Earthquake recurrence as a record breaking process

Jörn Davidsen,[1,2,3] Peter Grassberger,[2,4] and Maya Paczuski[2,5]

Received 23 February 2006; revised 10 April 2006; accepted 27 April 2006; published 6 June 2006.

[1] Extending the central concept of recurrence times for a point process to recurrent events in space-time allows us to characterize seismicity as a record breaking process using only spatiotemporal relations among events. Linking record breaking events with edges between nodes in a graph generates a complex dynamical network isolated from any length, time or magnitude scales set by the observer. For Southern California, the network of recurrences reveals new statistical features of seismicity with robust scaling laws. The rupture length and its scaling with magnitude emerges as a generic measure for distance between recurrent events. Further, the relative separations for subsequent records in space (or time) form a hierarchy with unexpected scaling properties. **Citation:** Davidsen, J., P. Grassberger, and M. Paczuski (2006), Earthquake recurrence as a record breaking process, *Geophys. Res. Lett.*, *33*, L11304, doi:10.1029/2006GL026122.

1. Introduction

[2] Fault systems as the San Andreas fault in California are prime examples of self-organizing systems in nature which are characterized by an internal dynamics that increases the inherent order of the system [*Rundle et al.*, 2002]. Very often they consist of interacting elements, each of which stays quiescent until its internal state variable reaches a trigger threshold leading to a rapid discharge or "firing". Since the internal state variables evolve in time in response to external driving sources and inputs from other elements, the firing of an element may in turn trigger a discharge of other elements. In the context of fault systems, this corresponds to stress discharge in the form of earthquakes, or the deformation and sudden rupture of parts of the earth's crust driven by convective motion in the mantle.

[3] Self-organizing systems very often exhibit dynamics that is strongly correlated in space and time over many scales. In particular, the complex spatiotemporal dynamics of fault systems manifests itself in a number of generic, empirical features of earthquake occurrence including clustering, fault traces and epicenter locations with fractal statistics, as well as scaling laws like the Omori and Gutenberg-Richter laws [*Turcotte*, 1997; *Rundle et al.*, 2003], giving rise to a worldwide debate about their explanation. Resolving this dispute could conceivably require measuring the internal state variables — the stress and strain everywhere within the earth along active faults — and their exact dynamics. This is (currently) impossible. Yet, the associated earthquake patterns are readily observable making a statistical approach based on the concept of spatiotemporal point processes feasible, where the description of each earthquake is reduced to its size or magnitude, its epicenter and its time of occurrence. Describing the patterns of seismicity may shed light on the fundamental physics since these patterns are emergent processes of the underlying many-body nonlinear system.

[4] Recently, such an approach has brought to light new properties of the clustering of seismicity in space and time [*Bak et al.*, 2002; *Corral*, 2003, 2004; *Davidsen and Goltz*, 2004; *Davidsen and Paczuski*, 2005; *Baiesi and Paczuski*, 2005], which can potentially be exploited for earthquake prediction [*Goltz*, 2001; *Tiampo et al.*, 2002; *Baiesi*, 2006]. One aim has been to evaluate distances between subsequent events, including temporal and spatial measures. The observed spatiotemporal clustering of seismicity suggests that subsequent events are to a certain extent causally related. It further suggests that the usual mainshock/aftershock scenario — where each event has at most one correlated predecessor — is too simplistic and that the causal structure of seismicity could extend beyond immediately subsequent events, especially since the determination of the sequence is largely arbitrary depending on the size of the region considered and the completeness of the record of events.

[5] In this work we quantify the spatiotemporal clustering of seismicity in terms of a sparse, directed network, where each earthquake is a node in the graph and links connect events with their recurrences. This general network picture allows us to characterize clustering by using only the spatiotemporal structure of seismicity, without any additional assumptions.

