Operational Test (OPTEST) Report **10 APRIL 2012**

The Global Ocean Forecast System, Version 3.0 (GOFS 3.0) or the Hybrid Coordinate Ocean Model (HYCOM)

Author: Frank L. Bub, NAVOCEANO NP3M

OPTEST PANEL

Chris DeHaan, NAVOCEANO
Bruce Lunde, NAVOCEANO
E. Joseph Metzger, NRL Stennis

REPORT SECTIONS

0. EXECUTIVE SUMMARY

This Operational Test (OPTEST) for the Global Ocean Forecast System (GOFS) demonstrates that the HYbrid Coordinate Ocean Model (HYCOM, GOFS 3.0) is slightly more skilled than the Global Navy Coastal Ocean Model (GNCOM, GOFS 2.6) that it is to replace. Skill, however, is not the most important reason to move to HYCOM as it represents a new basis for global prediction. There are a number of important technical and scientific improvements in HYCOM that will lead to greater forecast skill and the future coupling of the ocean, atmosphere, ice, waves and land in the Earth Systems Prediction Capability (ESPC).

The OPTEST results are based on eleven-month statistical comparisons between observed and modeled temperature, salinity, and sonic layer depth (SLD) data. The original plans included multiple case-studies by NAVOCEANO Oceanography Subject Matter Experts (OSMEs) but a heavy workload by our OSMEs and slow progress in setting up data streams for this sort of evaluation precluded this effort. As with NCOM, we will gain experience with HYCOM as we support Fleet operations, gain a better understanding of HYCOM skills, and weaknesses, and provide feedback to NRL on our findings.

A scorecard approach is used, with a plus or minus one (±1) assigned to HYCOM or GNCOM according to four standard statistical metrics: bias, correlation coefficient, root mean square difference, and percentage of differences within an acceptable range. Five ocean properties are scored: SLD, temperature at the surface and 100m, salinity at the surface and 100m. Five ocean areas that are important to Navy operations have been evaluated (Figure 1.1): the Western North Atlantic, Eastern North Atlantic, North West Indian Ocean, Western North Pacific, and Eastern North Pacific. The statistics are based on the eleven-month period beginning

January 2011 and ending November 2011, allowing us to look at the models during all four northern hemisphere seasons.

On average, HYCOM outscores GNCOM for 3% of the metrics, meaning we can say that HYCOM provides as good as or slightly better products compared with GNCOM, driven mostly by better temperature and salinity results at the surface. GNCOM actually outscored HYCOM for SLD metrics.

An objective of this OPTEST is to demonstrate that model skill declined very little over the period of the forecast. An average 5% reduction in the "tolerance" metric over 96-hour is a strong indication that HYCOM forecasts lose skill very slowly.

In its current configuration, HYCOM requires 624 processors for 13-hours on the DSRC system Einstein for a 4-day NCODA/hindcast and 5-day forecast.

In summary, HYCOM represents an improvement over GNCOM and we recommend that the AMOP declare GOFS 3.0 operational.

1. INTRODUCTION.

1.1. Objective. The objective of this OPTEST was to demonstrate that, for Navy operations, the skill of the HYbrid Coordinate Ocean Model (HYCOM) is equal to or better than that of the Global Navy Coastal Ocean Model (GNCOM). GNCOM is known as Version 2.6 of the Global Ocean Forecast System (GOFS 2.6) and the next generation is GOFS 3.0 or HYCOM.

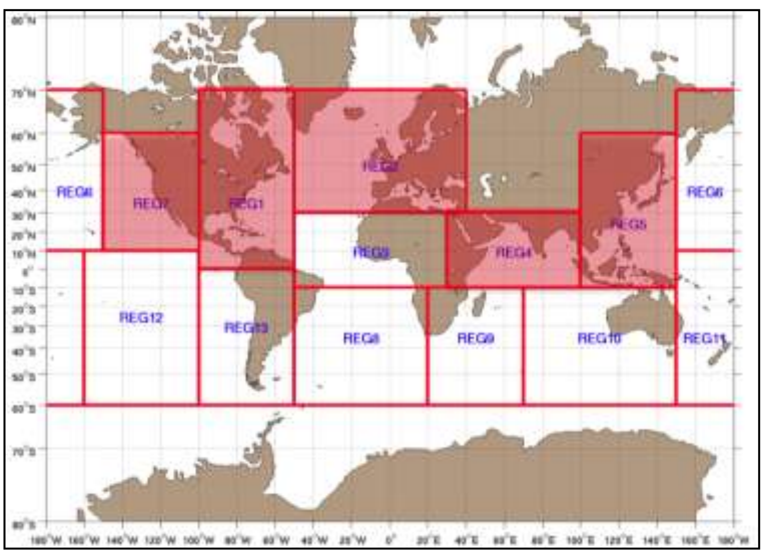


Figure 1.1. Map of GNCOM & HYCOM ocean regions. Shaded areas were evaluated for this OPTEST.

1.2. The Approach. The OPTEST has been heavily based on statistical comparisons between concurrent observed and model temperature, salinity, and sonic layer depth (SLD) data. These data are collected during the model run by the AutoMetrics program and analyzed by a series of Matlab scripts.

A scorecard approach similar to that developed by FNMOC is used, with a plus or minus one (±1) assigned to HYCOM or GNCOM according to four standard statistical metrics (bias, correlation coefficient, root mean square difference, and percent of differences between observed and modeled values within an acceptable range) for five ocean properties (SLD, temperature at the surface and 100m, salinity at the surface and 100m). The 100m depth was selected as it is normally within the thermocline-- a challenging region to forecast. We are assuming that skill at 100m should be an excellent indicator of overall model skill. Section 2 provides significant results and recommendations; Section 3 describes the model system; Section 4 discusses the findings; Section 5 briefly describes AutoMetrics; Section 6 summarizes the results and conclusions; Appendix A describes the analyses performed in greater detail; and Appendix B provides eleven-month time series of the metrics for each property and region evaluated.

Five ocean areas that are important to Navy operations are evaluated (Figure 1.1): the Western North Atlantic (REG1, WLANT); Eastern North Atlantic (REG2, ELANT); North West Indian Ocean (REG4, WIO); the Western North Pacific (REG5, WPAC); and the Eastern North Pacific (REG7, EPAC). Thus, for each month's scorecard, there is a potential maximum of 4 metrics times 5 properties times 5 regions = 100 points in favor of HYCOM if it were to outscore GNCOM in every region on every metric.

The OPTEST data collection was initiated on 01 December 2010. Twelve months, from December 2010 to November 2011 were evaluated, allowing us to span all four seasons in the northern hemisphere. HYCOM was updated by adding a diurnal atmospheric forcing module in early January 2011, so statistics are based on the eleven-month period beginning mid-January and ending November 2011.

1.3. The Report. This report was completed in March 2012. It will be presented to the Administrative Modeling Oversight Panel (AMOP) for Milestone III approval and a recommendation that GOFS 3.0 (HYCOM) be declared an operational model in the NAVOCEANO suite.

If so approved, HYCOM will replace the unclassified version of GNCOM during summer 2012, providing Navy support in regions where a higher resolution regional RNCOMs are not available. HYCOM will also be used for boundary conditions for nested RNCOM domains. The classified replacement of GNCOM will occur after replacement of the DOD Supercomputing Resource Center (DSRC) system Pascal toward the end of 2012.

2. RESULTS AND RECOMMENDATIONS.

Table 2.1. Summary of January to November 2011 monthly scores. A positive value means HYCOM scored above GNCOM the given number of times, a zero means the results were within ±1% of each other, and negative means GNCOM outscored HYCOM.

JAN-NOV 2011 SUMMARY	1 WLANT	2 ELANT	4 WIO	5 WPAC	7 EPAC	PROPERTY TOTALS	
SLD	-14	11	-16	-1	5	-15	-7%
T 00m	34	-9	-17	-4	24	28	13%
S 00m	5	-6	6	19	6	30	14%
T 100m	24	5	0	-19	9	19	9%
S 100m	0	-1	-21	-9	1	-30	-14%
REGION TOTALS	49	0	-48	-14	45	32	3%
TOTAL PERCENT	22%	0%	-22%	-6%	20%	3%	

2.1. Summary of Results. An extensive set of analyses and a large number of graphics were generated during the analysis. These were assembled as monthly Power Point files, which will be made available as unattached appendices. This section gives an overview of the results, based on the scorecard approach; with more in-depth discussions of the findings in Section 5. Table 2.1 and Figure 2.1 present "the bottom line."

For each region or column on Table 2.1, there is the potential for 4 metrics times 5 properties or 20 points for each of 11 months, resulting in a total of 220 points for each regional total. Thus, in WLANT, the bottom line shows that HYCOM outscores GNCOM 49/220 = 22% of the metrics. In WIO, GNCOM has the advantage by 22%, and so on. On the rightmost column, each of the property scores is summed, so for SLD, GNCOM outscores HYCOM by 15/220 or 7%. At the surface, HYCOM outscores GNCOM by 13% for temperature and 14% for salinity. At 100m, HYCOM outscores GNCOM by 9% for temperature and GNCOM outscores HYCOM by 14% for salinity.

On average, HYCOM outscores GNCOM by 32/1100 or 3% of the metrics, meaning we can say that HYCOM is provides as good as or slightly better products compared with GNCOM, driven mostly by better temperature and salinity results at the surface.

Figure 2.1 presents this same information as a bar graph, evaluated by property. The right (green) column in Table 2.1 is repeated by the rightmost bars. Figure 2.2 plots the total monthly scores for all properties, by region. During June and July, as summer progresses we note a shift from HYCOM to GNCOM as the higher scoring model. HYCOM skill in the Indian Ocean seems to be a major issue, which we will discuss further later.

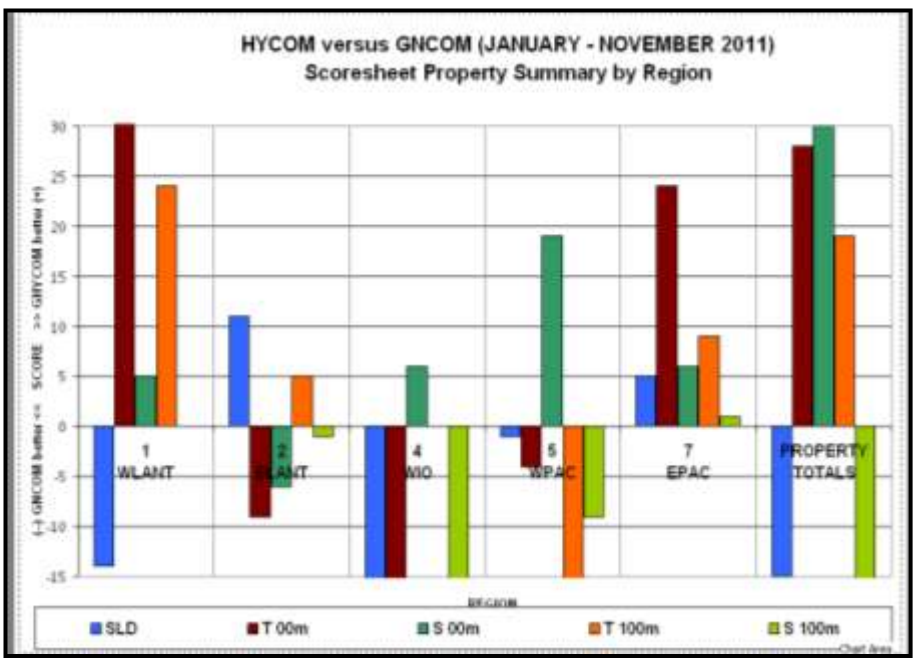


Figure 2.1. Bar graph of HYCOM scores for each property evaluated, by region, from Table 2.1. Positive scores (above the zero line) indicate a net "win" for HYCOM. All regions are summed by property to the right.

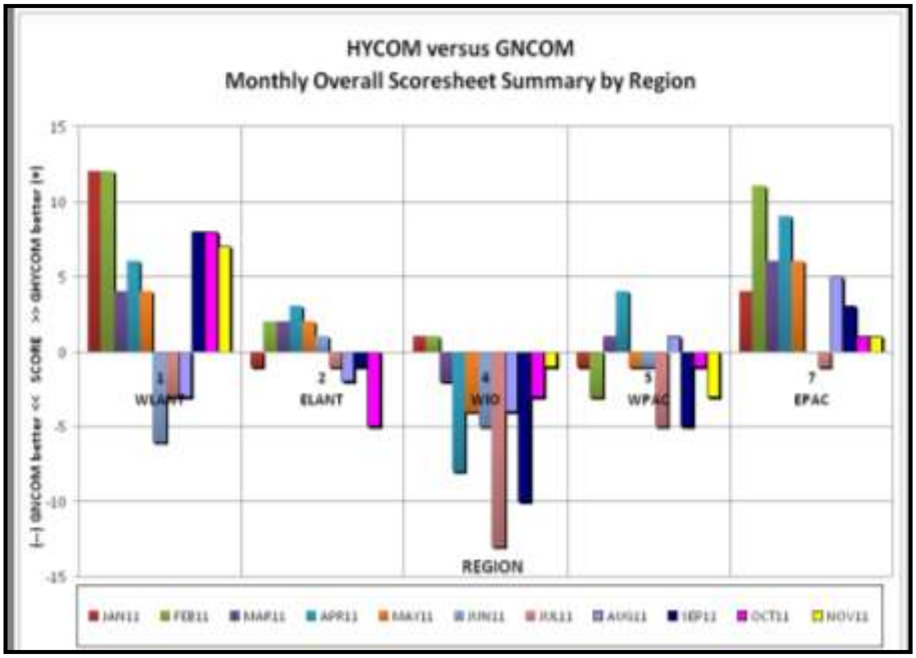


Figure 2.2. Bar graph of month-by-month HYCOM scores, by each of five regions. All regions are summed by month to the right.

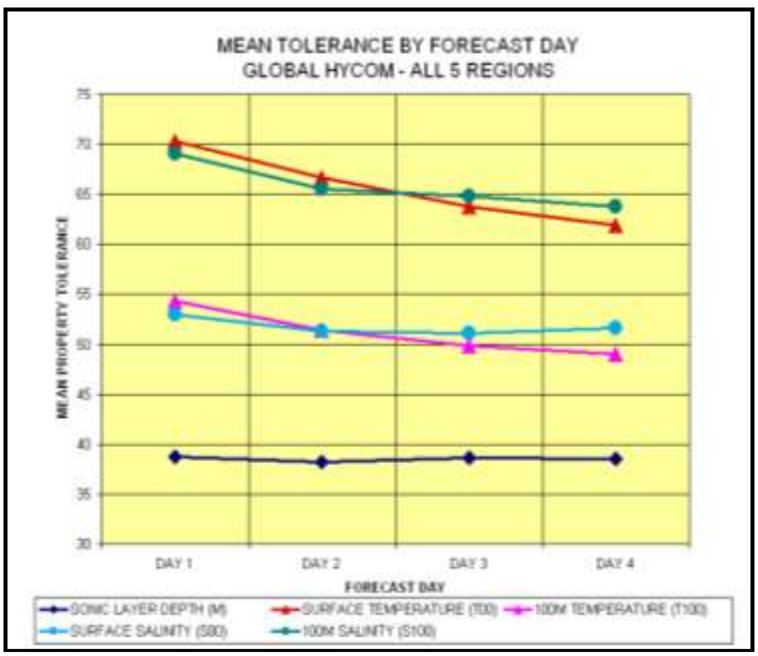


Figure 2.3. Mean tolerance scores across all regions for each of the properties and each forecast day.

A secondary objective of this OPTEST was to demonstrate that model skill declined very little over the period of the forecast. Figure 2.3 suggests that this is true. For instance, the 72-96-hour forecast for model minus observed temperature differences within the ±0.5°C range at the surface (red line) drops about 8% from 70% to 62% over the four-day period. By this metric, SLD tolerance dropped less than 1%, salinity at the surface dropped about 1%, and temperature and salinity at 100m each dropped about 5%. **An average 4% reduction in this metric over the 96-hours period is a strong indication that HYCOM forecasts are losing their skill very slowly.**

2.2. Data Used. Figure 2.4 is an example of the data coverage for the five ocean regions during May 2011, a typical sampling month. On average, there were 2,709 (REG1), 1,617 (REG2), 1,926 (REG4), 5,307 (REG5), and 1,207 (REG7) SLD comparisons.

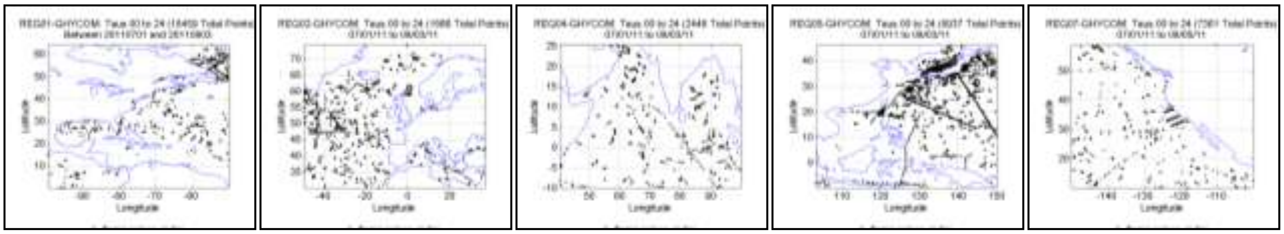


Figure 2.4. May 2011 data coverage for (left to right) regions 1, 2, 4, 5, and 7.

2.3. Unresolved Issues. This OPTEST brings out a number of issues worthy of further investigation. Why, for example, do we appear to do so poorly for the SLD forecasts? In part, this may be caused by the way SLD is calculated. An upcoming improvement to HYCOM, the Improved Synthetic Ocean Profile (ISOP) system is addressing this issue. The results are compared at model depths, which are at 5-10m increments in the main SLD region (see discussion with Table 5.1). Both models do a particularly poor job forecasting SLD in region 7, the Eastern North Pacific. Here, there appears to be some very deep observed SLDs (greater than 250m) that the model does not accept. Similarly, why does GNCOM seem to do better for both temperature and salinity at 100m than HYCOM? When we look at the actual metrics, the model and observed differences are small but since we normalize by the observed mean, the differences are just enough to score GNCOM higher. Why is GNCOM a definite winner in the WIO metrics? There is no clear answer here. While these and other issues will be discussed in section 5, all will not be resolved for this OPTEST report. Over the upcoming years, the NAVOCEANO Ocean Forecasters will continue to gain experience with GOFS 3.0 and beyond, providing feedback and recommendations to NRL.

2.3. Why Replace GNCOM? Irrespective of the fact that there may not be a large improvement indicated by this scorecard approach, HYCOM still represents a next generation capability for global prediction that will enable a long list of current and future enhancements. We need to consider the following:

- HYCOM is the ocean modeling system of the present and future. The National Ocean Partnership Program (NOPP) has made over a 10-year investment with funding and participation by ONR, NOAA, NSF, NRL, academia, and others. It is a consortium effort, with many high-level ocean modelers reviewing and exercising the system. The software is open source so modeling centers all over the world are installing the system and providing feedback to the developers. Put another way, there are many centers using HYCOM and only NAVOCEANO is using NCOM. There is no additional R&D being invested in NCOM. Currently, HYCOM is the ocean model proposed for the Earth System Prediction Capability (ESPC). An operational HYCOM will lead to the retirement of the Navy Layered Ocean Model (NLOM) and the Modular Ocean Data Assimilation System (MODAS).

- HYCOM physics result in a number of important improvements over NCOM. There is more to be learned about the hybrid coordinate approach from an operational standpoint, but all indications are that it will work better in areas where static NCOM layering has problems. The 1/12-degree HYCOM is approaching an eddy resolving vice eddy permitting system. This means that, for ASW operations where sharp ocean fronts and well defined eddies will affect acoustic propagation, this is a step forward. When we move to 1/25 degree, HYCOM should demonstrate skill in coastal regions where GNCOM has had problems. At that resolution, the intention will be to retire some of the 1/36-degree regional NCOM domains.

- HYCOM is integrated with the Navy Coupled Ocean Data Assimilation system (NCODA), the Community Ice Model (CICE), and the Earth Systems Modeling Framework (ESMF). NCODA gives us a way to move assimilation forward from MVOI to 3DVAR and eventually 4DVAR. Coupling CICE and HYCOM will provide better Arctic and Antarctic forecasts for ocean and ice as already demonstrated by the Arctic Cap Nowcast/Forecast System (ACNFS). These capabilities are planned for GOFS 3.1. ESMF is the universal vehicle for real-time data exchanges between ocean, atmosphere, wave, and ice models—a key element for the future of the ESPC. Future versions of HYCOM will include wave-current coupling with WaveWatch III and more realistic three-dimensional barotropic tides instead of the linearly added baroclinic tides that GNCOM uses (GOFS 3.5). We look forward to the advancements that the Improved Synthetic Ocean Profile (ISOP) project will provide.

On the other hand, GNCOM has not been a dead-end effort. Many of the improvements that went into HYCOM were derived from NCOM work and experience. It is a mature, stable, well-validated system that NAVOCEANO has successfully used to serve Navy operations throughout the world. When considering computational requirements, the 1/8-degree global NCOM is much less expensive than HYCOM, running a 4-day forecast in about 2 hours on 256 processors, while HYCOM runs a 5-day forecast for about 13 hours on 624 processors. In addition, the 1/12-degree HYCOM will triple our storage and transfer requirements (and the 1/25-degree version will quadruple them again for a factor of about 12 above today's GNCOM). While we have asked NAVOCEANO N6 and OIS to plan for these increases, there still may be throughput and data access limitations within NAVOCEANO and out to customers. Global NCOM is forced by the half-degree NOGAPS fields from FNMOC for a down-step of about 1:4. As we increase ocean model resolution, it seems prudent to do the same for the atmospheric forcing. So, for the 1/12-degree HYCOM, we require a 1/3-degree atmosphere (which is about what is planned for NAVGEM) and for the 1/25-degree version, we'd like a 1/6-degree atmosphere.

In summary, HYCOM represents an improvement over GNCOM and we recommend that the AMOP declare GOFS 3.0 operational.

3. THE HYCOM MODEL.

3.1. HYCOM. This description is condensed from the Naval Research Laboratory "Validation Test Report for the Global Ocean Forecasting System (GOFS) V3.0 – 1/12° HYCOM/NCODA: Phase I" by Metzger *et al.*, dated 26 November 2008 (NRL/MR/7320—08-9148).

The HYbrid Coordinate Ocean Model (HYCOM) was developed as a 10-year National Ocean Partnership Program (NOPP) project with NRL Stennis as one of the lead organizations. Assuming an acceptable OPTEST, HYCOM will replace the 1/8 degree eddy-permitting Global Navy Coastal Ocean Model (GNCOM, GOFS 2.6) by 2012. Future versions of HYCOM (GOFS 3.5 & 4.0) will include Polar Regions coupling with the Community Ice CodE (CICE), 3D Variational Assimilation scheme (3DVAR), integrated barotropic tides, and Wave Watch III model coupling.

HYCOM is an eddy-resolving, primitive equation general circulation model, with an approximate horizontal resolution of 1/12 or 0.08 degrees. The grid is a Mercator projection from 78.6°S to 47°N, linked to an Arctic dipole grid with poles over the Canada and Russia to avoid a convergence singularity at the North Pole. The mid-latitude and polar resolutions are approximately 7.5km (4nm) and 3.5km (2nm), respectively. The model employs 32 "hybrid" vertical surfaces that may be isopycnal (constant density), equal pressure (nearly level), or terrain following (sigma). The isopycnal surfaces are best in the deep stratified ocean, the pressure surfaces in the mixed layer and unstratified regions, and the sigma surface in shallow water. Depending on the dynamic situation, one of these three approaches is used at each time step and each point in the domain. Smooth transitions are optimally distributed by using the layered continuity equation. This hybrid approach takes advantage of the best points of each coordinate system and, in particular, results in a better representation of upper ocean physics. The model is designed to use a number of vertical mixing algorithms. The current version uses the K-Profile Parameterization (KPP) approach. A full description of the model and relevant algorithms is available in the HYCOM Software Design Document (SDD, NRL/MR/7320—09-9166).

HYCOM is forced by the Fleet Numerical Meteorology and Oceanography Center (FNMOC) 0.5° Navy Operational Global Atmospheric Predictions System (NOGAPS) at 3-hour time steps extending to 120-hours. NOGAPS fields include air temperature at 2m, surface specific humidity, net surface shortwave and longwave radiation, total precipitation, ground/sea temperature, downward surface shortwave radiation, eastward and northward winds at 10m, mean sea level pressure, and dewpoint temperature at 2m. These fields are either input directly or used to calculate heat and buoyancy fluxes, surface wind stress, or stability based on temperature and humidity.

HYCOM assimilates observed data using the Navy Coupled Ocean Data Assimilation System (NCODA) using a three-dimensional Multi-Variant Assimilation System (MVOI). Simultaneously analyzed variables include temperature, salinity, geopotential, and vector velocities. Observations include data from satellites (altimetry or SSH from Jason, ERS) or sea surface temperature (SST from AVHRR, GOES, MSG, AMSR-E), and ice concentration (from SSM/I, DMSP). Buoys, profiling floats, gliders, and ships collect *in situ* surface and profile observations of temperature and salinity. Surface SST and SSH data are projected downward as synthetic profiles to provide a 3D field where there are no *in situ* data. All data are run through a quality-control process (OCEAN-QC) where they are assigned a probability of correctness score that affects their weighting during assimilation. NCODA is a cycling system in which the previous day's 24-hour forecast is compared with and revised by the assimilated observations as an increment to the current day's initial or analysis field of ocean conditions.

HYCOM has been running with its current configuration in a pre-operational mode since early January 2011 and 624 processors are required for a 13-hour daily run on the NAVOCEANO Cray XT5 system Einstein. This includes a 4-day hindcast with NCODA assimilation and a 5-day (120-hour) forecast. To ensure that the model products reach Fleet users as soon as possible, the NCODA and hindcast runs are scheduled to start at 1900Z and complete after the 00Z FNMOC NOGAPS fields are available at approximately 0500Z daily. The 120-hour forecast series then runs and completes about 0900Z.

4. AUTOMETRICS.

Condensed from Dykes, James D., "Implementation of the Automated Numerical Model Performance Metrics System," 24 July 2011, NRL/MR/7320—11-9353.

The AutoMetrics system was the result of a project under the Rapid Transition Program to transition software to NAVOCEANO that provided information that oceanographers could use to routinely assess the performance of numerical ocean prediction models.

In situ observed ocean temperature and salinity surface and profile data are collected and matched with similar model data at corresponding locations and times. The relationship between observations and model output consists of one-to-many, as there are multiple forecasts spanning a given time and place, meaning one observation can be compared to up to five forecasts (within taus 00-24, 24-48, 48-72, etc.). Because the ocean is sparsely and randomly observed, the timings of both observations and model output are considered arbitrary and the system must continually check for both to provide the most up-to-date matches. Ultimately a data base will be built with a set of matches that can be statistically analyzed, providing indications of model skill in both space and time.

Collections of temperature and salinity observed and model profiles are then used to calculate matched sound speed. From that Navy Reference Publication 33 (RP-33) parameters such as sonic layer depth (SLD) and below-layer gradient (BLG) are computed, forming additional sets of matches for statistical analyses. Matlab programs have been prepared to read these match files and create comparisons scorecards, as described in Appendix A. AutoMetrics software have been incorporated into the ARCOAS environment for oceanographer support to ASW.

5. DISCUSSION AND FINDINGS.

5.1. Possible Evaluation Errors. To start, we need to consider some of the problems with the "bulk statistical" approach that has been taken. There has been no attempt to weight observations according to their inherent errors. All data are passed through the OceanQC program and scored but no profiles have been removed. The Matlab analyses does an initial statistical analysis and removes "outliers" that lie outside two standard deviations (96%) before the final statistics are generated. In AutoMetrics, all observations are interpolated to the model levels listed in Table 5.1. Thus, at 40m a one-level difference in SLD will be at least 5m. Between 50 and 100m, the vertical differences will be an increment of 10m. This is considered a major contributor to SLD errors.

Table 5.1. The 40 levels of the HYCOM model NetCDF output (in meters).

0	2	4	6	8	10	12	15	20	25
30	35	40	45	50	60	70	80	90	100
125	150	200	250	300	350	400	500	600	700
800	900	1000	1250	1500	2000	2500	3000	4000	5000

Horizontal errors are also introduced when an observation is linearly interpolated to the distance-weighted profile data at surrounding grid point prior to comparison. Harley Hurlburt (NRL Stennis) provided some perspective on how model resolution is specified: GNCOM is characterized by a mid-latitude resolution of 1/8° or 14km (common with older models), while HYCOM uses the equatorial resolution of 2/25° or about 9km. The equatorial resolution of GNCOM is about 19.5km, so HYCOM is 2.2 times finer than GNCOM at the equator. Near mid-latitudes, the HYCOM resolution is about 7km, or twice that of GNCOM. Thus, in the Kuroshio Current or Gulf Stream, the HYCOM spatial error could be as much as 5km while that of GNCOM could approach 10km. In dynamic frontal regions, this could be significant.

Similarly, temporal errors of 1.5 hours are introduced with a model time step of 3-hours. This could be significant in areas of large internal tides such as the Western Pacific (region 5).

Sonic Layer Depth (SLD) calculations are based on a relatively simple RP-33 algorithm that we know is flawed. Work is underway to improve this calculation. We suspect some of the problems with the observed SLD calculations in region 7 are due to weaknesses in the RP-33 algorithm, where it is most likely picking up the base of a weak secondary sound channel instead of the SLD that the models "see."

These results are bulk statistics, computed for full ocean regions that span many water masses and features (Figure 2.1). It is very unlikely that we would get the same results if we were to evaluate consistent water mass subsets of these regions.

When we apply a statistical approach, some other issues need to be considered: (1) are the data actually normally distributed or Gaussian? (2) Are the data stationary (not varying in time)? (3) Are the data stochastic (truly random), and (4) are the differences significant. Except for (4), none of these have been investigated in detail. Simple tests for the significance of the bias comparisons using 95% confidence intervals are presented in section 5.2.

Table 5.2. Summary of January to November 2011 monthly scorecards. A positive value means HYCOM scored above GNCOM the given number of times, a zero means they were within ±1% of each other, and negative means GNCOM outscored HYCOM.

