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, FL

Corresponding author address:

Carol Anne Clayson

Department of Meteorology

and

Geophysical Fluid Dynamics Institute

404 Love Building

Florida State University

Tallahassee, FL 32306

Phone: 850-644-0922

Email: clayson@met.fsu.edu

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 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 does help 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 a more detailed explanation of how a superensemble forecast is developed, refer to Williford (2002).

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. Since the same numerical models are used for tropical cyclone forecasts in both basins, compatibility is not an issue since the same versions of the models are applied to each basin. Based on the theories outlined above, such an experiment would be expected to lead to improved tropical cyclone forecasts.

2. Methods

Numerical models chosen in developing superensemble forecasts for the 2004 Eastern Pacific tropical cyclone season are outlined in Table 1. Detailed information about these numerical models can be found at www.nhc.noaa.gov/modelsummary.shmtl. The experiments presented in this paper were conducted as if real-time hurricane track and intensity 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 1 represents the collection of models which were determined to provide the best forecasts. No cross-validation experiments were conducted in any experiments since 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 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 hour forecasts while late models were used for 84-120 hour forecasts. 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 numerical models between 2002 and 2004. Therefore, using a 2002/2003 training set for forecasting cyclones in 2004 would be appropriate.

3. Results

The first experiment conducted involved using 2002/2003 Eastern Pacific training to forecast 2004 Eastern Pacific tropical cyclones. The previous forecast training set comprised of 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. Figure 1 shows the root-mean-square track and intensity errors, respectively, for the superensemble along with other model and forecast errors for the 2004 Eastern Pacific hurricane season. Figure 1 indicates that the superensemble track forecasts perform relatively well compared to other models for 12-48 hour forecasts. During this period, superensemble performance only slightly lags behind the performance of the ensemble models GUNS and GUNA. During the 60-120 hour forecasts, however, superensemble forecast skill falls significantly compared to the skill 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 skill pattern is also noted when examining the intensity errors of the superensemble and other models and forecasts. For 12-60 hour forecasts, the superensemble performed well, alternating between the best intensity model and the second-best intensity model. However, 72-120 hour forecasts reflect a significant drop in skill for the superensemble, as it trails several of its member models during these times.

In order to assess the effects of an increased training set on the superensemble forecast, a second experiment was conducted involving a combined 2002/2003 Eastern Pacific training set and a 2002/2003 Atlantic training set. This training set included 811 forecast instances. Figure 2 shows root-mean-square track and intensity errors for the superensemble and other numerical models. Figure 2 shows that 12-60 hour 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 hour 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 training can be helpful if the models used for both basins are the same.

Since the addition of an Atlantic training set improved superensemble forecasts, a third experiment was conducted to see whether the additional training set (the 2002/2003 Atlantic training set) alone would result in better forecasts than the training sets used in the two previous experiments. The 2002/2003 Atlantic training set included 429 previous forecast instances. Figure 3 shows the results of using this training set to forecast Eastern Pacific tropical cyclone track and intensity, respectively. The track results, as shown in Figure 6, indicate that 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 actually a bit worse than forecasts using the combined training set; however, forecasts during the later hours show remarkably improved skill over previous forecasts. Therefore, 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 track and intensity.

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? Intuitively, it does not make logical sense that forecasts for a given basin would be better because a forecaster uses training 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 comprised of only 40 more training cases than 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 non-chaotic patterns associated with model performance at any given time. Finally, a comparison of the seasonal 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 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.

Figures 4-9 show the biases of all models outlined in Table 1 at all synoptic times. The charts are broken down to show the biases of the models in the 2002/2003 Eastern Pacific training set, the 2002/2003 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 figure in detail with regards to the sign and magnitude of model bias in comparison to actual 2004 model biases, Table 2 provides correlation coefficients for early and late-time model training biases with the 2004 real-time model biases at corresponding times. For 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, indirect correlation while the Eastern Pacific training biases and the 2004 real-time model biases have a strong, direct correlation. Therefore, based on correlation analysis, one would expect that the Atlantic training would tend to provide better track forecasts since, in most cases, the biases associated with that training correlates better with the biases of the real-time models.

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 training-set biases. However, with an almost one to one correlation between Eastern Pacific training-set biases and real-time biases, one would reasonably expect a much better intensity forecast using Eastern Pacific training than when using Atlantic training (correlation of 0.595). However, even though the use of Pacific training does improve early-time intensity forecasts, the amount of improvement over intensity forecasts when using an Atlantic training set is small, as seen in comparing Figures 1b and 3b.

