Machine Learning and Data Mining Towards a Quantitative Assessment of Submarine Slope Failure Predictors

Jeffrey Obelcz[1*], Warren T. Wood[2]

(1) NRC Postdoctoral Fellow, Naval Research Lab, Stennis Space Center, MS, USA

(2) Naval Research Lab, Stennis Space Center, MS, USA

*Corresponding author- jbobelcz@gmail.com

CSDMS 2018

U.S. NAVAL RESEARCH LABORATORY

The National Academies of SCIENCES ENGINEERING MEDICINE

Abstract

Submarine slope failure is a ubiquitous process and dominant pathway for sediment and organic carbon flux from continental margins to the deep sea. Slope failure occurs over a wide range of temporal and spatial scales, from small (10^4-10^5 m^3/event), sub-annual failures on heavily sedimented river deltas to margin-altering and tsunamigenic (10-100 km^3/event) open slope failures occurring on glacial-interglacial timescales. Despite their importance to basic (closing the global source-to-sink sediment budget) and applied (submarine geohazards) research, submarine slope failure frequency and magnitude on most continental margins remains poorly constrained. This is primarily due to difficulty in 1) directly observing events, and 2) reconstructing age and size, particularly in the geologic record. The state of knowledge regarding submarine slope failure preconditioning and triggering factors is more qualitative than quantitative; a vague hierarchy of factor importance has been established in most settings but slope failures cannot yet be forecasted or hindcasted from *a priori* knowledge of these factors.

A new approach to address the knowledge gaps outlined above is using machine learning to quantitatively identify triggering and preconditioning factors that are most strongly correlated with submarine slope failure occurrence. This requires three general steps: 1) compile potential predictors of slope failure occurrence (gridded and interpolated at desired resolution), 2) compile predictands (specific values that we wish to predict), and 3) recursively test predictor/predictand correlation with observed data until the strongest correlations are found. Potential predictors can be parsed into categories such as morphology (gradient, curvature, roughness), geology (clay fraction, grain size, sedimentation rate, fault proximity), and triggers (seismicity, significant wave height, river discharge). Predictands (i.e. training and validation data) are various proxies for slope failure occurrence, including depth change between bathymetric surveys and sediment shear strength. The initial test sites are heavily sedimented, societally important river deltas, as they host both frequent slope failures and ample predictor/predictand measurements. Once predictors that strongly correlate with submarine slope failure occurrence are identified, this approach can be applied in more data-poor settings to further our current understanding of global submarine slope failure distribution, frequency, and magnitude.

