Ensemble machine learning accurately predicts coastal shoreline change rate

Sagar Gautam[1], Anastasia Piliouras[2], and Umakant Mishra[1]
[1]Computational Biology & Biophysics, Sandia National Laboratories, Livermore, CA, 60439
[2]Department of Geosciences, Pennsylvania State University, University Park, PA 16802

BACKGROUND

Arctic coastal environments are experiencing rapid changes due to the changing climate. However, our knowledge of the existing environmental factors that control and predict coastal soil erosion rates is limited due to the lack of comprehensive field measurements and understanding of the processes involved. Previous efforts to predict coastal erosion rates have utilized various methods, resulting in a wide range of predictions. In this study, we applied an ensemble machine learning approach with a large number of field observations and high-resolution environmental datasets to investigate the relationships between coastal erosion rate and environmental factors and predict the Arctic coastal erosion rates in North Slope of Alaska. Our results can improve our understanding of the environmental controls of coastal erosion rates and constrain the range in coastal erosion rates.

OBJECTIVES

- Identify dominant environmental controllers influencing coastal erosion rates.
- Derive empirical relationships between dominant environmental factors and erosion rates.
- Use the derived relationships to predict coastal erosion rates and compare the prediction accuracy of simpler model with the machine learning predictions.

DATA SOURCES

- Historical long-term (1950–2010) coastal erosion rates were obtained from Gibbs et al. (2017).
- Digital elevation model was obtained from the Alaska database (Alaska DGGS, 2018), and land cover map was collected from the North Slope Science Initiative (2013).
- Mean annual ground temperature was collected from Nicolsky and Romanovsky (2017), geomorphology data was obtained from Lara et al. (2018).
- Surface lithology, ecological landscape unit, and ice content data were obtained from Jorgenson et al. (2014).
- Coastal observation points (n = 11,546) were first divided into categories of exposed (33%) and sheltered (67%) as a first order control on coastal change rates.

METHODS

- Machine learning approach is a family of algorithms which do not assume any mechanistic nature to the data and instead seek to "learn" a function that best maps input parameters to an output. We used gradient boosting machine (GBM), extreme gradient boosting (XGB), and random forest (RF) machine learning approaches to predict the coastal erosion rates.
- The GBM algorithm which was proposed by Friedman (2001), uses simple regression model "weak learners" and iteratively combine this simple model to obtain "strong learner" with improved accuracy by reducing the bias and the variance. GBM model include two major user defined parameters; number of tree and tree depth.
- The XGB algorithm is based on classification and regression (Chen et al., 2015). It is the extended version of GBM, which iteratively combines the weak learners to obtain the strong learner. XGB uses a more structured model framework to control overfitting and to ensure computational efficiency.
- RF is a tree-based approach that works by building a set of regression trees and averaging the results for final prediction (Breiman, 2001). RF works on a rationale that the combination of learning models (tree-based ensemble) increases the prediction accuracy. The final result is a single prediction constructed as a weighted average over all these individually suboptimal trees.

