PROBABILISTIC PREDICTIONS OF EQUILIBRIUM RIPPLE GEOMETRY FOR TIME-DEPENDENT SEAFLOOR MODELING

Allison M. Penko, U.S. Naval Research Laboratory, allison.penko@nrlssc.navy.mil
Ryan E. Phillip, U.S. Naval Research Laboratory, ryan.phillip@nrlssc.navy.mil

ABSTRACT
We present a new equilibrium ripple predictor using a machine learning approach that outputs a probability distribution of wave-generated equilibrium wavelengths and statistics including an estimate of ripple height, the most probable ripple wavelength, and sediment and flow parameterizations. The Bayesian Optimal Model System (BOMS) is an ensemble machine learning system that combines two machine learning algorithms and two deterministic empirical ripple predictors with a Bayesian meta-learner to produce probabilistic wave-generated equilibrium ripple wavelength estimates in sandy locations.

BACKGROUND

Sand ripples are geomorphic features on the seafloor that affect bottom boundary layer dynamics including wave attenuation and sediment transport and are critical for time-dependent ripple and sediment transport models (Traykovski [2007], Nelson and Voulgaris [2015], Soulsby et al. [2012], Penko et al. [2017]). Decades of work focused on studying the equilibrium geometry of ripples generated by constant wave propagation over a sandy bed has resulted in many empirical formulations derived from laboratory and field observations (e.g., Soulsby and Whitehouse [2005], Faraci and Foti [2002], Mogridge et al. [1994], Nielsen [1981], and others). Typically, the data used to derive these deterministic equilibrium ripple predictors have a large spread and the predictions therefore have high uncertainty. Nelson et al. [2013] compiled over 50 years of laboratory and field observations to produce one of the most recent empirical formulations for the prediction of equilibrium ripple length and height given constant wave forcing. To date, this empirical deterministic predictor is the most accurate available.

METHODOLOGY

The Bayesian Optimal Model System (BOMS) (Phillip et al. [2022]) combines two ML base models with two empirical ripple predictors into an ensemble stacked system including a Bayesian meta-learner to produce probabilistic equilibrium ripple wavelength predictions and geometry statistics including an estimate of ripple height, the most probable ripple wavelength, and sediment and flow parameterizations.

The four base models within the first layer of the stacked generalized system include two machine learning algorithms and two deterministic empirical equilibrium ripple wavelength formulations. After extensive testing of several machine learning models, an optimized Gradient Boosting Regressor (Friedman [2002]) and a non-optimized XGBoost Regressor (Chen and Guestrin [2016]) were chosen to allow for the minimization of overfitting without significantly increasing the bias.

Including empirical equilibrium equations as base models allows for the representation of the present state-of-the-art deterministic ripple geometry predictions based on over 50 years of research and observations. The Nelson et al. [2013] and Traykovski et al. [1999] equations were chosen because while they are both functions of wave-orbital excursion, Traykovski et al. [1999] includes an additional dependence on sediment settling velocity and wave period. The two predictors incorporate a variety of independent variables and resulted in predictions with the highest skill.

The system was trained with the 50+ year dataset compiled by Nelson et al. [2013]. The compilation contains observations from both field and laboratory studies. Parameters included in the compilation are ripple height, ripple wavelength, median grain size, wave orbital velocity, semi-orbital excursion, water depth, wave period, and water density. The data was filtered to include only equilibrium ripples using the equilibrium ripple criteria established in Nelson et al. [2013]. Additionally, the dataset is filtered to only include wave-generated orbital ripples with wavelengths less than a specific threshold. A maximum ripple wavelength threshold of 1m for laboratory data and 1.5m for field data was applied to filter out potential megaripples that have longer wavelengths due to being forced by other processes (tides, infragravity waves, etc.). The final equilibrium ripple filtered dataset consisted of 3,622 data points. Lastly, null values were removed and a method was employed to predict ripple height values when they were not included in the dataset. A KNNImputer class was used to estimate missing ripple height values and serves as an additional output of deterministic ripple height in addition to the probabilistic ripple length predictions.

A Bayesian Linear Regression (BLR) is used as the meta-learner. Bayesian inference determines the posterior distribution of the model features from a prior probability. The final probabilistic predictions are generated from the posterior distribution with Markov Chain Monte Carlo sampling.

RESULTS

Ten-fold cross validations were performed on both layers of the stacked system as well as the ripple height imputer method. The KNNImputer method predicted ripple heights with an adjusted R-squared (R_{adj}^2) of 0.88 and root mean square error (RMSE) of 0.014 m. The comparisons of each of the base model layers resulted in R-squared (adjusted), RMSE, and Bias values as presented in Table 1. A final ten-fold cross validation was performed on the final predictions from the meta-learner in BOMS. The predictions resulted in an R-squared value of 0.93 and an average root mean square error (RMSE) of 8.0 cm (Figure 1).