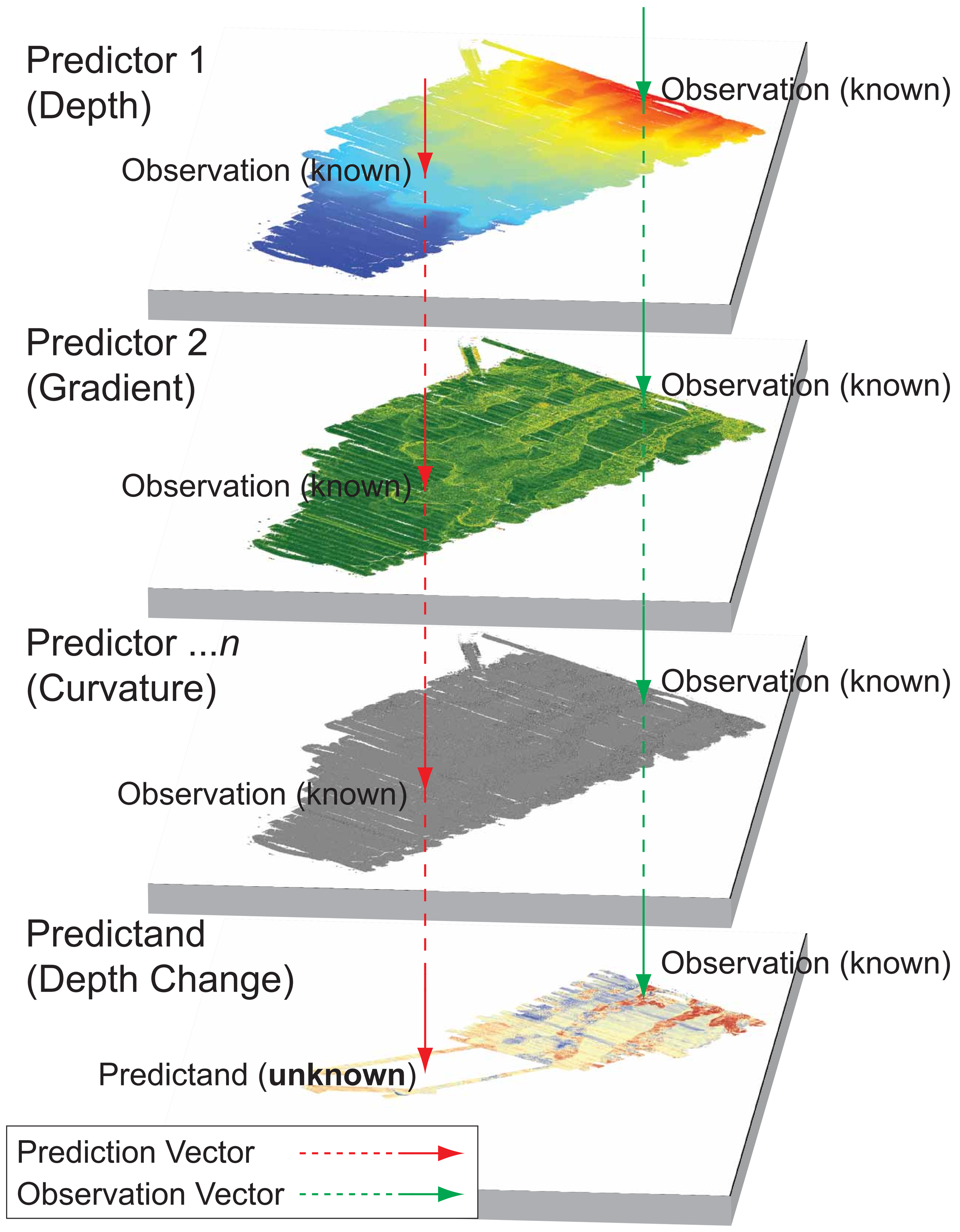
Problem

- Submarine slope failure difficult to observe, leading to difficulties forecasting; hindcasts difficult as well due to ambigious evidence of occurrence
- Triggering and preconditioning factors of slope failure qualitatively, but not quantitatively known for most settings

QUESTION: Can we quantitatively assess the importance of preconditioning/triggering factors of submarine slope failure, and use this knowledge to predict future failures?

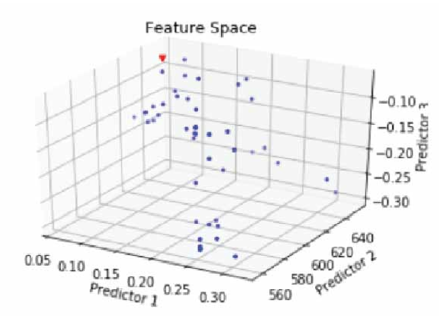
Approach: Machine Learning

Goal: Leverage relationship between **predictand** (undersampled parameter to be predicted) and **predictors** (parameters correlated with predictand) to provide "best guess" of predictands between known observations



Method: K-Nearest Neighbor (KNN), an unsupervised machine learning/data mining technique. At every grid point with a predictand and *n* predictors:

