

# User-based skill assessment techniques for operational hydrodynamic forecast systems

*A Zhang, KW Hess and F Aikman III, Coast Survey Development Laboratory, National Ocean Service/NOAA, Silver Spring, Maryland, USA*

---

NOAA's National Ocean Service (NOS) applies hydrodynamic models for the development, transition and implementation of operational forecast systems in US estuaries, ports, lakes and the coastal ocean. These models and systems have applications in the support of safe and efficient marine navigation and emergency response, as well as marine geospatial and ecological applications. As part of the forecast system development, skill assessment is required to evaluate each model system. A forecast model has to be ensured that it has been developed and implemented in a scientifically sound and operationally robust way, and its performance should meet or exceed target benchmarks before its products are approved for release to the public.

This paper discusses a new approach for the development of standards and procedures for skill assessment of operational forecast systems, which are designed primarily to support marine transportation and emergency response. Accordingly, this skill assessment focuses on variables, statistics and targets that are of the most use to these communities. The components of skill assessment include: data analysis techniques, definition of model run scenarios and statistical variables, and NOS target values and acceptance criteria. In addition, a large suite of skill assessment software has been developed to compute the new skill assessment scores using data files containing modelled variables and observed data. This software ingests model forecasts and observational data, fills in gaps, creates time series, extracts skill-specific parameters, and computes the skill scores. The software has been tested and applied to all NOS operational forecast systems.

---

---

## AUTHORS' BIOGRAPHIES

Frank Aikman III, PhD is Chief of the Marine Modeling & Analysis Programs, Coast Survey Development Laboratory, of NOAA's Ocean Service. He has been involved in coastal and estuarine observations, circulation modelling and prediction for over thirty years, and has published over 30 peer-reviewed papers and book chapters and many more reports and proceedings. He is presently very keen on development of techniques and models for accurate real-time forecasting of water levels, currents, temperature and salinity and the practical application of these tools to marine navigation, emergency response and ecological issues.

Kurt W Hess, PhD, is a senior oceanographer who has worked for many years at NOAA in the fields of numerical circulation modelling and environmental forecasting. He is presently a contractor for the Chesapeake Research Consortium of Annapolis, Maryland, and is working closely with CSDL on the software for vertical datum transformation (VDatum) and on other modelling and hydrographic surveying techniques.

Aijun Zhang, PhD, is at the Center for Operational Oceanographic Products and Services of NOAA's Ocean Service. He has worked for many years at NOAA in the fields

of numerical circulation modelling, environmental forecasting, and oceanographic data processing and analysis. Presently his primary focus is on the transition of research and development models to operations and maintenance of NOS operational forecast systems.

## INTRODUCTION

The National Ocean Service (NOS) of the National Oceanic and Atmospheric Administration (NOAA) has been at the forefront in developing, implementing and maintaining operational oceanographic observing systems (ie, the Physical Oceanographic Real-Time System (PORTS) and the National Water Level Observation Network (NWLON)) and operational coastal ocean forecasting systems.<sup>1, 2</sup> NOS presently focuses on developing estuarine and coastal hydrodynamic models, driven by real-time *in-situ* and remote data and the outputs of operational weather forecast models to produce nowcasts and short-term forecasts of water levels, currents, temperature and salinity in order to support NOS's mission to promote safe and efficient navigation (eg, forecast water level fields for under-keel clearance), to assist in emergency response, such as circulation and density fields for oil spill trajectory forecasts, and other NOS programmes including marine geospatial and ecological applications.<sup>3</sup>

There are currently nine water bodies in which NOS operational forecast systems (OFS) are functioning (Fig 1), including Chesapeake Bay, the Port of New York and New Jersey, Galveston Bay, the St Johns River, and the five Great Lakes.

The Chesapeake Bay Operational Forecast System (CBOFS) was established in the summer of 2001<sup>4, 5</sup> and is based on a 2-dimensional, barotropic version of the MECCA model.<sup>6</sup> Its design and implementation was in response to user demands for short-term forecasts of water levels. The New York-New Jersey Operational Forecast System (NYOFS) was made operational early in 2003<sup>7, 8</sup> and is based on a 3-dimensional, barotropic version of the Princeton Ocean Model (POM).<sup>9</sup> Mariners in the Port of New York and New Jersey had asked for short-term forecasts of water levels and currents at a number of critical locations. The Galveston Bay Operational Forecast System (GBOFS) is a fully 3-dimensional, baroclinic application of the POM model and provides information on water temperature and salinity as well as water levels and currents.<sup>10</sup> The St Johns River Operational Forecast System (SJROFS) is also based on a fully 3-dimensional, baroclinic model,<sup>11</sup> but in this case the Estuarine Fluid Dynamics Code<sup>12</sup> is used. SJROFS was transitioned from an original (non-operational) application by the St Johns River Water Management District.<sup>13</sup>

The five Great Lakes Operational Forecast Systems (GLOFS) are based on fully 3-dimensional, baroclinic applications of the POM model and were transitioned to opera-

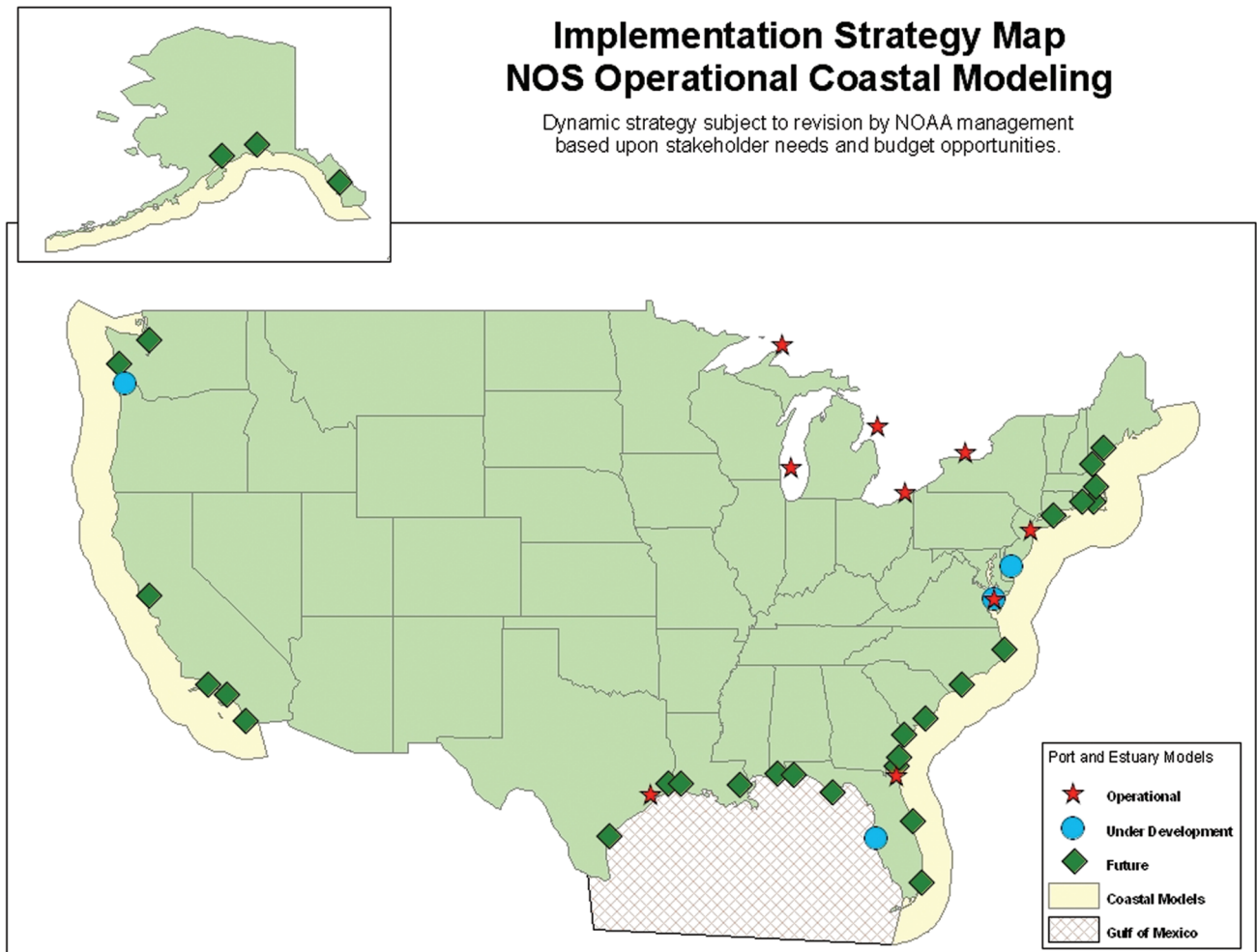


Fig 1: Implementation strategy map for National Ocean Service Operational Forecast Systems (OFS). The nine currently operational systems are represented by stars; systems under development by ovals; and future locations by diamonds<sup>1</sup>

tional status in NOS from the Ohio State University and the Great Lakes Environmental Research Laboratory.<sup>14, 15</sup> All nine OFS provide hourly updated nowcasts and 24–30h forecasts that are updated every six hours. Their outputs and products are available at <http://www.tidesandcurrents.noaa.gov>.

