

# Continuous chatter of the Cascadia subduction zone revealed by machine learning

Bertrand Rouet-Leduc<sup>1,2\*</sup>, Claudia Hulbert<sup>1,2</sup> and Paul A. Johnson<sup>1</sup>

**Tectonic faults slip in various manners, which range from ordinary earthquakes to slow slip events to aseismic fault creep. Slow slip and associated tremor are common to many subduction zones, and occur down-dip from the neighbouring locked zone where megaquakes take place. In the clearest cases, such as Cascadia, identified tremor occurs in discrete bursts, primarily during the slow slip event. Here we show that the Cascadia subduction zone is apparently continuously broadcasting a low-amplitude, tremor-like signal that precisely informs of the fault displacement rate throughout the slow slip cycle. Using a method based on machine learning previously developed in the laboratory, we analysed large amounts of raw seismic data from Vancouver Island to separate this signal from the background seismic noise. We posit that this provides indirect real-time access to fault physics on the down-dip portion of the megathrust, and thus may prove useful in determining if and how a slow slip may couple to or evolve into a major earthquake.**

After the discovery of tectonic tremor in Japan by Obara in 2002<sup>1</sup>, Rogers and Dragert identified episodic transient slow slip and tremor on the Cascadia fault beneath Vancouver Island, Canada<sup>2</sup>. The slow slip occurs roughly every 13 months, when the overriding North American plate lurches southwesterly over the subducting Juan de Fuca plate (Fig. 1). Subduction zones are divided into regions that slip during earthquakes and regions that slowly slip, primarily aseismically<sup>3</sup>. Slow slip in Cascadia is located down-dip from the seismogenic rupture zone, cyclically loading the locked fault<sup>4–9</sup> that has not broken since 1700<sup>10</sup> (Fig. 1). Slow, aseismic slip on the plate interface is accompanied by discrete bursts of tectonic tremor, which potentially radiate from strong asperities<sup>2,11</sup>. Recently, smaller slow slips that take place intermittently between the large, periodic ones were identified, and indicate that slow slip occurs over a spectrum of timescales<sup>12,13</sup>.

Elastic wave radiation by rupture propagation provides an accurate model of regular earthquakes<sup>14,15</sup>. In contrast, the physics of slow slip and associated tremor is not well understood<sup>16</sup>. Several questions remain open regarding the slip frictional physics, rupture propagation, the relationship between slip modes<sup>16,17</sup> and the distribution of energy<sup>18</sup>. Slow slip has been observed to precede large subduction earthquakes<sup>19–23</sup>, which suggests that they may modulate or be part of the nucleation process of major fault ruptures. Studies also found a large discrepancy between the slow slip surface displacement and the cumulated tremor energy–seismic moment calculated from identified tremor episodes, which are several orders of magnitude below the estimated energy release inferred from global positioning system (GPS) measured displacement<sup>7,18,24</sup>.

We find that the Cascadia megathrust emits a tremor-like signal apparently at all times, which may account for most of this missing energy. Statistical characteristics of this signal are a fingerprint of the displacement rate of the fault. These results resemble our previous work on laboratory faults<sup>25</sup>, which suggests that some aspects of the underlying physics may scale from a laboratory fault to Earth. If this is the case, analysis of the continuous seismic signal may help unravel fundamental questions about slow earthquakes: what is the mechanism behind the quasi-dynamic slow slip propagation and

can large earthquakes be triggered by slow slip events? In laboratory experiments of slow slip, we found that the full continuous seismic signal emitted by artificial faults is imprinted with rich information regarding their physical state<sup>26</sup>. In the laboratory, analysis of continuous seismic waves enables us to estimate fault friction and shear displacement accurately and reveals precursory information regarding the timing and magnitude of upcoming slip events<sup>25,26</sup>.

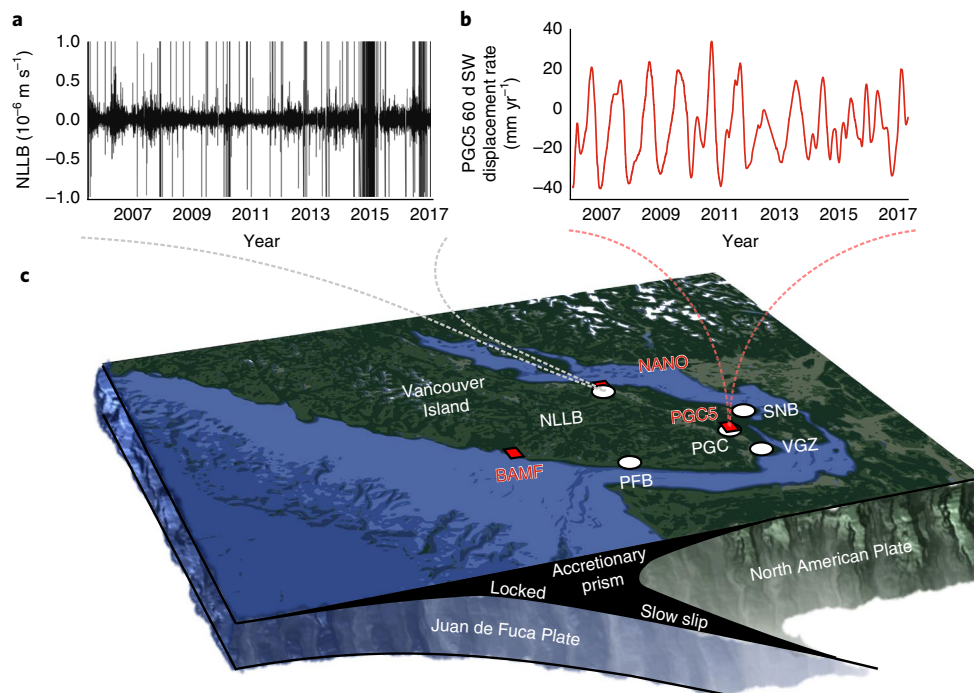
## GPS displacement can be estimated from seismic data

Analyses based on earthquake and tremor catalogues offer important insight into slip processes, but discard most of the continuously recorded seismic data. The approach we describe harnesses the power of supervised machine learning using other reliable geophysical data sets as a label (or target) to extract information from continuous seismic waves, in contrast to labels built by hand or heuristics. More specifically, we pose the problem as a regression between the statistical characteristics of the continuous seismic data and the surface GPS displacement rate. We describe the data-processing and machine-learning procedure below and provide a detailed description in Methods.

