

Title: The Predictable Chaos of Slow Earthquakes

Authors: A. Gualandi¹, J.-P. Avouac², S. Michel³ and D. Faranda^{4,5}

Affiliations

¹ National Aeronautics and Space Administration, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA, 4800 Oak Grove Dr, Pasadena, CA 91109

² California Institute of Technology, Pasadena, CA, USA

³ Laboratoire de Géologie, Ecole Normale Supérieure, Paris, France

⁴ LSCE-IPSL, CEA Saclay l'Orme des Merisiers, CNRS UMR 8212 CEA-CNRS-UVSQ, Université, Paris-Saclay, 91191 Gif-sur-Yvette, France

⁵ London Mathematical Laboratory, London, UK

* Correspondence to: adriano.geolandi@gmail.com.

Abstract

Regular and slow earthquakes result from unstable frictional slip and are described by a potentially chaotic dynamics¹⁻⁴. Assessing the predictability of regular earthquakes is a challenge because observations generally cover a short time compared to the long time needed to generate successive ruptures of a same fault. This limitation does not apply to slow earthquakes, which have much shorter recurrence times⁵. We therefore assess the predictability of slow slip events (SSEs) on the Cascadia subduction imaged using geodetic times series between 2007 and 2017⁶. These SSEs

show scaling laws and a systematic segmentation similar to regular earthquakes⁷. Using the embedding theory^{8,9} and extreme value theory¹⁰, we demonstrate a low-dimensional (<5) chaotic behavior. We calculate properties of the underlying strange attractor and find that the system has a predictability horizon of the order of 2 to 55 days. While longer-term deterministic prediction seems intrinsically impossible, short-term prediction of SSEs is shown to be possible, for example using the instantaneous dimension of the strange attractor which systematically peaks to high values at the onset of large slow slip events. Regular earthquakes might similarly be predictable but with a short predictable horizon which we infer to be of the order of their durations.

Main Text

Laboratory experiments have shown that it is possible to forecast the time to failure of both slow and fast ruptures¹¹⁻¹³, raising hopes that natural earthquakes might similarly be predictable. However, a fully deterministic failure prediction for natural earthquakes is still yet to come¹⁴: we currently do not know if natural earthquakes result from deterministic chaos, and, if so, what would be their predictability horizon. Detecting chaotic behavior in geophysical time series is a difficult task, mainly because of the short and noisy data generally available¹⁵. While historical observations have been used to argue for chaos¹⁶, characterizing the chaotic behavior of regular earthquakes has been a challenge because of the short history of documented earthquakes on a same fault segment. A possible chaotic behavior has been inferred in the case of foreshocks¹⁷ and Low Frequency Earthquakes (LFEs)¹⁸, a class of small and slow earthquakes detectable with seismology. Here, we focus on Slow Slip Events (SSEs), which are mostly aseismic recurring events with recurrence time much shorter than that of typical earthquakes (months/years instead of decades/centuries). They have been documented with geodetic techniques in various tectonic

contexts, in particular along major subduction megathrusts⁵, where they can reach moment magnitudes comparable to that of large earthquakes ($M_w > 7$). SSEs are strikingly similar to regular earthquakes: they evolve into large pulse-like ruptures¹⁹ and follow similar scaling laws, exhibiting systematic segmentation^{7,20}. These characteristics and the possibility to image them using geodetic time series make them a most suitable system to study the dynamics of frictional sliding at the natural scale. We focus on the Cascadia subduction, where multiple SSEs ruptured repeatedly 13 major segments between 2007 and 2017⁷ (Fig. 1). Our goal is to characterize if the slip time series reveal a deterministic dynamics or are the result of a stochastic process.

We calculate the slip potency rate time series for all the 13 segments ($\dot{P}_s(t)$, $s = 1, \dots, 13$, Fig. 2, Fig. S1, and Section Slip potency rate in the Methods) as the 1-day interval time derivative of the low-pass filtered slip potency. This is the scalar observable that we use to assess the dynamics of the system. We choose this variable because sliding velocity is considered to be the main observable variable governing the evolution of friction in laboratory experiments²¹. The slip potency rate shows episodic bursts with the most extreme events occurring quite regularly (Fig. 2 and S1). We refer to the space spanned by the subset of variables needed to explain this system as the phase space, and to the trajectories in this phase space as the attractor of the system, being aware that we are dealing with only a region of the whole phase space, representative of the slow slipping phenomenon.

A dynamical system is commonly characterized by its dimension, a geometrical property of the underlying attractor which quantifies the number of degrees of freedom (dof) needed to explain the observed dynamics²². For the stick-slip behavior considered here, the EVT (Section Extreme Value Theory in the Methods) indicates a low average attractor dimension $3.1 < D < 3.5$ (green dots in Fig. 3). Similar low values (correlation dimension $\nu \lesssim 5$) are obtained when using ET

(Section Embedding Theory in the Methods). However, the Signal-to-Noise Ratio (SNR) is high enough only for segments located in the northern section of the megathrust (segments #1 and #2). This is evident when surrogate data²³ are used to test the significance of the observed low dimensionality of the system (Fig. 3, Section Surrogate Data in the Methods and Fig. S2 in the Supplementary Material).

According to the EVT, the dynamics revealed by the filtered data requires at least 4 variables (because D is between 3 and 4). Interacting fault segments are commonly represented using a spring-slider system. The variables of a spring-slider system representing stick-slip friction in the laboratory, which would be similar to a system representing a single segment, must involve slip, slip rate and at least one additional ‘state’ variable to allow fault healing. Note that, even in the case of such a controlled system, the underlying physics and the number of state variables needed to explain laboratory friction remain matters of debate²⁴.

In absence of coupling with other spring-slider systems, the expected number of dof is thus two plus the number of state variables, and the dimension should be between dof-1 and dof. For a non-inertial single spring-slider, adding a second state variable is a viable way to enact chaotic behavior, and a Kaplan-Yorke dimension of 2.119 ± 0.001 has been derived². SSEs are generally thought to result from coupling between fault slip and fluid flow^{25,26}. A single fault segment obeying rate- and state-friction with slip coupled to fluid flow can be represented by a spring-slider system with 5 dofs, i.e. its dynamics can be described by 5 variables: loading shear stress, slip rate, one state variable, porosity, and pore pressure²⁵. To describe SSEs we can assume a quasi-static behavior, and the number of variables then drops to 4, a value consistent with the number of dof recovered from our analysis.