OFS are presently under development for the Columbia River, Delaware Bay, and Tampa Bay. Additionally, CBOFS is being upgraded to a fully 3-dimensional, baroclinic model. The Columbia River system is based on the Eulerian-Lagrangian Circulation (ELCIRC) model<sup>16</sup> and the new Delaware and Tampa Bay systems, as well as the CBOFS upgrade, will be based on the Regional Ocean Modeling System (ROMS).<sup>17, 18</sup> Model systems developed outside of NOS are subjected to skill assessment according to the NOS standards, as are all OFS developed within NOS.<sup>19, 20, 21</sup> There is currently no data assimilation being done in the OFS but it is expected that assimilation of real-time observed data will be incorporated in the future.

Skill assessment of ocean forecasting systems is still in an early stage of development compared to atmospheric systems, due primarily to the sparseness of oceanic observations.<sup>22</sup> For a general approach Dee<sup>23</sup> presented guidelines for standard validation documents and processes. Similarly, Refsgaard and Henriksen<sup>24, 25</sup> presented a generalised framework for quality assurance guidelines for hydrological models, including a consistent terminology (for terms such as model confirmation, code verification, model calibration, and model validation), a methodological foundation for bridging the gap between scientific philosophy and pragmatic modelling, and some application examples as well. To the extent possible, their approach has been followed by the authors. For example, it is assumed that the specific hydrodynamic model used has been validated for momentum and mass conservation, that it has an established domain of applicability, and that the authors have sought users (not model developers) to set performance criteria.

With regard to specific applications of model skill assessment, several approaches to the validation of coastal and estuarine circulation models are worth consideration. Vested *et al*<sup>26</sup> used some traditional statistical variables, such as the root mean square error (RMSE) and model mean error (ie, bias); the authors use both these parameters. Wilmott *et al*<sup>27, 28</sup> discuss an Index of Agreement (IoA) for comparing observed and modelled time series, and the IoA was used by Warner *et al*.<sup>29</sup> This index, however, produces large values (eg, > 0.85) for signals which are of the same shape but different amplitude; moreover, it is difficult to explain what exactly the score represents.

Dingman & Bedford<sup>30</sup> used statistical tests of significance (mean error, variance, skewness) to assess skill, but introduced non-dimensional skill scores (eg, 1 - mean error/mean value) as a way of determining whether deviations between model and data are real or whether they are simply related to the noise. The authors use mean error, variance (ie, standard deviation), and skewness (actually, positive and negative outlier frequency). Werner<sup>31</sup> used a visual analysis of the model response in the frequency domain. Although the response at certain frequencies can be useful in assessing performance, it is difficult to translate into a skill value.

A Pattern Correlation Coefficient (PCC) was used by Robinson *et al*<sup>32</sup> to evaluate salinity fields. Their PCC was

designed to show the skill of one forecast method (produced by a hydrodynamic model) over another (eg, persistence or climatology). Unfortunately, the point measurements the authors use for verification do not lend themselves easily to spatial pattern analysis. In sum, they have used many of the traditional skill assessment approaches described in the literature. However, as with most of the above methods, the assessment variables are difficult to translate into information from which a mariner can make decisions.

Information derived from ecological forecasting, often based on coupled physical, biological and chemical models, is increasingly in demand. Skill assessment approaches, standard statistics, and corresponding error criteria need to be developed; a recent issue of the *Journal of Marine Systems* is instructive in this regard as it addresses skill assessment of coupled marine (ecological) models with respect to plankton ecosystems and biogeochemistry, harmful algal blooms, food webs, and water quality.<sup>33</sup> Stow *et al*<sup>34</sup> reviewed several skill approaches and metrics, including univariate comparison of predictions and observations such as RMSE and average error (used by the authors), average absolute error (which is related to the authors' central frequency) and reliability index, correlation coefficient, and modelling efficiency (not used by the authors because they are difficult to interpret).

Another approach reviewed<sup>34</sup> is binary discriminator testing (eg, correct negative fraction and correct positive fraction, positive predictive value and negative predictive value); these are similar to the authors' positive and negative outlier frequency. Also, they describe multivariate comparisons (eg, cost functions), multivariate pattern evaluation, and comparison of spatial maps; these the authors have not used because they are limited to point data. Gregg *et al*<sup>35</sup> gave a comprehensive list of ocean biological models with data assimilation and their applications of skill assessment, in both ecosystem/biogeochemical and fisheries efforts. The majority use RMSE and/or graphical presentation of errors. However, the authors suggest that the errors in the observational and modelled datasets can influence skill assessment in different ways; it is assumed that the errors in both types are equivalent.

Wallhead *et al*<sup>36</sup> use skill metrics such as mean squared error to assess the skill or uncertainty of plankton model predictions. Stumpf *et al*<sup>37</sup> evaluate the skill of an operational harmful algal bloom forecast model on the west Florida shelf using such semi-quantitative metrics as percentage of correct forecasts (PCF), false positives, and false negatives. They used these scores because they lacked accurate time series of observations. Fitzpatrick<sup>38</sup> provides metrics he used to assess the skill of water quality models, including the standard RMSE and standard error, the less useful relative error and relative absolute error (which the authors note are unreliable for small values of the observed quantity), and correlation coefficient, and semi-quantitative metrics, such as PCF used by Stumpf *et al*.<sup>37</sup> Sheng and Kim<sup>39</sup> use a few quantitative metrics to evaluate a water quality model of the Indian River Lagoon, mainly absolute relative error and PCF, both of which are of limited value.

To summarise, several classic statistical variables mentioned above, such as RMSE, standard deviation, correct negative fraction and correct positive fraction, and positive predictive value and negative predictive value, are incorporated

in the NOS standard skill assessment software. However, some parameters such as correlation coefficient and non-dimensional skill scores are difficult for the mariner to use in decision-making, and due to the spatial sparseness of real-time observations, patterns of variables are difficult to obtain and translate into useful information.

The authors' NOS challenge was to develop and apply a robust and pragmatic set of standard analysis and metrics to quantitatively assess the performance of an operational ocean forecast system in a simple but meaningful way that relates to the purpose for which the system was developed. They have therefore followed the model validation process proposed by Dee<sup>23</sup> and Refsgaard.<sup>24</sup> However, following Williams *et al*<sup>40</sup> it was decided to limit use of metrics to those statistics which could be easily explained to the marine navigation user community (primarily commercial pilots and ship drivers, commercial and recreational fishers, and recreational boaters) and did not require a mathematical background to interpret.

The NOS procedures for developing and implementing operational forecast hydrodynamic model systems are described<sup>21</sup> and the skill parameters selected for use by NOS have been discussed.<sup>20</sup> Once the parameters were established, an automated computer program to test model-generated forecasts was designed and implemented. The program reads observational data and forecast system output, fills in missing data, creates new data structures

for comparison purposes, carries out the comparisons, and presents the results in tabular form. Details of the automated testing system are described in Zhang *et al*.<sup>41</sup> This paper presents a summary of the NOS approach to skill assessment.

## SKILL ASSESSMENT OVERVIEW

For the authors' purposes, skill assessment is the objective judgment of a model's performance (ie, its ability to reproduce observed variability and to predict future variability) using both objective standards and measures against other prediction methods. Some skill assessment statistics were designed to show how the prediction method could be improved. The standards described here are considered to be 'user-based', that is, established to show how well a model meets user needs (eg, for the users in the coastal navigation community, these needs include water levels for under-keel clearance and currents for manoeuvrability in port areas), and are not necessarily influenced by the model's capabilities. The OFS are not designed to forecast storm surge or inundation.

