PREDICTABILITY OF LARGE SUBDUCTION EARTHQUAKES: INSIGHTS FROM ANALOG MODELLING AND MACHINE LEARNING

Fabio Corbi[1,2,3]

and: F. Funiciello[2]; J. Bedford[3]; M. Rosenau[3]; L. Sandri[4]; A. Gualandi[5]

[1]Department of Earth Sciences, Institute of Geological Sciences, Freie Universität Berlin,, Germany, [2]Dip. Scienze, Laboratory of Experimental Tectonics, Università "Roma TRE", Rome, Italy, [3]Helmholtz Centre Potsdam – GFZ German Research Centre for Geosciences, Potsdam, Germany, [4]Istituto Nazionale di Geofisica e Vulcanologia, Bologna, Italy, [5]Department of Geology and Planetary Sciences, California Institute of Technology, Pasadena, CA, USA

DISPLAY BASED ON:

- Corbi F., Bedford J., Sandri L., Funiciello F., Gualandi A., Rosenau M. (2020) Predicting imminence of analog megathrust earthquakes with Machine Learning: Implications for monitoring subduction zones. Geophysical Research Letters, DOI: 10.1029/2019GL086615.
- Corbi F., Sandri L., Bedford J., Funiciello F., Brizzi S., Rosenau M., Lallemand S. (2019) Machine Learning can predict the timing and size of analog earthquakes. Geophysical Research Letters, DOI:10.1029/2018GL081251.

DATA AVAILABLE OPEN ACCESS AT:

- Corbi, F., Sandri, L., Bedford, J., Funiciello, F., Brizzi, S., Rosenau, M., & Lallemand, S. (2019b) Supplementary material to "machine learning can predict the timing and size of analog earthquakes". GFZ Data Services. https://doi.org/10.5880/fidgeo.2018.071.

fabio.corbi3@gmail.com

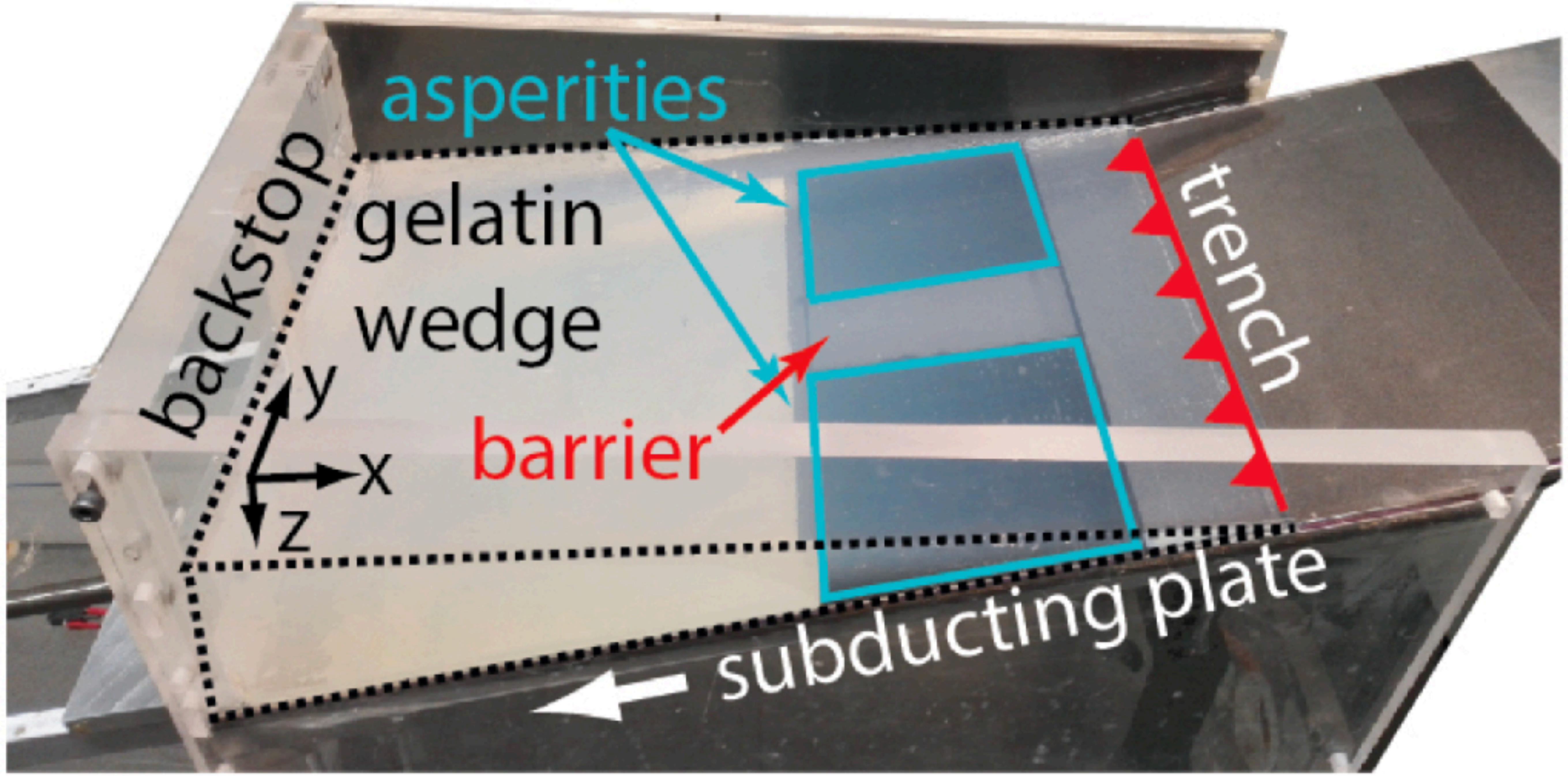
EGU 2020 DISPLAY

Motivation

Megathrust earthquakes are among the deadliest and costliest natural hazards.