2. Method

[6] The key advance we propose is to generalize the notion of a subsequent event to a record breaking event, one which is closer in space than all previous ones, up to that time. Consider a pair of events, A and B, occurring at times $t_A < t_B$. Earthquake B is a record with respect to A if no intervening earthquake happens in the spatial disc centered on A with radius $\overline{AB}$ during the time interval $[t_A, t_B]$. Each record B can be characterized by the distance $l = \overline{AB}$ and the time interval $T = t_B - t_A$ between the two events. Since the spatial window is centered on A, this definition of a record ensures that any record B is closer in space to A than all other events in the time interval $[t_A, t_B]$ and, thus, the notion of a record is justified. Naturally, A's subsequent event is a

[1] Max-Planck-Institut für Physik Komplexer Systeme, Dresden, Germany.
[2] Also at Perimeter Institute for Theoretical Physics, Waterloo, Ontario, Canada.
[3] Now at British Antarctic Survey, Cambridge, UK.
[4] John-von-Neumann Institute for Computing, Jülich, Germany.
[5] Complexity Science Group, Department of Physics and Astronomy, University of Calgary, Calgary, Alberta, Canada.

0094-8276/06/2006GL026122

record. Yet, it is generally not the only one and, by construction, each later record breaks all previous ones giving rise to a hierarchical cascade of records. We denote this sequence of records for a given event A as its *recurrences*. Note that each earthquake has its own sequence of records or recurrences that follow it in time.

[7] Our definition of recurrent events is based solely on spatiotemporal relations between events and minimizes the influence of the observer by avoiding the use of any space, time, or magnitude scales other than those explicitly associated with the earthquake catalog (i.e., its magnitude, spatial, and temporal ranges). Even the influence of the later scales is rather small since, for example, an increase in the spatial-temporal coverage of the catalog does not generally turn a record-breaking event in a non-record breaking event, thus, conserving the property of a record. Our definition further allows us to discuss spatial and temporal clustering, without introducing any artificial scales, or making any arbitrary assumptions about the form of seismic correlations. Also, as time goes on, one wants to be more strict in declaring B a recurrence of A, or related to A in a meaningful way, which is precisely what our definition achieves.

[8] To construct a network we represent each earthquake as a node, and each recurrence by a link between pairs of nodes, directed according to the time ordering of the earthquakes. Distinct events can have different numbers of in-going and out-going links — called in- and out-degree — which designate their relations to the other events. The out-going links from any node define the structure of recurrences in its neighborhood and characterize the spatiotemporal dynamics of seismicity, or its clustering with respect to that event. The overall structure of the network describes the clustering of seismic activity in the region that is analyzed.

[9] To test the suitability and robustness of our method to characterize seismicity, we study a "relocated" earthquake catalog from Southern California (data available at http://www.data.scec.org/ftp/catalogs/SHLK/) which has improved relative location accuracy within groups of similar events, the relative location errors being less than 100 m [*Shearer et al.*, 2003]. The catalog is assumed to be homogeneous from January 1984 to December 2002 and complete for events larger than magnitude m_c = 2.5 in the rectangle (120.5°W, 115.0°W) × (32.5°N, 36.0°N) giving N = 22217 events. In order to test for robustness and the dependence on magnitude, we analyze this sub-catalog and subsets of it that are obtained by selecting higher threshold magnitudes, namely m = 3.0, 3.5, 4.0 giving N = 5857, 1770, 577 events, or a shorter period from January 1984 to December 1987 giving N = 4744 events for $m = m_c$.

3. Results and Discussion

[10] Figure 1 shows the probability density function $P_m(l)$ of distances, l, of recurrent events for different thresholds m. The typical or characteristic distance, $l^*(m)$, where the distribution peaks, increases with magnitude. For sufficiently large l, all distributions show a power law decay with an exponent ≈1.05 up to a cutoff. This cutoff is induced by the finite size of the region that we consider.