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 intensity forecasts as compared to forecasts made using the Eastern Pacific training set. Therefore, it is reasonable to expect that the assignment of coefficients in these cases must have had a significant impact on forecast skill. Unfortunately, it is impossible to explain, at this time, why such results would occur since the superensemble method is designed to evaluate model bias and then assign coefficients to individual models based on overall skill.

4. Conclusions

The overall goal of this experiment was to see whether multi-basin training would be beneficial to superensemble forecasts if the characteristics between the training set models and the real-time models are similar. Such multi-basin training, at least in this instance, does appear to be beneficial to superensemble forecasts. It would be interesting and intriguing to see whether multi-basin 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 multi-basin 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 forecasts, 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 superensemble. First, the use of multi-basin training does appear to help forecasts generally because of an increase in overall training cases. As long as the models used in both basins are the same, using multi-basin 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 storm to storm. If such a study indicates that tropical cyclone model biases are not quantifiable on these scales, the validity of the superensemble method could be called into question.

REFERENCES

Demaria, Mark. "Summary of the NHC/TPC Tropical Cyclone Track and Intensity Guidance Models." 13 July 2005. <http://www.nhc.noaa.gov/modelsummary.shtml>.

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.

Williford, C. E., 2002: Real-time Superensemble Tropical Cyclone Prediction. Ph.D. Dissertation, The Florida State University, Tallahassee, FL 32306, 144 pp.

FIGURE CAPTIONS

Figure 1: 2004 RMS Tropical Cyclone (a) Track and (b) Intensity Errors (Pacific Training Set).

Figure 2: 2004 RMS Tropical Cyclone (a) Track and (b) Intensity Errors (Combined Training Set).

Figure 3: 2004 RMS Tropical Cyclone (a) Track and (b) Intensity Errors (Atlantic Training Set).

Figure 4: Latitude Model Biases for (a) Pacific Training, (b) Atlantic Training, and (c) Real-time Models at Early Times,

Figure 5: Longitude Model Biases for (a) Pacific Training, (b) Atlantic Training, and (c) Real-time Models at Early Times,

Figure 6: Intensity Model Biases for (a) Pacific Training, (b) Atlantic Training, and (c) Real-time Models at Early Times.

Figure 7: Latitude Model Biases for (a) Pacific Training, (b) Atlantic Training, and (c) Real-time Models at Late Times.

Figure 8: Longitude Model Biases for (a) Pacific Training, (b) Atlantic Training, and (c) Real-time Models at Late Times.

Figure 9: Intensity Model Biases for (a) Pacific Training, (b) Atlantic Training, and (c) Real-time Models at Late Times.

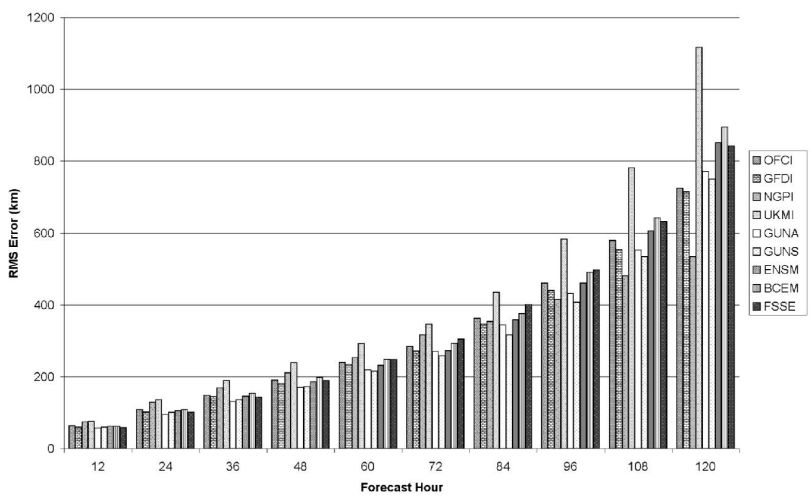
Table 1: Numerical Models Used in Producing Superensemble Forecasts for the 2004 Eastern Pacific Tropical Cyclone Season.

