NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

A WEIGHTED CONSENSUS APPROACH TO TROPICAL CYCLONE 96-H AND 120-H TRACK FORECASTING

by

James R. Hughes

March 2007

Thesis Advisor:	Russell L. Elsberry
Second Reader:	Mark A. Boothe

Approved for public release; distribution is unlimited

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			*Form Approved OMB No. 0704-0188*
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington DC 20503.			
1. AGENCY USE ONLY *(Leave blank)*	**2. REPORT DATE** March 2007	**3. REPORT TYPE AND DATES COVERED** Master's Thesis	
4. TITLE AND SUBTITLE A Weighted Consensus Approach to Tropical Cyclone 96-h and 120-h Track Forecasting			**5. FUNDING NUMBERS**
6. AUTHOR(S) James R. Hughes			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000			**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A			**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**
11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.			
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited			**12b. DISTRIBUTION CODE** A

13. ABSTRACT (maximum 200 words)

A long-range (96 h – 120 h) weighted position consensus for tropical cyclone tracks is evaluated for 24 western North Pacific storms in 2006. The first weighted position technique simply weights the 96-h, 108-h, and 120-h dynamical model positions inversely to their distances from the 60-h, 66-h, and 72-h consensus positions. The second weighted consensus technique uses the same weighting factors but is applied to the forecast motion vectors to assess 96 h – 120 h track errors.

The weighted position consensus yields modest reductions in error relative to an unweighted position consensus at 96 h – 120 h and produces smoother track forecasts. Weighted position consensus errors are reduced when the COAMPS model and the Air Force Weather Agency MM5 model are removed from the unweighted consensus used to form the weighting factors. Including the Japan and ECMWF model tracks also improves the weighted position consensus performance. The weighted motion vector consensus achieves dramatic improvements over an unweighted position consensus (9.9% at 96 h and 5.6% at 120 h). Most of the improvement over an unweighted position consensus is from using a motion vector consensus rather than a position consensus since large improvements are also achieved with an unweighted motion vector consensus.

14. SUBJECT TERMS Numerical Weather Prediction; Tropical Meteorology; Tropical Cyclone Track; Tropical Cyclone Prediction; Consensus Forecasting			**15. NUMBER OF PAGES** 107
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT** UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

A WEIGHTED CONSENSUS APPROACH FOR TROPICAL CYCLONE 96-H AND 120-H TRACK PREDICTION

James R. Hughes
First Lieutenant, United States Air Force
B.S., University of Wisconsin-Madison, 2003

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN METEOROLOGY

from the

NAVAL POSTGRADUATE SCHOOL
March 2007

Author: James Robert Hughes

Approved by: Dr. Russell L. Elsberry
Thesis Advisor

Mark A. Boothe
Second Reader

Dr. Philip A. Durkee
Chairman, Department of Meteorology

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

A long-range (96 h – 120 h) weighted position consensus for tropical cyclone tracks is evaluated for 24 western North Pacific storms in 2006. The first weighted position technique simply weights the 96-h, 108-h, and 120-h dynamical model positions inversely to their distances from the 60-h, 66-h, and 72-h consensus positions. The second weighted consensus technique uses the same weighting factors but is applied to the forecast motion vectors to assess 96 h – 120 h track errors.

The weighted position consensus yields modest reductions in error relative to an unweighted position consensus at 96 h – 120 h and produces smoother track forecasts. Weighted position consensus errors are reduced when the COAMPS model and the Air Force Weather Agency MM5 model are removed from the unweighted consensus used to form the weighting factors. Including the Japan and ECMWF model tracks also improves the weighted position consensus performance. The weighted motion vector consensus achieves dramatic improvements over an unweighted position consensus (9.9% at 96 h and 5.6% at 120 h). Most of the improvement over an unweighted position consensus is from using a motion vector consensus rather than a position consensus since large improvements are also achieved with an unweighted motion vector consensus.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

First, I thank my advisor Dr. Elsberry for developing an interesting thesis topic and helping me along the way with his guidance and timely feedback. Mark Boothe also deserves many thanks for his guidance, assistance with data, and for producing excellent schematics for my presentation. I thank Buck Sampson of the Naval Research Lab - Monterey for answering many questions and providing a table used in the thesis. Additionally, I acknowledge the Joint Typhoon Warning Center for providing data for this study. I thank Mike Fiorino for providing documentation on his vortex tracker. I also thank LCDR Sanabia for sharing her work with the weighted consensus technique to help me get started with my research.

