

NOTES AND CORRESPONDENCE

Investigating the Utility of Using Cross-Oceanic Training Sets for Superensemble Forecasting of Eastern Pacific Tropical Cyclone Track and Intensity

MARK R. JORDAN II, T. N. KRISHNAMURTI, AND CAROL ANNE CLAYSON

Department of Meteorology, The Florida State University, Tallahassee, Florida

(Manuscript received 1 March 2007, in final form 14 November 2007)

ABSTRACT

This paper examines how combining training-set forecasts from two separate oceanic basins affects the resulting tropical cyclone track and intensity forecasts in a particular oceanic basin. Atlantic and eastern Pacific training sets for 2002 and 2003 are combined and used to forecast 2004 eastern Pacific tropical cyclones in a real-time setting. These experiments show that the addition of Atlantic training improves the 2004 eastern Pacific forecasts. Finally, a detailed study of training-set and real-time model biases is completed in an effort to determine why cross-oceanic training may have helped in this instance.

1. Introduction

Since its inception in 1998, T. N. Krishnamurti's superensemble technique has led to significant improvements in the forecasting of temperature regimes and precipitation patterns (Krishnamurti et al. 1999). One of the technique's greatest accomplishments, however, has been its capability of improving tropical cyclone track and intensity forecasts. The superensemble proved its worth in 2004, as it provided the best track and intensity forecasts for Atlantic tropical cyclones at all synoptic times. The superensemble technique is unlike other numerical models since the technique does not try to independently model the atmosphere using governing equations, observational data, and differencing techniques. Instead, the superensemble takes many previous forecasts from several different models (training) and corrects the biases of each model using a least squares minimization technique. Subsequently, each model is then weighted using multiple linear regression. These weights and bias corrections are then applied to the suite of current model forecasts, and a single superensemble (biases removed and weighted) forecast is

produced. For the experiments in this paper and in real-time practice at The Florida State University, weights and bias corrections are applied to tropical cyclone forecasts uniformly, regardless of cyclone strength or location in a particular basin. Williford (2002) provides more details on how a superensemble forecast is developed.

For tropical cyclone superensemble forecasting, experiments have shown that larger training sets tend to provide better forecasts than smaller training sets. However, obtaining large enough training sets can be a challenge. The physical and dynamical characteristics of numerical models are regularly changed, so one potentially has the problem of using an outdated training set that incorrectly characterizes the biases and relative strength of an updated model. Furthermore, some tropical cyclone seasons are relatively inactive and, therefore, may not provide a sufficient number of forecast cases for a future season. This paper investigates the potential utility of combining training data from two separate ocean basins, the Atlantic and eastern Pacific, to provide a larger training set to forecast eastern Pacific tropical cyclones during the 2004 season. Because the same numerical models are used for tropical cyclone forecasts in both basins, compatibility is not an issue as the same versions of the models are applied to each basin. Based on the theories outlined above, such an experiment could lead to improved tropical cyclone forecasts.

Corresponding author address: Carol Anne Clayson, Dept. of Meteorology, and Geophysical Fluid Dynamics Institute, The Florida State University, 404 Love Bldg., Tallahassee, FL 32306.
E-mail: clayson@met.fsu.edu

TABLE 1. List of member model acronyms and their associated descriptions (Rhome 2007).

Model acronym	Description of model
OFCI	National Hurricane Center's official forecast (interpolated)
NGPI	U.S. Navy Operational Global Atmospheric Prediction System (NOGAPS) model (interpolated)
GFDI	Geophysical Fluid Dynamics Laboratory model (interpolated)
UKMI	Met Office Model (interpolated)
GUNS	Ensemble Mean of GFDI, UKMI, and NGPI
GUNA	Ensemble mean of the National Centers for Environmental Prediction's (NCEP) GFDI, UKMI, NGPI, and Global Forecast System model interpolated
BCEM	Bias-corrected ensemble mean of the member models used in this study
ENSM	Ensemble mean of the member models used in this study
SHF5	5-day Statistical Hurricane Intensity Forecast model
DSHP	Decay-Statistical Hurricane Intensity Prediction Scheme model