HYCOM OPTEST SCORESHEET - HYCOM VERSUS GNCOM - JAN 2011 -NOV 2011

Evaluated: A. model-observed difference or BIAS UPDATED: 27-Mar-12
B. correlation coefficient (CC)
C. root mean square difference (RMSD)
D. percent of differences in acceptable range - taus 00-24 forecasts

Properties: Sonic Layer Depth (SLD), Temperature and Salinity at 00 & 100m

Scoring: [+1] if HYCOM "better" (lower BIAS or RMSD, higher CC or IN RANGE), [0] if nearly even, [-1] if GNCOM better. TOTAL 4 per property

Score summary: 5 regions x 5 properties x 4 grades = 100 potential points/month

JAN11	1 WLANT	2 ELANT	4 WIO	5 WPAC	7 EPAC	PROPERTY TOTALS
SLD	2	0	0	0	4	6
T 00m	4	-1	-1	1	0	3
S 00m	3	0	4	2	1	10
T 100m	1	0	-2	-2	-2	-5
S 100m	2	0	0	-2	1	1
REGION TOTALS	12	-1	1	-1	4	15

FEB11	1 WLANT	2 ELANT	4 WIO	5 WPAC	7 EPAC	PROPERTY TOTALS
SLD	2	2	2	-2	1	5
T 00m	4	0	-1	1	3	7
S 00m	2	0	4	2	1	9
T 100m	3	0	-2	-2	4	3
S 100m	1	0	-2	-2	2	-1
REGION TOTALS	12	2	1	-3	11	23

MAR11	1 WLANT	2 ELANT	4 WIO	5 WPAC	7 EPAC	PROPERTY TOTALS
SLD	-2	2	1	0	0	1
T 00m	2	-2	-2	0	3	1
S 00m	2	0	0	1	1	4
T 100m	2	2	1	-2	2	5
S 100m	0	0	-2	2	0	0
REGION TOTALS	4	2	-2	1	6	11

APR11	1 WLANT	2 ELANT	4 WIO	5 WPAC	7 EPAC	PROPERTY TOTALS
SLD	-2	2	-4	1	2	-1
T 00m	3	0	-1	1	4	7
S 00m	4	1	-2	3	2	8
T 100m	1	2	1	-2	3	5
S 100m	0	-2	-2	1	-2	-5
REGION TOTALS	6	3	-8	4	9	14

MAY11	1 WLANT	2 ELANT	4 WIO	5 WPAC	7 EPAC	PROPERTY TOTALS
SLD	0	-3	0	-2	1	-4
T 00m	4	0	-2	0	4	6
S 00m	-1	3	2	2	2	8
T 100m	0	0	-1	-2	-1	-4
S 100m	1	2	-3	1	0	1
REGION TOTALS	4	2	-4	-1	6	7

JUN11	1 WLANT	2 ELANT	4 WIO	5 WPAC	7 EPAC	PROPERTY TOTALS
SLD	-4	2	1	-1	2	0
T 00m	3	1	-2	-1	2	3
S 00m	-3	-1	-3	3	-1	-5
T 100m	0	-1	1	0	-1	-1
S 100m	-2	0	-2	-2	-2	-8
REGION TOTALS	-6	1	-5	-1	0	-11

JUL11	1 WLANT	2 ELANT	4 WIO	5 WPAC	7 EPAC	PROPERTY TOTALS
SLD	-4	2	-4	-2	-2	-10
T 00m	2	0	-2	-1	1	0
S 00m	-3	-1	-3	2	-1	-6
T 100m	3	-2	-2	-2	0	-3
S 100m	-1	0	-2	-2	1	-4
REGION TOTALS	-3	-1	-13	-5	-1	-23

AUG11	1 WLANT	2 ELANT	4 WIO	5 WPAC	7 EPAC	PROPERTY TOTALS
SLD	-4	2	-4	3	-2	-5
T 00m	2	-2	-2	-1	4	1
S 00m	-3	-2	1	2	1	-1
T 100m	3	0	3	-2	2	6
S 100m	-1	0	-2	-1	0	-4
REGION TOTALS	-3	-2	-4	1	5	-3

SEP11	1 WLANT	2 ELANT	4 WIO	5 WPAC	7 EPAC	PROPERTY TOTALS
SLD	-2	0	-4	2	0	-4
T 00m	3	-2	-2	-2	1	-2
S 00m	3	-1	-2	-1	0	-1
T 100m	4	2	0	-2	2	6
S 100m	0	0	-2	-2	0	-4
REGION TOTALS	8	-1	-10	-5	3	-5

OCT11	1 WLANT	2 ELANT	4 WIO	5 WPAC	7 EPAC	PROPERTY TOTALS
SLD	0	-2	-4	2	-2	-6
T 00m	3	-1	0	-1	2	3
S 00m	1	-3	2	1	-1	0
T 100m	4	2	1	-1	0	6
S 100m	0	-1	-2	-2	2	-3
REGION TOTALS	8	-5	-3	-1	1	0

NOV11	1 WLANT	2 ELANT	4 WIO	5 WPAC	7 EPAC	PROPERTY TOTALS
SLD	0	4	0	-2	1	3
T 00m	4	-2	-2	-1	0	-1
S 00m	0	-2	3	2	1	4
T 100m	3	0	0	-2	0	1
S 100m	0	0	-2	0	-1	-3
REGION TOTALS	7	0	-1	-3	1	4

JAN-NOV 2011 SUMMARY	1 WLANT	2 ELANT	4 WIO	5 WPAC	7 EPAC	PROPERTY TOTALS	
SLD	-14	11	-16	-1	5	-15	-7%
T 00m	34	-9	-17	-4	24	28	13%
S 00m	5	-6	6	19	6	30	14%
T 100m	24	5	0	-19	9	19	9%
S 100m	0	-1	-21	-9	1	-30	-14%
REGION TOTALS	49	0	-48	-14	45	32	3%
TOTAL PERCENT	22%	0%	-22%	-6%	20%	3%	

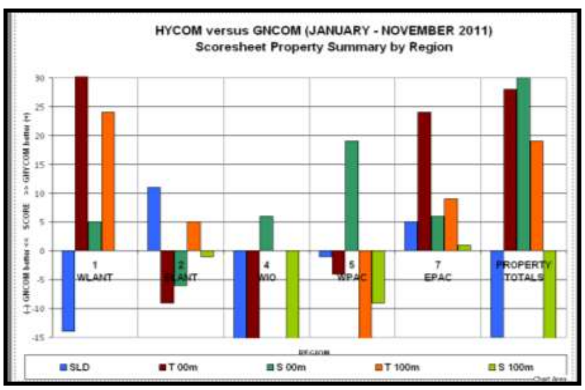
Finally, this is an OPTEST that relies on capabilities available to the NAVOCEANO watch team. The scientific, more controlled evaluations have been conducted by NRL and reported in the VTRs. The NAVOCEANO toolkit includes the production of daily "profile" files that are run through a very simple pass/fail quality control (QC) algorithm in the Real Time Data Handling System (RTDHS) and collated for the more stringent NCODA OceanQC software and assimilation. The AutoMetrics programs collect the concurrent profile and surface observations in the "profile" files and find concurrent (in space and time) model data. As a result, the GNCOM and HYCOM data sets can be different—some days are missing and the sizes of each day's comparisons can vary. Part of this apparent variation occurs because HYCOM is collecting 4-day forecasts and GNCOM only 3-days, but AutoMetrics continues to evolve and there are other differences that require further investigation.

5.2. In-Depth Discussion of the Results. See Appendix A for a complete development of the scoring system. For the scorecard, we have assigned +1 if HYCOM is better, 0 if normalized scores are within ±1%, and -1 if GNCOM is better. The bias and RMSD are normalized by the observed mean, while correlation coefficients and the tolerances are compared as computed. To determine whether the ±1% differences are statistically significant, we looked at 95% confidence intervals (95%CI) for the bias/mean ratio during the OPTEST period. If there is an overlap of the HYCOM and

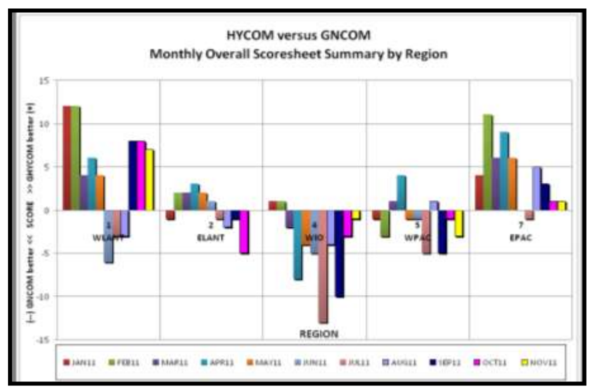
GNCOM ±95% confidence intervals, there is no significant difference, resulting in a score of 0. We evaluated the Region 1 confidence intervals and the “overlap” ranges for the selected properties are mixed, varying between ±1% and ±5%, so the ±1% constraint seemed reasonable, if not a bit harsh. If ±5% were used, most scores would come out 0.

Table 5.2 shows that HYCOM outscores GNCOM in a total of 32 of the metric measurements. For five regions, four metrics, and four properties, there is a total of 100 possible metrics per month or 1,100 for the eleven-month period. This means that the 32 total in favor of HYCOM is 3% of 1,100. By property, the SLD score for HYCOM is -15 or -7% of 220 measures (meaning GNCOM scores better in 7% of the measures), temperature at the surface +28 or +13%, salinity at the surface +30 or +14%, temperature at 100m +19 or +9%, and salinity at 100m -30 or -14%. These results are borne out by the far right bars in Figure 5.2. Looking at the metrics by region, WLANT (region 1) scores +49 of 220 or +22%, ELANT (region 2) 0 or 0%, WIO (region 4) -48 or -22%, WPAC (region 5) -14 or -6%, and EPAC (region 7) +32 or +20%. The WIO scores will be discussed in Section 6.

Figure 5.3 breaks down the scores by each of the properties, with each set of bar representing the monthly scores by regions and the summary to the right showing how the property score changes over time. These results are somewhat difficult to interpret so we will wrap up the results by looking at eleven-month bulk statistics as presented in Table 5.3.

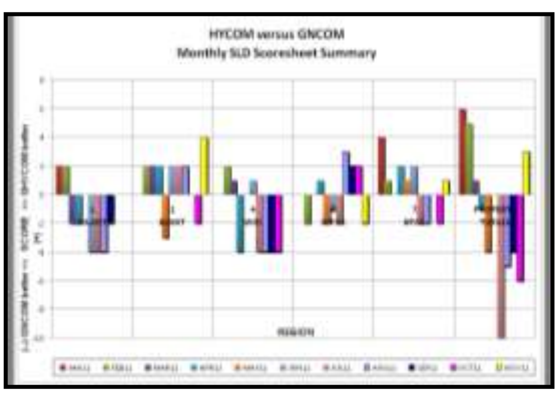


(a)

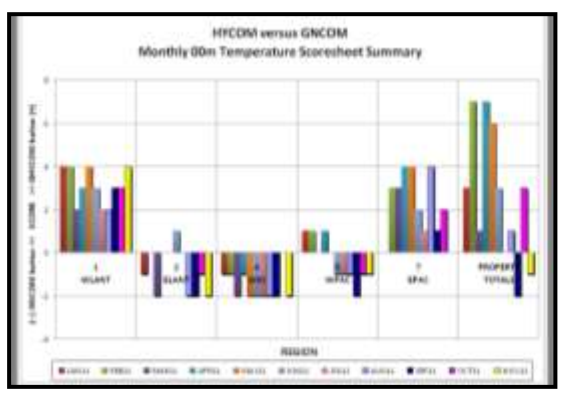


(b)

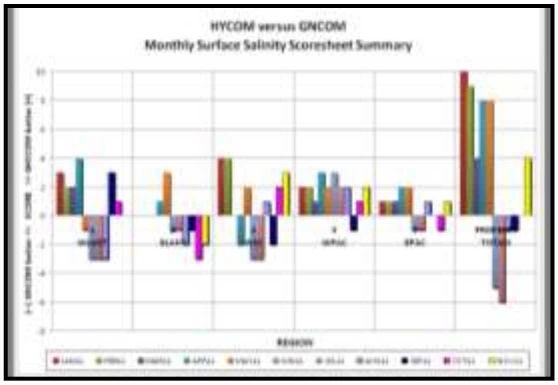
Figure 5.2. (repeats of Figures 3.1 and 3.2.) (a) Region-to-region bar graphs of HYCOM scores for each property from Table 5.2. Positive scores (above the zero line) indicate a net “win” for HYCOM. All regions are summed to the right. (b) Bar graphs of total monthly HYCOM scores, by region. All regions are summed to the right.

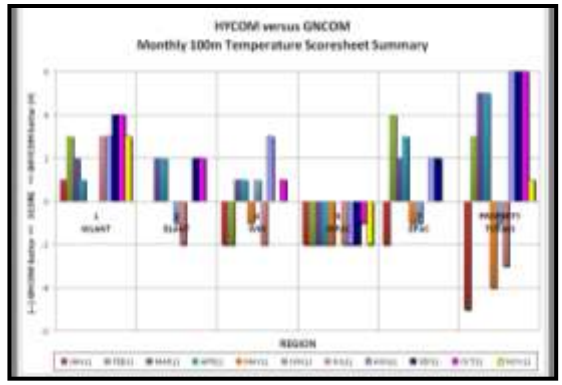


(a)



(b)

(c) 

(d) 

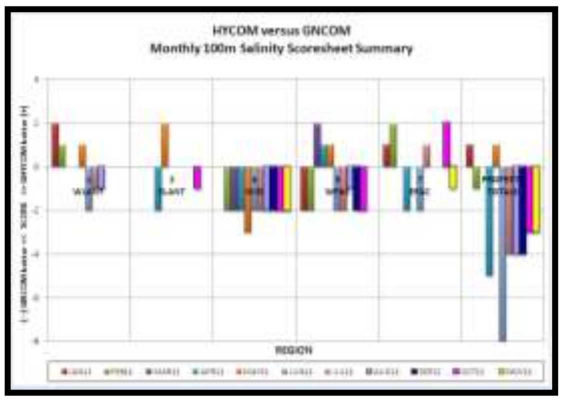
(e) 

Figure 5.3. Bar graph of HYCOM scores on Table 5.2 for each property, by region and month, for (a) SLD, (b) Surface temperature, (c) surface salinity, (d) 100m temperature, and (e) 100m salinity.

Table 5.3 summarizes the forecast day 1 (taus 00-24) results and scores for each region and all eleven months averaged together. Note that January 2011 data are not included for regions 1 and 7. The metrics presented are derived from the time series plots in Appendix B. The first column lists the region and the second is the property. The "DATA" column shows the HYCOM eleven-month model and observed values for SLD, T00, S00, T100, and S100. Note, for example, that the mean HYCOM SLD for all months and all regions is 46.3m and the observed mean is 57.3m for a mean HYCOM difference or bias of -11.0m.

The "CORRELATION" (correlation coefficient, CC) column shows the mean values for HYCOM (A), the A-B difference, and the score for each region. A higher CC is considered better. For example, for SLD, the mean CC is 0.60 (considered "fair"). The A-B difference is 0.00 meaning the GNCOM mean score is 0.60. The sum of region scores is 0 so there is no clear winner for SLD CC.

The "BIAS" column shows the mean HYCOM (model "A") bias (model mean minus observed mean), followed by the difference between the "A" (HYCOM) minus "B" (GNCOM) biases, and the HYCOM score for this metric. A lower bias is considered a better result. For example, the mean HYCOM SLD bias is -11.0m, which is -19% of the observed value, the A-B difference is +6.2m, so the GNMCOM bias is -17.2m. The sum of the region's score for SLD bias is +1. This means that the HYCOM result for SLD bias is slightly better than that for GNCOM. The table entries are coded green for a HYCOM win, tan for neutral, and red for GNCOM win.

Table 5.3. Summary sheet of monthly metrics, by property. See text for discussion.

SUMMED COMPARISONS OF METRICS BETWEEN HYCOM [A] AND GNCOM [B]

JANUARY TO NOVEMBER 2011 — 3/27/12

GREEN - HYCOM SCORES BETTER
TAN - NEUTRAL
RED - GNCOM SCORES BETTER

DAY 1

		DATA		BIAS SCORING			CORRELATION			RMSD			TOLERANCE						NET	
REG	PROP	MOD_MN	OBS_MN	A_MN	A-B	SCORE	A_MN	A-B	SCORE	A_MN	A-B	SCORE	A_MN	A-B	SCORE	DAY 2	DAY 3	DAY 4	SCORE	
SONIC LAYER DEPTH (M)																				
1	SLD	22.0	20.2	1.9	1.6	-1	0.77	-0.02	-1	19.4	5.3	-1	59.3	-3.1	-1	0.3	-0.7	-0.8	-4	NO JAN
2	SLD	40.8	55.1	-14.3	0.8	1	0.74	0.05	1	58.4	9.1	1	41.7	-1.3	-1	-1.3	-0.4	-1.0	2	
4	SLD	58.3	50.2	8.1	11.9	-1	0.52	0.00	0	25.4	5.5	-1	22.1	0.5	0	-0.6	0.9	1.5	-2	
5	SLD	74.8	70.5	4.4	16.2	1	0.71	-0.03	-1	26.2	0.8	-1	28.9	0.4	0	0.0	-0.7	-0.7	-1	
7	SLD	35.7	90.6	-55.0	0.7	1	0.26	0.01	1	94.0	7.0	1	41.4	-4.5	-1	-0.5	0.7	0.3	2	NO JAN
MEANS:		46.3	57.3	-11.0	6.2	1	0.60	0.00	0	44.7	5.5	-1	38.7	-1.6	-3	-0.4	0.0	-0.1	-3	
	PERCENT OF OBS:			-19%	-17.2			0.60	SI:	78%	39.1			40.3	DYS 2-4	38.3	38.6	38.5		
														NET SCORE FOR SLD:				-3		
SURFACE TEMPERATURE (T00)																				
1	T00	19.58	19.54	0.04	-0.06	0	0.95	0.04	1	0.66	-0.44	1	64.8	16.9	1	-6.2	-9.9	-12.8	3	
2	T00	14.27	14.25	0.02	0.16	0	0.95	0.00	0	0.55	0.08	0	69.3	-6.8	-1	0.2	-0.3	-2.0	-1	
4	T00	28.65	28.63	0.02	0.00	0	0.84	-0.03	-1	0.50	0.07	0	77.3	-6.7	-1	-3.9	-7.7	-9.4	-2	
5	T00	27.47	27.44	0.04	-0.02	0	0.95	0.00	0	0.54	0.03	0	74.3	-3.8	-1	-4.5	-8.3	-10.3	-1	
7	T00	13.20	13.09	0.12	-0.02	0	0.90	0.03	1	0.70	-0.22	1	65.8	14.0	1	-3.9	-6.2	-7.7	3	
MEANS:		20.63	20.59	0.05	0.01	0	0.92	0.01	1	0.59	-0.10	2	70.3	2.7	-1	-3.7	-6.5	-8.4	2	
	PERCENT OF OBS:			0.2%	0.04			0.91	SI:	2.9%	0.69			67.6	DYS 2-4	66.6	63.8	61.9		
														NET SCORE FOR T00:				2		
100M TEMPERATURE (T100)																				
1	T100	11.17	11.36	-0.19	-0.16	-1	0.93	0.01	0	0.96	-0.09	0	42.5	0.0	0	-2.8	-3.8	-5.9	-1	
2	T100	10.95	11.06	-0.11	0.05	0	0.95	0.00	0	0.50	0.06	0	75.5	-2.0	-1	0.0	-0.8	-2.0	-1	
4	T100	22.87	22.96	-0.09	-0.19	0	0.73	0.00	0	1.55	0.00	0	30.8	2.5	1	-1.4	-2.9	-2.8	1	
5	T100	25.18	25.20	-0.03	0.32	1	0.93	0.01	0	0.81	-0.05	0	52.5	7.8	1	-4.6	-6.9	-6.7	2	
7	T100	11.41	11.41	0.00	-0.02	0	0.94	0.00	0	0.52	-0.01	0	70.5	1.7	1	-5.8	-8.0	-9.4	1	
MEANS:		16.32	16.40	-0.08	0.00	0	0.90	0.00	0	0.87	-0.02	0	54.4	2.0	2	-2.9	-4.5	-5.4	2	
	PERCENT OF OBS:			-0.5%	-0.08			0.89	SI:	5.3%	0.89			52.4	DYS 2-4	51.4	49.9	49.0		
														NET SCORE FOR T100:				2		
SURFACE SALINITY (S00)																				
1	S00	34.16	33.61	0.55	0.18	0	0.85	0.26	1	0.98	-0.88	1	25.0	2.9	1	-1.1	-1.8	-2.4	3	
2	S00	35.14	35.17	-0.03	-0.01	0	0.95	0.02	1	0.36	0.08	0	72.3	2.3	1	-3.4	-4.5	-1.0	2	
4	S00	34.68	34.52	0.17	0.04	0	0.81	-0.01	-1	0.46	-0.08	0	54.5	2.7	1	-1.4	-3.3	-2.3	0	
5	S00	34.27	34.21	0.06	-0.03	0	0.66	-0.04	-1	0.38	0.04	0	59.9	-16.7	-1	0.6	2.0	1.6	-2	
7	S00	32.43	32.30	0.13	-0.06	0	0.84	0.13	1	0.62	-0.40	1	52.9	4.5	1	-3.0	-1.5	-2.2	3	
MEANS:		34.14	33.96	0.18	0.02	0	0.82	0.07	1	0.56	-0.25	2	52.9	-0.9	3	-1.7	-1.8	-1.3	6	
	PERCENT OF OBS:			0.5%	0.15			0.75	SI:	1.6%	0.81			53.8	DYS 2-4	51.3	51.1	51.7		
														NET SCORE FOR S00:				6		
100M SALINITY (S100)																				
1	S100	34.52	34.58	-0.06	-0.04	0	0.91	0.02	1	0.32	-0.02	0	52.3	-5.2	-1	-0.5	0.2	-2.1	0	
2	S100	36.07	36.09	-0.02	0.00	0	0.95	0.00	0	0.11	0.00	0	91.3	-0.9	0	-4.5	-3.7	-2.9	0	
4	S100	35.25	35.17	0.08	0.14	0	0.68	-0.21	-1	0.39	0.20	0	46.2	-23.2	-1	-2.1	-1.0	-1.4	-2	
5	S100	34.94	34.96	-0.02	0.06	0	0.85	-0.03	-1	0.18	0.03	0	71.7	-6.2	-1	-4.5	-8.1	-10.0	-2	
7	S100	33.77	33.77	0.00	-0.01	0	0.91	-0.01	0	0.15	-0.01	0	84.1	3.5	1	-6.4	-8.7	-10.3	1	
MEANS:		34.91	34.91	0.00	0.03	0	0.86	-0.05	-1	0.23	0.04	0	69.1	-6.4	-2	-3.6	-4.3	-5.3	-3	
	PERCENT OF OBS:			0.0%	-0.03			0.91	SI:	0.7%	0.19			75.5	DYS 2-4	65.5	64.9	63.8		
														NET SCORE FOR S100:				-3		
		NET METRIC SCORES: PERCENT OF TOTAL:			BIAS	1 4%		CC:	1 4%	RMSD:		3 12%		TOL:	-1 -4%		OVERALL		4 3%	
										MEANANCES			57.1			-2.5	-3.4	-4.1		
																54.6	53.7	53.0		

The “RMSD” (root mean squared difference) column shows the same information: mean RMSD for HYCOM, the HYCOM minus GNCOM (A-B) difference, and the RMDS scores. So, for SLD, the mean RMSD for all regions, all months is 44.7m, with the A-B difference of +5.5m, meaning the GNCOM mean is 39.1m. The net score for all regions is -1 in favor of GNCOM. The RMSD/OBS_MEAN ratio or scatter index (SI) for all HYCOM data is 78%.

The “TOLERANCE” scores, or mean percent of model-minus-observed differences within a specified range, are shown in the final column. A higher value is considered better. The selected tolerance ranges are ±5m for SLD, ±0.5°C for temperature, and ±0.2 psu for salinity. The mean value for HYCOM is under “A_MN” followed by the HYCOM minus GNCOM (A-B) difference and the score. The DAY2, DAY3, and DAY4 columns show how the HYCOM tolerance changes over subsequent daily forecasts (taus 24-48, 48-72, and 72-96). For example, the mean SLD tolerance for all regions indicates that 38.7% of the HYCOM data are within the ±5m range. The A-B difference of -1.6% means that 40.3% of the GNCOM SLDs are within this range. The net, five-region score for tolerance is -3 in favor of GNCOM. Over the 96-hour forecast period, the

HYCOM mean SLD tolerances changes from 38.7% (00-24) to 38.3% (24-48) to 38.6% (48-72) to 38.5% (72-96).

Table 5.4. Summary sheet of monthly metrics, by region.

SUMMED COMPARISONS OF METRICS BETWEEN HYCOM [A] AND GNCOM [B]
JANUARY TO NOVEMBER 2011 3/27/12

GREEN - HYCOM SCORES BETTER
TAN - NEUTRAL
RED - GNCOM SCORES BETTER

		DATA		BIAS SCORING			CORRELATION			RMSD			TOLERANCE DAY 1						NET	
REGION	PROP	MOD_MN	OBS_MN	A_MN	A-B	SCORE	A_MN	A-B	SCORE	A_MN	A-B	SCORE	A_MN	A-B	SCORE	DAY 2	DAY 3	DAY 4	SCORE	
1	SLD	22.0	20.2	1.9	1.6	-1	0.8	0.0	-1	19.4	5.3	-1	59.3	-3.1	-1	0.3	-0.7	-0.8	-4	NO JAN
1	T00	19.6	19.5	0.0	-0.1	0	0.95	0.04	1	0.7	-0.4	1	64.8	16.9	1	-6.2	-9.9	-12.8	3	
1	T100	11.2	11.4	-0.2	-0.2	-1	0.93	0.01	0	1.0	-0.1	0	42.5	0.0	0	-2.8	-3.8	-5.9	-1	
1	S00	34.2	33.6	0.6	0.2	0	0.9	0.3	1	1.0	-0.9	1	25.0	2.9	1	-1.1	-1.8	-2.4	3	
1	S100	34.5	34.6	-0.1	0.0	0	0.9	0.0	1	0.3	0.0	0	52.3	-5.2	-1	-0.5	0.2	-2.1	0	
						-2			2			1			0		NET REGION 1:		1	
2	SLD	40.8	55.1	-14.3	0.8	1	0.7	0.1	1	58.4	9.1	1	41.7	-1.3	-1	-1.3	-0.4	-1.0	2	
2	T00	14.3	14.3	0.0	0.2	0	0.95	0.00	0	0.6	0.1	0	69.3	-6.8	-1	0.2	-0.3	-2.0	-1	
2	T100	11.0	11.1	-0.1	0.1	0	0.95	0.00	0	0.5	0.1	0	75.5	-2.0	-1	0.0	-0.8	-2.0	-1	
2	S00	35.1	35.2	0.0	0.0	0	1.0	0.0	1	0.4	0.1	0	72.3	2.3	1	-3.4	-4.5	-1.0	2	
2	S100	36.1	36.1	0.0	0.0	0	1.0	0.0	0	0.1	0.0	0	91.3	-0.9	0	-4.5	-3.7	-2.9	0	
						1			2			1			-2		NET REGION 2:		2	
4	SLD	58.3	50.2	8.1	11.9	-1	0.5	0.0	0	25.4	5.5	-1	22.1	0.5	0	-0.6	0.9	1.5	-2	
4	T00	28.7	28.6	0.0	0.0	0	0.84	-0.03	-1	0.5	0.1	0	77.3	-6.7	-1	-3.9	-7.7	-9.4	-2	
4	T100	22.9	23.0	-0.1	-0.2	0	0.73	0.00	0	1.6	0.0	0	30.8	2.5	1	-1.4	-2.9	-2.8	1	
4	S00	34.7	34.5	0.2	0.0	0	0.8	0.0	-1	0.5	-0.1	0	54.5	2.7	1	-1.4	-3.3	-2.3	0	
4	S100	35.3	35.2	0.1	0.1	0	0.7	-0.2	-1	0.4	0.2	0	46.2	-23.2	-1	-2.1	-1.0	-1.4	-2	
						-1			-3			-1			0		NET REGION 4:		-5	
5	SLD	74.8	70.5	4.4	16.2	1	0.7	0.0	-1	26.2	0.8	-1	28.9	0.4	0	0.0	-0.7	-0.7	-1	
5	T00	27.5	27.4	0.0	0.0	0	0.95	0.00	0	0.5	0.0	0	74.3	-3.8	-1	-4.5	-8.3	-10.3	-1	
5	T100	25.2	25.2	0.0	0.3	1	0.93	0.01	0	0.8	-0.1	0	52.5	7.8	1	-4.6	-6.9	-6.7	2	
5	S00	34.3	34.2	0.1	0.0	0	0.7	0.0	-1	0.4	0.0	0	59.9	-16.7	-1	0.6	2.0	1.6	-2	
5	S100	34.9	35.0	0.0	0.1	0	0.9	0.0	-1	0.2	0.0	0	71.7	-6.2	-1	-4.5	-8.1	-10.0	-2	
						2			-3			-1			-2		NET REGION 5:		-4	
7	SLD	35.7	90.6	-55.0	0.7	1	0.3	0.0	1	94.0	7.0	1	41.4	-4.5	-1	-0.5	0.7	0.3	2	
7	T00	13.2	13.1	0.1	0.0	0	0.90	0.03	1	0.7	-0.2	1	65.8	14.0	1	-3.9	-6.2	-7.7	3	
7	T100	11.4	11.4	0.0	0.0	0	0.94	0.00	0	0.5	0.0	0	70.5	1.7	1	-5.8	-8.0	-9.4	1	
7	S00	32.4	32.3	0.1	-0.1	0	0.8	0.1	1	0.6	-0.4	1	52.9	4.5	1	-3.0	-1.5	-2.2	3	
7	S100	33.8	33.8	0.0	0.0	0	0.9	0.0	0	0.2	0.0	0	84.1	3.5	1	-6.4	-8.7	-10.3	1	
						1			3			3			3		NET REGION 7:		10	
		NET METRIC SCORES:			BIAS	1		CC:	1		RMSD:	3		TOL:	-1		NET OVERALL:		4	
		PERCENT OF TOTAL:				4%			4%			12%			-4%				3%	

In the discussions that follow, we arbitrarily grade the correlation coefficient (CC) and tolerance results which range from 0.0 to 1.0 or 0% to 100% as "excellent" if 85-100%, "good" if 65-85%, "fair" if 50-65%, and "poor" if below 50%.

These day-to-day tolerance results are plotted in figure 5.4 to illustrate that HYCOM skill decays very little over the 96-hour forecast period, as further discussed in section 5.3.

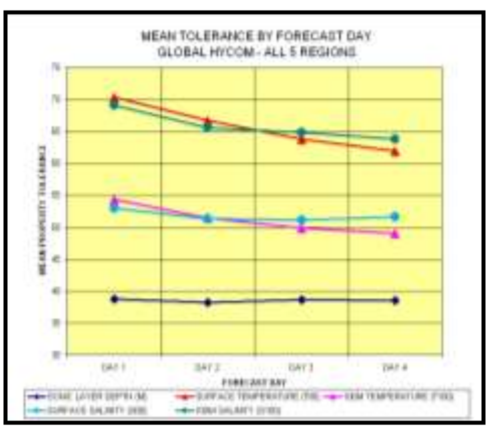


Figure 5.4. Mean HYCOM tolerance scores for each of the properties for forecast days 1-4.