Early Latitude	Late Latitude	Early Longitude	Late Longitude	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

	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 (Int.)	Atlantic training/ Eastern Pacific 2004 models (Int..)
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

Table 2: Model Bias Correlation Coefficients

a



b

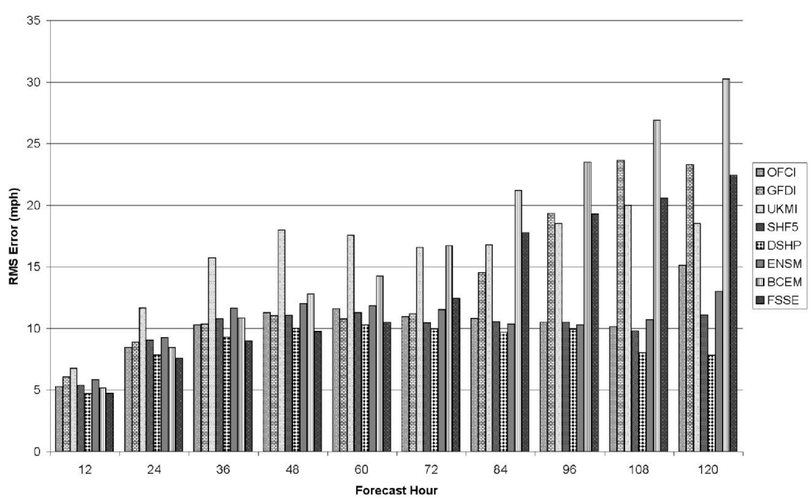
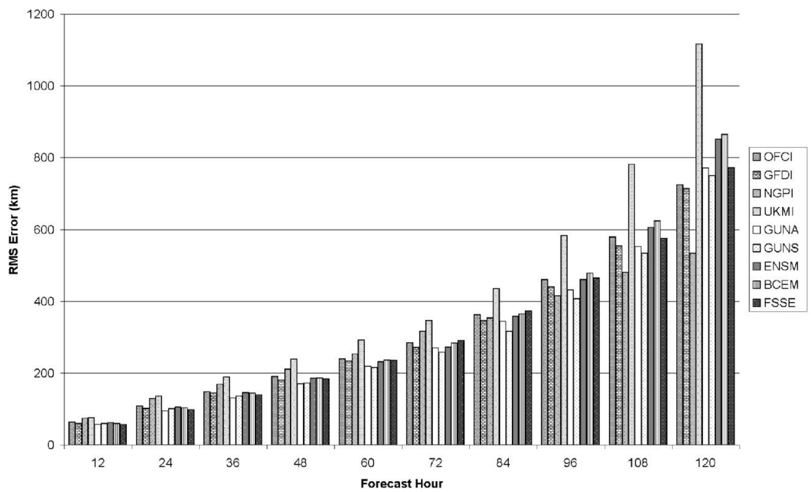


Figure 1: 2004 RMS Tropical Cyclone (a) Track and (b) Intensity Errors (Pacific Training Set)

a



b

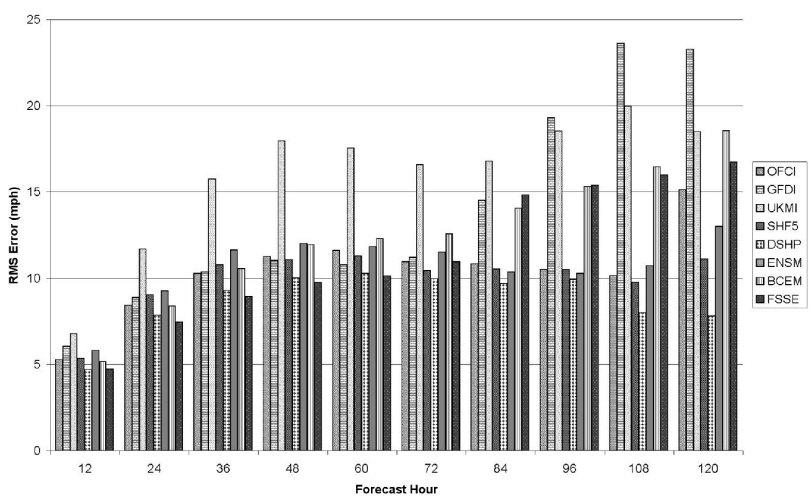
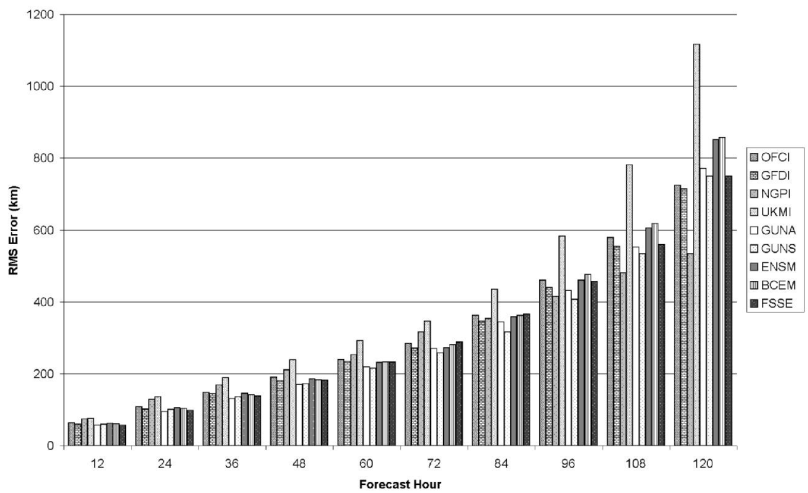