Previous research involving the modification of training sets for improving superensemble forecasts has primarily focused on developing location-specific training sets for tropical cyclone forecasting. Williford (2002) separates Atlantic basin training into subsections based on location within the Atlantic basin. Then, superensemble forecasts are generated for real-time tropical cyclones based on where the cyclone forms in the basin. For example, a real-time tropical cyclone that forms and moves through the Gulf of Mexico is forecasted using only Gulf of Mexico storms in the training set. The results of these experiments show marginal improvements in the overall tropical cyclone track and intensity forecasts. The main problem with maintaining location-specific training sets involves the small size of the training sets individually, especially because it is difficult to obtain a large number of instances in each training set before numerical models undergo significant modifications.

2. Methods

Table 1 defines the acronyms used to identify each model. The numerical models chosen in developing superensemble forecasts for the 2004 eastern Pacific tropical cyclone season are outlined in Table 2. Detailed information about these numerical models can be found online (at www.nhc.noaa.gov/modelsummary.shtml). The experiments presented in this paper were conducted as if real-time hurricane track and intensity

TABLE 2. Numerical models used in producing superensemble forecasts for the 2004 eastern Pacific tropical cyclone season.

Early lat	Late lat	Early lon	Late lon	Early intensity	Late intensity
NGPI	NGPI	OFCI	OFCI	OFCI	OFCI
UKMI	UKMI	GFDI	UKMI	GFDI	GFDI
GUNS	GUNS	NGPI	GUNA	UKMI	UKMI
		GUNA		SHF5	SHF5
				DSHP	DSHP

forecasts were being developed. Many preliminary experiments were conducted to determine the distribution of models that would be used in the actual experiments, and the distribution of models in Table 2 represents the collection of models that was determined to provide the best forecasts.

The early and late designations for latitude, longitude, and intensity models refer to the synoptic times for which each group of models was used. Early models were used for 12–72-h forecasts while late models were used for 84–120-h forecasts. The 2002 and 2003 tropical cyclone forecasts for both the Atlantic and eastern Pacific were used in the training set for forecasting the 2004 eastern Pacific tropical cyclone season. Prior to the beginning of work on this project, it was determined that no significant changes had been made to the numerical models between 2002 and 2004, except for the GFDI model. However, given the large number of changes that are often made to numerical models over a couple of years, the consistency of the member models is good in this situation. Therefore, using a 2002–03 training set for forecasting cyclones in 2004 would be appropriate.

No cross-validation experiments were conducted in any experiments because the use of such methods would invalidate the purpose of recreating a real-time scenario. Cross validation involves using the training associated with future storms to determine how forecast-year training may differ from past-years' training sets. While such cross-validation experiments can be useful in a research mode, those types of experiments provide no guidance for improving real-time forecasts. In determining error calculations, a homogeneous forecast sample was used so that all models and forecasts could be compared equally. The number of cases involved in the error calculations is different from published error results found on the National Hurricane Center's Web site because superensemble forecasts are not made for tropical depressions and storms that have made landfall, regardless of strength. Therefore, the National Hurricane Center has a higher number of cases in its error calculations.

TABLE 3. The 2004 tropical cyclone RMS track errors (km) for various member models and the ensemble mean (ENSM) of those member models.

Hour	OFCI	GFDI	NGPI	UKMI	GUNS	GUNA	ENSM
12	62.8	59.2	74.5	75.5	56.5	59.0	61.4
24	107.9	101.8	128.6	136.2	94.4	99.9	105.2
36	148.1	143.9	168.1	188.4	130.3	135.8	145.6
48	190.0	180.1	210.3	238.4	169.9	171.5	185.0
60	239.3	233.6	253.0	291.9	219.0	215.1	231.7
72	284.3	271.5	315.8	346.9	269.8	257.5	272.0
84	362.4	346.5	354.2	434.3	343.5	316.3	357.9
96	460.7	439.6	415.0	582.5	430.7	407.0	459.7
108	579.1	554.9	480.7	781.0	552.5	533.8	605.0
120	724.6	715.0	534.3	1116.6	770.8	750.1	851.6