5.3. Boring Down by Property Metrics.

5.3.1. SLD Score Summary. The January-November 2011, 5-region bulk mean HYCOM model SLD is 46.3m with an observed mean of 57.3m. The HYCOM bias is -11.0m, which is 19% of the observed value. Note that if we removed the questionable region 7 values, this would result in a model SLD of 49.0m with an observed mean of 49.0m, giving us an excellent +0.0m (0%) bias. The HYCOM minus GNCOM bias difference is +6.2m, meaning the GNCOM bias is a "worse" -17.2m. The sum of all region SLD bias scores is +1 in favor of HYCOM.

Is this difference significant? The HYCOM bias has a ±1.7m 95% confidence interval (95%CI) for a range of -9.3 to -12.7m and the GNCOM bias 95%CI is also ±1.7m for a range of -15.5 to 18.9. There is no overlap in these ranges so we can assume the net bias results are significant. The ratio between the 95%CI and mean observed SLD is 1.7/57.3 or 3%. For the SLD tolerance range, we used 5m, resulting in a 5.0/57.3 ratio of 9%. See a discussion of this information at the end of this section.

The mean HYCOM SLD CC is a "fair" 0.60. The A-B difference is 0.00 meaning the GNCOM mean CC is also 0.60. The sum of region scores is 0 in favor of neither model. The mean HYCOM SLD RMSD is 44.7m, with the A-B difference of +5.5m, meaning the GNCOM RMSD is a lower (better) 39.1m. The net score for all regions is -1 in favor of GNCOM. The scatter index for all HYCOM data is 78%, a number we'd like to see improved. The mean HYCOM SLD tolerance indicates that 38.7% of the HYCOM data are within the ±5m range. A "poor" result that we'd like improved. The A-B difference of -1.6% means that a slightly better 40.3% of the GNCOM SLD differences are within this range. The net SLD tolerance score is -3 in favor of GNCOM. Over the 96-hour forecast period, the HYCOM mean SLD tolerances changes from 38.7% (00-24) to 38.3% (24-48) to 38.6% (48-72) to 39.5% (72-96) for a net 96-hour change in this metric of -0.1% (Figure 5.4).

Bottom line—the net four-metric, five-region SLD score is -3 of 20, for 15% in favor of GNCOM. There is a 0% change in skill over 96-hours.

5.3.2. Surface Temperature (T00) Score Summary. The bulk mean HYCOM model T00 is 20.63°C with an observed mean of 20.59°C. The HYCOM bias is an excellent +0.05°C, which is 0.2% of the observed value. The HYCOM minus GNCOM bias difference is +0.01°C, meaning the GNCOM bias is a similar +0.04°C. The sum of all region T00 bias scores is 0, meaning neither model forecasts surface temperature any better.

The HYCOM T00 bias has a ±0.01°C 95%CI for a range of 0.04 to 0.06°C and the GNCOM bias 95%CI is also ±0.01°C for a range of 0.03 to 0.05°C. There is an overlap so we can assume the net bias results are not significant. The ratio between the 95%CI and mean observed T00 is 0.01/20.6 or 0%. For the temperature tolerance range, we used 0.5°C, resulting in a 0.5/20.6 ratio of 2%.

The mean HYCOM T00 CC is an "excellent" 0.92, with the A-B difference of 0.01 so the GNCOM mean CC is 0.91. The sum of region CC scores is +1 in favor of HYCOM. The mean HYCOM T00 RMSD is 0.59°C, with the A-B difference of -0.10°C. This means the GNCOM RMSD is a worse 0.69°C. The net T00 RMSD score for all regions is +2 in favor or HYCOM. The scatter index for all HYCOM data is an excellent 2.9%. The mean HYCOM T00 tolerance indicates that a "good" 70.3% of the HYCOM data are within the ±0.5°C range. The A-B difference of +2.7% means that 67.6% of the GNCOM T00s are within this range. The net T00 tolerance score is -1 in favor of GNCOM. Over the 96-hour period, the HYCOM mean T00 tolerances changes from 70.3% (00-24) to 66.6% (24-48) to 63.8% (48-72) to 61.9% (72-96) for a 96-hour change in this metric of -8.4%. (Figure 5.4).

Bottom line—the net 4-metric, five-region T00 score is +2 of 20, for 10% in favor of HYCOM. There is an 8% decay in the HYCOM surface temperature forecast skill over 96-hours.

5.3.3. 100m Temperature (T100) Score Summary. The bulk mean HYCOM model T100 is 16.32°C with an observed mean of 16.40°C. The HYCOM bias is an excellent -0.08°C, which is just 0.5% of the observed value. The HYCOM minus GNCOM bias difference is 0.0°C, meaning the GNCOM bias is a similar -0.08°C. The sum of all region T100 biases is 0, with no clear winner.

The HYCOM T100 bias has a ±0.06°C 95%CI for a range of -0.02 to -0.14°C and the GNCOM bias 95%CI is also ±0.06°C for the same range of -0.02 to -0.14°C. There <u>is an overlap</u> so we can assume the net T100 bias results are <u>not significant</u>. The ratio between the 95%CI and mean observed SLD is 0.06/16.3 or 0%. For the temperature tolerance range, we used 0.5°C, resulting in a 0.5/16.3 ratio of 3%.

The mean HYCOM T100 CC is an "excellent" 0.90, with the A-B difference of 0.00 so the GNCOM mean CC is also 0.90. The sum of region T100 CC scores is also 0. The mean HYCOM T100 RMSD is 0.87°C, with the A-B difference of -0.02°C. This means the GNCOM RMSD is a slightly worse 0.89°C. This 100m RMSD is 180% of the T00 value-- an expected result in this highly variable area of the upper thermocline. The net RMSD score for all regions is also 0. The scatter index for all HYCOM data is an excellent 5.3%. The mean HYCOM T100 tolerance indicates that a "poor" 54.4% of the HYCOM data are within the ±0.5°C range (fair). The A-B difference of +2.0% means that 52.4% of the GNCOM T100 differences are within this range. The net T100 tolerance score is +2 in favor of HYCOM. Over the 96-hour period, the HYCOM mean T100 tolerances change from 54.4% (00-24) to 51.4% (24-48) to 49.4% (48-72) to 49.0% (72-96) for a 96-hour change in this metric of -5.4% (Figure 5.4).

Bottom line—the net 4-metric, five-region T100 score is +2, for 10% points in favor of HYCOM. There is a 5% decay in HYCOM 100m temperature forecast skill over 96-hours.

5.3.4. Surface Salinity (S00) Score Summary. The bulk mean HYCOM model S00 is 34.14 psu with an observed mean of 33.96 psu. The HYCOM bias is an excellent +0.18 psu, which is 0.5% of the observed value. The HYCOM minus GNCOM bias difference is +0.02 psu, meaning the GNCOM bias is a slightly better +0.16 psu. The sum of all regional S00 bias scores is 0.

The HYCOM S00 bias has a ±0.02 psu 95%CI for a range of 0.16 to 0.20 psu and the GNCOM bias 95%CI is also ±0.02 psu for a range of 0.14 to 0.16 psu. It is close, but there is <u>no overlap</u> so we can assume the net S00 bias results <u>are significant</u>. The ratio between the 95%CI and mean observed S00 is 0.02/34.14 or 0%. For the salinity tolerance range, we used 0.2 psu, resulting in a 0.2/34.14 ratio of 1%.

The mean HYCOM S00 CC is a "good" 0.82, with the A-B difference of +0.07 so the GNCOM mean CC is a worse 0.75. The sum of region S00 scores is +1 in favor of HYCOM. The mean HYCOM S00 RMSD is 0.56 psu, with the A-B difference of -0.25 psu. This means the GNCOM RMSD is a worse 0.81 psu. The net S00 RMSD score for all regions is +2 in favor of HYCOM. The scatter index for all HYCOM data is an excellent 1.6%. The mean HYCOM S00 tolerance indicates that a "fair" 52.9% of the HYCOM data are within the ±0.2 psu range. The A-B difference of -0.9% means that 53.8% of the GNCOM S00s are within this range. The net S00 tolerance score is +3 in favor of HYCOM. Over the 96-hour period, the HYCOM mean S00 tolerances changes from 53.8% (00-24) to 51.3% (24-48) to 51.1% (48-72) to 51.7% (72-96) for a 96-hour change in this metric of -1.3% (Figure 5.4).

Bottom line—the net 4-metric, five-region S00 score is +6, or 30% in favor of HYCOM. There is an 1% decay in the HYCOM surface salinity forecast skill over 96-hours.

5.3.5. 100m Salinity (S100) Score Summary. The bulk mean HYCOM model S100 is 34.91 psu with an observed mean of 34.91 psu. The HYCOM bias is an excellent 0.0 psu, which is 0.0% of the observed value. The HYCOM minus GNCOM bias difference is +0.03 psu, meaning the GNCOM bias is a slightly "worse" -0.03 psu. The sum of all region S100 bias scores is 0.

The HYCOM S100 bias has a ±0.02 psu 95%CI for a range of -0.02 to +0.03 psu and the GNCOM bias 95%CI is also ±0.02 psu for a range of -0.05 to -0.01 psu. There is no overlap so we can assume the net bias results are significant. The ratio between the 95%CI and mean observed S100 is 0.02/34.91 or 0%. For the salinity tolerance range, we used 0.2 psu, resulting in a 0.2/34.91 ratio of 1%.

The mean HYCOM S100 CC is an "excellent" 0.86, with the A-B difference of -0.05 so the GNCOM mean CC is an even better 0.91. The sum of region scores is -1 in favor of GNCOM. The mean HYCOM S100 RMSD is 0.23 psu, with the A-B difference of +0.04 psu. This means the GNCOM RMSD is a lower or better 0.19 psu. The net S100 RMSD score for all regions is again 0. The scatter index for all HYCOM data is an excellent 0.7%. The mean HYCOM S100 tolerance indicates that a "fair" 69.1% of the HYCOM data are within the ±0.2 psu range. The A-B difference of -6.4% means that a better 75.5% of the GNCOM S100 differences are within this range. The net tolerance score is -2 in favor of GNCOM. Over the 96-hour period, the HYCOM mean S100 tolerances changes from 69.1% (00-24) to 65.5% (24-48) to 64.9% (48-72) to 63.8% (72-96) for a net 96-hour change in this metric of -5.3% (Figure 5.4).

Bottom line—the net four-metric, five-region S100 score is -3 of 20, for 15% in favor of GNCOM. There is a 5% decay in HYCOM 100m salinity forecast skill over 96-hours

Adding up the four-metrics, five-region results in a net BIAS score of +1 in favor of GNCOM, +1 for CC, +3 for RMSD, and -1 for TOLERANCE. The result is a total of +4. This 4/100 = 4% score is very similar to the 3% result from Table 5 which summed months by month.

To determine whether the ±1% range for neutral scores is reasonable, we compared the ratio between the 95% confidence intervals for bias over the observed mean as indicators of significant difference. These were 3% for SLD and 0% for T00, T100, S00, and S100, so the ±1% neutral range seems OK. To determine whether the tolerance ranges used were realistic, we also looked at the ratio between the selected values divided by the observed means (5m for SLD, 0.5°C for temperature, and 0.2 psu for salinity). These were 9% for SLD, 2% for T00, 3% for T100, 1% for S00, and 1% for S100. Except for SLD, these values all seem consistent with the 95%CI results of approximately 1% of the observed data. Because of the 5-10m differences between model layers noted above, reducing SLD tolerance below 5m does not seem reasonable.

5.4. A Review of the Metrics by Month.

The eleven-month time series of HYCOM and GNCOM SLD (left) and surface temperature (right) means and the various metrics are plotted on Figures 5.5 through 5.9. While the surface salinity and 100m temperature and salinity time series provide some interesting observations and issues, they will not be reviewed in this report. Interested readers are referred to Appendix B which presents all metrics time series for the 4 metrics and 5 regions. The means that are captioned in these plots are summarized in Tables 5.3 and 5.4 and discussed above.

In the upper left of each of the figures (panel a) are plotted the property mean values for each month. CIRCLES show HYCOM, with RED the model monthly mean and GREEN the observed mean. Squares show GNCOM, with BLUE the model monthly mean and MAGENTA the observed mean. Statistics for the total period are presented under each graphic. In the upper right (b) are plotted the model minus observed BIAS. The RED CIRCLE is for HYCOM and the BLUE SQUARE is for GNCOM bias. The GREEN line shows the absolute GNCOM (B) bias minus absolute HYCOM (A) bias differences. A LOWER bias is better, so a positive difference or green line above zero means HYCOM is better. The tolerance ranges (5m for SLD, etc.) are drawn as dashed lines on either side of the solid black zero line.

In the middle left (c) are plotted the CORRELATION COEFFICIENTS (CC), with a RED CIRCLE for HYCOM and a BLUE SQUARE for GNCOM. A HIGHER CC is better, so RED above BLUE means HYCOM is better. In the middle right (d) is the ROOT MEAN SQUARE DIFFERENCES (RMSD), with a RED CIRCLE for HYCOM and a BLUE SQUARE for GNCOM. A LOWER RMSD is better, so RED below BLUE means HYCOM scores better. The black dashed line indicates the tolerance for this metric.

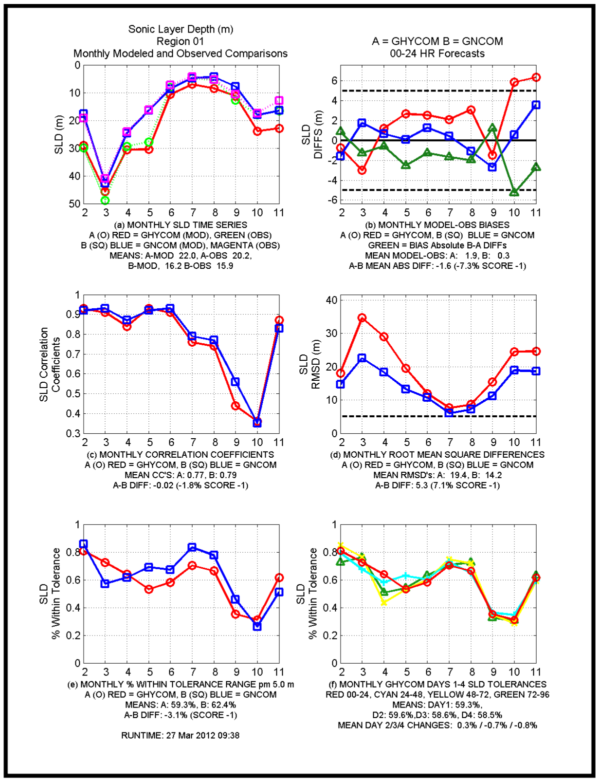
In the lower left (e) are plotted the percent of the differences that are within the given TOLERANCE range, with a RED CIRCLE for HYCOM and a BLUE SQUARE for GNCOM. A HIGHER value is better, so RED above BLUE means HYCOM scores higher. In the Lower Right (f) are plotted the HYCOM tolerances for each forecast day, with RED for the 00-24, CYAN 24-48, YELLOW 48-72, GREEN 72-96-hour forecasts. The objective is to demonstrate decay in skill over 4-day forecast period.

Net scores for the eleven-month period are indicated in the captions. When the bias (b) or RMSD (d) values are normalized by the mean observed values, the ratio is provided as a percentage. This will help to explain some of the seemingly inconsistent scores.

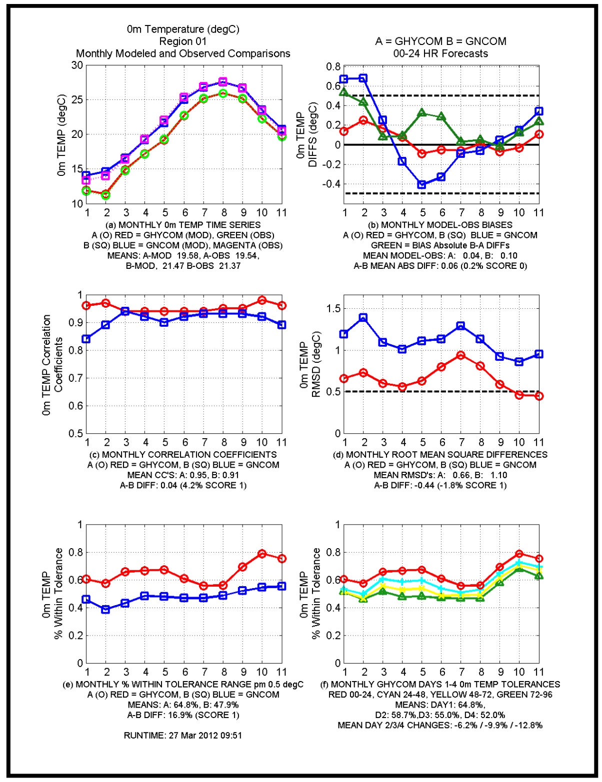
Note that the vertical scales in panels (a), (b), (c) and (d) vary according to the range of the metrics plotted. Panels (e) and (f) range from 0.0 to 1.0. In the discussions that follow, (*) will suggest that further investigation is warranted although in most cases that work will not occur as part of this OPTEST.

5.4.1. Western Atlantic (WLANT) Metrics Time Lines: Observed Region 1 SLD deepens from 30m JAN to 45m MAR, steadily rises or shoals to about 5m to 10m JUL-AUG, then deepens again to 25m by OCT (Figure 5.5.a). Note that the observations for both models do not agree, with the GNCOM data that are 5-10m shallower (*). The biases for both models are within the acceptable ±5m band except for a +6m HYCOM bias OCT-NOV. GNCOM generally scores better as the green line lies below zero. The FEB to JUN SLD CCs are an "excellent" 0.90 through JUN, then dip to an abysmal 0.35 OCT (*), rises back to an "excellent" 0.85 NOV. HYCOM SLD RMSD is 20m FEB, 35m MAR, improves to about 10m by JUL as the SLD shoals, and rises back to 25m OCT-NOV. The GNCOM RMSD is lower or "better" by 5-10m. HYCOM tolerance (percent of SLD differences within ±5m) is a "good" 80% in FEB, drops to a "fair" 50% MAY, rises back to 70% JUL-AUG, drops to 35% by OCT and rises again to 60% NOV. The GNCOM tolerance follows a similar pattern but remains slightly higher than HYCOM during the summer. The OCT drop in CC and tolerance require further investigation (*).

Mean Region 1 observed SST behaves nicely as it rises from 12-15°C FEB to 25-27°C JUL (Figure 5.5.b) then cools to 20-21°C by NOV. GNCOM observed data are consistently warmer than HYCOM by about 2°C (*). Both models have excellent SST biases within the ±0.5°C except for GNCOM at +0.7°C during JAN-FEB. HYCOM actually remains within ±0.2°C for the year but the net score is 0 because the differences are not statistically significant. HYCOM SST CC is "excellent" as it remains at or above 0.90 for the whole year. GNCOM is 0.10 (winter) to 0.05 (summer) lower. HYCOM SST RMSD is about 0.7°C JAN-FEB, drops to 0.5°C APR, rises to 1.0°C JUL, then drops to below 0.5°C by NOV. GNCOM RMSD is consistently 0.5°C higher. HYCOM tolerance (percent of SST within ±0.5°C) is a consistently "fair" 60-70% JAN-AUG and rises to a "good" 80% by OCT. GNCOM tolerance remains in the "poor" 40-60% range through the period, lower or "worse" than HYCOM by 15-20%.



(a)

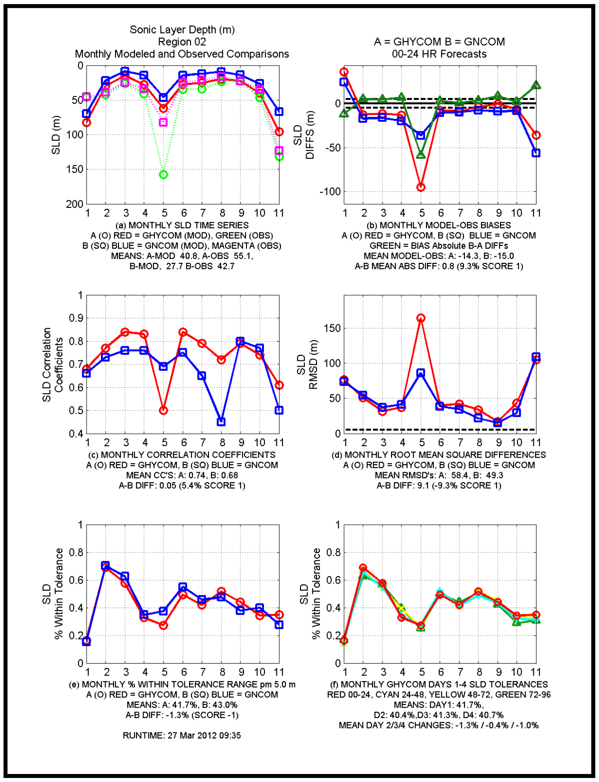


(b)

Figure 5.5. Eleven-month time series of the (a) Sonic Layer Depth (SLD) and (b) T00 metrics for Region 1 (WLANT). Note that the Region 1 SLD plot does not include January 2011.

5.4.2. Eastern Atlantic (ELANT) Metrics Time Lines: Mean observed SLD in region 2 rises from 50m JAN to 30-40m MAR-SEP, and drops to 130m by NOV, except for an anomalous jump to 80m (GNCOM) or 160m (HYCOM) during MAY (*) (Figure 5.6.a). Model SLDs are initially deeper at 70-80m, resulting in a 40m deep bias JAN. The bias then shifts to a more familiar 10-15m shallow for the rest of the year, with GNCOM about 5m more shallow than HYCOM. The MAY observations draw model SLDs down somewhat to 40-60m. Except for JAN, MAY, and NOV, the HYCOM-GNCOM bias differences are about +5m, meaning the models are consistent. HYCOM CC rises from a "fair" 0.70 JAN to a "good" 0.85 MAR to JUN with the exception of a "poor" 0.50 MAY. The GNCOM CC is 0.05-0.15 lower or worse through JUN, drops to a "poor" 0.45 AUG (*), then joins HYCOM SEP-OCT. Both model's RMSD values start at a high 70m JAN and improve slightly to 40m by MAR-JUL (except for MAY), then 10m by SEP, then rapidly rising to 100m by NOV. Tolerances are a "poor" 20% JAN rising to a "good" 70% FEB, dropping back to 35% APR-MAR, rising to 55% JUN, then slowly dropping to 30% by NOV.

The mean region 2 observed SST is 10°C in JAN, drops to 8°C in FEB, then rises steadily to about 20°C by AUG and drops again to 12°C by NOV (Figure 5.6.b). The HYCOM 11°C mean for MAY (*) supports the suggestion that there were some bad SST observations which the models fortunately attempts to "ignore." Both models track the observed data well, with SST bias well within the ±0.5°C envelope. HYCOM bias is nearly within nearly ±0.2°C throughout but the net score is 0 because the model differences are not significant. For both models, SST CC remains an "excellent" 0.90-1.00 throughout the year. RMSD is an excellent 0.3°C in JAN-MAR, steadily increases to 0.8°C by JUL, then improves to about 0.5°C SEP-NOV. GNCOM is 0.1-0.2°C lower or better than HYCOM but, again the score is 0 because this is not significant. Tolerance is an "excellent" 90% JAN-FEB, dropping to a "fair" 55% by JUN, then rising again to about 75% by NOV. GNCOM tolerances are 0-10% higher.

(a)


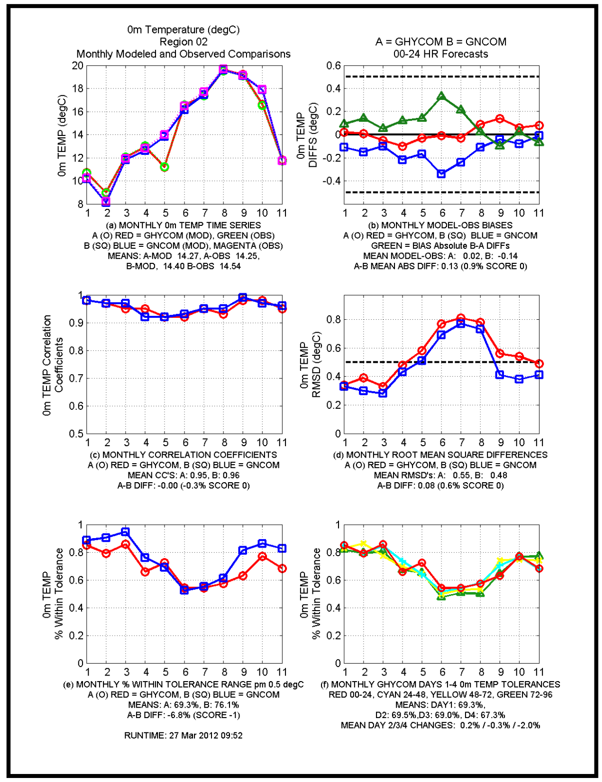
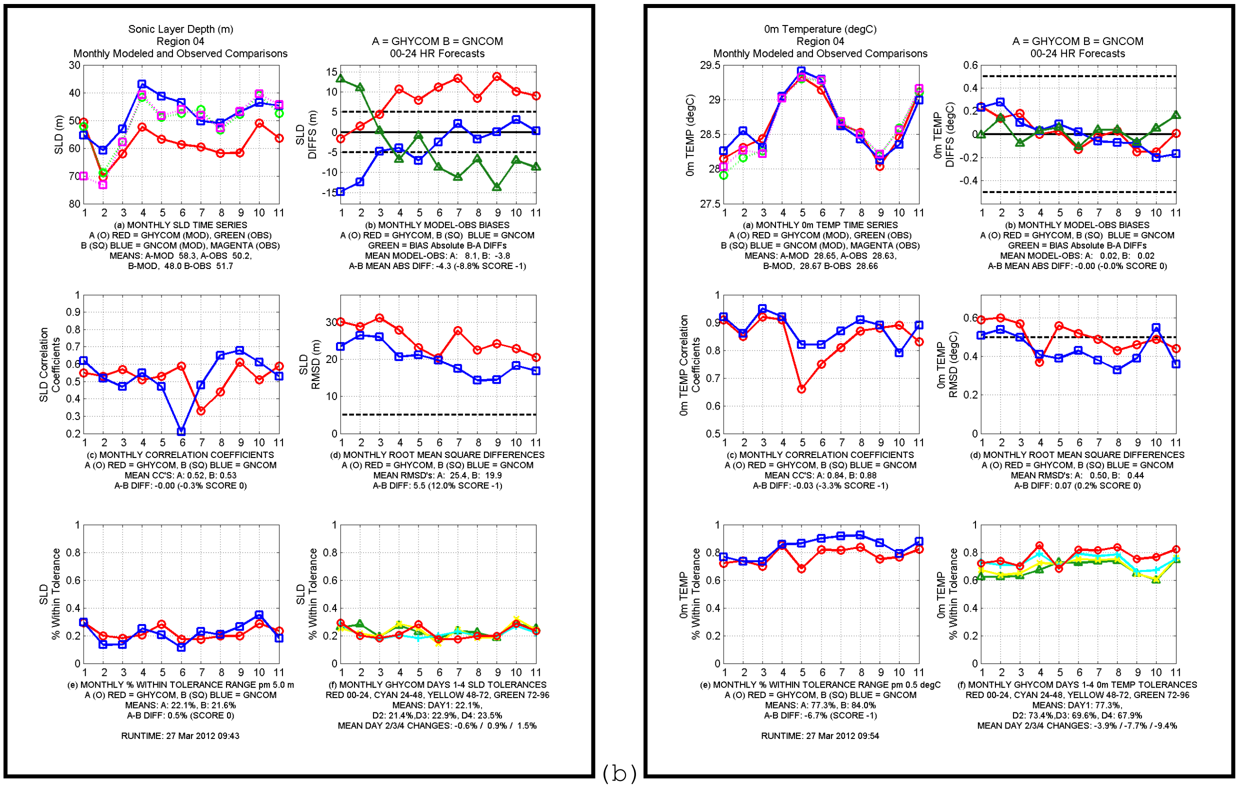
(b)


Figure 5.6. Same as Figure 5.5 for Region 2 (ELANT).

5.4.3. Western Indian Ocean (WIO) Metrics Time Lines: There are mixed results for SLD in region 4 with interesting contrasts in the GNCOM and HYCOM data. The observed JAN SLD is 50-55m for the GNCOM data and 70m for HYCOM, then a very similar 70m FEB, rising to 40m in APR, falling to 50m in AUG, and then 45m in NOV (Figure 5.7.a). The HYCOM and GNCOM SLDs follow this pattern with GNCOM a bit shallower than the observations and HYCOM is about 10m too deep. As a result, the HYCOM bias hovers around +10m after APR while the GNCOM bias ranges from -5m in MAR to near 0m from JUN on. The HYCOM CC is a "fair" 0.50-0.60 throughout the period except for a "poor" dip to 0.30 in JUL (*). The GNCOM CC is a "fair" 0.50-0.60 JAN to MAY, drops to a very "poor" 0.20 in JUN (*), rises back to 0.65 AUG-SEP, then falls to 0.50 in NOV. The HYCOM RMSD starts at 30m in JAN, steadily dropping to 20m by NOV. GNCOM is consistently better by 5-10m. SLD tolerances for both models are an abysmal 10% to 30% throughout the period (*).



(a) (b)

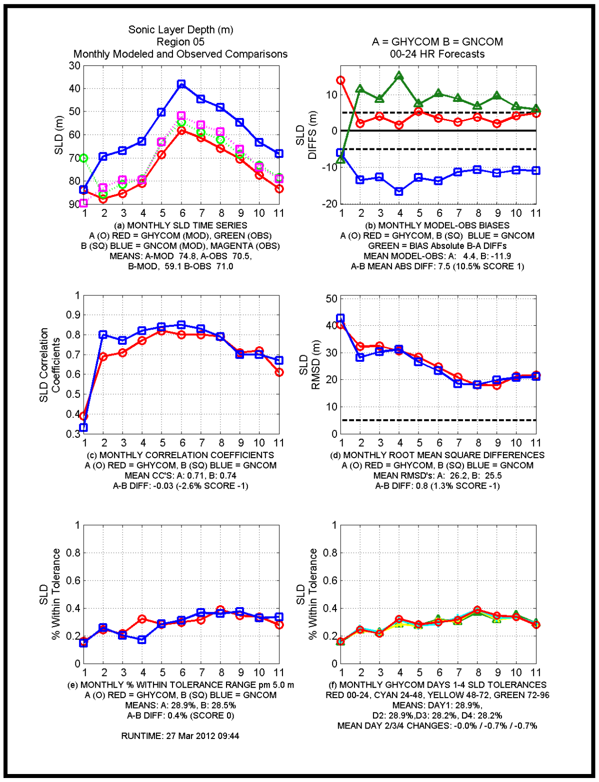
Figure 5.7. Same as Figure 5.5 for Region 4 (WIO).