Model	R^2_{adj}	$RMSE(m)$	$Bias(m)$
Gradient Boosting Regression	0.93	0.082	-0.0011
XGBoost Regression	0.92	0.087	-0.0004
Traykovski et al. (1999)	0.57	0.199	-0.0603
Nelson et al. (2013)	0.50	0.215	-0.0879

Table 1 - Performance metrics of the base models when subject to 10-fold cross validation on the training dataset.

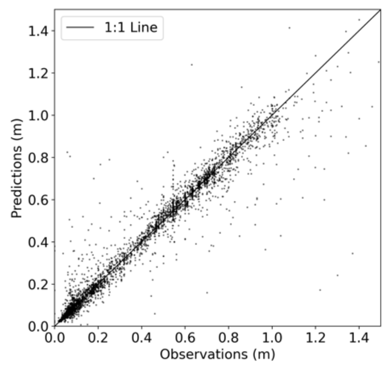

Figure 1 - Scatter plot visualizing the results from ten-fold cross validation of the stacked model system, BOMS.

BOMS was also tested on three unique field data sets not included in the training data. The RMSE from the field comparisons ranged from 5-10cm with biases $\mathcal{O}$(1cm). During both cross validation and testing on three unique field datasets, BOMS provided more accurate wavelength predictions than each individual base model and other common ripple predictors.

APPLICATIONS

Due to its ability to provide probabilistic distributions of ripple length, BOMS is well suited for providing distributions of equilibrium wavelengths that can be used to drive time-dependent seafloor roughness models. For example, to stochastically model the changes in seafloor ripples, we can combine a point process model with BOMS to produce probability distributions of ripple wavelengths over time given a timeseries of wave conditions (Figure 2).

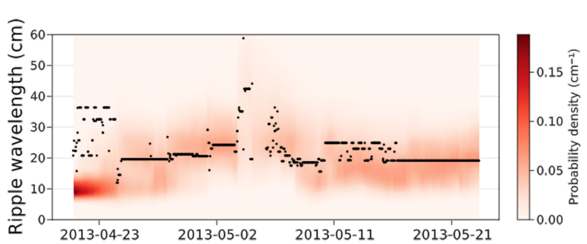

Figure 2 - Time series plot of the observed ripple wavelength (black dots) and the probability density of predicted ripple wavelengths (red shaded) at a location in 8m water depth off the coast of Panama City, FL.

CONCLUSIONS

The Bayesian Optimal Model System (BOMS) predicts probabilistic distributions of wave-generated equilibrium sand ripple wavelengths and estimations of a deterministic ripple height using machine learning techniques in combination with pre-existing deterministic empirical equilibrium ripple predictors. Overall, BOMS provides more accurate predictions of ripple wavelength during both cross-validation and testing compared to the performance of each independent base model and other common equilibrium ripple predictors. Practical applications of BOMS include probabilistic ripple geometry forecasting and the coupling with time-dependent ripple and sediment transport models.

REFERENCES

Chen, Guestrin (2016): XGBoost: A Scalable Tree Boosting System, pp. 785-794.
Faraci, Foti (2002): Geometry, migration and evolution of small-scale bedforms generated by regular and irregular waves. Coastal Engineering, 47, pp. 35-52.
Friedman (2002): Stochastic Gradient Boosting. Computational Statistics & Data Analysis, 38, pp. 367-378.
Mogridge, Davies, Willis (1994): Geometry prediction for wave-generated bedforms, Coastal Engineering 22, pp. 255-286.
Nelson, Voulgaris, Traykovski (2013): Predicting wave-induced ripple equilibrium geometry, Journal of Geophysical Research: Oceans, 118, pp. 3202-3220.
Nelson, Voulgaris (2015): A spectral model for estimating temporal and spatial evolution of rippled seabeds, Ocean Dynamics, 65, pp. 155-171.
Nielsen (1981): Dynamics and geometry of wave-generated ripples. Journal of Geophysical Research: Oceans, 86, pp. 6467-6472.
Penko, Calantoni, Hefner (2017): Modeling and observations of sand ripple formation and evolution during TREX13, IEEE Journal of Oceanic Engineering, 42, pp. 260-267.
Phillip, Penko, Palmsten, DuVal (2022): A machine learning approach to predicting equilibrium ripple wavelength, Environmental Modelling & Software, https://doi.org/10.1016/j.envsoft.2022.105509.
Soulsby, Whitehouse (2005): Prediction of ripple properties in shelf seas-Mark 2 predictor for time evolution-Final technical report.
Soulsby, Whitehouse, Marten (2012): Prediction of time-evolving sand ripples in shelf seas, Continental Shelf Research, 38, pp. 47-62.
Traykovski, Hay, Irish, Lynch (1999): Geometry, migration, and evolution of wave orbital ripples at LEO-15, Journal of Geophysical Research: Oceans, 104, pp. 1505-1524.
Traykovski (2007): Observations of wave orbital scale ripples and a nonequilibrium time-dependent model. Journal of Geophysical Research: Oceans, 112.