In a system formed of interacting fault segments, the loading term would depend on the variables associated with neighboring fault segments, which would then need to be introduced in the state vector describing the system's evolution. In the case of a 1-dimensional along-strike segmentation, as observed in Cascadia, we would expect a dimension as high as 12, with the extra 8 dof carried in by the two adjacent segments. Interactions with neighboring segments is probably critical at the onset and arrest of SSEs. The dimension D , being an average in the phase space, does not capture the information about transient instabilities, though. Recent advances in EVT applied to dynamical systems theory have proven that it is possible to characterize these transient instabilities via two instantaneous properties²⁷: the instantaneous dimension (d) and the instantaneous extremal index (θ). These quantities refer to the state of the system in a given location of the phase space. Their distribution and temporal evolution is shown in Fig. S3 of the Supplementary Material. The instantaneous dimension, d , indicates the number of variables needed to explain the dynamics of the system in a specific phase space location. We expect to find high values of d when a metastable state is approached. This is apparent from Fig. 4a, where a Poincaré section of the attractor is plotted with d color coded. The maximum retrieved instantaneous dimension is in the range [11,18] for the segments with high SNR. Considering the tentative segmentation here adopted and the low SNR, we consider this result in good agreement with the maximum expected dimension of 12 as in a 1-dimensional spring-and-slider chain. Note that the number of dof, and hence the dimension of the strange attractor, would be expected to be larger in case of interactions between non-adjacent segments, or in case of along-dip segmentation.

From the instantaneous extremal index (θ) (Fig. 4b) we can estimate a range of values for the metric entropy²⁸, which characterizes how fast on average close-by trajectories in the phase space diverge, and we can thus estimate the predictability time t^* of the system (see Extreme Value

Theory in the Methods). We find t^* ranging from 2.2 to 55 days for the 13 segments (Table S1). We exclude the possibility that the apparently chaotic dynamics is a result of the filtering (Section Surrogate Data in the Methods). We conclude that SSEs on the northernmost segments of the Cascadia subduction zone result from deterministic chaos. Our analysis implies that it should be possible to forecast the onset of large SSE ahead of time. This could in principle be achieved based on an explicit deterministic representation of the system dynamics or using some machine learning algorithm that would implicitly simulate the system dynamics. It is notable that the increase of d seems to constitute a reliable precursor of the large SSEs (segments #1 and #2, Fig. 2 and Fig. S1 in the Supplementary Material). The causal filter adopted here introduces a group delay larger than the predictability horizon time, meaning that this approach cannot be used for real-time forecasting. Operational forecasting based on this approach would require an alternative noise reduction, or more accurate data in order to avoid the filtering step or reduce the window size of the filter.

In conclusion, SSEs can be described as a deterministic, albeit chaotic, system rather than as a stochastic process. The deduced predictability horizon is of the order of days/weeks, equivalent to a fraction of the typical duration of large SSEs in our dataset. If the dynamics derived from the filtered time series is representative of the true underlying dynamics, long-term prediction of SSEs (i.e., over a time horizon much longer than their duration) seems intrinsically impossible. This implies that for long-term predictions a stochastic approach remains the best tool at our disposal, while short-term predictions should be feasible in a deterministic sense. As SSEs might be regarded as earthquakes in slow motion⁹, regular earthquakes might be similarly chaotic and predictable. Assuming a similar system dynamics, we infer a predictable horizon of the order of the typical duration of large earthquakes, i.e. tens to hundreds of seconds.

136 **References and Notes**

- (1) J. Huang and D. L. Turcotte, "Are earthquakes an example of deterministic chaos?," *Geophys. Res. Lett.*, vol. 17, no. 3, pp. 223-226, 1990.
- (2) T. W. Becker, "Deterministic Chaos in two State-variable Friction Sliders and the Effect of Elastic Interactions," in *GeoComplexity and the Physics of Earthquakes*, Washington, D.C., American Geophysical Union, 2000, pp. 5-26.
- (3) M. Anghel, Y. Ben-Zion and R. Rico-Martinez, "Dynamical System Analysis and Forecasting of Deformation Produced by an Earthquake Fault," *Pure Appl. Geophys.*, vol. 161, pp. 2023-2051, 2004.
- (4) N. Kato, "Earthquake Cycles in a Model of Interacting Fault Patches: Complex Behavior at Transition from Seismic to Aseismic Slip," *Bulletin of the Seismological Society of America*, vol. 106, no. 4, pp. 1772-1787, 2016.
- (5) S. Schwartz and J. Rokosky, "Slow slip events and seismic tremor at circum-Pacific subduction zones," *Reviews of Geophysics*, vol. 45, no. 3, 2007.
- (6) S. Michel, A. Gualandi and J.-P. Avouac, "Interseismic Coupling and Slow Slip Events on the Cascadia Megathrust," *Pure Appl. Geophys.*, 2018.
- (7) S. Michel, A. Gualandi and J.-P. Avouac, "Similar scaling laws for earthquakes and Cascadia Slow Slip Events," *Nature*, accepted for publication.

- (8) F. Takens, "Detecting Strange Attractors in Turbulence," in *Lecture Notes in Mathematics*, 2nd ed., Springer, 1981, p. 366.
- (9) T. Sauer, J. A. Yorke and M. Casdagli, "Embedology," *Journal of Statistical Physics*, vol. 65, no. 3-4, pp. 579-616, 1991.
- (10) V. Lucarini, D. Faranda, A. C. M. Freitas, J. M. Freitas, M. Holland, T. Kuna, M. Nicol, M. Todd and S. Vaienti, *Extremes and Recurrence in Dynamical Systems*, Hoboken, New Jersey: John Wiley & Sons, 2016.
- (11) B. Rouet-Leduc, C. Hulbert, L. Lubbers, K. Barros, C. Humphreys and P. A. Johnson, "Machine Learning Predicts Laboratory Earthquakes," *Geophysical Research Letters*, vol. 44, pp. 9276-9282, 2017.
- (12) C. Hulbert, B. Rouet-Leduc, P. A. Johnson, C. X. Ren, J. Rivière, D. C. Bolton and C. Marone, "Similarity of fast and slow earthquakes illuminated by machine learning," *Nature Geoscience*, vol. 12, pp. 69-74, 2019.
- (13) F. Corbi, L. Sandri, J. Bedford, F. Funiciello, S. Brizzi, M. Rosenau and S. Lallemand, "Machine Learning Can Predict the Timing and Size of Analog Earthquakes," *Geophysical Research Letters*, vol. 46, no. 3, pp. 1303-1311, 2019.
- (14) L. R. Sykes, B. E. Shaw and C. H. Scholz, "Rethinking Earthquake Prediction," *Pure and Applied Geophysics*, vol. 155, no. 2-4, pp. 207-232, 1999.
- (15) W. Marzocchi, F. Mulargia and G. Gonzato, "Detecting low-dimensional chaos in geophysical time series," *JGR – Solid Earth*, vol. 102, no. B2, pp. 3195-3209.