**Data and processing.** We rely on seismic data from the Canadian National Seismograph Network (CNSN)<sup>27</sup>. To measure the displacement rate presumed to be taking place at the down-dip transitional section of the subduction zone where slow slips occur, we use as a proxy the data from nearby GPS stations from the Western Canada Deformation Array, preprocessed by the United States Geological Survey<sup>28</sup>. We considered the total horizontal displacement (east/west + north/south GPS components), such that a negative displacement is in the direction of the subduction (northeast), and a positive displacement is in the direction of the slip (southwest). Both the seismic and GPS stations we used in this study are shown in Fig. 1. The seismic data are recorded or re-sampled at 40 Hz, and the GPS data are processed at one sample per day.

We begin by processing the seismic and GPS data. The raw GPS data are extremely noisy, and therefore we process them to emphasize the slow slip events (Fig. 1b). The denoised displacement rate

<sup>1</sup>Los Alamos National Laboratory, Geophysics Group, Los Alamos, New Mexico, USA. <sup>2</sup>These authors contributed equally: Bertrand Rouet-Leduc, Claudia Hulbert. \*e-mail: [bertrandrl@lanl.gov](mailto:bertrandrl@lanl.gov)



**Fig. 1 | Map and schematic of the region analysed: Vancouver Island and the subduction zone. a, b,** Our goal is to determine if, as in the laboratory, statistical characteristics (features) of the full continuous seismic data (**a**) can be used to estimate the GPS displacement rate accurately (**b**). Note that the seismic data shown are clipped to eliminate earthquakes so that waveform structure between slip events used in the analysis can be seen. **c,** The seismic stations are shown in white and the GPS stations in red. SW, southwest.

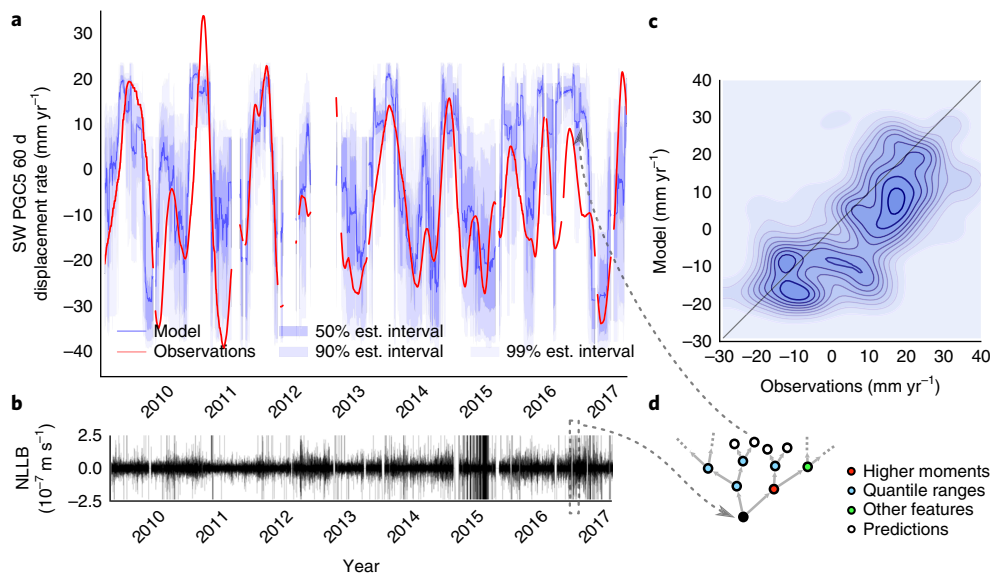
is calculated as the moving average of the slope of a least-squares linear regression over the time window of  $n_{\text{GPS}}$  days considered. The Juan de Fuca plate is subducting at a rate of about  $40 \text{ mm yr}^{-1}$  to the northeast<sup>29,30</sup>. Therefore, when the surface (smoothed) GPS displacement rate shown in Fig. 1b is close to  $-40 \text{ mm yr}^{-1}$  (that is,  $40 \text{ mm yr}^{-1}$  in the northeast direction), the transition zone of the megathrust is momentarily locked (the overriding plate is moving at the same velocity as the subducting plate). When the GPS displacement rate is greater than  $-40 \text{ mm yr}^{-1}$  the transition zone is slowly slipping, and it does so at various velocities with peaks than can approach  $+40 \text{ mm yr}^{-1}$  (that is,  $40 \text{ mm yr}^{-1}$  towards the southwest) at the surface, which presumably corresponds to peak slip velocities at the plate interface of  $80 \text{ mm yr}^{-1}$ .

We process the seismic data as follows. For each station and for each day, we correct the seismic signal for the instrument response, remove the average and detrend the data. We make several copies of the seismic signal, bandpassed at frequencies between 8 and 13 Hz, in increments of 1 Hz (8–9 Hz, 9–10 Hz, 10–11 Hz, 11–12 Hz and 12–13 Hz). Although tremor exists down to frequencies of approximately 1 Hz, in the 8–13 Hz band the signal is nonetheless strong, and contributions from microseisms that could affect the analysis are very weak<sup>31</sup>. After bandpassing, the signal is clipped ( $10^{-7} \text{ m s}^{-1}$ ) to remove the impulsive signals that arise from regular earthquakes<sup>32</sup>. This is done because we are interested in analysing the background signal, and want to avoid the influence of earthquake waveforms (Fig. 1a).

