



Center for Operational Oceanographic Products and Services
NATIONAL OCEAN SERVICE

Validation of Skill Assessment Software for NOS Operational Oceanographic Modeling Systems

Luke Judd

Aijun Zhang, Carolyn Lindley, Lucila Bloemendaal

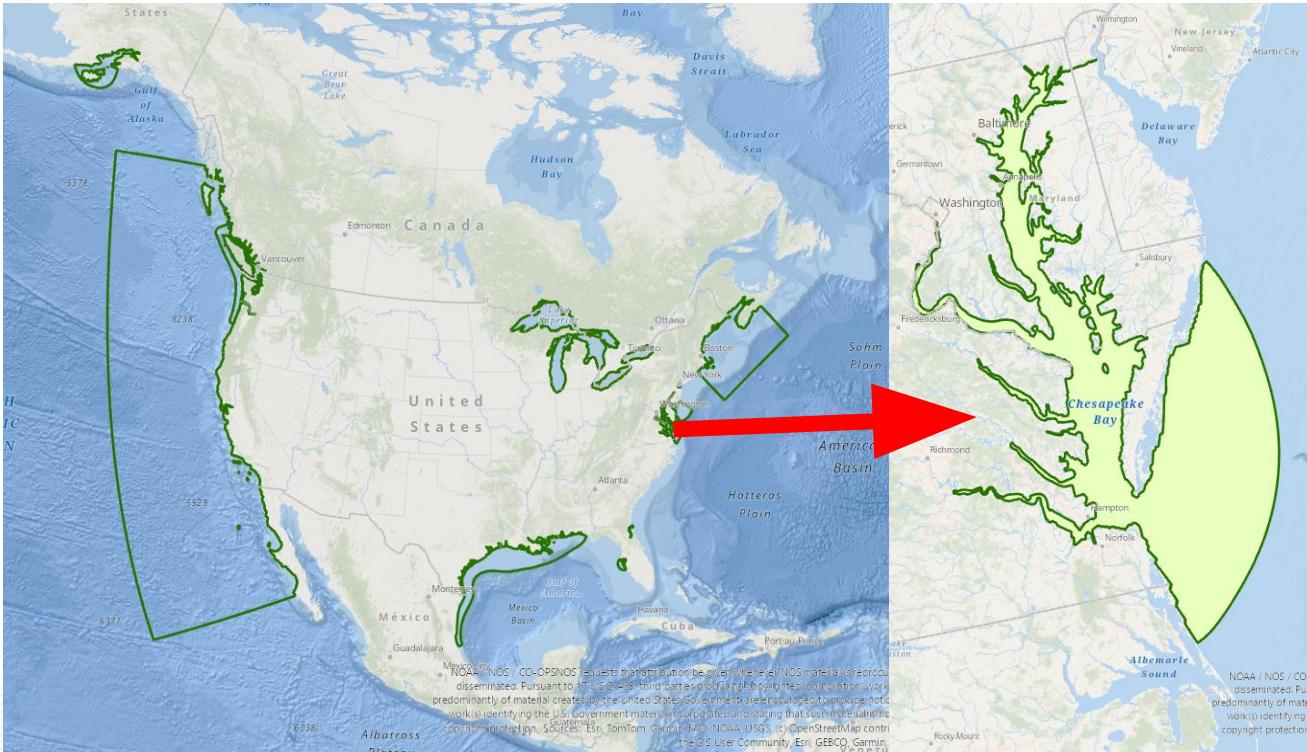
NOAA/NOS

CO-OPS Oceanographic Division

June - Aug, 2025



OFS Overview



Major Services and Stakeholders

Safe & efficient navigation

Navigation management (ports managers, pilots, etc.)

Marine Channel Forecasting

Water levels for under-keel clearance

Currents for USCG right-of-way Decision-making and ship maneuverability

Emergency response (OR&R, USCG, FEMA)

Inundation dashboard

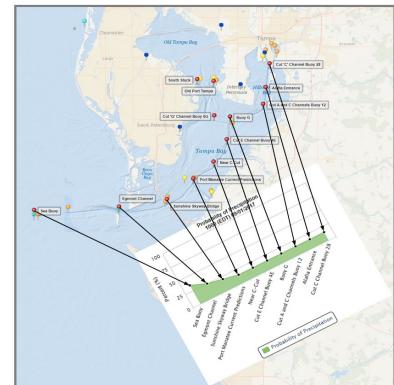
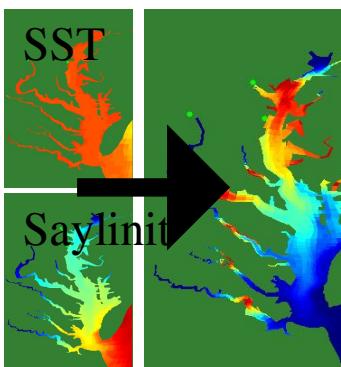
Hazardous spills

Search & Rescue

Enable management of the coastal zone

Ecosystem and water quality modeling

Coastal and fishery management



NOS Operational Forecast System Evaluation

(Standardized Skill Assessment Software)

Objective

- Measure the performance of model simulations (including tidal simulation, hindcast, nowcast, and forecast) by comparing with observations
- All models should be assessed and pass the NOS skill assessment standards before transitioning to operation

Functions

- Data acquisition and processing (observations and model outputs)
 - Time interval conversion and Gap-filling
 - Concatenation of model outputs
 - Filtering
 - Tidal harmonic analysis and prediction
 - Extracting extremes/events and slack waters
- Compute statistical variables
- Generate skill assessment score tables

NOAA Technical Report NOS CS 24

IMPLEMENTATION OF MODEL SKILL ASSESSMENT
SOFTWARE FOR WATER LEVEL AND CURRENT IN
TIDAL REGIONS

Aijun Zhang
Kurt W. Hess
Eugene Wei
Edwards Myers

March 2006



noaa National Oceanic and Atmospheric Administration

U.S. DEPARTMENT
OF COMMERCE
Gale E. Norton, Secretary
Office of Coast Survey
Captain Roger L. Parsons, NOAA

