is shown in Fig. 9(b). The proportion is visibly higher in areas with high heat flow, with typical values $p_F > 0.2$, while in areas with low heat flow the typical proportion is very small, $p_F < 0.1$. The increased production of foreshocks in hot regions is confirmed by the analysis of Fig. 10(b) that shows the value of p_F averaged for different combinations of strain rate tensor's style S and second invariant I_2 . This result corroborates the earlier regional finding of Zaliapin & Ben-Zion (2013b) that the number and proportion of foreshocks increases with the heat flow in southern California, the observation of McGuire *et al.* (2005) on large proportion of foreshocks within the swarms along the Pacific Rise Transform fault, as well as findings of Kagan *et al.* (2010) and Chu *et al.* (2011) based on fitting a branching model to the global seismicity.

Spatial patterns similar to those reported in Figs 9(a) and (b) are also seen for other examined cluster characteristics. For instance, Fig. A3(d) shows a worldwide map of the aftershock magnitude gap Δ_A defined for families with aftershocks as the difference between the magnitudes of the main shock and the largest aftershock. This analysis is done for all families with main shock magnitude $m \geq 5$. The gap is generally larger within cold regions, with typical value of $\Delta_A \approx 0.8$, while in hot regions it is typically smaller, $\Delta_A \approx 0.55$. This observation is corroborated in Fig. 10(c) that shows the aftershock magnitude gap as a function of strain rate tensor parameters (S, I_2); the domain of low gap values is similar of that of high heat flow shown in Fig. 3(c). The comparison of p_F and Δ_A with heat flow and strain rate tensor parameters using Spearman's correlation and GLM approach is summarized in Tables 5, 6 and Fig. D1(s)–(x).

We note that the values of the aftershock magnitude gap reported here are lower than the value $\Delta_A \approx 1$ suggested by the Båth law (Båth 1965; Shcherbakov & Turcotte 2004). This deflation is artificial and is due to the fact that we consider families with main shock magnitude $m \geq 5$, which is only one unit above the magnitude cut-off $m_{\min} = 4$ selected for this study. Notably, the difference in the magnitude gap Δ_A between hot and cold areas is only seen for intermediate-magnitude clusters with $m < 6$; the difference disappears for large-magnitude clusters with $m > 6$ (results not shown).

The magnitude gap is affected by the catalogue completeness magnitude, since a higher completeness magnitude leads to smaller

observed values of Δ_A (while clusters with larger magnitude gap may artificially become singles). Hence, the reported difference in magnitude gap might be influenced to some extent by inferior catalogue quality in hot areas. However, Zaliapin & Ben-Zion (2013b) reported lower magnitude gap in hot regions in a local study in southern California, where the quality of catalogues is comparable in both cold and hot regions. We therefore believe that the magnitude gap difference between hot and cold areas is a real phenomenon that will be confirmed in future studies with better catalogue quality.

4.5 Structure of earthquake families

Consider a tree T that represents an earthquake family as described in Section 3. The tree consists of a collection of vertices $V = \{v_i\}$, $i = 1, \dots, N$ each of which represents an earthquake, and edges $E = \{e_i\}$, $i = 2, \dots, N$ such that edge e_i connects earthquake i to its parent that also belongs to the tree T . Here we index the family earthquakes in the order of their occurrence time: $i < j$ if and only if $t_i < t_j$. By construction (see Section 3), the parent of the first event in the family does not belong to the same family, and hence the first event does not have an associated edge within T . All other events have a single parent from the same family. Hence each tree consists of N vertices and $N - 1$ edges. We refer to the first event in the family as the root. Denote by $C(i)$ the number of children of vertex i within T , and by $N_p = \#\{i: C(i) > 0\}$ the number of parental vertices within T . We consider two statistics of a time-oriented rooted tree that represent an earthquake family: the average family branching B and the average leaf depth d (Zaliapin & Ben-Zion 2013b). The average family branching B is the average number of offspring per parental vertex of the tree T :

$$B = \frac{1}{N_p} \sum_i C(i). \quad (9)$$

The average leaf depth d is the average number of edges between a leaf and the tree root. Namely, if d_i denotes the number of edges between vertex i and the tree root, then

$$d = \frac{1}{N - N_p} \sum_{i:C(i)=0} d_i. \quad (10)$$

It is natural to expect the leaf depth and family branching to be negatively correlated. An intuitive justification for such reciprocal relation comes from the observation that for a tree with constant branching and leaf depth, that is with $C(i) = C$ given that $C(i) > 0$ and $d_i = D$ given that $C(i) = 0$, we have $N - N_p = C^D$.

Zaliapin & Ben-Zion (2013b) showed that the values of B and d are strongly coupled with the heat flow in southern California. Specifically, the average leaf depth increases while the average branching decreases as the heat flow increases. The same general trend is observed on the global scale. Fig. 11 shows the values of d and B averaged for different family sizes N in hot and cold regions. It is seen that the average leaf depth d is significantly larger, and the family branching B is significantly smaller, in hot areas compared

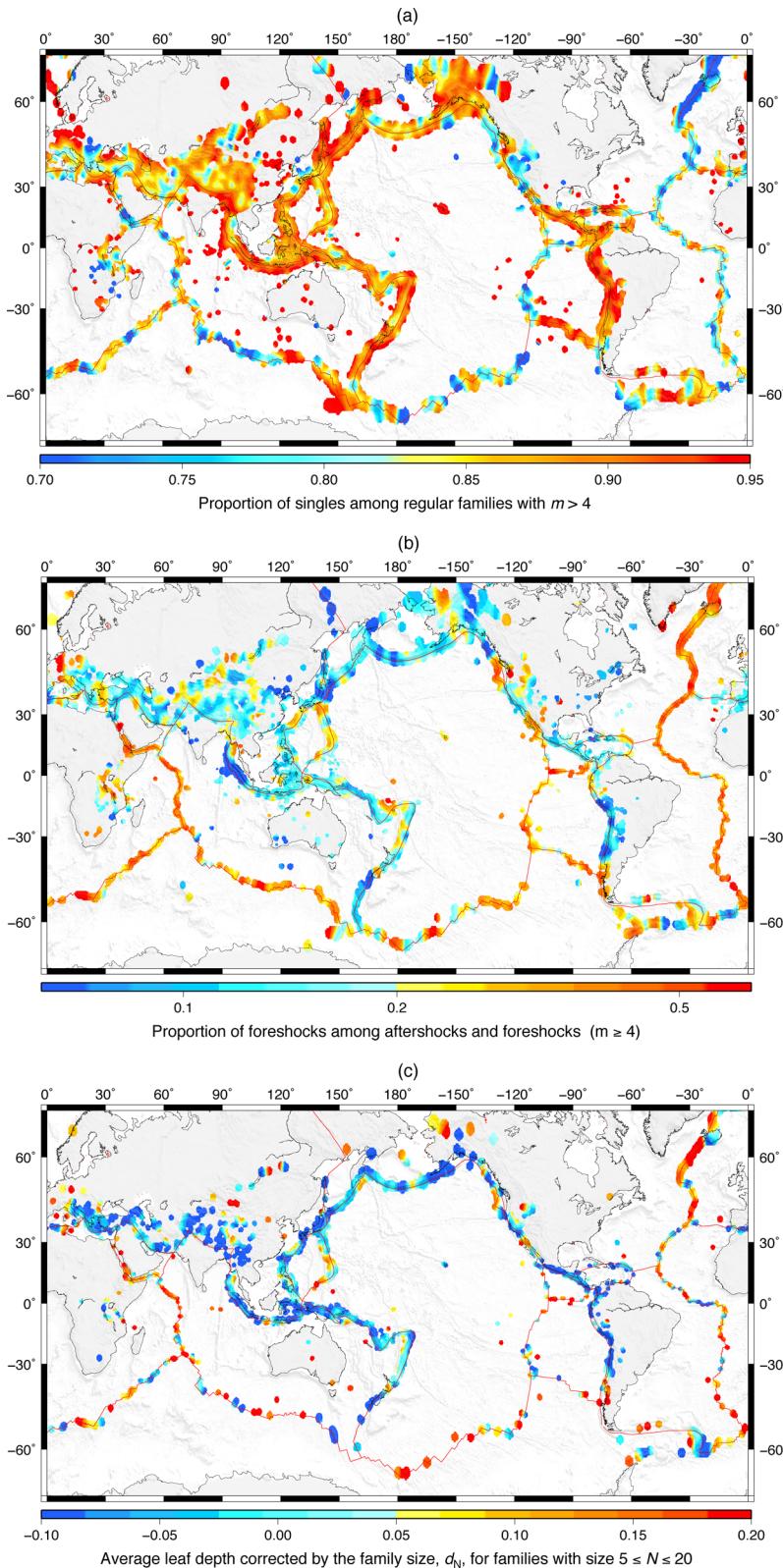


Figure 9. Global spatial distribution of selected earthquake cluster statistics. (a) Proportion p_S of singles among regular clusters. (b) Proportion p_F of foreshocks among foreshocks and aftershocks. (c) Average leaf depth corrected for cluster size, d_N , for families with size $5 \leq N \leq 20$.

to cold areas. We also notice that (i) the difference between hot and cold regions (difference between red and blue lines) is increasing with the family size and (ii) the values of both statistics increase with the family size.