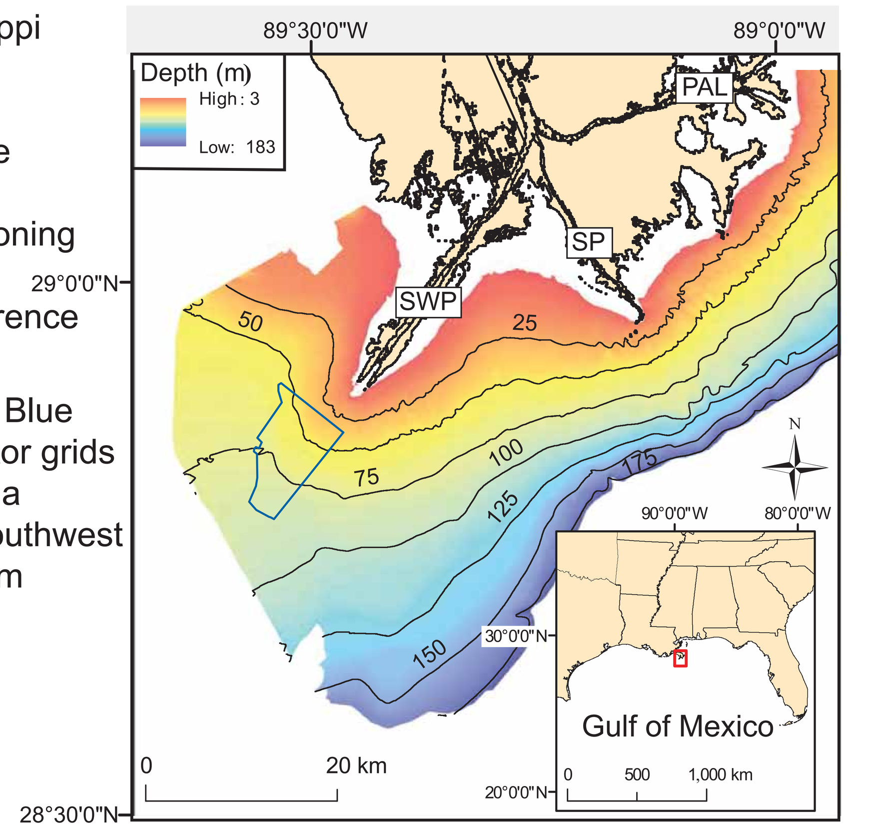
- Predictand vector is compared to all observation vectors
- The *k* (user-defined parameter) observation vectors closest in parameter space to the predictand vector are averaged and resulting value is assigned to the predictand



- This ensures predictand values are interpolated from **geologically** similar data, even if not spatially proximal, i.e. Mississippi (Gulf of Mexico) and Yellow River Deltas (South China Sea)

Test Case: Mississippi River Delta Front

- Largest distributary of Mississippi River sediment/water[1]
- Why chosen:
  - Societally important, therefore well-studied (data-rich)
  - Many slope failure preconditioning and triggering factors (predictors) and annual recurrence interval slope failures[2]



Mississippi River Delta Front. Blue outline shows location of predictor grids to left. Abbreviations: PAL-Pass a Loutre, SP-South Pass,SWP-Southwest Pass. Depth contours are in 25 m increments.

Predictors

Red = Utilized
Black = Planned

Preconditioning factors for slope failure

- Gradient (utilized)
- Curvature (utilized)
- Water depth (utilized)
- Sedimentation Rate
- Proximity to faults
- Total Organic Carbon
- Porosity[3] (utilized)
- Shear strength (utilized)

