Validation of Skill-Assessment Software for NOAA NOS Operational Oceanographic Modeling Systems

The National Ocean Service (NOS) currently validates Operational Forecast System (OFS) model output against real-time observations using a standardized Fortran-based skill-assessment package that computes accuracy and error metrics. The NOS Shared Cyber-Infrastructure and Skill Assessment (SCI-SA) project aims to develop a next generation Python-based system, delivering the same statistics from a simpler user interface with expanded and enhanced features.

This study applied both systems to Chesapeake Bay OFS output from January to June, 2024. Observations and matching model series for water level, currents, water temperature, and salinity were extracted for nowcast and forecast runs, and identically processed in accordance with the standard operating procedures for NOS skill assessments. Each package produced a comprehensive record at every station for the specified variable and run type, containing bias, root mean square error (RMSE), standard deviation, central and outlier frequencies, error-duration metrics, and related statistics. These paired records were directly compared and rigorously tested for statistical equivalence.

Visual data analysis exposed two discrepancies for all variables: a time series offset when station files were missing and a fixed vertical offset caused by model ingestion techniques. After correcting these issues and matching preprocessing settings, graphical differences diminished and every station, variable, and run type returned statistically similar skill assessment scores.

This experiment confirms that the Python workflow reproduces legacy results with full fidelity while offering a cleaner user interface and straightforward output. This validation framework provides a foundation for future NOS skill-assessment comparisons and advances our ability to deliver modern, continuously evaluated coastal ocean forecasts.