National Oceanic and
Atmospheric Administration
John H. Dunnigan,
Administrator, Under Secretary
VADM USN (Ret.), Under Secretary

National Ocean Service
John H. Dunnigan,
Administrator, Under Secretary
Coast Survey Development
Laboratory
Mary Erickson

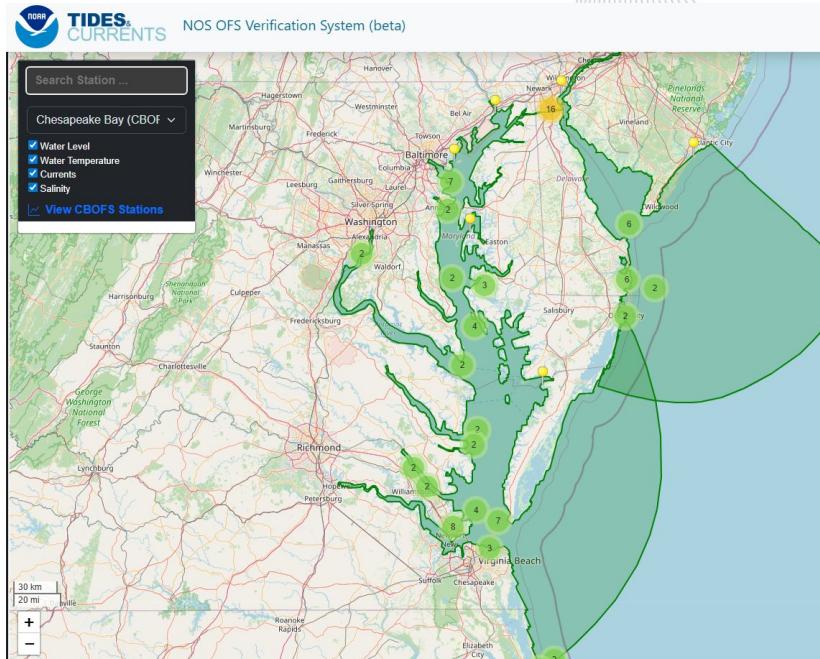
Statistical Variable Definition

<u>Variable</u>	<u>Explanation</u>
Error	The error = the predicted value - the observed
RMSE	Root Mean Square Error
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. MDPO is the length of time of the longest Positive Outliers event.
MDNO(X)	Maximum Duration of Negative Outliers. MDNO is the length of time of the longest Negative Outliers event.
WOF(X)	Worst Case Outlier Frequency. Fraction (percentage) of errors that exceed X , and model forecast is worse than tidal prediction.

Next Gen Skill Assessment System

Funded by BIL, NOS Shared Cyber-Infrastructure and Skill Assessment (SCI-SA) project

- More observation data, larger OFS boundaries, different core models
- Aims to replicate Fortran package outputs and accuracy
- Publicly available
 - <https://github.com/NOAA-CO-OPS/Next-Gen-NOS-OFS-Skill-Assessment>
- Active development cycles/WebApp development
- Simplified interface/setup and easier modification of run types



Project Timeline and Goals

Overall Goal: Validate Python package against Fortran package outputs, identify possible sources of errors/bugs

Weeks 1-2:

- Setup server access, begin running both packages. Understand different assessment calls and compare output formats

Weeks 3-4:

- Begin 6 month package runs, research into plotting and statistical methods. First scripts to convert data into useable format

Weeks 5-6:

- Comparison and plotting scripts, first look at accuracy scores. Identification of mismatch in model outputs, subsequent package updates and fixes

Weeks 7-8:

- Updated package runs, new analysis, expanding comparison methods. Script polishing and concatenation, final 6 month comparison run

Weeks 9-10:

- Abstract preparation, CO-OPS presentation, project documentation

Result Examples from the 2 Packages

ID	NODE	obs_water_depth	mod_water_depth	rmse	r	bias	bias_perc	bias_dir	central_freq	pos_outlier_freq	neg_outlier_freq	bias_standard_dev		
0	8571421	29	0	0.1	1.33	0.99	0.32	2.27	97.03	pass	0	pass	1.3	
1	8571892	22	0	0.1	1.14	1	0.68	5.11	99.47	pass	0	pass	0.91	
2	8573364	10	0	0.4	0.97	0.99	0.64	3.56	100	pass	0	pass	0.73	
3	8573927	1	0	0.3	0.82	1	0.11	0.85	100	pass	0	pass	0.81	
4	8575512	14	0	0.1	2.08	1	-1.71	-11.09	89.72	fail	0	pass	1.18	
5	8577330	27	0	0.1	0.88	1	0.2	1.5	99.65	pass	0	pass	0.86	
6	8632200	47	0	0.2	0.83	0.99	0.02	0.14	99.49	pass	0	pass	0.83	
7	8637689	45	0	0.4	1.58	0.99	1	7.04	95.1	pass	0	pass	1.23	
8	8638610	60	0	0.2	1.24	0.99	0.67	4.67	98.49	pass	0	pass	1.05	
9	44041	191	0	0.2	1.18	1	0.42	2.69	99.97	pass	0	pass	1.11	
10	44042	162	1	0.8	0.67	1	0.4	3.11	99.96	pass	0	pass	0.54	
11	44043	158	1	0.8	1.39	1	1.27	7.86	99.86	pass	0	pass	0.57	
12	44058	163	1	0.9	0.76	1	0.4	2.9	99.99	pass	0	pass	0.65	
13	44062	161	1	1	0.68	1	0.43	3.8	100	pass	0	pass	0.53	
14	44063	159	1	1	1.05	1	0.75	5.91	100	pass	0	pass	0.73	
15	44064	166	1	1.1	0.8	1	0.49	3.65	99.63	pass	0.03	pass	0.62	
16	44072	164	1	0.9	0.52	0.98	0.29	3.61	100	pass	0	pass	0.43	
17	CBCM2	9	0	0.3	1.4	0.99	1.18	5.88	97.63	pass	0	pass	0.75	
SCENARIO: SEMI-OPERATIONAL NOWCAST														
T			42831	13.435										
t			42831	13.188										
T-t	3.0	c	24h	42831	0.247	0.573	0.517	0.0	100.0	0.0	0.0	0.0	1.00	1.00
SCENARIO: SEMI-OPERATIONAL FORECAST														
T006-t006	3.0	c	24h	718	0.183	0.557	0.527	0.0	100.0	0.0	0.0	0.0	1.00	
T012-t012	3.0	c	24h	718	0.195	0.564	0.529	0.0	100.0	0.0	0.0	0.0	1.00	
T018-t018	3.0	c	24h	718	0.191	0.574	0.542	0.0	100.0	0.0	0.0	0.0	1.00	
T024-t024	3.0	c	24h	717	0.199	0.581	0.546	0.0	100.0	0.0	0.0	0.0	1.00	
T030-t030	3.0	c	24h	719	0.194	0.580	0.547	0.0	100.0	0.0	0.0	0.0	1.00	
T036-t036	3.0	c	24h	717	0.181	0.560	0.530	0.0	100.0	0.0	0.0	0.0	1.00	
T042-t042	3.0	c	24h	721	0.170	0.566	0.540	0.0	100.0	0.0	0.0	0.0	1.00	
T048-t048	3.0	c	24h	718	0.182	0.573	0.543	0.0	100.0	0.0	0.0	0.0	1.00	