Next we focus on the spatial distribution of the average leaf depth d and family branching B . The values of both statistics depend on the family size (Fig. 11), which can contaminate spatial analysis as the family size N significantly varies from region to region, as

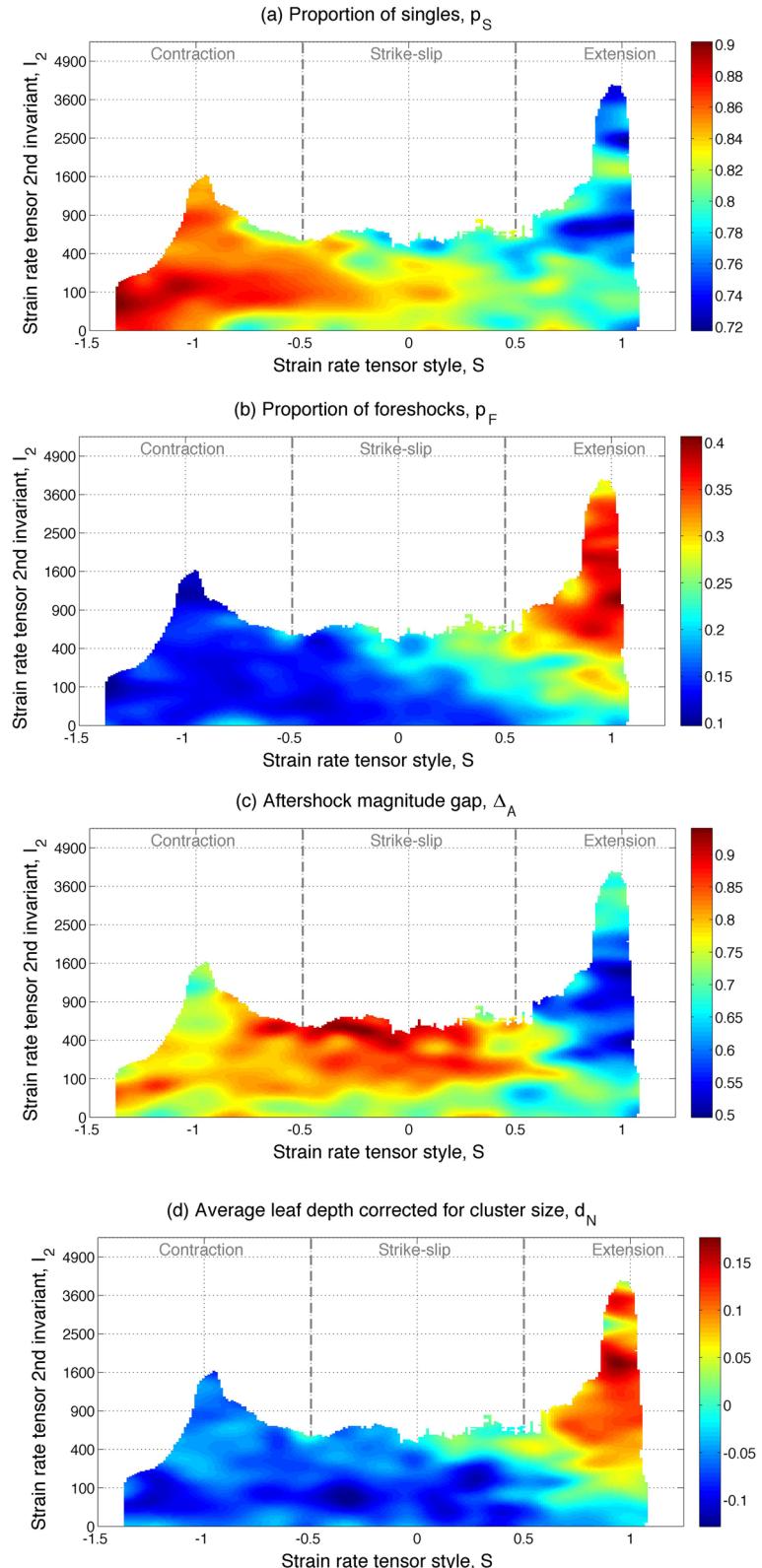


Figure 10. Average values of selected cluster statistics as a function of strain rate tensor's style S and second invariant I_2 . (a) Proportion p_S of singles among clusters. (b) Proportion p_F of foreshocks among foreshocks and aftershocks. (c) Aftershock magnitude gap Δ_A . (d) Average leaf depth corrected for cluster size, d_N , for families with size $5 \leq N \leq 20$.

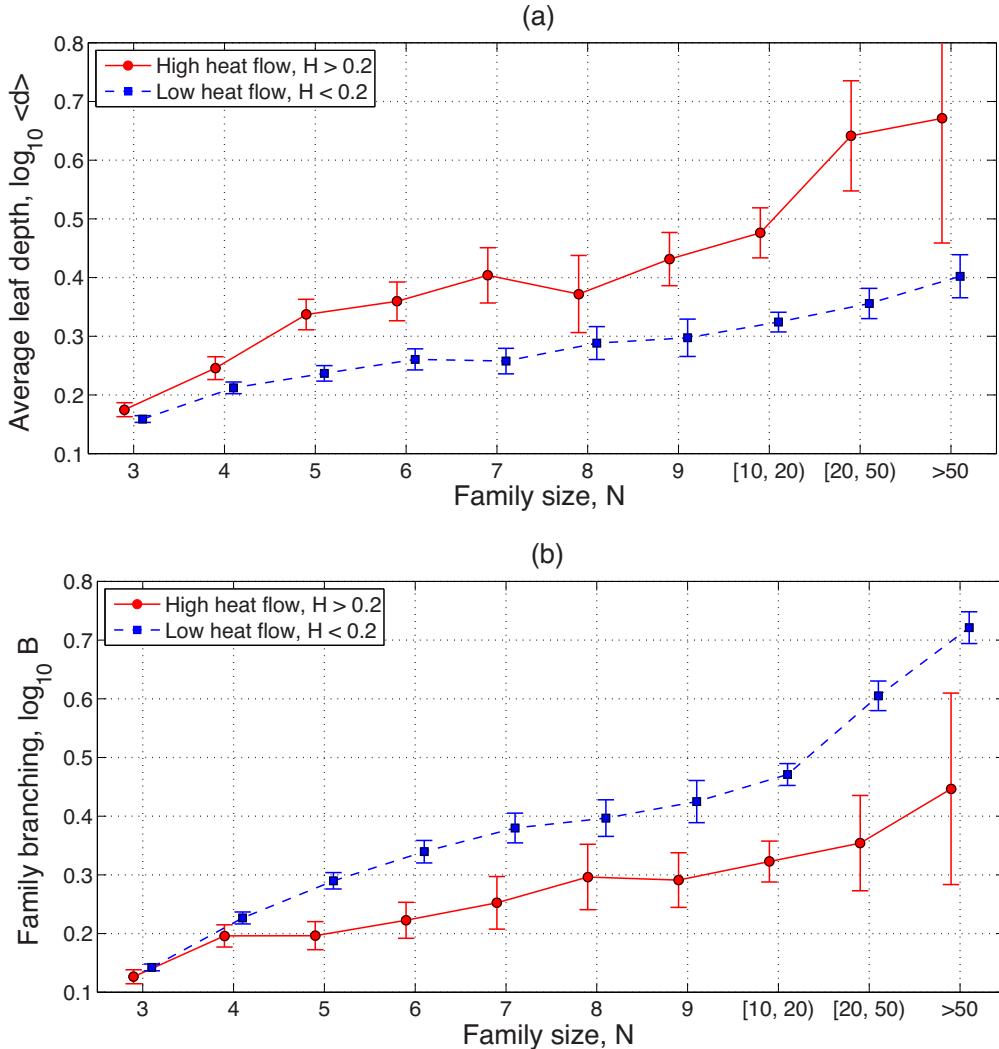


Figure 11. Average leaf depth d (a) and family branching B (b) as a function of family size N for regions with high ($H > 0.2$, red solid line and circles) and low ($H < 0.2$, blue dashed line and squares) values of the heat flow H . Notice positive trend in both examined characteristics with family size N .

documented in Fig. 8. A least-square regression analysis suggests that the examined statistics have the following relation to the family size N in the intermediate size range $5 \leq N \leq 20$:

$$\log_{10}d = 0.35 \log_{10}N + d_N, \quad \log_{10}B = 0.5 \log_{10}N + B_N, \quad (11)$$

where d_N and B_N are respective (non-Gaussian) regression errors with zero mean. Fig. 12 confirms that the average values of d_N and B_N are fairly independent of the family size in the range $5 \leq N \leq 20$. By regression construction, the error variables d_N and B_N describe variability of the initial statistics d and B , respectively, not explained by the family size N . Fig. 9(c) show the global spatial distribution of d_N for families with sizes $5 \leq N \leq 20$. This analysis confirms our earlier observation: despite some geographic fluctuations cold regions have a typical value $\log_{10} d_N \approx -0.05$ that is consistently smaller than a typical value of the hot regions, $\log_{10} d_N \approx 0.15$. The spatial distribution of B_N (Fig. A3e) has an inverted pattern: despite some geographic fluctuations cold regions have a typical value $\log_{10} B_N \approx 0$ that is consistently larger than a typical value of hot regions, $\log_{10} B_N \approx -0.15$. Figs 10(d) and A4(b) show the average values of d_N and B_N , respectively, as functions of strain rate tensor parameters S and I_2 . The comparison of d_N and B_N with heat flow and strain rate tensor parameters using Spearman's

correlation and GLM approach is summarized in Tables 5, 6 and Figs D1(y)–(ad). The results confirm that heat flow exerts the primary control on the values of these two statistics.

5 DISCUSSION

Clarifying whether earthquake dynamics follows universal laws or exhibits different forms related to physical properties of the lithosphere is among the main problems of statistical seismology. This study supports earlier results mentioned in the introduction on the existence of non-universal region-specific behaviour of seismicity. This is done by extending the analysis of Zaliapin & Ben-Zion (2013a,b) of earthquake clusters in southern California to the global scale using data from the NCEDC worldwide catalogue for the period 1975–2015. One general difficulty in demonstrating robust differences in properties of earthquakes in different regions is varied quality of seismic catalogues in different areas. This problem may be overcome by applying techniques and parameters not sensitive to variable location errors and completeness magnitudes (Zaliapin & Ben-Zion 2015). We return to this issue below.