- (16) J. Huang and D. L. Turcotte, "Evidence for chaotic fault interactions in the seismicity of the San Andreas fault and Nankai trough," *Nature*, vol. 348, pp. 234-236, 1990.
- (17) A. De Santis, G. Cianchini, E. Qamili and A. Frepoli, "The 2009 L'Aquila (Central Italy) seismic sequence as a chaotic process," *Tectonophysics*, vol. 496, no. 1-4, pp. 44-52, 2010.
- (18) D. R. Shelly, "Periodic, Chaotic, and Doubled Earthquake Recurrence Intervals on the Deep San Andreas Fault," *Science*, vol. 328, no. 5984, pp. 1385-1388, 2010.
- (19) N. M. Bartlow, S. Miyazaki, A. M. Bradley and P. Segall, "Space-time correlation of slip and tremor during the 2009 Cascadia slow slip event," *Geophysical Research Letters*, vol. 38, no. 18, 2011.
- (20) A. G. Wech, K. C. Creager, H. Houston and J. E. Vidale, "An earthquake-like magnitude-frequency distribution of slow slip in northern Cascadia," *Geophysical Research Letters*, vol. 37, no. 22, 2010.
- (21) J. H. Dieterich, "Modeling of rock friction: 1. Experimental results and constitutive equations," *Journal of Geophysical Research*, vol. 84, no. B5, pp. 2161-2168, 1979.
- (22) J. Theiler, "Estimating the Fractal Dimension of Chaotic Time Series," *The Lincoln Laboratory Journal*, vol. 3, no. 1, pp. 63-86, 1990.

- (23) J. Theiler, S. Eubank, A. Longtin, B. Galdrikian and J. D. Farmer, "Testing for nonlinearity in time series: the method of surrogate data," *Physica D: Nonlinear Phenomena*, vol. 58, no. 1-4, pp. 77-94, 1992.
- (24) J.-C. Gu, J. R. Rice, A. L. Ruina and S. T. Tse, "Slip motion and stability of a single degree of freedom elastic system with rate and state dependent friction," *Journal of the Mechanics and Physics of Solids*, vol. 32, no. 3, pp. 167-196, 1984.
- (25) P. Segall and J. Rice, "Dilatancy, compaction, and slip instability of a fluid-infiltrated fault," *J. Geophys. Res. Solid Earth*, vol. 100, no. B11, pp. 22155-22171, 1995.
- (26) P. Segall, A. Rubin, A. Bradley and J. Rice, "Dilatant strengthening as a mechanism for slow slip events," *J. Geophys. Res. Solid Earth*, vol. 115, no. B12, 2010.
- (27) D. Faranda, G. Messori and P. Yiou, "Dynamical proxies of North Atlantic predictability and extremes," *Scientific Reports*, vol. 7, no. 41278, 2017.
- (28) D. Faranda and S. Vaienti, "Correlation dimension and phase space contraction via extreme value theory," *Chaos*, vol. 28, no. 041103, 2018.
- (29) E. Lorenz, "Dimension of weather and climate attractors," *Nature*, vol. 353, pp. 241-244, 1991.
- (30) D. Broomhead, J. Huke and M. Muldoon, "Linear Filters and Non-Linear Systems," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 54, no. 2, pp. 373-382, 1992.

- (31) P. Rapp, A. Albano, T. Schmah and L. Farwell, "Filtered noise can mimic low-dimensional chaotic attractors," *Physical Review E*, vol. 47, no. 4, pp. 2289-2297, 1993.
- (32) P. Grassberger and I. Procaccia, "Characterization of Strange Attractors," *Physical Review Letters*, vol. 50, no. 5, pp. 346-349, 1983.
- (33) J. Theiler, "Spurious dimension from correlation algorithms applied to limited time-series data," *Phys. Rev. A*, vol. 34, no. 3, pp. 2427-2432, 1986.
- (34) M. Thompson, "A Comparison of Methods for Computing Autocorrelation Time, Technical Report, no. 1007," University of Toronto Statistics Department, Toronto, 2010.
- (35) L. Cao, "Practical method for determining the minimum embedding dimension of a scalar time series," *Physica D: Nonlinear Phenomena*, vol. 110, no. 1-2, pp. 43-50, 1997.
- (36) V. Lucarini, D. Faranda and J. Wouters, "Universal behaviour of extreme value statistics for selected observables of dynamical systems," *J. Stat. Phys.*, vol. 147, pp. 63-73, 2012.
- (37) J. Pickands III, "Statistical inference using extreme order statistics," *Ann. Stat.*, pp. 119-131, 1975.
- (38) A. Freitas, J. Freitas and M. Todd, "Hitting time statistics and extreme value theory," *Probab. Theory Rel.*, vol. 147, pp. 675-710, 2010.
- (39) R. L. Smith and I. Weissman, "Estimating the Extremal Index," *J. R. Statist. Soc. B*, vol. 56, no. 3, pp. 515-528, 1994.

- (40) M. Süveges, "Likelihood estimation of the extremal index," *Extremes*, vol. 10, no. 1-2, pp. 41-55, 2007.
- (41) D. Prichard, "The correlation dimension of differenced data," *Physics Letters A*, vol. 191, no. 3-4, pp. 245-250, 1994.
- (42) D. Prichard and J. Theiler, "Generating Surrogate Data for Time Series with Several Simultaneously Measured Variables," *Physical Review Letters*, vol. 73, no. 7, 1994.
- (43) A. Wolf, J. Swift, S. H. and J. Vastano, "Determining Lyapunov Exponents from a Time Series," *Physica D: Nonlinear Phenomena*, vol. 16, no. 3, pp. 285-317, 1985.