Analysis Breakdown

Comparison Script does the following steps based on user input:

1. Converts all fortran skill assessment tables to CSVs that match python formatting
2. Runs comparison module for forecast and nowcast, generates overall skill summary
 - a. Variable mean at each station and across OFS
 - b. Statistical tests on each variable output
 - c. Pass Fail flag verification
3. Generates time series plots of python vs fortran package extracted model outputs for each station, variable, call

- ★ Conversion script preprocessing for WL station offsets and fortran time series shift

scenario	variable	mean_pass	mean_total	mean_accuracy
nowcast	currents	6	8	0.75
nowcast	salinity	7	8	0.875
nowcast	water_level	5	8	0.625
nowcast	water_temperature	8	8	1
forecast	currents	3	8	0.375
forecast	salinity	7	8	0.875
forecast	water_level	4	8	0.5
forecast	water_temperature	7	8	0.875
forecast	ALL	21	32	0.65625
nowcast	ALL	26	32	0.8125
Total Mean Accuracy:		47	64	0.734

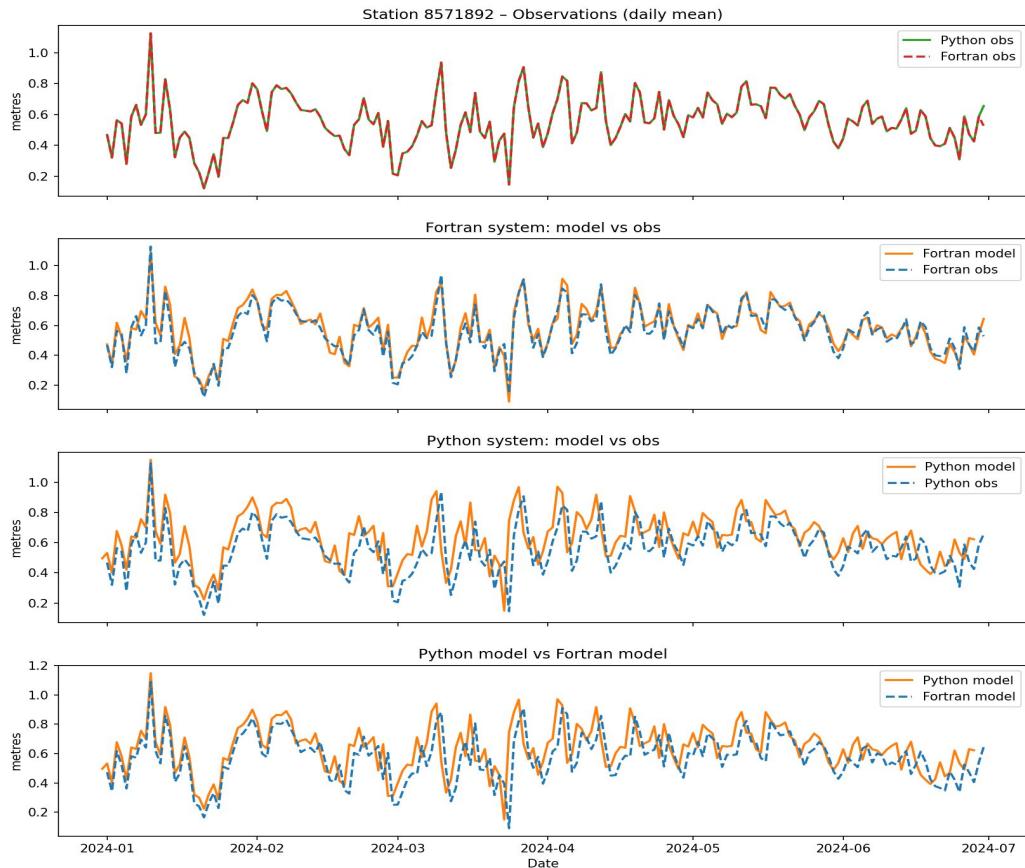
scenario	variable	stats_pass	stats_total	stats_accuracy
nowcast	currents	0	8	0
nowcast	salinity	5	8	0.625
nowcast	water_level	4	8	0.5
nowcast	water_temperature	6	8	0.75
forecast	currents	0	8	0
forecast	salinity	5	8	0.625
forecast	water_level	3	8	0.375
forecast	water_temperature	5	8	0.625
forecast	ALL	13	32	0.40625
nowcast	ALL	15	32	0.46875
Total Stats Accuracy:		28	64	0.438