### Relevant variables and data

Based on comments from the user community, the primary variables in terms of importance to navigation in US coastal waters and ports are (see Table 1):

Group	Variable	Symbol
Group 1	Water level	H, h
	Current speed	U, u
	Current direction	D, d
	Salinity	S, s
	Water temperature	T, t
Group 2	Amplitude of high water	AHW, ahw
	Amplitude of low water	ALW, ahw
	Time of high water	THW, thw
	Time of low water	TLW, tlw
	Amplitude of maximum flood current	AFC, afc
	Amplitude of maximum ebb current	AEC, aec
	Time of maximum flood current	TFC, tfc
	Time of maximum ebb current	TEC, tec
	Direction of current at maximum flood	DFC, dfc
	Direction of current at maximum ebb	DEC, dec
	Time of start of current slack before flood	TSF, tsf
	Time of end of current slack before flood	TEF, tef
	Time of start of current slack before ebb	TSE, tse
Time of end of current slack before ebb	TEE, tee	
Group 3	Water level at forecast projection time of nn hrs	Hnn, hnn
	Current speed at forecast projection time of nn hrs	Unn, unn
	Current direction at forecast projection time of nn hrs	Dnn, dnn
	Salinity at forecast projection time of nn hrs	Snn, snn
	Water temperature at forecast projection time of nn hrs	Tnn, tnn

Table 1: Data series groups and the variables in each. Note that upper case letters indicates a prediction series (eg, H), and lower case (eg, h) indicates a reference series (observation or astronomical prediction). Slack water is defined as a current speed less than 1/2 knot. The direction is computed only for current speeds above 1/2 knot. Group 1 values have a uniform time interval and represent a time series of observations or model-generated data. Group 2 contains a series of values at a specific stage of the tide. Group 3 values represent a fixed forecast projection time

- The magnitude of the water level at all times and locations for under-keel clearance,
- The times and amplitudes of high and low water for under-keel clearance,
- The speed and direction of the currents at all times and locations, but especially at channel junctions, for manoeuvring,
- The times, amplitudes, and directions of the maximum flood and ebb currents (for manoeuvring),
- The start times and end times of slack water (slack water is defined as a current speed of less than 0.5kt, or  $0.26\text{m s}^{-1}$ ) before flood and ebb at all locations, but especially at channel junctions for planning turns in confined areas, and
- Water density, since it contributes to buoyancy, for under-keel clearance and cargo loading capacity. Density is usually defined in terms of salinity and temperature.

Skill assessment will therefore be focused on the model system's accuracy in simulating the above variables when the system is run under a set of specific conditions, or scenarios.

There are three scenarios under which the model is run to produce the data for skill assessment, and they are discussed in the order they would occur during model development. The scenarios begin with:

- (1) **Astronomical Tide Only**, in which only tidal boundary conditions (and possibly internal tidal body forcing) are used in the simulations; atmospheric forcing and river flows are absent and density is spatially uniform and constant in time. This scenario is important because in most coastal regions tidal variations are generally dominant, they may account for a significant part of the error, and because there are extensive data available for validation. Modelled time series can be harmonically analysed to produce constituent amplitudes and phases for comparison with accepted values. These values provide information on the model's behaviour in frequency space and can also illuminate the role of friction and non-linear processes.
- (2) **Hindcast scenario**, a long simulation is made using the best available gap-filled data for observed boundary water levels, atmospheric forcing, river flows, and ocean densities.
- (3) **Semi-Operational Nowcast scenario and the Semi-Operational Forecast scenario**, simulations are made in an operational environment (ie, running daily with real-time input) in which they will occasionally encounter missing observations and forecasts. The system must be able to handle these conditions without significant loss of accuracy. The length of time each scenario is to be run is, ideally, 365 days in order to capture all expected seasonal conditions. In addition, observational data at 6 min intervals normally are available. The output of the model runs is processed and saved in groups of individual values so the relevant variables (Table 1) can be calculated.

## SKILL ASSESSMENT STATISTICS

### Statistic definitions

Although no single set of statistics can quantify model performance perfectly, several, easily-calculated quantities have been chosen that provide relevant information on the

important categories of model behaviour. A summary of relevant terms and quantities is shown in Table 2. For a global assessment of errors (here the error is defined as the predicted value minus the observed value), both the SM and the frequency with which errors lie within specified limits (herein termed the Central Frequency, CF) are used. The SM will indicate how well the model reproduces the observed mean and the CF indicates how often the error is within acceptable limits.

The RMSE and Standard Deviation (SD) are required to be calculated, but these statistics have limited use since errors are not always expected to be normally distributed and CF is easier to explain to users lacking a technical background. A plot of the errors in hourly water levels at Baltimore, Maryland, (Fig 2) shows that the frequency of errors near the mean is greater than the normal (Gaussian) curve and smaller for values between one and two standard deviations away from the mean. The CF concept has been previously used in the NOS for data quality assurance standards.<sup>40</sup>

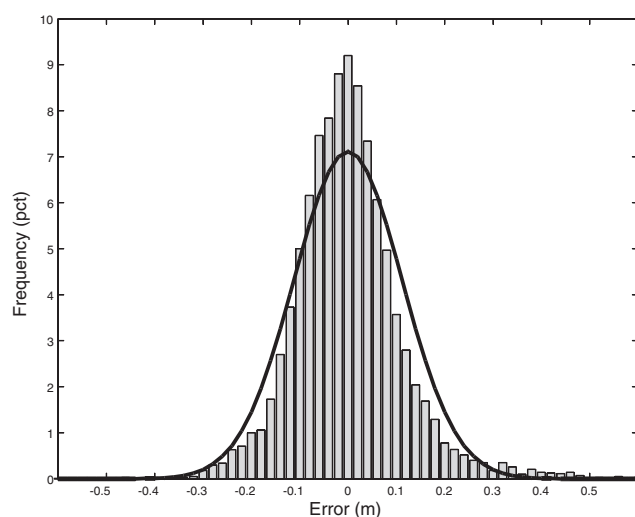


Fig 2: Hourly water level errors at Baltimore, Maryland, for 1996. The solid line is a Gaussian distribution

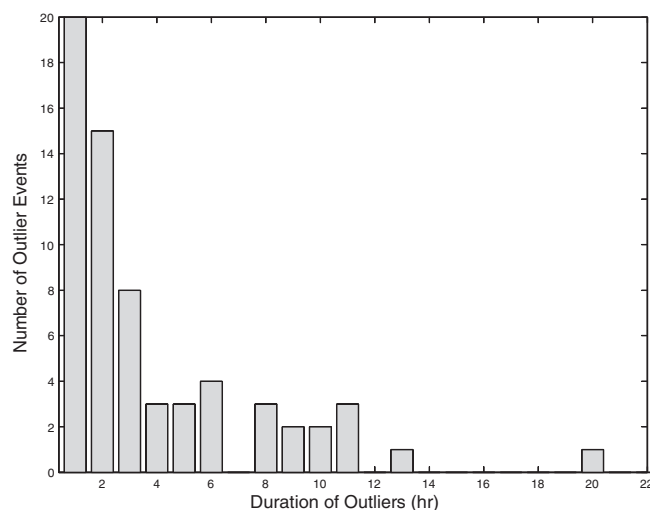


Fig 3: Number and duration of water level outlier events at Baltimore, Maryland. The number of events of duration less than 1h is quite large and has been truncated at 20

Variable	Explanation
Error	The error is defined as the predicted value, $p$ , minus the reference (observed or astronomical tide value, $r$ ): $e_i = p_i - r_i$ .
SM	Series Mean. The mean value of a series $y$ . Calculated as $\bar{y} = \frac{1}{N} \sum_{i=1}^N y_i$
RMSE	Root Mean Square Error. Calculated as $RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N e_i^2}$
SD	Standard Deviation. Calculated as $SD = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (e_i - \bar{e})^2}$
CF(X)	Central Frequency. Fraction (percentage) of errors that lie within the limits $\pm X$ .
POF(X)	Positive Outlier Frequency. Fraction (percentage) of errors that are greater than $X$ .
NOF(X)	Negative Outlier Frequency. Fraction (percentage) of errors that are less than $-X$ .
MDPO(X)	Maximum Duration of Positive Outliers. A positive outlier event is two or more consecutive occurrences of an error greater than $X$ . MDPO is the length of time (based on the number of consecutive occurrences) of the longest event.
MDNO(X)	Maximum Duration of Negative Outliers. A negative outlier event is two or more consecutive occurrences of an error less than $-X$ . MDNO is the length of time (based on the number of consecutive occurrences) of the longest event.
WOF(X)	Worst Case Outlier Frequency. Fraction (percentage) of errors that, given an error of magnitude exceeding $X$ , either (1) the simulated value of water level is greater than the astronomical tide and the observed value is less than the astronomical tide, or (2) the simulated value of water level is less than the astronomical tide and the observed value is greater than the astronomical tide.
PCD	The Principal Component Direction is computed as follows. PCD is counterclockwise from east and may represent either the flood or ebb direction. For an eastward current, $u$ , and northward current, $v$ , the PCD [Preisendorfer (1988) <sup>45</sup> , Eqn. 2.9] is $PCD = \frac{1}{2} \arctan \left( \frac{2 \sum_{i=1}^N (u_i - \bar{u})(v_i - \bar{v})}{\sum_{i=1}^N (u_i - \bar{u})^2 - \sum_{i=1}^N (v_i - \bar{v})^2} \right) + m \frac{\pi}{2}$ where $m$ is either 0 or 1, whichever gives a PCD that maximizes the variance, $s^2$ (Preisendorfer, 1988; Eqn. 2.6), defined as $s^2 = \cos^2(PCD) \sum_{i=1}^N (u_i - \bar{u})^2 + 2 \sin(PCD) \cos(PCD) \sum_{i=1}^N (u_i - \bar{u})(v_i - \bar{v}) + \sin^2(PCD) \sum_{i=1}^N (v_i - \bar{v})^2$
R	R is the ratio of the standard deviations of the current. $R = \frac{s^2(PCD + \frac{\pi}{2})}{s^2(PCD)}$

Table 2: Skill assessment statistics. The variables and statistics used in the skill assessment are explained

The frequency of times of poor performance is determined by analysing the outliers, which are errors whose absolute value exceed specified limits. The Positive Outlier Frequency (POF) measures how often the nowcast/forecast value is larger than the observation by some fixed amount. The duration of an outlier is defined as the number of consecutive occurrences of the outlier. Fig 3 shows the distribution of outliers and durations in water level errors at Baltimore, Maryland. The Maximum Duration of Positive Outliers (MDPO) indicates whether there are long periods when the model over predicts.

