

Evaluation Report: HYCOM Forced by NAVGEM & NOGAPS Comparisons for June and July 2013 Using AUTOMETRICS Data

Frank Bub, NP1M, NAVOCEANO

21 August 2013

Summary of Results: We saw no major differences between the two HYCOM versions based on an AutoMetrics evaluation using data from the months of June and July 2013.

Bottom line: the score card data indicate that NAVGEM forcing results in an improved HYCOM.

Recommendation: Proceed with the transition to forcing HYCOM with the NRL-calibrated NAVGEM fields. NAVOCEANO implemented NAVGEM forcing on the operational HYCOM on Tuesday, 20 August 2013.

As with the HYCOM OPEVAL, a score card approach was used to compare 5 physical properties: sonic layer depth (SLD), along with temperature and salinity at the surface (SST, SSS) and 100m (T100, S100). Metrics were (a) model minus observed BIAS, (b) correlation coefficient, (c) root mean square difference (RMSD), and (d) tolerance or percentage of differences within a given range. The 6 HYCOM regions of Navy interest (Figure 1) resulted in the metrics in Table 1.

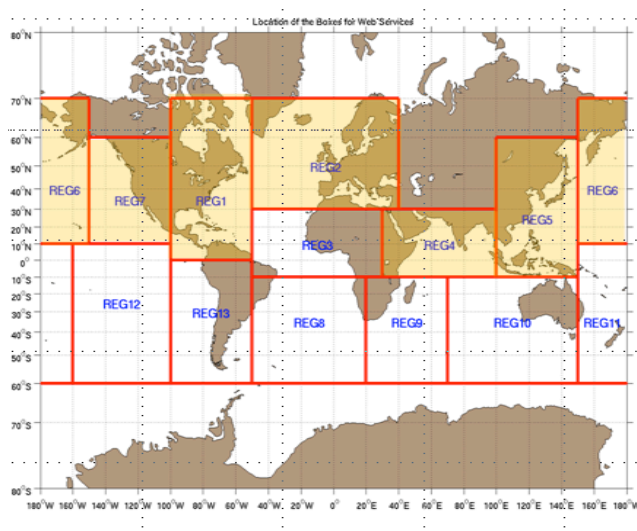


Figure 1. HYCOM regions evaluated.

The NOGAPS HYCOM is our current operational model. The NAVGEM version was run by NRL on the NAVOCEANO DSRC system HAISE after a year of NAVGEM forcing calibration. The resulting June and July calibration factors were used in this pre-operational run. Comparison data were collected using AutoMetrics, as described in the Global Ocean Forecast System (GOFS) OPEVAL report (GOFS3.0_optest_report_10APR12_final). A complete discussion of the metrics and approach used for this evaluation is available in this report.

Table 1. Forecast Day 1 (taus 00-24) Metrics Summary Sheets for the period 1 JUNE to 31 JULY 2013. Model [A] is HYCOM with NAVGEM and model [B] is HYCOM with NOGAPS forcing. Positive scores indicate a “win” for [A]. Neutral scores occur when the differences are within $\pm 1\%$ (bias and RMSD are normalized by mean observed values).

Approved for public release; distribution is unlimited.

REGION 01 [A] GHYCOM-NAVGEN versus [B] GHYCOM-NOGAPS					
	BIAS	CORR	RMSE	TOLERANCE	TOTAL FOR A [GHYCOM (NAVGEN)]
SLD	-1	+0	+1	+1	SUM: +1
T000	+0	+0	+0	+1	SUM: +1
T100	+0	+0	+0	+1	SUM: +1
S000	+0	+0	+0	+1	SUM: +1
S100	+0	+1	+0	+1	SUM: +2

TOTALS	-1	+1	+1	+5	SUM: +6 (30.0%)
REGION 02 [A] GHYCOM-NAVGEN versus [B] GHYCOM-NOGAPS					
	BIAS	CORR	RMSE	TOLERANCE	TOTAL FOR A [GHYCOM (NAVGEN)]
SLD	+1	+1	+1	+0	SUM: +3
T000	+0	+1	+0	+1	SUM: +2
T100	+0	+0	+0	+1	SUM: +1
S000	+0	+0	+0	-1	SUM: -1
S100	+0	-1	+0	+0	SUM: -1

TOTALS	+1	+1	+1	+1	SUM: +4 (20.0%)
REGION 04 [A] GHYCOM-NAVGEN versus [B] GHYCOM-NOGAPS					
	BIAS	CORR	RMSE	TOLERANCE	TOTAL FOR A [GHYCOM (NAVGEN)]
SLD	+0	+0	+0	+1	SUM: +1
T000	+0	+0	+0	-1	SUM: -1
T100	+0	+0	+0	-1	SUM: -1
S000	+0	+0	+0	+1	SUM: +1
S100	+0	+0	+0	-1	SUM: -1

