SHOALING FORECASTING AT THE MOUTH OF THE MISSISSIPPI RIVER WITH MACHINE LEARNING MODELS

Magdalena Asborno, Ph.D., Applied Research Associates, Inc., Magdalena.Asborno@usace.army.mil
Jacob Broders, Applied Research Associates, Inc. jbroders@ara.com
Kenneth N. Mitchell, Ph.D., US Army Engineer Research and Development Center, Kenneth.N.Mitchell@usace.army.mil
Michael A. Hartman, Ph.D., US Army Engineer Research and Development Center, Michael.A.Hartman@usace.army.mil
Lauren. D. Dunkin, US Army Engineer Research and Development Center, Lauren.M.Dunkin@usace.army.mil

BACKGROUND

The primary coastal outlet for the Mississippi River at the Gulf of Mexico, known as Southwest Pass (SWP), is one of the most highly utilized commercial deep-draft waterways in the United States (Figure 1). Disruptions in navigation due to hard-to-predict accumulation of sediments in SWP affect the access of deep-draft vessels to four of the nation's top 15 ports measured by tonnage that connect the U.S. Midwest with global markets, and handle around 500M tons of cargo annually. The SWP is maintained by the U.S. Army Corps of Engineers (USACE) at a depth of 50 feet (15.2 meters). The USACE spends on the order of $100M annually on dredging operations to maintain a reliable shipping channel throughout SWP, and the unpredictability of rapid-onset shoaling has been known to drive annual costs to more than twice that amount (Hartman, et al., 2022). Presently, USACE New Orleans District project managers rely on rules of thumb with seasonal river stage trends and thresholds to get 10 to 14 days of lead time for shoaling conditions at SWP.

PURPOSE AND OVERVIEW

This work covers the development of a machine learning regression model to increase both the lead times for and accuracy of shoaling forecasts and associated dredging requirements in SWP. The machine learning regression models are embedded in a multi-variate, multi-step time-series forecasting framework. Results obtained with a Random Forest (RF), and two Artificial Neural Networks (ANN) (a Multi-Layered Perceptron (MLP), and a Long Short Term Memory (LSTM)) are compared, for different scenarios of input days, to forecast 45-day channel shoaling volumes. An increase trend in daily shoaling values indicates a need to mobilize dredges to the SWP area. Model performance is evaluated with Root Mean Squared Error, normalized by total volume of shoaling in the 45 days forecasted (nRMSE). In the absence of metrics to evaluate the state of the practice, an univariate Auto-Regressive Integrated Moving Average (ARIMA) model is used as baseline.

The variable to predict is based on daily volumes of sediment accumulated in the 35-mile stretch of the Mississippi River between Mile 13.4 Above Head of Passes, and Mile 22 Below Head of Passes (Figure 1). Historical values of sediment accumulated on SWP between 2012-2022 are obtained from the Corps Shoaling Analysis Tool 2.5 (CSAT) (Dunkin, Coe, & Ratcliff, 2019), and used as proxy for dredging needs. This time-series is not stationary, thus CSAT estimates were transformed to 7-day rolling averages. The result is used both as variable to predict and input to the forecasting model.

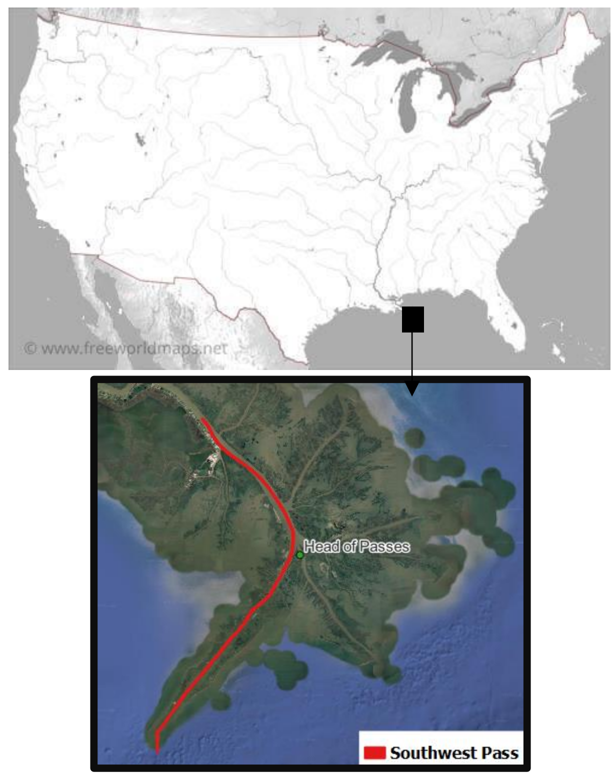

Figure 1. Southwest Pass geographical context

DATA

Following a physics-informed approach, the data preselected to feed the time-series forecasting model is in line with variables that may affect shoaling (USGS, 2018) (Nel, Dalu, & Wasserman, 2018). In addition to the variable to predict, 99 potential input variables were considered. One of them was the week of the year when the data was collected. In addition, a pool of 98 variables from 57 stations located along the Mississippi and Ohio Rivers are automatically collected from USGS and RiverGages websites through APIs. Type of variables include: river stage, discharge, turbidity, water temperature, precipitation, relative humidity, and air temperature. Precipitation and relative humidity were not correlated to the variable to predict, and were removed from the pool. Only variables available for more than 90% of the historical period of record were considered. The remaining 39 variables are subject to replacement of erroneous values by linear interpolation, a minimum-maximum scaling, 7-day rolling averaging (in line with the data preparation of the variable to predict). Feature selection is completed through a decision-tree based gradient boosting regressor (XGB). The top-6 variables ranked by importance are used to feed the sediment forecasting regression model. This approach allows the framework to automatically adapt to potential future changes in shoaling behavior and data availability.