Figure 2: 2004 RMS Tropical Cyclone (a) Track and (b) Intensity Errors (Combined Training Set)

a



b

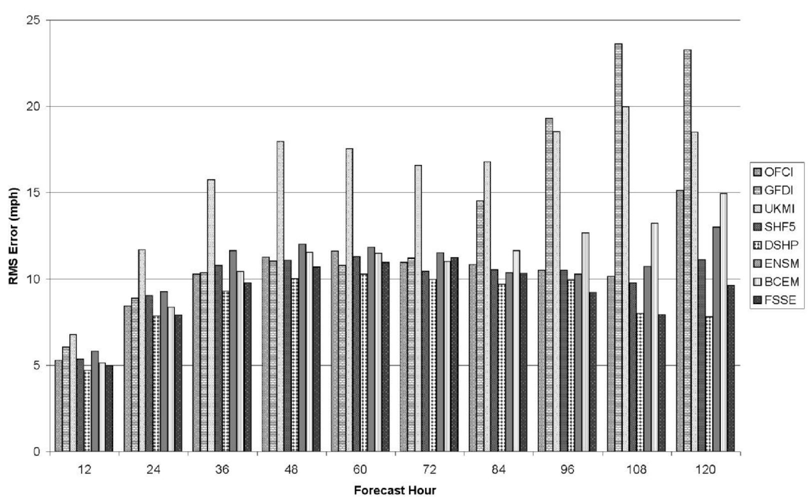


Figure 3: 2004 RMS Tropical Cyclone (a) Track and (b) Intensity Errors (Atlantic Training Set)