Limitation

The available seismic and geodetic record is shorter than mega-events recurrence interval.



Strategy

Use physically self consistent experiments that mimic multiple subduction megathrust seismic cycles monitored with optimal geodetic-like instrumentation (PIV is equivalent to GPS) and test earthquake predictability with machine learning.

Data

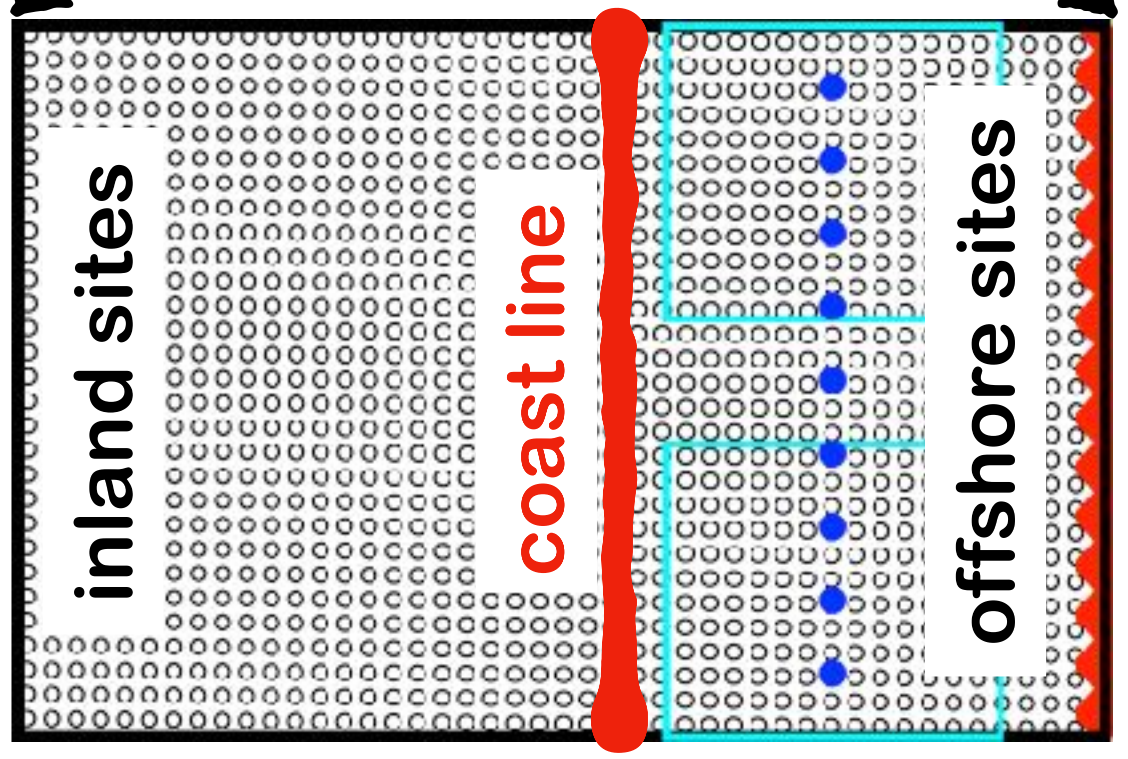
Data come from a seismotectonic analog model (photo above) that represents a subduction zone characterized by two asperities of equal size and friction. The model produces analog earthquakes equivalent to magnitude Mw 6.2–8.3 when scaled to nature, with a coefficient of variation in recurrence intervals of 0.5, similar to real subduction earthquakes. Data consist of about 7 min recording of incremental surface displacement capturing 40 seismic cycles. See Corbi et al. 2013 and Corbi et al., 2017b for more information about the model. We show the space-time rupture distribution in next 2 slides.

Machine learning procedure

Input

Extract 95 features

About 1500 synthetic GPS stations



Analog model (top view)

