UC San Diego
JACOBS SCHOOL OF ENGINEERING
Computer Science and Engineering

Experimental Design for Bathymetry Editing

Julaiti Alafate
jalafate@ucsd.edu

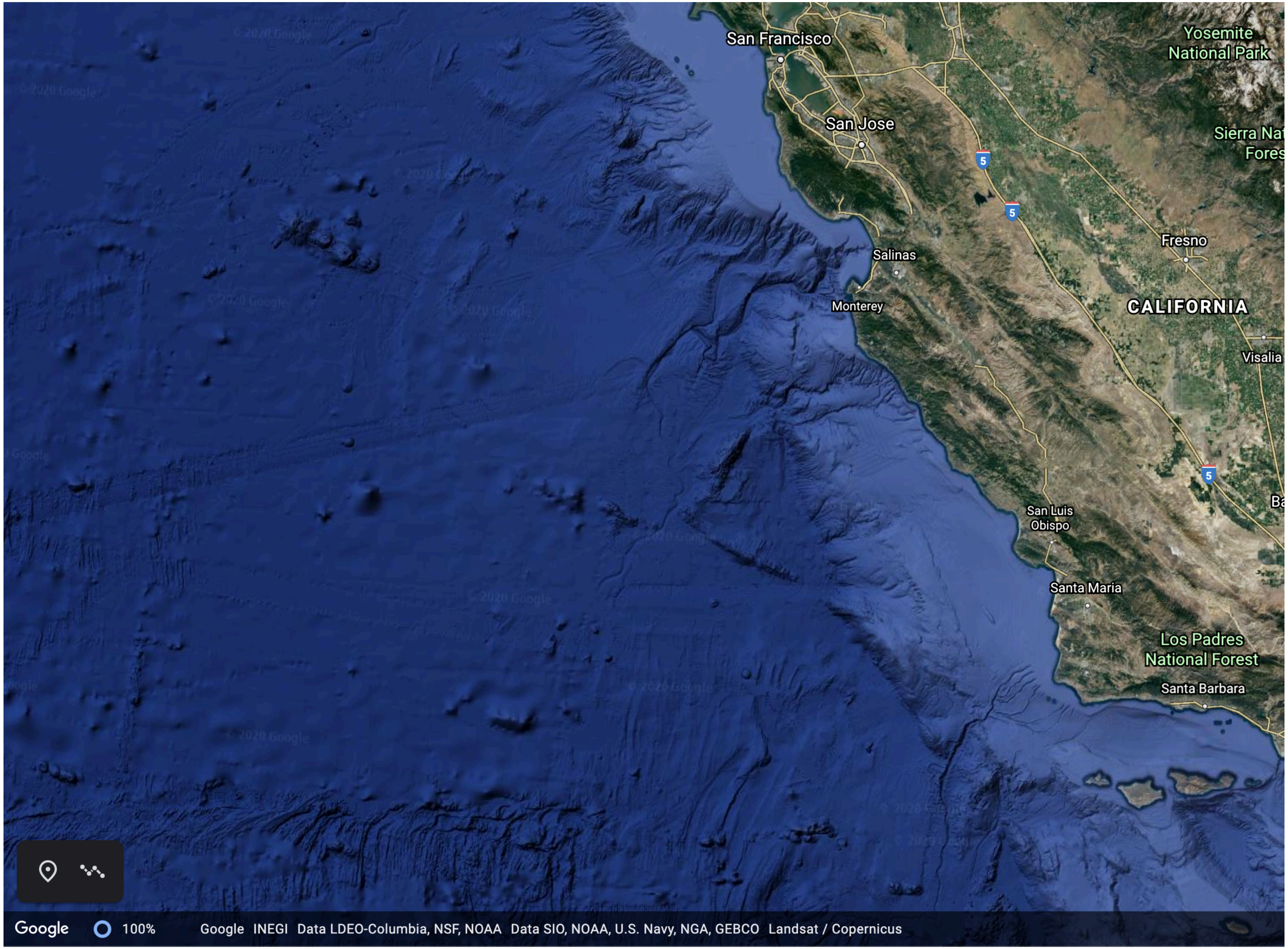
Yoav Freund
yfreund@ucsd.edu

David Sandwell