3. Results

The first experiment conducted involved using 2002–03 eastern Pacific training set to forecast 2004 eastern Pacific tropical cyclones. The previous forecast training set comprised 382 forecast instances. For this and all other experiments, the number of actual forecasts made for hours 12, 24, 36, 48, 60, 72, 84, 96, 108, and 120 were 138, 116, 99, 82, 69, 53, 43, 33, 23, and 14, respectively. Tables 3–6 show the root-mean-square track and intensity errors for the superensemble along with other model and forecast errors for the 2004 eastern Pacific hurricane season. Tables 3–6 indicate that the superensemble track forecasts perform relatively well compared to other models for 12–48-h forecasts. During this period, superensemble performance only slightly lags behind the performance of the ensemble models GUNS and GUNA. During the 60–120-h forecasts, however, superensemble forecast errors increase significantly compared to the errors of other models and forecasts. Many of the numerical models and ensemble models tend to outperform the superensemble during this time frame. The same type of error pattern is also

noted when examining the intensity errors of the superensemble and other models and forecasts. For 12–60-h forecasts, the superensemble performed well, alternating between the best intensity model and the second-best intensity model. However, 72–120-h forecasts show a significant drop in accuracy for the superensemble, as it trails several of its member models during these times.

To assess the effects of an increased training set on the superensemble forecast, a second experiment was conducted involving a combined 2002–03 eastern Pacific training set and a 2002–03 Atlantic training set. This training set included 811 forecast instances. Tables 3–6 show root-mean-square track and intensity errors for the superensemble and other numerical models. Tables 3–6 show that 12–60-h superensemble track forecasts improve slightly over the model's performance when using only Pacific training; however, the superensemble still trails the GUNS and GUNA ensemble models during these periods. For 72–120-h track forecasts, the superensemble shows significant improvement over its performance when using just Pacific training, although the model still demonstrates less skill than OFCI, GFDI, and the ensemble models (GUNS and GUNA). On the other hand, superensemble intensity forecasts improve remarkably with the use of combined training. The superensemble has the overall best intensity forecasts during hours 12–60. During hours 72–120, the superensemble's skill decreases relative to other numerical models and forecasts. During this period, the superensemble is routinely outperformed by DSHP, SHF5, and OFCI. Even though the skill of the superensemble noticeably decreases during the later hours, the improvement in the forecasts during the early hours significantly improves the superensemble's overall yearly performance since most of the verified forecasts occur during the early hours. These results appear to indicate that the use of combined-basin train-

TABLE 4. The 2004 tropical cyclone RMS track errors (km) for the BCEM and the Florida State Superensemble (FSSE) associated with various training sets.

Hour	BCEM (Pacific training only)	BCEM (Atlantic training only)	BCEM (combined training)	FSSE (Pacific training only)	FSSE (Atlantic training only)	FSSE (combined training)
12	61.7	60.5	59.8	58.6	57.4	57.5
24	108.1	103.3	102.9	100.1	97.8	98.0
36	153.6	141.6	143.9	142.5	137.7	139.2
48	197.8	182.3	186.5	188.4	182.0	183.5
60	248.0	232.2	236.6	247.2	232.1	235.2
72	293.6	280.8	283.1	305.5	287.9	290.4
84	375.7	362.1	364.7	400.3	365.1	373.4
96	489.8	476.3	478.7	496.3	456.2	465.4
108	642.2	617.7	623.5	631.9	560.0	575.0
120	894.8	857.1	865.1	841.8	749.3	772.6

TABLE 5. The 2004 tropical cyclone RMS intensity errors ($m s^{-1}$) associated with various member models and the ensemble mean (ENSM) of those member models.

Hour	OFCI	GFDL	UKMI	SHF5	DSHP	ENSM
12	2.4	2.7	3.0	2.4	2.1	2.6
24	3.8	4.0	5.2	4.0	3.5	4.1
36	4.6	4.6	7.0	4.8	4.1	5.2
48	5.0	4.9	8.0	4.9	4.5	5.3
60	5.2	4.8	7.8	5.0	4.6	5.3
72	4.9	5.0	7.4	4.6	4.4	5.1
84	4.8	6.4	7.5	4.7	4.3	4.6
96	4.7	8.6	8.2	4.7	4.4	4.6
108	4.5	10.5	8.9	4.3	3.6	4.8
120	6.7	10.3	8.2	4.9	3.5	5.8

ing can be helpful if the models used for both basins are the same.