The Negative Outlier Frequency (NOF) measures how often the nowcast/forecast value is smaller than the observation by some fixed amount. The Maximum Duration of

Negative Outliers (MDNO) indicates whether there are long periods when the model under predicts. The MDPO and MDNO are computed with data without gaps. For tidal water levels, the ‘worst case’, from a model-based nowcast/forecast viewpoint, is when actual water level turns out to be low but the model erroneously predicted much higher water levels and the user would have been better off using the astronomical tide for the water level prediction. This is called the Worst Case Outlier Frequency (WOF).

Target frequencies and durations

Many of the statistics described above have an associated target frequency of occurrence. For regions with predomi-

nantly tidal variations, NOS has decided on the targets for the distribution of errors in the general form:

$$CF(X_1) \geq 90\%, POF(X_2) \leq 1\%, NOF(X_2) \leq 1\%. \quad (1)$$

The set of conditions (1) means that 90% of the errors will be within the range  $-X_1$  to  $+X_1$ , only 1% of the errors will exceed  $X_2$ , and 1% of the errors will be less than  $-X_2$ . The acceptable error limits ( $X$ ) are defined in Table 3. To insure that the positive and negative outliers track relatively large errors, we take  $X_1 = X$  and  $X_2 = 2X$ . For tidal water levels only,

$$WOF(2X) \leq 0.5\%. \quad (2)$$

Note that for a normal (Gaussian) distribution, the requirement that  $CF(X) \geq 90\%$  implies that  $SD = 0.608X$  and that  $POF(2X) = 0.05\%$ . However, errors cannot be expected to be Gaussian (see Fig 2).

Other statistics are expressed as limits on the duration of errors. For example, the target time duration (defined as the length of time bracketing consecutive occurrences of an outlier) is:

$$MDPO(2X) \leq L \text{ and } MDNO(2X) \leq L, \quad (3)$$

where  $L$  is the target time limit in hours.

The above target frequencies (1) and durations (3), plus SM, RMSE, SD, and WOF (for water levels only) are required for the assessment of nearly all variables, and are collectively called the Standard Suite of Statistics. The acceptable error criteria,  $X$ , and maximum allowable duration,  $L$ , have been selected after discussions with the primary users of the forecasts, the mariners, and with others from the maritime community. For these users, the most important factors to consider are under-keel clearance and manoeuvrability in narrow channels and in port areas. A summary of the acceptable error criteria,  $X$ , and maximum allowable duration,  $L$ , used in NOS is listed in Table 3.

In areas without significant tidal variations (eg, the Great Lakes) or for model simulations where the tides are excluded (eg, for looking at subtidal, wind-driven water levels), there may be different requirements. Although water levels and currents at uniform time intervals can be assessed with the same error and duration parameters, the times and amplitudes (ie,

water level magnitude, current speed) of extrema in water levels and currents may require modification. For example, given a time series of water levels, in lieu of a high and low water level, an extreme event can be defined when the water level rises above a specified critical value, or falls below a specified critical value. Significant current events can be defined in a similar way. In the selection of events, the critical values can be either specified as constant values or calculated from the observations using the series mean and standard deviation to select a small number of events (this amounts to dynamically varying criteria). Also, the duration of outliers may be related to meteorological events or to a lake's natural period.

#### Additional requirements

Two additional comparisons useful for model assessment include tidal harmonic constants and the skill of competing prediction methods. In regions with significant tidal variations, tidal harmonic constants (amplitudes and phases) of 37 constituents (the NOS standard) are derived from an astronomical tide-only model simulation of water levels and currents and compared with NOS accepted values, which are either obtained from NOS' historical harmonic constant data, or derived directly from the observations. A least-squares program created by NOS<sup>42</sup> should be used to analyse the modelled 365-day time series (1h intervals are suitable).

This comparison of tidal harmonic constants is for model checking only and there are no target values. For water currents, the Principal Component Direction (PCD) of the modelled output and the value of the ratio ( $R$ ) of the standard deviations across and along the PCD must be computed (see Table 2 for definitions of these quantities). Comparisons are then made between the harmonic constants from the simulated, along-PCD currents and the harmonic constants obtained from the data along its (possibly different) PCD. If the SD of the cross-PCD currents is large compared to the SD of the along-PCD currents (ie,  $R$  greater than 0.25), a similar comparison must be made for them. The values of  $R$  are required for both the modelled and observed time series.

For a comparison of the skill of competing forecast methods, the standard suite of statistics for the appropriate variables must be evaluated for the model-based forecast and for at least one other forecast method. The requirement is that the model-based forecast should be better than the forecast

Variable Name	Abbreviation	X	L (h)
Water level	H, Hnn,	15 cm	24
Extrema water level amplitude	AHW,ALW	15 cm	24
Extrema water level time	THW,TLW	0.5 hr	25
Current speed	U, Unn	0.26 m/s	24
Extrema current speed	AFC, AEC	0.26 m/s	24
Extrema current time	TFC, TEC	0.5 h	25
Slack water time	TSE, TEF, TSE, TEE	0.25 h	25
Current direction	D, Dnn,	22.5 deg	24
Extrema current direction	DFC, DEC	22.5 deg	25
Water temperature	T, Tnn	7.7°C	24
Salinity	S, Snn	3.5	24

Table 3: Acceptable error limits ( $X$ ) and the maximum duration limits ( $L$ ) for variables in tidal regions

based on other methods. If the proposed model's forecast is not an improvement over the others, the model's implementation should be reconsidered. For water levels and currents in tidal regions, these other forecast methods include

- (1) the astronomical tidal prediction, and
- (2) a tide plus persistence forecast.

The astronomical tide forecast is a prediction at each station using accepted harmonic constants for that station. A tide plus persistence water level forecast is constructed by adding an offset value, based on an observed offset at that station during some time period before the forecast is made, to the tide prediction at each station. For currents, the offset may be a mean current.

Note that for salinity, temperature, and non-tidal water levels, an astronomically-based forecast is not possible, but a comparison forecast can possibly be made by using another method such as another hydrodynamic model, climatological variation, a normal-mode prediction, and the observed persistence.

## SKILL ASSESSMENT SOFTWARE

In the above section, the general procedures of skill assessment are described, the relevant data variables, model run scenarios and statistical variables are defined, and target frequencies and associated acceptance error criteria for each variable are discussed. However, the skill assessment scores are, admittedly, difficult to compute. Therefore, a skill assessment software package<sup>41</sup> has been developed to compute the skill assessment scores automatically using data files containing observed, nowcast, and forecast variables. These data are processed and analysed using several techniques which include tidal prediction, harmonic analysis, gap filling and interval conversion, Fourier filtering, and other methods. These techniques also include ways of concatenating nowcasts and forecasts and in extracting water level and current extrema. The software was developed to run in the NOS operational environment, but versions of the software have been successfully ported to other users after minimal modification.

### Data analysis techniques

Observational data and model output generally must be processed and analysed using several techniques before skill assessment computation can begin. For example, observed time series that have gaps in the data are filled in one of three possible ways. Model-generated series, which are usually produced from numerous individual runs, must be concatenated to form a continuous series. For each type, the entire series is analysed for harmonic constants (in tidal regions) and extrema (eg, high water, maximum flood current) values. Specific methods for each process are discussed below.

## GAP FILLING AND TIME INTERVAL CONVERSION

Data gaps often exist in observations, and the extraction of extrema cannot be accomplished in a time series with gaps. Therefore, the program utilises three gap filling methods: linear, cubic spline, and singular value decomposition (SVD). If

a gap is small (eg, less than  $Time_1$ , where  $Time_1$  is a user-specified variable with a value of, for example, 1h), a simple linear interpolation is appropriate. If a gap is large (eg, greater than  $Time_1$  and less than  $Time_2$ , where  $Time_2$  is a user-specified variable with a value of, for example, 6h), either a cubic spline or SVD interpolation is used, depending on user input. If a gap is greater than  $Time_2$ , the gap is filled with a value of -999.99.