We compress the seismic data into a list of statistical features. These seismic features are obtained by scanning the processed seismic data using moving time windows of length  $n_{\text{seismic}}$  days. The features are built to capture the distribution of the seismic data within each time window. For each station and for each band of 1 Hz from 8 to 13 Hz, we compute the range of the data, the normalized and non-normalized higher order moments of the data, as well as interquantile ranges that effectively capture the second moment

of the distribution of the seismic data, but are more robust to noise (Methods gives a detailed list of all the features). Statistical features are calculated for hourly or daily segments of the seismic signals. These hourly or daily features are then averaged over the time interval considered. This is accomplished using time windows with fixed duration. These time windows correspond to a fraction of the length of the slow slip cycle, which is approximately 13 months (a time window is drawn to scale in Fig. 2). We find that the window length  $n_{\text{seismic}}$  has little impact on our results (we varied the window from 1 hour to 60 days (Supplementary Figs. 1 and 2)). Features averaged over  $n_{\text{seismic}}$  hours or days are used to estimate the average displacement rate over  $n_{\text{GPS}}$  days, with the first day of both windows coinciding. We index the figures and our machine-learning database with the last hour or day of the GPS window. In the following, ‘seismic features’ refers to features averaged over a time window  $n_{\text{seismic}}$ . We know from our previous studies of laboratory shear experiments that distribution-related features are the most important to probe fault characteristics.

**Constructing the model.** Once all the seismic features are calculated for all the seismic stations, and over the different frequency intervals, they are used as inputs to build our machine-learning model. The model attempts to uncover the underlying function that relates these seismic features to GPS displacement rates, and thus formulates the problem as a regression. The model is built in a training phase in which the algorithm has access to the input data (seismic features) and the output or label (GPS displacement rate). The model is then evaluated on a new data set, the testing set, with access only to the seismic features and no information about the GPS displacement rate. For the testing set, the GPS displacement data are only used to evaluate the performance of the model, that is, the agreement between the model estimates and the true GPS displacement values. We use the first four years of the seismic features and GPS data as the training set (from 2005 to 2008), and the



**Fig. 2 | Estimating the GPS displacement rate from the continuous seismic data.** **a**, Smoothed GPS displacement rate (red). Estimate (est.) from the machine-learning model (blue) with estimation intervals (shades of blue) using statistical features of the full continuous seismic data (Methods) from the six considered seismic stations (PGC, NLLB, SNB, VGZ, PFB and MGB) as input. The figure shows the testing set, for which the algorithm only has access to the seismic data (for example, as in **b**). Gaps indicate missing data. **b**, Continuous seismic data from station NLLB over the same time interval. **c**, Distribution of observed versus predicted displacement rates; the contours show empirical isodensity, from 10 to 90%. The Pearson correlation coefficient between the estimates and actual displacement rate is 0.66, which shows that the continuous seismic waves contain rich information about the fault's state, apparently at all times. **d**, Schematic of the approach.

following eight years (from 2009 to 2017) as the testing set. The model's performance in testing is evaluated by using the Pearson correlation coefficient, a standard regression metric that measures the linear correlation between two quantities. We also build independent models to estimate the GPS displacement rate from the seismic data for other GPS stations and for various window sizes (Supplementary Figs. 1–3).

We use the random forest algorithm<sup>33</sup>, an ensemble of decision trees, to directly and explicitly build an estimator of GPS displacement rate from features of the seismic data. This algorithm allows us to probe the model to identify which features are critical to estimate the fault's displacement, in contrast to more 'blackbox' approaches, for example, neural networks. We rely on Bayesian optimization to set the model's hyperparameters (these include window length, tree depth and so on) by a fivefold cross-validation (Methods gives details on random forests and hyperparameter optimization).

Figure 2 shows the random forest model estimations of displacement rate on the testing set for the GPS station PGC5 as a blue bold line for  $n_{\text{seismic}} = n_{\text{GPS}} = 60$  days. The actual GPS displacement rate is shown in red. Assuming the GPS data are a proxy for the displacement rate of the down-dip plate interface, these results show that continuous seismic waves can be used to estimate directly the fault's displacement rate. In contrast, a model built exclusively from catalogued bursts of tremor is unable to estimate displacement rates, except when slip rates are very high (Supplementary Fig. 4b,c).

### The megathrust broadcasts a signature of its displacement

In Fig. 3 we show the most informative statistical feature of the seismic signal that the random forest identifies and uses to build an accurate model of displacement rate: the 60–40% inter-quantile range from 8 to 9 Hz at the NLLB seismic station, for  $n_{\text{seismic}} = n_{\text{GPS}} = 60$  days, and for  $n_{\text{seismic}} = 1$  hour and  $n_{\text{GPS}} = 60$  days. This feature accounts for most of the fault displacement rate estimates shown in Fig. 2, but Fig. 3 simply shows the evolution of