I kindly thank my wife who has been helpful throughout my thesis work as a source of support. I highly value her feedback and advice.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

A. MOTIVATION

The United States Department of Defense has a wealth of assets in the western North Pacific that are vulnerable to the damaging winds, heavy rain, and storm surges associated with tropical cyclones (TCs). As examples of these assets, the major United States Air Force (USAF) installations are displayed in Figure 1.1. Aircraft, personnel, and infrastructure are threatened by TC activity, e.g., the 2005 typhoon tracks in Figure 1.2 give an idea of the TC activity in the western North Pacific. It is clear that all of the major USAF installations are within the realm of TC influence. Because of the potential for damaging impacts, it is important to know when and where TCs will occur. Because of the large hazard posed by TCs, it is thus important to predict, as well as possible, their movement and strength. Military commanders require sufficient notification of hazardous weather to secure assets, plan military operations, move ships etc. Thus, long-range (96 h, 108 h, and 120 h) TC forecasts are important for military planning. Numerical weather prediction (NWP) model tracks provide a tool for forecasting these parameters. A consensus of numerical model tracks is used as a basis for TC track prediction. A weighted consensus for long-range (96 h, 108 h, 120 h) TC forecasts is used in this study that will hopefully lead to improved guidance for the Joint Typhoon Warning Center (JTWC) in forecasting TCs threatening the installations depicted in Figure 1.1.

Supertyphoon Pongsona provides an example of typhoon impact as well as an example of the need to improve forecast skill by JTWC that may result from superior NWP guidance. Information concerning Supertyphoon Pongsona is taken from the National Weather Service (NWS) service assessment available online at http://www.weather.gov/os/assessments/pdfs/Pongsona.pdf. In December 2002, Supertyphoon Pongsona became the third strongest typhoon to ever strike the island of Guam. The eye of the storm passed over Andersen AFB in northern Guam. Although the base was spared from the worst of the damage, the peak winds were around 150 kt and the rain accumulation exceeded 19 inches. The eyewall passed over the heavily populated region farther south and brought rainfall accumulations over 23 inches and winds exceeding 160 kt. Preliminary damage estimates at the time of the report surpassed $700 million.

The JTWC provides the guidance for the local National Weather Service office in Guam. The JTWC track forecast was too far east and generally under-estimated the intensity of Supertyphoon Pongsona by about 25 kt, which resulted in less than desirable notification to the public of the impending supertyphoon. Part of the warning problem could be attributed to the reluctance of JTWC to change their track forecasts even at the prompting of the local NWS office.

It is proposed that better track guidance from a weighted consensus technique may aid the JTWC in predicting the landfall of a supertyphoon. A more accurate long-range track forecast from the weighted consensus technique may have given forecasters a heads-up that the typhoon would not recurve as soon as indicated by the unweighted consensus track forecast.