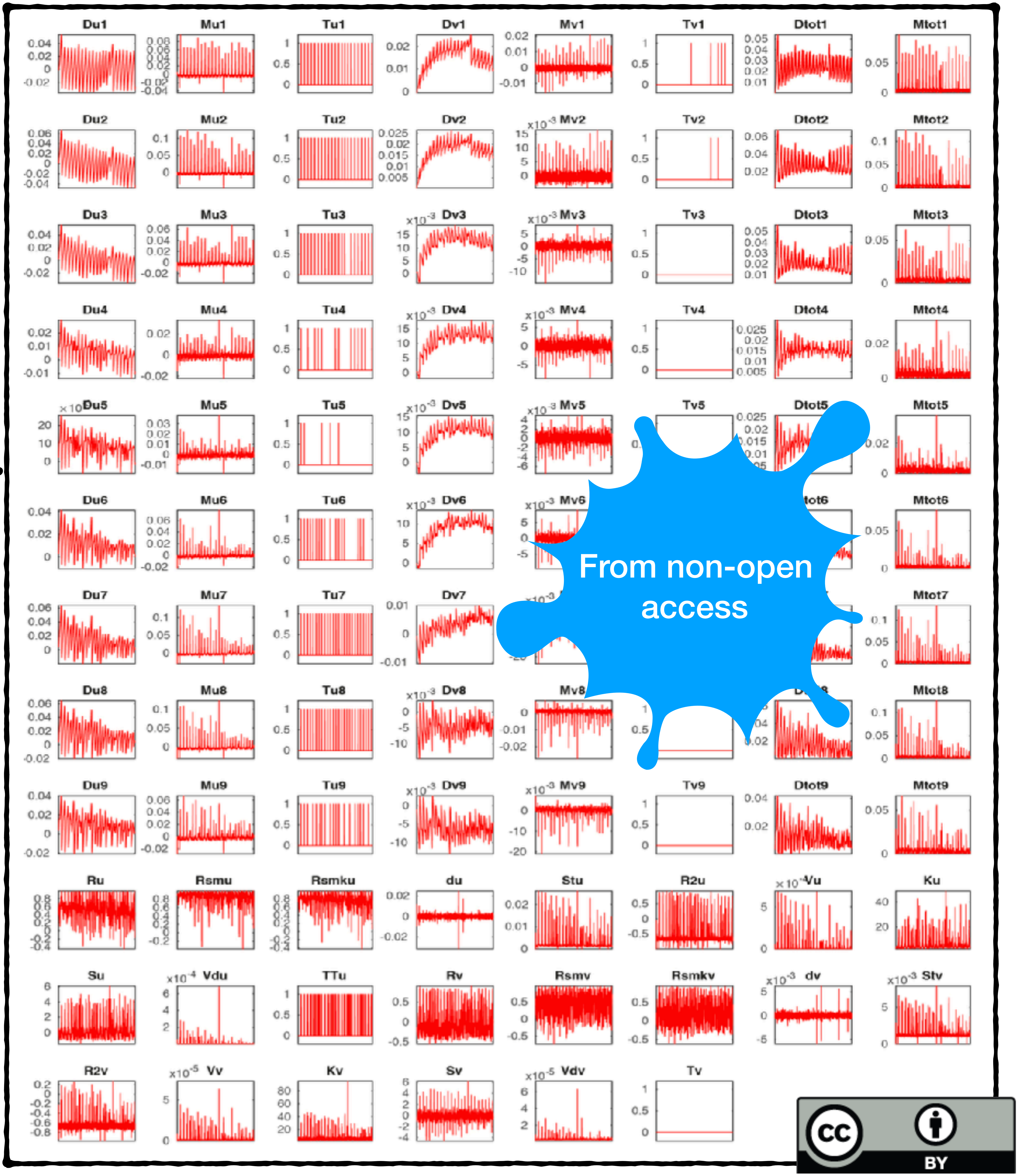
Create a model with Gradient Boosting

Feed the model with actual GPS observations

Output

Predict time to failure at various points distributed along the margin (blue points)

Training on 60% of data
Testing on 40% of data
No leakage



From Corbi et al., 2019

Machine learning can decipher the spatially and temporally complex surface deformation history and predict the timing and size of analog earthquakes.

Observed space-time rupture pattern

Bluish colors indicate that machine learning suggest an analog earthquake is impending at a given location.

What data do we need for predicting analog earthquakes?

Binary classification approach

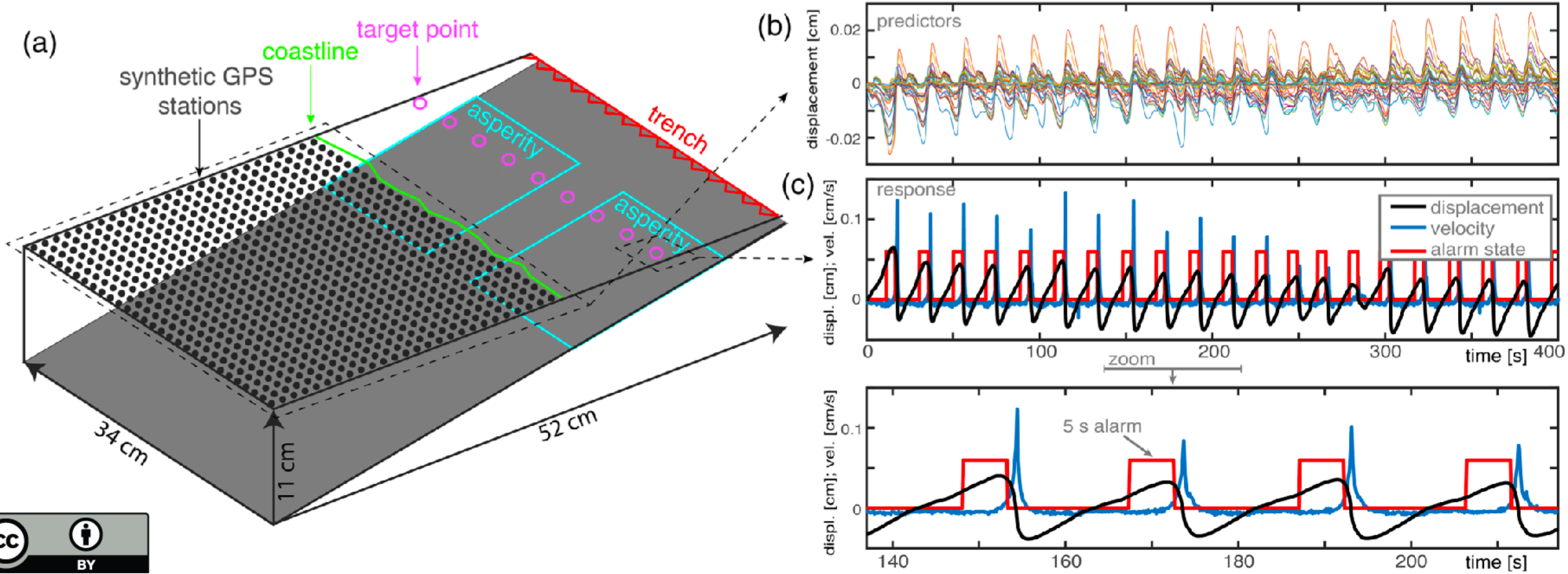
Which region of the margin is the most informative?
How important is the space-time coverage?
How in advance we can predict the slip onset?