Analysis Breakdown

- Each suite of tests is run for every variable at every station in both nowcast and forecast
- 8 tables and 152 “equivalent” tags to analyze

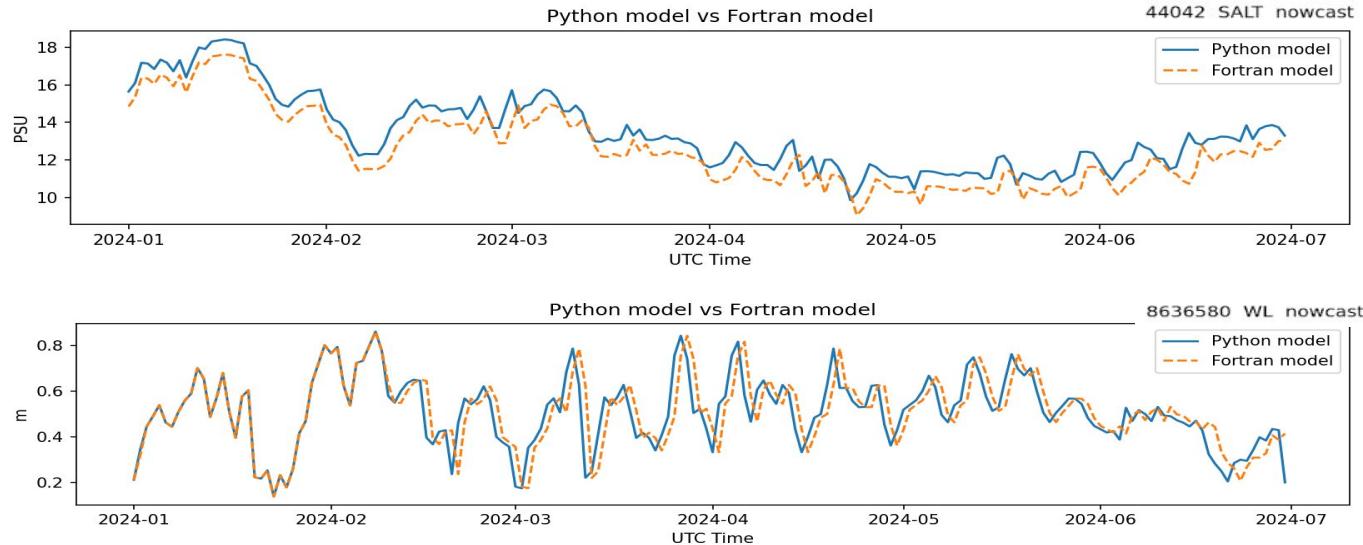
variable	metric	n	44041_ft	44041_py	44042_ft	44042_py	pass_ratio	overall_equiv	mean_ft	mean_py
water_temperature	rmse	16	1.183	1.18	0.573	0.67	1	YES	1.0078125	1.036875
water_temperature	r	16	1	1	1	1	1	YES	0.995625	0.99625
water_temperature	bias	16	0.418	0.42	0.247	0.4	1	YES	0.3303125	0.37875
water_temperature	bias_perc	16	2.67434421	2.69	1.872914771	3.11	0.75	YES	2.458908285	2.831875
water_temperature	central_freq	16	100	99.97	100	99.96	1	YES	98.83125	98.833125
water_temperature	pos_outlier_freq	16	0	0	0	0	0.9375	YES	0	0.001875
water_temperature	neg_outlier_freq	16	0	0	0	0	1	YES	0	0
water_temperature	bias_standard_dev	16	1.107	1.11	0.517	0.54	1	YES	0.79875	0.79875
variable	metric	n	p_ttest	p_wilcoxon	p_ks	cohen_dz	p_tost	equivalent	agreement	p_mcnemar
water_temperature	rmse	16	0.007889667	0.063777302	0.99989097	0.765776602	1.51601E-24	YES		
water_temperature	r	16	0.333170136	0.317310508		1	0.25	2.2623E-21	YES	
water_temperature	bias	16	0.016756992	0.011323905	0.716411657	0.67276063	9.16777E-19	YES		
water_temperature	bias_perc	16	0.014758848	0.006286621	0.716411657	0.688567716	0.047218031	YES		
water_temperature	central_freq	16	0.901310581	0.858516744	0.95225628	0.031530423	1.00828E-30	YES		
water_temperature	pos_outlier_freq	16	0.333170136	0.317310508		1	0.25	0.833414932	NO	
water_temperature	neg_outlier_freq	16				1			NO	
water_temperature	bias_standard_dev	16		1	0.378193546	0.99989097	3.04343E-16	2.52616E-10	YES	
water_temperature	central_freq_pass_fail	16						YES	1	1
water_temperature	pos_outlier_freq_pass_fail	16						YES	1	1
water_temperature	neg_outlier_freq_pass_fail	16						YES	1	1