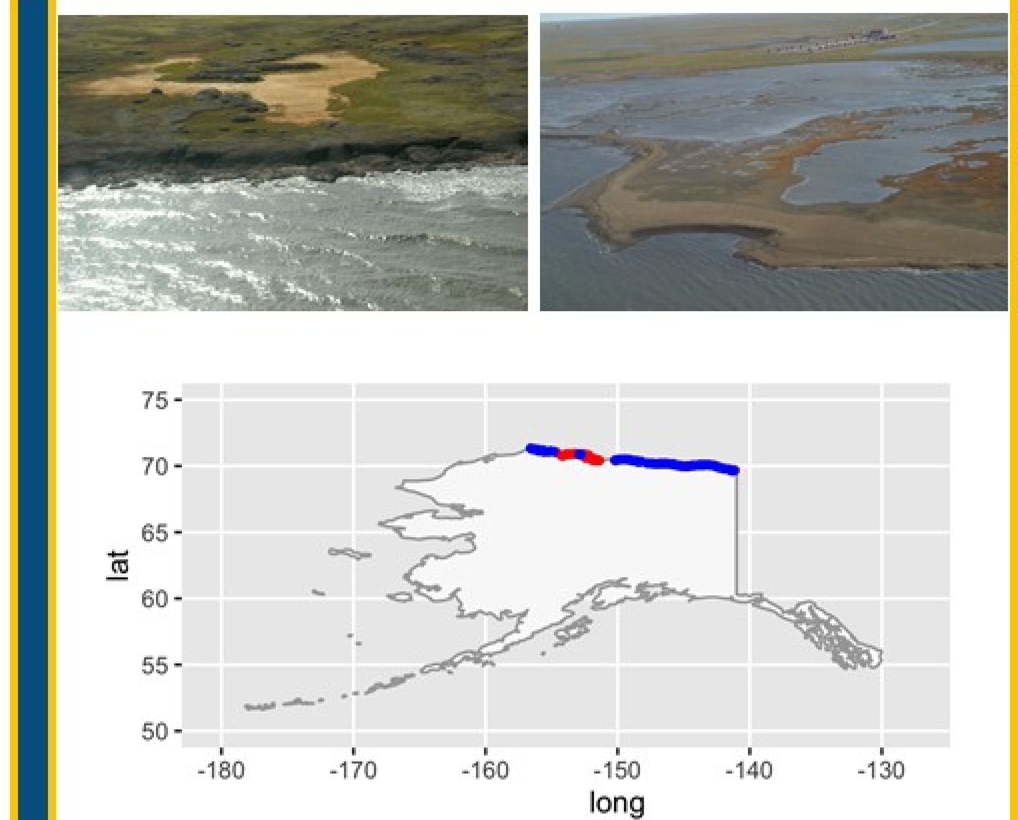
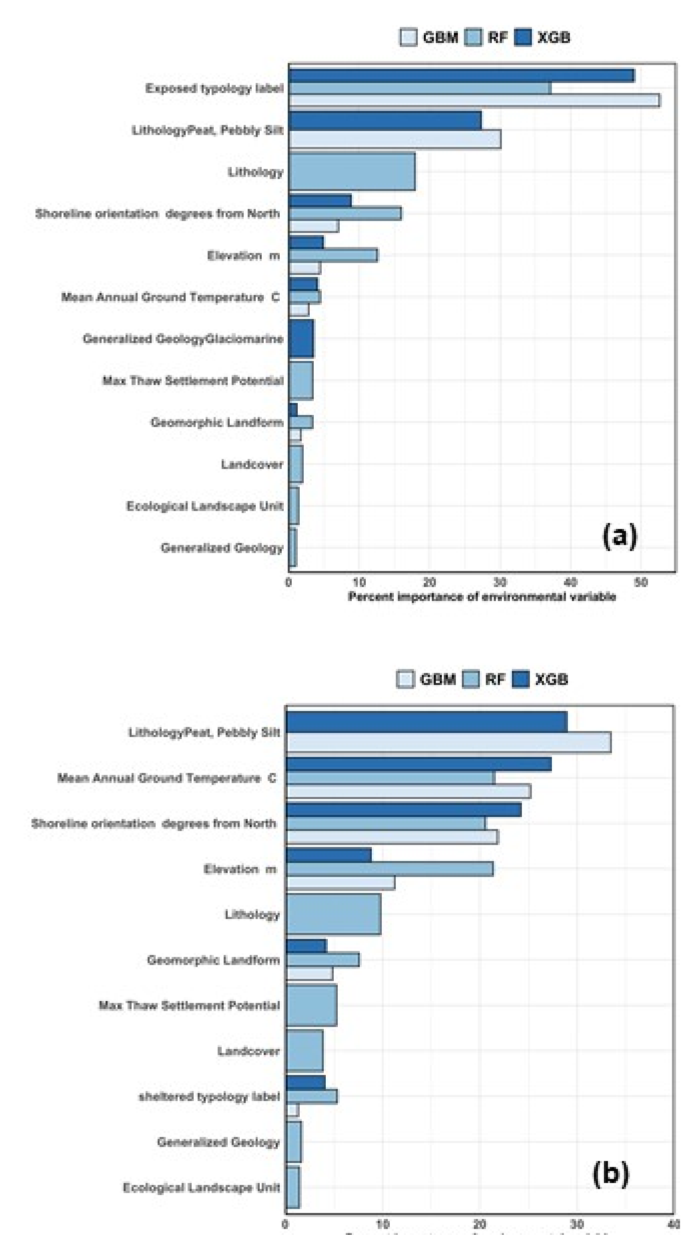


