Horizon Marine Inc. Numerical Ocean Model Nowcast and Forecast Skill Assessment in the Gulf of Mexico

Robert Leben and Kevin Corcoran
Colorado Center for Astrodynamics Research
University of Colorado, Boulder

Submitted to:

Steve Anderson
Horizon Marine, Inc.

September 1, 2005

Corresponding Author:

Robert R. Leben
431 UCB
Boulder, CO 80309-0431
Phone: 303 492-4113
Fax: 303 492-2825
Email: leben@colorado.edu

EXECUTIVE SUMMARY

Based an evaluation relative to the Loop Current (LC) Eddy Ulysses fronts depicted in the HMI EddyWatch™ reports, the nowcast/forecast skill of the HMI and HYCOM data assimilative models have been assessed for the time period from September 1 through December 15, 2004. The most useful statistical metric developed for this skill assessment was the RMS over all stations and experiments of the difference between the modeled and EddyWatch™ LC eddy frontal distance as a function of nowcast/forecast day. Smaller RMS values imply a better relative skill that includes both the random and systematic offsets between the modeled and observed fronts. A summary plot of these statistics is shown in Figure 1. The key points of this evaluation are:

- The LC Eddy Ulysses velocity fronts depicted by the HMI data assimilation model exceeded the skill exhibited by the HMI model temperature and sea surface height (SSH) fronts and the skill displayed by the HYCOM model nowcast/forecast LC fronts based on either velocity, temperature or SSH.
- The skill of the HYCOM LC eddy fronts based on temperature approached the skill of HMI LC eddy fronts based velocity, but did not exceed it.
- The relative ranking of the HMI model skill from highest to lowest is velocity, SSH and temperature tracking of the LC eddy front, respectively.
- The relative ranking of the HYCOM model skill from highest to lowest is temperature, SSH and velocity tracking of the LC eddy front, respectively.

We also made an assessment of the HMI and HYCOM models following the methodology of Kantha et al. (in press), which has been used for skill assessment of the University of Colorado version of the Princeton Ocean Model (CUPOM). While this methodology is imperfect, the results generally support the above conclusions. We report those results herein; however, there are significant difficulties associated with the metrics and statistics used. A critique of the methodology is included in the discussion of this report to address the utility of the statistical analysis reported in Kantha et al. (in press) and whether a similar approach should be used for future model nowcast/forecast skill assessments. Obviously site-specific statistics are important to the offshore industry, and should be made available; however, care must be exercised when reporting these statistics, so that they reasonably represent a realistic measure of skill in the models under consideration.

For example, the conclusion from the CUPOM analysis by Kantha et al. (in press) that "on the average, the model reproduces the LC from to about 11-12 km in the hindcast and nowcast modes, and 19 km in the forecast mode" results more from a judicious choice of the averaging time period than actual model skill and ignores significant problems with how the offset is calculated and the large standard deviations about the mean. In fact, the CUPOM nowcasts and 4-week forecasts of the LC thermal front are expected to be *within 20 km of the mean relative offset* to the EddyWatch™ LC fronts only about 28% and 18% of the time, respectively, based on the limited statistics presented in Kantha et al., (in press), assuming a normal distribution of the model and EddyWatch™ differences. Using similar assumptions, the HMI velocity front nowcasts and 4-week forecasts are expected to be within 20 km of the mean relative offset to the EddyWatch™ fronts 46% and 30% of the time, respectively.

We caution that making any determination of the relative skill between these two models may not be appropriate because the statistics are based on different time periods and LC configurations. It is apparent, however, that the interpretation of the statistics derived from a skill assessment study is just as important as the statistics employed. We believe that the RMS statistics presented in this report are a good step in that direction.

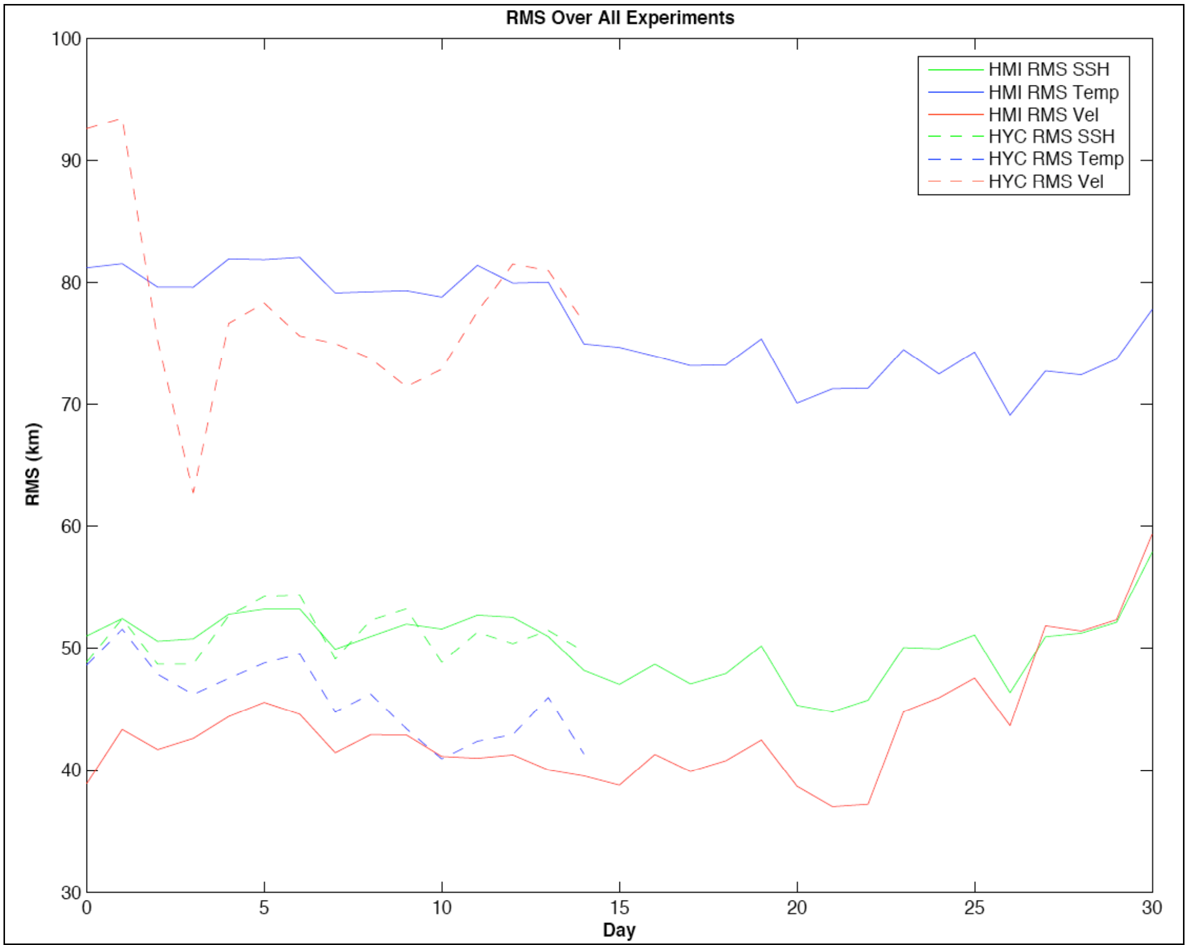

Figure 1. Comparison of rms over all experiments of the rms over all stations of the difference between EddyWatch™ fronts and HMI Model and HYCOM fronts as a function of the nowcast/forecast time period and frontal tracking technique.

1. INTRODUCTION

The Horizon Marine Inc. (HMI) data assimilative numerical ocean model is evaluated to objectively determine the skill of nowcasts and forecasts of Loop Current Eddy frontal boundaries in the Gulf of Mexico. EddyWatch™ reports from HMI are used as the observational "truth" for the frontal locations. These reports show the surface Loop Current/Loop Current Eddy front as determined from the available satellite tracked drifting buoy locations and GOES SST imagery. Loop Current fronts are objectively determined from each model nowcast/forecast field by tracking contours of sea surface height (SSH), and the isopleths of velocity magnitude at 50 meters and isotherms of temperature at 200 meters. Statistical differences between the observational "truth" and the model simulation give an objective measure of the model's nowcast/forecast skill. We will compare the HMI model results to similar results obtained from the HYCOM model.

A brief overview of the two models used in this study follows:

HMI Model

Horizon Marine Inc. is an operational oceanography service company that produces ocean nowcasts and forecasts using the Accurate Environmental Forecasting, Inc. Eddy Forecast System (EFS). The EFS is a forecast system for the Gulf of Mexico (GOM) that predicts the movements and locations of Loop Current eddy events. The EFS is broken down into two components: the initialization component and the prediction component. The initialization component requires 3-dimensional fields of temperature, salinity, and velocity to initialize the GOM eddy field. Based on this data the prediction component of the EFS numerically solves for the partial differential equations that model ocean dynamics. This predicts the movement and progression of the 3-dimensional initial fields of temperature, salinity, and velocity through time. The initialization component of the model requires some manual intervention. Experienced forecasters determine the approximate location of the LC eddy and then calculate appropriate density and velocity fields. These fields are used in the prediction component of the forecast system. The prediction component can be run using two different models. One model is a variant of the Princeton Ocean Model (POM), and the other is a multiprocessor version of the MIT model. Both models use 4 km horizontal resolution and 50 z-levels. The model that is used depends on the eddy configuration and computer time available.

HYCOM Model

The HYCOM consortium is a multi-institutional effort funded by the National Ocean Partnership Program (NOPP) to develop and evaluate a data-assimilative hybrid isopycnal-sigma-pressure (generalized) coordinate ocean model (called HYbrid Coordinate Ocean Model or HYCOM). HYCOM development is a collaborative effort between the University of Miami, the Naval Research Laboratory (NRL) and the Los Alamos National Laboratory (LANL). HMI is a member of the HYCOM consortium.

The model referred to in this study as the HYCOM model is a regional subset of the North Atlantic data from the Atlantic HYCOM system, which has a horizontal resolution of 1/12°. The system assimilates daily Modular Ocean Data Assimilation (MODAS) SSH analyses of the available satellite altimeter observations. The Cooper and Haines (1996) technique is used to project the surface information into the interior of the ocean.

A relaxation to the MODAS SST analysis is also included. Each week a 14-day forecast is performed using the Atlantic HYCOM system.