First Comparison



scenario	variable	mean_pass	mean_total	mean_accuracy
nowcast	currents	0	7	0
nowcast	salinity	3	7	0.4285714286
nowcast	water_level	2	7	0.2857142857
nowcast	water_temperature	6	7	0.8571428571
forecast	currents	1	7	0.1428571429
forecast	salinity	4	7	0.5714285714
forecast	water_level	4	7	0.5714285714
forecast	water_temperature	7	7	1
forecast	ALL	16	28	0.5714285714
nowcast	ALL	11	28	0.3928571429
Total Mean Accuracy:		27	56	0.482
scenario	variable	stats_pass	stats_total	stats_accuracy
nowcast	currents	0	7	0
nowcast	salinity	0	7	0
nowcast	water_level	2	7	0.2857142857
nowcast	water_temperature	3	7	0.4285714286
forecast	currents	0	7	0
forecast	salinity	2	7	0.2857142857
forecast	water_level	3	7	0.4285714286
forecast	water_temperature	5	7	0.7142857143
forecast	ALL	10	28	0.3571428571
nowcast	ALL	5	28	0.1785714286
Total Stats Accuracy:		15	56	0.268
scenario	variable	bin_pass	bin_total	bin_accuracy
nowcast	currents	2	3	0.6666666667
nowcast	salinity	3	3	1
nowcast	water_level	1	3	0.3333333333
nowcast	water_temperature	3	3	1
forecast	currents	2	3	0.6666666667
forecast	salinity	3	3	1
forecast	water_level	3	3	1
forecast	water_temperature	3	3	1
forecast	ALL	11	12	0.9166666667
nowcast	ALL	9	12	0.75
Total Bin Accuracy:		20	24	0.833

First Comparison



- Manual verification that obs data is identical
- Extracted model outputs from the two packages are different
 - Use of model outputs must be identical for an accurate comparison
- Data offsets and time shift both present

Fixing Errors, Package Updates

-Small bug with how python package handled data gaps, did not fill correctly

- Arose from missing files in NCEI Thredds

- Python package error with model data ingestion causing vertical offset

- Missing files:

- 20240209 t00z, 20240209 t06z

- 20240209 t12z, 20240209 t18z

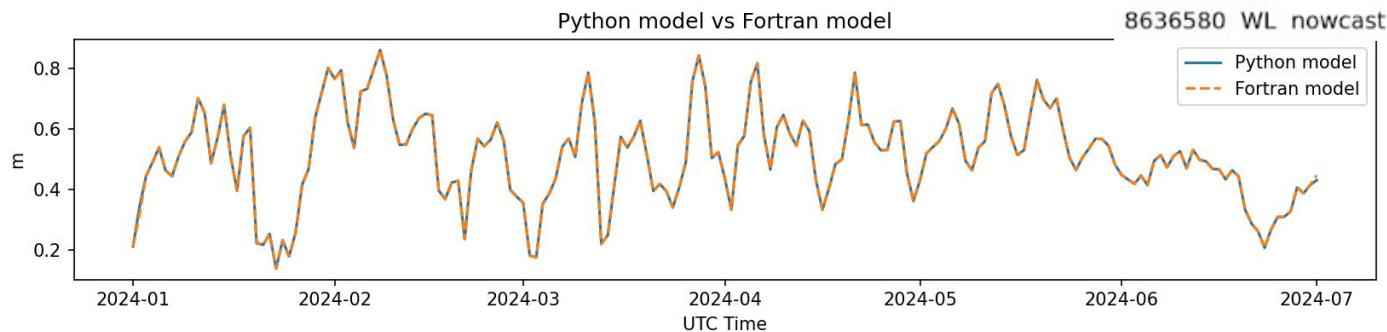
- 20240603 t06z, 20240604 t18z

- 20240605 t06z, 20240607 t12z

- 20240607 t18z

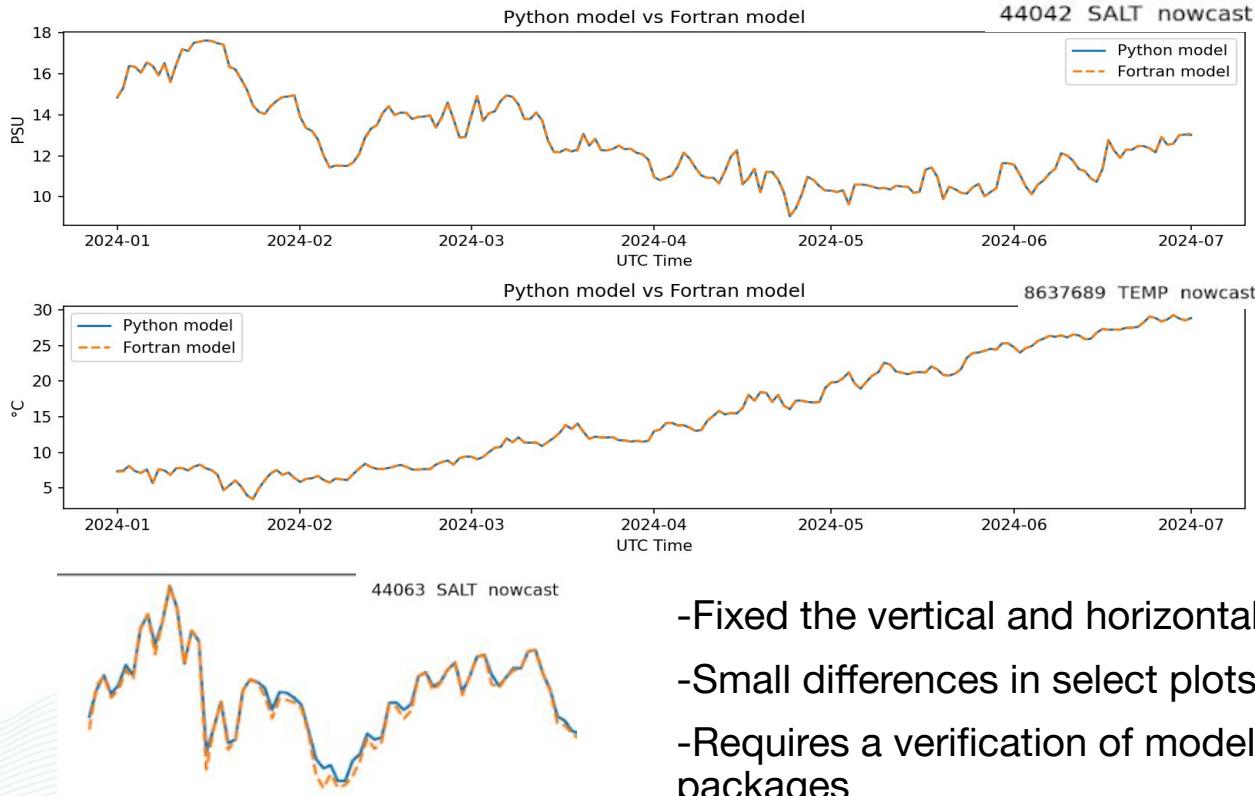
```
@@ -118,7 +118,7 @@ def format_temp_salt(prop, model, ofs_ctlfile, model_var, i):
    model[model_var][:, int(ofs_ctlfile[2][i]),
                     int(ofs_ctlfile[1][i])]
 120 120
 121 121 -     model_obs = model_obs + ofs_ctlfile[3][i]
 121 121 +     model_obs = model_obs #+ ofs_ctlfile[3][i]
 122 122   elif prop.ofsfiletype == 'stations':
 123 123     # Dimensions: time x siglay x station
 124 124     model_time = np.array(model["time"])
@@ -127,7 +127,7 @@ def format_temp_salt(prop, model, ofs_ctlfile, model_var, i):
    model[model_var][:, int(ofs_ctlfile[2][i]),
                     int(ofs_ctlfile[1][i])]
 128 128
 129 129 -     model_obs = model_obs + ofs_ctlfile[3][i]
 129 130 +     model_obs = model_obs #+ ofs_ctlfile[3][i]
 130 131   else:
 131 131     model_obs = None
 133 133   elif prop.model_source=="roms":
@@ -138,15 +138,15 @@ def format_temp_salt(prop, model, ofs_ctlfile, model_var, i):
    model_time = np.array(model["ocean_time"])
    model_obs = np.array(model[model_var][:, int(ofs_ctlfile[2][i]),
                                         i_index,j_index])
 138 138 -     model_obs = model_obs + ofs_ctlfile[3][i]
 138 141 +     model_obs = model_obs #+ ofs_ctlfile[3][i]
 141 142   elif prop.ofsfiletype == 'stations':
 142 143     # Dimensions: time x station x s_rho
 144 144     model_time = np.array(model["ocean_time"])
 145 145     if int(ofs_ctlfile[1][i]) > -999:
```