only onshore deformation data

RUSboosting

Alarm

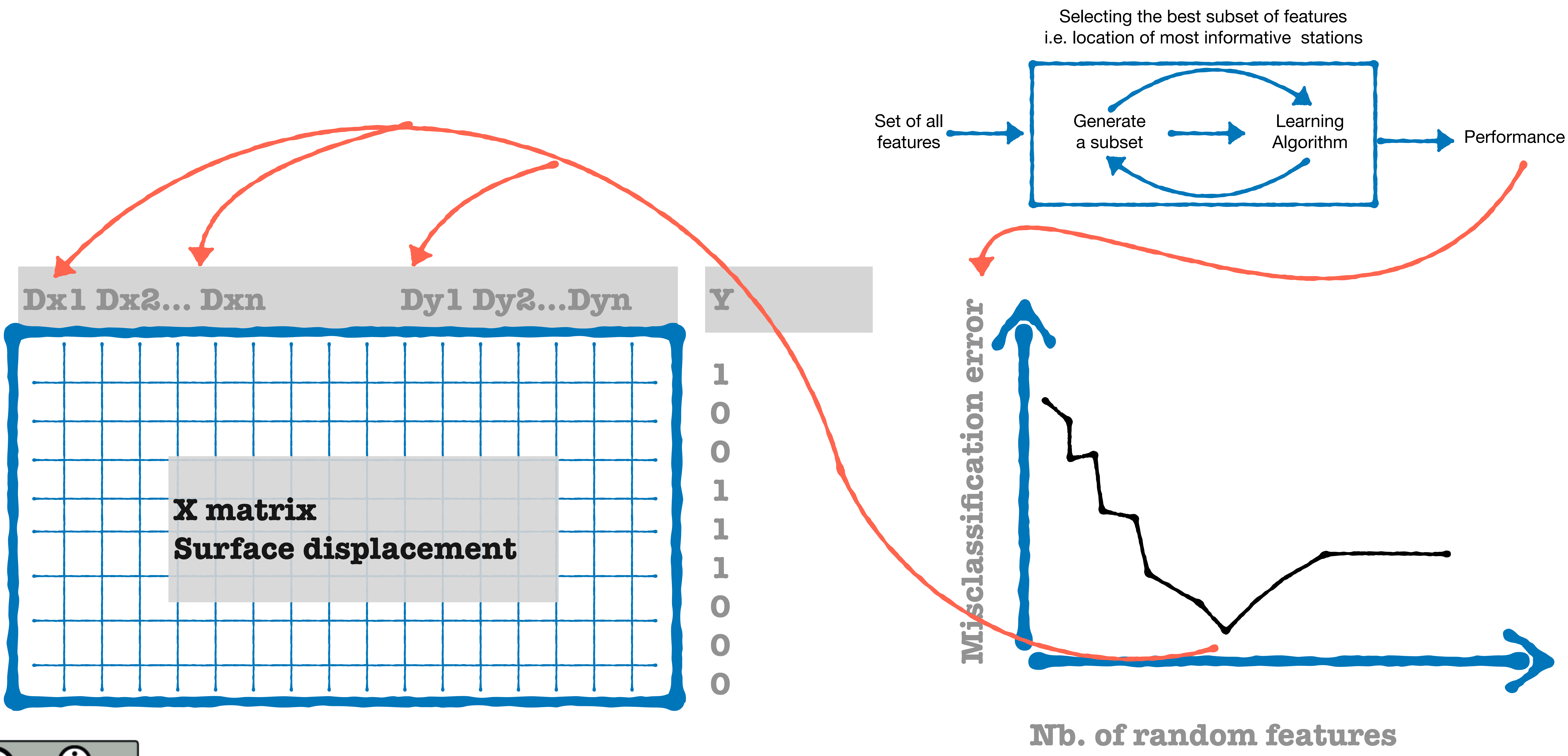
No-alarm



Corbi et al., 2020

What data do we need for predicting analog earthquakes?