137

138 **Acknowledgments:**

139

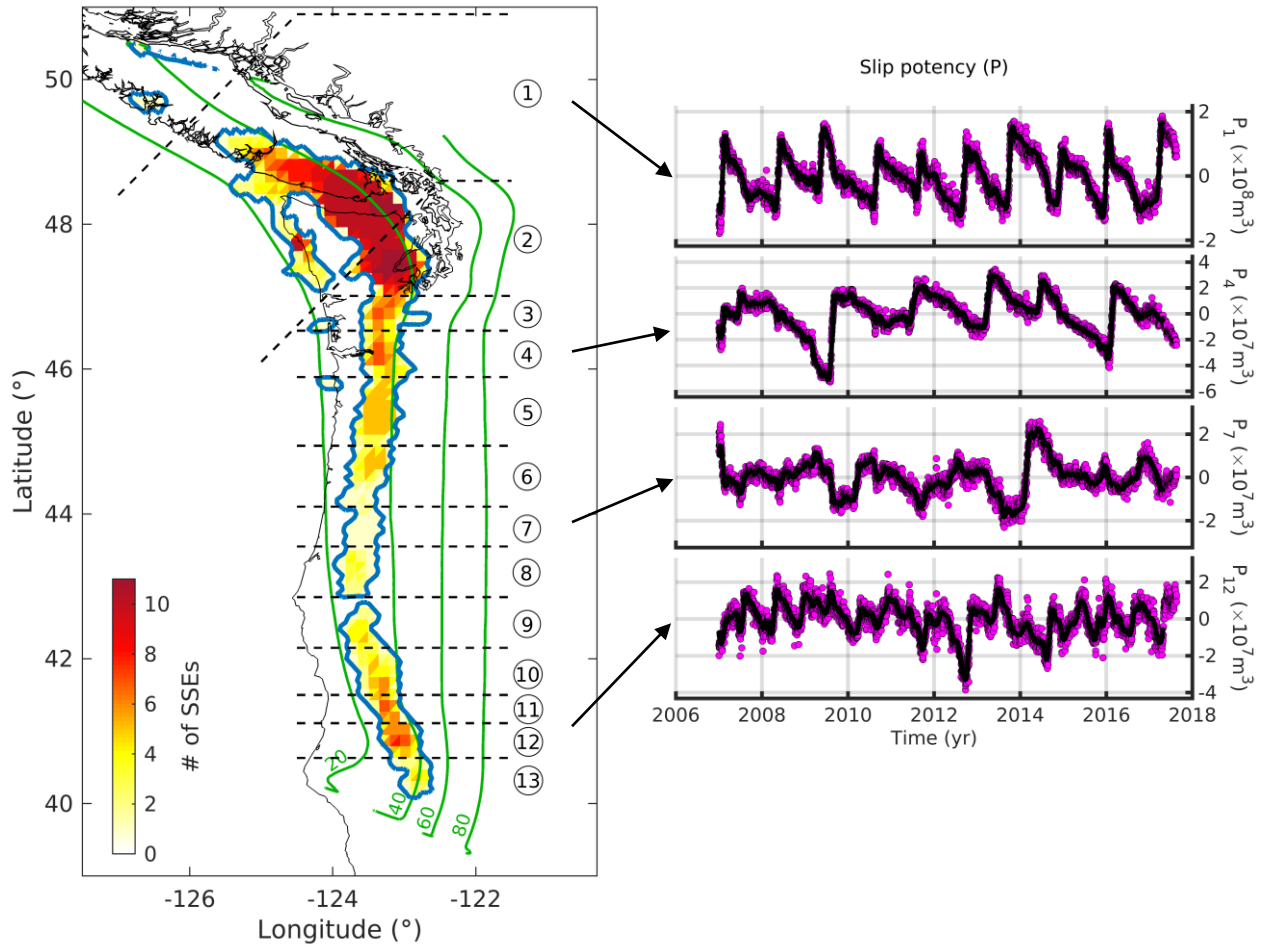


Fig. 1: Left panel: Modified from ref. 7. Blue contour line: SSES region. Color palette: Number of times a specific segment has ruptured in the time interval [2007.0, 2017.632] following the available SSES catalog⁶. Dashed black lines indicate the segmentation from ref. 6, here adopted. Black continuous line shows the coast for reference. Green lines are isodepths in km. Right panel: Slip potency ($P(t)$) for four selected segments. The calculation is performed over the entire area belonging to the segment and colored in the left panel. Magenta dots: Slip potency for segments #1, #4, #7, and #12. Black dots: Causal low-pass filtered slip potency. The causal filter introduces a 115 days time delay. For visual purpose, the filtered time series is shifted back

to the starting date 2007.0 (see Section Sleep potency rate in the Supplementary Material for more details).

141

142

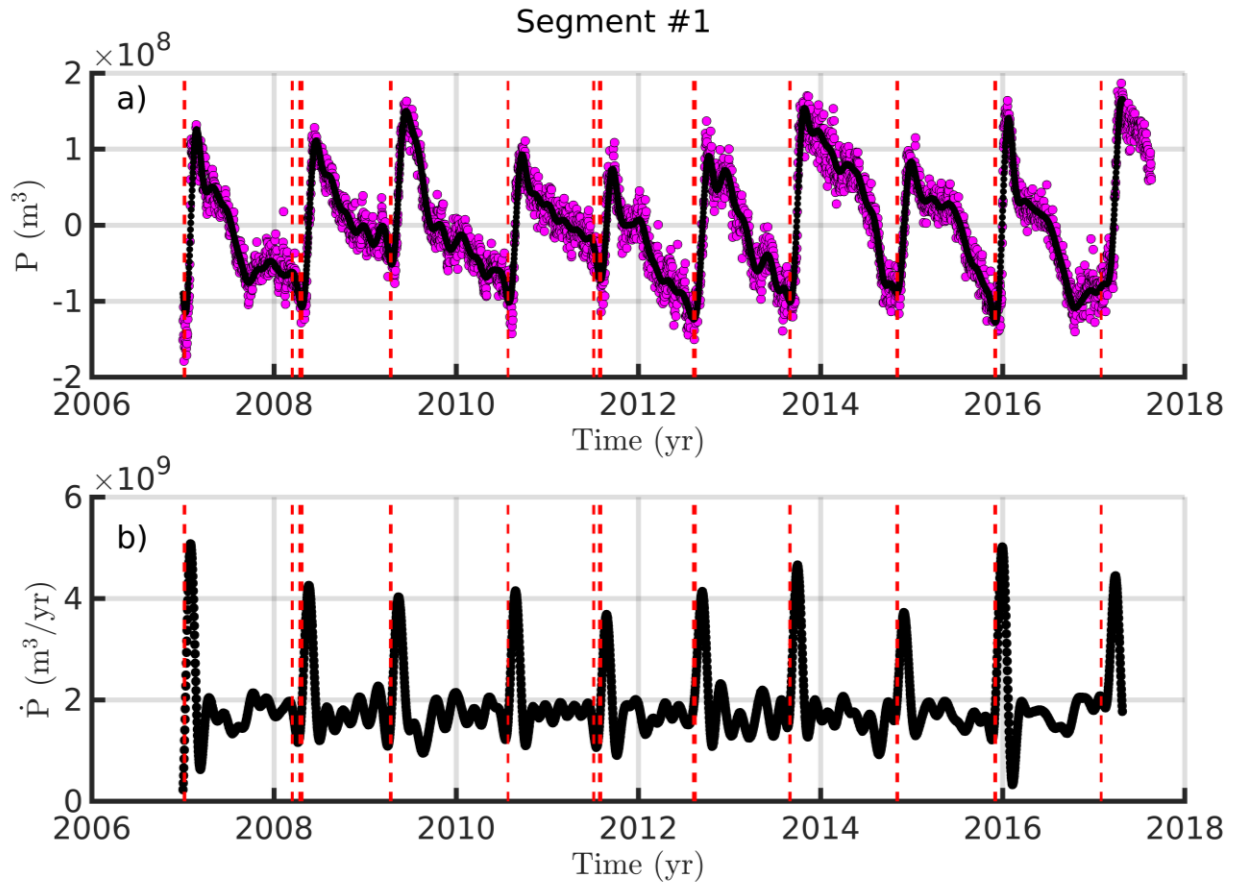


Fig. 2: a) Causally filtered slip potency (black dots) for segment #1 (same as in Fig. 1). Red dashed lines indicate the epochs for which the instantaneous dimension d is larger than its 95-th quantile up to that epoch and for which the slip potency is smaller than its 50-th quantile up to that epoch. b) Slip potency rate (black dots) for segment #1, calculated as the derivative of the causally filtered slip potency (panel a) using a 1 day interval time step, plus the long-term slip potency rate. Red dashed lines as in panel a).