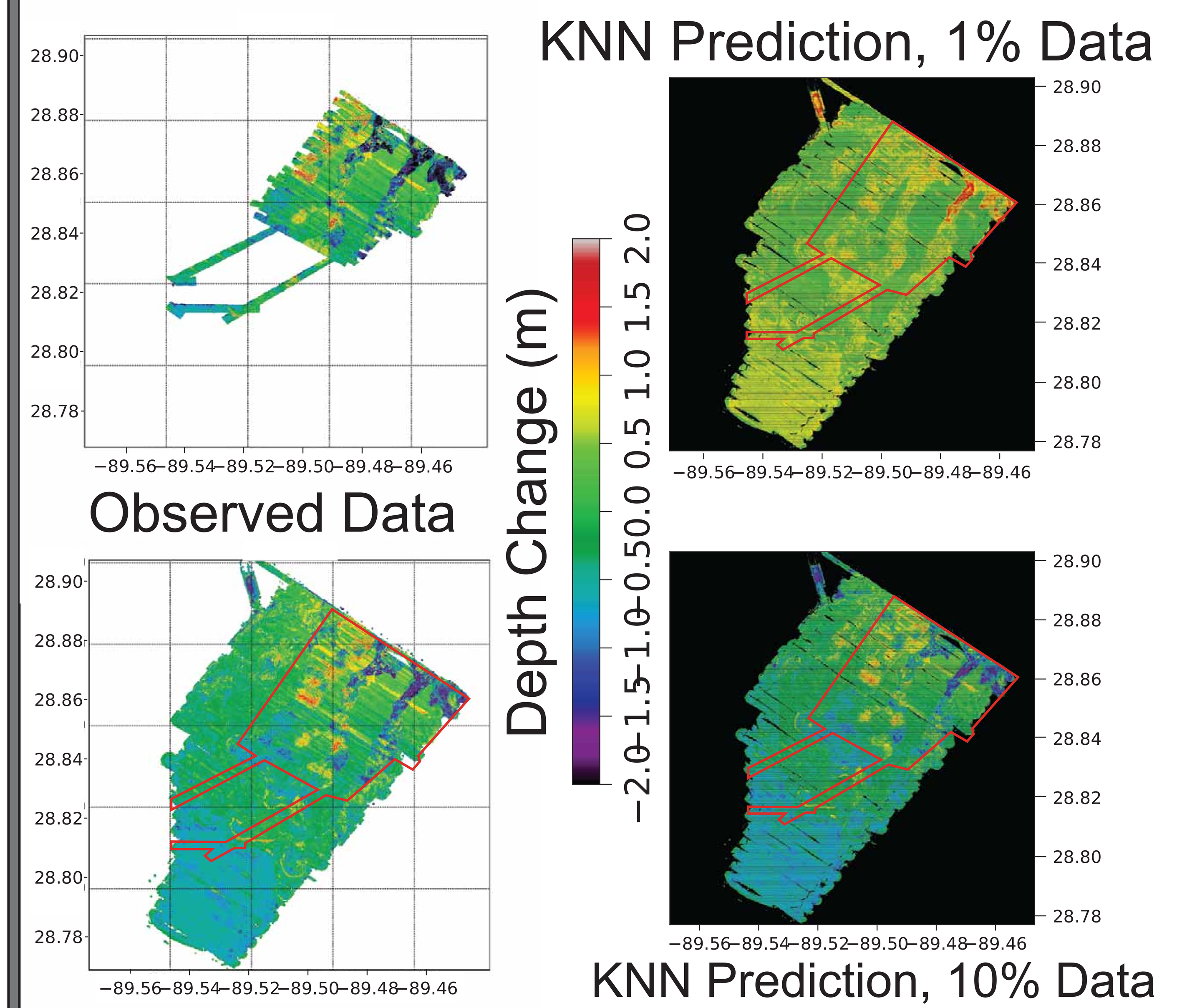
Triggering factors of slope failure

- Tropical Storm/Hurricane passage (Δ seafloor pressure)
- River Floods
- Sedimentation Rate (oversteepening)

Predictands

- Bathymetric change (utilized) (elevation change = slope failure)
- Landslide scarps/deposits (derived from bathymetry)
- Landslide deposits (identified via subbottom profiling)

