SYMPOSIUM OP'24
Ocean Predict
In partnership with unesco Intergovernmental Oceanographic Commission
2021 2030 United Nations Decade of Ocean Science for Sustainable Development

# An approach for good modelling and forecasting of sea surface salinity in a coastal zone using machine learning LASSO regression models built with sparse satellite time-series datasets

The risks of upstream seawater intrusion from coastal zones, particularly to the environment and people's health are gradually becoming serious issues that require proactive environmental monitoring and good modelling approaches. However, the temporal resolutions of relevant contemporary all-weather satellites that detect sea surface salinity (SSS) are unable to support real-time applications that can provide the required early warning information for mitigating such risks. Our current practical knowledge of the efficiency of machine learning (ML) least absolute shrinkage and selection operator (LASSO) regression models built with relatively sparse all-weather satellite data for achieving relatively accurate predictor variable selection, collinearity detection, and high SSS prediction accuracy that can provide early warning information for mitigating such risks is still limited. Therefore, we utilized relatively sparse time series all-weather satellite datasets consisting of 6 potential predictor variables (PPVs), wind speed (WS), high wind speed (HWS), sea surface temperature (SST), absolute dynamic topography (ADT), sea level anomalies (SLAs) and precipitation (PRECIP) (January 2016-December 2020) to construct a ML LASSO model to predict SSS on a tropical coast. We determined the best combination of lookback (LB) and h-step-ahead (H) parameter values for building a relatively accurate ML LASSO model with the datasets. We predict and validate the monthly SSS values for January-December 2021. We show that the LB:24 and H:12 parameter, with an RMSE of 0.54437, are the best for building such a relatively accurate LASSO model. We show that the WS, HWS, and SLAs are the most important PPVs. However, we show the limitations of such a LASSO model in achieving relatively accurate predictor variable selection and collinearity detection. We show practical solutions to such limitations by utilizing L0-regularized regression (L0) model to assist the LASSO model to achieve a relatively high SSS prediction accuracy. We predict the monthly SSS values and validate them with the observed SSS to obtain RMSE of 0.7428, and MAPE of 1.9039%. A MAPE value that is approximately 5 times less than 10% implies a high SSS prediction accuracy that can be replicated to provide useful early warning information for mitigating such risks in coastal areas. The results imply that the good practice for using such datasets to build a relatively accurate ML LASSO model for SSS forecasting should begin with rigorous supervised-automatic deletion of observation records with null values and outliers, followed by unbiased selection of appropriate parameter values, identification of important predictor variables, and collinearity assessment.

*Opeyemi Ajibola-James,a,b,* and Francis I. Okeke,a,**aUniversity of Nigeria Enugu Campus, Faculty of Environmental Studies,Department of Geoinformatics and Surveying, Enugu, NigeriabGeo Inheritance Limited,Department of Data Science, Geoinformatics and Surveying, Port Harcourt, Nigeria*Corresponding and First Author, Email: o.ajibolajames@geoinheritance.com**Second Author, Email: francis.okeke@unn.edu.ng*

**Correspondence*