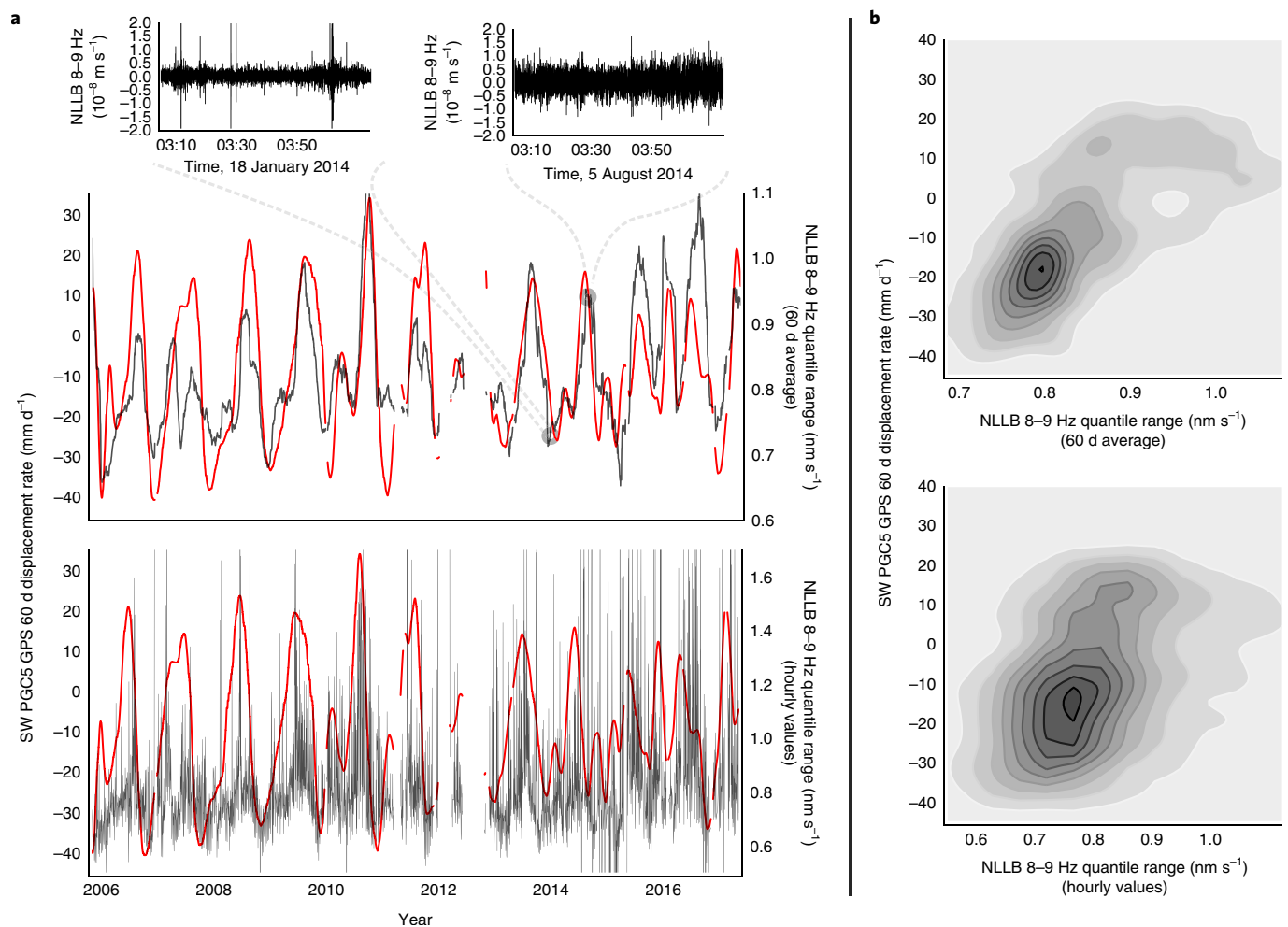
this feature and GPS displacement rate over time, with no machine learning involved.

The most informative features are proportional to the GPS displacement rate, increasing whenever the North American plate slips towards the southwest. Calculating the tremor 'power' over a window of seismic signal enables a precise estimation of the average displacement rate of the plate. This result demonstrates that a characteristic of the statistical distribution of the continuous seismic data is a fingerprint of the GPS slip, and shows that the accuracy of the random forest model cannot be an artefact.

A comparison between this continuous feature and catalogued tremor events from the PNSN catalogue<sup>8</sup> shows that they are proportional when slip rates are high (Supplementary Fig. 4a). This is not the case at low slip rates, when tremor might be more difficult to identify due to the lower amplitudes. This suggests that catalogued tremor episodes and the continuous tremor signal we identified have a common origin. We are currently working to locate the newly identified signal.

When analysing earthquake cycles from a laboratory fault, one can derive a similar machine-learning model that relates fault displacement to features of the continuous seismic signal emitted from the fault. In the case of the laboratory, the most important feature is the power of the seismic signal. In the laboratory, determining the power of the seismic signal at any given moment is sufficient to accurately infer the frictional state of the fault at that same moment, which leads to an equation of state between the seismic power and friction<sup>25</sup>.

Interestingly, in the case of the Cascadia slow earthquakes, the statistical features identified as most important by the machine-learning model are very similar to those found in the laboratory<sup>25</sup>. The most important features are related to the power of the signal, but with extreme values removed, using interquantile ranges instead of higher-order moments. Extreme values are, in general, related to anthropogenic noise or small earthquakes, and quantiles help eliminate them and enable the model to focus on the acoustic power of



**Fig. 3 | Strong correlations between the statistics of the seismic data and the GPS displacement rate.** **a**, Smoothed displacement rate at the PGC5 GPS station (red). Best feature of the seismic data (black), the 60–40 interquantile range from 8 to 9 Hz at the NLLB seismic station, averaged over 60 days (upper plot) or 1 h (lower plot). The upper insets show the seismic data when the interquantile range is low (left) versus when it is high (right). This change in the interquantile range over the fault cycle leads to accurate slip-rate estimates. **b**, Density plots that show the empirical distribution of displacement rate at the PGC5 GPS station versus the best feature of the seismic data (same as in **a**) for all the data considered (from 2005 to 2017). The y axis is the average GPS displacement rate within a window of 60 days, and the x axis is the corresponding value of the statistical feature during the same 60 day window for the upper plot, and from hourly values at the start of the GPS window for the lower plot. Gaps indicate missing data.

the continuous background tremor. The interquantile range feature has units of velocity and thus its square is a measure of energy and power for fixed and equal time windows. Although the continuous signal identified probably comes from a set of heterogeneous processes distributed spatially at the plate interface and possibly in the overriding plate<sup>6</sup>, which involves a suite of fault patches, the fact that very similar statistical features are found in the laboratory and Earth suggests that the underlying physics scales from a laboratory fault to faults in Earth. Enhanced fluid pressure on or near the plate interface is often appealed to as well, and potentially gives rise to a large pore pressure and a decrease in effective pressure, in line with the lab experiments. There is also evidence that the region becomes weaker due to a reduced strength in the overriding lower crust caused by a thermally induced brittle–ductile transition<sup>34</sup>. In short, the processes that generate tremor signals in Cascadia bear similarities with those in a simple laboratory fault, despite the considerable differences in frictional complexity.