We use the nearest-neighbour approach (Baisei & Paczuski 2004; Zaliapin *et al.* 2008; Zaliapin & Ben-Zion 2013a) to partition the

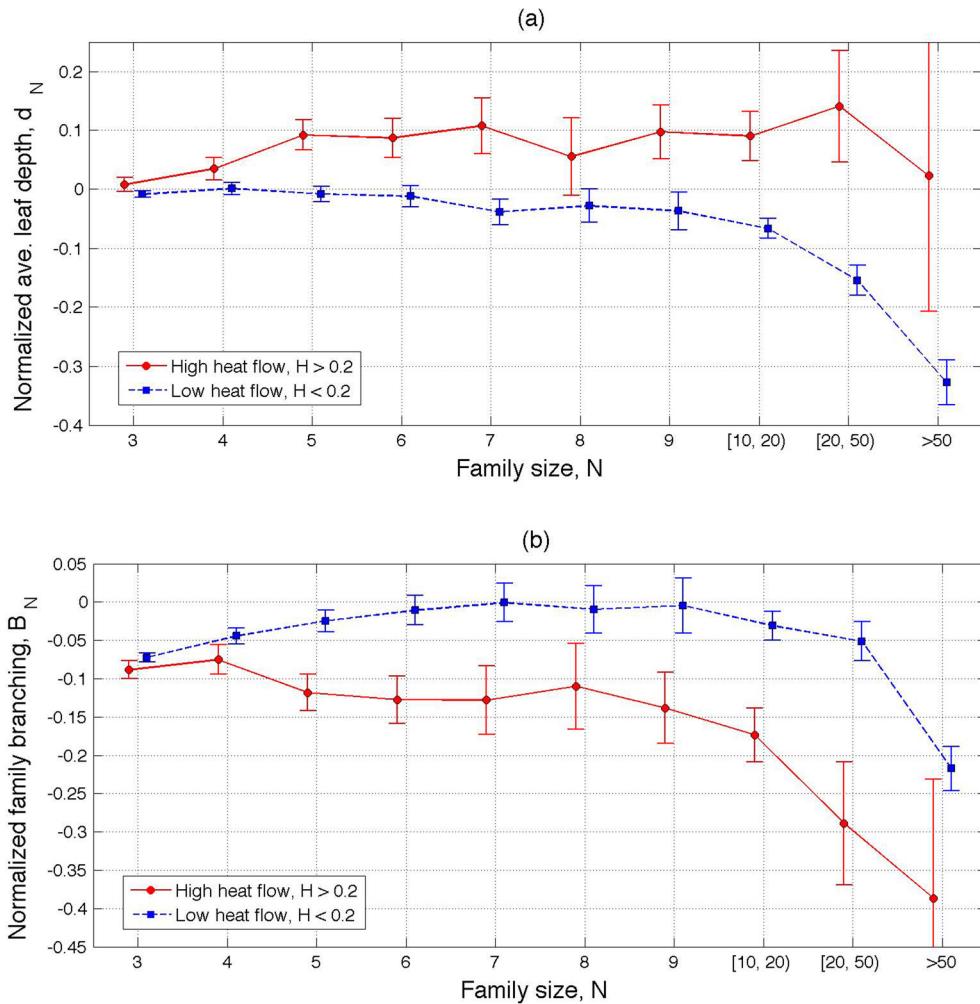


Figure 12. Average leaf depth d_N corrected for cluster size (panel a) and family branching B_N corrected for cluster size (panel b) as a function of family size N for regions with high ($H > 0.2$, red solid line and circles) and low ($H < 0.2$, blue dashed line and squares) values of the heat flow H . Notice the absence of N -dependent trend in both examined characteristics with family size for the intermediate range $5 \leq N \leq 20$.

earthquakes reported in the NCECD global catalogue into individual clusters. We then compare the worldwide space distribution of various cluster statistics with global heat flow production (Bird *et al.* 2008) and style of lithospheric deformation indicated by an estimated strain rate tensor (Kreemer *et al.* 2014). Our comparison is based on (i) spatial maps of selected characteristics in seismically active regions (Figs 1, 2, 6 and 9), (ii) averaged values of the examined characteristics as a function of the strain rate tensor style S and second invariant I_2 (Figs 3, 7 and 10), (iii) Spearman rank correlation analysis (Table 5, Appendix C) and (iv) Generalized Linear Model analysis (Table 6, Fig. D1).

We demonstrate that multiple statistics of earthquakes and seismicity clusters have spatially dependent distribution, tightly correlated with the global heat flow production: (1) earthquake intensity Λ (Figs 1a, 3b and D1a); (2) average nearest-neighbour earthquake distance η_{bg} within the background mode (Figs 4, 6a, 7a and D1d), (3) proportion p_B of background events (Figs 4, 6b, 7b and D1g), (4) quality Q of background/cluster mode separation (Figs 4, A3c, 7c and D1j); (5) threshold η_0 that separates the background and cluster modes (Figs 4, A3b and D1m); (6) rate of temporal decay of offspring events (Figs 4 and 5); (7) intensity of repeaters (Fig. 5); (8) Cluster size of intermediate magnitude ($m < 6$) main

shocks (Fig. 8); (9) proportion p_S of singles among regular families (Figs 9a, 10a and D1p); (10) proportion p_F of foreshocks among foreshocks and aftershocks (Figs 9b, 10b and D1s); (11) average leaf depth corrected for the family size, d_N (Figs 9c, 10d, 11a, 12a and D1y); (12) average family branching corrected for the family size, B_N (Figs A3e, A4b, 11b, 12b and D1ab) and (13) aftershock magnitude gap Δ_A (Figs A3d, 10c, D1y). The results are summarized in Tables 4–6.

The cluster structure and statistics estimated by our technique are subject to artefacts related to catalogue uncertainties (Zaliapin & Ben-Zion 2015). We address potential effects of catalogue incompleteness, varying earthquake intensity, and maximal magnitude on each of the examined statistics and design the analysis to minimize the possible artefacts. Notably, some of our observations (e.g. increased size of small clusters in Fig. 8c and decreased proportion of singles in Figs 9a and 10a) demonstrate a trend that goes against possible artefacts of catalogue uncertainties. Furthermore, the results of this study are consistent with those of a local analysis of southern California (Zaliapin & Ben-Zion 2013a,b) obtained with a high quality catalogue (median location error of 500 m) by Hauksson *et al.* (2012) and much lower minimal magnitude of analysis, $m_{min} = 2$. The combination of our

Table 4. Summary of examined earthquake and cluster statistics.

Statistic	Average value ^a in Cold regions ($H < 0.2$)		Section	Figures
	Hot regions ($H > 0.2$)			
Earthquake intensity, Δ	1.36	0.33	2.1, 4.1	1a, 3b
Background nearest-neighbour distance, $\log_{10}(\eta_{\text{bg}})$	-4.44	-3.91	4.2	4, 6a, 7a
Proportion of background events, p_{bg}	0.57	0.67	4.2	4, 6b, 7b
Quality of mode separation, Q	0.93	0.96	4.2	4, A3c, 7c
Threshold between background and cluster modes, $\log_{10}(\eta_0)$	-5.40	-5.07	4.2	4, A3b
Rate of temporal decay of offspring	low	high	4.1	4, 5
Intensity of repeaters	low	high	4.1	5
Cluster size, N (for main shocks $m < 6$)	1.28	1.44	4.3	8
Proportion of singles, p_S	0.87	0.82	4.3	9a, 10a
Proportion of foreshocks, p_F	0.19	0.33	4.4	9b, 10b
Aftershock magnitude gap, Δ_A	0.79	0.67	4.4	10c, A3d
Size-corrected leaf depth, d_N	-0.02	0.07	4.5	9c, 10d, 11a, 12a
Size-corrected branching, B_N	-0.07	-0.16	4.5	11b, 12b, A3e, A3b

^aThe average values are given here to illustrate the trend of changes between cold and hot zones. All reported differences are highly significant according to ANOVA test (not shown). The average differences reported here are typically lower than those observed for individual hot/cold regions in the worldwide maps and maps in the strain rate tensor coordinates (S, I_2).

Table 5. Spearman's correlation between the examined cluster statistics and heat flow, deformation type.

Variable	Heat flow, H	Tensor style, S	Strain rate second invariant, I_2
Earthquake intensity, Δ	-0.38 ^a	-0.40	0.12
Background nearest-neighbour distance, $\log_{10}(\eta_{\text{bg}})$	0.46	0.43	0.001 (0.88)
Proportion of background events, p_{bg}	0.23	0.07	0.16
Quality of mode separation, Q	0.34	0.27	0.09
Threshold between background and cluster modes, $\log_{10}(\eta_0)$	0.25	0.30	-0.10
Proportion of singles, p_S	-0.29	-0.21	-0.04 (10^{-7})
Proportion of foreshocks, p_F	0.38	0.19	0.37
Aftershock magnitude gap, Δ_A	-0.12	-0.09	-0.03 (10^{-4})
Size-corrected leaf depth, d_N	0.33	0.22	0.13
Size-corrected branching, B_N	-0.30	-0.19	-0.09

^aThe P -values are indicated in parentheses; no P -value is indicated in case $P < 10^{-10}$. See Appendix C for definitions.

Table 6. Generalized Linear Model Analysis^a: Coefficient of determination for a GLM $E[Y] = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 I_{\{X > \text{mean}(X)\}}$.