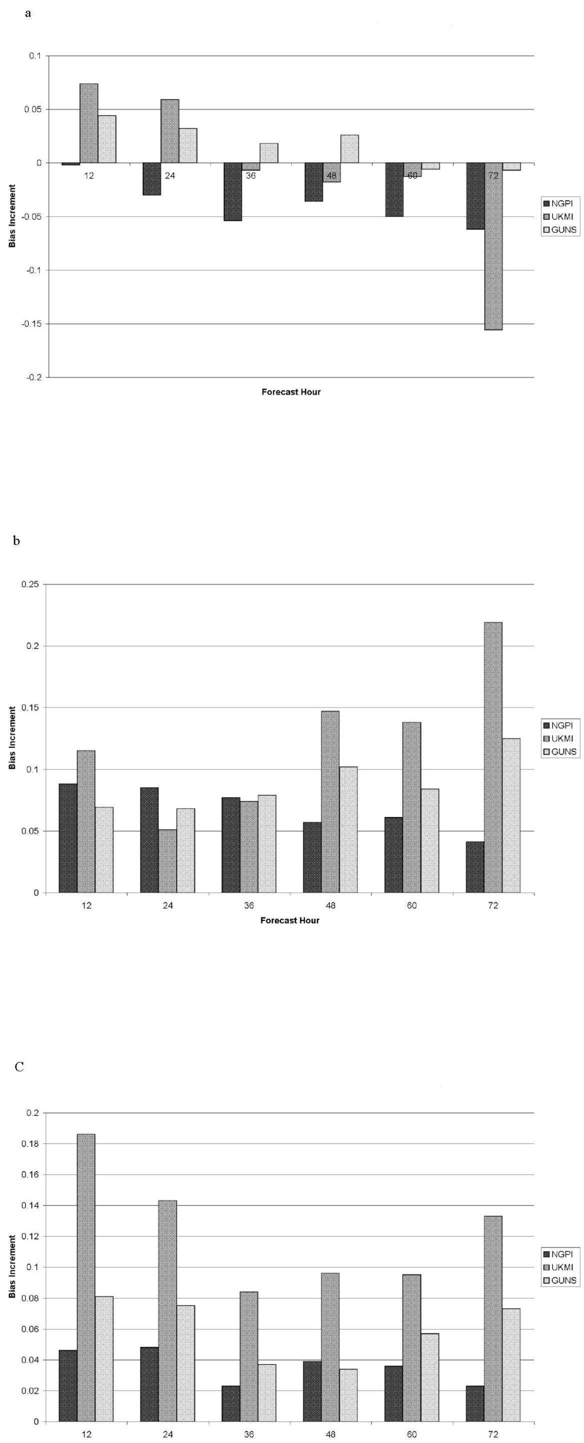
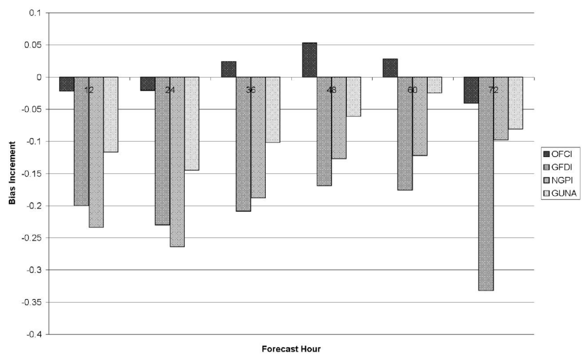
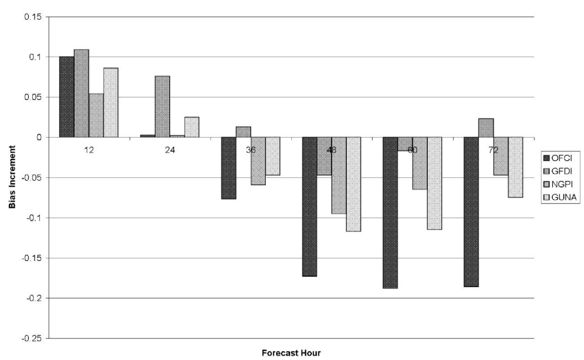


Figure 4: Latitude Model Biases for (a) Pacific Training, (b) Atlantic Training, and (c) Real-time Models at Early Times

a



b



C

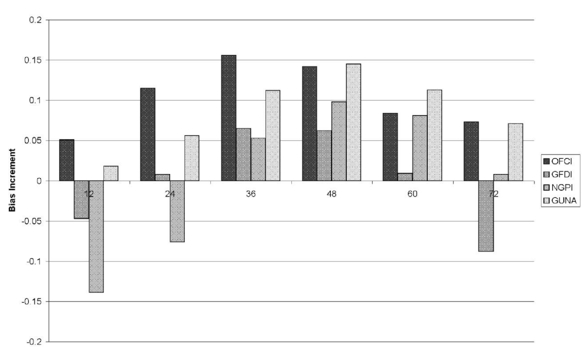


Figure 5: Longitude Model Biases for (a) Pacific Training, (b) Atlantic Training, and (c) Real-time Models at Early Times