Our results indicate that tremor in Cascadia occurs all the time, or nearly all the time. The point is made more clearly in Fig. 3, in which we show that the background tremor energy tracks the GPS

displacement rate on an hourly basis. This means there is sufficient tremor in any hourly segment of the seismic data we analysed to obtain physically meaningful tremor statistics that inform us of the fault slip rate. As the continuous tremor and the GPS slip rate appear to be proportional (Fig. 3b), our results point towards tremor being driven by the background slow slip. This proportionality holds even for low slip rates, when the displacement rate at the surface is in the direction of the subduction and slightly lower than the Juan de Fuca subducting rate. These results also indicate that the slip can occur below the GPS threshold and that the seismic signal may be a more sensitive indicator of slip than GPS, as suggested by others<sup>4,12</sup>. Our continuous tremor-like signal tracks the state of the slowly slipping fault on an hourly basis, whereas GPS displacement rates are too noisy to determine the fault slip when calculated on windows smaller than a few days (Fig. 3 and Supplementary Fig. 1).

## Conclusions

The continuous tremor-like signal we identify tracks the slow slip rate, apparently at all times, and so provides real-time access to the physical state of the slowly slipping portion of the megathrust. As



the slow earthquakes transfer stress to the adjacent locked region where megaquakes originate, careful monitoring of this tremor-like signal may provide new information on the locked zone, with the potential to improve earthquake hazard assessment in Cascadia.

### Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at <https://doi.org/10.1038/s41561-018-0274-6>.

Received: 5 June 2018; Accepted: 14 November 2018;

Published online: 17 December 2018

### References

- Obata, K. Nonvolcanic deep tremor associated with subduction in southwest Japan. *Science* **296**, 1679–1681 (2002).
- Rogers, G. & Dragert, H. Episodic tremor and slip on the Cascadia subduction zone: the chatter of silent slip. *Science* **300**, 1942–1943 (2003).
- Perfettini, H. et al. Seismic and aseismic slip on the Central Peru megathrust. *Nature* **465**, 78–81 (2010).
- Wech, A. G., Creager, K. C. & Melbourne, T. I. Seismic and geodetic constraints on Cascadia slow slip. *J. Geophys. Res. Solid Earth* **114**, B10316 (2009).
- Bartlow, N. M., Miyazaki, S., Bradley, A. M. & Segall, P. Space–time correlation of slip and tremor during the 2009 Cascadia slow slip event. *Geophys. Res. Lett.* **38**, L18309 (2011).
- Kao, H. et al. Spatial–temporal patterns of seismic tremors in northern Cascadia. *J. Geophys. Res. Solid Earth* **111**, B03309 (2006).
- Kao, H., Shan, S.-J., Dragert, H. & Rogers, G. Northern Cascadia episodic tremor and slip: a decade of tremor observations from 1997 to 2007. *J. Geophys. Res. Solid Earth* **114**, B00A12 (2009).
- Wech, A. G. & Creager, K. C. Automated detection and location of Cascadia tremor. *Geophys. Res. Lett.* **35**, L20302 (2008).
- Wang, K., Hu, Y. & He, J. Deformation cycles of subduction earthquakes in a viscoelastic earth. *Nature* **484**, 327–332 (2012).
- Goldfinger, C., Nelson, C. H., Johnson, J. E. & Party, T. S. S. Holocene earthquake records from the Cascadia subduction zone and northern San Andreas fault based on precise dating of offshore turbidites. *Ann. Rev. Earth Planet. Sci.* **31**, 555–577 (2003).
- Wech, A. G. & Bartlow, N. M. Slip rate and tremor genesis in Cascadia. *Geophys. Res. Lett.* **41**, 392–398 (2014).
- Frank, W. B. Slow slip hidden in the noise: the intermittence of tectonic release. *Geophys. Res. Lett.* **43**, 125–10,133 (2016).
- Hawthorne, C. J. & Rubin, A. M. Short-time scale correlation between slow slip and tremor in Cascadia. *J. Geophys. Res. Solid Earth* **118**, 1316–1329 (2013).
- Brace, W. F. & Byerlee, J. D. Stick–slip as a mechanism for earthquakes. *Science* **153**, 990–992 (1966).
- Scholz, C. H. *The Mechanics of Earthquakes and Faulting* (Cambridge Univ. Press, Cambridge, 2002).
- Obata, K. & Kato, A. Connecting slow earthquakes to huge earthquakes. *Science* **353**, 253–257 (2016).
- Satake, K., Shimazaki, K., Tsuji, Y. & Ueda, K. Time and size of a giant earthquake in Cascadia inferred from Japanese tsunami records of January 1700. *Nature* **379**, 246–249 (1996).
- Aguiar, A. C., Melbourne, T. I. & Scrivner, C. W. Moment release rate of Cascadia tremor constrained by GPS. *J. Geophys. Res. Solid Earth* **114**, B00A05 (2009).
- Ito, Y. et al. Episodic slow slip events in the Japan subduction zone before the 2011 Tohoku–Oki earthquake. *Tectonophysics* **600**, 14–26 (2013).
- Hasegawa, A. & Yoshida, K. Preceding seismic activity and slow slip events in the source area of the 2011 Mw 9.0 Tohoku–Oki earthquake: a review. *Geosci. Lett.* **2**, 6 (2015).
- Kato, A. et al. Propagation of slow slip leading up to the 2011 Mw 9.0 Tohoku–Oki Earthquake. *Science* **335**, 705–708 (2012).
- Radiguet, M. et al. Triggering of the 2014 Mw 7.3 Papanoa earthquake by a slow slip event in Guerrero, Mexico. *Nat. Geosci.* **9**, 829–834 (2016).
- Ruiz, S. et al. Intense foreshocks and a slow slip event preceded the 2014 Iquique Mw 8.1 earthquake. *Science* **345**, 1165–1169 (2014).
- Kostoglodov, V. et al. The 2006 slow slip event and nonvolcanic tremor in the Mexican subduction zone. *Geophys. Res. Lett.* **37**, L24301 (2010).
- Rouet-Leduc, B. et al. Estimating fault friction from seismic signals in the laboratory. *Geophys. Res. Lett.* **45**, 1321–1329 (2018).
- Hulbert, C. et al. Laboratory earthquake prediction illuminates connections between the spectrum of fault slip modes. *Nat. Geosci.* <https://doi.org/10.1038/s41561-018-0272-8> (2018).
- IRIS DMC FDSNWS Dataslect Web Service (IRIS, Geological Survey of Canada, accessed 1 September 2017); <http://service.iris.edu/fdsnws/dataslect/1/>
- Murray, J. R. & Svarc, J. Global positioning system data collection, processing, and analysis conducted by the US Geological Survey Earthquake Hazards Program. *Seismol. Res. Lett.* **88**, 916–925 (2017).
- Gripp, A. E. & Gordon, R. G. Young tracks of hotspots and current plate velocities. *Geophys. J. Int.* **150**, 321–361 (2002).
- Xue, M. & Allen, R. M. The fate of the Juan de Fuca plate: implications for a Yellowstone plume head. *Earth Planet. Sci. Lett.* **264**, 266–276 (2007).
- Zhang, J. et al. Cascadia tremor spectra: low corner frequencies and earthquake-like high-frequency falloff. *Geochem. Geophys. Geosyst.* **12**, Q10007 (2011).
- Bensen, G. et al. Processing seismic ambient noise data to obtain reliable broad-band surface wave dispersion measurements. *Geophys. J. Int.* **169**, 1239–1260 (2007).
- Breiman, L. Random forests. *Machine Learning* **45**, 5–32 (2001).
- Holtkamp, S. & Brudzinski, M. Determination of slow slip episodes and strain accumulation along the Cascadia margin. *J. Geophys. Res. Solid Earth* **115**, B00A17 (2007).
- Beyreuther, M. et al. Obspy: a python toolbox for seismology. *Seismol. Res. Lett.* **81**, 530–533 (2010).
- Pedregosa, F. et al. Scikit-learn: machine learning in python. *J. Machine Learning Res.* **12**, 2825–2830 (2011).