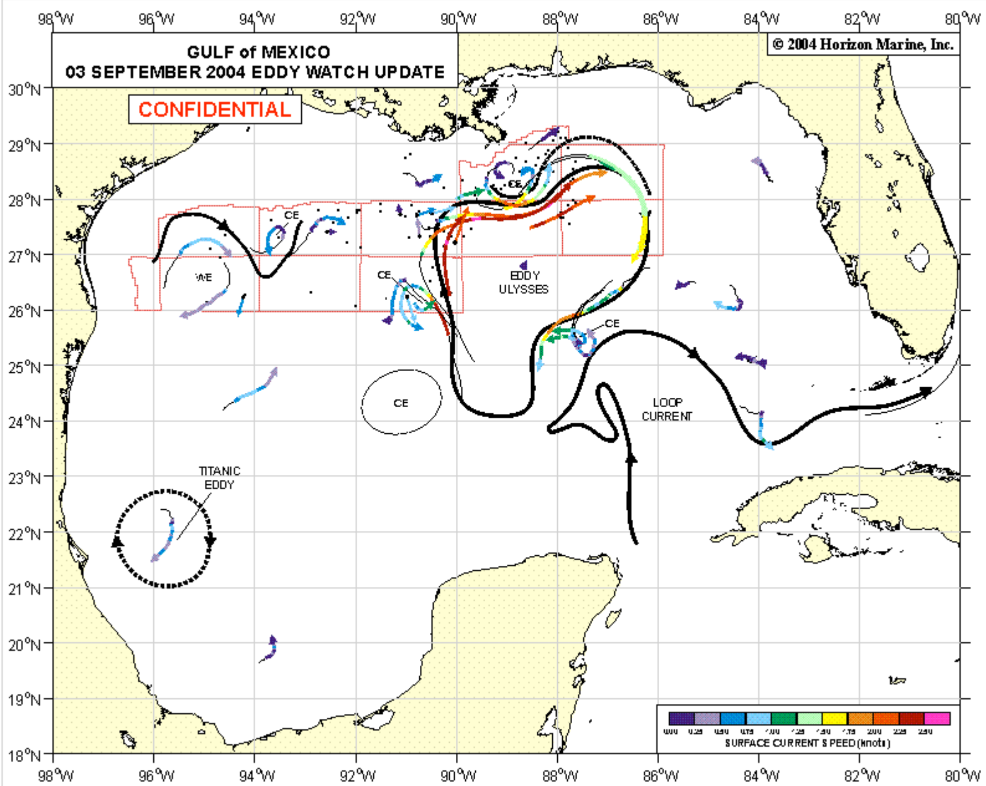

Figure 2. Sample EddyWatch™ Report for September 3, 2004. Digitized version is shown in Figure 3.

2. DATA AND METHODS

The study time period spans the 3.5-month time period from September 3 through December 15, 2004. During this time period a large anticyclonic eddy, "Eddy Ulysses", separated from the Loop Current. The study data archive includes HMI EddyWatch™ reports, and HMI Model and HYCOM Model nowcast/forecast fields from within the study time period. Data processing and analysis methods were developed in MATLAB® to facilitate data visualization, statistical analysis, and model skill assessment. Using MATLAB® the data from the two models was slightly modified and cropped to facilitate analysis. The extent of this is descried in HMI Model Data Archive and the HYCOM Data Archive.

HMI EddyWatch™ Archive

A total of 62 EddyWatch™ Reports were archived. The report archive includes daily reports from September 3 through October 15 (with the exception of September 6th, which was Labor Day), and the October 17, 20, 21, 25, 28, November 1, 4, 8, 11, 15, 18,

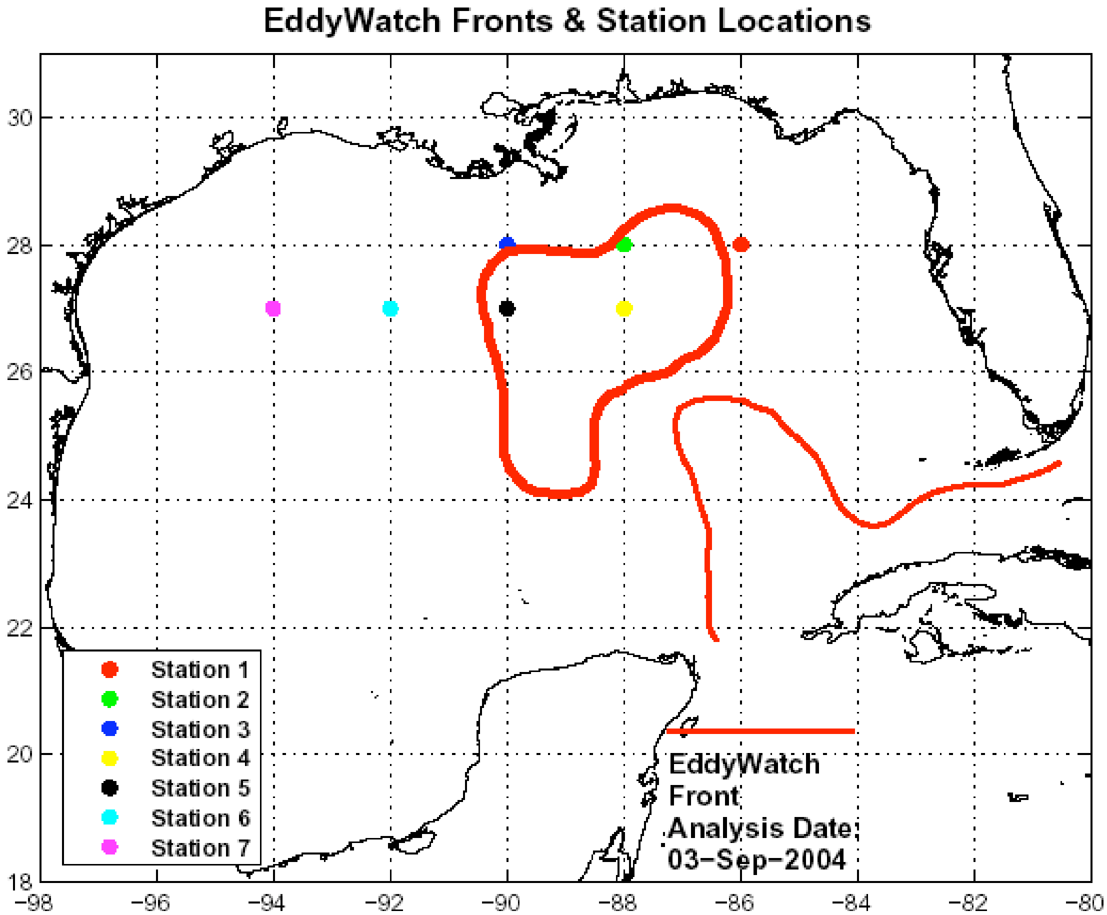

Figure 3. Sample plot of digitized fronts from EddyWatch™ Report for September 3, 2004 that is shown in Figure 1.

22, 23, 24, 29, December 2, 6, 9, 13, 16 reports. All of the EddyWatch™ Reports used in this study are included in Appendix A. A sample report from September 3rd is shown in Figure 2. The hand-digitized version of the Loop Current and Eddy Ulysses fronts from that report are shown in Figure 3.

HMI Model Data Archive

A total of 10 HMI nowcast/forecasts were available for model skill assessment during the study time period from September 1 through December 15, 2004. The time lines for each of the nowcast/forecast experiments are shown in Figure 4. The model data as it was delivered to CCAR is shown in blue with the spin up date shown in green and the issued date shown in red. For analysis purposes the time series shown in black were created using linear interpolation to fill skipped days in the original model nowcast/forecast archives. Additional changes were made to experiments two, three, five, and ten. They had missing data that was dealt with in the following ways. Experiment two was shorter than the rest of the experiments so eight additional days were added that were duplicates of the final good day in the HMI model. For experiments three and five there were no missing days, however, the data for the final day was blank. To remedy this the final model day was made a duplicate of the final model day with good data. After linear interpolation of the every-other-day model data to a once-per-day record, the final three days of both experiments are identical. A similar problem was found with experiment ten, however, the final three days of the distributed HMI model had no data, and so after linear interpolation of the data gaps there are seven identical days in the interpolated record.

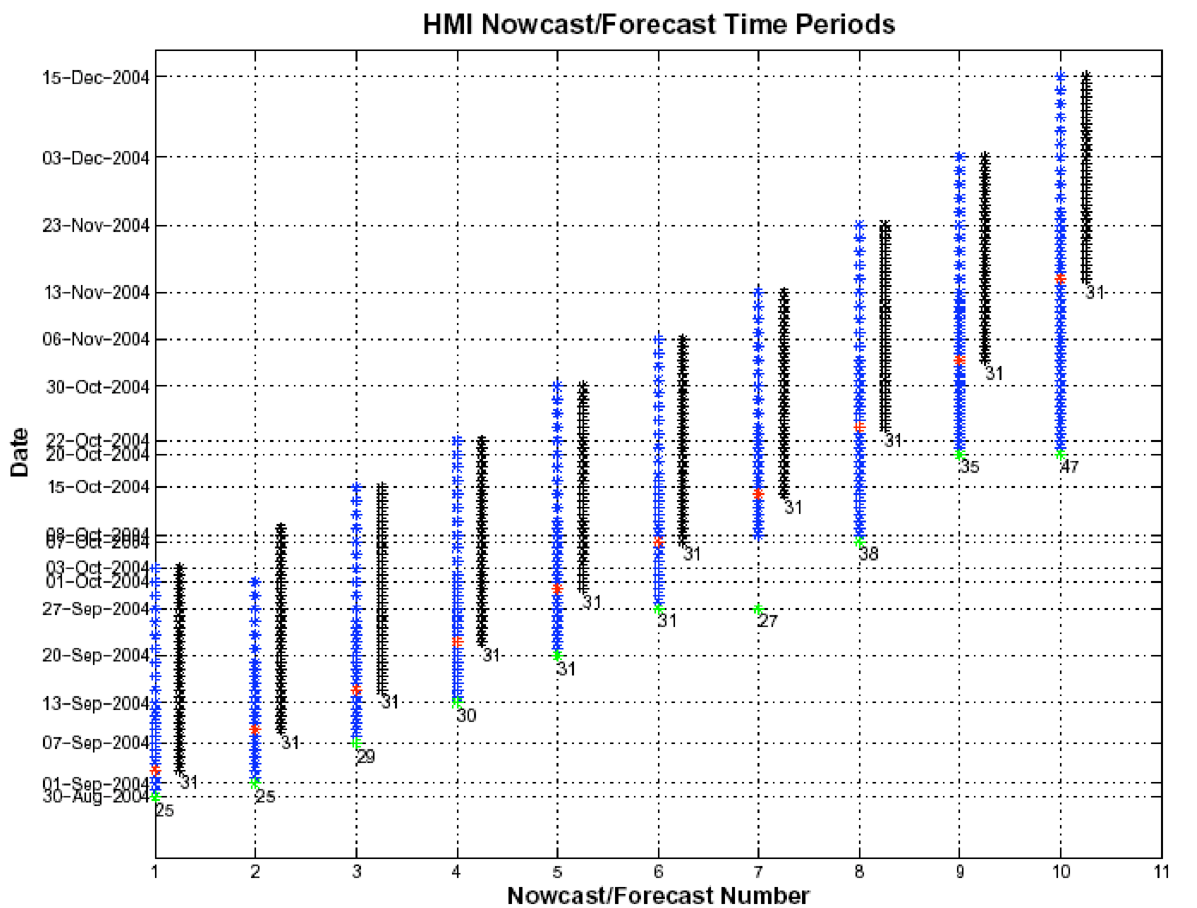

Figure 4. Time spans of HMI Model nowcast/forecast experiments available during study time period are plotted above. Model data is shown in blue with the spin up data shown in green, and the issued date shown in red. The linearly interpolated data set that was analyzed is shown in black.