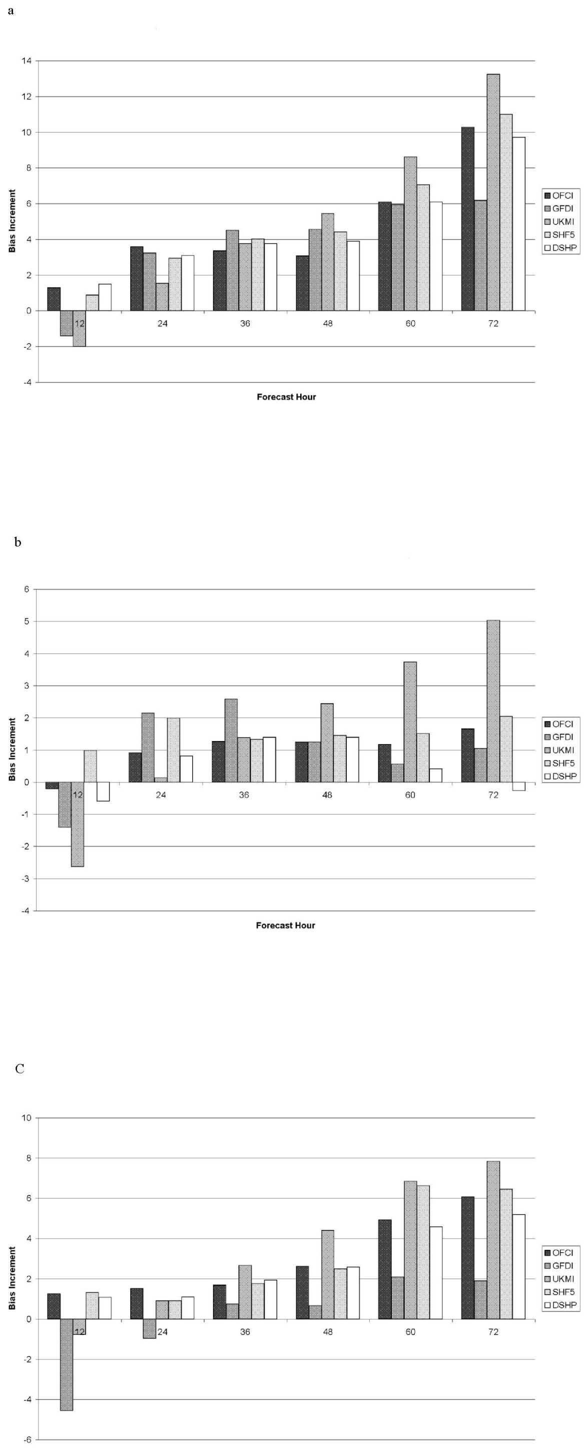
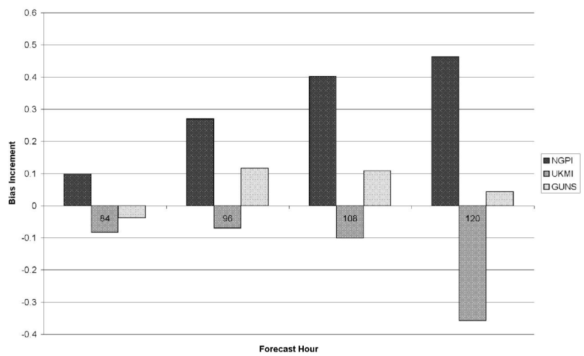
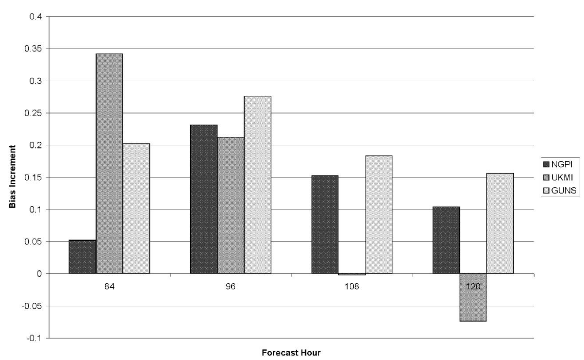


Figure 6: Intensity Model Biases for (a) Pacific Training, (b) Atlantic Training, and (c) Real-time Models at Early Times