Response, Y	Predictor, X		
	Heat flow, H	Tensor style, S	Strain rate second invariant, I_2
Earthquake intensity, Δ	0.20	0.16	0.12
Background nearest-neighbour distance, $\log_{10}(\eta_{\text{bg}})$	0.29	0.19	0.05
Proportion of background events, p_{bg}	0.09	0.09	0.06
Quality of mode separation, Q	0.18	0.09	0.04
Threshold between background and cluster modes, $\log_{10}(\eta_0)$	0.09	0.08	0.02
Proportion of singles, p_S	0.04	0.03	0.06
Proportion of foreshocks, p_F	0.20	0.08	0.11
Aftershock magnitude gap, Δ_A	0.04	0.02	0.01
Size-corrected leaf depth, d_N	0.14	0.08	0.05
Size-corrected branching, B_N	0.10	0.06	0.03

^aSee Appendix D for definitions and further detail.

global results and those obtained in the detailed regional study of Zaliapin & Ben-Zion (2013a,b) indicate clear dependency of seismic clustering on the heat flow in the region. The results are consistent with those obtained by Yang & Ben-Zion (2009) and Enescu *et al.* (2009) by different statistical analyses, and with theoretical expectations based on a viscoelastic damage rheology model (Ben-Zion & Lyakhovsky 2006).

The overall picture emerging from these studies indicate that there exist two primary types of earthquake clustering. (i) Brittle fracture in cold regions results in *burst-like clusters* characterized by a prominently large main shock that happens in the very beginning of the sequence and triggers multiple offspring of smaller

magnitude occurring in a small number of generations and decaying until merging with the background seismicity. Heterogeneity of stress/strength field (which might create fracture barriers) and generally larger failure threshold in cold regions reduce the triggering potential, which particularly affects small-to-intermediate magnitude events and results in lower overall offspring production, smaller cluster size, and increased proportion of singles. At the same time, large events ($m > 6$) have sufficient potential not only for overcoming the failure threshold but also for significantly disturbing the neighbouring stress/strength field and generating long slow decaying aftershock sequences. (ii) Brittle–ductile failure mechanisms in hot regions result in *swarm-like clusters* that lack a single

prominent main shock. Instead, they gradually develop, event-by-event, by triggering earthquakes of comparable magnitudes. The offspring span generally multiple generations in such clusters, but decay overall much faster creating a notable temporal gap between offspring activity in a fading cluster and future background events. The stress/strength field in hot areas is more homogeneous, and the failure threshold is generally lower than in cold regions, which facilitate triggering potential and allows small-to-medium magnitude events to have offspring. This leads to increased size of clusters (for small-to-intermediate main shocks) and decreased proportion of singles.

Our findings on preferential occurrence of swarm-like clusters in hot regions, (prominently including the mid-ocean ridges transform areas) and their general statistical properties, are consistent with previous large-scale analyses of oceanic swarms (e.g. McGuire *et al.* 2005; Roland & McGuire 2009). Furthermore, our results on seismic intensity, cluster size distribution, background proportions, and temporal decay of offspring parallel the findings of Kagan *et al.* (2010) and Chu *et al.* (2011) on variations of Epidemic-Type After-shock Sequence (ETAS) model parameters across different tectonic zones. Their tectonic zones 1 (active continents) and 4 (trenches) roughly correspond to the cold areas of our study; while zones 2 (slow-spreading ridges) and 3 (fast-spreading ridges) generally correspond to the hot areas.

We propose the effective viscosity of the lithosphere to be the main control of the style of earthquake clustering. This is consistent with interpretations that swarms reflect migration of fluids or creep (e.g. Hill 1977; Hainzl 2004; Hainzl & Ogata 2005; Vidale & Shearer 2006; Chen *et al.* 2012), as increasing heat flow and fluid content will reduce the effective viscosity and lead to brittle-ductile deformation that may include creep. However, the explanations based on fluid flow and slow slip events appeal to specific detailed micromechanisms involving in general many parameters (and expected to have additional consequences beyond swarm generation). In contrast, the simpler term effective viscosity involves a continuum-based macroscopic description of the behaviour in a region (Ben-Zion & Lyakhovsky 2006), not committing to any micro mechanism.

There is no sharp transition between the clusters of the two primary types, and global seismicity exhibits a wide variety of clustering forms. Nevertheless, the cold and hot environments are clearly distinguishable by the average values and distributions of multiple cluster characteristics. Our analysis also suggests that the type and intensity of lithospheric transformation, as measured by the strain rate tensor, play a secondary role in determining the earthquake cluster style (see, in particular, Tables 5, 6 and Fig. D1). Examining multiple complementary statistics not sensitive to artefacts produced by common catalogue deficiencies (Zaliapin & Ben-Zion 2015) allows us to have confidence that our main findings will remain valid in future analyses with improved catalogue quality and alternative cluster identification techniques (e.g. Roland & McGuire 2009; Zhang & Shearer 2016). Additional analyses of seismicity accounting for non-universal space-dependent properties, combined with geodetic data on aseismic deformation and modelling, can improve further the understanding of earthquake dynamics and provide refined information for seismic hazard assessments.

ACKNOWLEDGEMENTS

We are grateful to Peter Bird for making the heat flow data publicly available. We also thank Corne Kreemer for sharing with us

the global strain rate model data. The paper benefitted from comments by two anonymous reviewers and Editor Eiichi Fukuyama. Some plots were made using the Generic Mapping Tools version 4.5.8 (www.soest.hawaii.edu/gmt; Wessel & Smith 1991). This research was supported by the Southern California Earthquake Center (Contribution No. 6486). Southern California Earthquake Center is funded by National Science Foundation Cooperative Agreement EAR-1033462 & United States Geological Survey Cooperative Agreement G12AC20038.

REFERENCES

- Agresti, A., 2007. *An Introduction to Categorical Data Analysis*, Wiley, 400 pp.
- Ampuero, J.P. & Ben-Zion, Y., 2008. Cracks, pulses and macroscopic asymmetry of dynamic rupture on a bimaterial interface with velocity-weakening friction, *Geophys. J. Int.*, **173**(2), 674–692.
- Baiesi, M. & Paczuski, M., 2004. Scale-free networks of earthquakes and aftershocks, *Phys. Rev. E*, **69**, 066106, doi:10.1103/PhysRevE.69.066106.
- Bailey, I.W., Becker, T.W. & Ben-Zion, Y., 2009. Patterns of co-seismic strain computed from southern California focal mechanisms, *Geophys. J. Int.*, **177**(3), 1015–1036.
- Bailey, I.W., Ben-Zion, Y., Becker, T.W. & Holschneider, M., 2010. Quantifying focal mechanism heterogeneity for fault zones in central and southern California, *Geophys. J. Int.*, **183**(1), 433–450.
- Bak, P. & Tang, C., 1989. Earthquakes as a self-organized critical phenomenon, *J. geophys. Res.*, **94**, 15 635–15 637.
- Bath, M., 1965. Lateral inhomogeneities of the upper mantle, *Tectonophysics*, **2**, 483–514.
- Ben-Zion, Y., 2001. Dynamic ruptures in recent models of earthquake faults, *J. Mech. Phys. Solids*, **49**(9), 2209–2244.
- Ben-Zion, Y., 2008. Collective behavior of earthquakes and faults: continuum-discrete transitions, progressive evolutionary changes, and different dynamic regimes, *Rev. Geophys.*, **46**(4), RG4006, doi:10.1029/2008RG000260.
- Ben-Zion, Y. & Lyakhovsky, V., 2006. Analysis of aftershocks in a lithospheric model with seismogenic zone governed by damage rheology, *Geophys. J. Int.*, **165**(1), 197–210.
- Bird, P., Liu, Z. & Rucker, W.K., 2008. Stresses that drive the plates from below: definitions, computational path, model optimization, and error analysis, *J. geophys. Res.*, **113**, B11406, doi:10.1029/2007JB005460.
- Chen, X., Shearer, P.M. & Abercrombie, R.E., 2012. Spatial migration of earthquakes within seismic clusters in Southern California: evidence for fluid diffusion, *J. geophys. Res.*, **117**(B4), B04301, doi:10.1029/2011JB008973.
- Chu, A., Schoenberg, F.P., Bird, P., Jackson, D.D. & Kagan, Y.Y., 2011. Comparison of ETAS parameter estimates across different global tectonic zones, *Bull. seism. Soc. Am.*, **101**(5), 2323–2339.
- Corral, A., 2004. Long-term clustering, scaling, and universality in the temporal occurrence of earthquakes, *Phys. Rev. Lett.*, **92**, 108501, doi:10.1103/PhysRevLett.92.108501.
- Davidsen, J., Gu, C. & Baiesi, M., 2015. Generalized Omori–Utsu law for aftershock sequences in southern California, *Geophys. J. Int.*, **201**(2), 965–978.
- Dempster, A.P., Laird, N.M. & Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm, *J. R. Stat. Soc. B (Methodological)*, 1–38.
- Enescu, B., Hainzl, S. & Ben-Zion, Y., 2009. Correlations of seismicity patterns in Southern California with surface heat flow data, *Bull. seism. Soc. Am.*, **99**(6), 3114–3123.
- Gu, C., Schumann, A.Y., Baiesi, M. & Davidsen, J., 2013. Triggering cascades and statistical properties of aftershocks, *J. geophys. Res.*, **118**(8), 4278–4295.
- Gutenberg, B. & Richter, C.F., 1944. Frequency of earthquakes in California, *Bull. seism. Soc. Am.*, **34**(4), 185–188.