HYCOM Data Archive

A total of 9 HYCOM 14 day forecasts were available for model skill assessment during the study time period. A nowcast day was not supplied for each of the nine experiments

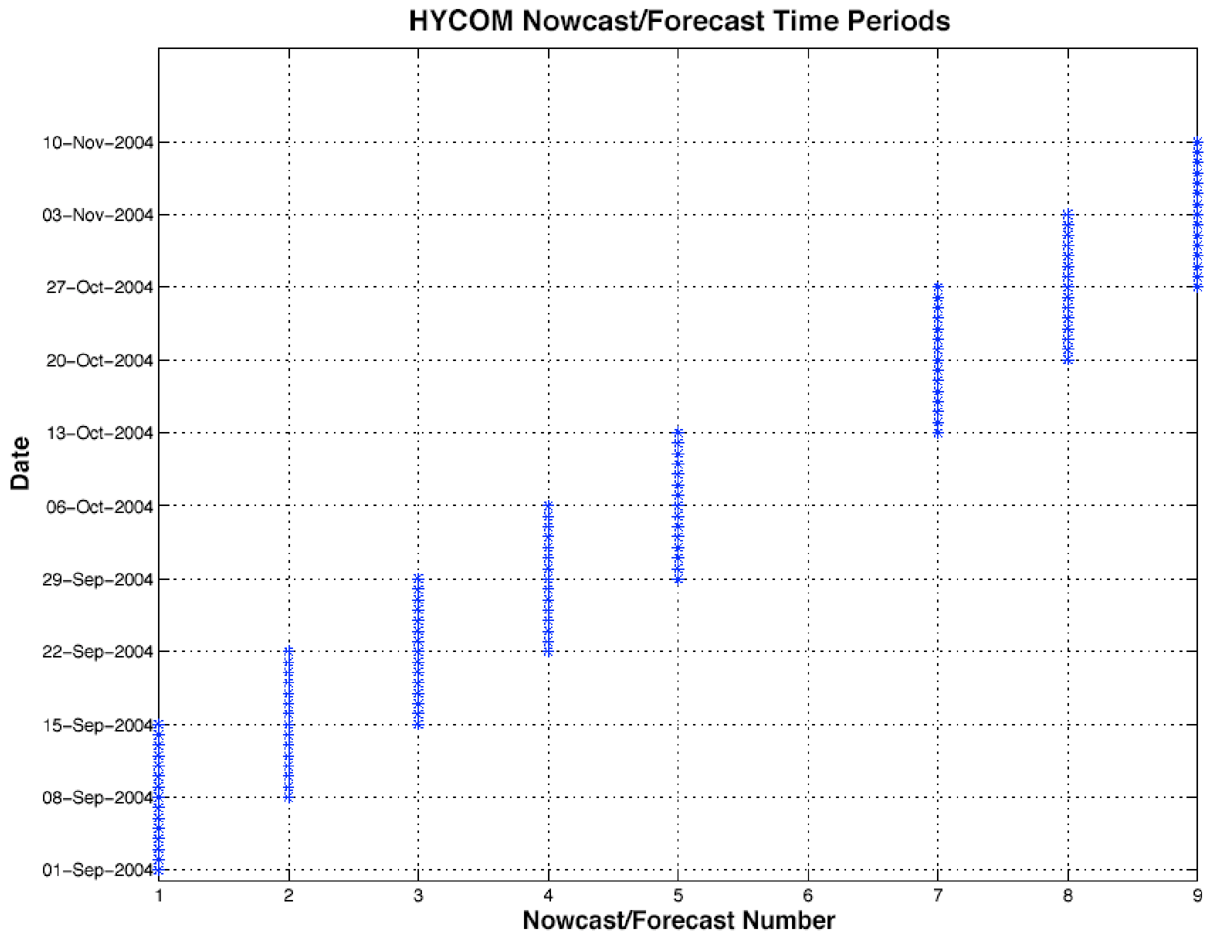

Figure 5. Time spans of HYCOM nowcast/forecast experiments available during study time period are plotted above.

so we created one by duplicating the first day of the forecast record. This simplified the analysis process. A graphical representation of the nowcast/forecast time periods is shown in Figure 5. The daily fields of SSH, sea surface temperature (SST) at 200 meters, and velocity at 50 meters were provided by the University of Miami HYCOM consortium database managers. One nowcast/forecast was partially missing from the HYCOM archive (experiment #6: October 6th to October 20th) and was not used in the study. In addition forecast 7 was missing the first day of its forecast so two duplicate days were added to the nowcast/forecast.

Model Skill Assessment Methodology

The skill assessment methodology applied in this study follows the techniques and methods developed to evaluate the University of Colorado Princeton Ocean Model (CUPOM) data assimilation system (Kantha and Choi, 2003; Kantha et al., submitted). The original tools developed for that work were not available to this study, so we have developed our own comprehensive skill assessment toolbox that includes modular databases and analysis tools that can be routinely applied to model skill assessment in the Gulf of Mexico.

The methodology is as follows:

1. Digitize the Loop Current and eddy fronts from the available Horizon Marine EddyWatch™ reports.
2. Define Loop Current and eddy fronts from the model nowcast/forecast fields using the following criteria:
    a. 15°C isotherm at 200m depth
    b. 17-cm SSH contour
    c. 75 cm/sec isopleth of velocity magnitude

    Note: Different tracking values may be required because of thermal biases and/or differing background means in the model nowcast/forecast fields.
3. Mask extraneous isotherms, contours and isopleths not associated with the Loop Current or Loop Current eddy front.
4. Calculate distances from station locations given in Table 1 to the Loop Current frontal boundaries defined by each EddyWatch™ front and all nowcast/forecast SSH, velocity and temperature fronts.

TABLE 1: Location of Frontal Analysis Reference Stations

Station	Latitude	Longitude
1	28°	86°
2	28°	88°
3	28°	90°
4	27°	88°
5	27°	90°
6	27°	92°
7	27°	94°

5. Calculate the magnitude difference between the EddyWatch™ and model frontal distances.
6. Analyze the time series and statistics of the frontal distance differences.

The data analysis tasks cannot be fully automated. Manual intervention is required in several of the steps; however, we have outlined protocols for the analyst to insure consistent processing of the data sets.

During a preliminary analysis it was determined that the standard tracking values used for the SSH, temperature and velocity would not be as accurate as they could be if we were to select new contours to track. For the HMI Model it was determined through visual analysis that 30 cm was more suitable for the SSH field, the 21°C contour was chosen for the isotherm at 200m depth, and 75 cm/sec was selected for the velocity magnitude at 50m depth. Similarly the contours selected for the HYCOM Model are as follows: 30cm, 13°C, and 65 cm/sec.

While evaluating the HMI Model and the HYCOM Model no comparisons to the Loop Current front were made, only the Loop Current Eddy Ulysses front was used. The Loop Current front was added only as reference in the plots and was not included in the distance analysis.

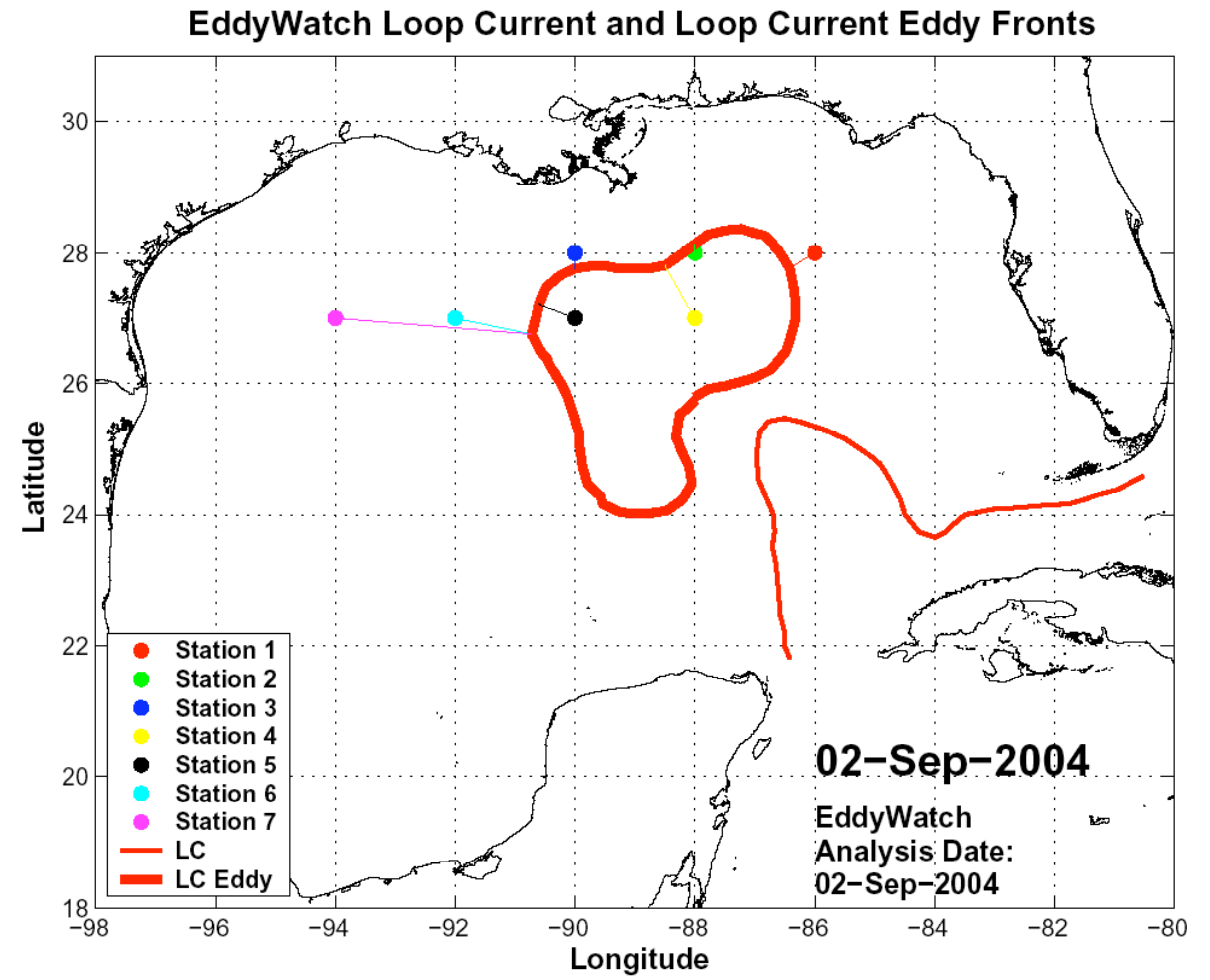

Figure 6. A sample chart used to check distance-to-front determined from the digitized EddyWatch™ report fronts. The original EddyWatch™ chart and the digitized version are shown in Figures 2&3, respectively.

3. RESULTS

Results will be summarized based on the specific tasks outlined in the work scope for this project.

EddyWatch™ LC/LCE Frontal Tracking

Task 1: Digitize the Loop Current and eddy fronts from the weekly Horizon Marine EddyWatch™ Reports.

A total of 62 EddyWatch™ Reports were archived. Separate fronts were identified for the Loop Current and Eddy Ulysses, which can be plotted individually or combined for LC/LCE frontal analyses. Daily values of the EddyWatch™ frontal distances were computed using the digitized EddyWatch™ Loop Current Eddy (LCE) front closest in time to the day under consideration. The distance from each of the station locations to the nearest LCE front were determined with only minimal masking required to insure the distances to the LCE were accurately represented. An animation of the entire record of fronts with color-coded lines showing station to nearest front distance is included on the CDROM #1. An example frame is shown in Figure 6.

Model Frontal Tracking

Task 2: Define Loop Current and eddy fronts from the models' (nowcast/analysis and forecasts) using the following criteria:

a. 15°C contour at 200m depth
b. Sea surface height at 17cm
c. Velocity magnitude of 75cm/sec at 50m depth

As noted in the Model Skill Assessment Methodology section, different tracking contours were selected. For the HMI Model the tracking contours were 30cm for the SSH field, 21°C isotherm at 200m depth and 75 cm/sec isopleth of velocity magnitude at 50m depth for frontal tracking skill assessment. The tracking contours for the HYCOM Model were 30cm, and13°C, and 65 cm/sec at 200m, respectively.