Updated Comparison



scenario	variable	mean_pass	mean_total	mean_accuracy	scenario	variable	stats_pass	stats_total	stats_accuracy	scenario	variable	bin_pass	bin_total	bin_accuracy
nowcast	currents	4	7	0.5714285714	nowcast	currents	0	7	0	nowcast	currents	3	3	1
nowcast	salinity	7	7	1	nowcast	salinity	5	7	0.7142857143	nowcast	salinity	3	3	1
nowcast	water_level	5	7	0.7142857143	nowcast	water_level	4	7	0.5714285714	nowcast	water_level	3	3	1
nowcast	water_temperature	7	7	1	nowcast	water_temperature	5	7	0.7142857143	nowcast	water_temperature	3	3	1
forecast	currents	0	7	0	forecast	currents	0	7	0	forecast	currents	3	3	1
forecast	salinity	7	7	1	forecast	salinity	5	7	0.7142857143	forecast	salinity	3	3	1
forecast	water_level	4	7	0.5714285714	forecast	water_level	3	7	0.4285714286	forecast	water_level	3	3	1
forecast	water_temperature	7	7	1	forecast	water_temperature	5	7	0.7142857143	forecast	water_temperature	3	3	1
forecast	All	18	28	0.6428571429	forecast	All	13	28	0.4642857143	forecast	All	12	12	1
nowcast	All	23	28	0.8214285714	nowcast	All	14	28	0.5	nowcast	All	12	12	1
Total Mean Accuracy:		41	56	0.732	Total Stats Accuracy:		27	56	0.482	Total Bin Accuracy:		24	24	1

Updated Comparison



- Fixed the vertical and horizontal offsets
- Small differences in select plots
- Requires a verification of model depth values across both packages