Observed and modeled SSTs in region 4 are consistent and don't change much, rising from 28°C JAN to 29.4°C MAY, dropping to 28.2°C SEP and rising to 29.0°C by NOV (Figure 5.7.b). Both HYCOM and GNCOM biases are very small, ranging from +0.2°C in JAN to near -0.2°C by NOV. CC is an "excellent" 0.85-0.95 JAN-APR, with HYCOM CC dropping TO 0.65 in MAY (*). Both model CCs remain in the "good" to "excellent" ranges of 0.80 to 0.90. HYCOM CC is about 0.05 lower or worse than GNCOM. Both model RMSEs are very good at 0.4-0.6°C with HYCOM about 0.05°C higher than GNCOM. Again, the net score is 0 as the differences are not significantly significant. Tolerances for both models are a "good" to "excellent" 80-90% throughout the period.

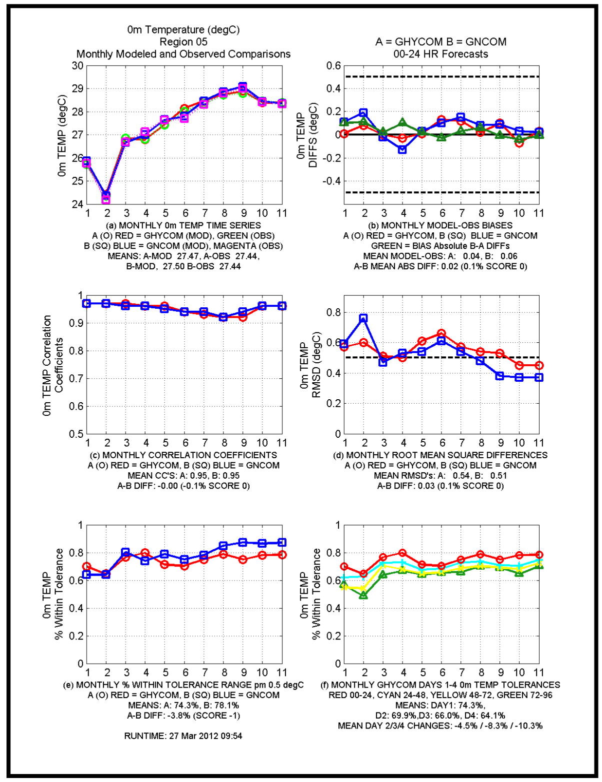
5.4.4.Western Pacific (WPAC) Metrics Time Lines: Region 5 observed SLD is 70m for the GNCOM data and 90m for HYCOM JAN, shoals to 60m by JUN, and deepens back to 80m by NOV (Figure 5.8.a). Except for JAN, the HYCOM bias is an excellent 0-5m too deep throughout the period while the GNCOM bias is 10-15m shallow. HYCOM CC starts at a very poor 0.35 (*) then rises rapidly to a "good" 0.70-0.80 FEB to AUG, then a "fair" 0.70 in OCT-NOV. The GNCOM CC is 0.05 to 0.10 higher or better. Both model RMSDs follow the same paths from 40m JAN to 20m by JUL through NOV. The SLD tolerances rise from a poor 20% to 40% from JUL to NOV (*).

Observed and modeled SST in region 5 are in excellent agreement as they drop from 26°C JAN to 24.5°C FEB, up to 27°C MAR, then steady rise to 29°C SEP, ending at 28°C in NOV (Figure 5.8.a). Bias for both models are "excellent," with all within 0.0 to 0.2°C. CC is an "excellent" 0.90-0.95 throughout the period. RMSD is 0.6°C in JAN, falling to about 0.4°C by NOV with HYCOM about 0.05°C higher. SLD tolerance for both models rises from a "fair" 70-80% in JAN to a "good" 80% in NOV, with GNCOM 3-5% higher than GNCOM.

The WPAC region is consistently the most highly observed area and, as a result, there are no issues with the metrics.



(a)

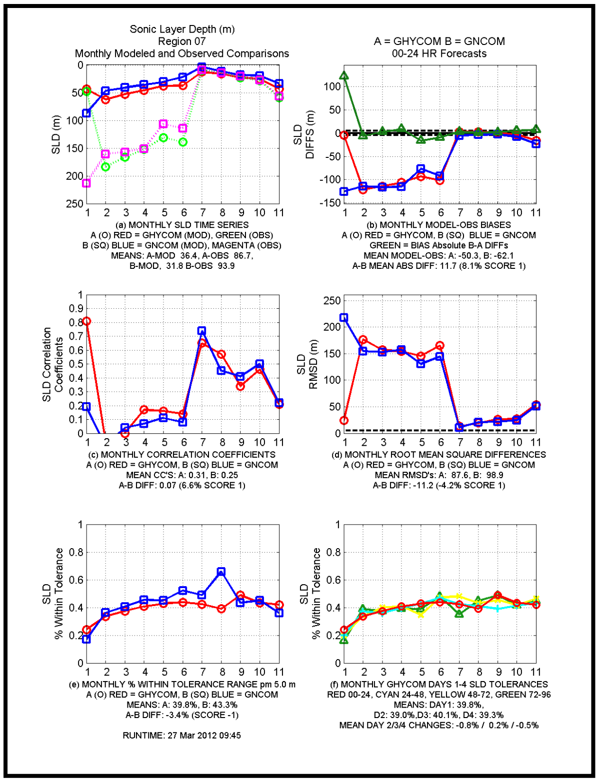


(b)

Figure 5.8. Same as Figure 5.5 for Region 5 (WPAC).

5.4.5. Eastern Pacific (EPAC) Metrics Time Lines: SLD in region 7 is a problem. For FEB to JUN both the model and observed SLDs for both models range from 160m to 130m, about 100m deeper than the model SLDs (*) until they rise to a more realistic 20m JUL (Figure 5.9.a). SLDs for both models rise continuously from 60m FEB to about 10m by JUL, then deepen to 50m by NOV. Our suspicion is that the RP-33 algorithm is calling the base of an observed secondary sound channel as the SLD. We looked at the MAY observations in 16 10°x10° sub-regions and the deep SLD values were generally consistent throughout region 7. As a result, the model biases were -120m to -100m FEB to JUN, then a much better -5m JUL to OCT. In FEB both model CCs are 0.00 to 0.10 FEB-JUN, rises to a "fair" 0.60 in JUL, then drops back to 0.20 by NOV. RMSD is 160m FEB-JUN, drops to about 10m in JUL, 20m AUG-OCT, and 50m NOV. SLD tolerance is a "poor" 40-50% throughout the period.

Observed region 7 SST starts at 14°C JAN, drops to 11.5°C APR, returns to 14°C by JUL, drops back to 13°C AUG, rises to 15°C OCT, and drops to 13°C by NOV (Figure 5.9.b). The HYCOM and GNCOM biases stay within the ±0.5°C range throughout the period except for a+1.0°C excursion by GNCOM during AUG-SEP. CC for both models is an "excellent" 0.90-0.95 JAN to MAY, "good" JUN to SEP, then back to 0.90 OCT-NOV, except for a 0.65 GNCOM value in SEP (*). HYCOM RMSD is 0.4°C JAN to APR, rises to 1.5°C by JUL, then drops to 0.50 by NOV. The GNCOM RMSD is 0.5 to 1.0°C higher or worse than HYCOM. HYCOM tolerance an excellent 80-100% JAN to APR, then drops to a poor 30% JUL, and rises back to 80% by NOV. GNCOM tolerance drops from 80% JAN to 30% JUL then follows the HYCOM pattern.

(a)


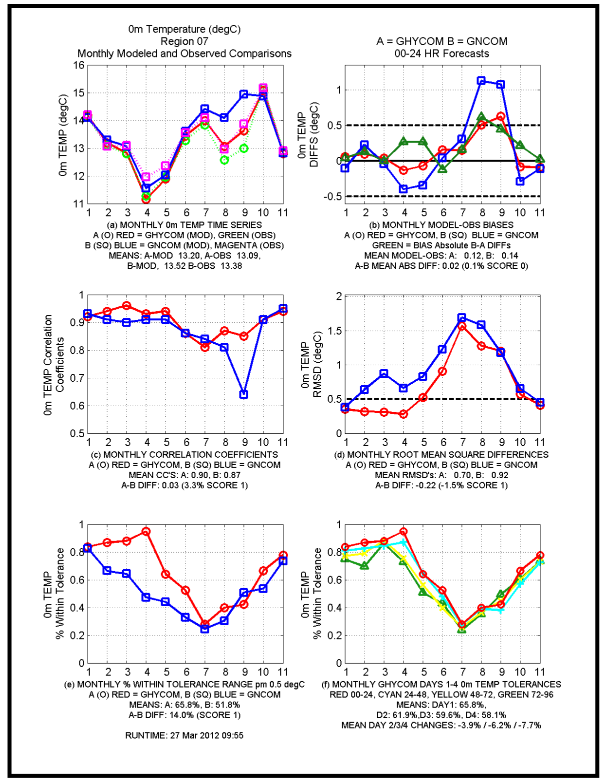
(b)


Figure 5.9. Same as Figure 5.5 for Region 7 (EPAC). The January SLD data are not included.

6. SUMMARY AND CONCLUSIONS

This OPTEST for the Global Ocean Forecast System (GOFS) is to demonstrate that, for Navy applications, the skill of the Hybrid Coordinate Ocean Model (HYCOM, GOFS 3.0) is equal to or better than that of the Global Navy Coastal Ocean Model (GNCOM, GOFS 2.6) that it is to replace. The statistical results show that HYCOM is slightly more skilled than GNCOM. Skill, however, is not the most important reason to move to HYCOM as it represents a new basis for global prediction. There are a number of important technical and scientific improvements in HYCOM that will lead to greater forecast skill and the future coupling of the ocean, atmosphere, ice, waves and land in the Earth Systems Prediction Capability (ESPC).

The OPTEST results are based on eleven-month statistical comparisons between observed and modeled temperature, salinity, and sonic layer depth (SLD) data. A scorecard approach is used, with a plus or minus one (±1) assigned to HYCOM or GNCOM according to four standard statistical metrics: bias, correlation coefficient, root mean square difference, and percentage of differences within an acceptable range. Five ocean properties are scored: SLD, temperature at the surface and 100m, salinity at the surface and 100m. Five ocean areas that are important to Navy operations have been evaluated (Figure 1.1): the Western North Atlantic, Eastern North Atlantic, North West Indian Ocean, Western North Pacific, and Eastern North Pacific. The statistics are based on the eleven-month period beginning January 2011 and ending November 2011, allowing us to look at the models during all four northern hemisphere seasons.

On average, HYCOM outscores GNCOM for 4% of the metrics, meaning we can say that HYCOM is provides as good as or slightly better products compared with GNCOM, driven mostly by better temperature and salinity results at the surface. GNCOM actually outscored HYCOM for SLD metrics.

An objective of this OPTEST is to demonstrate that model skill declined very little over the period of the forecast. An average 5% reduction in the “tolerance” metric over 96-hour is a strong indication that HYCOM forecasts lose skill very slowly.

The net four-metric, five-region SLD score is -3 of 20, for 15% in favor of GNCOM. There is a 0% change in skill over 96-hours. The T00 score is +2 of 20, for 10% in favor of HYCOM. There is an 8% decay in the HYCOM surface temperature forecast skill over 96-hours. The S00 score is +6, or 30% in favor of HYCOM. There is an 1% decay in the HYCOM surface salinity forecast skill over 96-hours. The T100 score is +2, for 10% points in favor of HYCOM. There is a 5% decay in HYCOM 100m temperature forecast skill over 96-hours. The S100 score is -3 of 20, for 15% in favor of GNCOM. There is a 5% decay in HYCOM 100m salinity forecast skill over 96-hours. Adding up the four-metrics, five-region results in a net BIAS score of +1 in favor of GNCOM, +1 for CC, +3 for RMSD, and -1 for TOLERANCE. The result is a total of +4 or 4/100 = 4% in favor of HYCOM.

By region, HYCOM outscored GNCOM in Region 1 (WLANT) by +1, in Region 2 (ELANT) by +2, and Region 7 (EPAC) by +10. On the other hand, GNCOM outscored HYCOM in Region 4 (NIO) by -5 and in Region 5 (WPAC) by -4.

In summary, HYCOM represents an improvement over GNCOM and we recommend that the AMOP declare GOFS 3.0 operational.

8. REFERENCES

The HYCOM Consortium Website: http://hycom.org/

Historical documentation: http://hycom.org/hycom/documentation

Some selected publications:

Chassignet, E.P., H.E. Hurlburt, E.J. Metzger, O.M. Smedstad, J. Cummings, G.R. Halliwell, R. Bleck, R. Baraille, A.J. Wallcraft, C. Lozano, H.L. Tolman, A. Srinivasan, S. Hankin, P. Cornillon, R. Weisberg, A. Barth, R. He, F. Werner, and J. Wilkin, 2009. U.S. GODAE: Global Ocean Prediction with the HYbrid Coordinate Ocean Model (HYCOM). Oceanography, 22(2), 64-75.

Hurlburt H.E., E.P. Chassignet, J.A. Cummings, A.B. Kara, E.J. Metzger, J.F. Shriver, O.M. Smedstad, A.J. Wallcraft, and C.N. Barron, 2008: Eddy-resolving global ocean prediction. In "Eddy-Resolving Ocean Modeling", M. Hecht and H. Hasumi, Eds., AGU Monograph Series, 353-382.

Chassignet, E.P., H.E. Hurlburt, O.M. Smedstad, G.R. Halliwell, P.J. Hogan, A.J. Wallcraft, R. Baraille, and R. Bleck, 2007: The HYCOM (HYbrid Coordinate Ocean Model) data assimilative system. J. Mar. Sys., 65, 60-83.

Chassignet, E.P., H.E. Hurlburt, O.M. Smedstad, G.R. Halliwell, A.J. Wallcraft, E.J. Metzger, B.O. Blanton, C. Lozano, D.B. Rao, P.J. Hogan, and A. Srinivasan, 2006: Generalized vertical coordinates for eddy-resolving global and coastal ocean forecasts. Oceanography, 19, 20-31.

NRL Validation Test Reports:

Metzger, E. J., H. E. Hurlburt, A.J. Wallcraft, J. F. Shriver, L. F. Smedstad, O. M. Smedstad, P. Thoppil, D. S. Franklin, 26 November 2008, "Validation Test Report for the Global Ocean Prediction System V3.0 – 1/12° HYCOM/NCODA: Phase I," NRL/MR/7320—08-9148.

Metzger, E. J., H. E. Hurlburt, A.J. Wallcraft, J. F. Shriver, T. L. Townsend, O. M. Smedstad, P. G. Thoppil, D. S. Franklin, G. Peggion, 23 February 2010, "Validation Test Report for the Global Ocean Prediction System V3.0 – 1/12° HYCOM/NCODA: Phase II," NRL/MR/7320—10-9236.

Metzger, E. J., P. G. Thoppil, O. M. Smedstad, D. S. Franklin, 05 November 2010, "Global Ocean Prediction System V3.0 Validation Test Report Addendum: Addition of the Diurnal Cycle," NRL/MR/7320—10-9305.

Metzger, E. J., P. G. Thoppil, G. Peggion, D. S. Franklin, O. M. Smedstad, 21 March 2012, "Global Ocean Prediction System V3.0 Validation Test Report Addendum: Provision of Boundary Conditions to the Relocatable Navy Coastal Ocean Model (NCOM)," NRL/MR/7320—12-9386.

Dykes, James D., 24 July 2011, "Implementation of the Automated Numerical Model Performance Metrics System," NRL/MR/7320—11-9353.

APPENDICES FOR THE HYCOM OPTEST REPORT

APPENDIX A. EVALUATION TECHNIQUES.

A.1. Summary of the Approach. Not all the graphics and analyses discussed here are in the OPTEST report. They can be viewed in the monthly Power Point slides for people who are interested. Note that this is a work in progress requiring a number of Matlab programs and analysis steps that are still under development and not ready for transition to automated operations. We start with a Matlab program that takes AutoMetrics files and extracts and formats temperature and salinity profiles, runs them through an initial quality control, computes skill statistics, and plots scatter plots of the concurrent (in time and space) observed (horizontal) versus model (vertical) data.

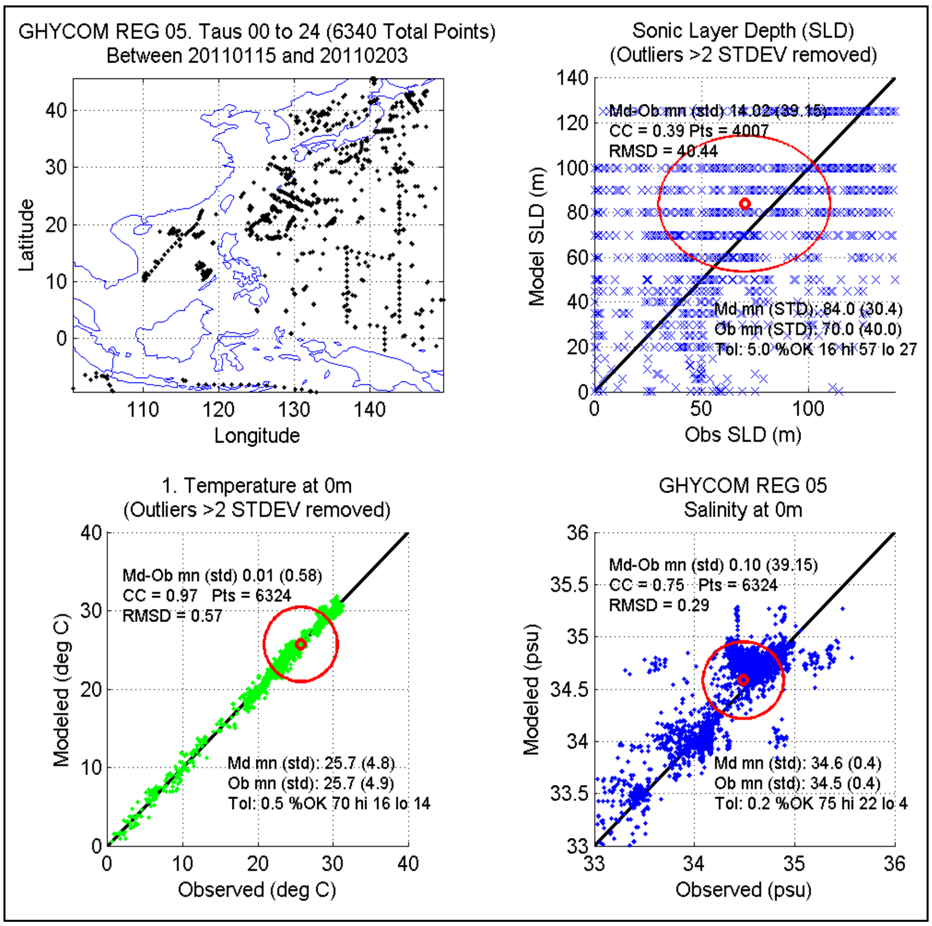


Figure A.1. (UL) Locations, periods, and time ranges of the analyzed data set. Scatter Plots of physical properties including (UR) Sonic Layer Depth (m) (SLD in BLUE, (LL) Surface Temperature (°C) in GREEN, and (LR) Salinity (psu) at surface in BLUE.

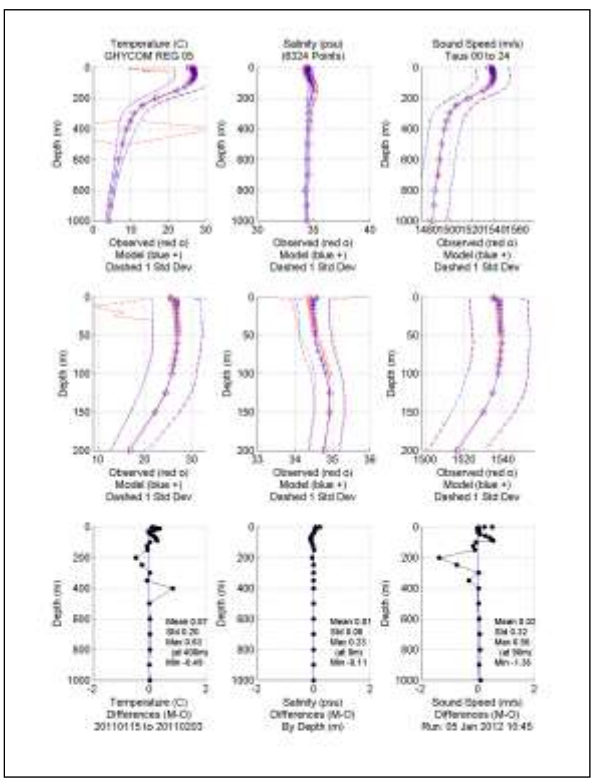
In the Figure A.1 scatter plots are indicated: the model; domain; the forecast period or day (first 24 hours (taus 00-24), second (24-48), third (48-72), or fourth (72-96)); total number of observation-model comparisons; and the period of the analysis (yyyymmdd). Outliers greater than the indicated model minus observation are removed before statistics are calculated. Note that plot scales are allowed to vary in accordance with the data although ordinate and abscissa are the same length. After the SLD plot, temperature and salinity at the surface (0.5m) and

100m depths (within the thermocline) are plotted on the first page; 10, 50, 150m on the second page; and 250, 500 and 1000m on the third.

On each plot, the BLACK diagonal indicates a "perfect fit," where every observed and model value pair would be equal. All differences are computed as MODEL MINUS OBSERVED values. The RED DOT shows the location of mean observed and modeled values. When the red dot is not on the diagonal, there is a model BIAS. If it is to the upper left of the diagonal, the model mean is larger than observed or there is a positive model bias. If it is in the lower right, the observation mean is larger, or bias is negative.

The RED ellipse encompasses one standard deviation of observations (horizontal axis) and model values (vertical axis) centered on the mean. Thus, a larger ellipse indicates more variability in the property. A circle means the variance of the model is reproducing the variance of the observations or "nature."

Statistics for the analysis are printed in the upper left including: The MEAN of model minus observed value (i.e., BIAS) followed by the standard deviation of the bias in parentheses. The CORRELATION COEFFICIENT of the comparison is based on the least-squares fit (Pearson method). The number of points used for each property and level is indicated. The ROOT MEAN SQUARE OF THE DIFFERENCE (RMSD) is shown. Statistics in the lower right include mean (standard deviation) of both the model and the observed data. The TOLERANCE values are discussed below.

(a) 

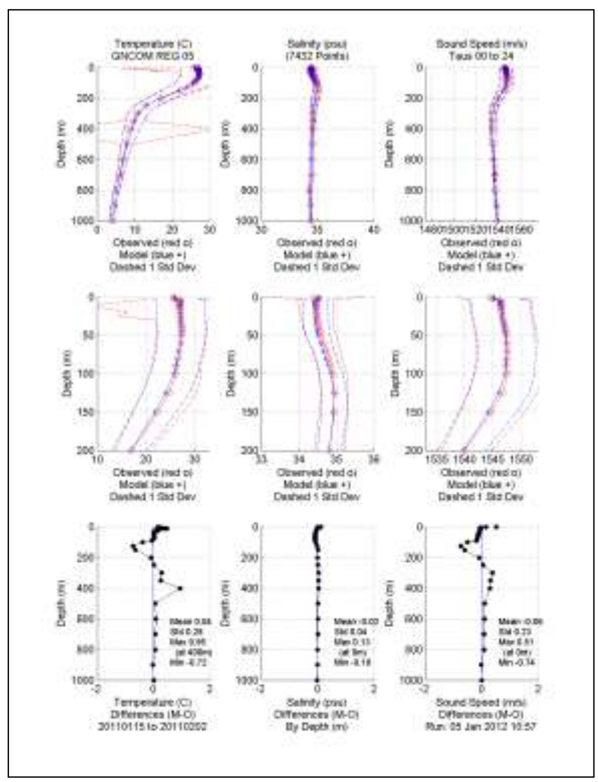
(b) 

Figure A.2. Comparing mean profiles for (a) HYCOM and (b) GNCOM. Temperature, salinity, and sound speed to 1000m are plotted in the upper graphs for the model data (BLUE +) and observed (RED o) data, averaged by layer. A one standard deviation envelope is dashed. The center graphs show the same data for the upper 200m. The lower graphs show model minus observed differences by layer. Text includes the mean, standard deviation, and maximum positive and negative differences (with depth).

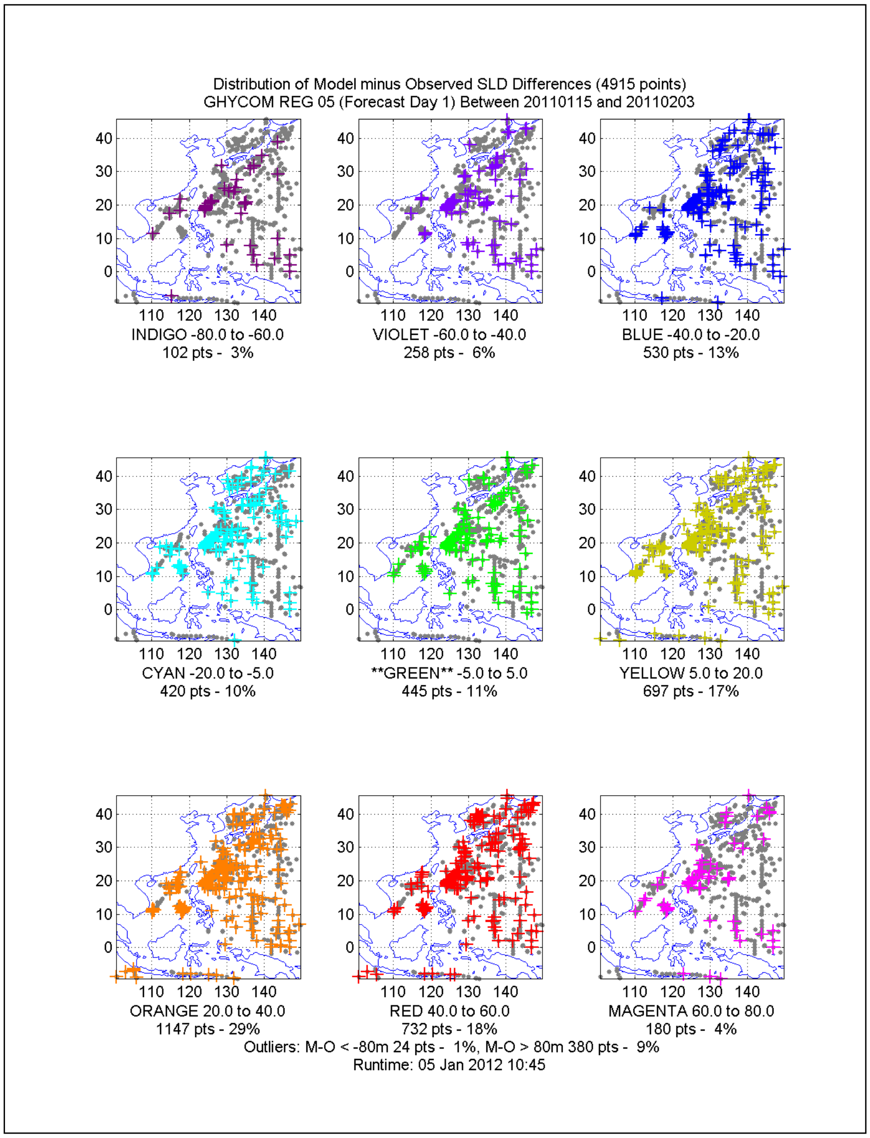


Figure A.3. Distribution of SLD Model minus observation differences, color-coded by bins. Bin boundaries are: -80, -60, -40, -20, -5, +5, +20, +40, +60, and +80m. The percentages of points in each range are given.

We are using Figure A.3 to look for ocean regions where there may be good or bad concentrations of forecasts. Boundaries of the green area (center plot) correspond to the tolerance range of ±5m discussed below.

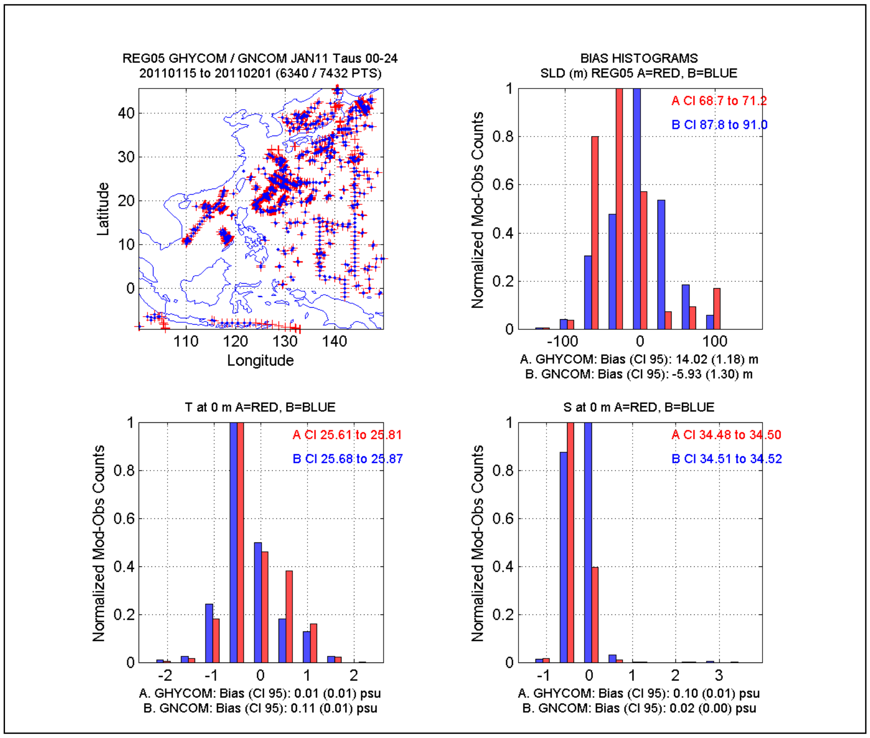


Figure A.4. Histograms of SLD and surface temperature and salinity biases for HYCOM (A, RED) and GNCOM (B, BLUE). The 95% confidence intervals are indicated, along with means and standard deviations of the metrics.

After running the above analyses for the first through fourth day forecasts (when available), the information is combined as a series of bar plots using another Matlab program. The main objectives are to compare the HYCOM and GNCOM results and demonstrate that there is little decay in model skill over the 96-hour period.