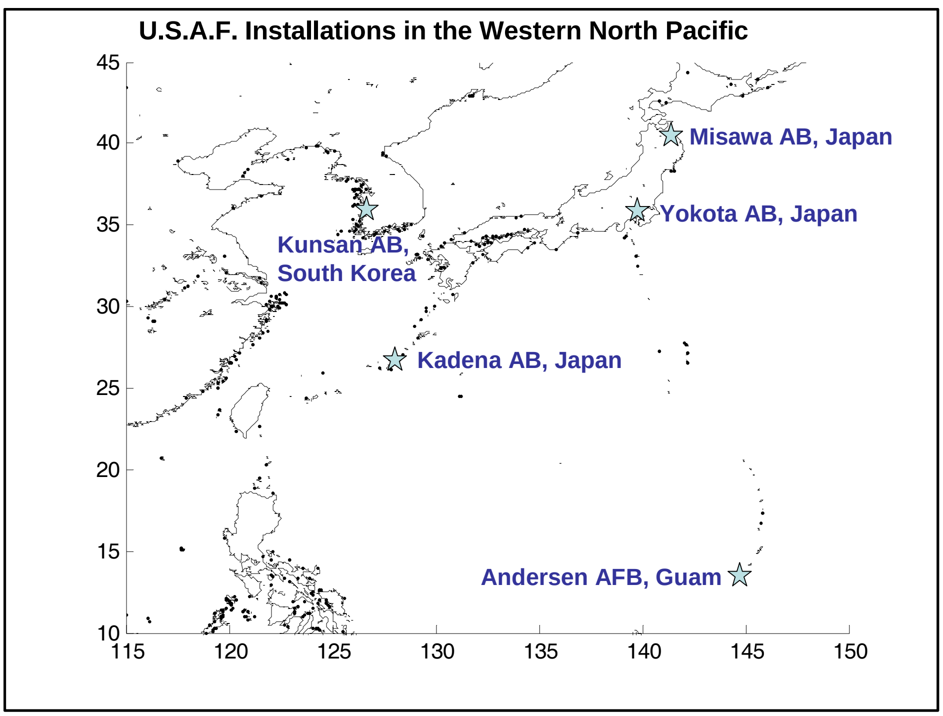


Figure 1.1	Major USAF installations in the western North Pacific.

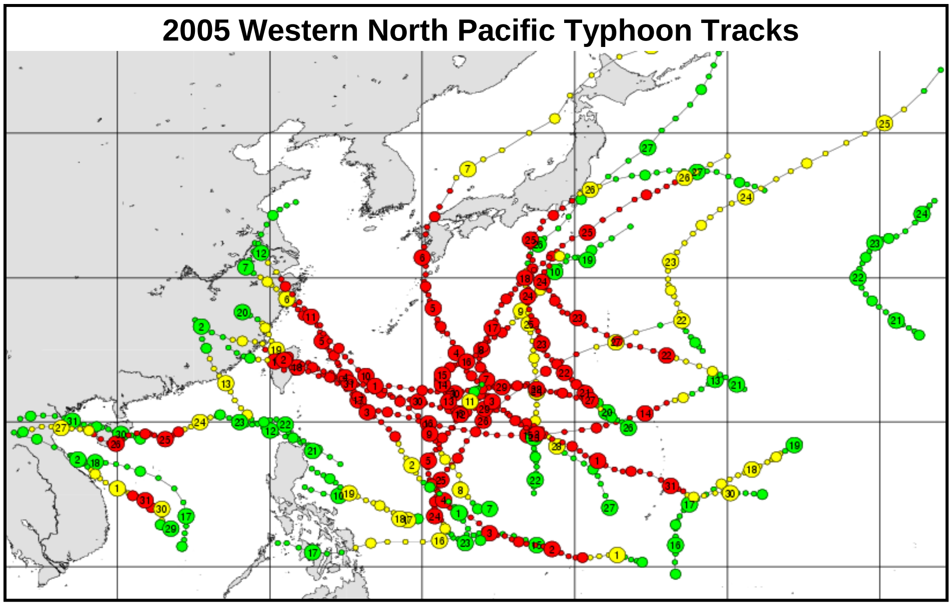


Figure 1.2 Best tracks for the 2005 North Pacific typhoons [Available online at http://agora.ex.nii.ac.jp/cgi-bin/dt/track_view.pl?basin=wnp&t=0&b=14&lang=en&type=1&size=128&tnum=23&max=99&ids=200501:200502:200503:200504:200505:200506:200507:200508:200509:200510:200511:200512:200513:200514:200515:200516:200517:200518:200519:200520:200521:200522:200523 (current as of 1 Mar 2007)].