The time intervals of observed and modelled time series might be different. The processing will convert all time series with different time intervals into equally-spaced time series with the same unique desired time interval (6 min) using the chosen interpolation method according to gap length and user-specified criteria.

### Filtering

Because of short period variations and noise (specifically in current data), filtering of values in a time series is sometimes necessary to accurately select the extrema (ie, maximum and minimum) values and times. A 30h Fourier low-pass filter is used in this software since it computes the amplitudes of the components of the signal at various frequencies and reduces the amplitudes at selected high frequencies. Simple smoothing is to be avoided because it reduces extrema amplitudes.

### Tidal prediction and harmonic analysis

Tidal prediction of water level and current is required for skill assessment in tidal regions. In tidal regions, a comparison of tidal harmonic constants for the 37 tidal constituents is necessary for the evaluation of water levels and currents. For some NWLON stations, observed tidal elevation harmonic constants can be obtained from the website of NOAA's Center for Operational Oceanographic Products and Services (CO-OPS). Otherwise, tidal elevation and current harmonic constants can be derived from observed and modelled time series using harmonic analysis programs.

Two analytical techniques, least squares harmonic analysis and Fourier harmonic analysis, are used in terms that depend on of the length of the analysed time series. The least squares method<sup>42</sup> is a method for deriving the tidal constituents from a water level or current time series by creating a matrix of covariance between each individual constituent time series and the observed time series. The matrix is inverted to solve for the amplitudes and phases of the harmonic constituents. The constituent with the highest correlation with the total observed signal is then subtracted from the observed time series, and the matrix is recalculated with a residual time series in place of the observed.

This method has the capability of solving for up to 175 tidal constituents, but it is not useful to analyse a time series which is shorter than 29 days. The Fourier harmonic analysis method<sup>43</sup> uses Fourier series summations to obtain the tidal constituents of water level or current data. This method has been programmed for data periods of either 15 or 29 days of continuous data time series.

## CONCATENATION

For nowcasts and forecasts, model outputs are normally stored in different files for model runs on different days and on different cycles in the day. Therefore, it is necessary to concatenate certain of these files to construct several continuous



time series for further analysis. In the discussion below, an example is considered of a model that is run four times a day (ie, four cycles per day) and, for each run, produces a 6h nowcast time series and a 36h forecast time series, each with a time interval of 0.1h. Thus, all data (model-derived and observed) is analysed at the 0.1h (6 min) interval.

To concatenate the nowcasts, the output from each cycle of each day is simply appended to the end of the previous cycle's output. This series will be continuous because each nowcast is initialised with the model output from the end of the previous cycle's nowcast. In the example of four cycles per day, each 6h nowcast is appended to the previous nowcast. Thus, the 6h to 12h nowcast is appended to the 0h to 6h nowcast, and so on.

The forecasts are concatenated in two ways. In the first method, the value at a single projection time (forecast time) in each forecast cycle's time series is selected. For example, the forecasted value at hour 3 from the second cycle is appended to the forecasted value at hour 3 of the first cycle, and so on. The time interval is 6h and the time associated with each value in any one series is the time that the projection is valid. With this method, a unique series can be constructed for each of the 36h of the forecast, and individual values can be compared to observations at the same time. In the second method, the first 6h of each cycle is appended to the first 6h of the previous cycle. This method produces a time series with the time interval of 0.1h, although there may be a discontinuity of values every 6h, corresponding to the joining of two distinct segments. This series can be used to find outliers and extrema.

### Extrema extraction

For skill assessment in tidal regions, the magnitudes and times of high and low waters and the magnitudes and times of maximum flood and ebb currents are required. If there is noise the time series may first need to be filtered, as explained above (using a Fourier low pass filter to filter out high frequency variations [noise]). The extrema are extracted by searching for the largest and smallest values within a given time window in the series.

Extrema series of magnitudes and the corresponding occurrence times from both observations and model simulations are compared to evaluate the ability of the model system to simulate high and low waters and maximum flood and ebb currents.

In non-tidal regions, such as in the Great Lakes, there are not the regular high and low waters at tidal periods, but there may be some irregular water level amplitude events caused by meteorological forcing. These events are defined by peak levels at least two standard deviations above or below the mean level for the time period being analysed. The paired observed and simulated extreme events are compared to each other to assess the ability of the forecast system to simulate the water level magnitudes and times of occurrence (see<sup>44</sup> for Lake Ontario examples).

## IMPLEMENTATION OF NOS' SKILL ASSESSMENT SOFTWARE

The skill assessment software package is designed to perform water level, current, temperature, and salinity skill assess-

ment for different model systems in both tidally-dominated and non-tidal regions. The main processes are included in the order they occur: (1) parameter setup, (2) acquisition of verified water level observations from the CO-OPS database via the Internet; (3) tide prediction and tidal constituent acquisition from the CO-OPS database via the Internet if necessary; (4) concatenation of model hindcasts, nowcasts, and forecasts to form continuous time series; (5) creation of a persistence forecast based on the tidal prediction (in tidal regions) and observations; (6) computation of the standard statistical variables to produce the skill assessment table; and (7) harmonic analysis and tidal constituent comparisons.

The skill assessment software package consists of several shell scripts and Fortran programs. It is a stand-alone package which has been designed to be as computer system independent as possible and easy to be installed and executed in either a Unix or Linux environment. There are several steps involved in running this package after copying the software to the user's local directory. These steps can either be automatically initiated together or be run separately without an impact on the execution of other steps. For generic purposes, all time series (observations, tidal predictions, and model output) required by the skill assessment programs are processed and reformatted into the same ASCII format. The user has to modify two control files to provide appropriate parameter values associated with a specific project before running any shell scripts. More detailed instructions for installation and execution of this software are included in Zhang *et al.*<sup>41</sup>

Table 4 shows an example of a skill assessment score table for water levels in a tidal region and Table 5 shows example tidal harmonic constant comparisons for tidal currents. Examples of skill assessment score tables for currents, salinity, and temperature can be found.<sup>20</sup> The skill assessment software has been well tested and used to assess the skill of water level, current, water temperature, and salinity predictions (where available) from all of the NOS operational forecast systems, including those for the Chesapeake Bay, New York Harbor, Galveston Bay, and the St Johns River, and for forecast systems for each of the five Great Lakes.

Since the tables are somewhat difficult to interpret at a glance, several examples of the NOS skill assessment skill scores are shown graphically in Fig 4 for four different locations and ports, including the Bayonne Bridge, NJ, from NYOFS; Baltimore, MD, from CBOFS; Buchman Bridge, FL, from SJROFS; and Eagle Point, TX, from GBOFS. These skill scores were computed by comparing the OFS outputs with observations (at 6min intervals) for the twelve month period of calendar year 2008.

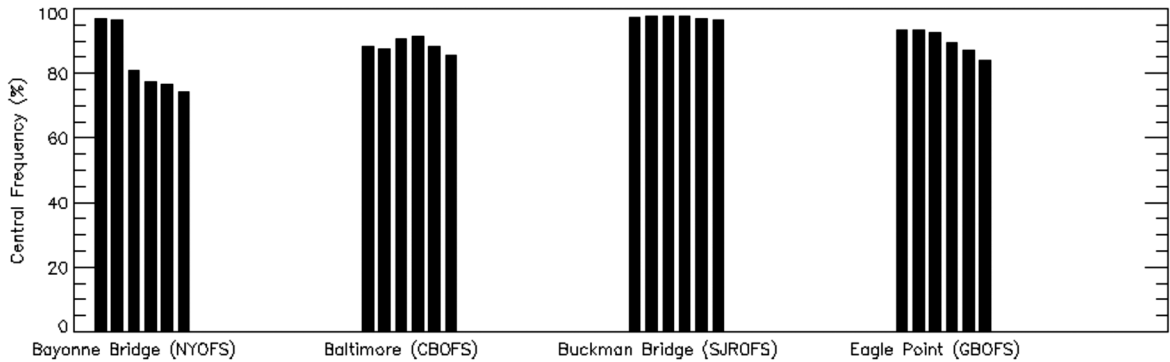
Three statistical parameters are shown: the CF of water level amplitude errors, the RMSE of water level amplitudes, and the CF of the error in the amplitude of high water and low water events. For each location and for the CF and RMSE of water level amplitudes the skill score is shown for the nowcast and for the forecasts in six hour increments (hours 0, 6, 12, 18 and 24). For the CF of the amplitude of high water and low water events, the skill of the nowcast and forecast of high and low water events is computed. These results are typical skill scores for these estuaries/ports.