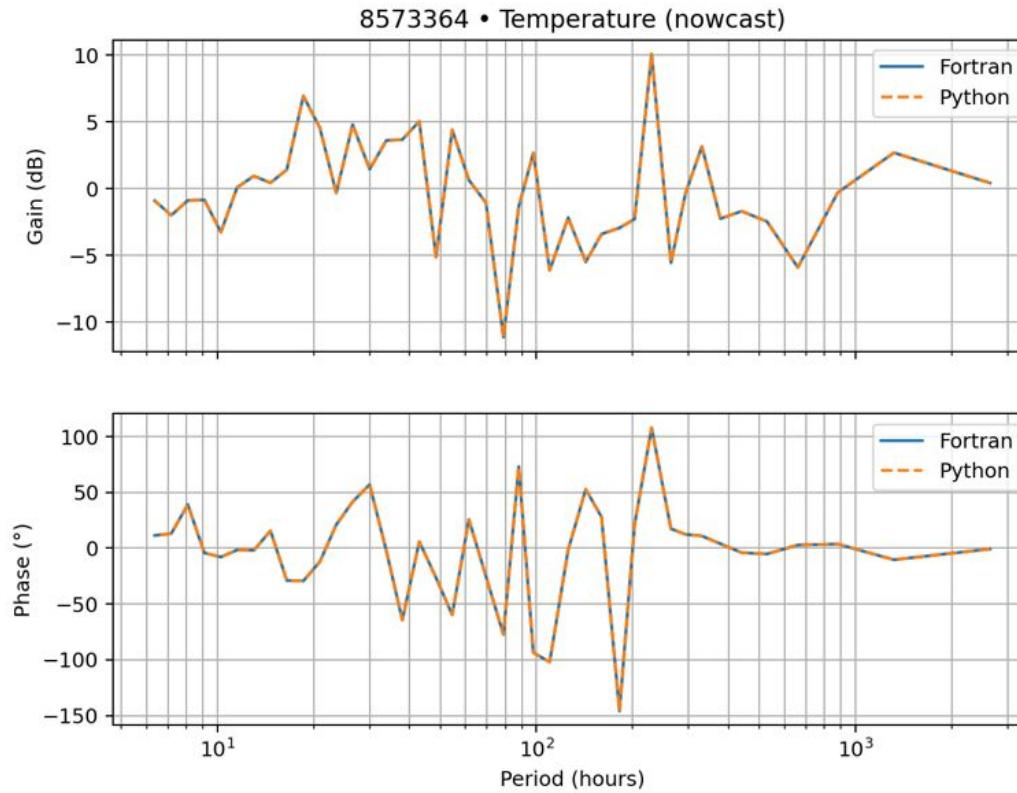
Final Comparison

8551910	8551910T_CB	"Reedy Point,DE"	-999	-999	-999	-999	8551910	0.0
39.55833	-75.57333	0.0 0.0 0.	29	19	38.227	-76.036	8571421	0.1
8573927	8573927T_CB	"Chesapeake City"	22	19	38.576	-76.069	8571892	0.1
39.52667	-75.81000	0.0 0.0 0.0	10	19	39.219	-76.249	8573364	0.4
8574680	8574680T_CB	"Baltimore"	1	19	39.528	-75.809	8573927	0.3
39.26667	-76.57833	0.0 0.0 0.0	-999	-999	-999	-999	8574680	0.0
8573364	8573364T_CB	"Tolchester"	14	19	38.985	-76.475	8575512	0.1
39.21333	-76.24500	0.0 0.0 0.0	27	19	38.318	-76.451	8577330	0.1
8571892	8571892T_CB	"Cambridge"	47	19	37.160	-76.002	8632200	0.2
38.57333	-76.06833	0.0 0.0 0.0	45	19	37.229	-76.478	8637689	0.4
8575512	8575512T_CB	"Annapolis"	60	19	36.948	-76.331	8638610	0.2
38.98333	-76.48000	0.0 0.0 0.0	-999	-999	-999	-999	8651370	0.0
8577330	8577330T_CB	"Solomons"	-999	-999	-999	-999	01482695	0.0
38.31667	-76.45167	0.0 0.0 0.0	191	19	37.218	-76.792	44041	0.2
8632200T	8632200T_CB	"Kiptopeke"	162	18	38.034	-76.337	44042	0.8
37.16500	-75.98833	0.0 0.0 0.0	158	17	39.152	-76.391	44043	0.8
8638863	8638863T_CB	"CBBT"	-999	-999	-999	-999	44056	0.0
36.96667	-76.11333	0.0 0.0 0.0	163	17	37.567	-76.257	44058	0.9
8638610	8638610T_CB	"Sewells"	161	18	38.556	-76.415	44062	1.0
36.94667	-76.33000	0.0 0.0 0.0	159	16	38.963	-76.448	44063	1.0

scenario	variable	mean_pass	mean_total	mean_accuracy
nowcast	currents	6	8	0.75
nowcast	salinity	7	8	0.875
nowcast	water_level	5	8	0.625
nowcast	water_temperature	8	8	1
forecast	currents	3	8	0.375
forecast	salinity	7	8	0.875
forecast	water_level	4	8	0.5
forecast	water_temperature	7	8	0.875
forecast	ALL	21	32	0.65625
nowcast	ALL	26	32	0.8125
Total Mean Accuracy:		47	64	0.734

- Verified that correct data (6 hour forecast) was being pulled from fortran tables
- Manually altered offset values
- Verified that python package was pulling correct model depth values
- Added comparison of bias percent statistic
- Overall accuracy saw slight increase

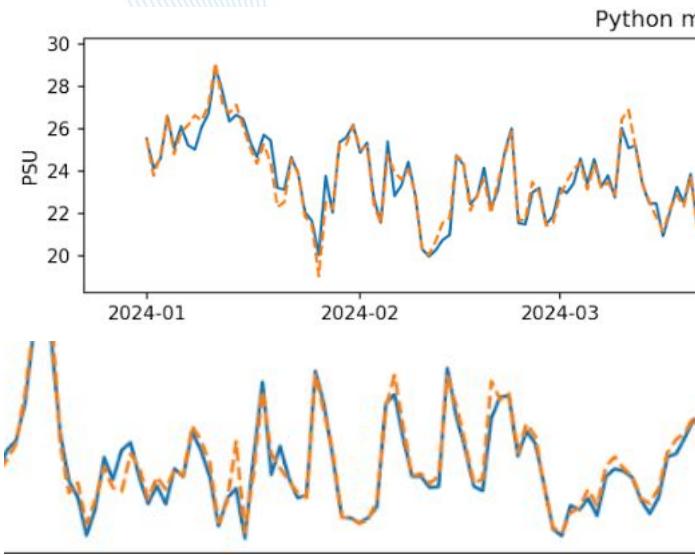
Bode Plots



Results + Disclaimers

- Time series plots are identical, supported by Bode plots
- Pass fail flags are 100 percent accurate
 - Numerical value may be different as calculation methods or thresholds are not consistent
- Statistical test accuracy is lower than mean comparison
 - TOST relies on a predefined relative tolerance value, updated in last comparison
- Further analysis needed of forecast methods, see next slide