### Acknowledgements

This work was funded by Institutional Support (LDRD) at Los Alamos and the DOE Office of Basic Research, Geoscience Program. We are grateful to H. Dragert, H. Kao, T. Melbourne, J. Gombert, D. Trugman, R. Guyer, A. Delorey, J. Murray, I. McBrearty and C. Lee for fruitful comments and discussions. We thank W. Frank for his extensive review of our work and his suggestions. We thank T. Cote, X. Jin and M. Kolaj from the CNSN for their data and help.

### Author contributions

All the authors devised the original study. B.R.L. and C.H. conducted the machine-learning work and all the authors contributed to writing the manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41561-018-0274-6>.

Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints).

Correspondence and requests for materials should be addressed to B.R.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© This is a U.S. government work and not under copyright protection in the U.S.; foreign copyright protection may apply 2018

## Methods

The methodology based on machine learning used in this manuscript to analyse continuous seismic data and map it to the bulk behaviour of fault(s) was first developed on laboratory data<sup>25,35</sup>. In the following we detail how we modified this method to be applicable to real seismic data. In particular, the features we extract from the continuous seismic data, which is to say the way in which we present the seismic data to the machine-learning algorithms, are modified somewhat extensively. The machine-learning models we apply are very similar, however.

### Step-by-step description of data analysis and machine-learning methods.

1. Processing of the seismic data. We obtained the raw seismic data for stations PGC, NLLB, SNB, VGZ, PF, and MGB from the CNSN<sup>27</sup>. We corrected the data for instrument response using the corresponding RESP files on the IRIS server, and used the simulate function from the obspy Python library<sup>36</sup>. We then centred and detrended the seismic data. We obtained one file for each day and for each station of corrected horizontal (east component) seismic velocity data, from 2005 to 2017.
2. Defining statistical features of the seismic data. This step is crucial, as it extracts the characteristics (features) of the seismic data that the machine-learning model will use. Here we created a bandpassed copy of each file of seismic data, in frequency bands from 8 Hz to 13 Hz at 1 Hz intervals. We then computed statistical features of the seismic data for each day, for each station and for each frequency band. These statistical features are: range of the seismic velocity data, root mean square, higher-order moments (variance, skew and kurtosis) and quantile ranges (more details in statistical features below).
3. Processing of the geodetic data. We used GPS data from the Western Canada Deformation Array, preprocessed by the United States Geological Survey<sup>28</sup>. This geodetic data comes from calculating the daily GPS positions in a fixed North America reference frame that uses the stable interior of the North American continent as a reference. No detrending was applied, because as we consider GPS displacement rates, our processing removes the trend (detrended GPS data can also be used with similar results). We considered the total horizontal displacement (east/west + north/south GPS components), such that a negative displacement is in the direction of the subduction (north-east), and a positive displacement is in the direction of the slip (southwest). The raw GPS data are noisy and, as a target (label) for our machine-learning regression, we used the denoised GPS displacement rate, as calculated by the moving average of the slope of a least-squares linear regression over the time window  $n_{\text{GPS}}$  considered. We relied on scikit-learn<sup>37</sup> to compute the slope coefficients of a linear fit within each of these windows. We then used a moving average to further smooth these coefficients. The length of the time windows used for the linear regression and the moving average were the same (for instance, 60 d for Fig. 1).
4. Moving window processing of the seismic features. The target of our regression is the GPS displacement rate smoothed over  $n_{\text{GPS}}$  days, and we smoothed the features of the seismic data as well, over  $n_{\text{seismic}}$  days. For the main results (Figs. 2 and 3a) we used  $n_{\text{GPS}} = n_{\text{seismic}} = 60$  d. We also investigated the resolution of the seismic data compared to the GPS data by applying shorter time windows for the features of the seismic data ( $n_{\text{GPS}} = 60$  d and  $n_{\text{seismic}} = 1$  hour for Fig. 3b).
5. Subwindows of seismic features. In addition to the statistical features smoothed over  $n_{\text{seismic}}$  days, as described above, we computed the evolution of these features within the window. To do so, we computed the differences between the average values of the features during the last half of each window and the first half of each window, that is, the difference between the averages of the features during the last  $n_{\text{seismic}}/2$  d and the averages of the features during the first  $n_{\text{seismic}}/2$  d of the window  $n_{\text{seismic}}$  days long.
6. Overlapping time windows. We scanned the data in increments of 1 d as we built features and the corresponding labels over  $n_{\text{seismic}}$  and  $n_{\text{GPS}}$  days and then added the resulting list of labelled features to the database at each time increment.
7. Creation of a machine-learning-friendly database. This step puts the data in a format that will actually be used by the machine-learning models. Each line of the database contains the features that describe each time window of seismic data, and the corresponding average GPS displacement rate that takes place at the same time ( $n_{\text{GPS}} = n_{\text{seismic}}$ ) or during a window of time that spans the window of seismic data or larger ( $n_{\text{GPS}} > n_{\text{seismic}}$ ) in which case the first day of each window coincides). Each line  $i$  is therefore a list  $\{x_{1,1}^i, x_{1,2}^i, \dots, x_{1,D}^i, x_{2,1}^i, x_{2,2}^i, \dots, x_{2,D}^i, y^i\}$ , with  $x_{j,1}^i$  the  $j$ th feature of the  $i$ th time window of seismic data,  $x_{j,2}^i$  the evolution of the  $j$ th feature between the two halves of the  $i$ th time window of seismic data,  $D$  the total number of features and  $y^i$  the average displacement rate during this time window.
8. Train-test split. The database built as described above is a time series of features of the seismic data and the corresponding average GPS displacement rate. As such, the train-test split must be of two contiguous pieces, owing to the autocorrelation of the system: one line of our database (that is, the features of a time window of acoustic data and corresponding label) is quite