Task 3: Perform visual quality check on frontal locations to insure that all fronts are represented consistently.

Movies of the Loop Current/Loop Current Eddy (LC/LCE) fronts from each of the model nowcast/forecast are included on the CDROM. Animations of the SSH, temperature and velocity LC/LCE fronts were created from the HMI Model and from the HYCOM Model nowcast/forecasts. These are provided so that the masking applied to determine the distance to the LC/LCE fronts could be evaluated on a case-by-case basis to insure proper tracking of the nearest front by the objective tracking technique. For this analysis it was not necessary to mask the HMI Model. Significant manual intervention was required to insure that all fronts were consistently tracked. In addition the LC of the HYCOM Model was not used in the analysis since the HMI Model was an eddy prediction model only.

One ambiguity in the modeled fronts that has not been addressed is the double isopleth of velocity magnitude that occurs within a LCE or the LC (for example see Figure 11). By definition the closest isopleth of velocity is taken. In cases when a station is within the LC or a LC eddy the nearest point is often on the inner isopleth and not the outer isopleth which more closely corresponds to the LC or LC eddy front.

Task 4: Generate charts of the fronts overlaid from the model and EddyWatch™.

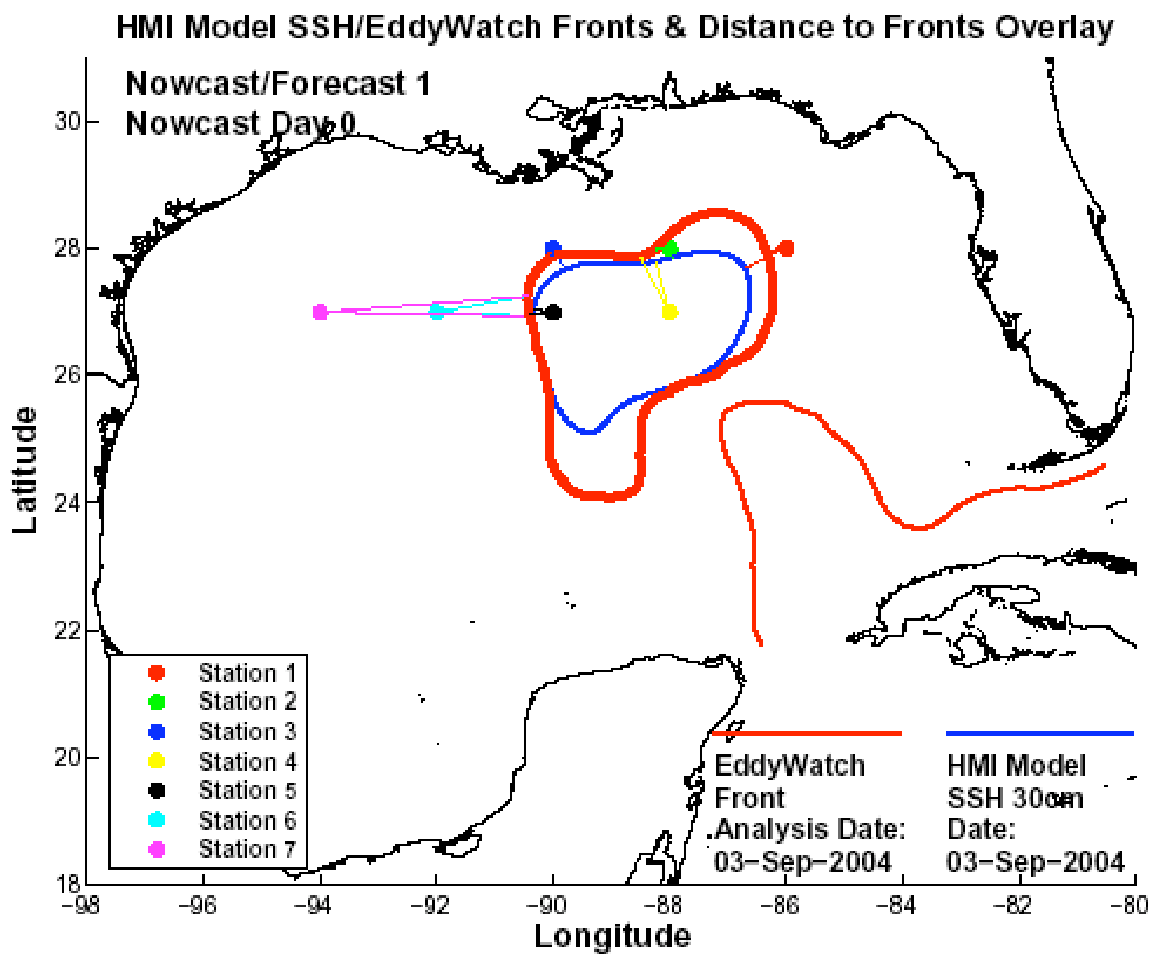

Figure 7: Sample plot of model LCE front overlaid on EddyWatch LC/LCE fronts.

Movies of the Loop Current/Loop Current Eddy (LC/LCE) fronts from each of the model nowcast/forecasts are overlaid with the corresponding EddyWatch™ fronts and are included on the CDROM. Individual charts are also included in the Appendix to this report. A sample is shown in Figure 7.

LCE Frontal Statistics and Time Series

Task 5: Compute frontal statistics from stations listed in Table 1 including mean distance and standard deviation of difference between model and EddyWatch™ fronts, and correlation of model and EddyWatch™ fronts.

Task 6: Plot the time series of frontal distances from EddyWatch™, models (analysis/nowcasts and forecasts), and the difference between models and EddyWatch™ for each station and frontal type.

Task 7: Produce tables containing the summary statistics of the frontal distances from the temperature, sea surface height, and velocity for each station and nowcast/forecast period.

The LC/LCE frontal statistics and time series comprise tasks 5 and 6. Sample statistics and time series plots are included on the following pages (Figures 8, 9 & 10). Complete archives of these plots are included in the Appendix. Task 7 is a tabulation of the results, which are given in Tables 2 through 6.

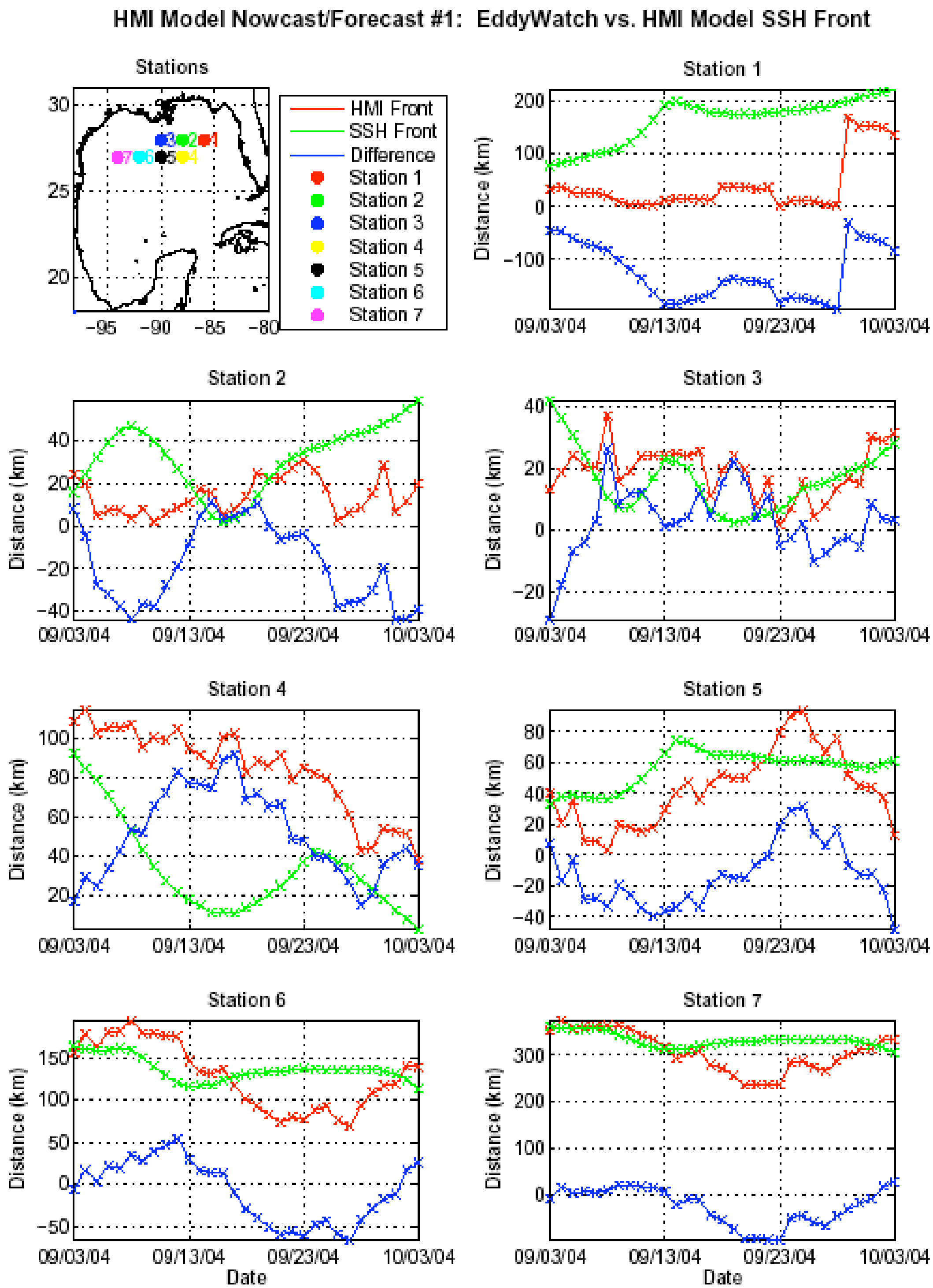

Figure 8. Sample station time series plot for a specific nowcast/forecast experiment.