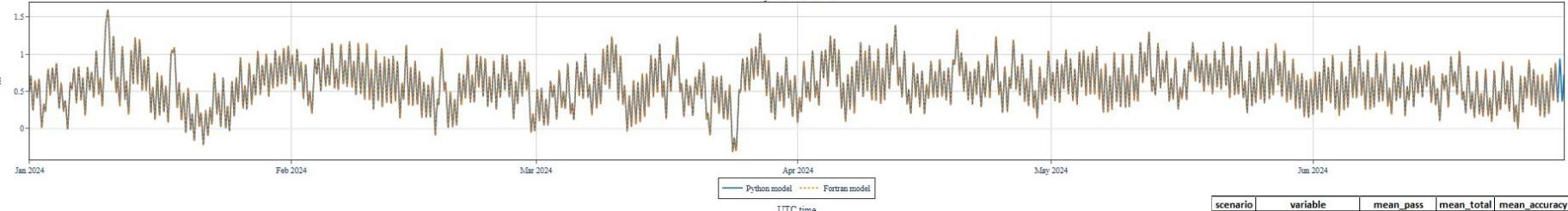
scenario	variable	stats_pass	stats_total	stats_accuracy
nowcast	currents	0	8	0
nowcast	salinity	5	8	0.625
nowcast	water_level	4	8	0.5
nowcast	water_temperature	6	8	0.75
forecast	currents	0	8	0
forecast	salinity	5	8	0.625
forecast	water_level	3	8	0.375
forecast	water_temperature	5	8	0.625
forecast	ALL	13	32	0.40625
nowcast	ALL	15	32	0.46875
Total Stats Accuracy:		28	64	0.438



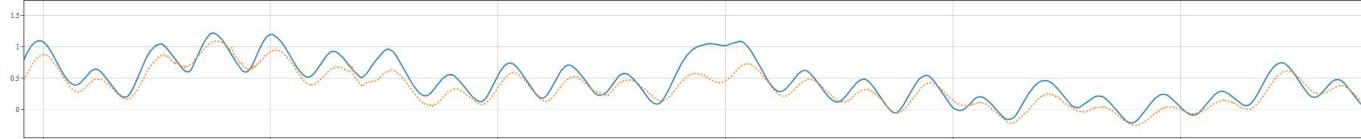
Conclusion and Next Steps

- Current errors and updates needed to the Python Package:
 - Add MSL datum to avoid conversion errors
 - Confirm that model depth is being pulled correctly at stations showing 0.0m
 - Fix gap filling error occurring when files are missing
- Modeling team is working on capabilities for different forecast horizons
 - Fortran runs for 6 hour periods from 6-48 hours
 - Updates in next python dev cycle include graphical representation of outputs for different forecast horizons
- Future Comparisons:
 - Ideally testing continues for other OFS across different time periods
 - This project ran mainly at cbofs for Jan - Jun 2024
 - Script is dynamic and has been tested at dbofs

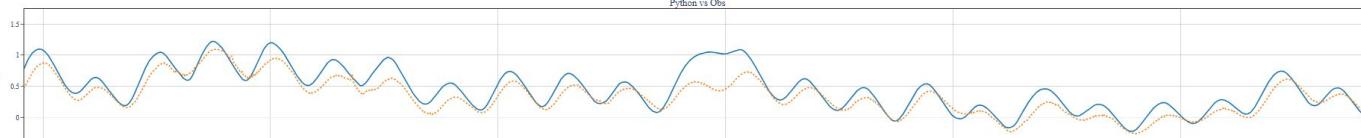
Python vs Fortran



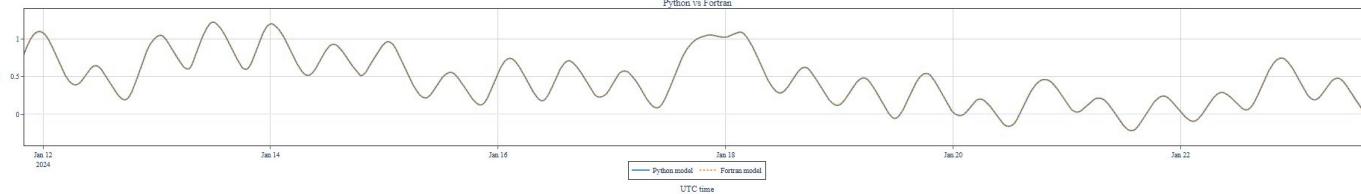
Fortran vs Obs



Python vs Obs



Python vs Fortran



scenario	variable	mean_pass	mean_total	mean_accuracy
nowcast	currents	6	8	0.75
nowcast	salinity	7	8	0.875
nowcast	water_level	6	8	0.75
nowcast	water_temperature	8	8	1
forecast	currents	4	8	0.5
forecast	salinity	7	8	0.875
forecast	water_level	6	8	0.75
forecast	water_temperature	7	8	0.875
forecast	All	24	32	0.75
nowcast	All	27	32	0.84375
Total Mean Accuracy:		51	64	0.797

scenario	variable	stats_pass	stats_total	stats_accuracy
nowcast	currents	2	8	0.25
nowcast	salinity	4	8	0.5
nowcast	water_level	4	8	0.5
nowcast	water_temperature	6	8	0.75
forecast	currents	3	8	0.375
forecast	salinity	4	8	0.5
forecast	water_level	3	8	0.375
forecast	water_temperature	7	8	0.875
forecast	All	17	32	0.53125
nowcast	All	16	32	0.5
Total Stats Accuracy:		33	64	0.516

scenario	variable	bin_pass	bin_total	bin_accuracy
nowcast	currents	3	3	1
nowcast	salinity	3	3	1
nowcast	water_level	3	3	1
nowcast	water_temperature	3	3	1
forecast	currents	3	3	1
forecast	salinity	3	3	1
forecast	water_level	3	3	1
forecast	water_temperature	3	3	1
forecast	All	12	12	1
nowcast	All	12	12	1
Total Bin Accuracy:		24	24	1
scenario variable mean_accuracy				
Total Accuracy:				0.711

Special Thanks

Dr. Roohr and the Lapenta Team at NOAA

My Mentors: Carolyn Lindley, Aijun Zhang, and Lucila Bloemendaal

The CO-OPS Modeling Team and Oceanographic Division

I've felt so welcomed at NOAA and learned so much this summer. You are all doing incredible work and inspiring the next generation of scientists, and I am honored to have worked alongside you all.



Center for Operational Oceanographic Products and Services

NATIONAL OCEAN SERVICE



Questions?