TOTALS	+0	+0	+0	-1	SUM: -1 (-5.0%)
REGION 05 [A] GHYCOM-NAVGEN versus [B] GHYCOM-NOGAPS					
	BIAS	CORR	RMSE	TOLERANCE	TOTAL FOR A [GHYCOM (NAVGEN)]
SLD	+1	+0	+0	+1	SUM: +2
T000	+0	+1	+0	-1	SUM: +0
T100	+0	+0	+0	+1	SUM: +1
S000	+0	+1	+0	-1	SUM: +0
S100	+0	+1	+0	+1	SUM: +2

TOTALS	+1	+3	+0	+1	SUM: +5 (25.0%)
REGION 06 [A] GHYCOM-NAVGEN versus [B] GHYCOM-NOGAPS					
	BIAS	CORR	RMSE	TOLERANCE	TOTAL FOR A [GHYCOM (NAVGEN)]
SLD	-1	-1	-1	+1	SUM: -2
T000	+0	+0	+0	+1	SUM: +1
T100	+0	+0	+0	+1	SUM: +1
S000	+0	+0	+0	-1	SUM: -1
S100	+0	+0	+0	+1	SUM: +1

TOTALS	-1	-1	-1	+3	SUM: +0 (0.0%)
REGION 07 [A] GHYCOM-NAVGEN versus [B] GHYCOM-NOGAPS					
	BIAS	CORR	RMSE	TOLERANCE	TOTAL FOR A [GHYCOM (NAVGEN)]
SLD	-1	+1	+1	+1	SUM: +2
T000	+0	+0	+0	+1	SUM: +1
T100	+0	+0	+0	+1	SUM: +1
S000	+0	+1	+0	+0	SUM: +1
S100	+0	+0	+0	+0	SUM: +0

TOTALS	-1	+2	+1	+3	SUM: +5 (25.0%)
SUMS FOR ALL 6 REGIONS [A] GHYCOM-NAVGEN versus [B] GHYCOM-NOGAPS					
	BIAS	CORR	RMSE	TOLERANCE	TOTAL FOR A [GHYCOM (NAVGEN)]
SLD	-1	+1	+2	+5	SUM: +7
T000	+0	+2	+0	+2	SUM: +4
T100	+0	+0	+0	+4	SUM: +4
S000	+0	+2	+0	-1	SUM: +1
S100	+0	+1	+0	+2	SUM: +3

TOTALS	-1	+6	+2	+12	SUM: +19 (16%)

Summary of the results from Table 1:

- Of 120 metrics (20 scores/region times 6 regions), the net score is +19 or 16% in favor of HYCOM with NAVGEM forcing.
- The SLD comparisons show an apparent 29% (7/24) gain in skill with NAVGEM forcing.
- The temperature and salinity scores for the surface and 100m depths also suggest that NAVGEM forcing improves HYCOM skill.
- Results are positive in favor of NAVGEM for 4 of 6 regions. The exceptions are a -1 in favor of NOGAPS in Region 4 (the North Indian Ocean) and no difference in region 6 (Central North Pacific).

A normalized $\pm 1\%$ “neutral” or “zero difference” zone was assigned to each of these metrics. To determine if this was realistic, we looked at the sizes of 95% confidence intervals (95%CI) for biases using a bootstrap analysis. For SLD, the 95%CI averaged about 5% of the mean observed SLD while temperature and salinity 95%CIs were usually less than $\pm 0.50\%$ and $\pm 0.20\%$ of the means, respectively. For the “tolerance” or percent within an expected value, we have been using $\pm 5\text{m}$ for SLD, $\pm 0.50^\circ\text{C}$ for temperature and ± 0.20 psu for salinity. The 95%CI results suggest these values are OK although $\pm 10\text{m}$ for SLD, $\pm 0.20^\circ\text{C}$ for temperature and ± 0.10 psu for salinity might be achievable long-term objectives for HYCOM modelers. Also, as one looks deeper in the ocean, all temperature and salinity difference ranges will shrink to nearly zero.

We also looked at comparison scores for each forecast day between 1 and 5 (the operational HYCOM forced with NOGAPS forecasts run to 7-days and the pre-operational NAVGEM version to 5-days). Here, we see an interesting increase in NAVGEM-forces scores as the forecast extends toward the end of the period.

Table 2. Summary of JUNE-JULY 2013 daily metric scores for each forecast day between 1 (taus 00-24) and 5 (taus 96-120). Column 1 repeats the scores in Table 1.