Variable Criterion	X -	L -	lmax -	SM -	RMSE -	SD -	NOF <1%	CF >90%	POF <1%	MDNO <L	MDPO <L	WOF <.5%
<b>SCENARIO 1: ASTRONOMICAL TIDE ONLY</b>												
H			4801	-0.002								
h			4801	-0.001								
H-h	15cm	24h	4801	-0.001	0.110	0.110	0.0	80.5	0.0	0.0	0.0	11.35
AHW-ahw	15cm	24h	40	-0.046	0.128	0.121	0.0	75.0	0.0	0.0	0.0	
ALW-hwl	15cm	24h	40	0.048	0.084	0.070	0.0	92.5	0.0	0.0	0.0	
THW-thw	.5h	25h	40	-0.145	0.278	0.241	0.0	85.0	0.0	0.0	0.0	
TLW-tlw	.5h	25h	40	-0.255	0.383	0.290	0.0	77.5	0.0	0.0	0.0	
<b>SCENARIO 2: HINDCAST</b>												
H			4801	-0.006								
h			4801	-0.087								
H-h	15cm	24h	4801	0.081	0.082	0.014	0.0	100.0	0.0	0.0	0.0	0.00
AHW-ahw	15cm	24h	42	0.084	0.097	0.049	0.0	97.6	2.4	0.0	0.0	
ALW-alw	15cm	24h	42	0.093	0.116	0.070	0.0	95.2	4.8	0.0	0.0	
THW-thw	.5h	25h	42	0.052	0.234	0.231	0.0	95.2	2.4	0.0	0.0	
TLW-tlw	.5h	25h	42	0.057	0.243	0.239	0.0	92.9	2.4	0.0	0.0	
<b>SCENARIO 3: TEST FORECAST</b>												
H00-h00	15cm	24h	76	0.075	0.082	0.034	0.0	98.7	0.0	0.0	0.0	0.00
H06-h06	15cm	24h	76	0.071	0.092	0.059	1.3	98.7	0.0	0.0	0.0	1.32
H12-h12	15cm	24h	75	0.069	0.092	0.061	1.3	98.7	0.0	0.0	0.0	1.33
H18-h18	15cm	24h	74	0.070	0.090	0.057	1.4	97.3	0.0	0.0	0.0	1.35
H24-h24	15cm	24h	73	0.067	0.090	0.059	1.4	97.3	0.0	0.0	0.0	1.37
AHW-ahw	15cm	24h	42	0.084	0.097	0.049	0.0	97.6	2.4	0.0	0.0	
HLW-alw	15cm	24h	42	0.093	0.116	0.070	0.0	95.2	4.8	0.0	0.0	
THW-thw	.5h	25h	42	0.052	0.234	0.231	0.0	95.2	2.4	0.0	0.0	
TLW-tlw	.5h	25h	42	0.057	0.243	0.239	0.0	92.9	2.4	0.0	0.0	
<b>SCENARIO 4: SEMI-OPERATIONAL NOWCAST</b>												
H			4801	-0.006								
h			4801	-0.087								
H-h	15cm	24h	4801	0.081	0.082	0.014	0.0	100.0	0.0	0.0	0.0	0.00
AHW-ahw	15cm	24h	42	0.084	0.097	0.049	0.0	97.6	2.4	0.0	0.0	
ALW-alw	15cm	24h	42	0.093	0.116	0.070	0.0	95.2	4.8	0.0	0.0	
THW-thw	.5h	25h	42	0.052	0.234	0.231	0.0	95.2	2.4	0.0	0.0	
TLW-tlw	.5h	25h	42	0.057	0.243	0.239	0.0	92.9	2.4	0.0	0.0	
<b>SCENARIO 5: SEMI-OPERATIONAL FORECAST</b>												
H00-h00	15cm	24h	76	0.075	0.082	0.034	0.0	98.7	0.0	0.0	0.0	0.00
H06-h06	15cm	24h	76	0.071	0.092	0.059	1.3	98.7	0.0	0.0	0.0	1.32
H12-h12	15cm	24h	75	0.069	0.092	0.061	1.3	98.7	0.0	0.0	0.0	1.33
H18-h18	15cm	24h	74	0.070	0.090	0.057	1.4	97.3	0.0	0.0	0.0	1.35
H24-h24	15cm	24h	73	0.067	0.090	0.059	1.4	97.3	0.0	0.0	0.0	1.37
AHW-ahw	15cm	24h	41	0.092	0.160	0.132	0.0	70.7	9.8	0.0	0.1	
ALW-alw	15cm	24h	41	0.069	0.174	0.162	0.0	70.7	7.3	0.0	0.1	
THW-thw	.5h	25h	41	0.846	0.917	0.357	0.0	9.8	31.7	0.0	0.3	
TLW-tlw	.5h	25h	41	0.868	0.906	0.261	0.0	4.9	22.0	0.0	0.1	
<b>COMPARISON: ASTRONOMICAL TIDE ONLY FORECAST</b>												
H-h	15cm	24h	4801	0.086	0.140	0.110	0.0	60.8	0.0	0.0	0.0	0.00
AHW-ahw	15cm	24h	42	0.078	0.137	0.114	0.0	61.9	0.0	0.0	0.0	
ALW-alw	15cm	24h	42	0.098	0.150	0.115	0.0	57.1	2.4	0.0	0.0	
THW-thw	.5h	25h	42	0.019	0.295	0.298	2.4	95.2	2.4	0.0	0.0	
TLW-tlw	.5h	25h	42	-0.033	0.248	0.249	2.4	95.2	0.0	0.0	0.0	
<b>COMPARISON: PERSISTENCE FORECAST</b>												
H00-h00	15cm	24h	76	0.087	0.099	0.049	0.0	96.1	1.3	0.0	0.0	1.32
H06-h06	15cm	24h	76	0.089	0.107	0.060	0.0	90.8	0.0	0.0	0.0	0.00
H12-h12	15cm	24h	75	0.098	0.114	0.058	0.0	92.0	0.0	0.0	0.0	0.00
H18-h18	15cm	24h	74	0.086	0.105	0.060	0.0	94.6	1.4	0.0	0.0	1.35

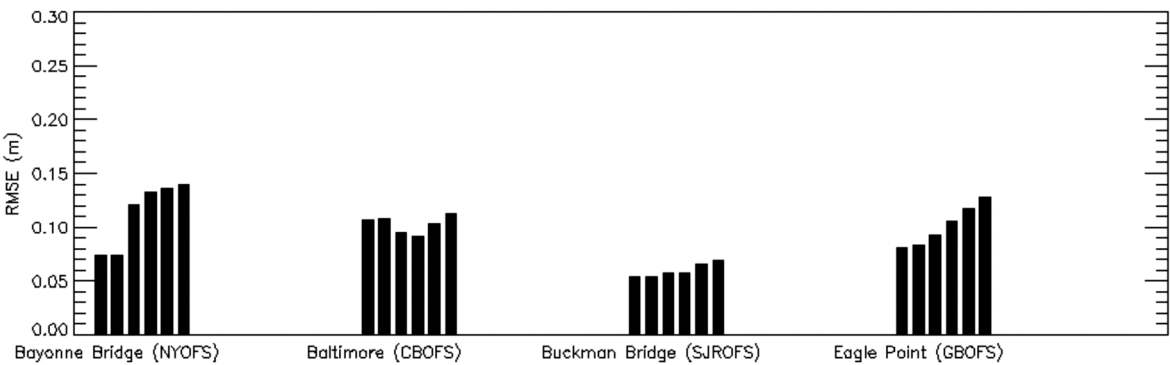
Continued Table

Variable Criterion	X	L	Imax	SM	RMSE	SD	NOF <1%	CF >90%	POF <1%	MDNO <L	MDPO <L	WOF <.5%
H24-h24	15cm	24h	73	0.080	0.106	0.070	1.4	94.5	1.4	0.0	0.0	2.74
AHW-ahw	15cm	24h	41	0.100	0.141	0.100	0.0	65.9	4.9	0.0	0.1	
ALW-alw	15cm	24h	42	0.090	0.168	0.143	0.0	66.7	4.8	0.0	0.1	
THW-thw	.5h	25h	41	0.866	0.931	0.345	0.0	7.3	26.8	0.0	0.2	
TLW-tlw	.5h	25h	42	0.850	0.896	0.288	0.0	4.8	19.0	0.0	0.2	

