Predictability of large subduction earthquakes: insights from analog models and machine learning

Fabio Corbi¹,²,³, Jonathan Bedford³, Laura Sandri⁴, Francesca Funiciello², Adriano Gualandi⁵, and Matthias Rosenau³

¹Freie Universität Berlin, Berlin, Germany.
²Università Roma Tre, Dip. Scienze, Laboratory of Experimental Tectonics, Rome, Italy.
³Helmholtz Centre Potsdam - GFZ German Research Centre for Geosciences, Potsdam, Germany.
⁴Istituto Nazionale di Geofisica e Vulcanologia, Sez. Bologna, Bologna, Italy.
⁵California Institute of Technology, Pasadena, CA, USA.

Despite the growing spatio-temporal density of geophysical observations, our understanding of the megathrust earthquake cycle continues to be limited by a series of factors, in particular the short observation time compared to mega-earthquake recurrence and the partial spatial coverage of geodetic data. Here, we attempt to compensate for these natural limitations by simulating dozens of seismic cycles in a laboratory-scale analogue model of subduction. The model creates analog earthquakes of magnitude Mw 6.2–8.3, with a coefficient of variation in recurrence intervals of 0.5, similar to real subduction megathrusts. Using a digital image correlation technique, we measure coseismic and interseismic deformation – this is akin to having a dense continuous geodetic network homogeneously distributed over the whole margin. We show how, by deciphering the spatially and temporally complex surface deformation history, machine learning can predict the timing and size of analog earthquakes. Then, we investigate data characteristics that maximize the performance of a machine learning binary classifier predicting slip-events imminence. We show how this framing can be used for designing an efficient geodetic network, and defining the minimum space-time coverage requirements for analog earthquake prediction. Converting the laboratory scale to the natural scale, we found that a 70-85 km wide coastal swath gives the most important information on slip imminence and that model performance is mainly influenced by the alarm duration, with density of stations and record length playing a secondary role. Under optimal monitoring conditions, about ten seismic cycles long record is enough to predict alarm periods in good agreement with those observed.