- Hainzl, S., 2004. Seismicity patterns of earthquake swarms due to fluid intrusion and stress triggering, *Geophys. J. Int.*, **159**(3), 1090–1096.
- Hainzl, S. & Ogata, Y., 2005. Detecting fluid signals in seismicity data through statistical earthquake modeling, *J. geophys. Res.*, **110**(B5), B05S07, doi:10.1029/2004JB003247.
- Hauksson, E., Yang, W. & Shearer, P.M., 2012. Waveform relocated earthquake catalog for Southern California (1981 to June 2011), *Bull. seism. Soc. Am.*, **102**(5), 2239–2244.
- Hicks, A., 2011. Clustering in multidimensional spaces with applications to statistical analysis of earthquake clustering, in *MSc Thesis*, Department of Mathematics and Statistics, University of Nevada, Reno, August, 2011.
- Hill, D.P., 1977. A model for earthquake swarms, *J. geophys. Res.*, **82**(8), 1347–1352.
- Kagan, Y.Y., 1999. Universality of the seismic moment-frequency relation, in *Seismicity Patterns, their Statistical Significance and Physical Meaning*, pp. 537–573, eds Wyss, M., Shimazaki, K. & Ito, A., Birkhäuser Basel.
- Kagan, Y.Y., Bird, P. & Jackson, D.D., 2010. Earthquake patterns in diverse tectonic zones of the globe, *Pure appl. Geophys.*, **167**(6–7), 721–741.
- Keilis-Borok, V.I., 1990. The lithosphere of the Earth as a nonlinear system with implications for earthquake prediction, *Rev. Geophys.*, **28**(1), 19–34.
- Kendall, M.G. & Stuart, A., 1973. *The Advanced Theory of Statistics, Volume 2: Inference and Relationship*, Griffin, ISBN 0-85264-215-6, Sections 31.19, 31.21.
- Kreemer, C., Blewitt, G. & Klein, E.C., 2014. A geodetic plate motion and global strain rate model, *Geochem., Geophys., Geosyst.*, **15**(10), 3849–3889.
- McGuire, J.J., Boettcher, M.S. & Jordan, T.H., 2005. Foreshock sequences and short-term earthquake predictability on East Pacific Rise transform faults, *Nature*, **434**(7032), 457–461.
- NCECD, 2015. *Northern California Earthquake Data Center*, UC Berkeley Seismological Laboratory, Dataset, doi:10.7932/NCECD.
- Omori, F., 1894. On after-shocks of earthquakes, *J. Coll. Sci. Imp. Univ. Tokyo*, **7**, 111–200.
- Reverso, T., Marsan, D. & Helmstetter, A., 2015. Detection and characterization of transient forcing episodes affecting earthquake activity in the Aleutian Arc system, *Earth planet. Sci. Lett.*, **412**, 25–34.
- Roland, E. & McGuire, J.J., 2009. Earthquake swarms on transform faults, *Geophys. J. Int.*, **178**(3), 1677–1690.
- Rubin, A.M., 2002. Aftershocks of microearthquakes as probes of the mechanics of rupture, *J. geophys. Res.*, **107**(B7), doi:10.1029/2001JB000496.
- Rubin, A.M. & Gillard, D., 2000. Aftershock asymmetry/rupture directivity among central San Andreas fault microearthquakes, *J. geophys. Res.*, **105**(B8), 19 095–19 109.
- Rundle, J.B., Turcotte, D.L., Shcherbakov, R., Klein, W. & Sammis, C., 2003. Statistical physics approach to understanding the multiscale dynamics of earthquake fault systems, *Rev. Geophys.*, **41**(4), 1019, doi:10.1029/2003RG000135.
- Schoenball, M., Davatzes, N.C. & Glen, J.M.G., 2015. Differentiating induced and natural seismicity using space-time-magnitude statistics applied to the Coso Geothermal field, *Geophys. Res. Lett.*, **42**, doi:10.1002/2015GL064772.
- Shcherbakov, R. & Turcotte, D.L., 2004. A modified form of Bath's law, *Bull. seism. Soc. Am.*, **94**, 1968–1975.
- Sornette, A. & Sornette, D., 1989. Self-organized criticality and earthquakes, *EPL (Europhys. Lett.)*, **9**(3), 197.
- Turcotte, D.L. & Malamud, B.D., 2004. Landslides, forest fires, and earthquakes: examples of self-organized critical behavior, *Physica A: Stat. Mech. Appl.*, **340**(4), 580–589.
- Utsu, T., 1970. Aftershocks and earthquake statistics (II)–Further investigation of aftershocks and other earthquake sequences based on a new classification of earthquake sequences, *J. Fac. Sci. Hokkaido Univ. Ser. VII*, **3**, 197–266.
- Utsu, T., Ogata, Y. & Matsuura, R., 1995. The centenary of the Omori formula for a decay law of aftershock activity, *J. Phys. Earth*, **43**(1), 1–33.
- Vidale, J.E., Boyle, K.L. & Shearer, P.M., 2006. Crustal earthquake bursts in California and Japan: their patterns and relation to volcanoes, *Geophys. Res. Lett.*, **33**, L20313, doi:10.1029/2006GL027723.
- Vidale, J.E. & Shearer, P.M., 2006. A survey of 71 earthquake bursts across southern California: exploring the role of pore fluid pressure fluctuations and aseismic slip as drivers, *J. geophys. Res.*, **111**, B05312, doi:10.1029/2005JB004034.
- Weertman, J., 1980. Unstable slippage across a fault that separates elastic media of different elastic constants, *J. geophys. Res.*, **85**(B3), 1455–1461.
- Wessel, P. & Smith, W.H., 1991. Free software helps map and display data, *EOS, Trans. Am. geophys. Un.*, **72**(441), 445–446.
- Yang, W. & Ben-Zion, Y., 2009. Observational analysis of correlations between aftershock productivities and regional conditions in the context of a damage rheology model, *Geophys. J. Int.*, **177**(2), 481–490.
- Zaliapin, I. & Ben-Zion, Y., 2011. Asymmetric distribution of aftershocks on large faults in California, *Geophys. J. Int.*, **185**(3), 1288–1304.
- Zaliapin, I. & Ben-Zion, Y., 2013a. Earthquake clusters in southern California I: identification and stability, *J. geophys. Res.*, **118**(6), 2847–2864.
- Zaliapin, I. & Ben-Zion, Y., 2013b. Earthquake clusters in southern California II: classification and relation to physical properties of the crust, *J. geophys. Res.*, **118**(6), 2865–2877.
- Zaliapin, I. & Ben-Zion, Y., 2015. Artifacts of earthquake location errors and short-term incompleteness on seismicity clusters in southern California, *Geophys. J. Int.*, **202** 1949–1968.
- Zaliapin, I. & Ben-Zion, Y., 2016. Discriminating characteristics of tectonic and human-induced seismicity, *Bull. seism. Soc. Am.*, in press.
- Zaliapin, I., Gabrielov, A., Keilis-Borok, V. & Wong, H., 2008. Clustering analysis of seismicity and aftershock identification, *Phys. Rev. Lett.*, **101**, 018501, doi:10.1103/PhysRevLett.101.018501.
- Zhang, Q. & Shearer, P.M., 2016. A new method to identify earthquake swarms applied to seismicity near the San Jacinto Fault, California, *Geophys. J. Int.*, **205**(2), 995–1005.

APPENDIX A: SELECTED CHARACTERISTICS OF LITHOSPHERE AND SEISMICITY

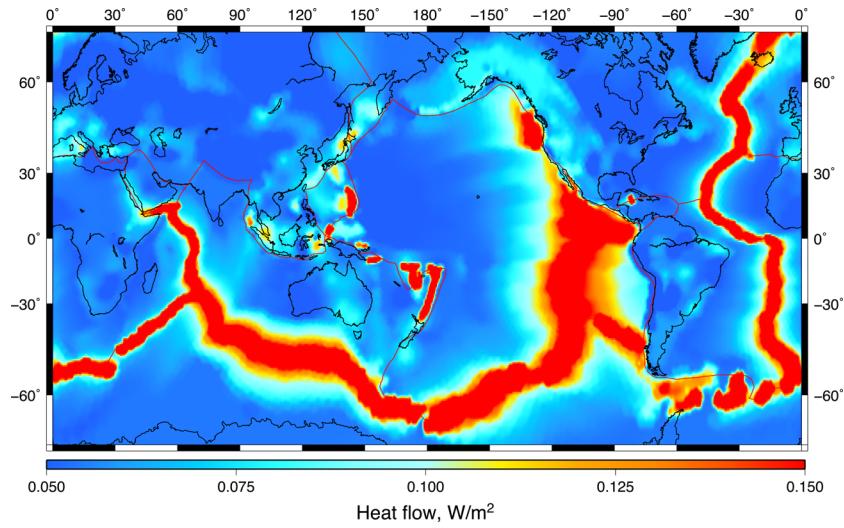


Figure A1. Global heat flow distribution, after (Bird *et al.* 2008). The map is clipped at $H = 0.16 \text{ W m}^{-2}$, while the maximal reported value if $H = 0.3 \text{ W m}^{-2}$.