Forecast Day:	1	2	3	4	5
Region	-----				
01	6	9	7	8	8
02	4	1	2	1	2
04	-1	-1	7	4	5
05	5	-5	4	8	8
06	0	-1	5	4	7
07	5	5	-1	2	-1

Sums:	19	8	24	27	29
Percentages:	16%	7%	20%	22%	24%

As was documented in the HYCOM OPEVALS report, we see very little decay in skill over the 5-day forecast periods. To illustrate this point, we present Figure 2, a set of Taylor diagrams for Region 4. If there were a loss in skill over the length of the 5-day forecasts, the colored dots would be spread out, with day-5 point found toward the upper left. Instead, all the circles are on top of each other.

For the 6 regions evaluated, SLD forecasts are usually the least skilled (i.e., off the black dot) and to the left of the center line, an indication that HYCOM under-forecasts SLD. This is common although not as bad as we once saw with Global NCOM. Temperature at the surface and 100m and salinity at 100m are normally very accurate with SSS somewhat lower in skill. This lower SSS skill is most prominent in Region 5 for some reason.

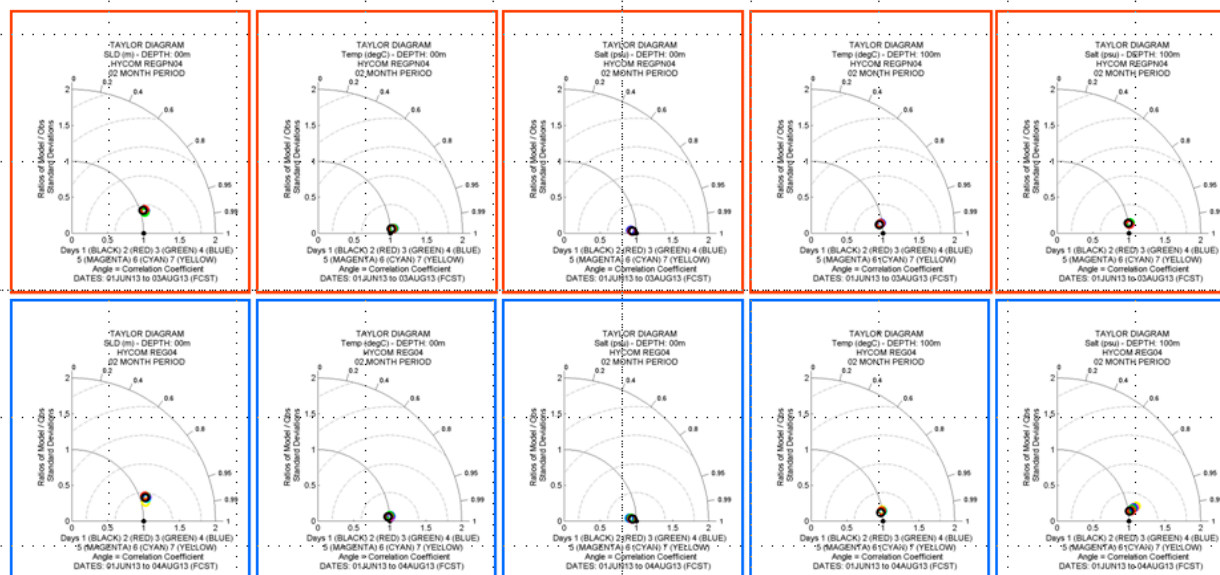


Figure 2. Taylor Plots of metrics. Top (red) is the NAVGEM series and bottom (blue) show NOGAPS. From, left to right are SLD, T00, S00, T100, and S100 plots. A “perfect” score would be on the black dot at the bottom center. The left vertical scale represent the ratio between observed and modeled RMSD and the curved right scale is the correlation coefficient ranging from 1.0 at the bottom to 0.0. Each day’s score is a different colored circle.

Table 3. Changes in the NOGAPS-forced HYCOM tolerance metric for a 30-month period for forecast days-2, -3 and -4, relative to the 6-region mean tolerance on forecast day-1

Compared to Day-1: Tolerance Metric	Day-2	Day-3	Day-4
SLD	-0.3	-0.6	-0.5 %
SST	-3.5	-6.0	-8.0
SSS	-1.1	-1.8	-1.9
T100	-2.2	-3.3	-4.2
S100	-1.8	-2.7	-4.1

Based on our long-term AutoMetrics evaluation of NOGAPS-forced HYCOM over the 30-months from February 2011 to July 2013, we can quantify the decay in skill over the period from forecast day-2 to day-4. We do this by looking at the change in tolerance scores (Table 3). We can interpret this to suggest that over a 4-day period, the HYCOM model loses a maximum of 8% in forecasting skill. As indicated by Table 1, The NAVGEM HYCOM score seems to improve relative to the NOGAPS-forced version.