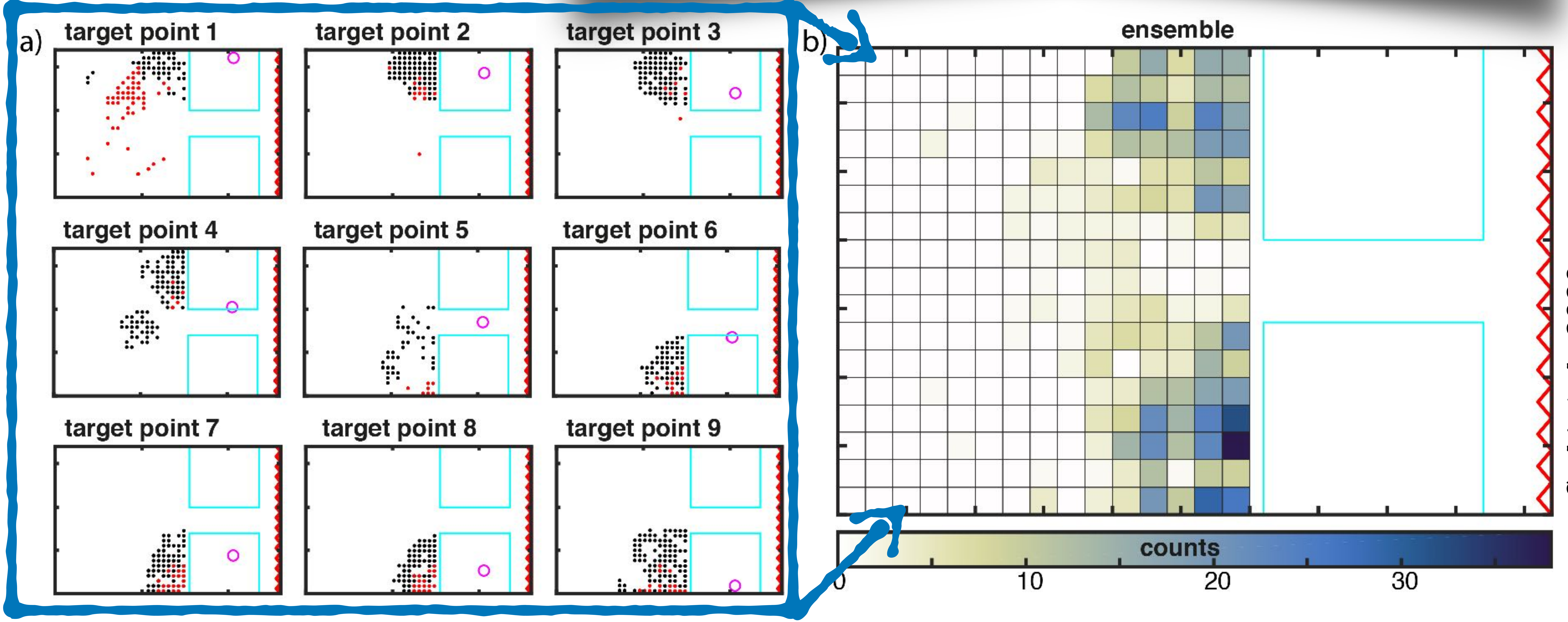
Features Selection helps identifying the most informative region



What data do we need for predicting analog earthquakes?

Features Selection helps identifying the most informative region

FS highlights a 70-80 km wide band adjacent and parallel to the coastline. Check Corbi et al. 2020 to see how this graph would look like if also offshore stations would be ideally available.