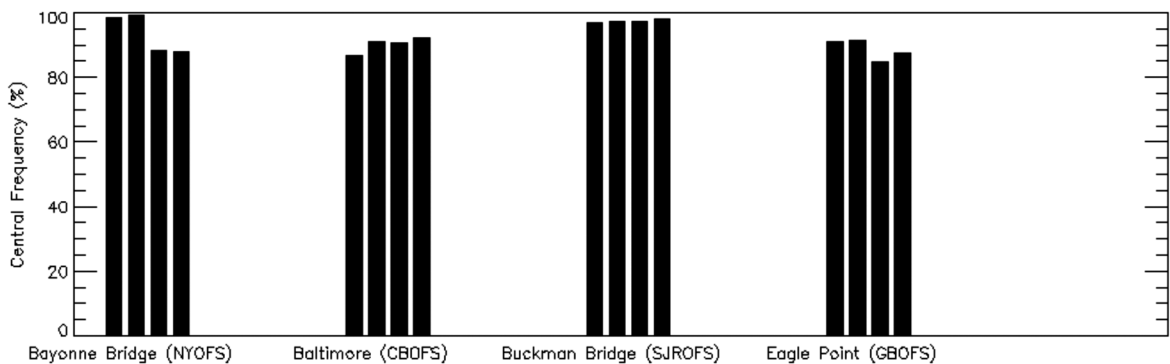
Table 4: Example of skill assessment scores for TIDAL WATER LEVELS



CF of Water Levels  
 parameters: nowcast H-h  
 forecast H00-h00 H06-h06 H12-h12 H18-h18 H24-h24  
 criteria: > 90%



RMSE of Water Levels  
 parameters: nowcast H-h  
 forecast H00-h00 H06-h06 H12-h12 H18-h18 H24-h24



CF of Amplitude of HW & LW Events  
 parameters: nowcast AHW-ahw ALW-alw  
 forecast AHW-ahw ALW-alw  
 criteria: > 90%

Fig 4: Examples of the NOS skill assessment skill scores for four different locations and ports, for the Bayonne Bridge, NJ – NYOFS; Baltimore, MD – CBOFS; Buchman Bridge, FL – SJROFS; and Eagle Point, TX – GBOFS and for the twelve month period of calendar year 2008. Three statistical parameters are shown, from top-to-bottom, for the CF of water level amplitudes, RMSE of water level amplitudes, and for the CF of the amplitude of high water and low water events

Observed Current (R=0.85)			Modeled Current (R=0.92)			Difference	
N	Constituent	Amplitude	Epoch	Amplitude	Epoch	Amplitude	Epoch
CURRENT ALONG PCD: DIRECTION = 272 DIRECTION = 282							
1	M2	0.580	210.6	0.607	213.4	0.027	2.8
2	S2	0.137	218.4	0.141	220.7	0.004	2.3
3	N2	0.123	184.9	0.129	187.0	0.006	2.1
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
33	L2	0.016	225.8	0.019	225.5	0.003	-0.3
34	2MK3	0.014	96.7	0.019	91.5	0.005	-5.2
35	K2	0.040	209.9	0.041	211.6	0.001	1.7
36	M8	0.000	207.4	0.000	207.5	0.000	0.1
37	MS4	0.010	154.3	0.012	147.2	0.002	-7.1
CURRENT ACROSS PCD: DIRECTION = 92 DIRECTION = 102							
1	M2	0.580	210.6	0.607	213.4	0.027	2.8
2	S2	0.137	218.4	0.141	220.7	0.004	2.3
3	N2	0.123	184.9	0.129	187.0	0.006	2.1
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
33	L2	0.016	225.8	0.019	225.5	0.003	-0.3
34	2MK3	0.014	96.7	0.019	91.5	0.005	-5.2
35	K2	0.040	209.9	0.041	211.6	0.001	1.7
36	M8	0.000	207.4	0.000	207.5	0.000	0.1
37	MS4	0.010	154.3	0.012	147.2	0.002	-7.1

Table 5: Comparison of tidal constituent amplitudes and epochs for TIDAL CURRENTS. In the abbreviated table, the amplitudes for currents are in m/s and the epochs are in degrees. Note that currents require PCD and values of R. Tidal epoch may be local or Greenwich, but must be consistent

## INTERPRETATION OF SKILL SCORES

At this time, NOS has not carried out an in-depth investigation of OFS errors using the skill scores. The authors are aware that errors can arise from four major sources: inaccuracies in the prediction of astronomical tides at the ocean boundaries; inaccuracies in the prediction of non-tidal forcing at the ocean boundaries; inaccuracies in a hydrodynamic model's simulation of the propagation of long waves into an estuarine region; and inaccuracies in the predicted local wind field. Generally, errors in water levels increase (decreasing CF and increasing RMSE) with distance from the entrance or, more accurately, the location of ocean forcing. Some exceptions occur for stations located up rivers or in small navigation channels, regardless of their distance from the ocean forcing. Also, errors are expected to grow as the forecast goes further into the future, but not always. For example, in the upper bay stations in the Chesapeake Bay (eg, Baltimore), the forecast water levels improve slightly in the first 6h to 12h. This may be due to the delayed influence of water level boundary conditions which require several hours for their influence to be felt in the upper bay.<sup>4</sup>

The daily operation of nine operational systems obviously generates a large amount of data, and except for the skill software NOS has not yet developed the tools for an intensive analysis of this information. However, the status of each system is being assessed in real time, and model run output and observations are being archived for future analysis. As

described,<sup>21</sup> NOS intends to analyse and document the skill of each forecast system (using the NOS skill assessment methodology) on a periodic basis, and plan for system improvements.

## SUMMARY

This paper discusses the standards and procedures for skill assessment of NOS operational hydrodynamic and estuarine operational forecast systems. These operational systems were developed primarily to assist the Coast Survey's mission of supporting marine transportation. The components of skill assessment include: time series of observed and predicted variables, data analysis techniques, model run scenarios, statistical variables, target values, and acceptance criteria. The comprehensive set of skill assessment statistics are calculated to provide relevant information on model behaviour and global assessment of errors. The standard suite of statistics includes the SM and the frequency with which errors lie within specified limits (CF), the RMSE and SD of error, the positive outlier frequency (POF), the maximum duration of positive outlier (MDPO), the negative outlier frequency (NOF) and maximum duration of negative outlier (MDNO).

The target frequency for the standard statistics and acceptance error limits for the time series are also given for reference to an application of a particular model system skill assessment. The corresponding skill assessment software has been developed to compute the skill assessment scores using

data files containing observed and modelled variables. This software ingests model forecasts and observational data, fills in gaps and missing data, creates specific time series such as concatenated forecasts, extracts skill-specific parameters such as times and amplitudes of high and low water, and computes the skill scores. This software has been tested and applied in all NOS operational forecast systems. It has also been ported to other users outside of NOAA.

Currently, this software works only for time series of water levels, currents, water temperature, and salinity at a specific vertical layer for a particular location. Many other variables and issues have not been covered, eg, evaluation of model forecast skill with respect to horizontal and vertical stratification, vertical profiles of properties, particle tracking, passive tracer dispersion, sediment transport, and so on. Therefore, there are opportunities to extend the current NOS skill assessment software in the future. Additional skill scores may be needed in large fresh water bodies such as the Great Lakes, and for rivers. Finally, as new approaches to skill assessment are developed, NOS plans to frequently re-evaluate its methods and practices and add some measures which inform on individual model performance.

## ACKNOWLEDGEMENTS

The following people have significantly contributed to the development of the scope of the NOS skill assessment approach and to the testing of the software and procedures. In alphabetical order they are John Kelley, Edward Myers, Richard Patchen, Richard Schmalz, Eugene Wei and Mark Vincent.