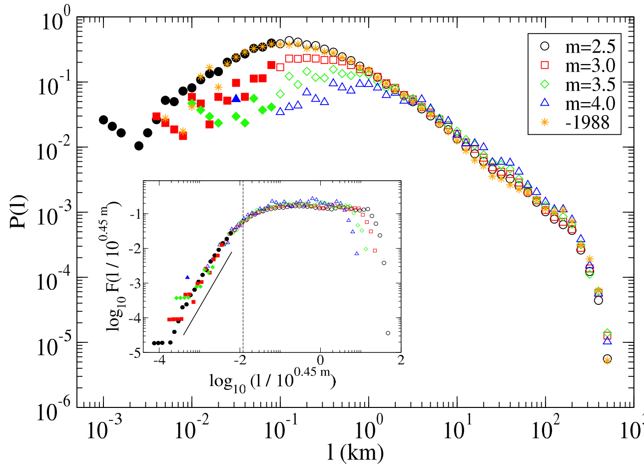

Figure 1. Distribution of distances l of recurrent events for sets with different magnitude thresholds m. The distribution for m = 2.5 up to 1988 is also shown. Filled symbols correspond to distances below 100 m and are unreliable due to location errors. The inset shows a data collapse, obtained by rescaling distances and distributions according to equation (1). The full straight line has slope 2.05; the vertical dashed line indicates the pre-factor L_0 in the scaling law for the characteristic distance, $l^* = L_0 \times 10^{0.45m}$.

[11] With a suitable scaling ansatz, the different curves in Figure 1 fall onto a universal curve, except at the cutoff, which is a man-made scale imposed on the geological system. The inset in Figure 1 shows results of a data collapse using

$$P_m(l) \sim l^{-1.05} F\left(l/10^{0.45m}\right). \qquad (1)$$

The scaling function F has two regimes, a power-law increase with exponent ≈2.05 for small arguments and a constant regime at large arguments. The transition point between the two regimes can be estimated by extrapolating them and selecting the intersection point, giving L_0 = 0.012 km. For the characteristic distance that appears in F we thus find $l^* \approx L_0 \times 10^{0.45m}$. This is close to the estimated behavior of the rupture length $L_R \approx 0.02 \times 10^{m\prime/2}$ km given by [*Kagan*, 2002] and remarkably close to $L_R = \sqrt{A_R} \approx 0.018 \times 10^{0.46m\prime}$ km given by [*Wells and Coppersmith*, 1994], where m' is the magnitude of the event and A_R its rupture area.

[12] The agreement between our result and that of [*Wells and Coppersmith*, 1994] suggests that the characteristic length scale of distances of recurrent events is the rupture length, defined in terms of the rupture area $l^* = L_R \equiv \sqrt{A_R}$. This is substantially supported by the remarkable fact that, for fixed m, $P_m(l)$ and thus l^* does not significantly vary with the length of the observation period despite huge differences in the number of earthquakes N — which is very different from a random process (J. Davidsen et al., Networks of recurrent events: Clustering and the underlying causal structure, submitted to *Phys. Rev. E*, 2006). As Figure 1 shows, $P_{2.5}(l)$ is largely unaltered if only the sub-catalog up to 1988 is analyzed. However, we find that this is not true for sub-catalogs of similar size generated by randomly deleting events. This is evidence

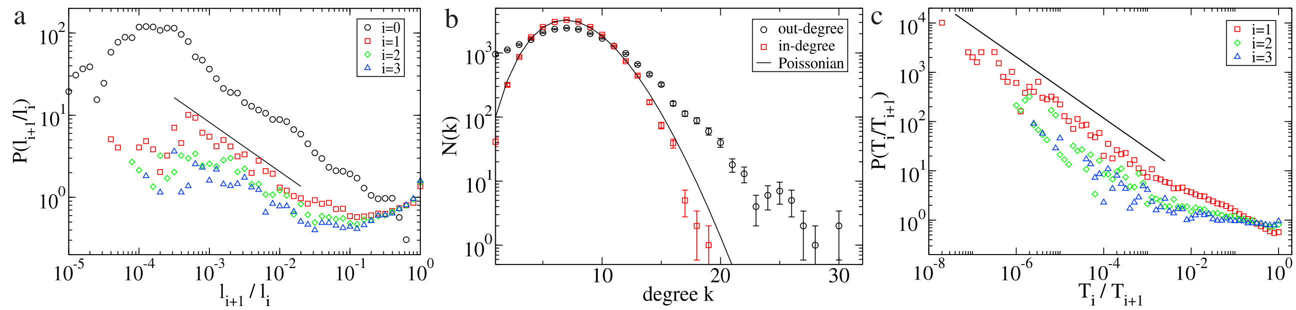

Figure 2. (a) Distribution of recurrence distance ratios l_{i+1}/l_i for threshold magnitude m = 2.5. The straight line corresponds to a decay with exponent 0.6. (b) Distributions of in- and out-degrees of the network for m = 2.5. (c) Distribution of recurrence time ratios T_i/T_{i+1} for m = 2.5. The straight line has slope −0.62.