Because the addition of an Atlantic training set improved the superensemble forecasts, a third experiment was conducted to see whether the additional training set (the 2002–03 Atlantic training set) alone would result in better forecasts than the training sets used in the two previous experiments. The 2002–03 Atlantic training set included 429 previous forecast instances. Tables 3–6 show the results of using this training set to forecast eastern Pacific tropical cyclone track and intensity. The track results, as shown in Tables 3 and 4, indicate that the forecast skill for hours 12–60 is not significantly different than the skill seen when using combined training. However, forecasts for hours 72–120 show significant improvement over the skill of using either combined training or solely Pacific training. While the superensemble still trails GUNS and GUNA in forecast skill, the model routinely shows improved skill over other dynamical, statistical, and subjective forecasts. Superensemble intensity forecasts using an Atlantic training set show similar characteristics as the track forecasts. In this instance, hour 12–60 forecasts are ac-

tually a bit worse than forecasts using the combined training set; however, forecasts during the later hours show remarkably improved skill over previous forecasts.

Therefore, the overall results indicate that the best track results are achieved through using an Atlantic training set while the best intensity results are achieved through using the combined Atlantic–eastern Pacific training set. Either way, however, the introduction of Atlantic training appears to have increased the skill of the superensemble over solely using an eastern Pacific training set for forecasting eastern Pacific tropical cyclone tracks and intensities.

Determining whether the addition of an Atlantic training set results in statistically significant improvements in the overall superensemble track and intensity forecasts requires using a paired *t* test. Table 7 shows various paired *t* tests conducted between track and intensity results using only Pacific training and track and intensity results versus using combined training and Atlantic training. These tests were conducted using the average forecast errors at each forecast hour. Four paired *t* tests were conducted, and the results indicate that track and intensity improvements using both combined and Atlantic training are significant at the 95% confidence level, with corresponding *P* values ranging from 0.016 to 0.054. Therefore, it is likely that the improvements seen with the additional training sets are more likely systematic in nature and not random.

A question arises at this point: Why would using an Atlantic training set, as opposed to an eastern Pacific training set, result in better eastern Pacific superensemble forecasts? It is counterintuitive that superensemble forecasts for a given basin would be better because of using training sets from a different basin. It is possible that the size of the training sets caused the outcomes seen in these experiments. However, the size of the two training sets was not appreciably different. The Atlantic training set is composed of only 40 more

TABLE 6. The 2004 tropical cyclone RMS intensity errors ($m s^{-1}$) for the BCEMs and the FSSEs associated with various training sets.

Hour	BCEM (Pacific training)	BCEM (Atlantic training)	BCEM (combined training)	FSSE (Pacific training)	FSSE (Atlantic training)	FSSE (combined training)
12	2.3	2.3	2.3	2.1	2.2	2.1
24	3.8	3.7	3.7	3.4	3.5	3.3
36	4.8	4.6	4.7	4.0	4.3	4.0
48	5.7	5.1	5.3	4.3	4.8	4.3
60	6.3	5.1	5.5	4.7	4.9	4.5
72	7.4	4.9	5.6	5.5	5.0	4.9
84	9.4	5.2	6.3	7.9	4.6	6.6
96	10.4	5.6	6.8	8.6	4.1	6.8
108	12.0	5.9	7.3	9.1	3.5	7.1
120	13.4	6.6	8.2	10.0	4.3	7.4

TABLE 7. Results of using a paired *t* test to determine the significance of the differences between various paired error sets.