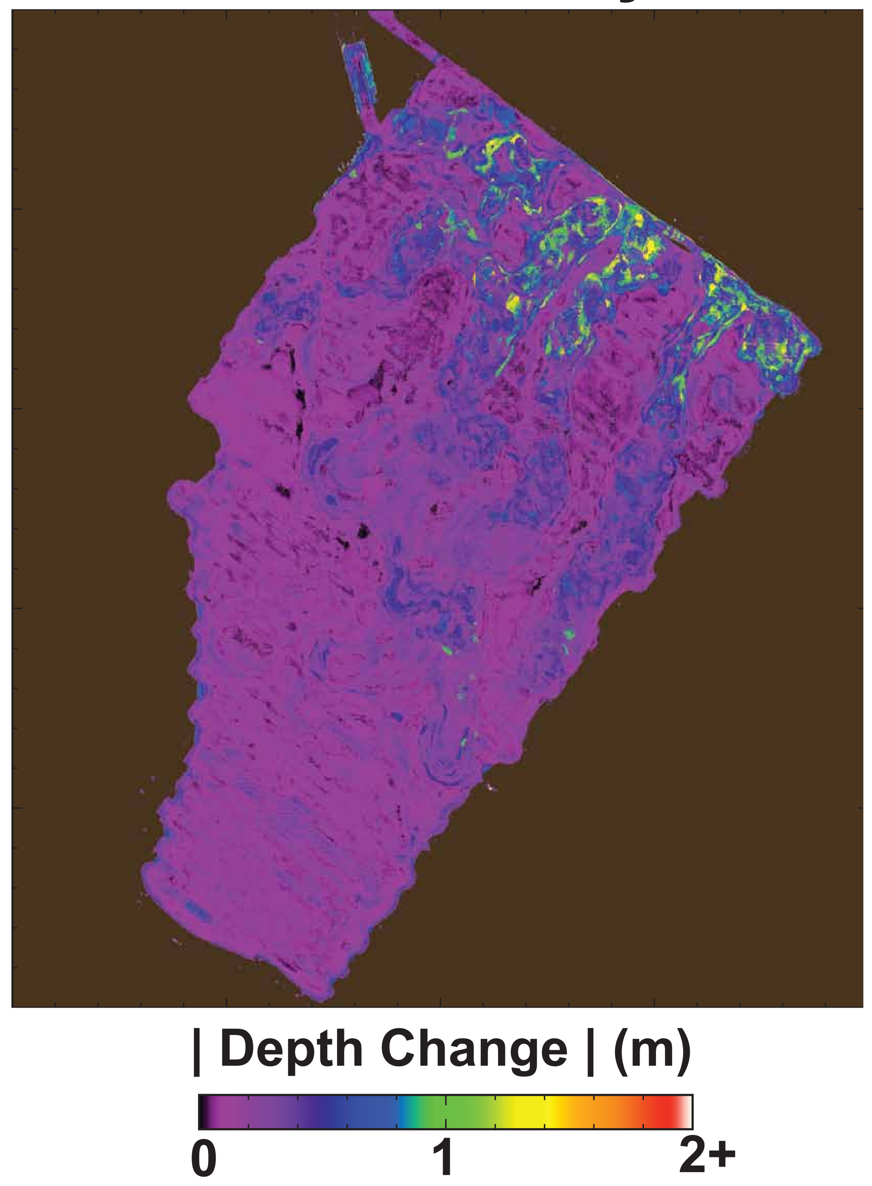
Results



KNN Prediction, 100% Data

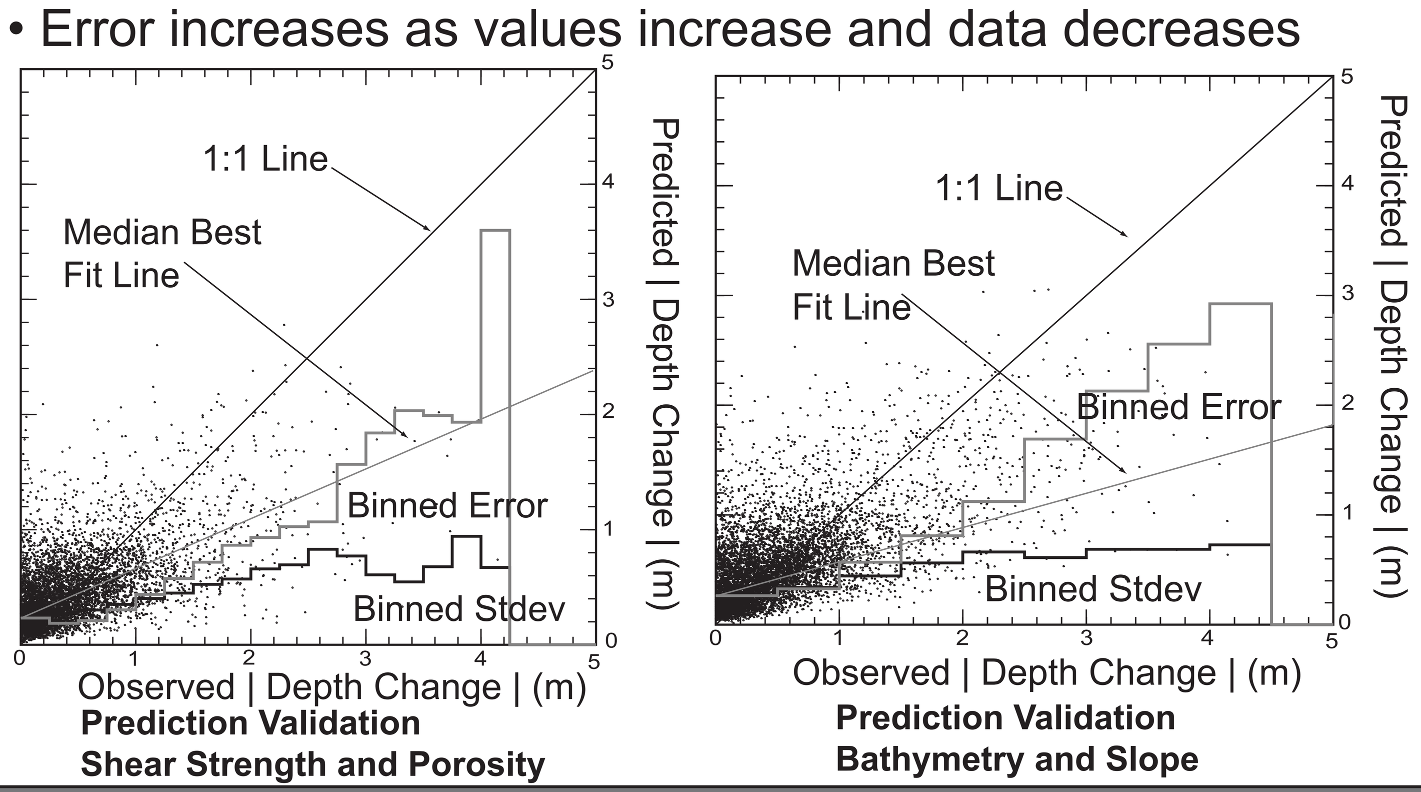
Clockwise from top left: (1) Observed depth change data, (2) Prediction using 1% of observation data (absolute value of depth change shown), (3) Prediction using 10% of observation data, (4) Prediction using 100% of observation data.