for a dynamical origin of the shape of $P_m(l)$. The comparison of the two different observation periods in Figure 1 further shows that l^* does not depend strongly on the total number of recurrences (or links) or on the average degree of the network, $\langle k \rangle$ = #links/#nodes ($\langle k \rangle$ = 6.56 (7.40) for events up to 1988 (2002) and m = 2.5), but clearly on m. The independence of the time span and consequent number of events implies that equation (1) is a robust, empirical result for seismicity.

[13] The identification $l^* = L_R$ is also consistent with the fact that the description of earthquakes as a point process breaks down at the rupture length. Below that scale, the relevant distance(s) between earthquakes is not given solely by their epicenters but also by the relative location and orientation of the spatially extended ruptures. Due to different orientations we expect randomness or lack of correlations between epicenters for distances below the rupture length. If events are happening randomly in space, or are recorded as happening randomly in space due to location errors, then $P_m(l)$ rises linearly. To see this consider a two dimensional disc of radius R, with one point at the center and N_R randomly distributed points. The probability that there will be no (other) point within a distance l of the center point is $(1 - l^2/R^2)^{N_R}$; therefore, the probability density for the closest point to be at distance l is $(2N_R l/R^2)(1 - l^2/R^2)^{N_R - 1}$. At small l, this will describe the distribution shown in Figure 1 and determine the scaling function F in equation (1). In fact, this is precisely what the earthquake data show for distances smaller than the rupture length (see the straight line with a slope of 2.05 in the inset of Figure 1 and the linear increase with slope 1 in the main part of Figure 1).

[14] The lengths l^* observed for the values of m we consider are larger than the length (≈100 m) at which we observe random behavior due to location errors. In fact, the data do not show any anomaly near 100 m. Moreover, $P_4(l)$ (blue triangles) does not change substantially if the epicenters in the catalog are randomly relocated by a small distance up to one kilometer. Yet, the maximum for $P_{2.5}(l)$ shifts to larger l with this procedure, destroying the scaling of $l^*(m)$. Since the smallest l^* that obeys the data collapse is ≈160 m, the data collapse we observe for the original data verifies that the relative location errors are indeed less than 100 m, or of that order. Furthermore, our observations indicate that spatial correlations between epicenters are already lost for distances 100 m < l < l^* (J. Davidsen et al., Networks of recurrent events: Clustering and the underlying causal structure, submitted to *Phys. Rev. E*, 2006). (Note that a systematic dependence of the location error on magnitude has not been reported in the literature and is also not present in the catalog at hand. It is unlikely that the characteristic length we see (l^*) is merely an artifact due to location error growing with magnitude.) Related to the distribution of distances of recurrent events is the distribution of distance ratios l_{i+1}/l_i in the cascade of recurrences to a given event. Here recurrences are ordered by time; recurrence i + 1 comes after i. We take l_0 = 448.5 km, which is the size of the region covered by the catalog (Figure 2a). By construction these ratios are always ≤ 1. We denote by $P_i(x)$ the probability density function that $l_{i+1}/l_i = x$ for each event that has an $(i + 1)^{th}$ recurrence. The data for i = 0 (black circles) scale over a wide region as $P_0(x) \sim x^{-\delta_r}$ with $\delta_r \approx 0.6$ – as already shown in [*Davidsen and Paczuski*, 2005]. This is indicated in Figure 2a by the straight line. Although each distribution $P_i(x)$ is different, the curves for $i \geq 1$ also show (more restricted) power law decay comparable to P_0. For $l_{i+1}/l_i \rightarrow 1$ they also show a peak, which becomes more pronounced with increasing i. This is due to recurrences occurring at almost the same distance. The observed exponent δ_r for the power law decay has a dynamical origin and is *not* determined by the spatial distribution of seismicity [*Davidsen and Paczuski*, 2005]: Purely based on the correlation dimension D_2, one would expect $P_1(x) \sim x^{D_2 - 1}$. For Southern California, this gives a growing dependence $P_1(x) \sim x^{0.2}$ rather than a decaying behavior. Thus, the exponent δ_r reflects the complex *spatiotemporal* organization of seismicity.