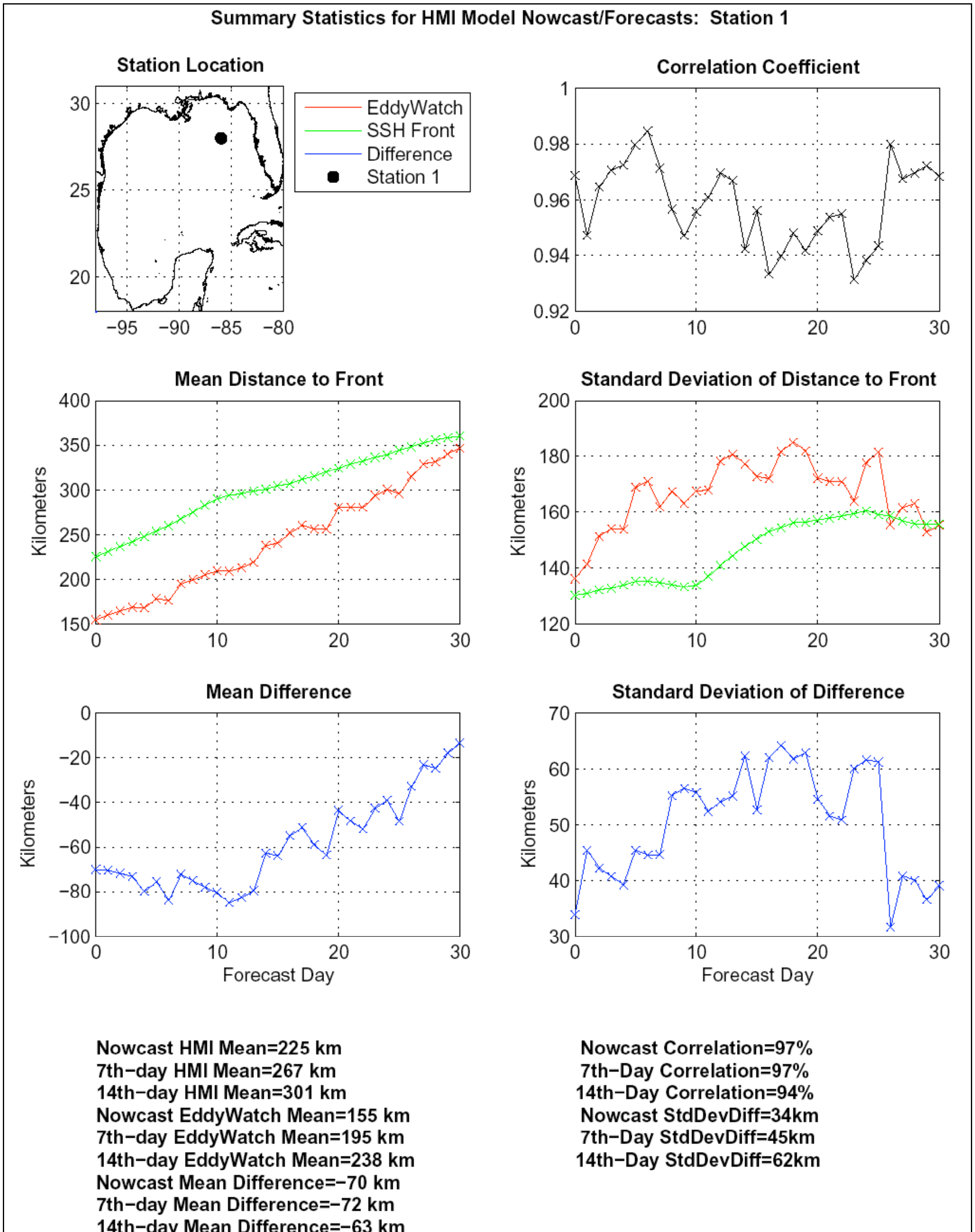

Figure 9: Sample summary statistics plot for HMI model SSH front (Note that the front used is identified in the Station Location legend).

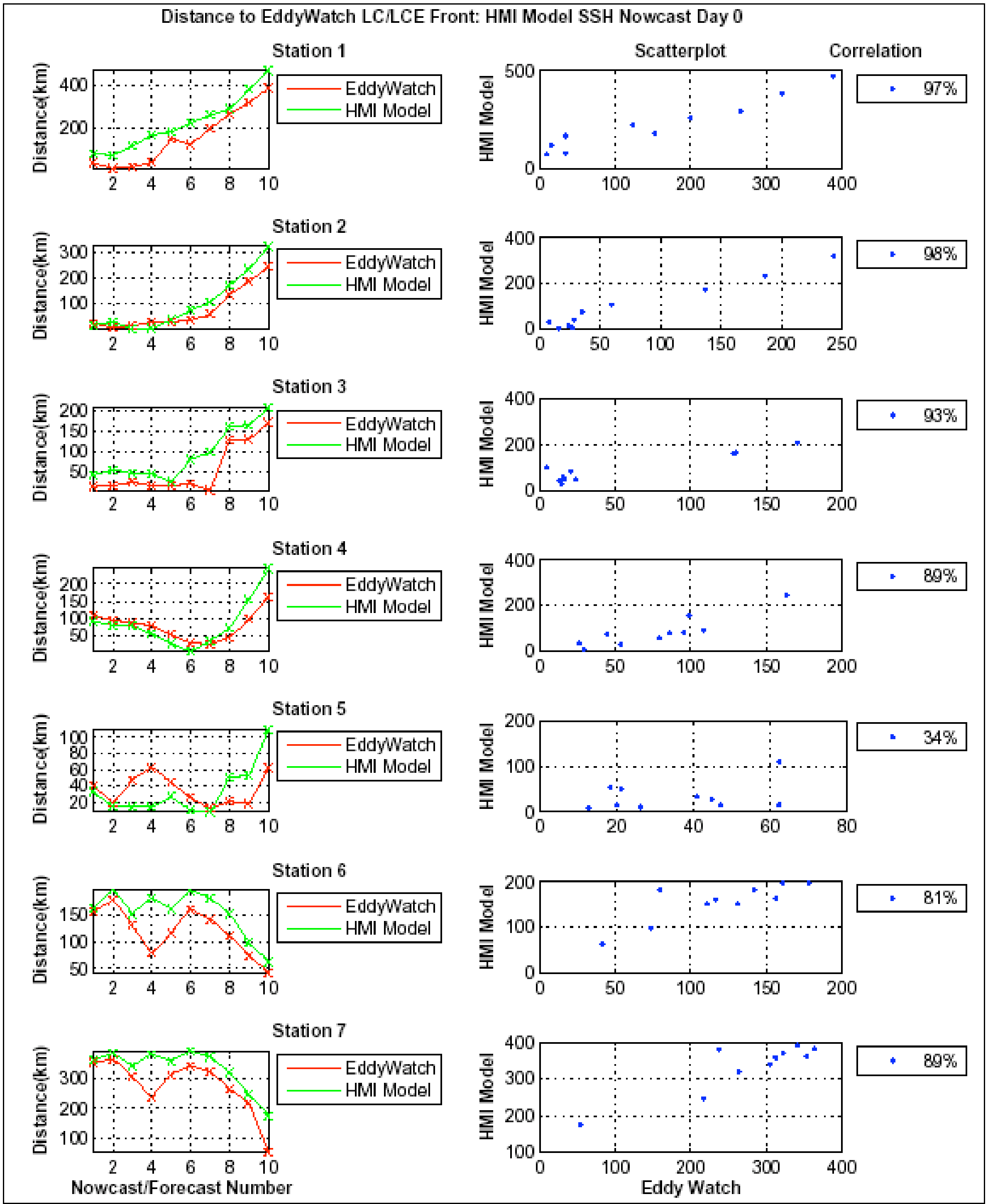

Figure 10: Sample summary statistics plot for calculating correlation coefficient.

TABLE 2: Comparison HMI Model SSH, Temperature and Velocity Frontal Statistics.

HMI Model SSH FRONTS										
HMI Model		Nowcast			7-day Forecast			14-day Forecast		
Nowcast/Forecasts		Total number = 10			Total number = 10			Total number = 10		
		Difference: EddyWatch-model			Difference: EddyWatch-model			Difference: EddyWatch-model		
Reference Station		mean	std	correl	mean	std	correl	mean	std	correl
#	Lat/Lon	(km)	(km)	coeff	(km)	(km)	coeff	(km)	(km)	coeff
1	28°N, 86°W	-70	34	0.97	-72	45	0.97	-63	62	0.94
2	28°N, 88°W	-24	32	0.98	-27	27	0.98	-15	27	0.99
3	28°N, 90°W	-38	23	0.93	-27	29	0.93	-16	21	0.97
4	27°N, 88°W	-7	38	0.89	-3	46	0.88	-7	49	0.9
5	27°N, 90°W	2	30	0.34	-4	25	0.81	-12	18	0.93
6	27°N, 92°W	-36	27	0.81	-20	27	0.83	-12	21	0.9
7	27°N, 94°W	-56	44	0.89	-50	41	0.96	-48	45	0.95
	mean	-33	33	0.83	-29	34	0.91	-25	35	0.94

HMI Model TEMPERATURE FRONTS										
HMI Model		Nowcast			7-day Forecast			14-day Forecast		
Nowcast/Forecasts		Total number = 10			Total number = 10			Total number = 10		
		Difference: EddyWatch-model			Difference: EddyWatch-model			Difference: EddyWatch-model		
Reference Station		mean	std	correl	mean	std	correl	mean	std	correl
#	Lat/Lon	(km)	(km)	coeff	(km)	(km)	coeff	(km)	(km)	coeff
1	28°N, 86°W	-109	29	0.98	-109	43	0.97	-99	57	0.95
2	28°N, 88°W	-55	42	0.98	-57	38	0.97	-49	31	0.99
3	28°N, 90°W	-68	27	0.92	-56	32	0.92	-43	26	0.97
4	27°N, 88°W	-19	67	0.72	-20	71	0.79	-20	73	0.86
5	27°N, 90°W	-12	50	0.02	-16	45	0.65	-19	39	0.83
6	27°N, 92°W	-70	31	0.7	-56	31	0.77	-49	24	0.89
7	27°N, 94°W	-94	51	0.86	-91	51	0.95	-91	74	0.95
	mean	-61	42	0.74	-58	44	0.86	-53	46	0.92

HMI Model VELOCITY FRONTS										
HMI Model		Nowcast			7-day Forecast			14-day Forecast		
Nowcast/Forecasts		Total number = 10			Total number = 10			Total number = 10		
		Difference: EddyWatch-model			Difference: EddyWatch-model			Difference: EddyWatch-model		
Reference Station		mean	std	correl	mean	std	correl	mean	std	correl
#	Lat/Lon	(km)	(km)	coeff	(km)	(km)	coeff	(km)	(km)	coeff
1	28°N, 86°W	-40	34	0.97	-45	49	0.96	-34	62	0.94
2	28°N, 88°W	-9	21	0.99	-17	17	0.99	0	24	0.98
3	28°N, 90°W	-14	25	0.92	-6	28	0.93	-1	17	0.98
4	27°N, 88°W	20	46	0.76	21	52	0.84	18	41	0.93
5	27°N, 90°W	9	22	0.45	9	20	0.83	14	24	0.86
6	27°N, 92°W	-12	27	0.8	3	30	0.8	11	20	0.81
7	27°N, 94°W	-35	46	0.9	-27	44	0.94	-24	47	0.94
	mean	-12	32	0.83	-9	34	0.90	-2	34	0.92

TABLE 3: HYCOM SSH, Temperature and Velocity Frontal Statistics.