Compared results	Degrees of freedom	Alpha	<i>T</i> statistic	<i>P</i> value	Are differences significant?
RMS track errors (Pacific training vs combined training)	9	0.05	2.93	0.016	Yes
RMS track errors (Pacific training vs Atlantic training)	9	0.05	2.88	0.018	Yes
RMS intensity errors (Pacific training vs combined training)	9	0.05	2.79	0.020	Yes
RMS intensity errors (Pacific training vs Atlantic training)	9	0.05	2.22	0.054	Yes

training cases than is the eastern Pacific training set. It is also possible that the biases and overall performance of the models and forecasts change depending on the overall synoptic pattern in a particular basin. Unfortunately, to this date no one has performed detailed studies as to whether there are nonchaotic patterns associated with model performance at any given time. Finally, a comparison of Pacific forecasts shows that in many cases, the BCEM actually performs worse than the ENSM. This result would indicate that improper bias corrections are occurring in the majority of forecasts, and incorrect bias corrections would have a significant impact on the overall superensemble performance. Therefore, the only possible way, at this time, to answer the original question is to examine the biases of training-set models during the training and forecast periods since the weighting of the numerical models used to produce the superensemble forecast is secondary to the primary bias-correction scheme.

Tables 8–13 show the biases of all of the models outlined in Table 2 at all synoptic times. The charts are broken down to show the biases of the models in the

2002–03 eastern Pacific training set, the 2002–03 Atlantic training set, and the actual biases used in the 2004 eastern Pacific forecasts. The natural goal would be to have a training set whose biases more closely resemble the biases of the models in real time. Instead of examining each table in detail with regard to the sign and magnitude of model bias in comparison to actual 2004 model biases, Table 14 provides correlation coefficients for the early and late time model training biases with the 2004 real-time model biases at the corresponding times. For the latitude and longitude model biases, the correlation coefficients indicate a higher correlation between the Atlantic training biases and the 2004 real-time model biases in all situations except for the correlations involving early time, longitude training. These correlations show that the Atlantic training biases and the 2004 real-time model biases have a strong, direct correlation while the eastern Pacific training biases and the 2004 real-time model biases have a strong, indirect correlation. Therefore, based on correlation analysis, one would expect that the Atlantic training would tend to provide better track forecasts because, in most cases,

TABLE 8. Latitude model biases for the early times (km).

Hour	NGPI	UKMI	GUNS
Pacific training			
12	-0.22	8.21	4.88
24	-3.33	6.55	3.55
36	-5.99	-0.78	2.00
48	-4.00	-2.00	2.89
60	-5.55	-1.42	-0.67
72	-6.88	-17.32	-0.78
Atlantic training			
12	9.77	12.77	7.66
24	9.44	5.66	7.55
36	8.55	8.21	8.77
48	6.33	16.32	11.32
60	6.77	15.32	9.32
72	4.55	24.31	13.88
Real time			
12	5.11	20.65	8.99
24	5.33	15.87	8.33
36	2.55	9.32	4.11
48	4.33	10.66	3.77
60	4.00	10.55	6.33
72	2.55	14.76	8.10

TABLE 9. Longitude model biases for the early times (km).

Hour	OFCI	GFDI	NGPI	GUNA
Pacific training				
12	-2.40	-21.80	-25.51	-12.75
24	-2.29	-25.07	-28.78	-15.81
36	2.62	-22.78	-20.49	-11.12
48	5.78	-18.42	-13.84	-6.65
60	3.05	-19.18	-13.30	-2.73
72	-4.47	-36.19	-10.68	-8.83
Atlantic training				
12	10.50	11.45	5.67	9.03
24	0.32	7.98	0.21	2.63
36	-8.09	1.37	-6.20	-4.94
48	-18.17	-4.94	-9.98	-12.29
60	-19.74	-1.79	-6.83	-12.08
72	-19.53	2.42	-4.94	-7.88
Real time				
12	5.56	-5.12	-15.15	1.96
24	12.54	0.87	-8.28	6.10
36	17.00	7.09	5.78	12.21
48	15.48	6.76	10.68	15.81
60	9.16	0.98	8.83	12.32
72	7.96	-9.59	0.87	7.74

TABLE 10. Intensity model biases for the early times ($m s^{-1}$).