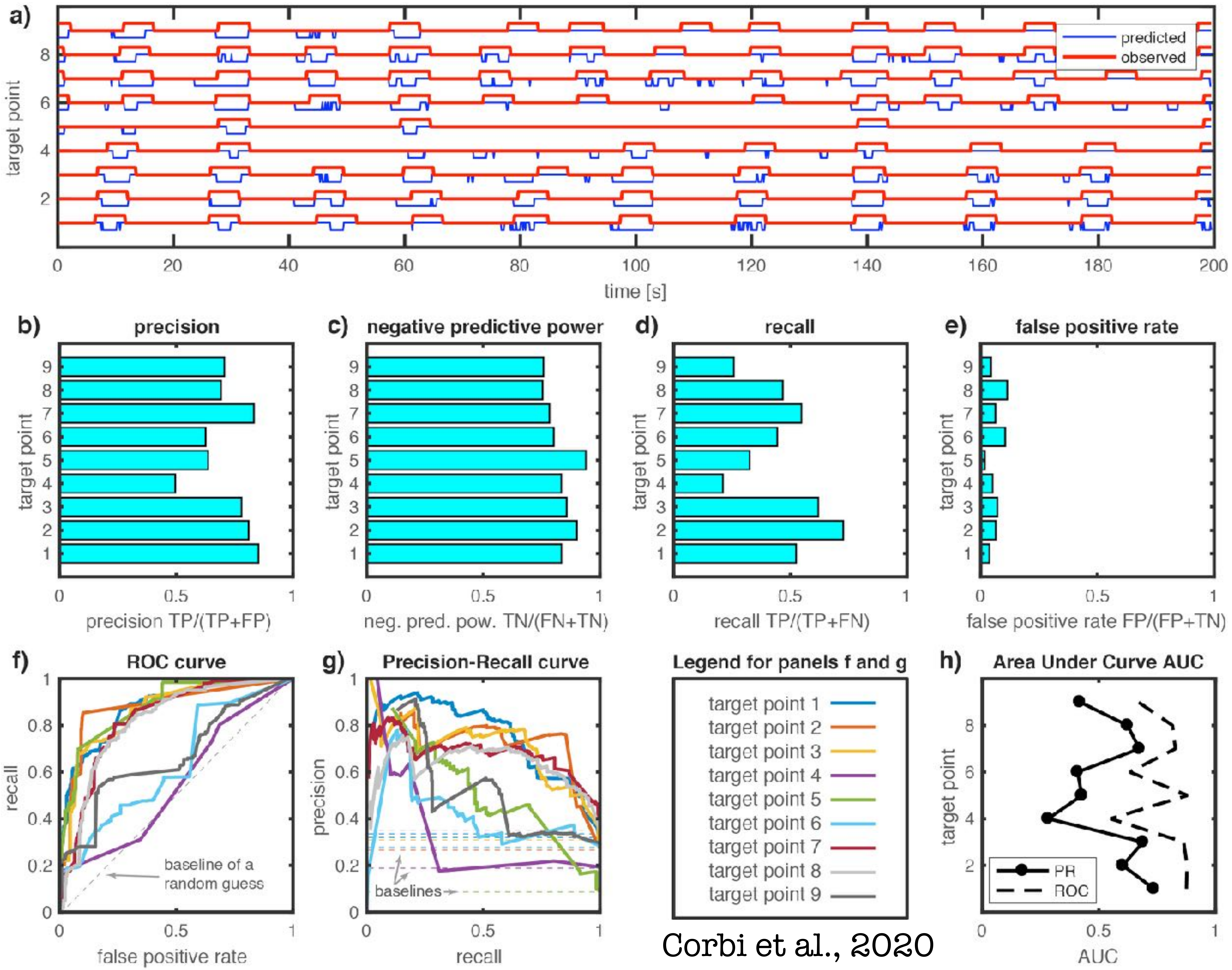


Corbi et al., 2020

Binary classification of analog earthquakes imminence

Testing

Quantification of prediction performances



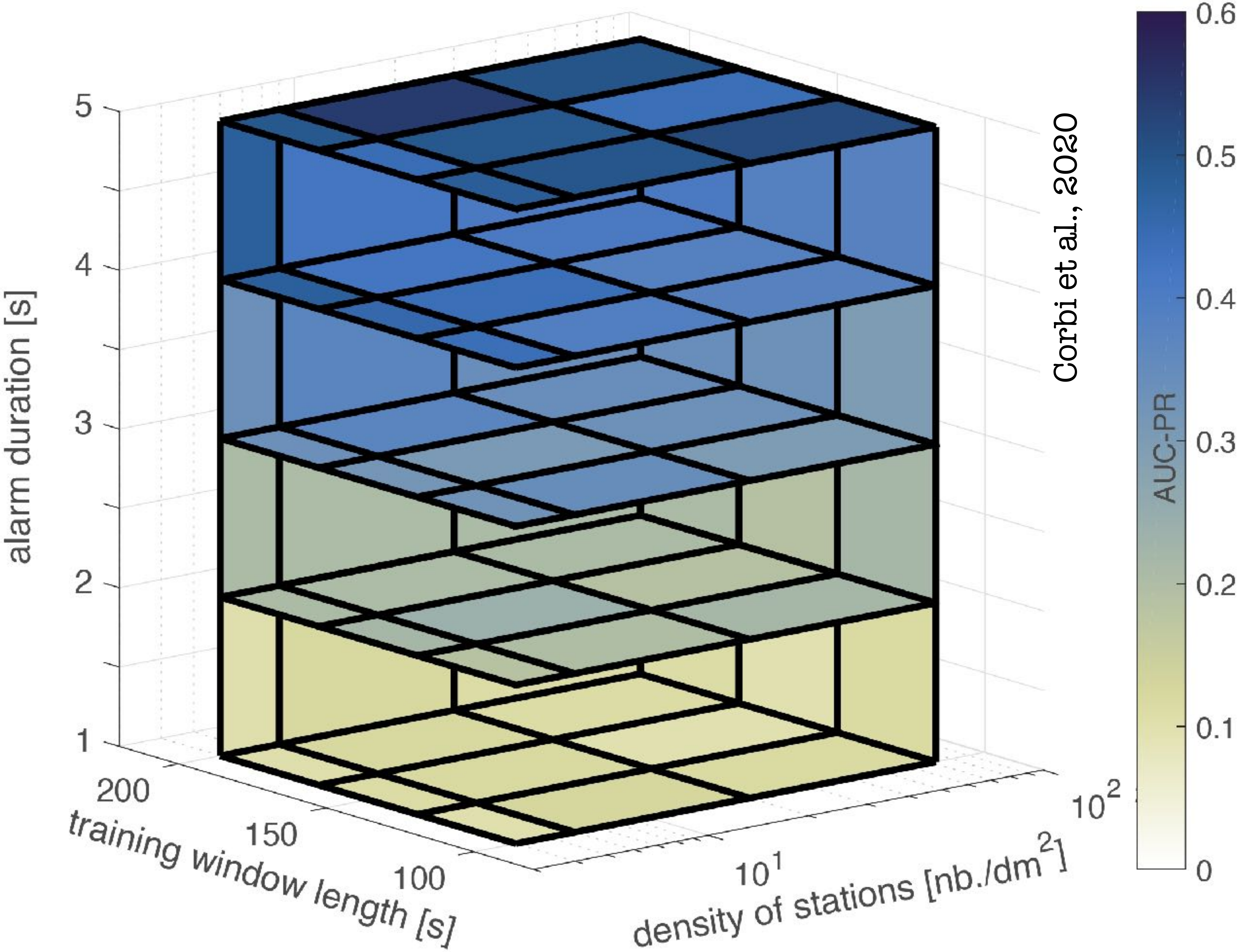
Corbi et al., 2020

ML predicted correctly the 80% of alarms (precision), and 90% of no-alarm predictions were reliable (negative predictive power).