143

144

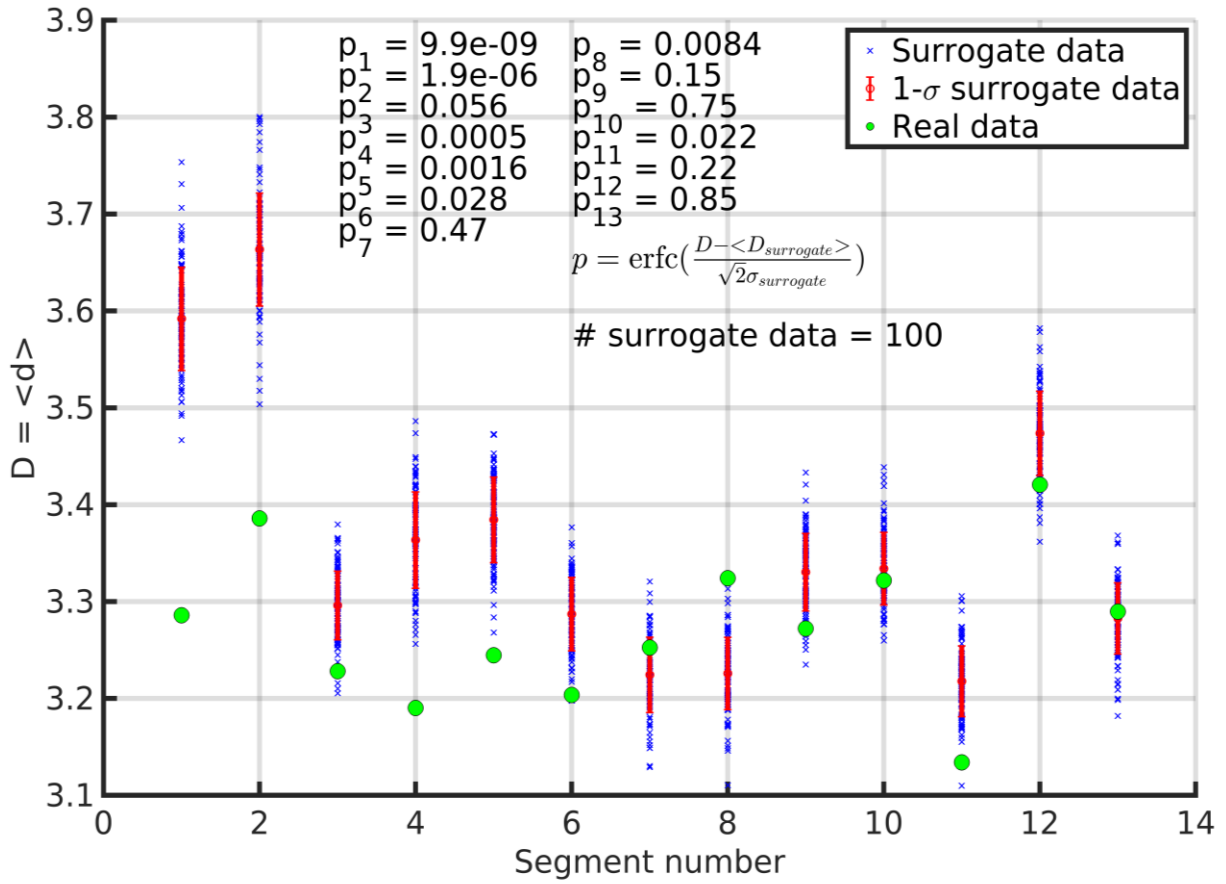


Fig. 3: The average attractor dimension D has been estimated as the average of the instantaneous dimension d . The surrogate data have been obtained randomizing the phases of the Fourier transformed slip potency and maintaining correlation between patches (see Section Surrogate Data in the Supplementary Material). The p -value is estimated after calculating D for every surrogate and comparing the data derived D with the distribution of the surrogate data derived $D_{\text{surrogate}}$. Only segments #1 and #2 show an average dimension significantly ($p < 0.001$) lower than the one derived from the surrogate data independently from the choice of the filter, of the norm to calculate distances, and of the quantile threshold to determine the exceedances (see Section Extreme Value Theory in the Supplementary Material and Fig. S2).