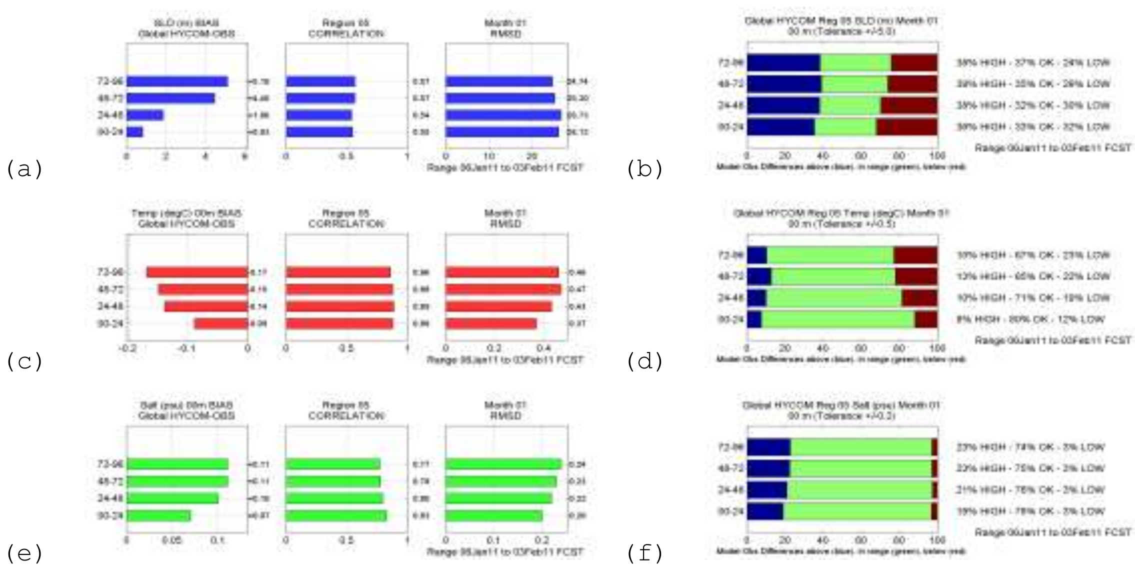


Figure A.5. Example of bar plots of Statistics for (a) the Sonic Layer Depth (BLUE), (c) Surface Temperature (RED), and (e) Surface Salinity (GREEN). From bottom to top, the first through fourth forecast days are plotted, with each value printed to the right. From left to right, the plots show model bias, correlation coefficient, and Root Mean Square of the Differences (RMSD). To the right (b,d,f) are plotted the tolerance or acceptable ranges of property model minus observed differences, with the BLUE section of the bar indicating the percent of differences that are above, GREEN for differences that are within, and RED for differences below the selected tolerance range. Values are indicated to the right.

The tolerance approach is based on a similar metric developed by the National Oceanic and Atmospheric Administration (NOAA) National Ocean Service (NOS) (see reference X). We set some values for expected product accuracy and then indicate how many of the model-minus-observed differences fall within these limits. The ranges selected for this analysis are: Sonic Layer Depth (±5.0 m), Temperature (±0.5 °C), Salinity (±0.2 psu), and Sound Speed (±2.0 m/s). While somewhat arbitrary (i.e., there is no Navy requirement that ocean models meet these values specifically), they do represent attainable accuracy goals. The best results would be all GREEN or 100% within the specified tolerance. Note that a net positive (negative) bias in the bar plots should result in larger BLUE (RED) bars on the tolerance plots. Also note the correspondence between GREEN tolerance bar and the GREEN central plot in Figure A.3.

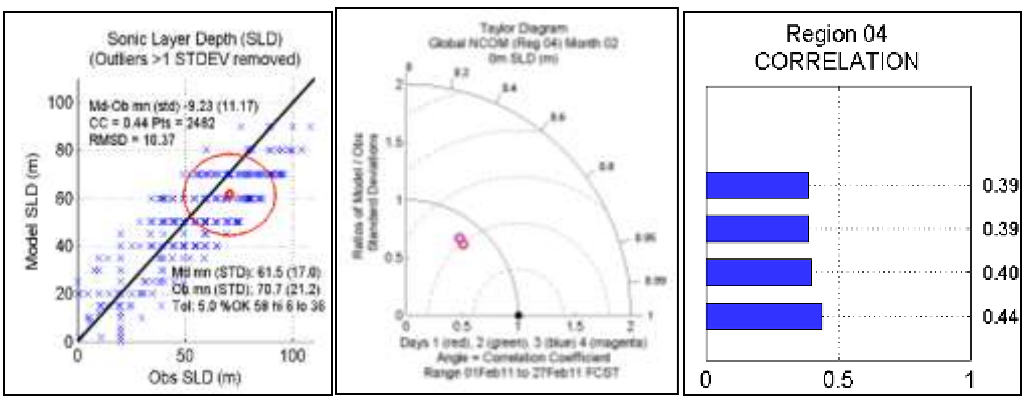


Figure A.5. A Taylor Plot of the SLD data. Dot colors show forecast days, with day 1 = RED, 2 = GREEN, 3 = BLUE, and 4 = MAGENTA.

On a Taylor plot, the ratios of the model to observed standard deviations are plotted on the radial lines. A perfect ratio of 1.0 would lie on the "1" arc. Points to the left (right) indicate that the model variance is less (more) than that found in the observations or "nature." Note that this corresponds to the ellipse axes on the scatter plots which means a circle would lie on the "1" arc. The correlation coefficient is indicated by the clockwise angle to the left of vertical on a log scale. A "perfect" solution (1:1 ratio and 1.0 correlation coefficient) would be at the BLACK dot.

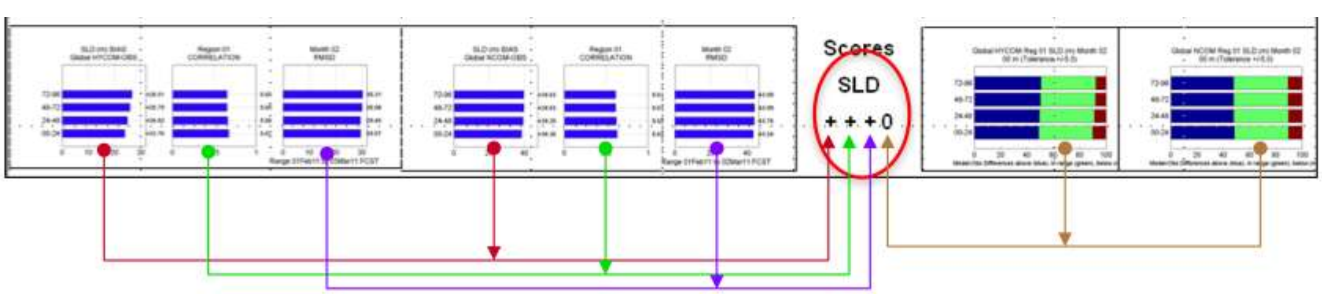


Figure A.6. Scoring example for comparisons between HYCOM and GNCOM.

Only the 00-24-hour forecast periods (the lowest bars) are evaluated for a score. For leftmost bias bars (RED connectors), a smaller value is scored +1 for HYCOM. For the center correlation coefficient bars (GREEN), a higher value is scored +1 for HYCOM. For the rightmost bars (PURPLE), a lower RMSE is scored +1 for HYCOM. For the tolerance plots (BROWN), a larger green bar (differences within the set range) is scored +1 for HYCOM. Any differences within ±1% are scored 0. Rather than assign scores by eyeball, they are automated as follows. For this discussion, A = HYCOM and B = GNCOM. We only score the tau 00-24 forecast comparisons.

```
% AUTO SCORE RUN FOR [A] GHYCOM versus [B] GNCOM
%   MAY11 (01MAY11-03JUN11)    [RUNTIME: 01 Sep 2011 16:08]
%   REGION 01:   A = GHYCOM,   B = GNCOM
% ---------------------------------------------------------------------------

% 1. METRIC SCORES: BIAS - Negative means A (HYCOM) bias lower (better) than B (GNCOM)

%  SLD REG DEPTH MON TAU:  A_MEAN  B_MEAN  A_METRIC (%)    B_METRIC (%)    A-B_METRIC  PCT_A_MN  SCORE
   05  01  0000  05  00    26.64   16.84    0.76 (  2.9)    0.02 (  0.1)    0.74     2.7%     -1
   05  01  0000  05  24    25.72   16.84    2.73 ( 10.6)    0.02 (  0.1)    2.71    10.5%     -1
   05  01  0000  05  48    25.00   16.84    1.78 (  7.1)    0.02 (  0.1)    1.76     7.0%     -1
   05  01  0000  05  72    27.37   16.84    1.23 (  4.5)    0.02 (  0.1)    1.21     4.4%     -1
```

On the first row of the bias metric (tau 00), we assume that a lower bias is better. We normalize the results as a ratio between the bias and mean observed SLD. For example, in the first row the A observed mean SLD is 26.64m and the bias (A_METRIC) is 0.76m for a bias/observed ration of 2.9%. The B observed mean SLD is 16.84m and the bias (B_METRIC) is 0.02m for a bias/observed ratio of 0.1%. The A-B bias difference is +0.74m for an A-B ratio difference is 2.7%. The positive difference indicates the A ratio is higher than the B ratio or B gives a better result in terms of a lower normalized bias, so the HYCOM score is -1.

```
% 2. METRIC SCORES: CORRELATION COEFFICIENT - Positive mean A (HYCOM) correlation higher (better) than B (GNCOM)

% SLD REG DEPTH MON TAU:  A_METRIC  B_METRIC   A-B_METRIC  PCT_A_METRIC  SCORE
   05  01  0000  05  00       0.93       0.94      -0.01   -1.1%     +0
   05  01  0000  05  24       0.91       0.94      -0.03   -3.3%     -1
   05  01  0000  05  48       0.90       0.94      -0.04   -4.4%     -1
   05  01  0000  05  72       0.95       0.94       0.01    1.1%     +1
```

On the first row of the correlation coefficient (CC) metric, we assume that a higher value is better. The A observed mean CC is 0.93 and the B CC is 0.94 for an A-B difference of -0.01. The difference is within 1% so the score is 0.

```
% 3. METRIC SCORES: RMSE - negative mean A (HYCOM) RMSE lower (better) than B (GNCOM)

% SLD REG DEPTH MON TAU:  A_MEAN  B_MEAN  A_METRIC (%)    B_METRIC (%)    A-B_METRIC  PCT_A_MN  SCORE
   05  01  0000  05  00:   19.29   12.46   19.29 ( 72.4)   12.46 ( 74.0)    6.83    -1.6%     +1
   05  01  0000  05  24:   19.09   12.46   19.09 ( 74.2)   12.46 ( 74.0)    6.63     0.2%     +0
   05  01  0000  05  48:   22.15   12.46   22.15 ( 88.6)   12.46 ( 74.0)    9.69    14.6%     -1
   05  01  0000  05  72:   19.25   12.46   19.25 ( 70.3)   12.46 ( 74.0)    6.79    -3.7%     +1
```

A lower RMSD is considered better. On the first row, the A_MEAN RMSD/observed ratio is 72.4% and the B_MEAN RMSD/observed ratio (also known as the “scatter index”) is 74.0%. The A-B difference is -1.6% so A is lower or better and HYCOM is scored +1.

```
% 4. METRIC SCORES: TOLERANCE RANGE - Positive mean A (HYCOM) range higher (better) than B (GNCOM)

% SLD  REG DEPTH MON TAU:  A_POINTS  B_POINTS  A_TOLE  B_TOLE   A-B_METRIC  PCT_A_METRIC  SCORE
   05  01  0000  05  00     1865      6212       56%      70%    -13.5%  -24.0%     -1
   05  01  0000  05  24     1520      6212       53%      70%    -16.9%  -32.0%     -1
   05  01  0000  05  48     1461      6212       55%      70%    -15.0%  -27.3%     -1
   05  01  0000  05  72     1287      6212       57%      70%    -12.7%  -22.2%     -1
```

A higher tolerance value is considered better. On the first row, the A metric is 56% and the B metric is 70% for a -13.5% difference, so B is better and HYCOM scores -1.

The results are summarized as follows. Scores for SLD, temperature at the surface and 100m, and salinity at the surface and 100m are added up for a possible total of +20 for HYCOM in each region. Here the sum is +2, meaning that HYCOM scores higher 10% of the time.

```
SUMMARY SCORE SHEET for REGION 01, [A] GHYCOM versus [B] GNCOM
   MAY11 (01MAY11-03JUN11, Taus 00-24)  [runtime 01 Sep 2011 16:08]
   A = GHYCOM, B = GNCOM

         BIAS   CORR   RMSE  TOLERANCE  TOTAL FOR A
   SLD    -1     0      1     -1        SUM: -1
   T000    1     1      1      1        SUM:  4
   T100    0     0      0     -1        SUM: -1
   S000   -1     1      0     -1        SUM: -1
   S100    0     0      0      1        SUM:  1
---------------------------------------------------
 TOTALS   -1     2      2     -1        SUM:  2
```

Five regions are evaluated for a total potential monthly HYCOM score of +100. The above results for each property and region are entered into an Excel table.

Table A.7. Summary Table for the December 2010 scores. Positive (negative) results indicate that HYCOM (GNCOM) scored higher.

DEC 2010	1 VLANT	2 ELANT	4 VIO	5 VPAC	7 EPAC	PROPERTY TOTALS
SLD	3	-1	1	3	3	9
T 00m	3	0	2	0	3	8
S 00m	-2	0	1	-3	-1	-5
T 100m	0	1	3	3	-1	6
S 100m	0	-1	0	0	1	0
REGION TOTALS	4	-1	7	3	5	18

This scorecard approach has been developed to parallel the atmospheric model metrics system used at FNMOC.

These results are summarized on a series of bar graphs that show model comparisons over time:

In the upper left of Figure A.9 are plotted the property MEAN values by month. CIRCLES are HYCOM, with RED the model monthly mean and GREEN the observed mean. SQUARES are GNCOM, with BLUE the model monthly mean and MAGENTA the observed mean. In the upper right are plotted the model minus observed BIAS. The RED CIRCLE is

HYCOM and the BLUE SQUARE is GNCOM bias. The BLACK + shows the HYCOM minus GNCOM bias differences. A LOWER bias is better, so negative difference means HYCOM is better.

In the middle left are plotted CORRELATION COEFFICIENTS (CC), with a RED CIRCLE for HYCOM and a BLUE SQUARE for GNCOM. A HIGHER CC is better, so RED above BLUE means HYCOM is better. In the Middle Right is the ROOT MEAN SQUARE DIFFERENCES (RMSD), with a RED CIRCLE for HYCOM and a BLUE SQUARE for GNCOM. A LOWER RMSD is better, so RED below BLUE means HYCOM scored better.

In the lower left are plotted the Percent of Differences within Given TOLERANCE Range, with a RED CIRCLE for HYCOM and a BLUE SQUARE for GNCOM. A HIGHER value is better, so RED above BLUE means HYCOM scores higher. In the Lower Right are plotted bar graphs for TOLERANCE Ranges for HYCOM forecast, with BLUE for the 00-24, CYAN 24-48, YELLOW 48-72, RED 72-96-hour forecasts. The objective is to demonstrate decay in skill over 4-day forecast period.

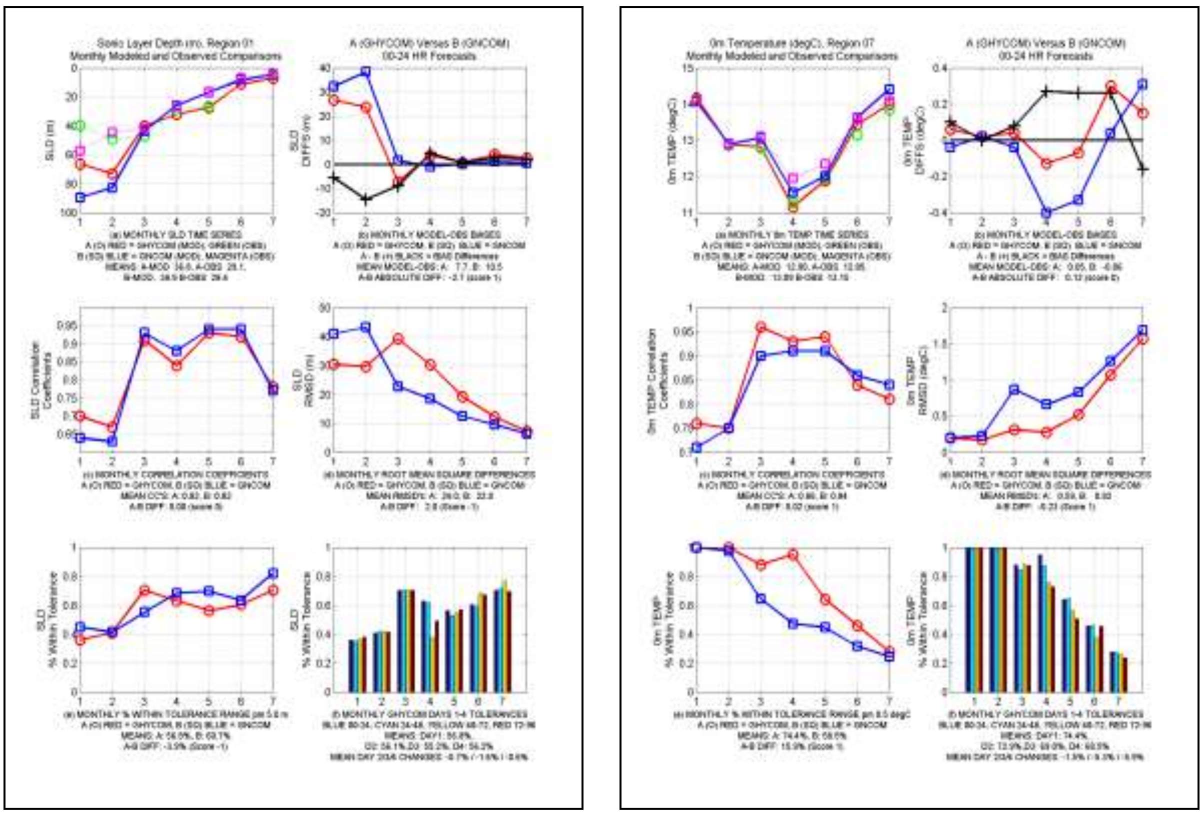
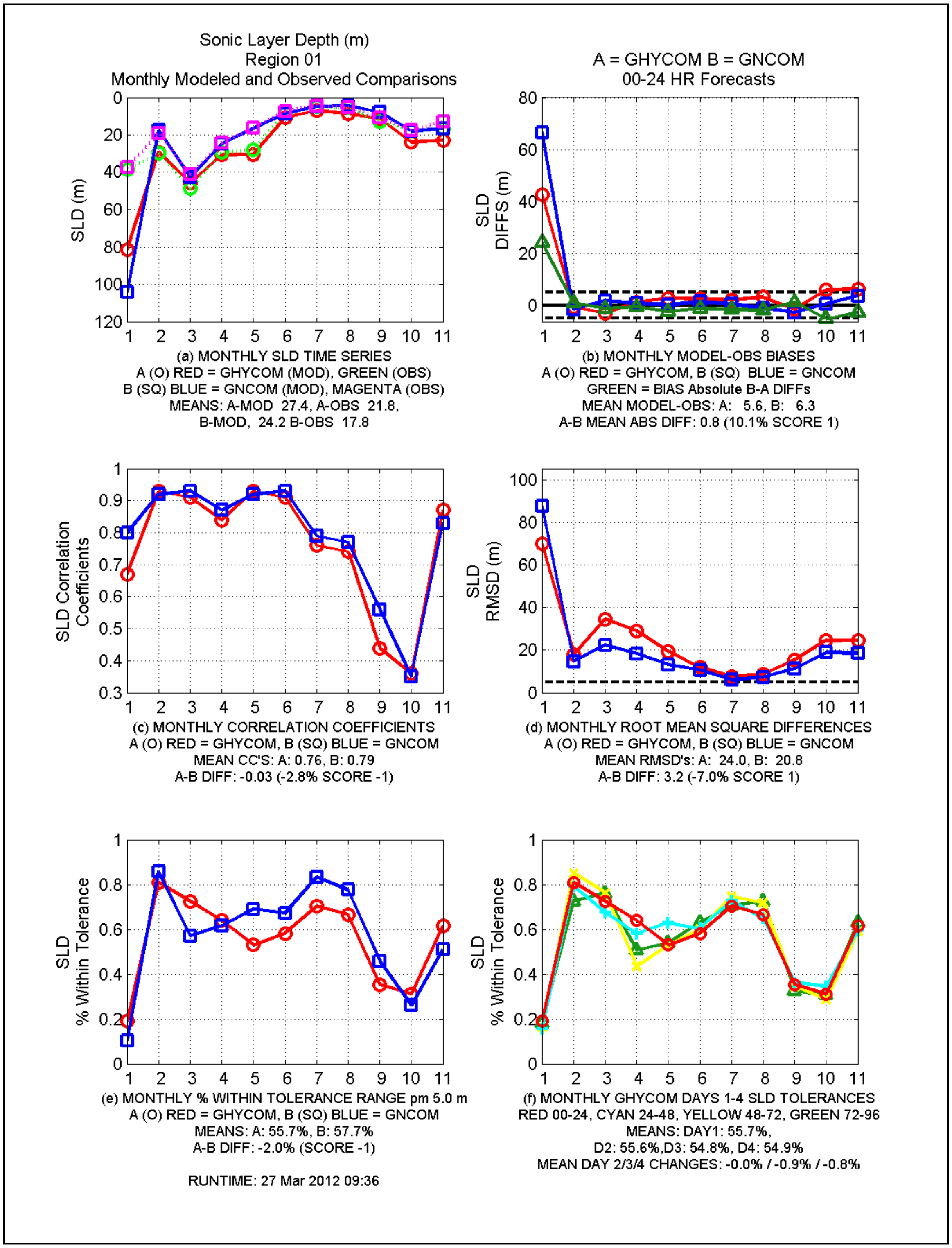


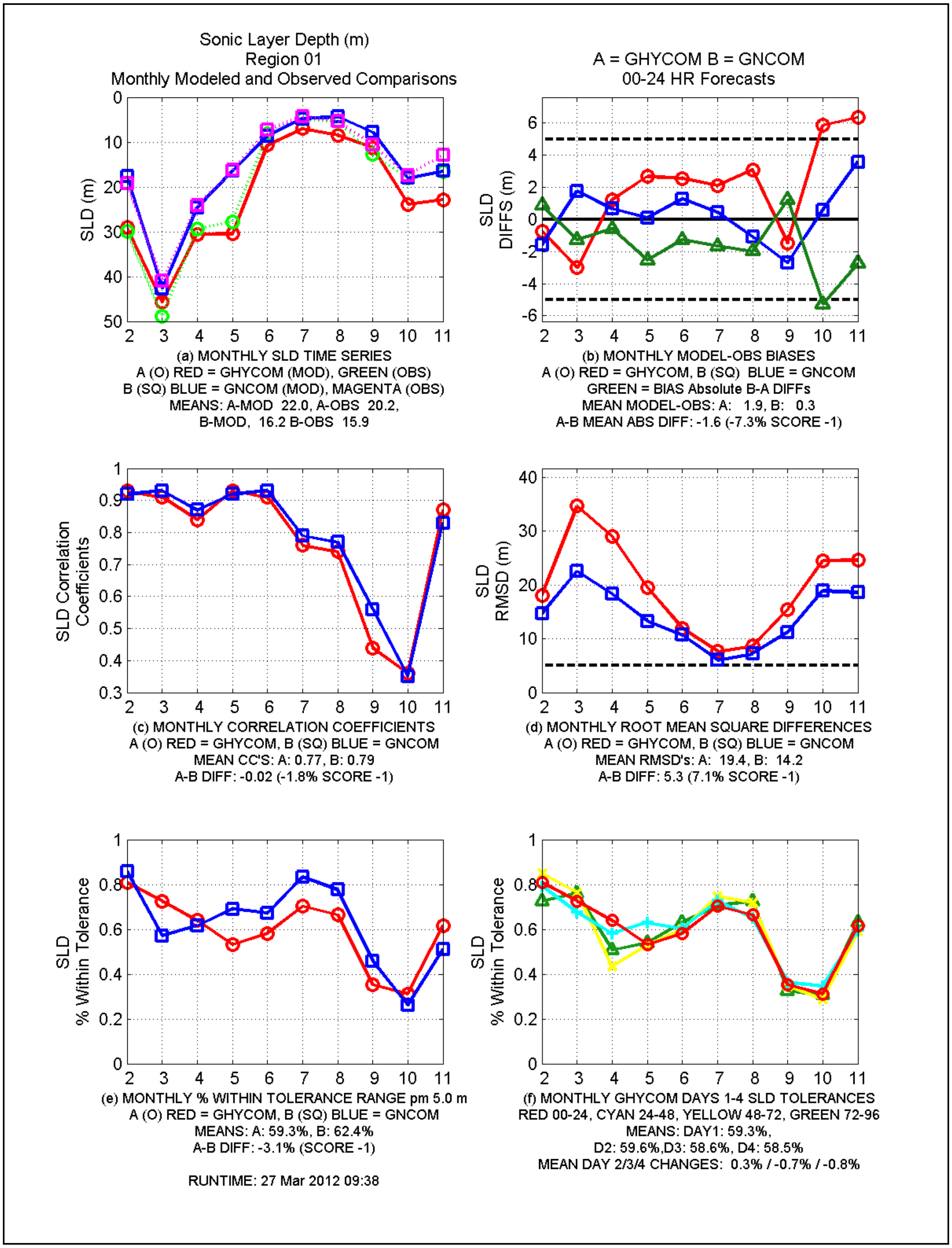
Figure A.9. Time series of metrics Region 1 SLD (left) and Region 7 surface water temperature. See the text for a description of the panels.

APPENDIX B. TIME SERIES PLOTS OF SONIC LAYER DEPTH, SURFACE TEMPERATURE, SURFACE SALINITY, 100M TEMPERATURE, AND 100M SALINITY. (Full month-by-month analyses and graphics are available upon request)

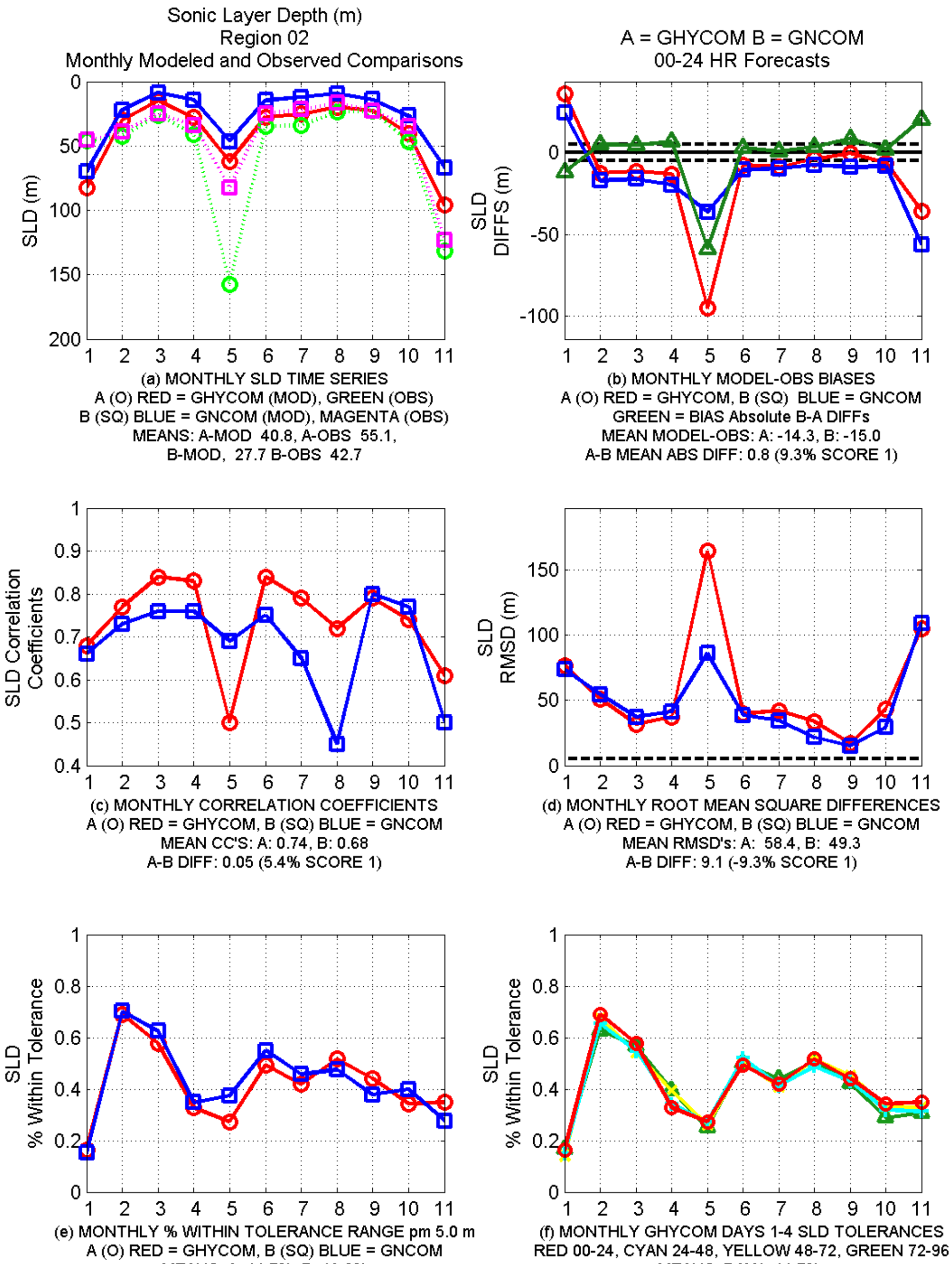
B.1.1.a. Sonic Layer Depths - Region 1 (Western Atlantic) with JANUARY 2011



B.1.1.b. Sonic Layer Depths - Region 1 (Western Atlantic) without JANUARY 2011



B.1.2. Sonic Layer Depths - Region 2 (Eastern Atlantic and Mediterranean)



(a) MONTHLY SLD TIME SERIES
A (O) RED = GHYCOM (MOD), GREEN (OBS)
B (SQ) BLUE = GNCOM (MOD), MAGENTA (OBS)
MEANS: A-MOD 40.8, A-OBS 55.1,
B-MOD, 27.7 B-OBS 42.7

(b) MONTHLY MODEL-OBS BIASES
A (O) RED = GHYCOM, B (SQ) BLUE = GNCOM
GREEN = BIAS Absolute B-A DIFFs
MEAN MODEL-OBS: A: -14.3, B: -15.0
A-B MEAN ABS DIFF: 0.8 (9.3% SCORE 1)

(c) MONTHLY CORRELATION COEFFICIENTS
A (O) RED = GHYCOM, B (SQ) BLUE = GNCOM
MEAN CC'S: A: 0.74, B: 0.68
A-B DIFF: 0.05 (5.4% SCORE 1)