HYCOM SSH FRONTS										
HYCOM Model		**Nowcast**			**7-day Forecast**			**14-day Forecast**		
Nowcast/Forecasts		Total number = 9			Total number =9			Total number = 9		
		Difference: EddyWatch-model			Difference: EddyWatch-model			Difference: EddyWatch-model		
Reference Station		mean	std	correl	mean	std	correl	mean	std	correl
#	**Lat/Lon**	(km)	(km)	coeff	(km)	(km)	coeff	(km)	(km)	coeff
1	28°N, 86°W	-60	55	0.90	-52	42	0.94	-59	48	0.86
2	28°N, 88°W	-14	41	0.72	-14	28	0.96	-21	38	0.92
3	28°N, 90°W	-58	41	0.44	-32	34	0.82	-24	36	0.93
4	27°N, 88°W	17	25	0.75	17	13	0.92	14	18	0.90
5	27°N, 90°W	3	30	0.41	8	27	-0.1	5	19	0.46
6	27°N, 92°W	-25	44	0.49	3	58	0.09	4	52	0.27
7	27°N, 94°W	-27	62	0.37	-8	75	0.13	-18	87	0.27
	mean	-23	43	0.58	-11	40	0.54	-14	43	0.66

HYCOM TEMPERATURE FRONTS										
HYCOM Model		**Nowcast**			**7-day Forecast**			**14-day Forecast**		
Nowcast/Forecasts		Total number = 9			Total number = 9			Total number = 9		
		Difference: EddyWatch-model			Difference: EddyWatch-model			Difference: EddyWatch-model		
Reference Station		mean	std	correl	mean	std	correl	mean	std	correl
#	**Lat/Lon**	(km)	(km)	coeff	(km)	(km)	coeff	(km)	(km)	coeff
1	28°N, 86°W	-67	48	0.92	-75	46	0.89	-81	53	0.83
2	28°N, 88°W	-24	23	0.93	-30	22	0.96	-39	32	0.94
3	28°N, 90°W	-76	25	0.87	-55	28	0.87	-44	31	0.95
4	27°N, 88°W	11	19	0.88	9	25	0.77	11	28	0.87
5	27°N, 90°W	3	36	-0.23	3	36	-0.51	8	25	0.2
6	27°N, 92°W	-45	34	0.58	-25	43	0.3	-29	21	0.84
7	27°N, 94°W	-51	39	0.61	-42	46	0.54	-57	43	0.88
	mean	-36	32	0.65	-31	35	0.55	-33	33	.79

HYCOM VELOCITY FRONTS										
HYCOM Model		**Nowcast**			**7-day Forecast**			**14-day Forecast**		
Nowcast/Forecasts		Total number = 9			Total number = 9			Total number = 9		
		Difference: EddyWatch-model			Difference: EddyWatch-model			Difference: EddyWatch-model		
Reference Station		mean	std	correl	mean	std	correl	mean	std	correl
#	**Lat/Lon**	(km)	(km)	coeff	(km)	(km)	coeff	(km)	(km)	coeff
1	28°N, 86°W	-10	68	0.8	-30	34	0.95	-81	80	0.6
2	28°N, 88°W	11	63	0.15	-35	48	0.83	-48	63	0.86
3	28°N, 90°W	-80	52	0.04	-104	82	0.72	-123	66	0.92
4	27°N, 88°W	48	33	0.54	12	38	0.59	-8	65	0.74
5	27°N, 90°W	-57	62	-0.26	-83	88	-0.22	-90	91	0.04
6	27°N, 92°W	-142	71	-0.08	-125	84	-0.32	-123	70	-0.27
7	27°N, 94°W	-155	69	0.06	-137	84	0.07	-130	78	0.33
	mean	-55	60	0.18	-72	65	0.37	-86	73	0.46

TABLE 4: HMI SSH Frontal Statistics – Difference with EddyWatch

HMI SSH Frontal Statistics – Difference with EddyWatch										
	Reference Station	**1**	**2**	**3**	**4**	**5**	**6**	**7**		
Nowcast	Latitude Longitude	28°N 86°W	28°N 88°W	28°N 90°W	27°N 88°W	27°N 90°W	27°N 92°W	27°N 94°W		
	Units								**mean**	**Rms**
	mean (km)	-70	-24	-38	-7	2	-36	-56	-33	
	std (km)	34	32	23	38	30	27	44		33
	correlation	0.97	0.98	0.93	0.89	0.34	0.81	0.89	**rmse** = 52	
1-Week	Units									
	mean (km)	-72	-27	-27	-3	-4	-20	-50	-29	
	std (km)	45	27	29	46	25	27	47		36
	correlation	0.97	0.98	0.93	0.88	0.81	0.83	0.96	**rmse** = 52	
2-Week	Units									
	mean (km)	-63	-15	-16	-7	-12	-12	-48	-25	
	std (km)	62	27	21	49	18	21	45		38
	correlation	0.94	0.99	0.97	0.9	0.93	0.9	0.95	**rmse** = 50	
3-Week	Units									
	mean (km)	-48	-24	-13	-15	-24	-5	-30	-23	
	std (km)	52	32	33	36	23	37	49		39
	correlation	0.95	0.98	0.94	0.97	0.9	0.72	0.93	**rmse** = 47	
4-Week	Units									
	mean (km)	-25	-15	3	-2	-22	-1	-34	-14	
	std (km)	40	45	45	35	36	44	87		50
	correlation	0.97	0.95	0.88	0.97	0.81	0.59	0.72	**rmse** = 54	

Note: The rmse values are the given by the formula:

$$rmse = \sqrt{\frac{1}{N_{stations}} \sum_{i=1}^{N_{stations}} \left(mean_i^{\,2} + std_i^{\,2} \right)}$$

Table 5: HMI Temperature Frontal Statistics – Difference with EddyWatch

HMI Temperature Frontal Statistics – Difference with EddyWatch										
	Reference Station	**1**	**2**	**3**	**4**	**5**	**6**	**7**		
Nowcast	Latitude Longitude	28°N 86°W	28°N 88°W	28°N 90°W	27°N 88°W	27°N 90°W	27°N 92°W	27°N 94°W		
	Units								**mean**	**Rms**
	mean (km)	-109	-55	-68	-19	-12	-70	-94	-61	
	std (km)	29	42	27	67	50	31	51		45
	correlation	0.98	0.98	0.92	0.72	0.02	0.7	0.86	**rmse** = 82	
1-Week	Units									
	mean (km)	-109	-57	-56	-20	-16	-56	-91	-58	
	std (km)	43	38	32	71	45	31	51		46
	correlation	0.97	0.97	0.92	0.79	0.65	0.77	0.95	**rmse** = 80	
2-Week	Units									
	mean (km)	-99	-49	-43	-20	-19	-49	-91	-53	
	std (km)	57	31	26	73	39	24	55		47
	correlation	0.95	0.99	0.97	0.86	0.83	0.89	0.95	**rmse** = 76	
3-Week	Units									
	mean (km)	-81	-52	-42	-29	-32	-45	-80	-51	
	std (km)	49	33	42	59	43	38	63		48
	correlation	0.96	0.99	0.94	0.95	0.86	0.64	0.9	**rmse** = 73	
4-Week	Units									
	mean (km)	-49	-41	-27	-22	-32	-48	-89	-44	
	std (km)	41	46	50	46	38	54	99		57
	correlation	0.97	0.96	0.89	0.97	0.9	0.23	0.62	**rmse** = 75	

Note: The rmse values are the given by the formula:

$$rmse = \sqrt{\frac{1}{N_{stations}} \sum_{i=1}^{N_{stations}} \left(mean_i^2 + std_i^2 \right)}$$

Table 6: HMI Velocity Frontal Statistics – Difference with EddyWatch

HMI Velocity Frontal Statistics – Difference with EddyWatch										
	Reference Station	**1**	**2**	**3**	**4**	**5**	**6**	**7**		
Nowcast	Latitude Longitude	28°N 86°W	28°N 88°W	28°N 90°W	27°N 88°W	27°N 90°W	27°N 92°W	27°N 94°W		
	Units								**mean**	**Rms**
	mean (km)	-40	-9	-14	20	9	-12	-35	-12	
	std (km)	34	21	25	46	22	27	46		33
	correlation	0.97	0.99	0.92	0.76	0.45	0.8	0.9	**rmse** = 40	
1-Week	Units									
	mean (km)	-45	-17	-6	21	9	3	-27	-9	
	std (km)	49	17	28	52	20	30	44		37
	correlation	0.96	0.99	0.93	0.84	0.83	0.8	0.94	**rmse** = 43	
2-Week	Units									
	mean (km)	-34	0	-1	18	14	11	-24	-2	
	std (km)	62	24	17	41	24	20	47		37
	correlation	0.94	0.98	0.98	0.93	0.86	0.81	0.94	**rmse** = 41	
3-Week	Units									
	mean (km)	-21	3	3	5	7	13	-12	0	
	std (km)	53	32	31	28	29	32	49		38
	correlation	0.95	0.98	0.94	0.98	0.85	0.77	0.92	**rmse** = 39	
4-Week	Units									
	mean (km)	10	14	26	6	9	14	-15	9	
	std (km)	43	44	44	32	33	48	94		52
	correlation	0.97	0.95	0.87	0.97	0.84	0.5	0.67	**rmse** = 54	

Note: The rmse values are the given by the formula:

$$rmse = \sqrt{\frac{1}{N_{stations}} \sum_{i=1}^{N_{stations}} \left(mean_i^{\,2} + std_i^{\,2} \right)}$$

4. DISCUSSION

The statistical metrics computed in the skill assessment methodology of Kantha et al., (in press) are based on the mean and standard deviation (std) of the EddyWatch™ and model station-to-LCE-frontal distance differences, and the correlations of the EddyWatch™ and model station-to-LCE-frontal distance times series. Summary statistics were reported by averaging these mean, std and correlation values over the seven stations listed in Table 1. These statistics for temperature, SSH and velocity tracking of the Loop Current eddy (LCE) front in the HMI and HYCOM nowcast/forecasts are shown in Tables 2 & 3, which are identical in form and content to those of the CUPOM skill assessment published in the confidential report by Kantha and Choi (2003).

Before discussing these statistics, a discussion of the station-to-LC Eddy Front distance difference metric is necessary to understand the complexity inherent in this simple calculation.

Station-to-LCE Front Distance Difference: This statistical metric is a measure of the relative offset between the EddyWatch™ LCE fronts and the model prediction. By definition the nearest point to a station will be on a line normal to the given front, which typically will be along lines in different directions given the differences in the LCE frontal boundaries exhibited between the EddyWatch™ and modeled fronts (see Figure 7). The discrepancy between the actual local offset and the estimated offset from the station-to-LCE frontal distance difference becomes larger as boundary shapes differ. This means that the metric is a less than perfect measure of the local relative distance between the fronts, especially when eddy boundary shapes differ dramatically. Nevertheless, the metric achieves a minimum value in the case where the EddyWatch™ and modeled fronts coincide or are locally self-similar in shape (i.e. locally differ only by a translation and scale factor), and the station location is completely inside or outside of the EddyWatch™ and model LCE fronts.

When the station is between the EddyWatch™ and model LCE fronts an ambiguity arise because of the scalar nature of the distance metric. For example, this discrepancy is most severe when a station is both between and exactly the same distance from the fronts. In that case, a zero mean difference is calculated although the fronts are actually separated by the sum of the distances. Ideally, these cases could be identified and the sum of the distances, not the differences, used to compute the scalar offset. This hasn't been implemented to keep our analysis consistent with the methodology of Kantha et al. (in press).

Another ambiguity that hasn't been addressed is the double isopleth of velocity magnitude that often occurs within a LCE or the LC. By definition the closest isopleth of velocity is taken, which in cases when a station is within the LC eddy the nearest point is often on the inner isopleth and not the outer isopleth that more closely corresponds to the LCE front. A nice example of this ambiguity is seen at stations 4&5 in Figure 13.

Now let us consider the mean, std and correlation skill assessment statistics keeping in mind the limitations of the station-to-LCE front distance difference metrics.

Mean and STD of Station-to-LCE Front Distance Differences: As noted above, the mean station-to-LCE front distance difference is a measure, albeit poor, of the relative offset between the two LCE fronts. It is not exact because of the influence of the geometry of the stations relative to the fronts and the shape of the individual fronts. There will also be a difference between the EddyWatch™ surface fronts and the modeled fronts at depth because of the physics. These effects can and do vary with time as the LCE propagates westward, so we expect to see secular changes in the mean values during the progression from the nowcast through the forecast time period. These secular trends are clearly seen in Tables 4, 5 and 6. Note that for each of the HMI model frontal tracking metrics (SSH, temperature, velocity) the relative offsets all increase with increasing forecast time. The velocity frontal tracking has the smallest relative offsets, followed by the SSH and then temperature. The tendency is for the offsets of the HMI fronts to be away from the LC or LCE center. Note, however, that we expect an offset in this direction between the EddyWatch™ fronts and model fronts, especially for the fronts based on SSH or temperature/velocity at depth. This is because the EddyWatch™ fronts are based on surface observations and there is a tendency for the surface signature of the LCE and LC front to lean to the right relative to the front at depth assuming one is looking in the direction of flow. The trends in the HYCOM means are more difficult to identify because of the limited 14-day forecast time period, however, they tend to agree with the HMI results.