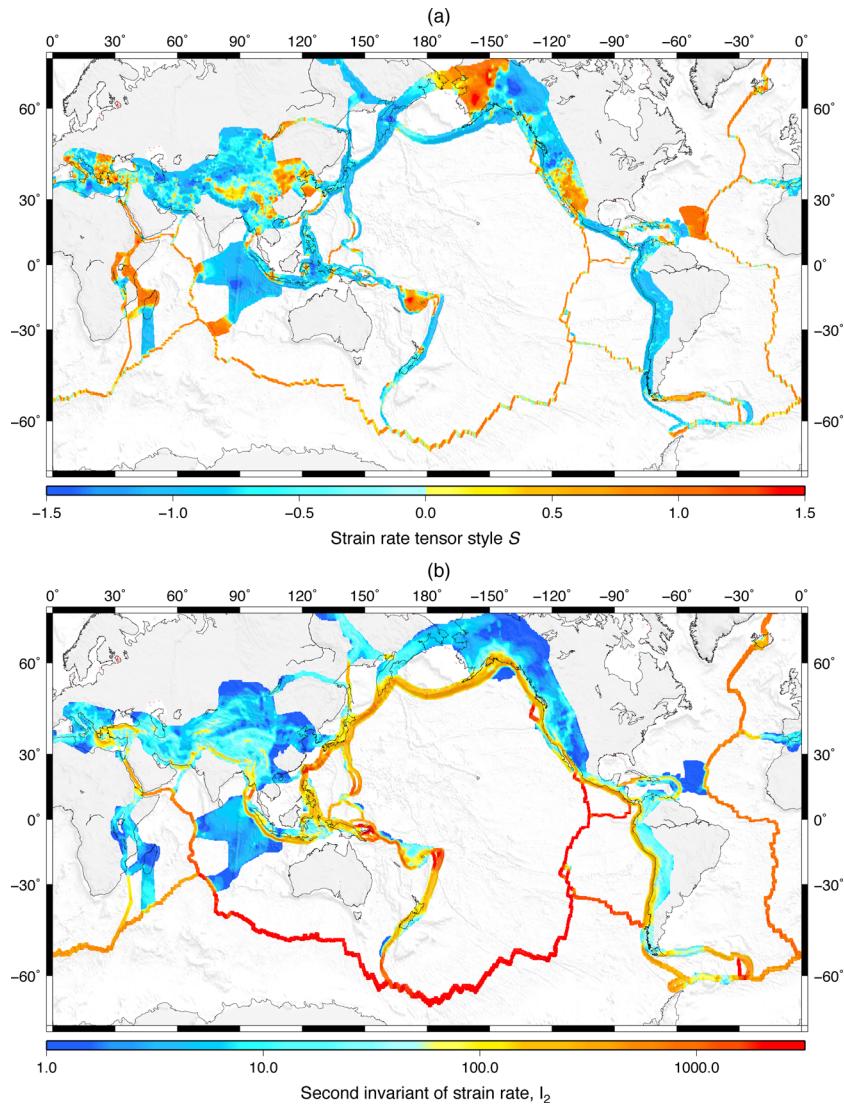


Figure A2. Characteristics of the strain rate tensor (Kreemer *et al.* 2014). (a) Strain rate tensor style S of eq. (3). (b) Strain rate tensor second invariant I_2 of eq. (2).

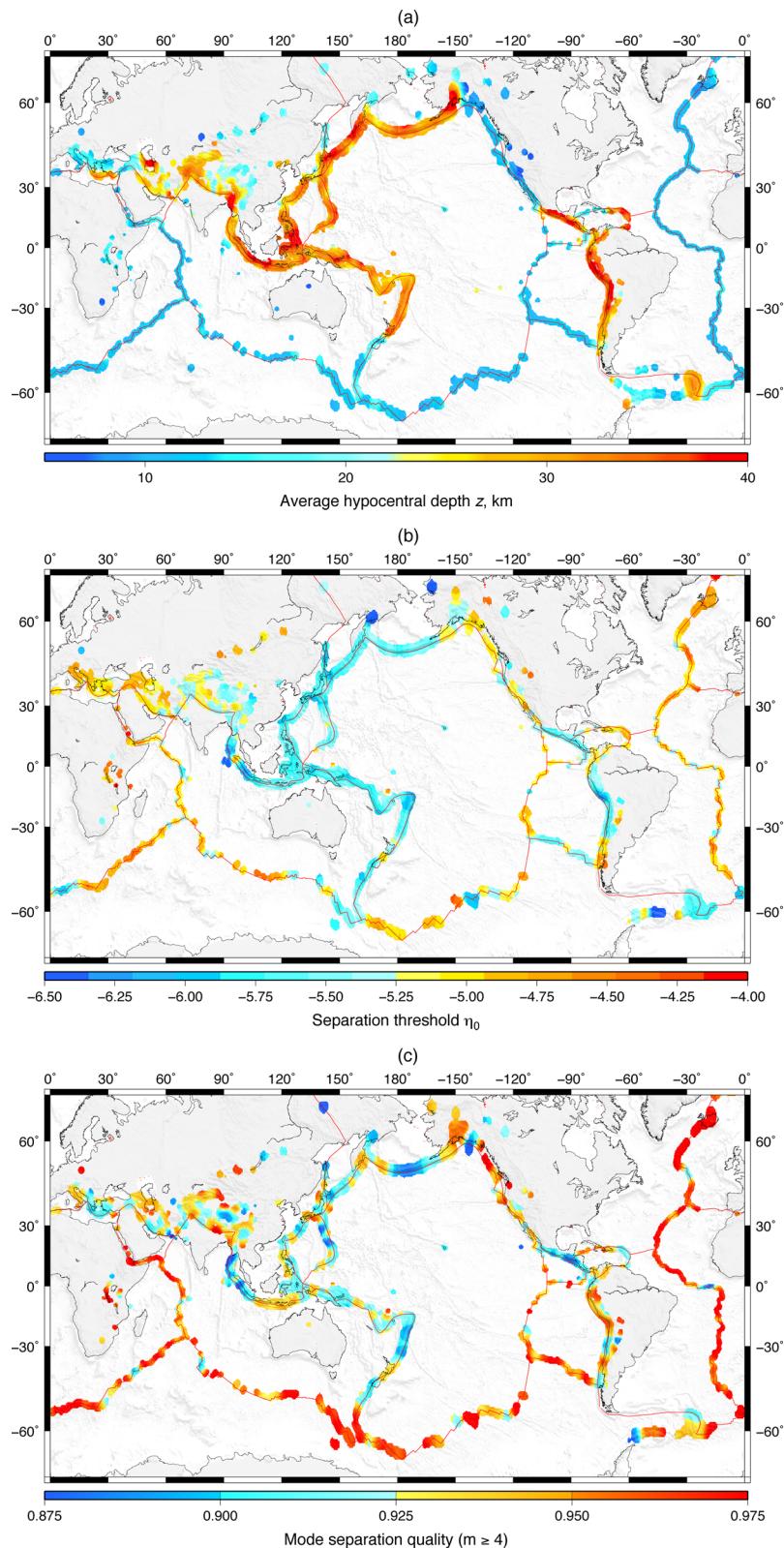
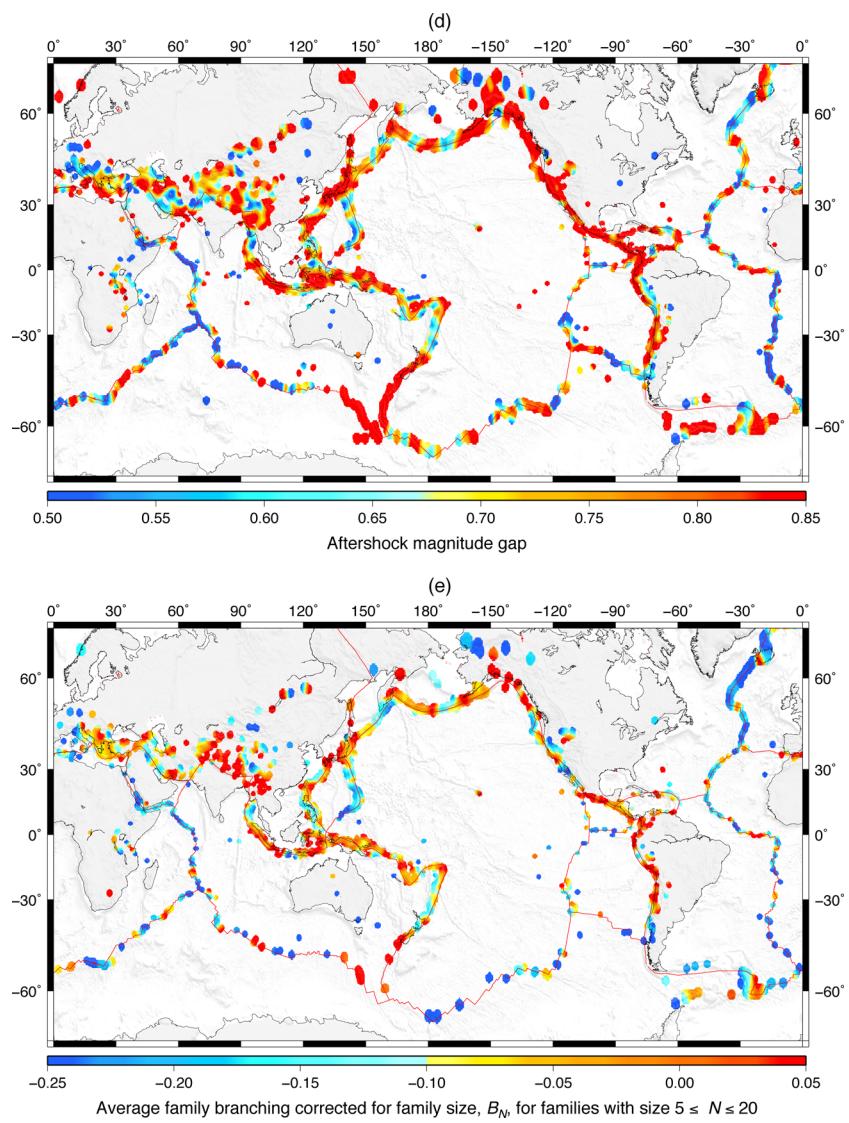


Figure A3. Worldwide spatial distributions of selected earthquake and cluster statistics. (a) Average hypocentral depth, z . (b) Threshold η_0 that separates the background and cluster modes. (c) Quality Q of separation between the background and cluster modes. (d) Aftershock magnitude gap Δ_A . (e) Average family branching B_N corrected for cluster size, for families with size $5 \leq N \leq 20$.