[15] A similar analysis can be made for the distribution of recurrence times. Figure 3 shows the probability density function $P_m(T)$ for different threshold magnitudes m. They all decay roughly as $1/T^{\alpha}$ with $\alpha \approx 0.9$ for intermediate times. The apparent scaling region in Figure 3 shows some curvature, though. Surprisingly, $P_m(T)$ is independent of m and the number of events in the considered catalog. This is very different from earlier results for waiting time distributions between subsequent earthquakes [*Bak et al.*, 2002; *Corral*, 2003] and reflects a new non-trivial feature of the spatiotemporal dynamics of seismicity that appears when events other than the immediately subsequent ones are considered.

[16] The relative times between subsequent recurrences in the hierarchy can be analyzed in the same way as distances were above. Figure 2 shows the probability density functions $P_i(T_i/T_{i+1})$ for the ratios of the times T_i/T_{i+1} for subsequent recurrences to a given event. Note

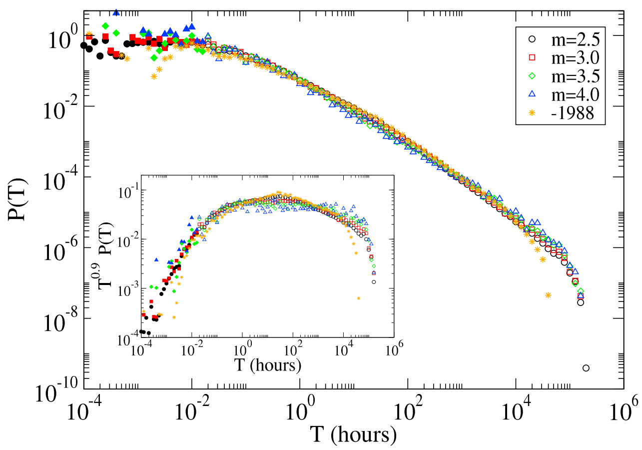

Figure 3. Distributions of recurrence times. The distribution for m = 2.5 up to 1988 is also shown. Filled symbols correspond to times below 90 seconds which are underestimated and unreliable due to measurement restrictions. The inset shows the rescaled distributions.

that these ratios are again $\leq$ 1, yet the order is inverted compared to the distance ratios. The broadest scaling regime materializes for T_1/T_2, again with exponent $\delta_T \approx \delta_r \approx 0.6$. The distributions for larger i follow roughly the same behavior for ratios $\ll$1, but deviate (less strongly than for the spatial data in Figure 2a) when the ratios tend to 1. Again it is obvious that this behavior cannot be explained by random events.

[17] The description of seismicity as a network of earthquake recurrences allows its characterization by means of the usual characteristics that are thought to be important for unweighted complex networks [*Albert and Barabási*, 2002]. One such network property is its degree distribution. Figure 2b shows the degree distributions for m = 2.5, which is compared to a Poisson distribution with the same mean degree $\langle k \rangle$ = 7.40 (solid line). A Poisson degree distribution would be expected in the limit of infinitely large networks if earthquakes epicenters were placed randomly in space and time (J. Davidsen et al., Networks of recurrent events: Clustering and the underlying causal structure, submitted to *Phys. Rev. E*, 2006). While the in-degree distribution roughly agrees with such a random network, the out-degree distribution shows significant deviations. In particular, the network keeps a preponderance of nodes with small out-degree as well as an excess of nodes with large out-degree compared to a Poisson distribution. This effect is independent of magnitude, as an analysis of subsets with higher magnitude threshold shows. Note, however, that $\langle k \rangle$ decreases with m, simply because the catalog size shrinks with m. In particular, we find $\langle k \rangle$ = 6.24, 5.20, 4.35 for m = 3.0, 3.5, 4.0, respectively. The non-trivial behavior of the out-degree distribution implies in particular that the network topology and, thus, the hierarchial cascade of recurrences or records captures important information about the spatiotemporal clustering of seismicity.