Kantha el al. (in press) interpreted the increasing offsets they observed as a function of forecast time to be a measure of the decrease in CUPOM forecast skill with time. We believe this is not advisable because the mean over stations has a tendency to average out errors and the mean values are not stationary in time because of the secular trends we have identified. For example, if an LCE forecast consistently overestimates the translation velocity of the eddy the mean of the mean station offsets may not identify this error as long as the stations are distributed both in front of and behind the eddy.

Kantha et al., (in press) also quoted standard deviation (std) values, which are more appropriate for skill assessment. We note, however, that for std values it is preferable to quote the root-mean-square (rms) of the station std values rather than a mean as was done in Tables 2 & 3 to follow the methodology of Kantha and Choi (2003). The more statistically rigorous rms of the station stds are included in Table 4 through 6 for the HMI model. Note that the actual rms std values are not significantly different than the mean std values quoted in Table 2 for the HMI model, so the results from Kantha and Choi (2003) and Kantha et al., (in press) can be used as an approximation.

The std distance difference values can be used to calculate the expectation that the modeled LCE front is within a certain distance of the EddyWatch™ LC fronts for a give probability distribution. This is a much better metric than the "on average" skill metrics quoted by Kantha et al. (in press) in which they stated that "on the average, the model reproduces the LC from to about 11-12 km in the hindcast and nowcast modes, and 19 km in the forecast mode, when temperature criterion is used to discern the front". In fact, as stated in the executive summary, the CUPOM nowcasts and 4-week forecasts of the LC thermal front are expected to be *within 20 km of the mean relative offset* to the EddyWatch™ LC fronts only about 28% and 18% of the time, respectively, based on the limited statistics presented in Kantha et al., (in press) and assuming a normal distribution of the model and EddyWatch™ differences. Using similar assumptions, the HMI velocity front nowcasts and 4-week forecasts are expected to be within 20 km of the mean relative offset to the EddyWatch™ fronts 46% and 30% of the time, respectively. HMI SSH fronts are 46% and 31%, respectively, which is comparable or better albeit with a greater mean relative offset. HMI temperature fronts were only 34% and 27%, respectively, but still better than the values quoted for CUPOM. We caution, however, that making any determination of the relative skill between these two models may not be appropriate because the statistics are based on different time periods and LC configurations.

The overall mean and mean(std) offset between the modeled and EddyWatch™ fronts over the nowcast to 2-week forecast (Tables 2 & 3) time periods can be used to make a relative ranking of the model skill for each of the tracking techniques. This ranking is shown in Table 7 below. We have quoted the range of mean and mean(std) values found in Tables 2 & 3 and the ranking based on the statistics used by Kantha et al. (in press).

Table 7. Relative ranking of skill.

Rank - Model/Front	mean range (km)	mean(std) range (km)
1. HMI/Velocity	-12 to -2	32 to 34
2. HMI/ SSH	-36 to -31	32 to 35
HYCOM/Temperature	-33 to -25	33 to 35
4. HYCOM/SSH	-23 to -14	40 to 43
5. HMI/Temperature	-61 to -53	42 to 46
6. HYCOM/Velocity	-55 to -86	60 to 73

Correlation Statistics of Station-to-LCE Front Distance: This is the least useful of the skill assessment metrics. Once again geometry has a significant influence on the statistic that makes comparisons of model skill difficult. The primary effect of geometry is to increase correlation with increasing distance of the front from the station location because of the commensurate increase in the percent of distance variance explained by being far from the front (see Figure 10). This simply means that stations far from the front tend to have higher correlation, which should not be a characteristic of a skill metric. Correlations at individual stations are also adversely affected by the ambiguities discussed earlier that are associated with station locations between the fronts. Another clear deficiency of this metric is that it is possible for a very poor model simulation to have good correlation just by being consistently wrong. An example would be the case where the model's LCE propagation speed was consistently too fast compared with observations. Over time the individual forecast experiments would deviate from "truth" in a linear and thus consistent manner and thereby exhibit a nearly perfect 100% correlation; however, by almost any other measure of skill the forecast would be deemed poor. This is the primary reason why higher correlations are found in the 14-day forecasts in Tables 2 & 3 than in the corresponding nowcasts. Because of these interpretation difficulties, we recommend using other metrics for skill assessment. Nevertheless, the HMI model generally exhibits higher correlation of the station-to-LCE fronts with EddyWatch™ than the HYCOM model.

RMS Station-to-LCE Front Distance Differences: The combination of the mean and standard deviation of the station-to-LCE front distance difference can be used to calculate an rms error value, which is equivalent to calculating the rms station-to-LCE frontal differences and root-mean-square summing over all stations and experiments. This rms value includes both the random and systematic offset of the model and EddyWatch™ fronts.

These values have been calculated from the tabulated mean and std values at each station and are identified as rmse in the Tables 4 through 6. The equivalent rms values have also been computed from the rms over all experiments of the rms over all stations of the difference between EddyWatch™ fronts and HMI Model and HYCOM fronts. These are shown as a function of the nowcast/forecast time period and frontal tracking technique in the summary plots in Figure 1. Summary plots including the statistics for each of the individual nowcast/forecast experiments and the rms over all experiments are shown in Figures 11 and 12.The rms error as a function of forecast day can also be computed for individual experiments to monitor the skill of an individual nowcast/forecast experiment. A sample frame from animations made to evaluate station to front range difference and RMS values for each model and experiment is shown in Figure 13. The animations are included on CCAR Report CD-ROM #1.

Many of the larger range deviations are associated with frontal lobes in the EddyWatch™ reports that correspond to cyclonic eddies on the periphery of the LC or a LCE. These cyclones displace the surface thermal front that the analyst uses tracks to draw the EddyWatch™ LC/LCE front while minimally affecting the tracking contours based on the SSH and subsurface velocity and SST fields. There are many issues associated with using the EddyWatch™ fronts for skill assessment; however, because the systematic errors affect both models under consideration it is unlikely that the EddyWatch™ standard is unfair to either model as long as the time period used for evaluation is the same.

The use of the EddyWatch™ fronts for skill assessment, however, does increase the “noise” of observed “truth”. This requires longer nowcast/forecasts time periods before model errors can overwhelm the observational error enough to clearly determine that a model solution is deviating significantly from “truth”. The 2-week forecasts from the HYCOM model are clearly too short for this to happen. The HMI model shows the characteristic increase in rms error at about 3-weeks for the velocity front tracking, which as we will note in the next section is the most skillful solution. The less skillful SSH and temperature frontal tracking begins to increase only near the very end of the 30-day nowcast/forecast interval (see Figure 11).

The summary conclusions regarding the relative ranking using the RMS skill metric will be made in the following section.

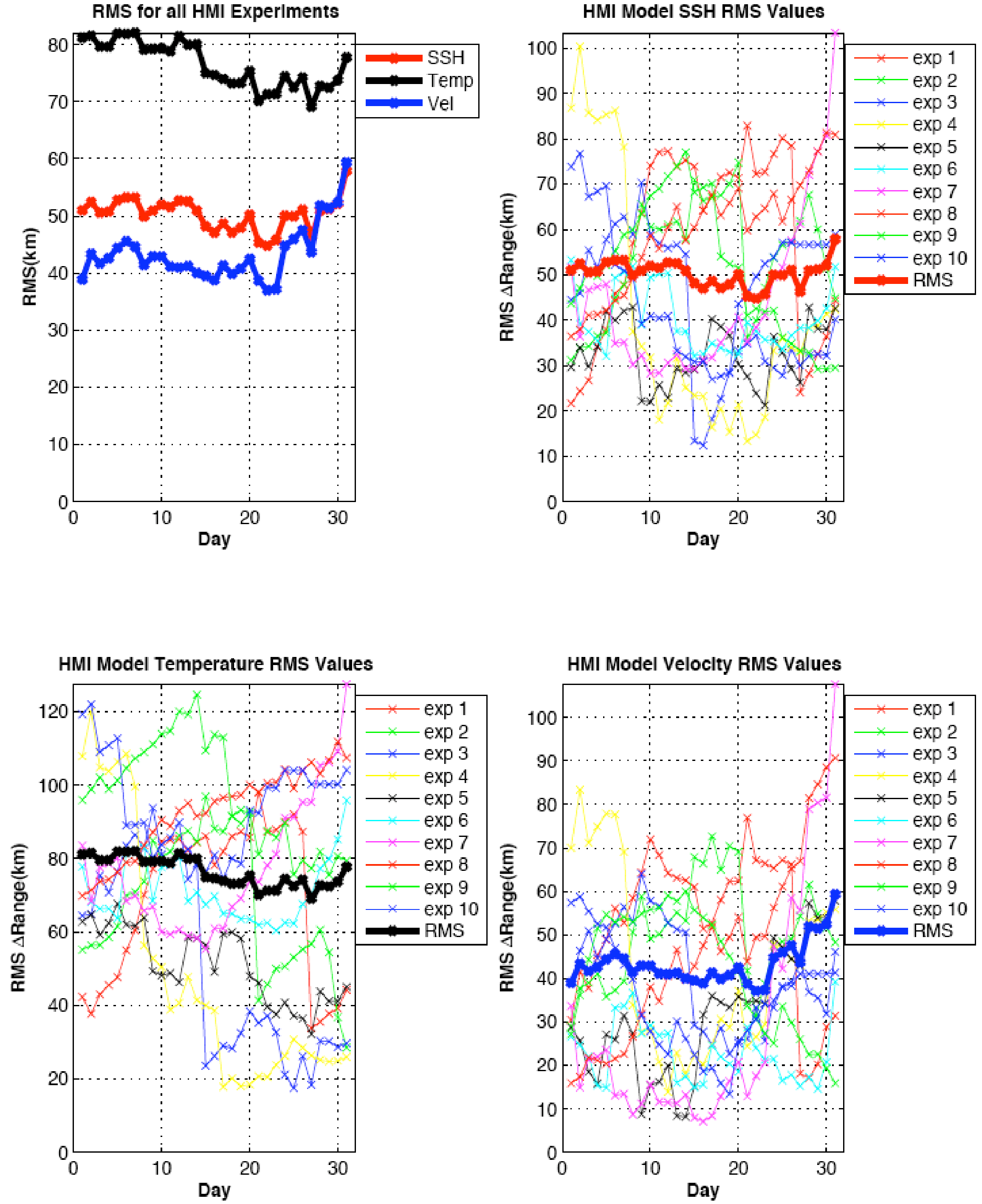

Figure 11. Comparison of rms over all experiments of the rms over all stations if the difference between EddyWatch™ and the HMI Model LCE fronts as a function of the nowcast/forecast time period and frontal tracking technique.