## REFERENCES

1. Aikman III F, Vincent M and Patchen R. 2007. *Development and evolution of operational forecast systems for the coastal and estuarine environment in NOAA's National Ocean Service*. Estuarine and Coastal Modeling, American Society of Civil Engineers, ML. Spaulding (Ed), Proceedings of the 10th International Conference, Newport, RI, November 5–7, 2007, 671–684.
2. Parker BB. 1998. *Development of model-based regional nowcasting/forecasting systems*. In: Spaulding ML and Blumberg AF (Eds), Estuarine and Coastal Modeling, American Society of Civil Engineers, New York, refereed proceedings of the 5<sup>th</sup> International Conference, Alexandria, VA, p 355–373.
3. Aikman III F and Rao DB. 1999. *A NOAA perspective on a coastal forecast system*. Coastal Ocean Prediction. Christopher NK Mooers (ed), American Geophysical Union, Coastal & Estuarine Studies, 56, 467–499.
4. Bosley KT and Hess KW. 2001. *Comparison of statistical and model-based hindcasts of subtidal water levels in Chesapeake Bay*. Journal of Geophysical Research, Oceans, 106(8), 16,869–16,885.
5. Gross TF, Bosley KT and Hess KW. 2000. *The Chesapeake Bay Operational Forecast System (CBOFS): Technical Documentation*. NOS Technical Report OCS/CO-OPS 1, 69p.
6. Hess KW. 2000. *MECCA 2 program documentation*. NOAA Technical Report NOS CS 5. 49p.
7. Wei E. 2003. *The New Port of New York and New Jersey Operational Forecast System NYOFS*. Bulletin of the American Meteorological Society, 84(94), 1184–1186.
8. Wei E and Chen M. 2002. *NOS experimental nowcast/forecast system for the Port of New York/New Jersey (NYEFS): Requirements, overview and skill assessment*. NOAA Technical Report NOS CS 15, 53p.
9. Blumberg AF and Mellor GL. 1987. *A description of a three-dimensional coastal ocean circulation model*. Three-dimensional coastal ocean models, Heaps (ed), American Geophysical Union, Washington, DC. 1–16.
10. Schmalz RA and Richardson PH. 2002. *Evaluation of the NOS experimental nowcast/forecast system for Galveston Bay*. NOAA Technical Report NOS CS 14, 107p.
11. Myers E, Aikman III F and Zhang A. 2004. *A forecast circulation model for the St Johns River, Florida*. Proceedings, 8<sup>th</sup> International Estuarine and Coastal Modeling Conference, ASCE, Malcolm Spaulding (ed), 144–156.
12. Hamrick JM. 1992. *A three-dimensional environmental fluid dynamics computational code: Theoretical and computational aspects*. Special Rept. 317, The College of William and Mary, Virginia Institute of Marine Sciences, VA.
13. Sucusy PV and Morris FW. 2001. *Calibration of a three-dimensional circulation and mixing model of the Lower St Johns River*. Technical Memorandum, St Johns River Water Management District, Palatka, Florida, 97p.
14. Schwab DJ and Bedford KW. 1994. *Initial implementation of the Great Lakes forecasting system: a real-time system for predicting lake circulation and thermal structure*. Water Pollution Research Journal of Canada, 29(2/3), 203–220.
15. Schwab D J and Bedford KW. 1999. *The Great Lakes forecasting system*. Coastal Ocean Prediction. Christopher NK Mooers (ed), American Geophysical Union, Coastal & Estuarine Studies, 56, 157–173.
16. Zhang Y-L, Baptista AM and Myers EP. 2004. *A cross-scale model for 3D baroclinic circulation in estuary-plume-shelf systems: I. Formulation and skill assessment*. Continental Shelf Research, 24, 2187–2214.
17. Shchepetkin AF and McWilliams JC. 2004. *The regional oceanic modeling system: a split-explicit, free-surface, topography-following-coordinate ocean model*. Ocean Modelling, 9 (4), 347–404.
18. Haidvogel DP, Arango H, Budgell WP, Cornuell BD, Curchister E, Di Lorenzo E, Fennel K, Geyer WR, Hermann AJ, Lanerolle L, Levin J, McWilliams JC, Miller AJ, Moore AM, Powell TM, Shchepetkin AF, Sherwood CR, Signell RP, Warner JC and Wilkin J. 2008. *Ocean forecasting in terrain-following coordinates: Formulation and skill assessment of the regional ocean modeling system*. Journal of Computational Physics, 227 (7), 3595–3624.
19. NOS. 1999. *NOS procedures for developing and implementing operational nowcast and forecast systems for ports*. NOAA Technical Report NOS CO-OPS 0020, 33p.
20. Hess KW, Gross TF, Schmalz RA, Kelley JGW, Aikman III F, Wei E and Vincent MS. 2003. *NOS standards for evaluating operational nowcast and forecast hydrodynamic model systems*. NOAA Technical Report, NOS CS 17, 48p.

21. Vincent M, Hess KW and Kelley JG. 2003. NOS procedures for developing and implementing operational nowcast and forecast hydrodynamic model systems. NOAA Technical Report NOS CO-OPS 39, 17 p.
22. ASCE. 2009. *Verification and validation of 3D free-surface flow models*. Edited by Sam SY Wang *et al.* American Society of Civil Engineers (ASCE) Environmental and Water Resources Institute (EWRI). 1801 Alexander Bell Drive, Reston, VA 20191-4400.
23. Dee DP. 1995. *A pragmatic approach to model validation*. In: Lynch DR and Davies AM (eds), *Quantitative skill assessment of coastal ocean models*, AGU, Washington. p1-13.
24. Refsgaard JC. 2001. *Towards a formal approach to calibration and validation of models using spatial data*. In: Grayson R and Blöschl G (eds), *Spatial patterns in catchment hydrology: Observations and modeling*, Cambridge University Press, p329-354.
25. Refsgaard JC and Henriksen HJ. 2004. *Modeling guidelines – terminology and guiding principles*. *Advances in Water Resources*, 27(1), 71-82.
26. Vested HJ, Nielson JW, Jensen HR and Kristensen KB. 1995. *Skill assessment of an operational hydrodynamic forecast system for the North Sea and Danish Belts*. In: Lynch DR and Davies AM (eds), *Quantitative Skill Assessment of Coastal Ocean Models*, AGU, Washington. p373-396.
27. Wilmott CJ. 1981. *On the validation of models*. *Physical Geography*, 2, 184-194.
28. Wilmott CJ, Ackleson SG, Davis RT, Feddema JJ, Klink KM, Legates DR, O'Donnell J and Rowe CM. 1985. *Statistics for evaluation and comparison of models*. *Journal of Geophysical Research*, 90, C5, 8995-9005.
29. Warner JC, Geyer WR and Lerczak JA. 2005. *Numerical modeling of an estuary: A comprehensive skill assessment*. *Journal of Geophysical Research*, 110(5), C05001.
30. Dingman JS and Bedford KW. 1986. *Skill tests and parametric statistics for model evaluation*. *Journal of Hydraulic Engineering*, 112 (2), 124-141.
31. Werner FE. 1995. *A field test case for tidally forced flows: a review of the tidal forum*. In: Lynch DR and Davies AM (eds), *Quantitative Skill Assessment of Coastal Ocean Models*, AGU, Washington. p269-283.
32. Robinson AR, Lermusiaux PFJ, Haley Jr PJ and Leslie WG. 2002. *Predictive skill, predictive capability and predictability in ocean forecasting*. *Proc. Oceans 2002 IEEE/MITS Conf*, 787-794.
33. Lynch DR, McGillicuddy Jr DJ and Werner FE. 2009. *Skill assessment for coupled biological/physical models of marine systems*. *Journal of Marine Systems*, 76, 1-3.
34. Stow CA, Jolliff J, McGillicuddy DJ, Doney SC, Allen JI, Friedrichs MAM, Rose KA and Wallhead P. 2009. *Skill assessment for coupled biological/physical models of marine systems*. *Journal of Marine Systems*, 76, 4-15.
35. Gregg WW, Friedrichs MAM, Robinson AR, Rose KA, Schlitzer R, Thompson KR and Doney SC. 2009. *Skill assessment in ocean biological data assimilation*. *Journal of Marine Systems*, 76, 16-33.
36. Wallhead PJ, Martin AP, Srokosz MA and Franks PJS. 2009. *Skill assessment via cross-validation and Monte Carlo simulation: An application to Georges Bank*. *Journal of Marine Systems*, 76, 134-150.
37. Stumpf RP, Tomlinson MC, Calkins JA, Kirkpatrick B, Fisher K, Nierenberg K, Currier R and Wynne TT. 2009. *Skill assessment for an operational algal bloom forecast system*. *Journal of Marine Systems*, 76, 151-161.
38. Fitzpatrick JJ. 2009. *Assessment skill of estuarine and coastal eutrophication models for water quality managers*. *Journal of Marine Systems*, 76, 195-211.
39. Sheng P and Kim T. 2009. *Skill assessment of an integrated modeling system for shallow coastal and estuarine ecosystems*. *Journal of Marine Systems*, 76, 212-243.
40. Williams RG, Bethem TD and Frey HR. 1989. Tampa Bay current prediction quality assurance miniproject. NOAA Technical Memorandum NOS OMA 50, National Ocean Service, Office of Oceanography and Marine Assessment, Rockville, Maryland, 11p.
41. Zhang A, Hess KW, Wei E and Myers E. 2006. Implementation of model skill assessment software for water level and current in tidal regions. NOAA Technical Report, NOS CS 24, 61p.
42. Zervas C. 1999. Tidal current analysis procedures and associated computer programs. NOAA Technical Report NOS CO-OPS 21, 101p.
43. Dennis RE and Long EE. 1971. A user's guide to a computer program for harmonic analysis of data at tidal frequencies. NOAA Technical Report, NOS 41, 31p.
44. Kelley JGW, Zhang A, Chu P and Lang G. 2008. Skill assessment of NOS Lake Ontario Operational Forecast System (LOOFS). NOAA Technical Memorandum NOS CS 13 40p.
45. Preisendorfer RW. 1988. Principal component analysis in meteorology and oceanography. *Developments in Atmospheric Science*, v17. Elsevier, Amsterdam. p425.