**Figure A3 (Continued).**

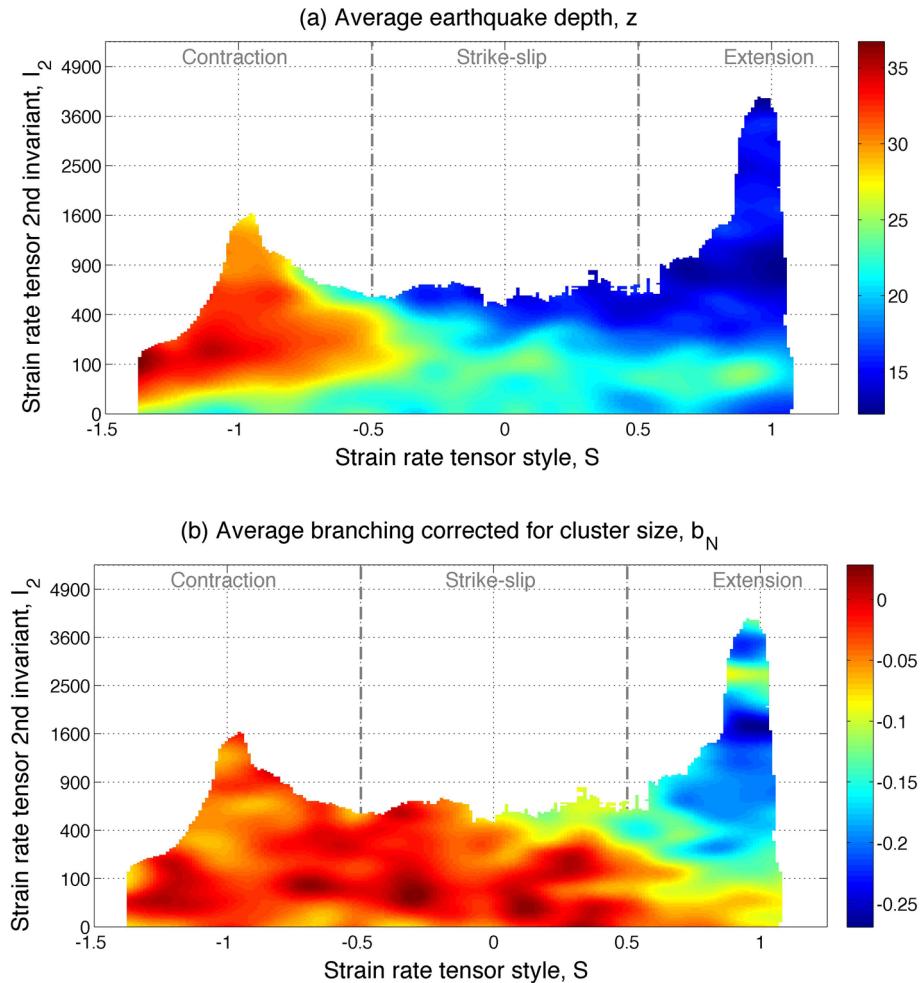


Figure A4. Average values of selected cluster statistics as a function of strain rate tensor's style S and second invariant I_2 . (a) Average hypocentral depth, z . (b) Average family branching B_N corrected for cluster size, for families with size $5 \leq N \leq 20$.

APPENDIX B: PRODUCING SPATIAL MAPS

In this study we produce spatial maps for selected characteristics of the lithosphere, earthquakes, and earthquake clusters. These characteristics can be partitioned into the following four types:

(i) *Individual earthquake characteristics*—magnitude m , hypocentral depth z , Baiesi-Paczuski distance η_{ij} to the parent, foreshock index $I_F(i)$ that equals to unity if earthquake i is a foreshock and zero otherwise, and background index $I_B(i)$ that equals to unity if earthquake i belongs to background population, and zero otherwise.

(ii) *Regional characteristics*—earthquake intensity Λ , maximal observed magnitude m_{\max} , the position η_{bg} of the background population, and quality Q of separation between background and cluster mode.

(iii) *Individual cluster characteristics*—index $I_S(k)$ of being a single that equals to unity if cluster k consists of a single event, and zero otherwise, size-corrected average leaf depth d_V (only defined for families), size-corrected average family branching B_N (only defined for families), and the aftershock magnitude gap Δ_A equal to

the difference between the magnitudes of the family main shock and the largest aftershock (only defined for families with aftershocks).

A spatial map of a selected characteristic is produced via the following steps:

(1a) The value of an individual earthquake characteristic is assigned to the event epicentre.

(1b) The value of an individual cluster characteristic is assigned to the epicentre of the family main shock (single is considered to be a main shock).

(1c) The value of a regional characteristic is estimated at the epicentre of each catalogue event with magnitude $m \geq 5$ within a circle with radius $r = 200$ km centred at this event.

(2) To obtain an averaged (or maximal) value of the selected characteristic at point x , the raw estimation from (1) is averaged (or maximized) within circles of radius $r = \min(r_{100}, 100$ km), where r_{100} is the radius of the circle centred at x that contains 100 events with magnitude $m \geq 5$. The points where the circle of radius 100 km contains less than 5 events of magnitude $m \geq 5$ are left transparent. The bandwidth of such adaptive averaging is inversely related to the seismic intensity, which leads to emphasizing detailed changes of the examined earthquake characteristics in high-intensity regions

while applying large-scale smoothing in low-intensity regions. The 100 km limit for the radius of the averaging circle produces coloured bands that are often wider than actual seismicity domains, in particular in the transform and divergent environments. This is done for visual convenience.

(3) An isotropic Gaussian filter is applied to a map from (2). This last step is only applied to smooth the maps. It does not disturb the global patterns of spatial variability.

APPENDIX C: SPEARMAN CORRELATION ANALYSIS

Spearman rank correlation is designed to detect possible dependency between non-Gaussian random samples. Recall that the rank $r(X_k)$ of an observation from a sample $\{X_i\}$, $i = 1, \dots, N$ is defined as the index of this observation in the sample ordered in increasing fashion, from the smallest to the largest observation. For example, the ranks of the observations from the sample $\{-5, 2, 1\}$ are $\{1, 3, 2\}$. In a sample with non-repeating values, the ranks are natural numbers from 1 to N .

Consider a paired random sample $\{X_i, Y_i\}$, $i = 1, \dots, N$. The Spearman correlation coefficient ρ for the sample is defined as the Pearson correlation between the ranked sample values:

$$r(X, Y) = \text{corr}[r(X_i), r(Y_i)].$$

Spearman correlation has several properties that make it convenient for establishing non-linear relations in non-Gaussian data:

- (i) The correlation can detect non-linear relations. Namely, $\rho(X, Y) = 1$ for any monotone increasing deterministic relation $Y = f(X)$ and $\rho(X, Y) = -1$ for any monotone decreasing deterministic relation $Y = f(X)$;
- (ii) The correlation is insensitive to monotone transformations of data. Namely, $\rho(X, Y) = \rho(f(X), g(Y))$ for any monotone increasing functions f and g ;
- (iii) $\rho(X, Y)$ is less sensitive to outliers than the conventional Pearson correlation;
- (iv) For Gaussian data, the Spearman correlation is close to Pearson correlation.

A straightforward way of establishing significance of Spearman correlation is bootstrap: generating multiple independent ranked paired samples and using them to approximate the distribution of ρ under the hypothesis of independence. An alternative, approximate, approach suggests that the quantity

$$t = \rho \sqrt{\frac{N-2}{1-\rho^2}} \quad (\text{C1})$$

under the null hypothesis of independence has the Student distribution with $N - 2$ degrees of freedom (Kendall & Stuart 1973). For large sample sizes this is very close to the standard normal distribution.

The bootstrap and analytical approaches give practically indistinguishable results for our sample sizes. We report the significance levels (see Table 5) according to the analytical approximation of eq. (C1).

APPENDIX D: GENERALIZED LINEAR MODEL ANALYSIS

Generalized Linear Models (GLMs) is an extension of linear regression framework to non-normal responses that uses the likelihood approach for model fitting. We use GLMs to complement the Spearman correlation analysis of Appendix C. Specifically, we see how the average values of the examined earthquake statistics can be predicted by non-linear functions of (i) heat flow H [W m^{-2}], (ii) strain rate tensor style S , and (iii) strain rate tensor second invariant I_2 . For every examined earthquake statistic Y , we fit three GLMs, one for each predictor X from the above list:

$$\mu \equiv E[Y] = \beta_0 + \beta_1 X + \beta_2 X^2 + \beta_3 I_{\{X > \text{mean}(X)\}}. \quad (\text{D1})$$

The model's right hand side combines a quadratic regression in X and the Heaviside function for the deviation of X from its average. The goodness of fit is measured by the coefficient of determination:

$$R^2 = 1 - \text{Var}[\text{model residuals}] / \text{Var}[Y]. \quad (\text{D2})$$

Here $\text{Var}[\cdot]$ denotes the sample variance. The value of R^2 is interpreted as the proportion of variance in Y explained by the model. In particular, $R^2 = 1$ means that the model does a perfect (deterministic) prediction of Y , while $R^2 = 0$ suggests that the right hand side of (D1) has no information about Y .