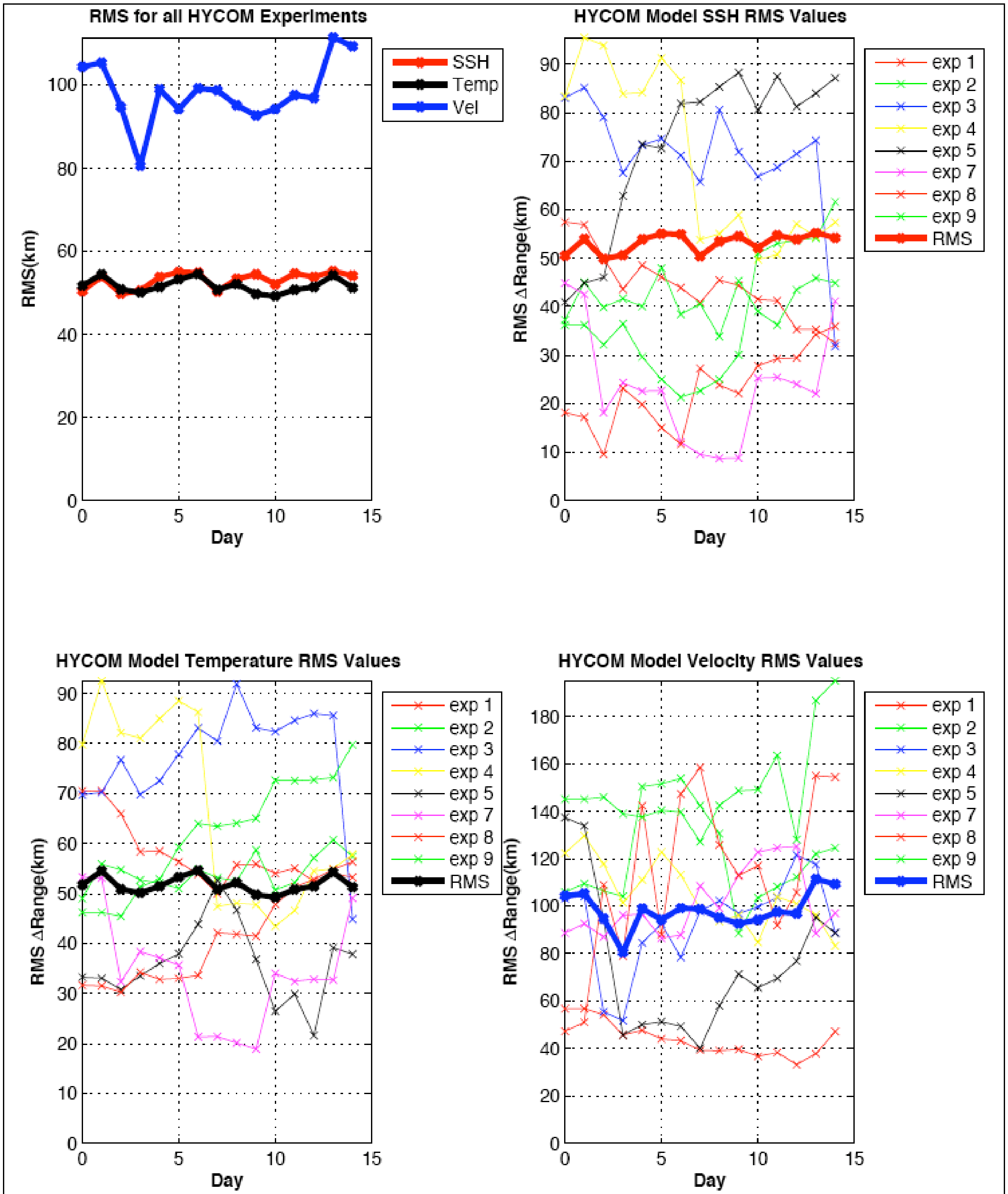

Figure 12. Comparison of rms over all experiments of the rms over all stations if the difference between EddyWatch™ and the HYCOM Model LCE fronts as a function of the nowcast/forecast time period and frontal tracking technique.

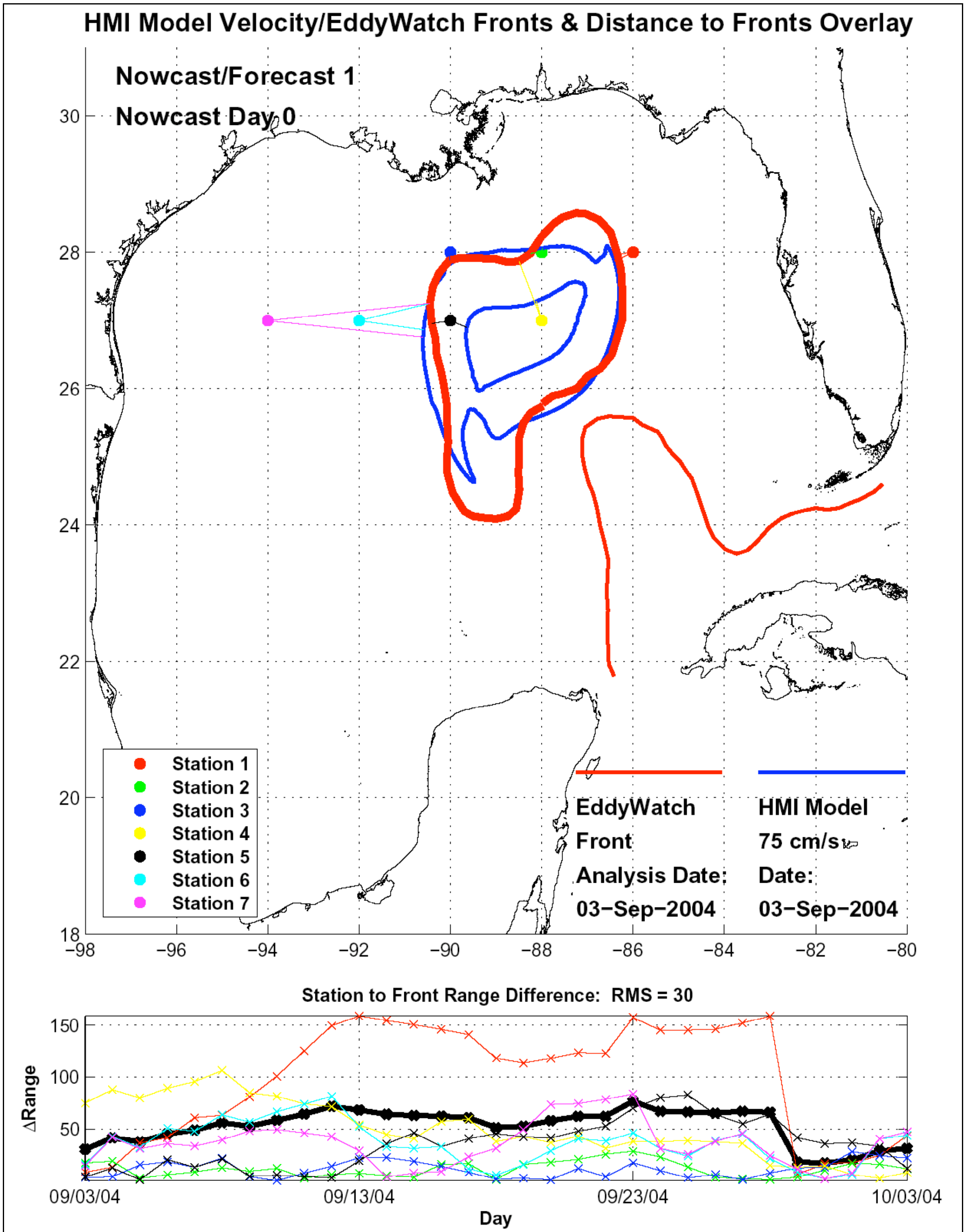

Figure 13. Sample frame from animations made to evaluate station to front range difference and RMS values for each model and experiment.

5. CONCLUSION

Based on an evaluation relative to the Loop Current (LC) Eddy Ulysses fronts depicted in the HMI EddyWatch™ reports, the nowcast/forecast skill of the HMI and HYCOM data assimilative models have been assessed for the time period from September 1 through December 15, 2004. The most useful statistical metric developed for this skill assessment was the RMS over all stations and experiments of the difference between the model and EddyWatch™ LC eddy fronts' distance as a function of nowcast/forecast day. Smaller RMS values imply better relative skill that includes both the random and systematic offsets between the modeled and observed fronts. The key points of this evaluation are:

- The LC Eddy Ulysses velocity fronts depicted by the HMI data assimilation model exceeded the skill exhibited by the HMI model temperature and sea surface height (SSH) fronts and the skill displayed by the HYCOM model nowcast/forecast LC eddy fronts based on either velocity, temperature or SSH.
- The skill of the HYCOM LC eddy fronts based on temperature approached the skill of HMI LC eddy fronts based velocity, but did not exceed it.
- The relative ranking of the HMI model skill from highest to lowest is velocity, SSH and temperature tracking of the LC eddy front, respectively.
- The relative ranking of the HYCOM model skill from highest to lowest is temperature, SSH and velocity tracking of the LC eddy front, respectively.

We also made an assessment of the HMI and HYCOM models following the methodology of Kantha et al. (in press), which has been used for the skill assessment of CUPOM. The relative rankings of the HMI and HYCOM models for the various frontal tracking techniques using that method (see Table 7) agree well with the conclusions based on the RMS statistics even though there are significant difficulties associated with the metrics and statistics used as has been discussed in this report.

A more difficult issue that needs to be addressed is how to improve the estimation of the relative offset of the "true" and modeled front. Significant improvements in the methodology of Kantha et al. (in press) can and should be made. A number of options exist, however, the ultimate implementation in software must be carefully considered and tested before systematic skill assessment of models of the Gulf of Mexico can become routine.

6. REFERENCES

Chassignet, E. P., L. T. Smith, G. R. Halliwell, and R. Bleck, 2003: "North Atlantic Simulation with the Hybrid Coordinate Ocean Model (HYCOM): Impact of the vertical coordinate choice, reference density, and thermobaricity, *J. Phys. Oceanogr.*, 33, 2504–2526.

Cooper, M. and K. Haines, 1996: "Altimetric Assimilation with Water Property Conservation", *J. Geophys. Res.,* 101, 1059–1077.

Kantha, L. K., and J. K. Choi, 2003: "A Data-Assimilative Model of the Gulf of Mexico and the Caribbean Sea for Operational Use: Development and Testing", confidential report to EJIP/CASE.

Kantha, L.K., J. K. Choi, K. J. Schaudt, and C. K. Cooper, (in press): A Regional Data-Assimilative Model for Operational Use in the Gulf of Mexico, In *New Developments in the Circulation of the Gulf of Mexico*, edited by W. Sturges, A. Lugo-Fernandez, W. Schmitz, Jr., and L-. L.- Oey.

APPENDICES

A number of CD-ROMs are included with this report. They are tabulated below.

CD-ROM #1: Report, Appendices, and Movies

- A copy of this report.
- Appendix of EddyWatch PDFs (sample in report: Figure 2).
- Appendix of Digitized EddyWatch PDFs (sample in report: Figure 3)
- Appendices of PDF plots of HMI and HYCOM for each of the following:
  - Frontal Differences (sample in report: Figure 8)
  - Summary Statistics (sample in report: Figure 9)
  - Correlation (sample in report: Figure 10)
  - RMS (sample in report: Figure 11, 12, 13)
- Movies
  - EddyWatch Report Movie
  - HMI and HYCOM Movies for each of the following:
    - Distance to Front (sample frame in report: Figure 6)
    - Overlay (sample frame in report: Figure 7)
    - Overlay RMS (sample frame in report: Figure 13)

The CD-ROMs #2 through #4 contain PDF versions of all of the movie frames so that they can be used individually in presentations and reports.

Supplementary CD-ROM #2: HMI_EddyWatch Overlay PDFs

Supplementary CD-ROM #3: HMI_EddyWatch Overlay RMS PDFs

Supplementary CD-ROM #3: HYCOM_EddyWatch Overlay/Overlay RMS PDFs