similar to the following window, especially given the windows overlap if  $n_{\text{seismic}} > 1$ . Therefore, a random train-test split is not appropriate, and instead a split in two contiguous sets must be done to ensure independence between the training and testing set. We used the first 4 yr of the seismic features and GPS data as the training set (from 2005 to 2008), and the following 8 yr (from 2009 to 2017) as the testing set.

9. Tuning the hyperparameters of the model. Before actually creating a model that relates features of the seismic data to the corresponding GPS displacement rate, we have to determine the space of functions that will be explored as possible models. This is done by tuning the hyperparameters of our model. These hyperparameters control how vast the explored function space will be during training, typically by setting how smooth the explored functions are. In the case of the random forest models used here, the hyperparameters determine the power of expression of the trees that constitute the model (see 'Hyperparameter optimization' below). The hyperparameters of the model are determined to maximize the performance of trial models in cross-validation: a set of hyperparameters was used to model a subset of the training data and evaluated on its performance on the rest of the training data.
10. Feature selection. We used a recursive feature elimination algorithm<sup>38</sup> to down select the features of the seismic data that will actually be fed to our machine-learning models.
11. Training of the machine-learning model. With a training database in hand and the complexity of our model determined (the hyperparameters), we can train our final model. Model training consists of obtaining the best fits of the training data given the complexity of the model. The previous step of optimizing this complexity on subsets of the training data ensures that this best model will not overfit the data. Indeed, too complex a model (prone to overfitting) would do poorly in cross-validation. Our final model is an ensemble of decision trees (scikit-learn implementation<sup>37</sup>) formed from a series of yes/no decisions based on a given list of features to arrive at a modelled (that is, predicted) label. We give more details on decision trees in the following subsection.
12. Assessing our final model. The performance of the final model was assessed using the testing set, the last 8 yr of the data. The metric used here, the Pearson correlation coefficient, measures the linear correlation between two variables. The same metric is used during the cross-validation step above and, provided the data are similar in training and testing (follow the same distribution), the performance in cross-validation is a good proxy for the performance in testing. A different model is made for each GPS station considered; we show the results of estimating PGC5 in the main text, and the results of estimating BAMF and NANO in the Supplementary Information.
13. Feature importance. Once we have an accurate final model, we can look for the best features identified by our model to try to understand how the model reached its estimations (see 'Feature importance' below).

**Statistical features of the seismic data used as inputs to the machine-learning models.** For each 1 Hz frequency band from 8 Hz to 13 Hz, from each station and for each day, we computed:

- The range of the data—the maximum of the velocity recorded by the seismic station minus the minimum of the velocity recorded by the seismic station.
- The root mean square of the seismic data.
- Non-centred moments (second, third and fourth moments) that characterize the distribution of the velocity recorded by the seismic station. The  $n$ th non-centred moment of a density function  $f$  is given by  $\int (t)^n f(t) dt$ .
- Centred moments (second, third and fourth moments) of variance, skewness and kurtosis. The  $n$ th centred moment of a density function  $f$  is given by  $\int (t - \mu)^n f(t) dt$ , with  $\mu$  the mean of the data.
- Quantile ranges of the velocity recorded by the seismic station (range between the 1st percentile and the 99th, 5th to 95th, 10th to 90th, 25th to 75th and 40th to 60th). These quantile ranges are closely related to the standard deviation of the velocity recorded by the seismic station, but exclude outliers and are therefore more robust to noise.
- Evolution of the features. All the previous features were calculated over portions of 1 d of seismic data, and smoothed over  $n_{\text{seismic}}$  days. We add as features the evolution of the above features during this window of  $n_{\text{seismic}}$  days by adding, for each feature, the value smoothed over the last  $n_{\text{seismic}}/2$  d of the window minus the value smoothed over the first  $n_{\text{seismic}}/2$  d of the window.

This results in  $13 \times 5 \times 6 \times 2 = 780$  features (13 features over 5 frequency bands over 6 stations, and their evolution within a time window) for each line of our database. Each line is labelled with the GPS displacement rate, calculated as described above.