dsandwell@ucsd.edu

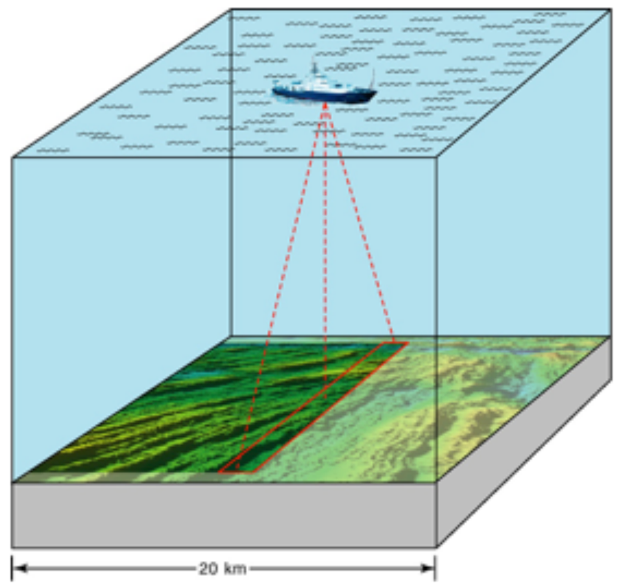
Brook Tozer
btozer@ucsd.edu

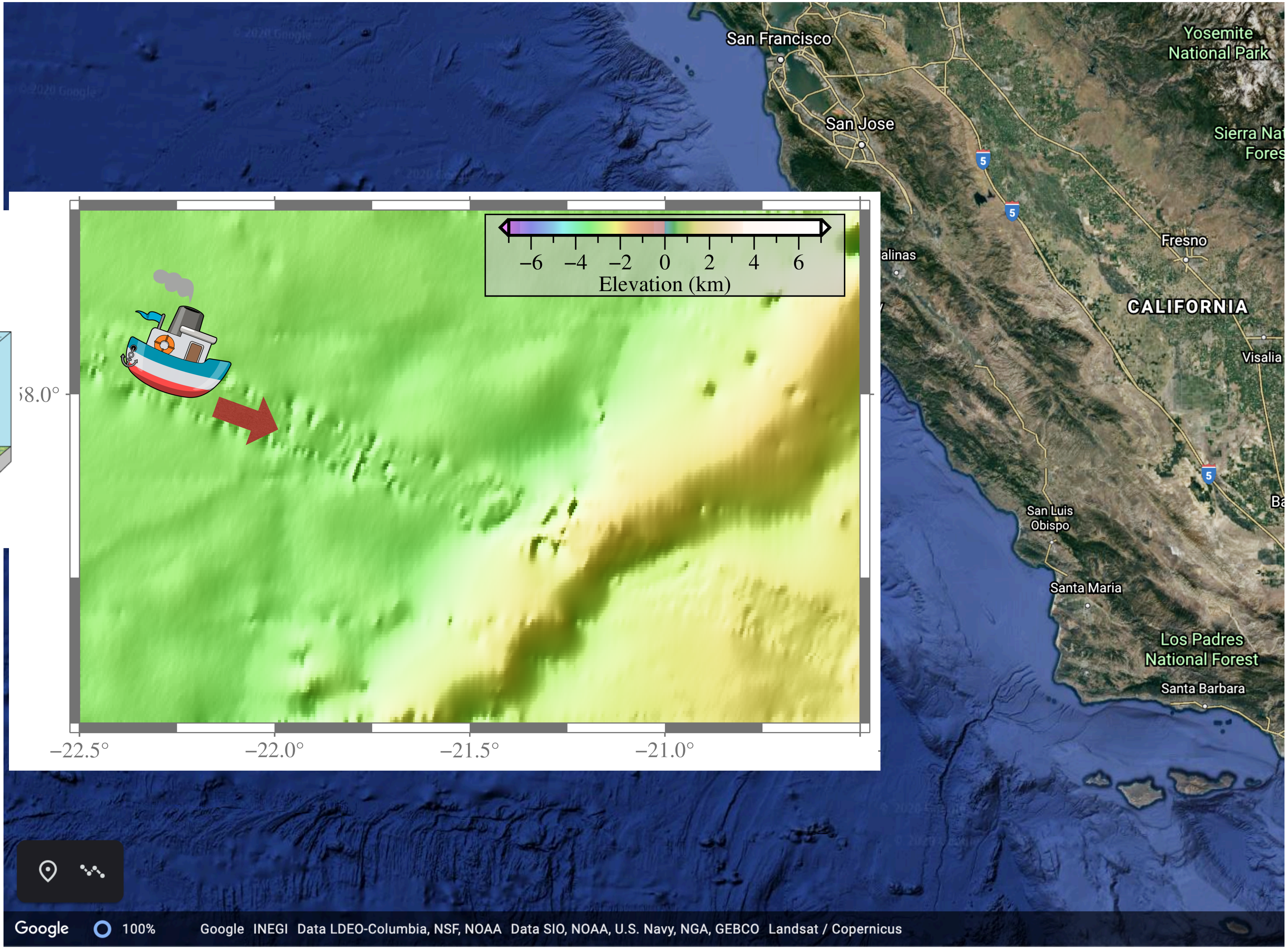
UC San Diego

Workshop on Real World Experiment Design and Active Learning at ICML 2020
July 18, 2020