METHODOLOGY

Machine learning models learn repetitive patterns from relatively big data. To allow for a single historical time-series to be used for supervised machine learning, the 10 years of historical data are broken down into smaller time-series (i.e. “instances”). The size of the instances considers a number of days used by the model as input, or “in-lag”, plus the 45-day prediction. For this work, models were applied and compared to the following scenarios of in-lags: 45; 60; 90; 120; 150, and 180 days. After removing the last 45-days used for model evaluation, for each scenario, as many as possible instances are created, each starting one day apart from each other. In this way, the machine learning regressor benefits from an increased number of instances to be trained and tested, without the need to generate synthetic data.

For each regression model type (RF, MLP, and LSTM) several architectures and parameter set-ups were tested. The combination that produced the higher average nRMSE for each type of regressor was selected. The MLP regressor was constructed with two hidden layers of 10 and 5 nodes respectively, using hyperbolic tan activation function, and stochastic gradient descent for optimization of weights. The RF had 75 trees, and mean squared error was the loss function. To prevent overfitting, the maximum depth is limited to three layers, and the minimum number of samples required to split an internal node is two. The sequential LSTM was made with a dimensionality of 50-75 LSTM layers all with hyperbolic tan activation function. The final layer is a dense layer with a linear activation. During the training phase of each epoch, the model has dropout layers after each LSTM layer to prevent overfitting on the training data.

For model evaluation purposes, in each in-lag scenario, the nRMSE is calculated over all evaluation instances, and over instances evidencing an increasing shoaling trend (nRMSEinc), indicative of dredging needs.

RESULTS

Results indicate that all multivariate machine learning models outperformed the univariate baseline. Moreover, to evaluate future dredging needs, it is particularly interesting to observe the behavior of model predictions in the subset of instances with increasing shoaling trends (Figure 2). In this context, the best results were obtained with the MLP model for the 60-day input scenario. Acting as a digital twin of predicted SWP sediment volumes, the proposed model will modernize and accelerate dredging operations decision-making at SWP. The approach adopted for SWP in this work may be applied to forecast shoaling needs at other critical coastal channels.

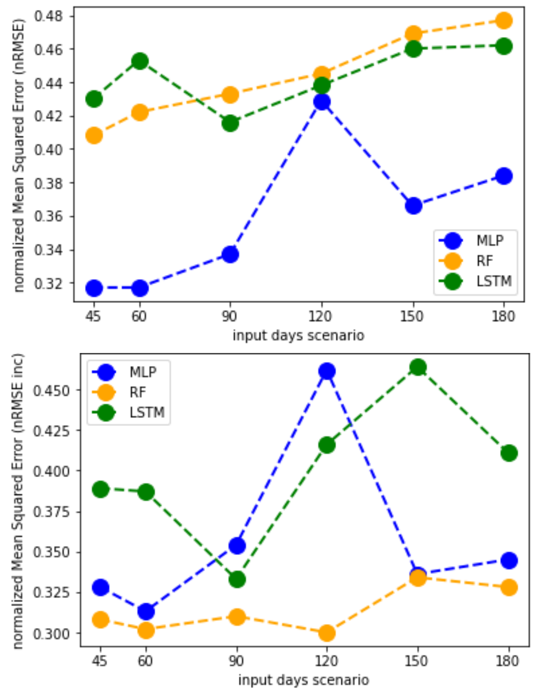

Figure 2 - Comparison of results obtained with RF, MLP, and LSTM model types and in-lag scenarios, based on nRMSE (top) and nRMSEinc (bottom).

REFERENCES

Dunkin, L. M., Coe, L. A., & Ratcliff, J. J. (2019). Corps Shoaling Analysis Tool : predicting channel shoaling. USACE-ERDC-CHL. doi:http://dx.doi.org/10.21079/11681/30382

Hartman, M. A., Mitchell, K. N., Dunkin, L. M., Lewis, J., Emery, B., Lenssen, N. F., & Copeland, R. (2022). Southwest Pass Sedimentation and Dredging Data Analysis. J. Waterway, Port, Coastal, Ocean Eng., 148(2).

Nel, H., Dalu, T., & Wasserman, R. (2018). Sinks and sources: Assessing microplastic abundance in river sediment and deposit feeders in an Austral temperate urban river system. Science of The Total Environment, 612, pp 950-956. doi:https://doi.org/10.1016/j.scitotenv.2017.08.298

USGS. (2018, June 9). Sediment and Suspended Sediment. Retrieved from Water Science School: https://www.usgs.gov/special-topics/water-science-school/science/sediment-and-suspended-sediment