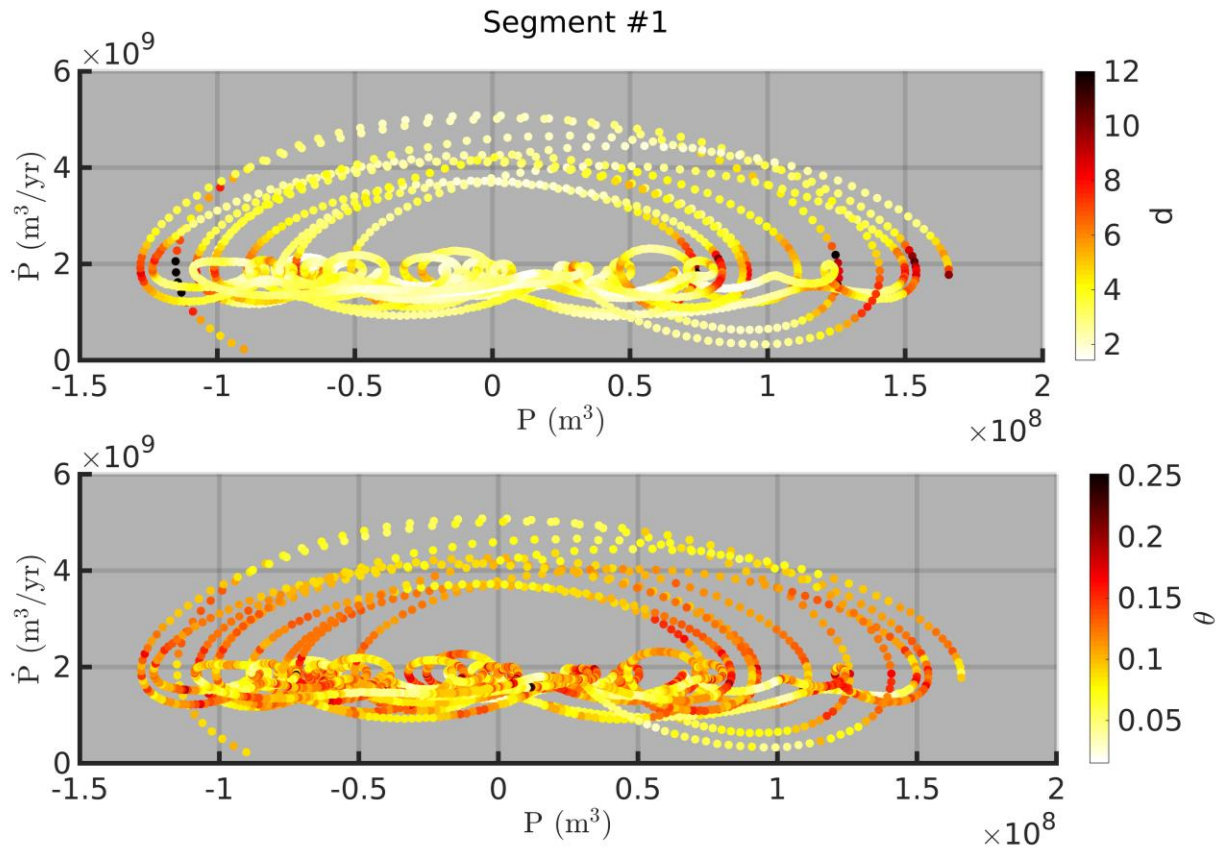


Fig. 4: We use the slip potency ($P(t)$) and the slip potency rate ($\dot{P}(t)$) as variables to reconstruct a Poincaré section of the attractor. The color palette indicates the instantaneous dimension d and the instantaneous extremal index θ in the top and bottom panels, respectively.

149 **Methods**

150 *Slip potency rate*

151 Given a subfault (n) of a given segment ($s = 1, \dots, 13$), we apply an equiripple low-pass filter to
152 the slip history ($u_{sn}^{rough}(t)$) from ref. 6 in order to obtain a smoothed slip history ($u_{sn}(t)$). The
153 passband frequency is $1/21 \text{ days}^{-1}$, the stopband is $1/35 \text{ days}^{-1}$, the passband ripple is 1 dB,
154 and the stopband attenuation is 60 dB. These filter specifications are used to create both a causal
155 and a non-causal filter. In particular, we use the Matlab® built-in function `filter` and the
156 function `FiltFiltM` (<https://www.mathworks.com/matlabcentral/fileexchange/32261-filterm>).
157 We then multiply the slip history by the area of each subfault (A_{sn}), getting the slip potency for
158 each subfault:

$$p_{sn}(t) = A_{sn}u_{sn}(t) \quad (\text{M1})$$

159

160 The total slip potency for the s -th segment is calculated as the slip potency integral over the entire
161 slipping area. In a discretized case where there are N_s subfaults belonging to the s -th segment, the
162 total slip potency is calculated as:

$$P_s(t) = \sum_{n=1}^{N_s} A_{sn} u_{sn}(t) \quad (\text{M2})$$

163 We also construct slip potency rate curves for each subfault:

$$\dot{p}_{sn}(t) = A_{sn}(v_{0sn} + \dot{u}_{sn}(t)) \quad (\text{M3})$$

164 and the total slip potency rate for each segment:

$$\dot{P}_s(t) = \sum_{n=1}^{N_s} A_{sn} (v_{0sn} + \dot{u}_{sn}(t)) \quad (\text{M4})$$

165 where v_{0sn} is the long-term reference loading velocity for a given patch belonging to a specific
166 segment as derived from ref. 6.

167 Non-causal filters cannot be adopted for real time applications because they use information from
168 future epochs. On the contrary, causal filters use only past and present information, but they
169 introduce a time delay in the filtered signal. The filter here adopted introduces the same time delay
170 of 115 days for every frequency. This information is relevant when discussing about the
171 predictability horizon of the system (see Main Text). The causally filtered slip potency for the 13
172 segments is shown in Fig. 1b for a selection of 4 segments, and in Fig. S1 for all of them. For
173 graphical purposes, the filtered time series have been shifted back in time in order to be in phase
174 with the original time series.

175 The choice of the variable to use for the dynamics reconstruction is important because it can bias
176 the correlation dimension estimation. In particular, the correlation dimension can be lowered when
177 using a variable strongly correlated only with a few variables of the system²⁹. Having in mind the
178 rate-and-state formalism for friction²¹, we hypothesize that the slip rate, and thus the slip potency
179 rate, is not only more strongly coupled than the slip potency to the other variables of the system,
180 but it is also coupled with more variables (e.g., the state variables). We find smaller values for the
181 correlation dimension of the attractor ν when using the slip potency as observable (Figs. S4 and
182 S5 and Section Embedding Theory in the Methods for their interpretation), confirming our
183 intuition that slip potency rate is a more suitable observable.

The effect of low-pass filters on dynamical systems' dimension estimation has been extensively studied. The filter here adopted is a finite-order nonrecursive filter. It has been proven theoretically that this class of filters should not modify quantities estimated via ET, but this might be not true for practical cases where finite time series are available³⁰. For both causally and noncausally filtered time series, we retrieved similar average dimensions, lower than those determined on non-filtered time series (see Fig. S6 for a comparison of causal and non-causal filter effects via EVT for the segment #1). In particular, for the segments with large SNR we notice that the application of a low-pass filter (both causal and non-causal) reduces the dimensionality of the system from values > 6 to values < 4 .

Filters applied to pure noise can potentially produce the spurious identification of finite correlation dimension³¹. For this reason, we have conducted the experiment with surrogate data also on filtered random time series (see Section Surrogate Data in the Methods).

Embedding Theory (ET)

We apply two methodologies based on ET to determine the correlation dimension of the strange attractor. The correlation dimension is defined as³²:

$$\nu = \lim_{r \rightarrow 0} \lim_{T \rightarrow \infty} \frac{\text{Log}(C(r, T))}{\text{Log}(r)} \quad (\text{M5})$$

where $C(r, T)$ is the correlation integral, r is a variable threshold distance, T is the time series length, and we use the base-10 logarithm Log .