[18] It is important to note that our results are robust with respect to modifications of the rules used to construct the network, for example, using spatial neighborhoods such that the construction becomes symmetric under time reversal or taking into account magnitudes. All such modifications have the drawback that they do not define a record breaking process consisting of recurrences to each event. Our results are also unaltered if we exclude links with propagation velocities larger than 6 km/sec (≈0.1% of all links).

4. Conclusions

[19] Our analysis shows that the description of seismicity by means of recurrences in space-time allows us to characterize its clustering behavior using only spatiotemporal relations between events and to identify new, robust scaling laws in the pattern of seismic activity. The pairs of recurrent events form a complex network with non-trivial statistics. The method allows us to detect the rupture length and its scaling with magnitude directly from earthquake catalogs without making any assumptions. Our results for the distributions of relative separations for the next recurrence in space and time should also have implications for seismic hazard assessment. Finally, our findings provide detailed benchmark tests for models of seismicity.

References

Albert, R., and A.-L. Barabási (2002), Statistical mechanics of complex networks, *Rev. Mod. Phys.*, *74*, 47.
Bak, P., K. Christensen, L. Danon, and T. Scanlon (2002), Unified scaling law for earthquakes, *Phys. Rev. Lett.*, *88*, 178501.
Baiesi, M. (2006), Scaling and precursor motifs in earthquake networks, *Physica A*, *360*, 534.
Baiesi, M., and M. Paczuski (2005), Complex networks of earthquakes and aftershocks, *Nonlinear Processes Geophys.*, *12*, 1.
Corral, A. (2003), Local distribution and rate fluctuations in a unified scaling law for earthquakes, *Phys. Rev. E*, *68*, 035102.
Corral, A. (2004), Long-term clustering, scaling, and universality in the temporal occurrence of earthquakes, *Phys. Rev. Lett.*, *92*, 108501.
Davidsen, J., and C. Goltz (2004), Are seismic waiting time distributions universal?, *Geophys. Res. Lett.*, *31*, L21612, doi:10.1029/2004GL020892.
Davidsen, J., and M. Paczuski (2005), Analysis of the spatial distribution between successive earthquakes, *Phys. Rev. Lett.*, *94*, 048501.
Goltz, C. (2001), Decomposing spatio-temporal seismicity patterns, *Nat. Hazards Earth Syst. Sci.*, *1*, 83.
Kagan, Y. Y. (2002), Aftershock zone scaling, *Bull. Seismol. Soc. Am.*, *92*, 641.
Rundle, J. B., K. F. Tiampo, W. Klein, and J. S. S. Martins (2002), Self-organization in leaky threshold systems: The influence of near-mean field dynamics and its implications for earthquakes, neurobiology, and forecasting, *Proc. Natl. Acad. Sci. U. S. A.*, *99*, 2514.
Rundle, J. B., D. L. Turcotte, R. Shcherbakov, W. Klein, and C. Sammis (2003), Statistical physics approach to understanding the multiscale dynamics of earthquake fault systems, *Rev. Geophys.*, *41*(4), 1019, doi:10.1029/2003RG000135.
Shearer, P., E. Hauksson, G. Lin, and D. Kilb (2003), Comprehensive waveform cross-correlation of southern California seismograms: 2. Event locations obtained using cluster analysis, *Eos Trans. AGU*, *84*, 46.
Tiampo, K. F., J. B. Rundle, S. McGinnis, S. J. Gross, and W. Klein (2002), Mean-field threshold systems and phase dynamics: An application to earthquake fault systems, *Europhys. Lett.*, *60*, 481.
Turcotte, D. L. (1997), *Fractals and Chaos in Geology and Geophysics*, 2nd ed., Cambridge Univ. Press, New York.
Wells, D. L., and K. J. Coppersmith (1994), New empirical relationships between magnitude, rupture length, rupture width, rupture area, and surface displacement, *Bull. Seismol. Soc. Am.*, *84*, 974.

J. Davidsen, British Antarctic Survey, High Cross, Madingley Road, Cambridge CB3 0ET, UK. (j.davidsen@bas.ac.uk)
P. Grassberger, John-von-Neumann Institute for Computing, Forschungszentrum Jülich, D-52425 Jülich, Germany.
M. Paczuski, University of Calgary, 2500 University Drive NW, Calgary, Alberta, Canada T2N 1N4.