Hour	OFCI	GFDI	UKMI	SHF5	DSHP
Pacific training					
12	0.57	-0.62	-0.89	0.39	0.67
24	1.59	1.44	0.69	1.30	1.37
36	1.49	2.00	1.67	1.78	1.67
48	1.37	2.03	2.42	1.96	1.73
60	2.71	2.63	3.83	3.13	2.71
72	4.57	2.74	5.88	4.89	4.32
Atlantic training					
12	-0.09	-0.62	-1.17	0.44	-0.26
24	0.40	0.95	0.06	0.89	0.36
36	0.56	1.15	0.61	0.59	0.62
48	0.55	0.55	1.08	0.65	0.62
60	0.52	0.25	1.66	0.67	0.18
72	0.74	0.46	2.23	0.91	-0.12
Real time					
12	0.56	-2.03	-0.34	0.58	0.48
24	0.67	-0.43	0.40	0.41	0.49
36	0.75	0.33	1.18	0.78	0.86
48	1.16	0.29	1.96	1.11	1.14
60	2.18	0.93	3.04	2.94	2.03
72	2.70	0.84	3.48	2.86	2.31

the biases associated with that training correlate better with the biases of the real-time models. It is important to note, however, that the bias-correlation improvements may not explain the total improvement in the track errors. Subsequent coefficient weighting, resulting from training biases, may also help improve track errors when using Atlantic training.

Using correlation analysis to explain the outcomes of various intensity forecasts, however, is somewhat more problematic. At all times, the intensity biases associated with the real-time models correlate better with the intensity biases associated with the eastern Pacific training. During the early times, high correlations exist between the real-time model biases and both sets of train-

TABLE 12. Longitude model biases for the late times (km).

Hour	OFCI	UKMI	GUNA
Pacific training			
84	-8.61	-31.17	-21.69
96	-13.41	-59.95	-29.87
108	-1.09	-59.95	-41.97
120	-3.71	-69.76	-52.43
Atlantic training			
84	-12.18	-35.49	-14.18
96	-11.34	-56.49	-15.96
108	2.52	-48.62	-18.69
120	23.10	-47.25	-12.81
Real time			
84	17.00	12.21	11.12
96	37.50	29.10	21.80
108	41.97	55.05	31.61
120	21.80	134.40	58.10

ing-set biases. However, with an almost one to one correlation between the eastern Pacific training-set biases and the real-time biases, one would reasonably expect a much better intensity forecast using the eastern Pacific training set than when using the Atlantic training set (correlation of 0.595). However, even though the use of Pacific training does improve the early time intensity forecasts, the amount of improvement over the intensity forecasts when using an Atlantic training set is small, as is seen in examining Tables 5–6.

During the later times, the correlations between the biases of the two training sets and the real-time model biases are much more similar in magnitude. The eastern Pacific–real-time correlation is approximately 0.46 while the Atlantic–real-time correlation is approximately 0.35. However, even though the eastern Pacific–real-time correlation is higher during this period, the Atlantic training set significantly improves the intensity forecasts as compared to forecasts made using the east-

TABLE 11. Latitude model biases for the late times (km).

Hour	NGPI	UKMI	GUNS
Pacific training			
84	10.99	-9.21	-4.22
96	29.97	-7.77	12.88
108	44.62	-11.21	11.99
120	51.39	-39.63	4.77
Atlantic training			
84	5.77	37.96	22.42
96	25.64	23.53	30.64
108	16.87	-0.22	20.31
120	11.54	-8.21	17.32
Real time			
84	-11.77	18.54	5.77
96	-25.86	29.97	48.84
108	-23.09	26.86	-2.44
120	-19.09	-21.65	-41.29

TABLE 13. Intensity model biases at the late times.

Hour	OFCI	GFDI	UKMI	SHF5	DSHP
Pacific training					
84	7.32	8.45	8.22	7.57	7.82
96	2.63	1.34	4.77	1.46	2.61
108	1.95	0.75	4.39	3.37	1.26
120	1.02	-0.29	3.35	1.30	0.57
Atlantic training					
84	0.07	0.26	1.99	0.13	-1.84
96	-0.25	0.06	1.58	0.14	-2.07
108	-0.07	0.21	1.64	-0.29	-2.28
120	0.60	1.21	2.00	-2.28	-0.56
Real time					
84	2.20	-0.37	3.68	2.31	2.29
96	2.04	-1.02	3.25	1.97	1.80
108	2.40	-0.61	3.13	1.82	1.50
120	1.54	0.64	1.43	-1.28	-0.30

TABLE 14. Model bias correlation coefficients.