202 The construction of the correlation integral typically involves two parameters: the delay time (τ)
 203 and the embedding dimension (m). Given a scalar time series $x(t)$, for example $\dot{P}_a(t)$, its values
 204 are used to construct an m -dimensional vector delaying in time the time series of an amount τ for
 205 $m - 1$ times:

$$\mathbf{F}(t, \tau, m) = [x(t), x(t - \tau), \dots, x(t - (m - 1)\tau)] \quad (\text{M6})$$

206 We can now define the correlation integral as³³:

$$C(r, T, w) = \frac{2}{T^2} \sum_{t_2=w}^T \sum_{t_1=1}^{T-t_2} H(r - \|\mathbf{F}(t_1 + t_2, \tau, m) - \mathbf{F}(t_1, \tau, m)\|) \quad (\text{M7})$$

207 where we introduce a cutoff parameter $w > 1$ to improve the convergence of the classical
 208 algorithm ($w = 1$) toward its limit $T \rightarrow \infty$, and H is the Heaviside function. Values of $w \sim \tau_{corr}$
 209 are recommended, where τ_{corr} is the autocorrelation time for the specific scalar time series under
 210 exam²⁹. We calculate τ_{corr} using the batch means method³⁴.

211 Basically, the algorithm counts how many m -dimensional vectors are closer than r at different
 212 times. If $m \geq 2\nu + 1$, then $\mathbf{F}$ is an embedding function of the strange attractor⁸, and we expect the
 213 correlation dimension to be independent of m . Figs. S4 and S5 show the ν vs $\text{Log}(r)$ curves for
 214 different m . The selected delay times τ are chosen in order to emphasize the scaling region. For
 215 every segment, we have tested values of $\tau = 2i + 1$ days, with i from 1 to 6, and $m = 4j$, with j
 216 from 1 to 5. In the case of a stochastic signal, we might expect that a scaling relationship between
 217 $C(r)$ and r does not hold, and larger values of ν are calculated when increasing m , i.e. ν does not
 218 saturate. Nonetheless, autocorrelated noise in short time series can fool the described algorithm²³.

219 For this reason, surrogate data are typically introduced, but instead of evaluating the plateau for ν ,
 220 which is quite subjective, we prefer to apply the methodology from EVT for this calculation. Given
 221 the time delays retrieved from ET (see Fig. S4), one accurate slip rate data per week might be
 222 sufficient to reconstruct an embedding for SSEs.

223 The second methodology derived from ET needs only the definition of the delay time in order to
 224 determine the minimum embedding dimension³⁵. The algorithm still uses the embedding function
 225 $\mathbf{F}$, but it exploits the nearest neighbors counting in the reconstructed phase space to detect false
 226 neighbors when changing the embedding dimension m .

227 The following two quantities are defined:

$$E(m) = \frac{1}{T - m\tau} \sum_{t=1}^{T-m\tau} \frac{\|\mathbf{F}(t, \tau, m+1) - \mathbf{F}_{n(t,m)}(t, \tau, m+1)\|}{\|\mathbf{F}(t, \tau, m) - \mathbf{F}_{n(t,m)}(t, \tau, m)\|} \quad (\text{M8})$$

$$E^*(m) = \frac{1}{T - m\tau} \sum_{t=1}^{T-m\tau} \|x(t + m\tau) - x_{n(t,m)}(t + m\tau)\| \quad (\text{M9})$$

228

229 where $1 \leq n(t, m) \leq T - m\tau$ is an integer such that $\mathbf{F}_{n(t,m)}(t, \tau, m)$ is the nearest neighbor of
 230 $\mathbf{F}(t, \tau, m)$. From these two quantities, the ratio at two subsequent embedding dimensions is
 231 calculated:

$$E_1(m) = \frac{E(m+1)}{E(m)} \quad (\text{M10})$$

$$E_2(m) = \frac{E^*(m+1)}{E^*(m)} \quad (\text{M11})$$

232 The quantity E_1 is relevant because if the studied time series is the result of a dynamic process, i.e.
 233 it comes from an attractor, then $E_1(m)$ saturates³⁵. In other words, we reach a value $\hat{m}$ such that
 234 E_1 stops changing increasing m above $\hat{m}$. The minimum embedding dimension will then be $\hat{m} +$
 235 1. The second quantity, E_2 , is introduced to check for randomness in the data. In theory, for
 236 stochastic time series E_1 should never saturate, but in practical cases it can be unclear if E_1 is
 237 slowly increasing with increasing m or not. If the time series of interest is the result of a stochastic
 238 process, we expect future data points to be independent from the previous ones. This means that
 239 $E_2(m) = 1$ for each m . If the studied time series is instead the result of a deterministic process,
 240 E_2 is not constant, and there must exist some m values such that $E_2(m) \neq 1$. We plot E_1 and E_2
 241 in Figs. S4 and S5 for the analysis on slip potency rate and slip potency time series, respectively.
 242 We show in Fig. S6 the results on both causally filtered (i.e., present values are depending only on
 243 the past and the present, such that the statistic E_2 can be used) and non-filtered time series for the
 244 example of segment #1. A minimum embedding dimension $m \simeq 10$ is detected from filtered data,
 245 implying a correlation dimension $\nu \leq (m - 1)/2 \lesssim 4.5$, while for non-filtered data E_2 remains
 246 almost constant at unitary values. This result is consistent with the results from EVT, for which a
 247 θ close to 1 is calculated for non-filtered time series (e.g., Fig. S6).

248

249 ***Extreme Value Theory (EVT)***

We use recent results of EVT in order to overcome some of the issues encountered when using ET algorithms. In particular, we would like: 1) a method to rigorously calculate a particular statistic (for example, the attractor's dimension), and, consequently, test if the calculated quantity is the result of a random process or not; and 2) a method to calculate also instantaneous properties of the attractor. The main idea behind the usage of EVT for the characterization of a dynamical system is to connect the statistics of extreme events to the Poincaré recurrence theorem^{27,36}.

Let us consider a generic point ζ on a strange attractor. Ultimately, the instantaneous dimension is a quantity which measures the density of neighbors in the phase space around ζ . In order to calculate this density, we can ask ourselves the following question: what is the probability to visit again a region of the phase space close to (i.e., in an arbitrary small radius ε from) ζ ? If we had access to all the possible states of the system (z), we could calculate the distances from the actual state under study, $\delta(z, \zeta)$. Then we would like to construct a random variable related to $\delta(z, \zeta)$ and use the second theorem of EVT³⁷ to be able to gain information about the density of neighbors around ζ . In fact, the second theorem of EVT states that, given a random variable Z with non vanishing probability distribution, we can set a threshold value q such that, for q sufficiently large, values of Z that exceed q (or exceedances) follow a Generalized Pareto Distribution (GPD).