**Random forests: ensemble of decision trees.** A number of machine-learning algorithms are suitable for modelling the relationship between an input data (here features derived from the continuous seismic signal) and the corresponding output label (here the GPS displacement rate during the same time window or contemporaneous window). Here we report our results using random forests<sup>33</sup>, an ensemble of weak learners (decision trees) as they achieved top performance

and retain an explicit link between the machine-learning model and the physics of the system. A random forest is built stochastically because, as we explain below, each tree is purposely limited in fitting power by only having access to part of the training data and part of the feature space. First, we briefly explain regression trees<sup>39</sup>, the weak learners that constitute the building blocks of random forest regression models. A decision tree is created by sequentially and stochastically creating nodes that best partition the data. To generate each decision (each node in the tree), the data that arrives at a node is split into two subsets (the right and left branches). This split  $s$  is made to partition the data available at the current node  $k$  into two subsets that are maximally different from each other with respect to the label of the regression (here the GPS displacement rate). Building a split corresponds to choosing a feature  $x_j$  and a threshold  $c$  that best partition the data. Then, the left subset comprises data for which  $x_j \leq c$  and the right subset comprises data for which  $x_j > c$ . A split at a node is decided by the maximum reduction in empirical variance between the data at the current node, and the two subsets partitioned by the split:

$$\Delta \text{Var}(s, k) = \text{Var}(S_k) - \frac{N_{k,L}}{N_k} \text{Var}(S_{k,L}) - \frac{N_{k,R}}{N_k} \text{Var}(S_{k,R}),$$

where  $k$  is the current node of the tree,  $S_k$  and  $N_k$  are, respectively, the labels and the number of data points at this current node,  $N_{k,L}$  and  $N_{k,R}$  are the number of data points in the left and right subsets generated by the split and  $S_{k,L}$  and  $S_{k,R}$  are their corresponding labels. The split chosen at node  $k$  is the one that maximizes this variance reduction. This enforces that the labels of the data within each of the two subsets generated by the split are as homogeneous as possible, and that these two subsets are as heterogeneous as possible one from another. The data partitioned according to the split generates two branches. In each branch, the process is repeated recursively to generate the entire decision tree structure. Decisions are generated until a limit is reached (for example, the number of data points at a node has reached a minimum size, or the tree has reached a maximum depth), at which point the tree constructs a leaf node. This creation of decisions is made on the training set. When a new point from the testing set falls on a leaf, following a path of decisions in the tree, it is given the average value of the label of that leaf.

A random forest is an ensemble of decision trees built by 'bootstrap aggregation', that is, by uniformly randomly sampling with replacement the training data each tree has access to when creating a node. To build a node, the set of features the algorithm has access to is also random. In other words, each tree is given a random subset of training data points with a random subset of features, and does its best to fit this partial data, according to the procedure described above. A particular decision tree that would have access to the full training data set would be prone to overfitting, but this bootstrap aggregation of weakened decision trees creates a stable and non-overfitting model, the random forest<sup>39</sup>. Given some input data point, the random forest outputs the average output of all its constituent decision trees.

**Feature importance.** We use the mean decrease in impurity (MDI) to measure the importance of a feature  $x_j$ :

$$\text{MDI}(x_j) = \frac{1}{N_T} \sum_t \sum_{s_{x_j}} \frac{N_k}{N} \Delta \text{Var}(s, k),$$

where  $k$  corresponds to the index of the current node in a decision tree,  $s_{x_j}$  to the splits made using the feature  $x_j$ , and  $t \in 1, \dots, T$  corresponds to the different decision trees. Therefore, the MDI for the feature  $x_j$  is the average drop in impurity of the data induced by this particular feature, over the  $T$  decision trees of the random forest. If this value is large, then the feature played a crucial role in partitioning the data into heterogeneous subsets, and is therefore crucial in the final model.

**Hyperparameter optimization.** Hyperparameters determine the parameter space that will be explored as possible models, that is, they control how vast the explored function space will be during training. In particular, a model with a large tree depth, able to split the data into many partitions, will be more complex than a model with a small tree depth. Random forests have the very convenient property of having few hyperparameters and of being quite insensitive to them due to their relative inability to overfit. We use a Bayesian optimization<sup>40</sup> approach to tune the hyperparameters of our random forest models: number of samples required for a split, minimum number of samples per leaf and number of trees (the first two control the depth of the trees and therefore their complexity). This Bayesian approach is not necessary to obtain good results using random forests, but is simply convenient. Models are compared by cross-validation: they are iteratively built on subsets of the training data, and evaluated on the remaining training data. They are then compared, the best model is identified (using Bayesian optimization) and its hyperparameters are selected for the final model.

## Data availability

All the data used are publicly available and can be found online. The seismic data are from the CNSN<sup>27</sup> ([www.earthquakescanada.nrcan.gc.ca](http://www.earthquakescanada.nrcan.gc.ca)), and the GPS data are from the Western Canada Deformation Array operated by the Geological Survey of Canada, preprocessed by the United States Geological Survey<sup>28</sup> ([https://earthquake.usgs.gov/monitoring/gps/Pacific\\_Northwest](https://earthquake.usgs.gov/monitoring/gps/Pacific_Northwest), NA-fixed trended data). The work flow described in Methods uses open source software (python and python packages including scikit-learn<sup>37</sup> and obspy<sup>36</sup>). We are currently unable to make the python script associated with this paper available, but we aim to make it available in the near future. Please contact [bertrandr@lanl.gov](mailto:bertrandr@lanl.gov) for details.

## References

35. Rouet-Leduc, B. et al. Machine learning predicts laboratory earthquakes. *Geophys. Res. Lett.* **44**, 9276–9282 (2017).
38. Gregorutti, B., Michel, B. & Saint-Pierre, P. Correlation and variable importance in random forests. *Stat. Comput.* **27**, 659–678 (2017).
39. Breiman, L., Friedman, J. H., Olshen, R. A. & Stone, C. J. *Classification and Regression Trees* (CRC Press, New York, 1999).
40. Jones, D. R., Schonlau, M. & Welch, W. J. Efficient global optimization of expensive black-box functions. *J. Global Optim.* **13**, 455–492 (1998).