Let's see how predictions are affected by alarm duration and space-time coverage (next slide)

		observation		
		positive (alarm)	negative (no-alarm)	
prediction	positive (alarm)	true positive TP	false positive FP	precision TP/(TP+FP)
prediction	negative (no-alarm)	false negative FN	true negative TN	negative pred. power TN/(FN+TN)
		true positive rate (recall) TP/(TP+FN)	false positive rate FP/(FP+TN)	

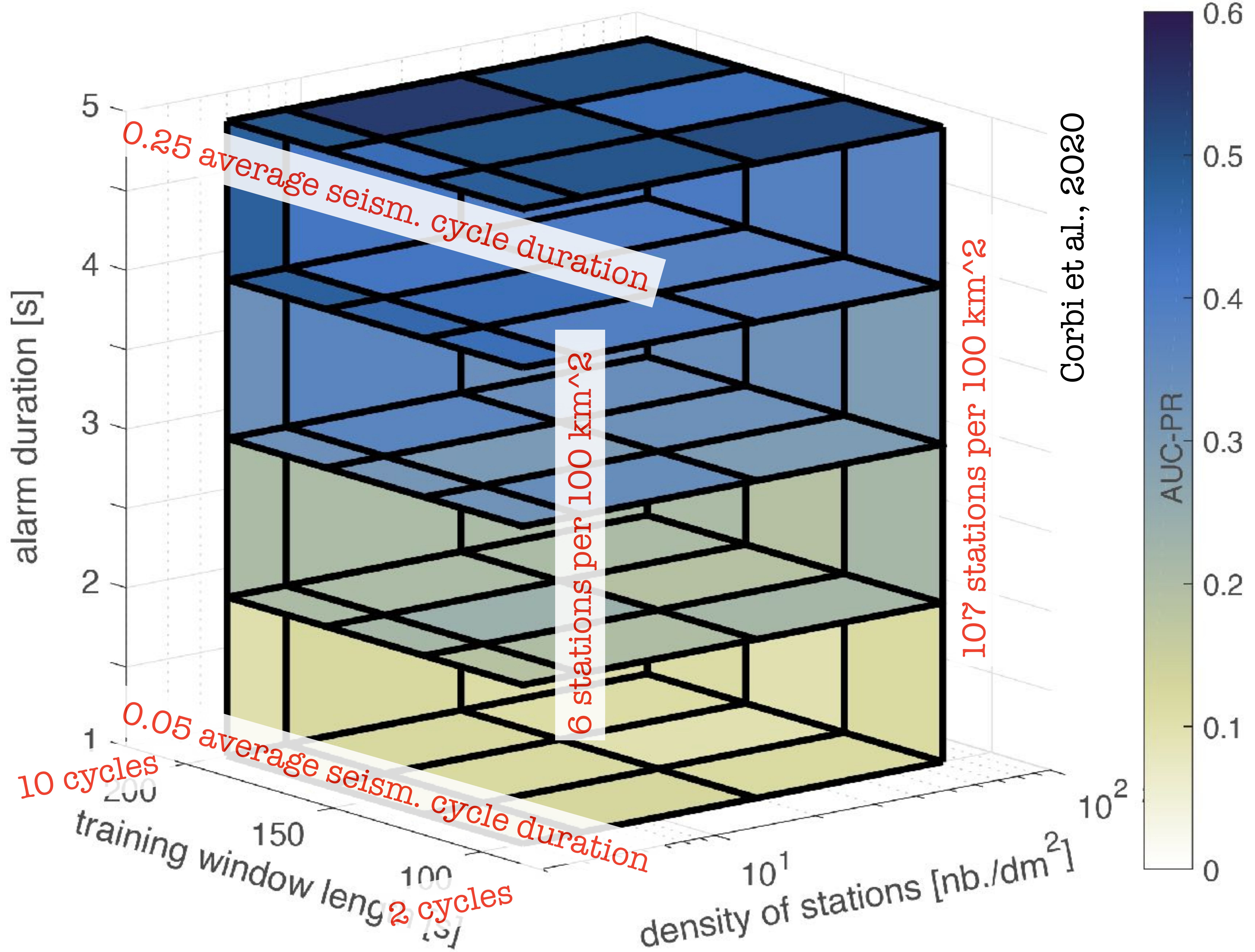
How space-time coverage and alarm duration influence predictions?



AUC-PR (area under the Precision Recall curve) is an useful metric for binary classification. For our application the AUC-PR is more informative than the ROC, being independent from the larger fraction of no-alarms of our time series.

Predictions are mainly influenced by alarm duration, with density of stations and record length playing a secondary role.