For real case scenarios, we might have only one scalar time series, and here we consider the univariate case. Similarly to what is done to characterize atmospheric flows²⁷, we assume that our observed scalar time series (i.e., the slip potency rate) approximates possible states of the system. The only requirement to apply this methodology is the observed time series to be sampled from an underlying ergodic system. A possible improvement would be to consider a multivariate case with slip potency rates and/or slip potencies from adjacent segments, but we defer this complication to future investigations. For our case of interest, to generate the pool of all possible z we consider the

slip potency rates of all subfaults belonging to the segment under exam (equation M3). We then select ζ to be equal to the slip potency rate at a certain epoch t , and calculate the distance from all other possible states that we have recorded at different times. Recent studies^{36,38} have shown that if we construct the random variable given by the negative log-distances, $Z = -\ln(\delta(z, \zeta))$, we can use the GPD shape parameter (σ) to calculate the instantaneous dimension as:

$$d(\zeta) = \frac{1}{\sigma(\zeta)} \quad (\text{M12})$$

This assumes that the exceedances follow a GPD. We visually inspected the Q-Q plots of the exceedances vs a GPD, and discrepancies between the observed quantiles and those predicted by a GPD are noticed in the high quantile of the distribution. This probably reflects the fact that not many extreme events have been observed, and a longer slip history may provide better results. Repeating this calculation for different values of ζ extracted from the pool of values z , we can then estimate the attractor dimension simply averaging over d : $D = \langle d(\zeta) \rangle$. This is a very powerful result for two reasons. We can calculate the attractor dimension 1) without the need of an embedding, and 2) simply setting one parameter: a threshold q on the negative log-distances. Here we have used $q = 0.98$ percentile and $q = 0.99$ percentile of the negative log-distance, and we have tested both L-1 and L-2 norm distances. The results are overall similar (green dots in Fig. S2). We can now apply the method of surrogate data to each segment, and compare the reconstructed attractor dimensions with the one calculated from real data. We refer to Section Surrogate Data in the Supplementary Material for more details about the surrogate data method. We notice that using a threshold $q = 0.99$ reduces the average dimension to values < 3 . The threshold quantile q does not affect the conclusion that a chaotic deterministic dynamics can be

293 successfully detected only for the northernmost segments. It affects though the interpretation of
294 the total number of degrees of freedom (dof) needed to explain the system. Longer time series will
295 help to resolve this ambiguity since the value of $q = 0.99$ may be too high for the amount of
296 available data.

297 Another quantity of interest that can be derived is the instantaneous extremal index of the system
298 $\theta(\zeta)$ (Fig. 4b). The instantaneous extremal index can be defined as the inverse of the persistence
299 in the phase space, where the persistence tells us how long the trajectory sticks in the proximity of
300 a certain point in the phase space. In other words, while d is related to the density of points in a
301 certain neighborhood of the phase space, i.e. how many times a certain region of the phase space
302 is visited, the persistence time indicates for how long the system stays in a region in the
303 neighborhood of a given state. If a state ζ is a fixed point, we expect an infinite persistence, i.e.
304 null θ . On the other hand, if we are studying a stochastic process we expect the persistence to tend
305 to 0, i.e. unitary θ . In other words, we expect that at a certain epoch t the system will be in a state
306 ζ , and then at a subsequent epoch $t + \Delta t$ to be in a region of the phase space far from the one
307 occupied at time t . Looking at the extremal index from this angle gives us the intuition that it can
308 be related to the predictability of the system.

309 The extremal index (Θ), a parameter in the range $[0,1]$, measures the degree of clustering of
310 extremes in a stationary process ($\dot{P}_S(t)$ in our case), and can be defined as the reciprocal of the
311 mean cluster size³⁹. A relationship between Θ and the metric entropy H of a system has been
312 recently demonstrated²⁸:

$$\Theta \sim 1 - e^{-H} \quad (\text{M13})$$

The relationship between Θ and θ is not as straightforward as the one between D and d , but we must have $\Theta \in [\theta_{min}, \theta_{max}]$. We can thus deduce a range of values for H . Given the values of θ that we have found (Fig. S3 of the Supplementary Material) and the fact that the metric entropy is equal to the sum of all the positive Lyapunov exponents, we deduce that there is at least one positive Lyapunov exponent in our system.

We calculate θ via a maximum likelihood estimator⁴⁰. When we look at the calculated θ (see Fig. S3 for non-causal filter, and S5 for causal and non-causal filter relative to segment #1), we see that both causally and non-causally filtered data show values far from 1. This already indicates that, no matter the causality or not of the filter, the system at the selected frequencies shows predictability features. We observe a different situation when performing the analysis on non-filtered data (Fig. S6). The values of θ are now very close to 1, implying a predictability horizon shorter than the sampling time. This is consistent with the fact that in such non-filtered slip potency rate time series the high-frequency noise dominates, which is a random process.

Surrogate Data

The concept behind surrogate data techniques is rooted in statistical hypothesis testing. The method requires to state a null hypothesis, and, using the words of ref. 23, "The idea is to test the original time series *against* the null hypothesis by checking whether the discriminating statistic computed for the original time series differs significantly from the statistics computed for each of the surrogate sets." The necessity to use surrogate data derives from the intrinsic finiteness of the available data: it is always possible to generate the observations with a particular random process.

334 The null hypothesis that we test consists in assuming that the data can be described via a linear
335 stochastic model. Surrogate data should be generated before any filtering⁴¹ thus we first generate
336 the surrogate data from the original slip potency time series, then filter both the actual and
337 surrogate data, and finally calculate the slip potency rate. Since we are using slip potency rates on
338 multiple subfaults, when shuffling the signal phases we want to preserve not only the
339 autocorrelation of each slip potency rate time series, but also the cross correlations between
340 subfaults, and we thus use a generalization of the phase-randomized Fourier transform algorithm⁴².
341 Despite the fact that filtering should not compromise the actual chaotic nature of the system⁴³, the
342 estimate of D might depend on the applied filter, and we consequently test the method on both
343 causally and non-causally filtered time series. With both filters we witness a reduction of the
344 attractor dimension with respect to non-filtered data.

345 If we perform the same analysis on the unfiltered time series we find extremal index values close
346 to 1, with predictability horizons smaller than the data sampling rate, indicating a random system
347 (Fig. S6 for an example relative to segment #1). This shows that the noise is dominating, masking
348 the SSEs dynamical structure. We further test the effects of the filter on the predictability of the
349 time series generating (pseudo-)random time series and applying the same filter to them. The
350 number of generated random time series is equal to the number of subfaults in segment #1, which
351 is the segment with the largest number of subfaults. We then generate surrogate data and calculate
352 the average dimension on both the filtered random time series and the surrogate data. The result
353 shows that we would not be able to reject the null hypothesis according to which the time series
354 were generated by a random process (Fig. S7).