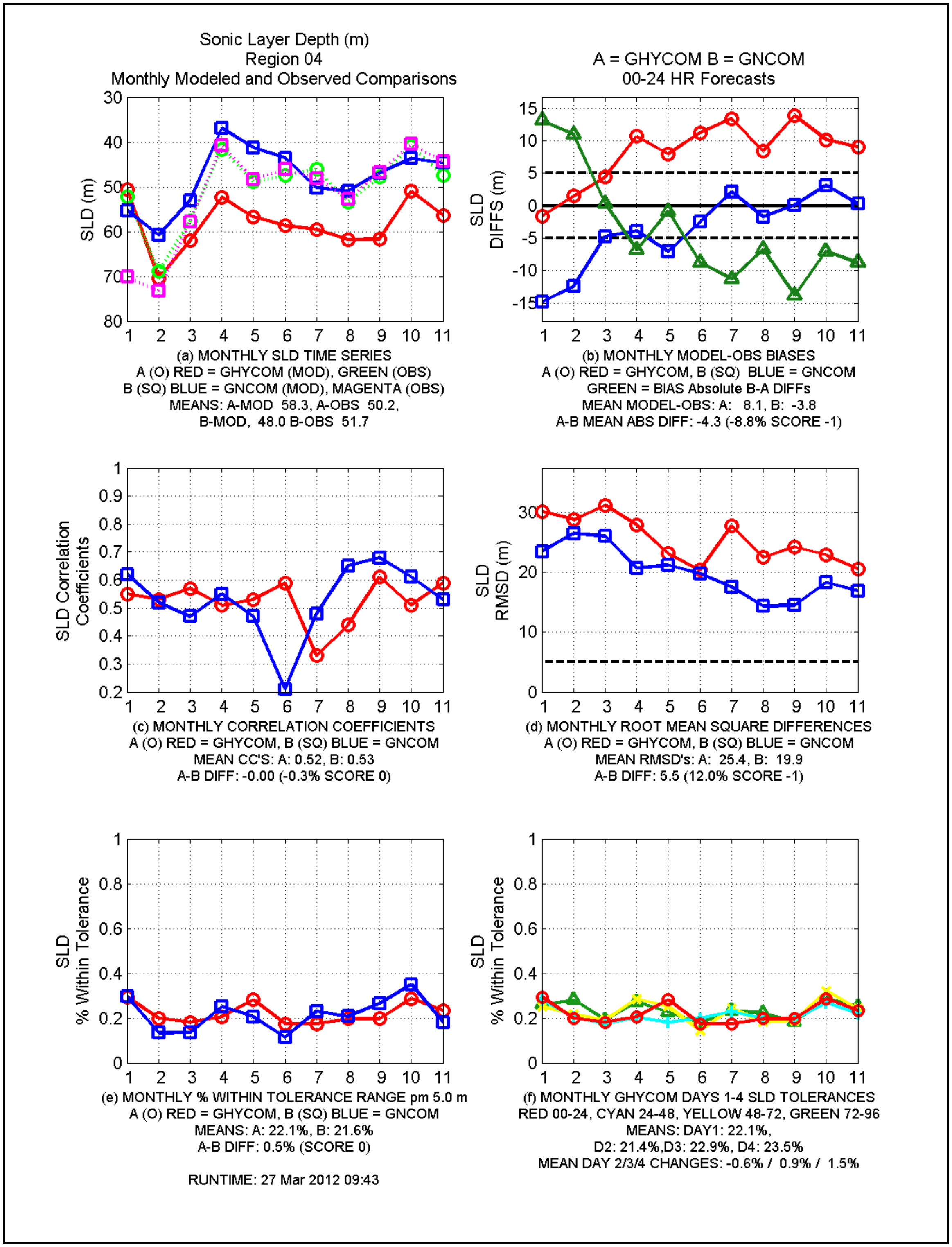
(d) MONTHLY ROOT MEAN SQUARE DIFFERENCES
A (O) RED = GHYCOM, B (SQ) BLUE = GNCOM
MEAN RMSD's: A: 58.4, B: 49.3
A-B DIFF: 9.1 (-9.3% SCORE 1)

(e) MONTHLY % WITHIN TOLERANCE RANGE pm 5.0 m
A (O) RED = GHYCOM, B (SQ) BLUE = GNCOM
MEANS: A: 41.7%, B: 43.0%
A-B DIFF: -1.3% (SCORE -1)

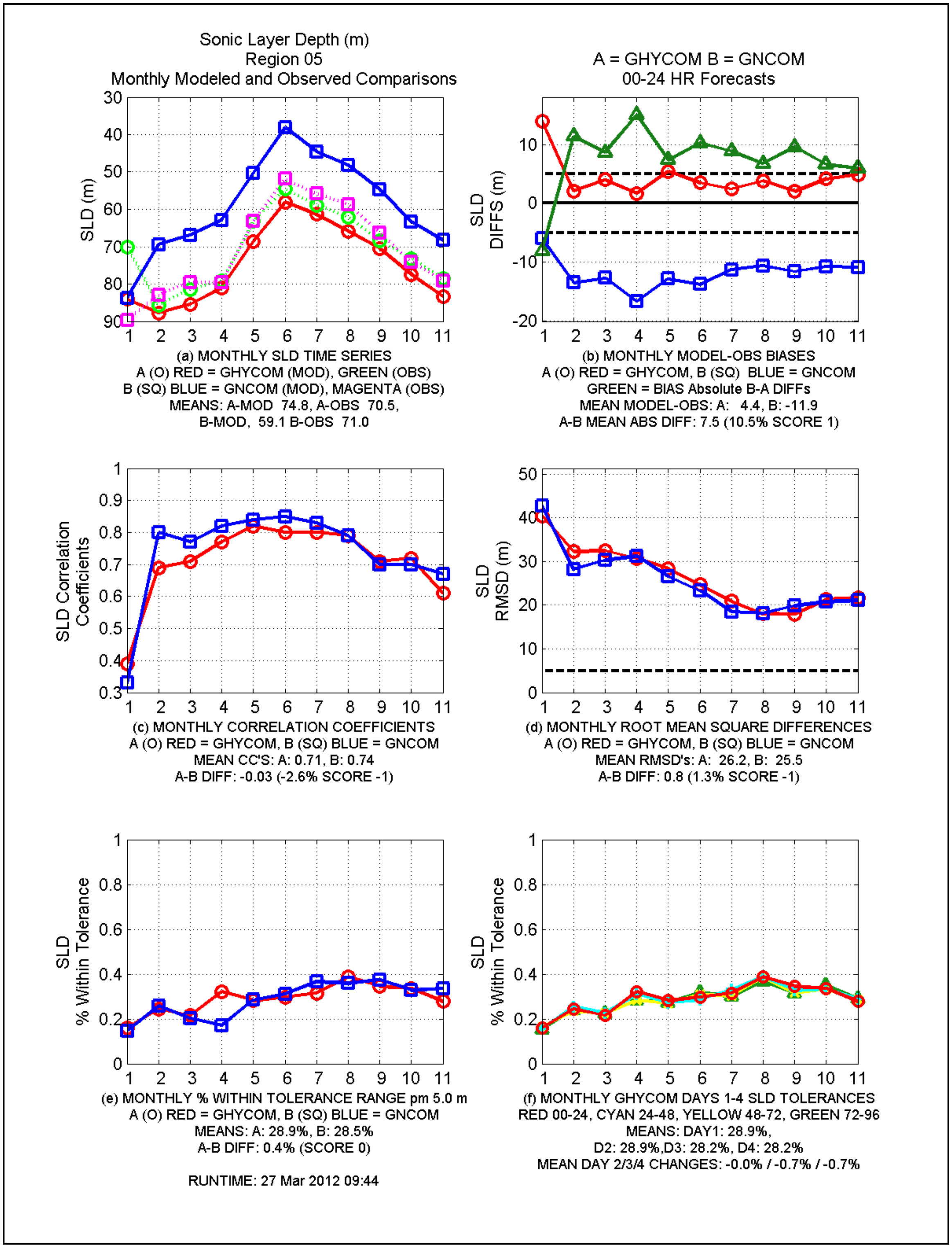
RUNTIME: 27 Mar 2012 09:35

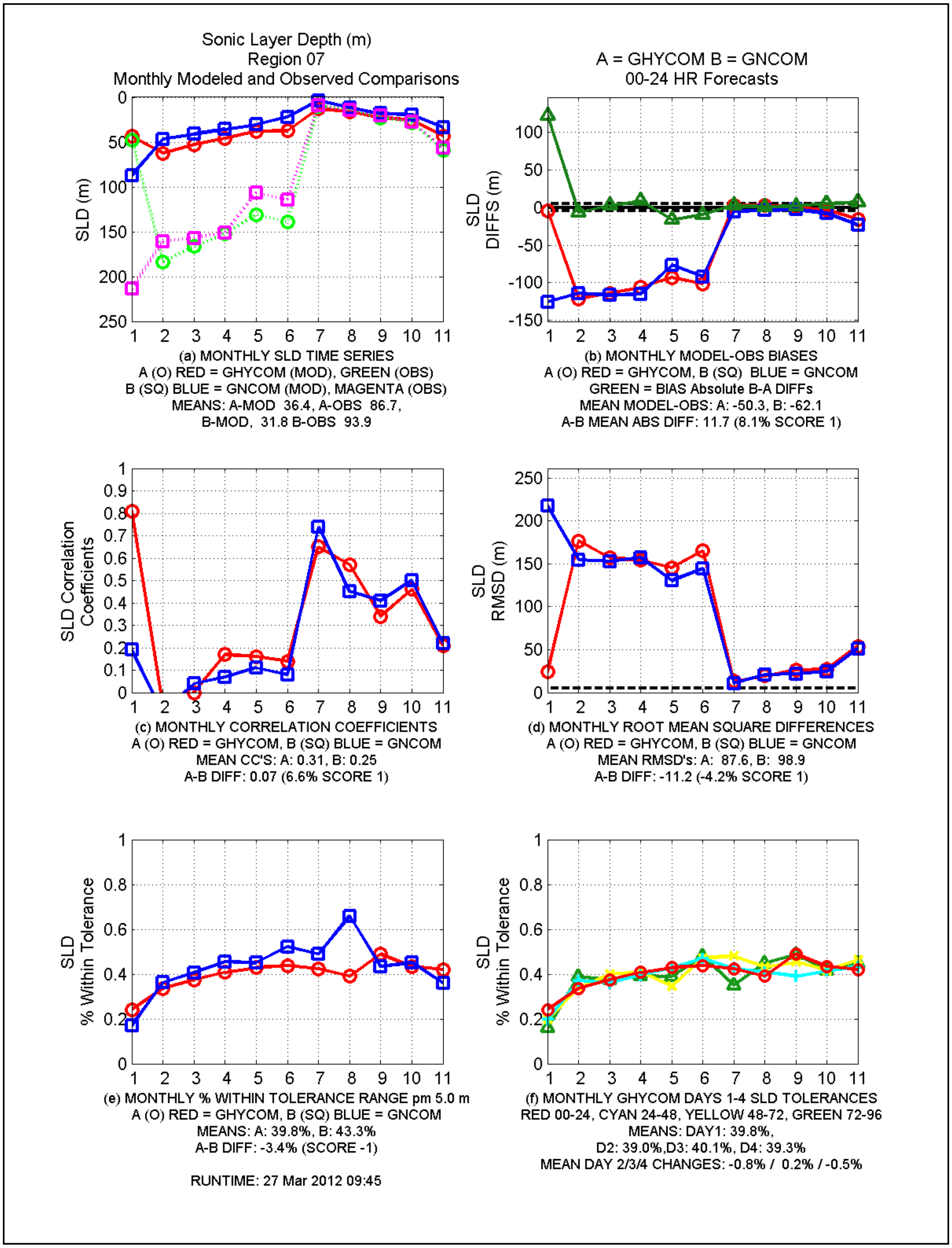
(f) MONTHLY GHYCOM DAYS 1-4 SLD TOLERANCES
RED 00-24, CYAN 24-48, YELLOW 48-72, GREEN 72-96
MEANS: DAY1: 41.7%,
D2: 40.4%,D3: 41.3%, D4: 40.7%
MEAN DAY 2/3/4 CHANGES: -1.3% / -0.4% / -1.0%

B.1.4. Sonic Layer Depths – Region 4 (Northwest Indian Ocean)



B.1.5. Sonic Layer Depths - Region 5 (Northwestern Pacific)





(a) MONTHLY SLD TIME SERIES
A (O) RED = GHYCOM (MOD), GREEN (OBS)
B (SQ) BLUE = GNCOM (MOD), MAGENTA (OBS)
MEANS: A-MOD 36.4, A-OBS 86.7,
B-MOD, 31.8 B-OBS 93.9

(b) MONTHLY MODEL-OBS BIASES
A (O) RED = GHYCOM, B (SQ) BLUE = GNCOM
GREEN = BIAS Absolute B-A DIFFs
MEAN MODEL-OBS: A: -50.3, B: -62.1
A-B MEAN ABS DIFF: 11.7 (8.1% SCORE 1)

(c) MONTHLY CORRELATION COEFFICIENTS
A (O) RED = GHYCOM, B (SQ) BLUE = GNCOM
MEAN CC'S: A: 0.31, B: 0.25
A-B DIFF: 0.07 (6.6% SCORE 1)

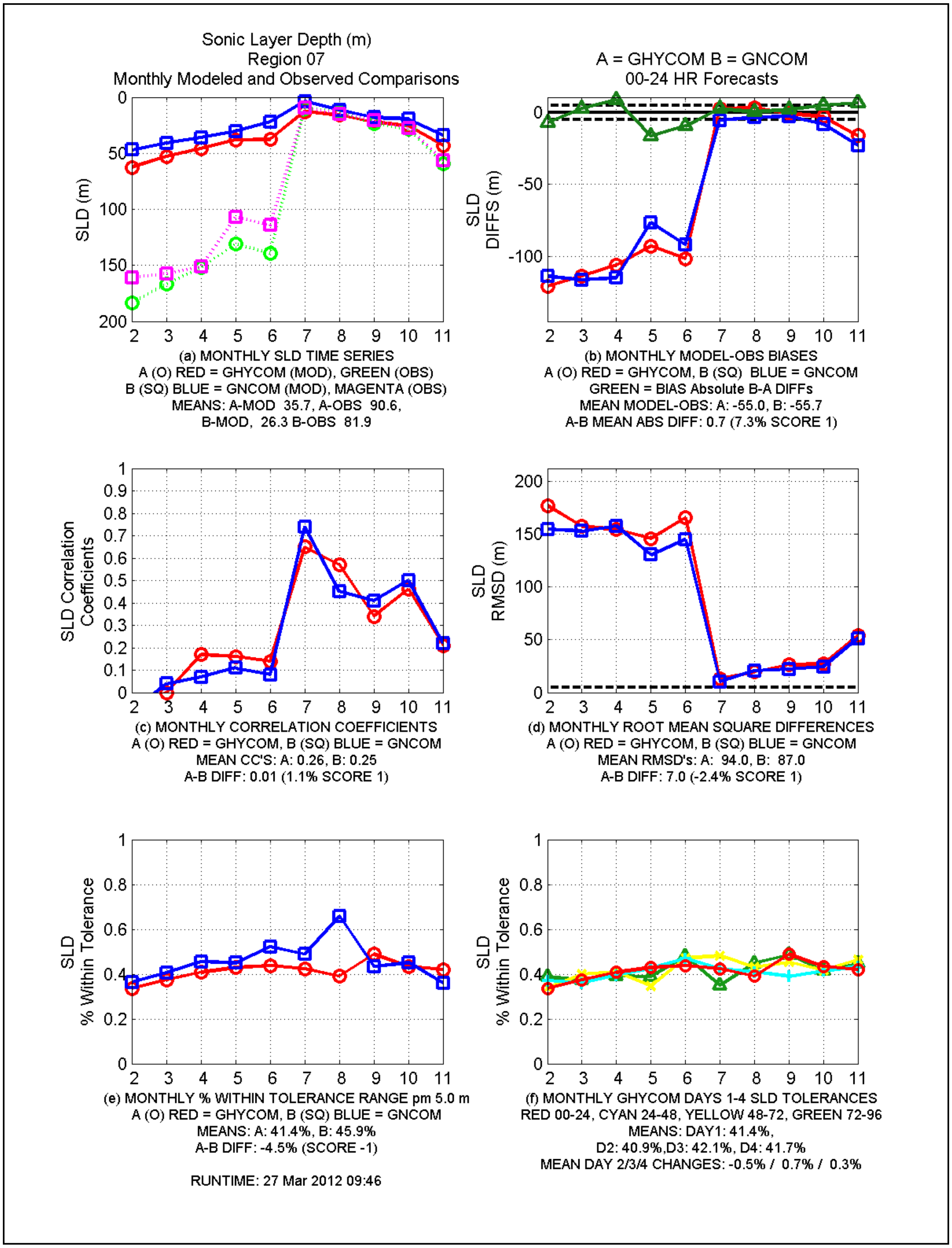
(d) MONTHLY ROOT MEAN SQUARE DIFFERENCES
A (O) RED = GHYCOM, B (SQ) BLUE = GNCOM
MEAN RMSD's: A: 87.6, B: 98.9
A-B DIFF: -11.2 (-4.2% SCORE 1)

(e) MONTHLY % WITHIN TOLERANCE RANGE pm 5.0 m
A (O) RED = GHYCOM, B (SQ) BLUE = GNCOM
MEANS: A: 39.8%, B: 43.3%
A-B DIFF: -3.4% (SCORE -1)

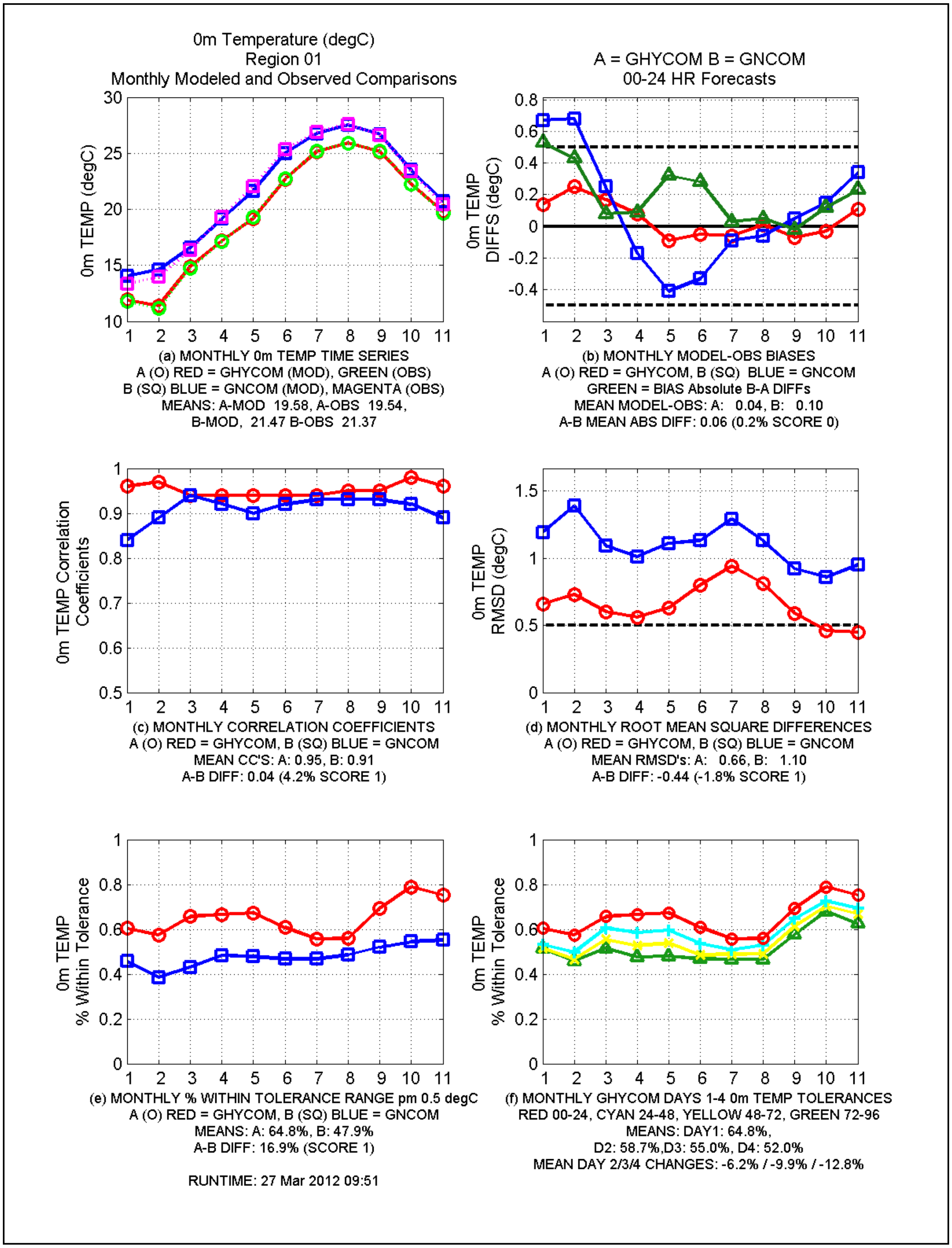
RUNTIME: 27 Mar 2012 09:45

(f) MONTHLY GHYCOM DAYS 1-4 SLD TOLERANCES
RED 00-24, CYAN 24-48, YELLOW 48-72, GREEN 72-96
MEANS: DAY1: 39.8%,
D2: 39.0%,D3: 40.1%, D4: 39.3%
MEAN DAY 2/3/4 CHANGES: -0.8% / 0.2% / -0.5%

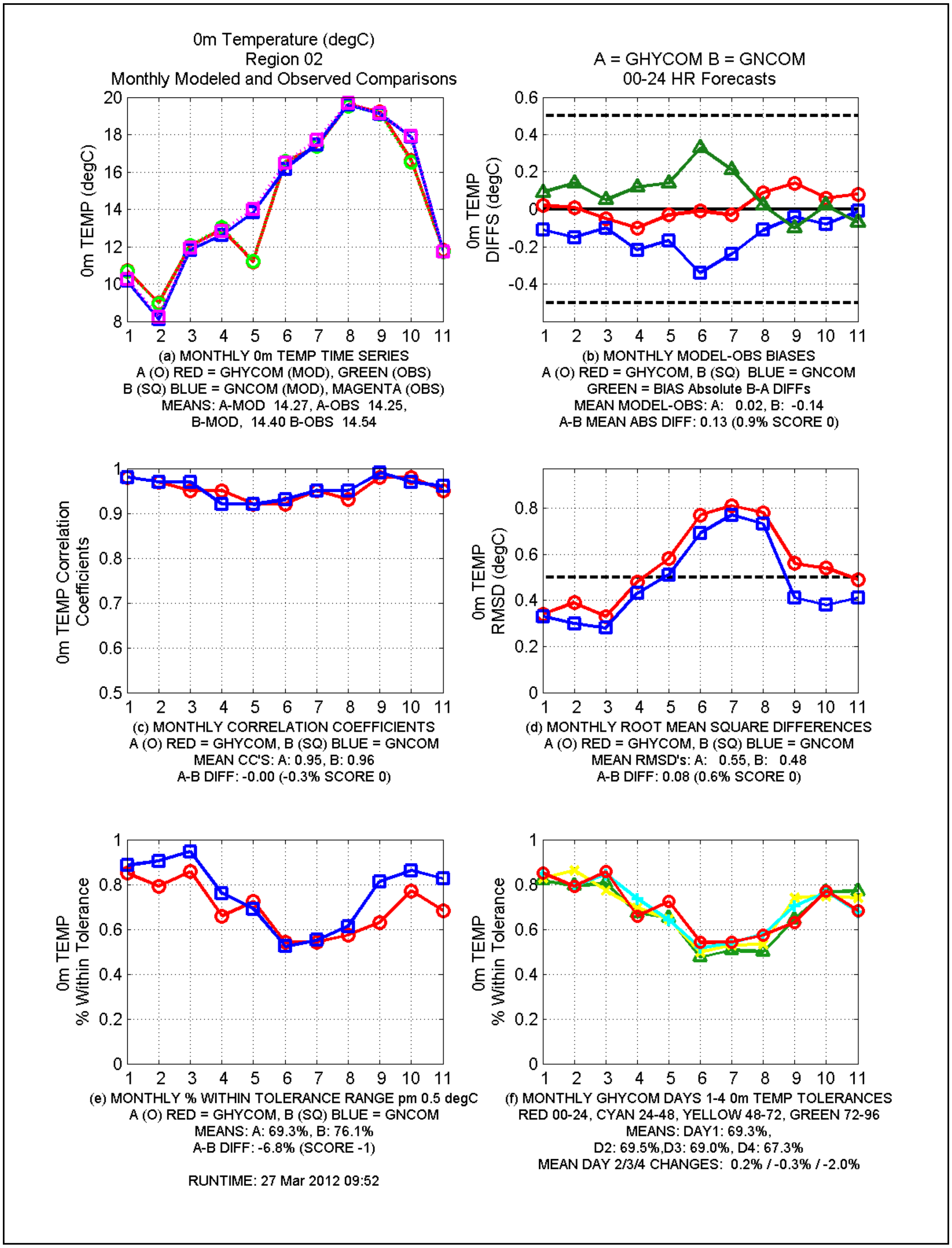
B.1.7.b. Sonic Layer Depths - Region 7 (Northeastern Pacific) without JANUARY



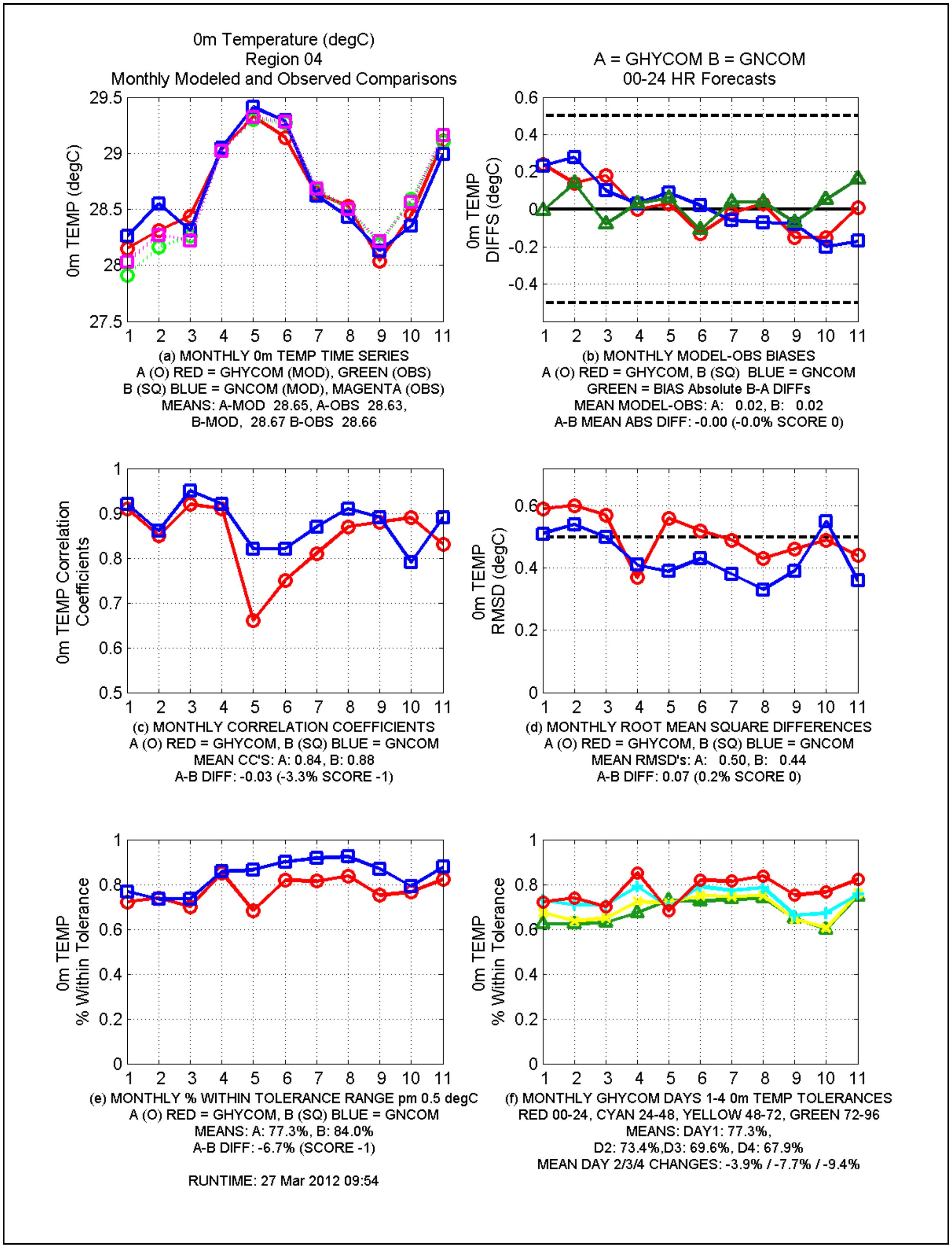
B.2.1. Temperature at 0m (surface) - Region 1 (Western Atlantic)



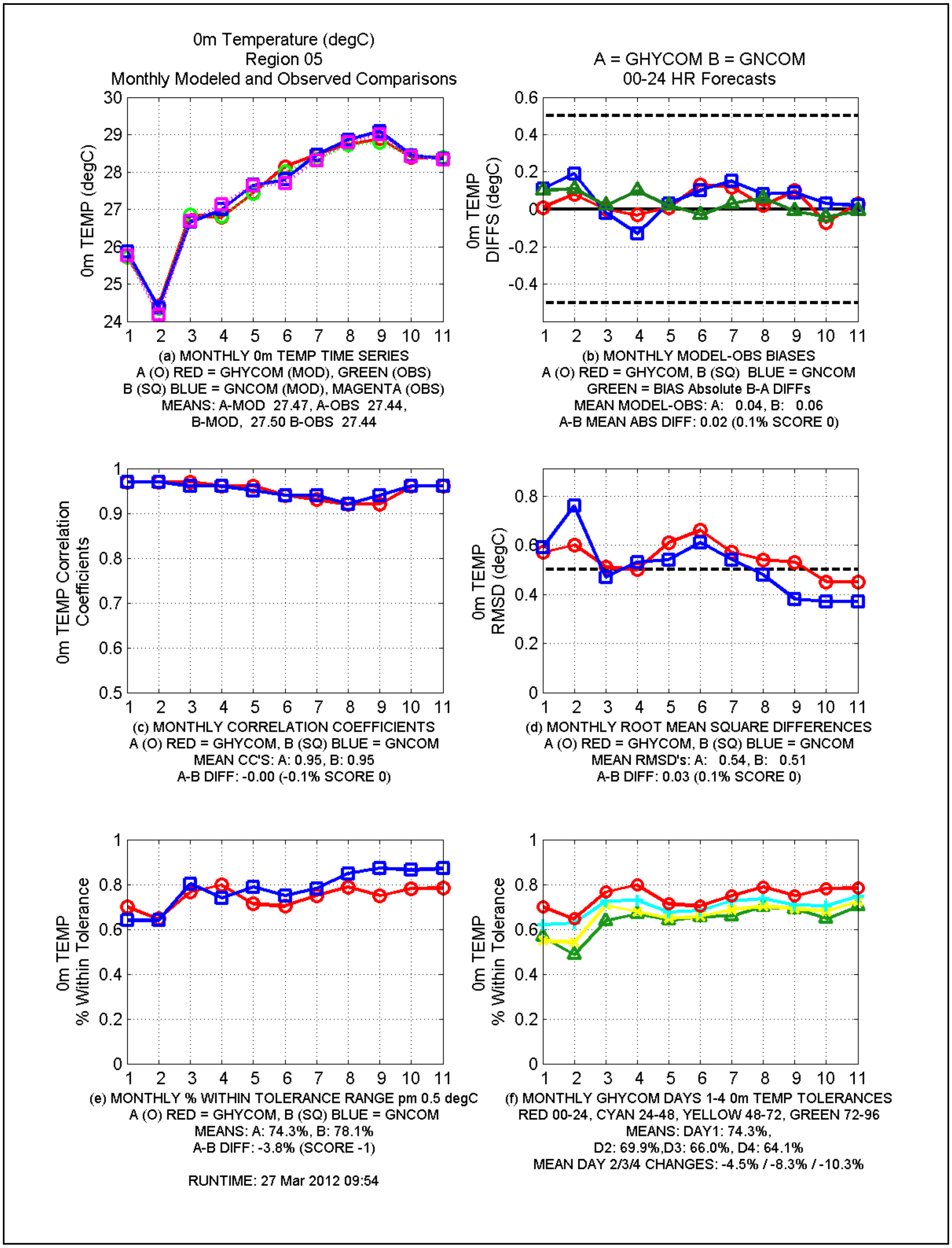
B.2.2. Temperature at 0m (surface) - Region 2 (Eastern Atlantic and Mediterranean)