a



b



C

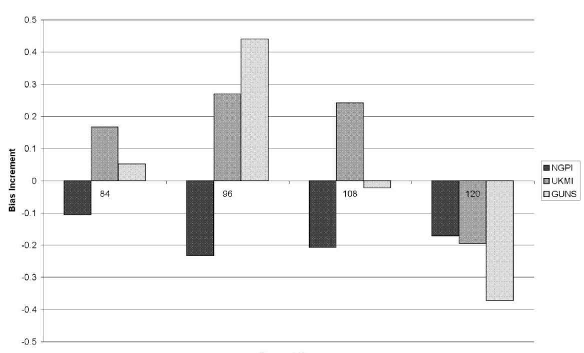
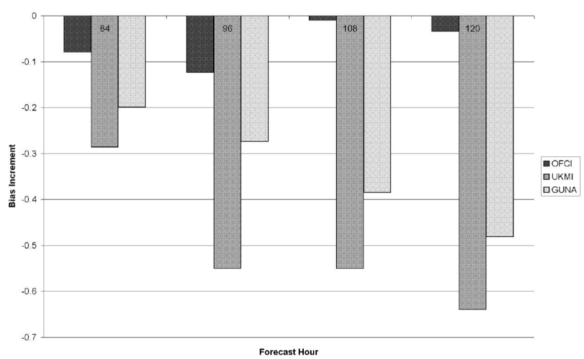
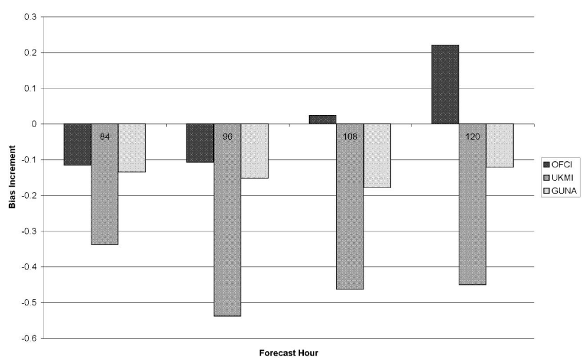


Figure 7: Latitude Model Biases for (a) Pacific Training, (b) Atlantic Training, and (c) Real-time Models at Late Times

a



b



C

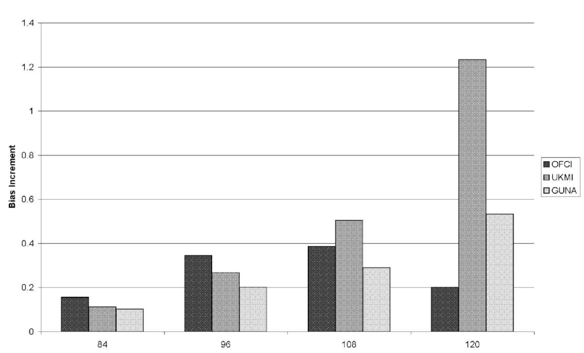
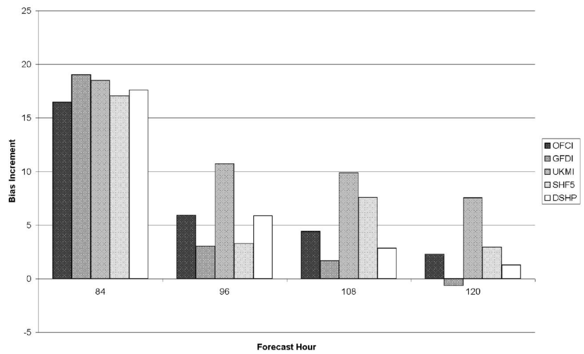
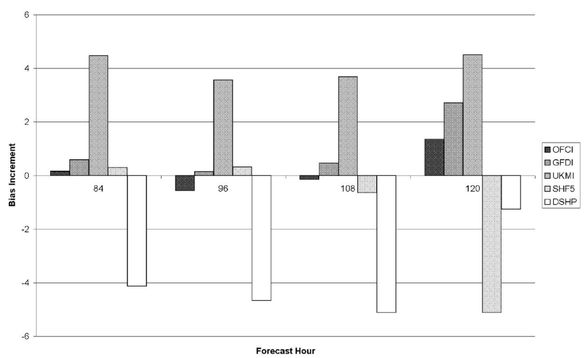


Figure 8: Longitude Model Biases for (a) Pacific Training, (b) Atlantic Training, and (c) Real-time Models at Late Times

a



b



C

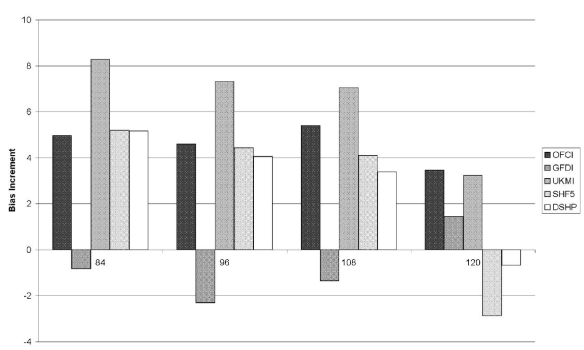


Figure 9: Intensity Model Biases for (a) Pacific Training, (b) Atlantic Training, and (c) Real-time Models at Late Times