Uncertainty and Validation



- Uncertainty estimated via standard deviation of top *k* predictors (left, prediction using bathymetry and slope)
- Uncertainty highest in areas of largest depth change
- Depth change predicted outside mudflow gully/lobe zones, not seen in observed data

- Ten-fold cross validation used to test predictive capability
- Error increases as values increase and data decreases

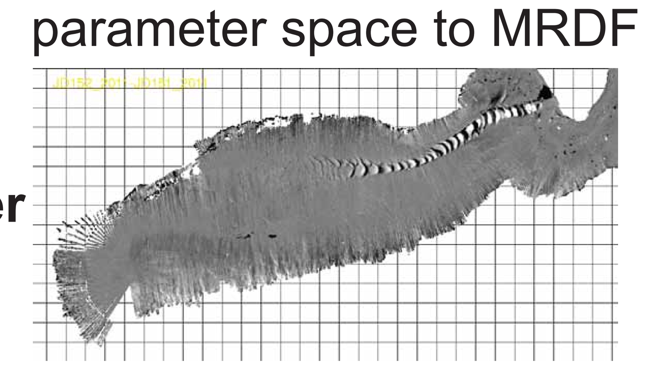


Interpretation

- KNN can be used to estimate depth change (proxy for slope failure) as "placeholder" for data
- Porosity, shear strength, and gradient showed strongest predictive capability of depth change in MRDF setting
- KNN reproduced general spatial trends of depth change (and lack thereof), but lots of mismatch in scale/geologic facies-significant room for improvement

Future Work

- Integrate more predictors into KNN algorithm (more = better)
- Find better predictand for submarine slope failure, depth change is imperfect
- Integrate combinative predictors into workflow (i.e. depth + porosity)
- Test algorithm on settings assumed to be close in parameter space to MRDF



(Right) Squamish River Delta (Canada) Depth Change Map

Acknowledgements and References Cited

The authors thank J.P. Walsh and Fugro Geosciences Inc. for bathymetric data. Squamish Delta bathymetry provided courtesy of Jaime Hizzett (University of Southampton).

[1]Allison, M.A., et al., 2012. Journal of Hydrology 432-433, pp. 84-97.
[2]Obelcz, J., et al., 2017. Geology 45(8), pp. 703-706.
[3]Martin, K.M., et al., 2015. Geophysical Research Letters 42, pp. 1-7.