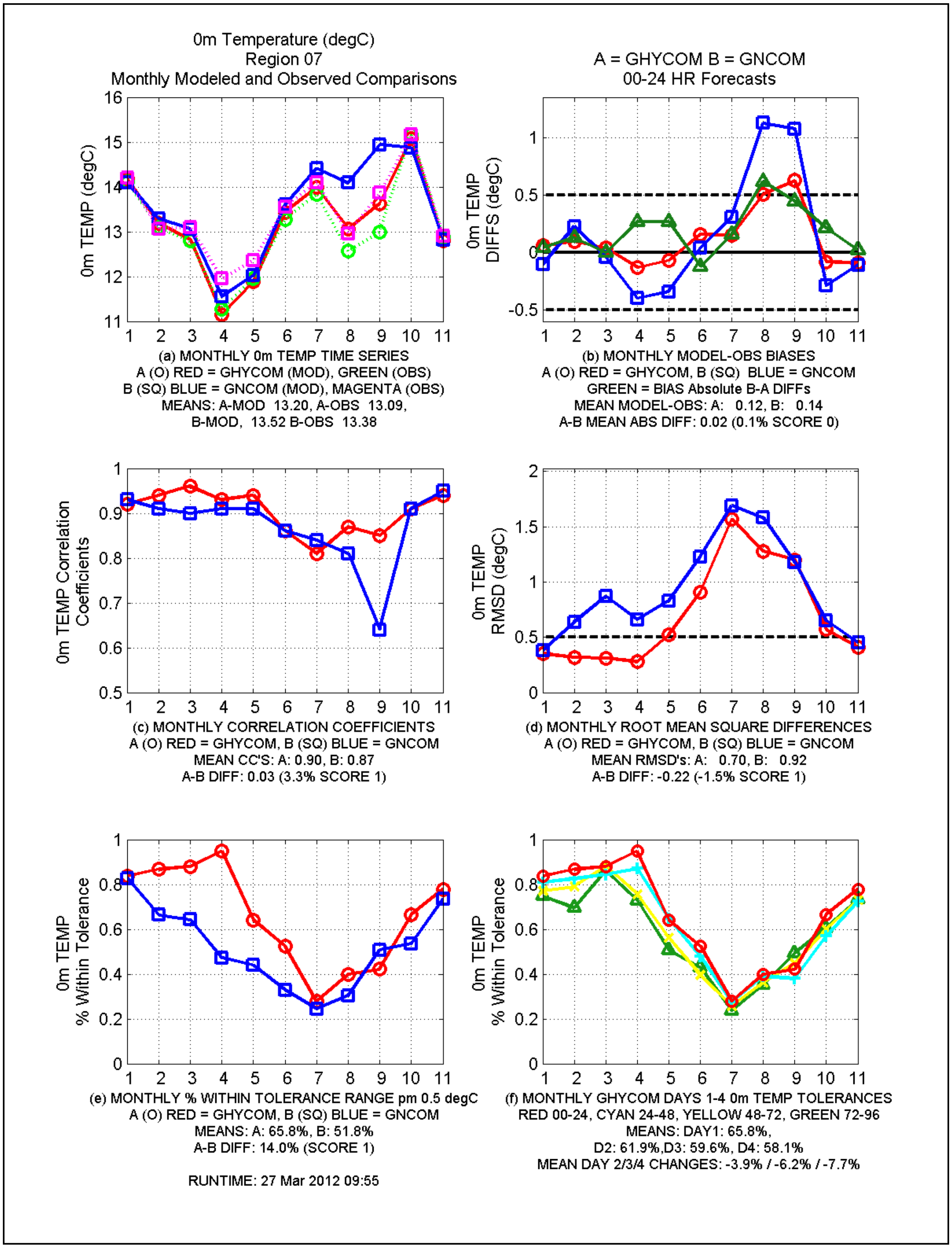
B.2.4. Temperature at 0m (surface) - Region 4 (Northwest Indian Ocean)



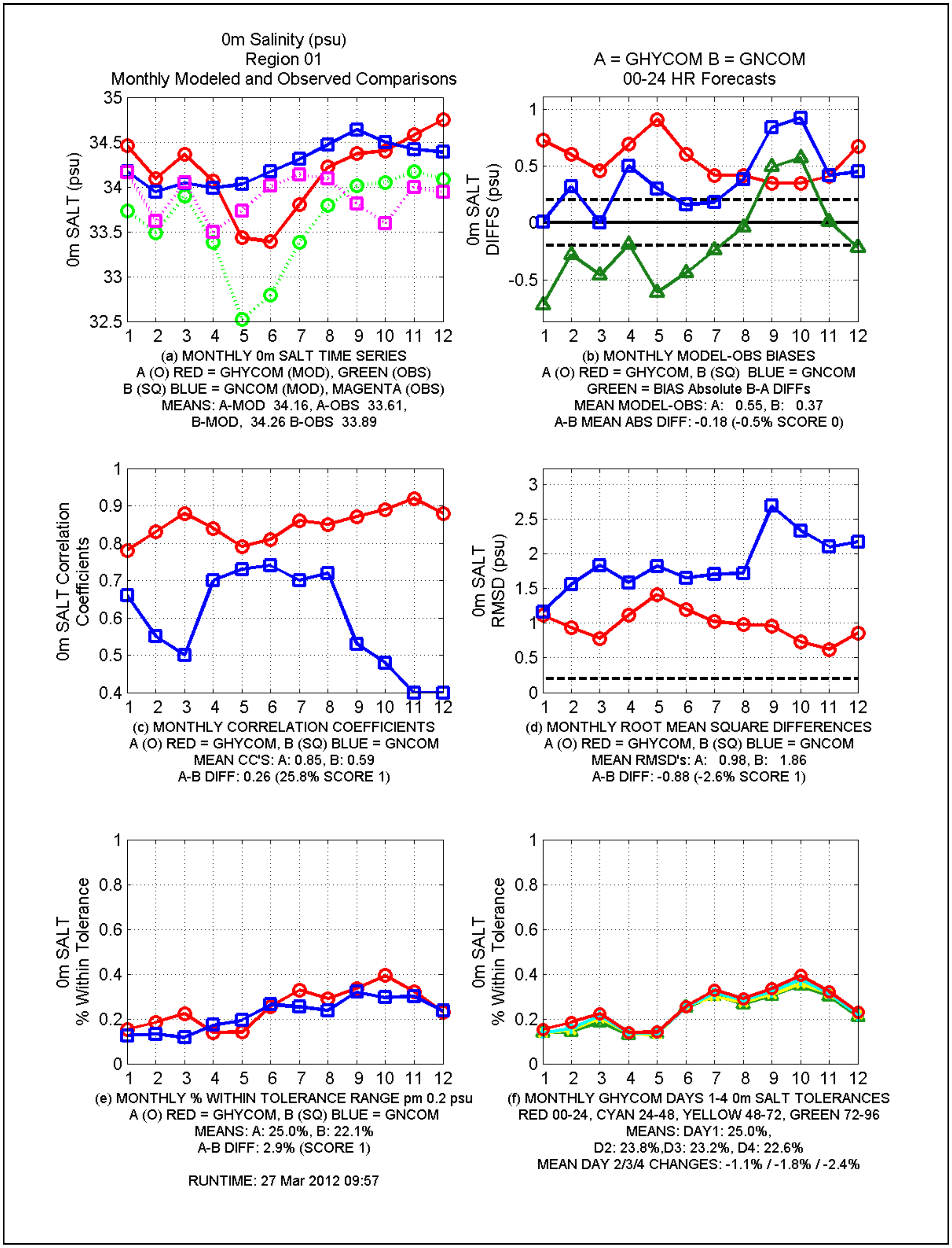
B.2.5. Temperature at 0m (surface) - Region 5 (Northwestern Pacific)



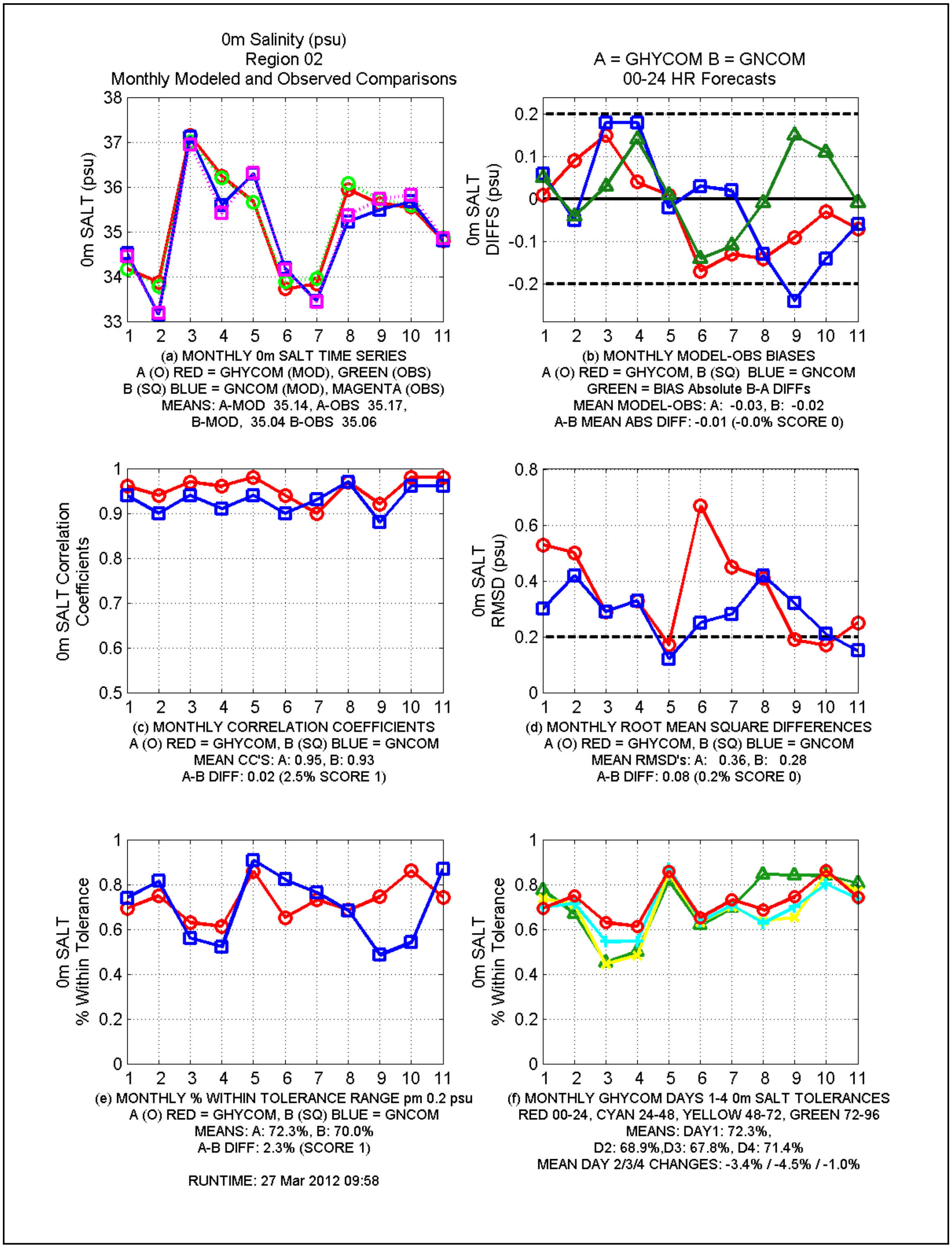
B.2.7. Temperature at 0m (surface) - Region 7 (Northeastern Pacific)



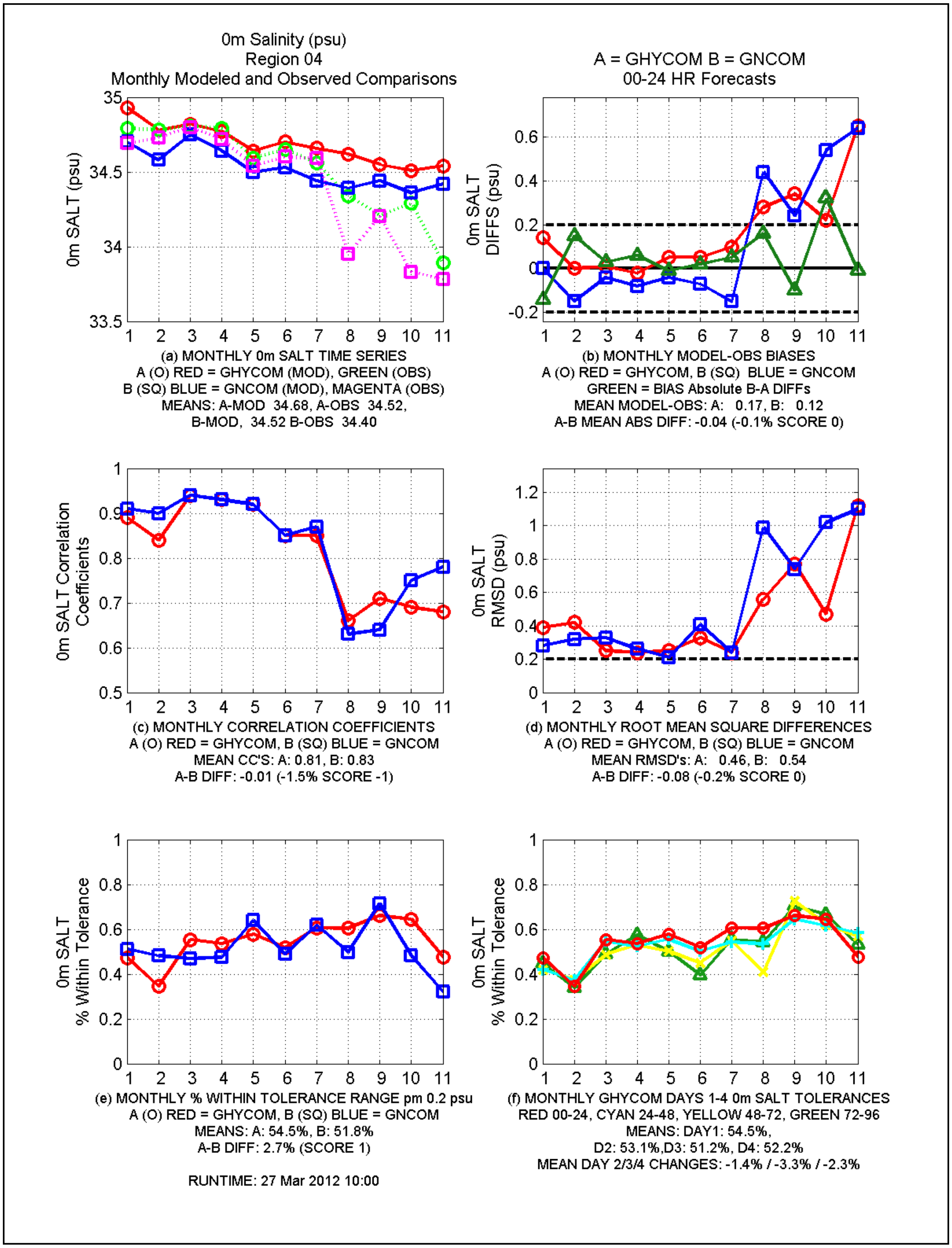
B.3.1. Salinity at 00m (surface) - Region 1 (Western Atlantic)



B.3.2. Salinity at 00m (surface) - Region 2 (Eastern Atlantic and Mediterranean)



B.3.4. Salinity at 00m (surface) - Region 4 (Northwest Indian Ocean)



(a) MONTHLY 0m SALT TIME SERIES
A (O) RED = GHYCOM (MOD), GREEN (OBS)
B (SQ) BLUE = GNCOM (MOD), MAGENTA (OBS)
MEANS: A-MOD 34.68, A-OBS 34.52,
B-MOD, 34.52 B-OBS 34.40

(b) MONTHLY MODEL-OBS BIASES
A (O) RED = GHYCOM, B (SQ) BLUE = GNCOM
GREEN = BIAS Absolute B-A DIFFs
MEAN MODEL-OBS: A: 0.17, B: 0.12
A-B MEAN ABS DIFF: -0.04 (-0.1% SCORE 0)

(c) MONTHLY CORRELATION COEFFICIENTS
A (O) RED = GHYCOM, B (SQ) BLUE = GNCOM
MEAN CC'S: A: 0.81, B: 0.83
A-B DIFF: -0.01 (-1.5% SCORE -1)

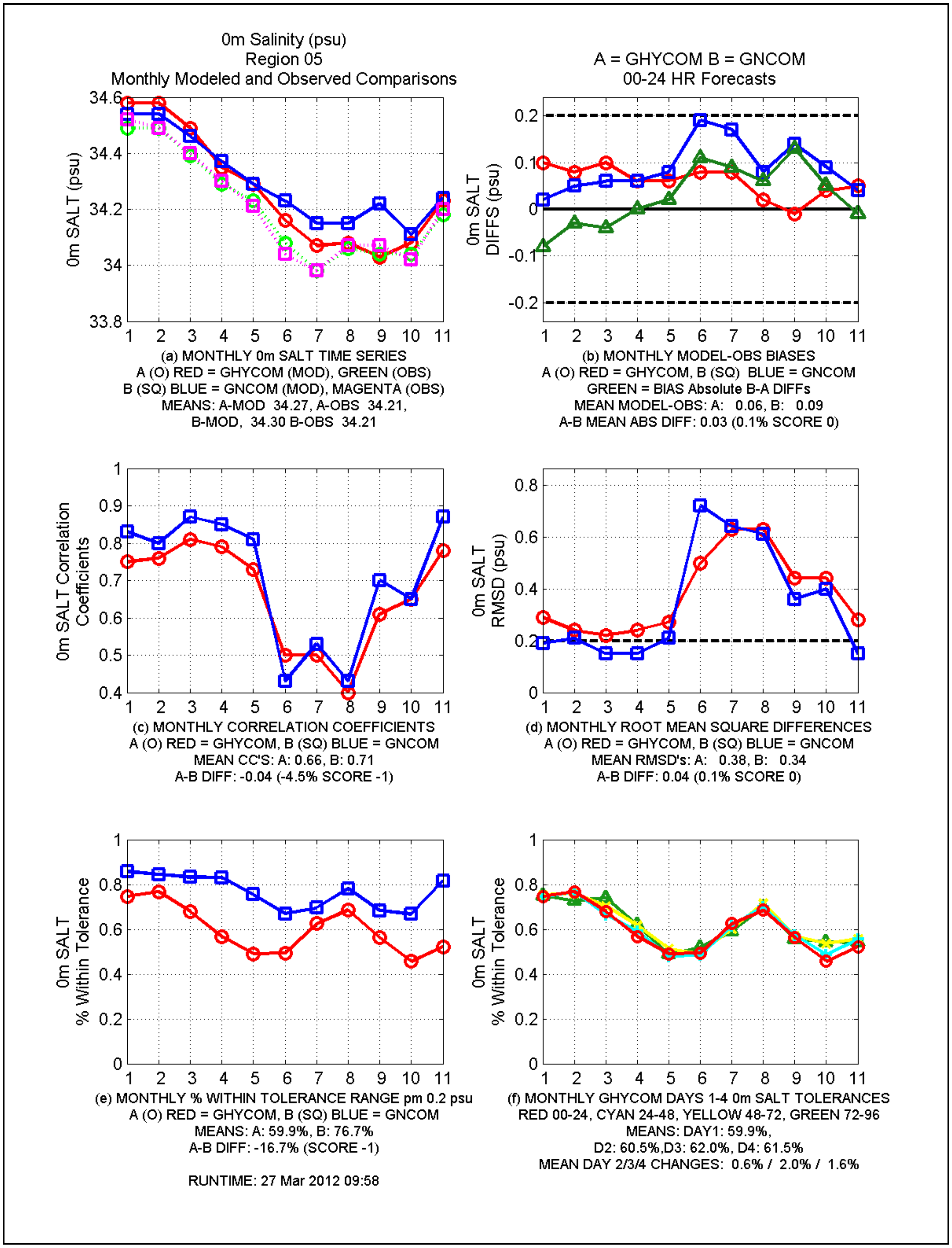
(d) MONTHLY ROOT MEAN SQUARE DIFFERENCES
A (O) RED = GHYCOM, B (SQ) BLUE = GNCOM
MEAN RMSD's: A: 0.46, B: 0.54
A-B DIFF: -0.08 (-0.2% SCORE 0)

(e) MONTHLY % WITHIN TOLERANCE RANGE pm 0.2 psu
A (O) RED = GHYCOM, B (SQ) BLUE = GNCOM
MEANS: A: 54.5%, B: 51.8%
A-B DIFF: 2.7% (SCORE 1)

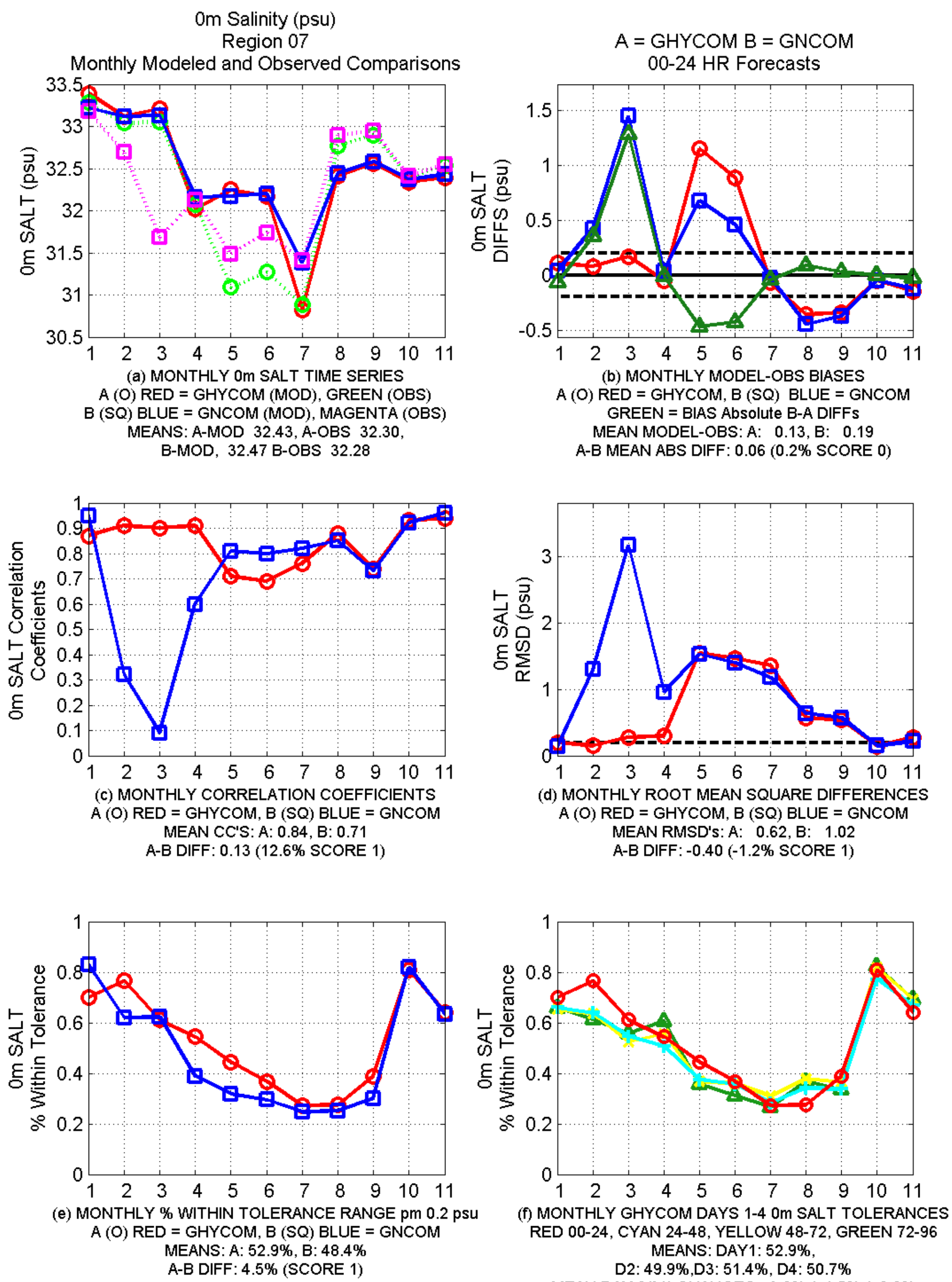
(f) MONTHLY GHYCOM DAYS 1-4 0m SALT TOLERANCES
RED 00-24, CYAN 24-48, YELLOW 48-72, GREEN 72-96
MEANS: DAY1: 54.5%,
D2: 53.1%,D3: 51.2%, D4: 52.2%
MEAN DAY 2/3/4 CHANGES: -1.4% / -3.3% / -2.3%

RUNTIME: 27 Mar 2012 10:00

B.3.5. Salinity at 00m (surface) - Region 5 (Northwestern Pacific)



B.3.7. Salinity at 00m (surface) - Region 7 (Northeastern Pacific)



(a) MONTHLY 0m SALT TIME SERIES
A (O) RED = GHYCOM (MOD), GREEN (OBS)
B (SQ) BLUE = GNCOM (MOD), MAGENTA (OBS)
MEANS: A-MOD 32.43, A-OBS 32.30,
B-MOD, 32.47 B-OBS 32.28

(b) MONTHLY MODEL-OBS BIASES
A (O) RED = GHYCOM, B (SQ) BLUE = GNCOM
GREEN = BIAS Absolute B-A DIFFs
MEAN MODEL-OBS: A: 0.13, B: 0.19
A-B MEAN ABS DIFF: 0.06 (0.2% SCORE 0)

(c) MONTHLY CORRELATION COEFFICIENTS
A (O) RED = GHYCOM, B (SQ) BLUE = GNCOM
MEAN CC'S: A: 0.84, B: 0.71
A-B DIFF: 0.13 (12.6% SCORE 1)

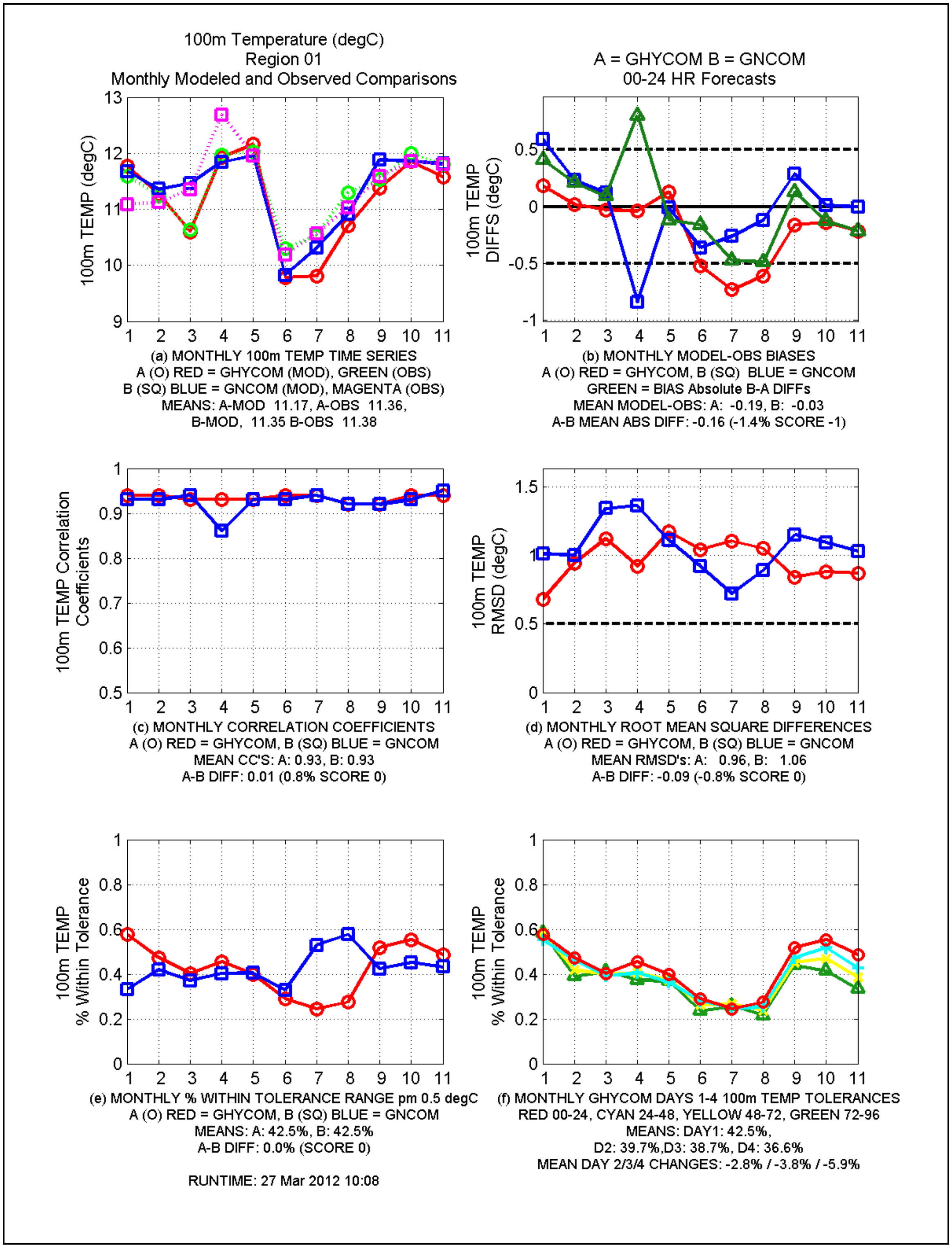
(d) MONTHLY ROOT MEAN SQUARE DIFFERENCES
A (O) RED = GHYCOM, B (SQ) BLUE = GNCOM
MEAN RMSD's: A: 0.62, B: 1.02
A-B DIFF: -0.40 (-1.2% SCORE 1)

(e) MONTHLY % WITHIN TOLERANCE RANGE pm 0.2 psu
A (O) RED = GHYCOM, B (SQ) BLUE = GNCOM
MEANS: A: 52.9%, B: 48.4%
A-B DIFF: 4.5% (SCORE 1)