Table 4. Summary of 6-Region NAVGEM-forced HYCOM metrics for June-July 2013. For BIAS and RMSD, the metric is divided by the mean observed value to come up with a mean percent. All difference are absolute forecast day-5 (D5) minus absolute day-1(D1). Differences less than 1% are indicated by “NO CHANGE.”

	DAY-1 MEAN	DAY-1 METRIC	PCT OF MEAN	DAY-5 METRIC	PCT OF MEAN	D1 - D5 METRIC	D1 - D5 PERCENT	DIFFERENCE D5 VS.D1
A.MODEL - OBSERVED BIAS								
SLD	34.37	-0.36	-6.6%	0.27	-4.7%	-0.09	-0.3%	D5 BETTER
SST	21.45	0.08	0.5%	0.07	0.3%	-0.01	-0.1%	NO CHANGE
SSS	33.57	0.21	0.6%	0.24	0.7%	0.03	0.1%	NO CHANGE
T100	16.52	-0.02	-0.1%	-0.01	0.0%	-0.01	-0.1%	NO CHANGE
S100	34.90	0.01	0.0%	0.00	0.0%	-0.01	0.0%	NO CHANGE
B.CORRELATION COEFFICIENT								
SLD		0.97		0.95		-0.02	-2.0%	D5 BETTER
SST		0.98		0.98		-0.01	-0.7%	NO CHANGE
SSS		0.91		0.90		-0.01	-1.2%	D1 BETTER
T100		0.98		0.98		0.00	-0.2%	NO CHANGE
S100		0.96		0.95		-0.01	-1.2%	D1 BETTER
C.RMSD								
SLD	34.37	31.22	101%	31.95	102%	0.73	2.1%	D1 BETTER
SST	21.45	0.86	4.4%	0.99	5.0%	0.13	0.6%	NO CHANGE
SSS	33.57	0.49	1.5%	0.52	1.6%	0.04	0.1%	NO CHANGE
T100	16.52	0.81	5.0%	0.95	5.9%	0.14	0.8%	NO CHANGE
S100	34.90	0.21	0.6%	0.23	0.7%	0.02	0.0%	NO CHANGE
D.TOLERANCE								
SLD		0.46		0.45		-0.01	-1.2%	D1 BETTER
SST		0.56		0.49		-0.07	-7.0%	D1 BETTER
SSS		0.60		0.57		-0.03	-3.3%	D1 BETTER
T100		0.54		0.48		-0.06	-6.0%	D1 BETTER
S100		0.76		0.72		-0.05	-4.5%	D1 BETTER
MEAN OF 4 SCORES								
						SLD	-1.4%	D1 BETTER
						SST	-2.1%	D1 BETTER
						SSS	-1.2%	D1 BETTER
						T100	-1.8%	D1 BETTER
						S100	-1.4%	D1 BETTER

Table 4 provides further insight into the 5-day forecast skill of the NAVGEM-forces HYCOM. Temperature at the surface and 100m lose the most skill—about 2% over 5 days. SLD, an important property for Navy applications, only decays by 1.4% over the 5-day period. The tolerance metric shows the greatest variance over the forecast period. In a number of measures, the model skill actually rises (see the “D5 BETTER” comments). We acknowledge that these are averages of averages and more specific scores are less consistent. Also, these are two summer months when SLD is normally shallow and SST is at its peak. A 12-month evaluation (or the 30-month coverage for NOGAPS-forced HYCOM) will reveal some seasonal shifts in skill.

Figure 5 is an example of how the metrics are determined. A full Power Point summary for the assessment of all 6 regions evaluated is available.

Figure 5. Bar graph comparisons between NAVGEM-forced HYCOM (red border) and NOGAPS-forced HYCOM (blue border) for forecast days 1 (bottom) to 5 or 7 (top bar). Left graphics show bias, correlation coefficient, RMSD and right panes are tolerance (green is within given range of $\pm 5.0\text{m}$, $\pm 0.5^\circ\text{C}$, or $\pm 0.2\text{psu}$ for SLD, temperature, or salinity, respectively). Horizontal blocks show metrics for SLD (blue bars), temperature at the surface and 100m (red), and salinity at the surface and 100m (green). Metrics are given to the right of each bar plot (zoom to see). Note that scales vary.



Conclusions: NAVGEM forcing of HYCOM provides a positive improvement in the model's skill. According to the metrics used, the model loses very little skill over the full forecast period. Replacing NOGAPS with NAVGEM should improve HYCOM skill.