We consider ten earthquake statistics and three predictors, which results in thirty models. The results are summarized in Table 6 and Fig. D1. The solid lines in Fig. D1 represent model fit; the amplitude of a line's jump corresponds to the importance of the last non-linear term $I_{\{X > \text{mean}(X)\}}$ —the difference in the average of Y for below-than-average and above-than-average values of X . The analysis suggests that the heat flow is the most powerful predictor among the three examined ones: there exists only one statistic for which the heat flow does not give the highest R^2 —the proportion of singles p_s .

We also observe that the quality of forecast is uniformly low—the coefficient of determination never exceeds 0.3. Nevertheless, all models with the heat flow as predictor are highly significant. This confirms the observations in the main part of the paper that heat flow is a significant control for the average value of the examined earthquake statistics.

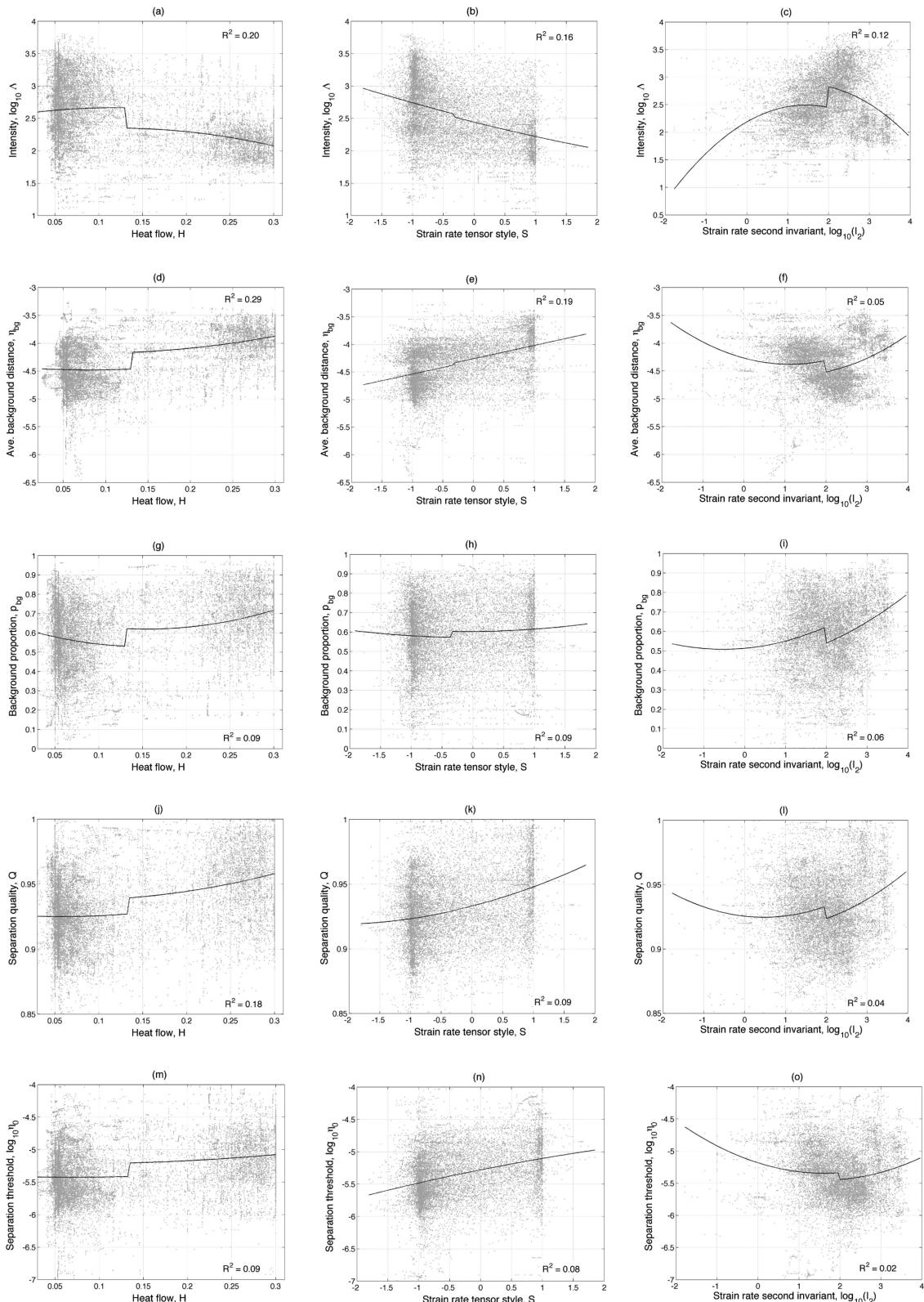
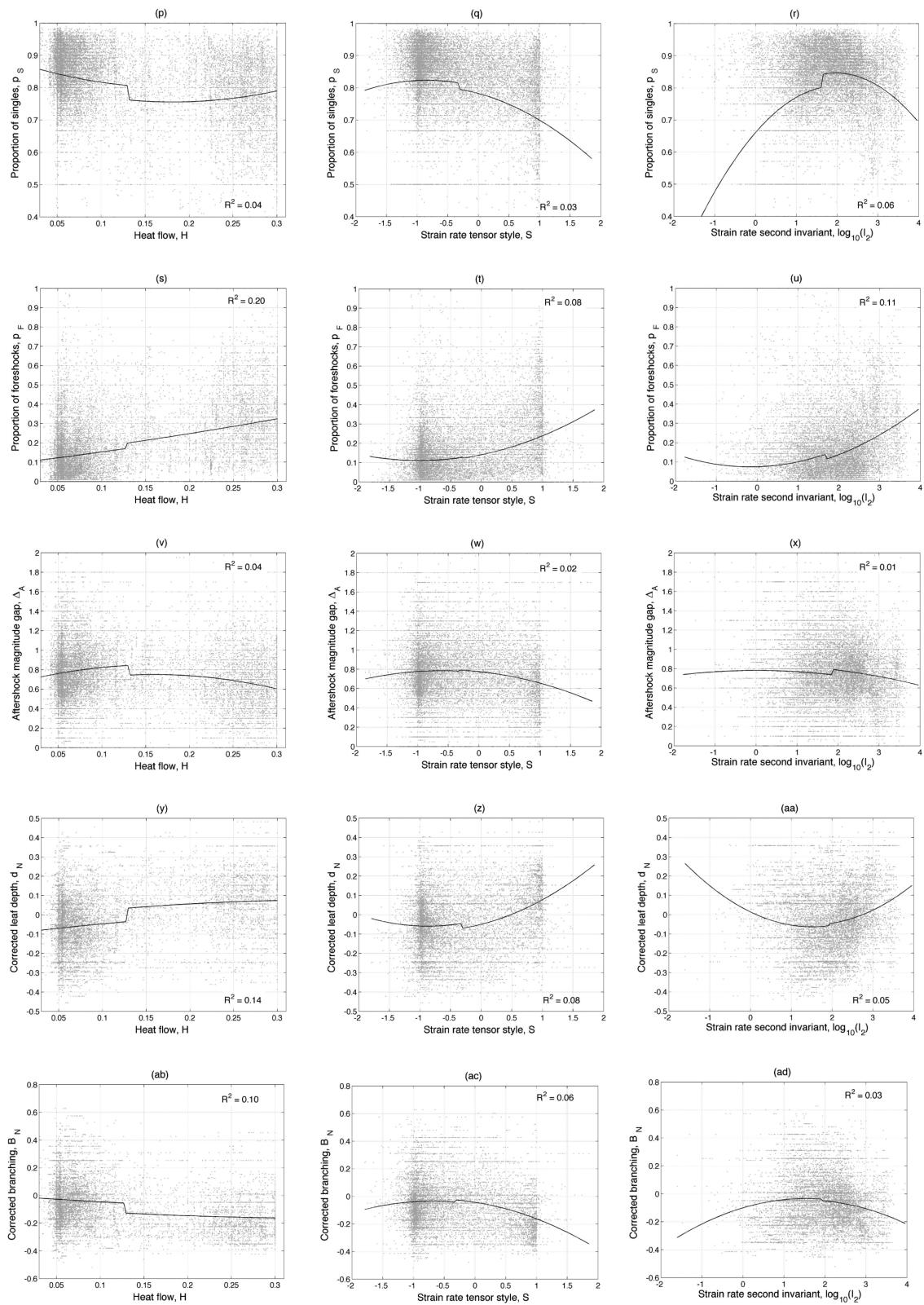


Figure D1. Generalized Linear Model (GLM) analysis of global seismicity. Each panel refers to forecasting a particular characteristic of seismicity using either heat flow H (first column—a, d, g, j, m, p, s, y, ab), strain rate tensor style S (second column—b, e, h, k, n, q, t, w, z, ac), or logarithm of strain rate tensor second invariant $\log_{10}(I_2)$ (third column—c, f, i, l, o, r, u, x, aa, ad). Grey dots—data points; each data point corresponds to a $0.5^\circ \times 0.5^\circ$ Earth surface cell. Black lines—GLM forecast. (a, b, c): Seismic intensity Λ [events/year]; the analysis is done for $\log_{10} \Lambda$; (d, e, f): Average distance in background mode, $\log_{10}(\eta_{bg})$; (g, h, i): Proportion of background events, p_{bg} ; (j, k, l): Quality of separation between cluster and background modes, Q ; (m, n, o): Threshold that separates the cluster and background modes, $\log_{10}(\eta_0)$; (p, q, r): Proportion of singles among clusters, p_S ; (s, t, u): Proportion of foreshocks among aftershocks and foreshocks, p_F ; (v, w, x): Difference between magnitude of the main shock and the largest aftershock, Δ_A ; (y, z, aa): Average leaf depth, corrected for family size, d_N ; (ab, ac, ad): Average family branching, corrected for family size, B_N .

**Figure D1 (Continued).**