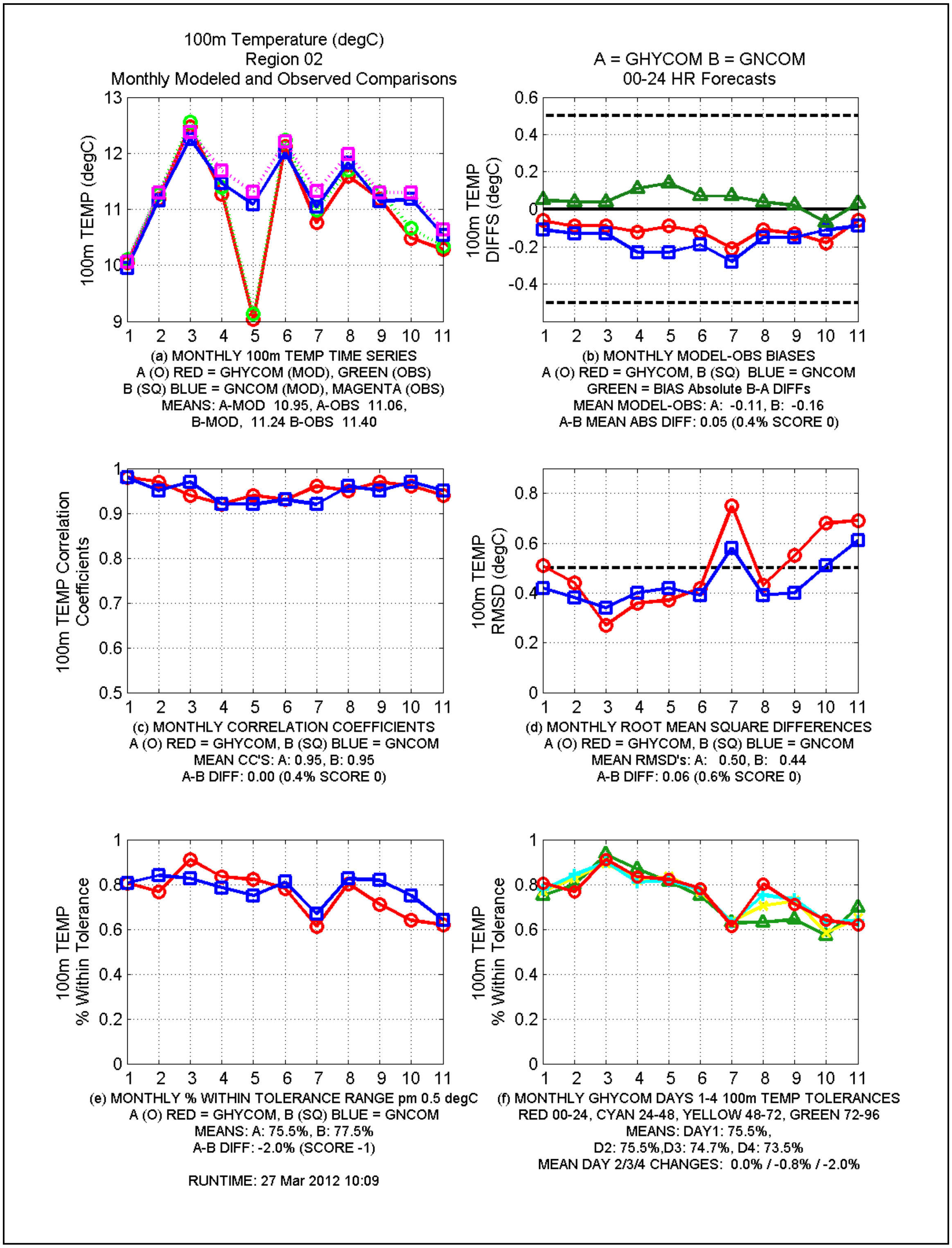
RUNTIME: 27 Mar 2012 09:59

(f) MONTHLY GHYCOM DAYS 1-4 0m SALT TOLERANCES
RED 00-24, CYAN 24-48, YELLOW 48-72, GREEN 72-96
MEANS: DAY1: 52.9%,
D2: 49.9%,D3: 51.4%, D4: 50.7%
MEAN DAY 2/3/4 CHANGES: -3.0% / -1.5% / -2.2%

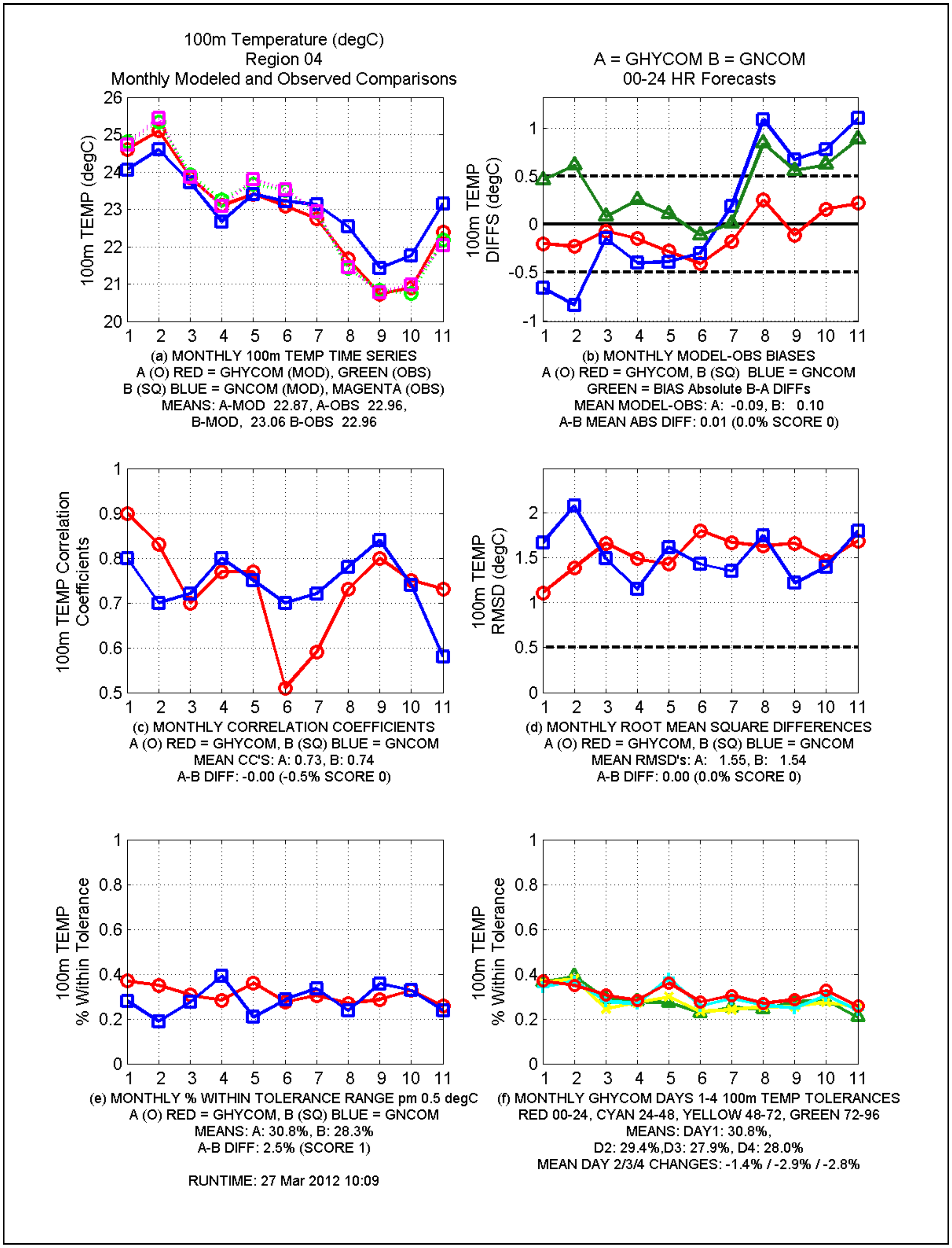
B.4.1. Temperature at 100m – Region 1 (Western Atlantic)



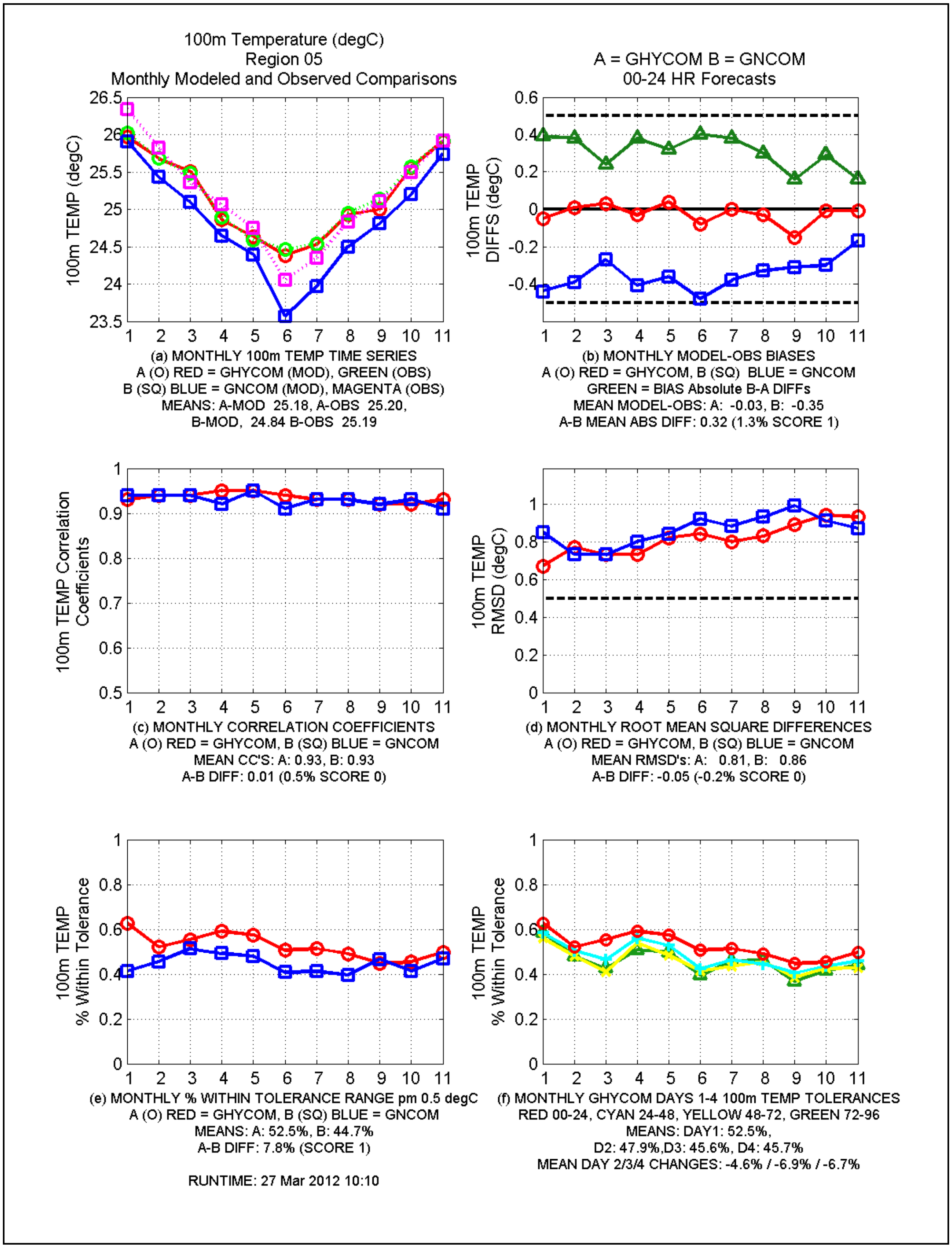
B.4.2. Temperature at 100m - Region 2 (Eastern Atlantic and Mediterranean)



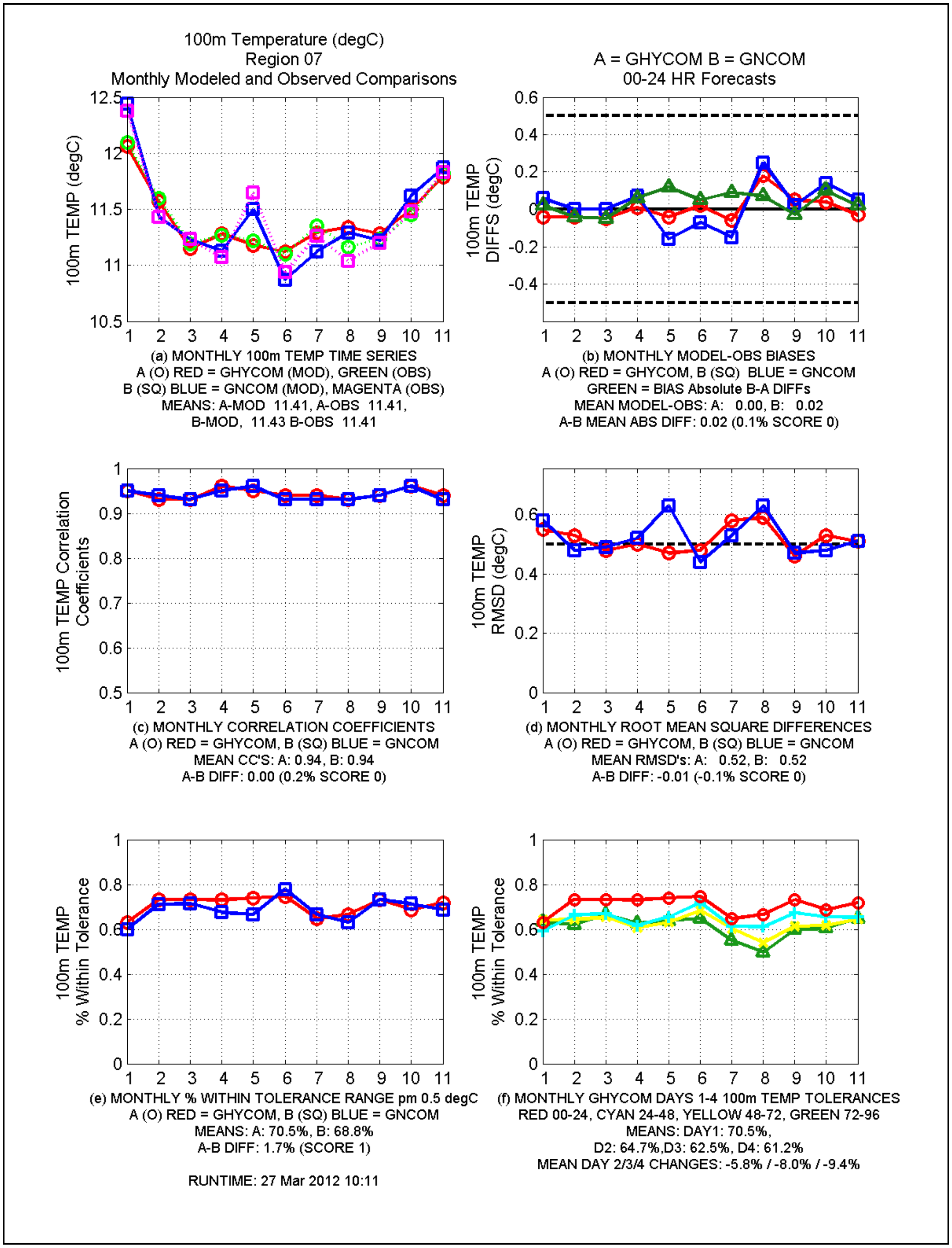
B.4.4. Temperature at 100m – Region 4 (Northwest Indian Ocean)



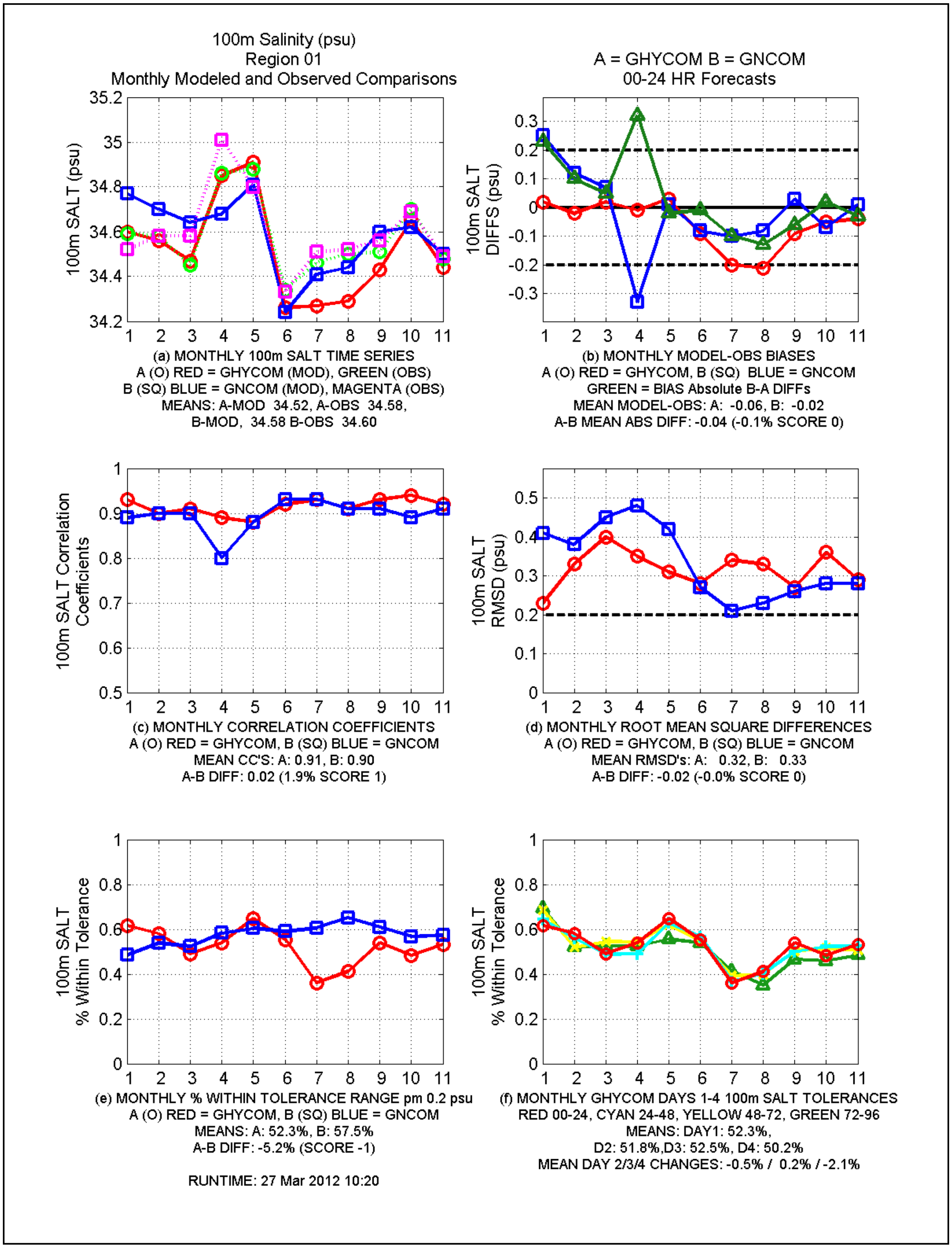
B.4.5. Temperature at 100m - Region 5 (Northwestern Pacific)



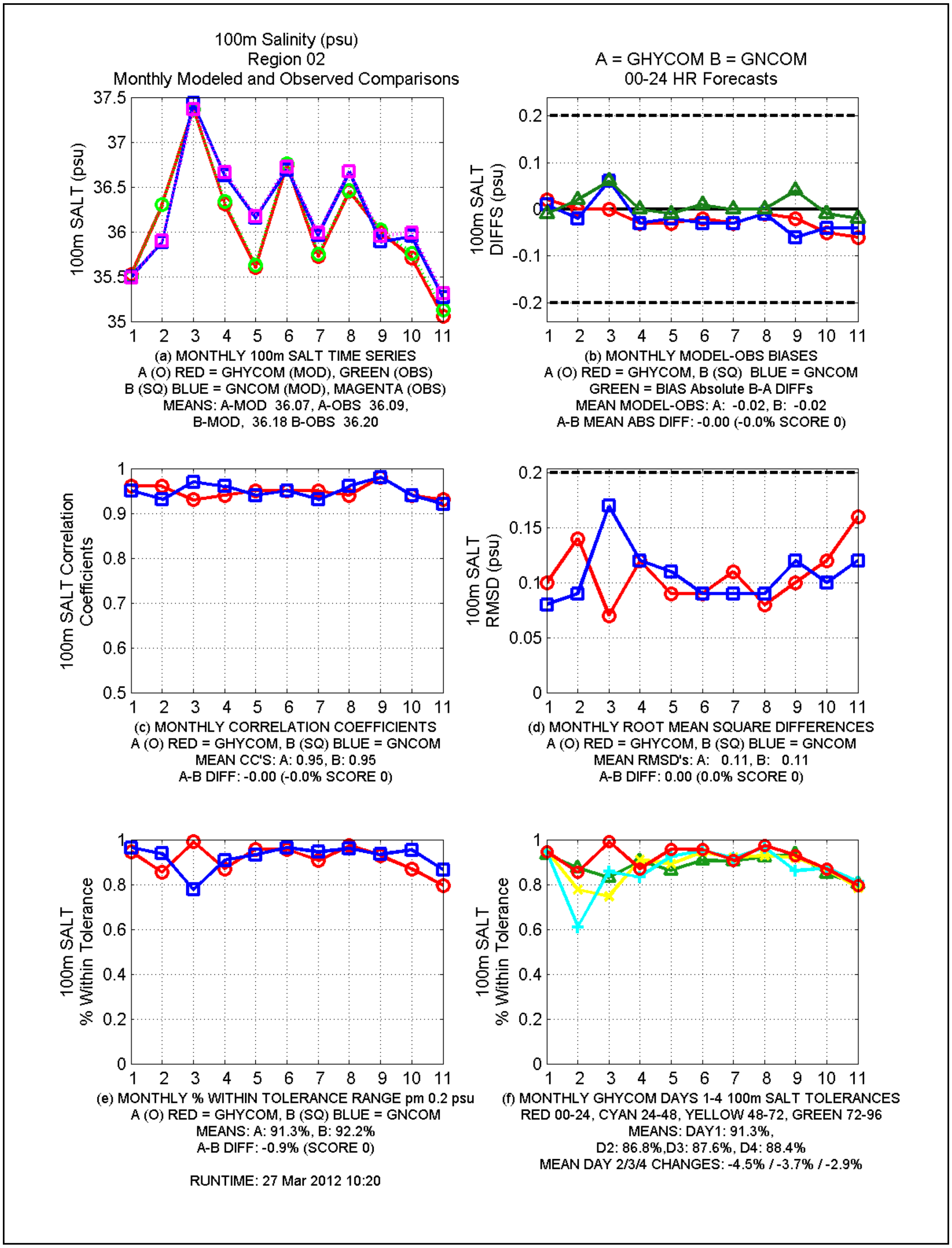
B.4.7. Temperature at 100m - Region 7 (Northeastern Pacific)



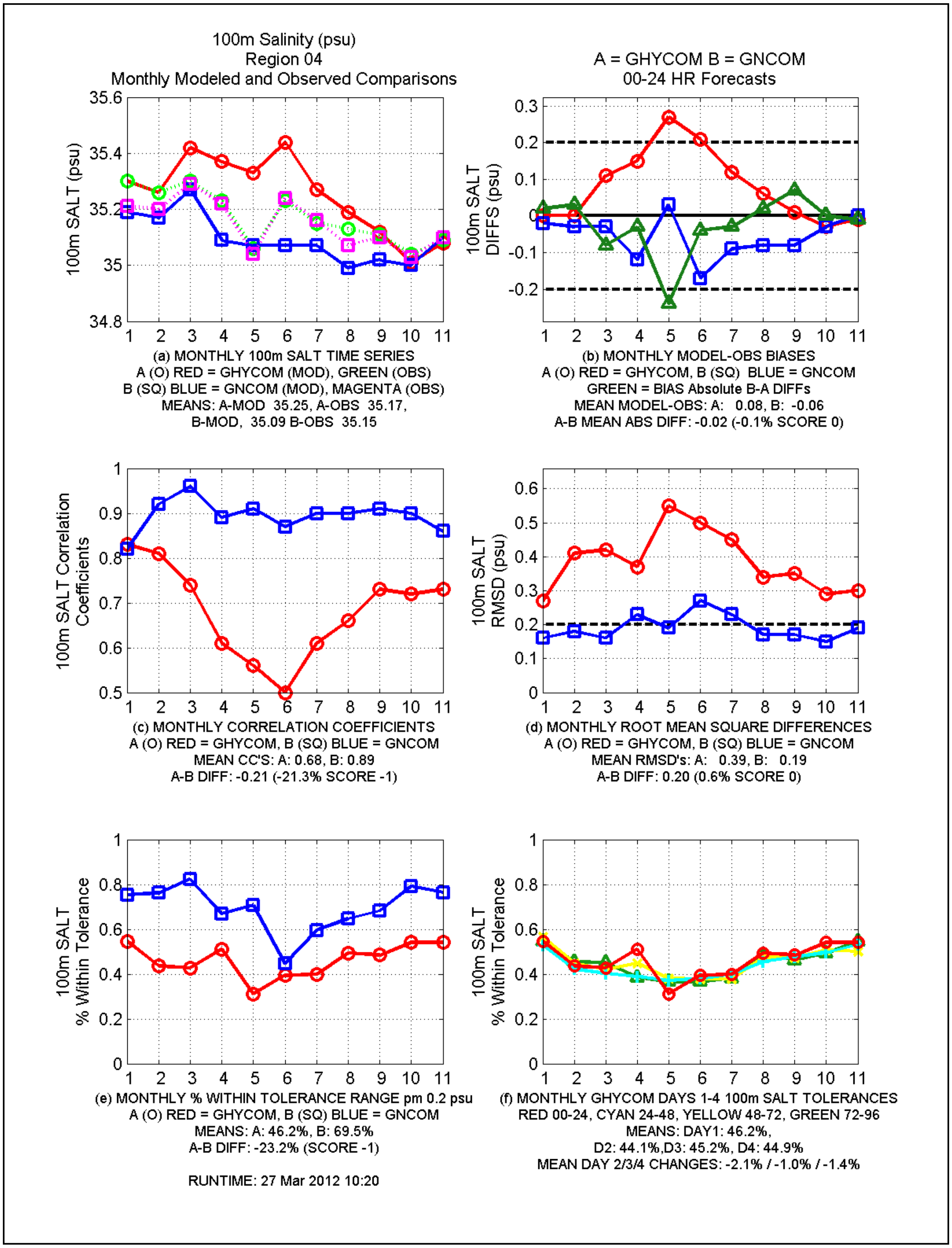
B.5.1. Salinity at 100m – Region 1 (Western Atlantic)



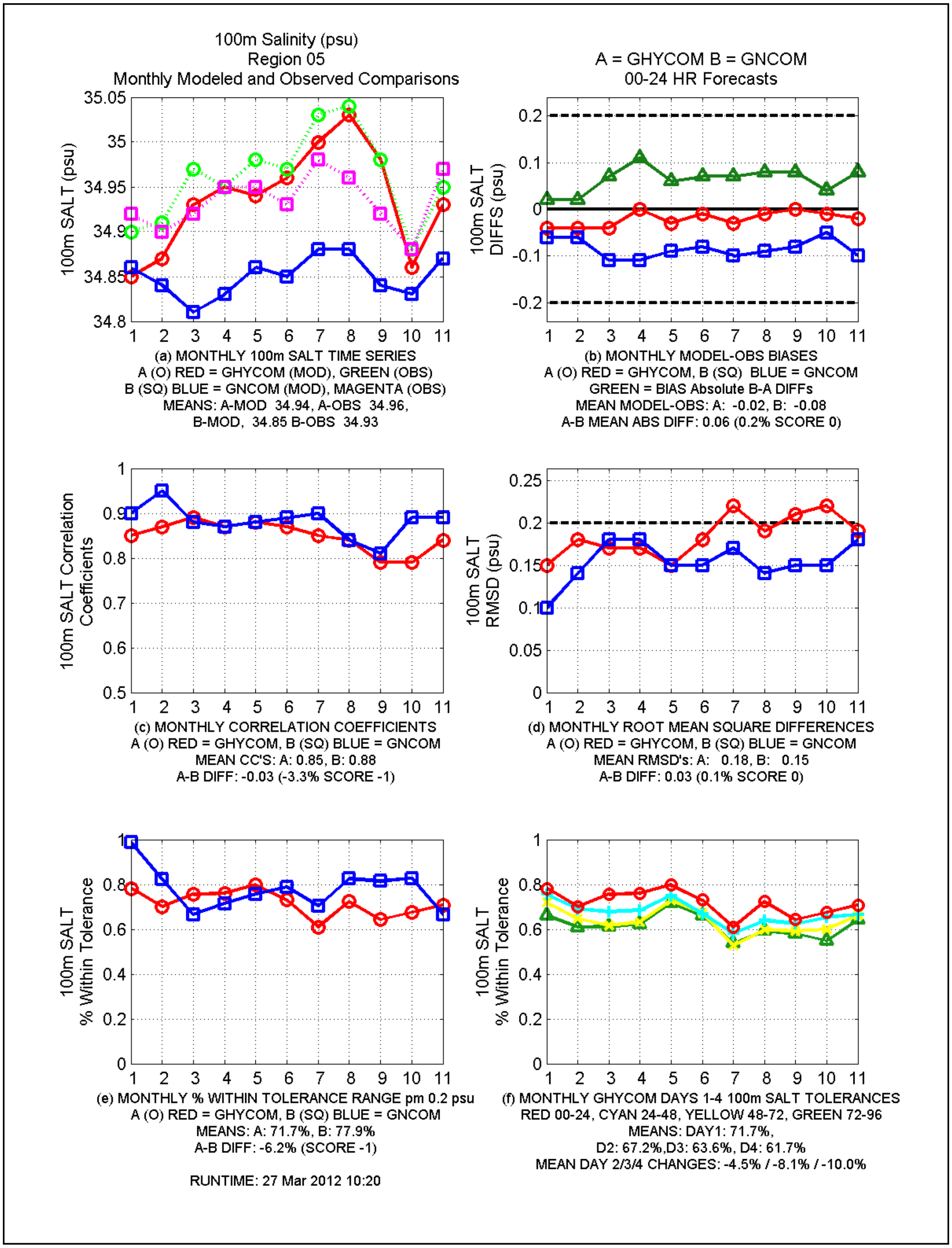
B.5.2. Salinity at 100m - Region 2 (Eastern Atlantic and Mediterranean)



B.5.4. Salinity at 100m - Region 4 (Northwest Indian Ocean)



B.5.5. Salinity at 100m - Region 5 (Northwestern Pacific)



B.5.7. Salinity at 100m - Region 7 (Northeastern Pacific)