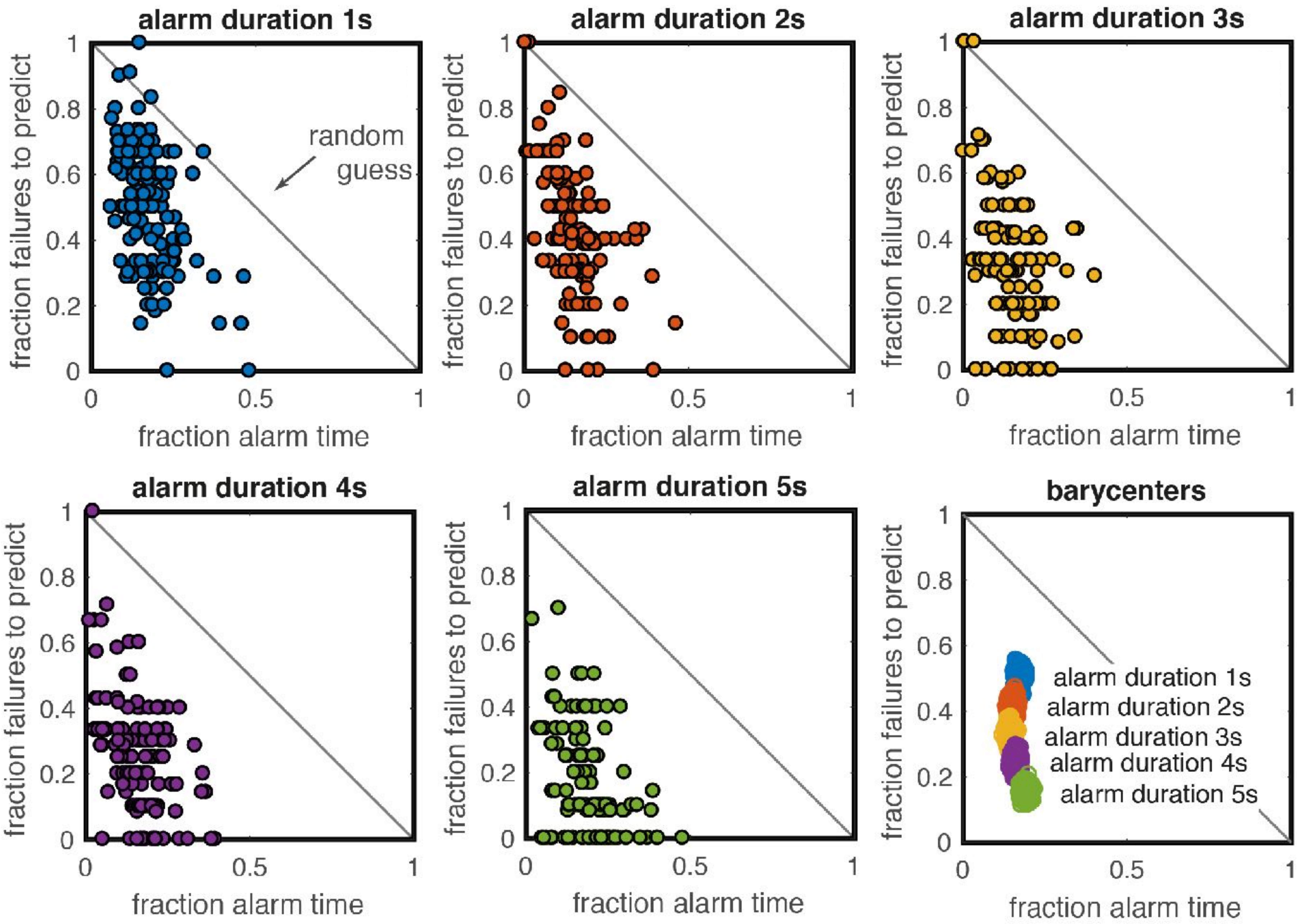
How space-time coverage and alarm duration influence predictions?



AUC-PR (area under the Precision Recall curve) is an useful metric for binary classification. For our application the AUC-PR is more informative than the ROC, being independent from the larger fraction of no-alarms of our time series.

Predictions are mainly influenced by alarm duration, with density of stations and record length playing a secondary role.

The observed improvement of predictions with alarm duration is not obvious



Corbi et al., 2020

We show that the improvement we get is more significant than what would be expected by chance using error diagrams. Each point on the graph represents the prediction at a given target point for different training window lengths and density of stations.

The downward shift of barycenters indicates that models with longer alarm durations are more precise while requiring almost the same number of declared alarms as models with short alarm durations

TAKE HOME MESSAGES

- **ML predicts the timing and size of analog earthquakes by deciphering the spatially and temporally complex surface deformation history.**
- **A 70/80 km wide band parallel to the coastline is the most important region to monitor.**
- **Length of time that we consider an event imminent plays a primary role in tuning the performances of a binary classifier predicting the imminence of analog earthquakes.**
- **A sharp, accurate binary analog earthquake prediction is unfeasible with the algorithm used in this study, even in a simplified system with a perfectly designed monitoring network. But predictions become reasonably good with observed earthquakes when tens of seismic cycles have been recorded and when the alarm duration is longer.**
- **These results can be further improved by tuning the network design and acquisition rates (paragraph 4.1 in Corbi et al 2020).**
- **Predictable slow slip events in some regions?**

THANK YOU

fabio.corbi3@gmail.com