San Francisco
San Jose
Yosemite National Park
Sierra Nat Fores
Fresno
Salinas
Monterey
CALIFORNIA
Visalia
San Luis Obispo
Santa Maria
Los Padres National Forest
Santa Barbara
Google
100%
Google INEGI Data LDEO-Columbia, NSF, NOAA Data SIO, NOAA, U.S. Navy, NGA, GEBCO Landsat / Copernicus

shipboard echo sounder
(high resolution ~100 m, poor coverage 11%)

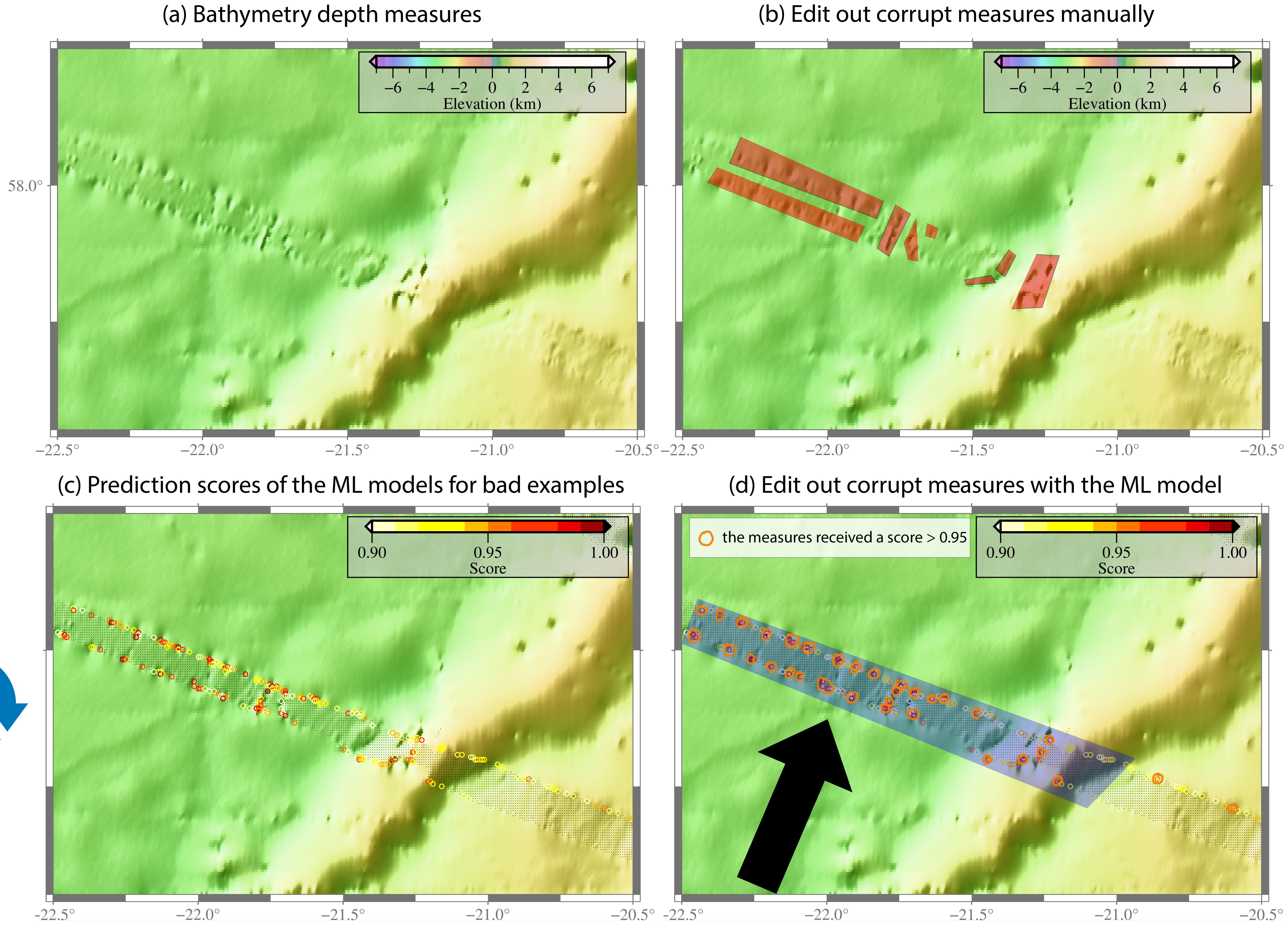




Computer-aided bathymetry data editing

Manual editing

ML-aid editing + Active learning



Summary

- Using active learning reduces the workload of the human data editors
- Other lesson we learned: real-world data is (often) *non-IID*
    - As a result, randomized train/test split leads to poor generalization
- See our paper for more details: `https://arxiv.org/abs/2007.07495`