	Eastern Pacific training–eastern Pacific 2004 models (lat)	Atlantic training–eastern Pacific 2004 models (lat)	Eastern Pacific training–eastern Pacific 2004 models (lon)	Atlantic training–eastern Pacific 2004 models (lon)	Eastern Pacific training–eastern Pacific 2004 models (intensity)	Atlantic training–eastern Pacific 2004 models (intensity)
Early times	0.268	0.466	0.792	−0.611	0.876	0.594
Late times	−0.296	0.346	−0.608	−0.384	0.464	0.350

ern Pacific training set. Therefore, it is reasonable to expect that the assignment of coefficients in these cases must have had a significant impact on the forecast skill.

Because correlation analysis does not reveal any information concerning the magnitude differences among the various sets of biases, the magnitude differences were calculated for the intensity biases to determine whether the magnitudes of the bias differences help to explain why the Atlantic training provided better intensity forecasts. In many cases, the magnitude differences between the Pacific training–Pacific real time and Atlantic training–Pacific real time were not substantial. However, the GFDI Pacific real-time biases are much closer in magnitude to the GFDI Atlantic training biases than the GFDI Pacific training biases. The same trend is noted when comparing biases associated with the UKMI and SHF5, but not to the same extent as with the GFDI biases. Furthermore, a bias magnitude spike is noted at hours 72 and 84 with the Pacific training biases that is not seen in either the Atlantic training biases or the Pacific real-time biases. This bias magnitude spike may also help to explain the success of the Atlantic training set when forecasting Pacific tropical cyclone intensity.

4. Conclusions

The overall goal of this experiment was to see whether multibasin training would be beneficial to superensemble forecasts if the characteristics between the training-set models and the real-time models are similar. Such multibasin training, at least in this instance, does appear to be beneficial to superensemble forecasts. It would be interesting and intriguing to see whether multibasin training sets continue to result in improved tropical cyclone forecasts in future seasons. The most captivating part of this experiment, however, does not result from proving that multibasin training could be useful in tropical cyclone forecasting. Instead, the most important aspects of this research involve showing that, at least in this case, Atlantic training is more beneficial to eastern Pacific tropical cyclone forecasting than eastern Pacific training is. Furthermore, while similar correlations between training-set biases and real-time model biases appear to improve the fore-

casts, exceedingly similar correlations are not always necessary to produce competitive forecasts, and higher correlations do not always result in better forecasts. These observations reveal a couple of important traits of superensembles. First, the use of multibasin training does appear to help the forecasts generally because of an increase in the overall training cases. As long as the models used in both basins are the same, using multibasin training should continue to be beneficial in reducing errors in the future. Next, the nature of the model-bias calculations in this paper seems to indicate two possibilities: Coefficient assignments are not as closely linked to model-bias evaluation as the superensemble method would lead one to believe, or the model-bias calculations, coefficient assignments, or both are the result of an attempt to implement order in a naturally chaotic system. One method to evaluate whether the latter idea is valid is to examine model bias as it relates to tropical cyclones. The key questions to answer involve whether model biases are consistent from cyclone to cyclone, from day to day, and from forecast to forecast.

Future studies should attempt to clarify the nature of biases in numerical models as they relate to tropical cyclone track and intensity. On a smaller scale, revisiting the idea of separating the training based on the location in the Pacific could result in better forecasts since the Pacific does not contain as many subsections as the Atlantic does, thereby allowing for more training instances in each training set. Also, it is reasonable to assume that model biases could be completely different for tropical cyclones near the Mexican coast, where terrain issues are present, as opposed to farther west in the open ocean, where terrain issues are not a factor.

REFERENCES

- Krishnamurti, T. N., C. M. Kishtawal, T. LaRow, D. Bachiochi, Z. Zhang, C. E. Williford, S. Gadgil, and S. Surendran, 1999: Improved skills for weather and seasonal climate forecasts from multimodel superensemble. *Science*, **285**, 1548–1550.
- Rhime, J., cited 2007: Technical summary of the NHC/TPC tropical cyclone track and intensity guidance models. [Available online at <http://www.nhc.noaa.gov/modelsummary.shtml>.]
- Williford, C. E., 2002: Real-time superensemble tropical cyclone prediction. Ph.D. dissertation, The Florida State University, Tallahassee, FL, 144 pp.