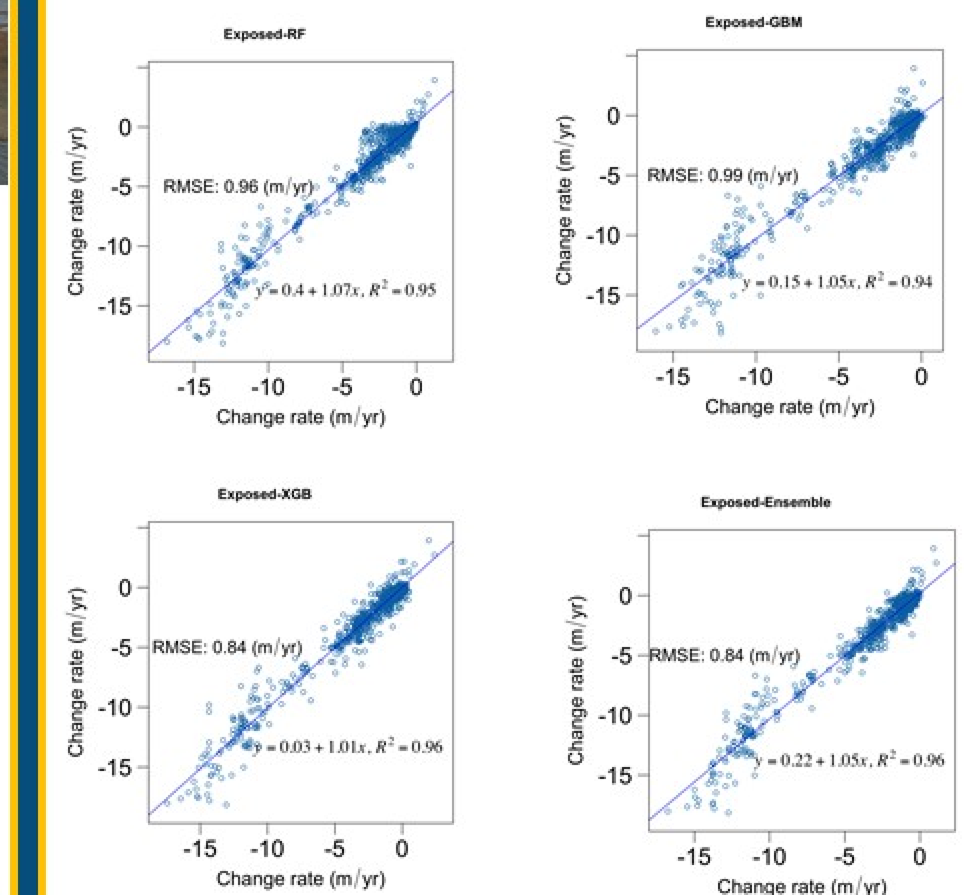
Figure: Exposed (upper left) and sheltered (upper right) types of coastal sites used in this study. Red dots show samples from exposed coastal settings and blue dots show samples from sheltered coastal settings.

RESULTS

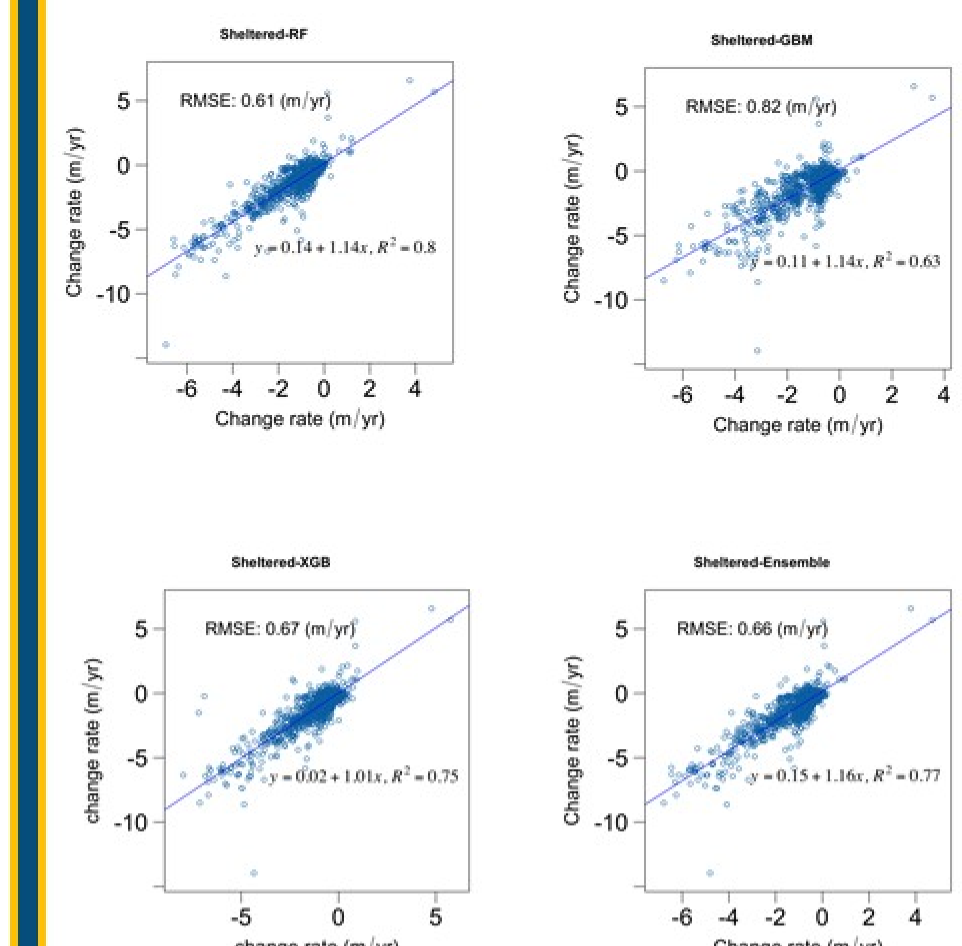
1. Importance of environmental factors in predicting coastal erosion rates at exposed (a) and sheltered (b) sites.



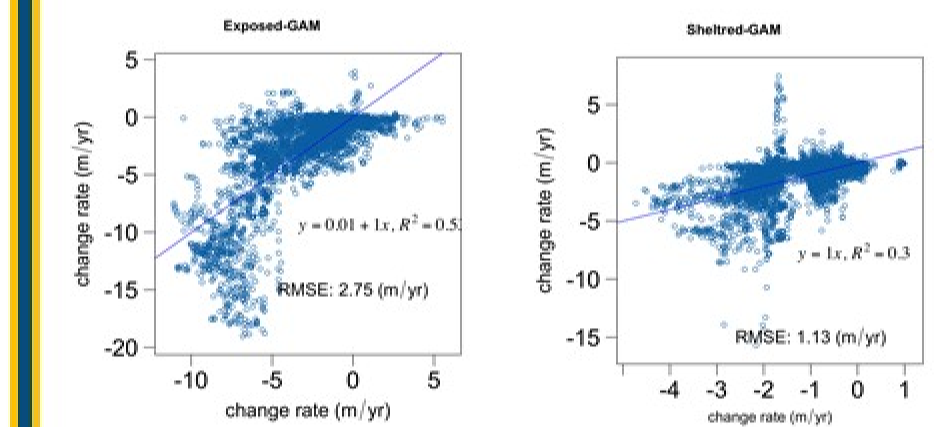
2. Prediction accuracy of different machine learning approaches at exposed sites



3. Prediction accuracy of different machine learning approaches at sheltered sites



4. Using Generalized Additive Modeling to derive non-linear equations of important environmental controllers of coastal erosion rates (ongoing analysis)



CONCLUSIONS

- Our results show that the ensemble of three machine learning approaches accurately predicted coastal erosion rates (R2 = 0.77-0.96).
- Among the 15 environmental factors considered, typology type, lithology, aspect, elevation, and temperature emerged as dominant predictors of coastal erosion rates.
- Machine learning models performed better in predicting erosion rates at exposed shoreline points compared to sheltered areas.
- In next steps, we will derive non-linear equations describing relationships between environmental factors and the coastal erosion rates, and compare the prediction accuracy obtained from these equations with machine learning approach.
- Overall, this study provides valuable insights into the environmental controls of coastal erosion rates and helps decrease the range in predicting coastal erosion rates. The empirical relationships we hope to produce can serve as potential benchmarks for evaluating representations of environmental controls in process-based models.

CITED STUDIES

Gibbs et al., 2017, U.S. Geological Survey data release, https://doi.org/10.5066/F72Z13N1.
Jorgenson et al. 2014. Permafrost database development Characterization and mapping for Northern Alaska. https://catalog.northslopescience.org/dataset/2236
Lara et al. 2018. Scientific Data, 5(1), 180058. doi:10.1038/sdata.2018.58
Nicolsky & Romanovsky 2017. Simulated permafrost dynamics across the Alaskan North Slope region in the 20th and 21st centuries. https://cida.usgs.gov/thredds/catalog.html?dataset=cida.usgs.gov/north slope_rcp45_gravel0.6m_styrofoam.
North Slope Science Initiative, 2013, Landcover/Vegetation Mapping for North Slope of Alaska. https://catalog.northslopescience.org/dataset/2450

ACKNOWLEDGMENT

This study was supported by a grant from U.S. Department of Energy; under Sandia National Laboratories contract No. DE-AC05-00OR227 as part of the InteRFACE project.