B. BACKGROUND

Goerss (2000) demonstrated the advantage of using a consensus of dynamical model tracks. The error using a consensus of model tracks is reduced from the individual model errors averaged over a sufficiently large sample. Additionally, the consensus spread gives a measure of forecast uncertainty. Consensus forecasts were introduced at JTWC in the early 1990's, but they were not used consistently until the late 1990's (Jeffries and Fukada 2002). Carr and Elsberry (2001) introduced the Systematic Approach to TC Forecasting Aid (SAFA) to JTWC in 2000, which highlights the importance of using a consensus for TC track forecasting (Jeffries and Fukada 2002). Although the selective consensus in SAFA did not add much value over a non-selective consensus, the introduction of SAFA led JTWC to use the consensus of numerical model tracks as a basis for TC track prediction, which has contributed to their improvement in track forecast accuracy (Jeffries and Fukada 2002). Previously, the skill of a tropical cyclone track forecast has been measured by comparison with a climatology and persistence (CLIPER) forecast, which required no meteorological understanding. The consensus is arguably a no-skill measure for forecasters at JTWC since it is a simple average of the selected model track positions at each forecast time. Forecasters must improve on this consensus to add value to a TC track forecast.

Numerical models have biases, and performance varies from model to model, which suggests that all models should not be given the same weighting when a consensus is formed. If one could determine which models have greater skill, those models would receive greater weightings than less skillful models. A weighted consensus could improve on a simple unweighted consensus.

Kumar et al. (2003) developed a multi-model superensemble for TC track and intensity prediction. The superensemble has a training phase in which the model performances are measured through a regression of the model track forecasts against the best-track positions. The second phase is the forecast phase in which the weights determined in the training phases are applied to model tracks to form a superensemble. Weights are determined every 12 h for each model from the initial time to a possible 6 days. Kumar et al. (2003) showed the superensemble applied to Pacific TC tracks was able to improve on individual model forecasts and an unweighted ensemble mean (consensus) forecast, especially for forecasts 3 days and longer.

Weber (2003) developed a statistical ensemble prediction system (STEPS) to produce weighted consensus track forecasts and probabilistic strike distributions. Analogous to the training period in Kumar et al. (2003), Weber determines the STEPS weights during an initialization period based on the performance of the models during the previous season as measured by selected storm parameters. Subsequently, STEPS is applied during a forecast period. The STEPS track predictions generally outperform the individual numerical model guidance and the National Hurricane Center forecasts, but performed comparably to a simple unweighted ensemble of selected models.

It would be desirable to produce a weighted consensus that is not reliant on an extended training period to statistically derive the weighting factors as in both the Kumar et al. (2003) and Weber (2003) studies. Numerical models are occasionally updated to implement improvements, which may invalidate the statistical weighting factors that were derived before the model changes. Also, statistical weighting techniques may suffer if model performance changes due to intraseasonal or interannual variability.

The Australian Government Bureau of Meteorology (BoM) began using a motion vector consensus in the 2005-2006 Southern Hemisphere TC season (Burton 2006). The implementation of a motion vector consensus was motivated by the erratic track changes that result from a reduction in the number of model tracks in the consensus with increasing forecast times. All model tracks are first translated to a common initial position. An average of the model motion vectors over an increment of time is then added to the initial position to determine the next consensus position; this process is repeated through the rest of the forecast interval. For example, the model track guidance in Figure 1.3a is quite disparate, which results in large jumps in the position consensus track positions at longer forecast times when the number of models is reduced (Figure 1.3b). A smoother TC track forecast results from the use of a motion vector consensus (Figure 1.3c) rather than a position consensus (Figure 1.3b) since the disparate track positions at longer lead times do not affect the motion vector consensus. Rather only a combination of the incremental motion vectors contributes to the consensus positions.

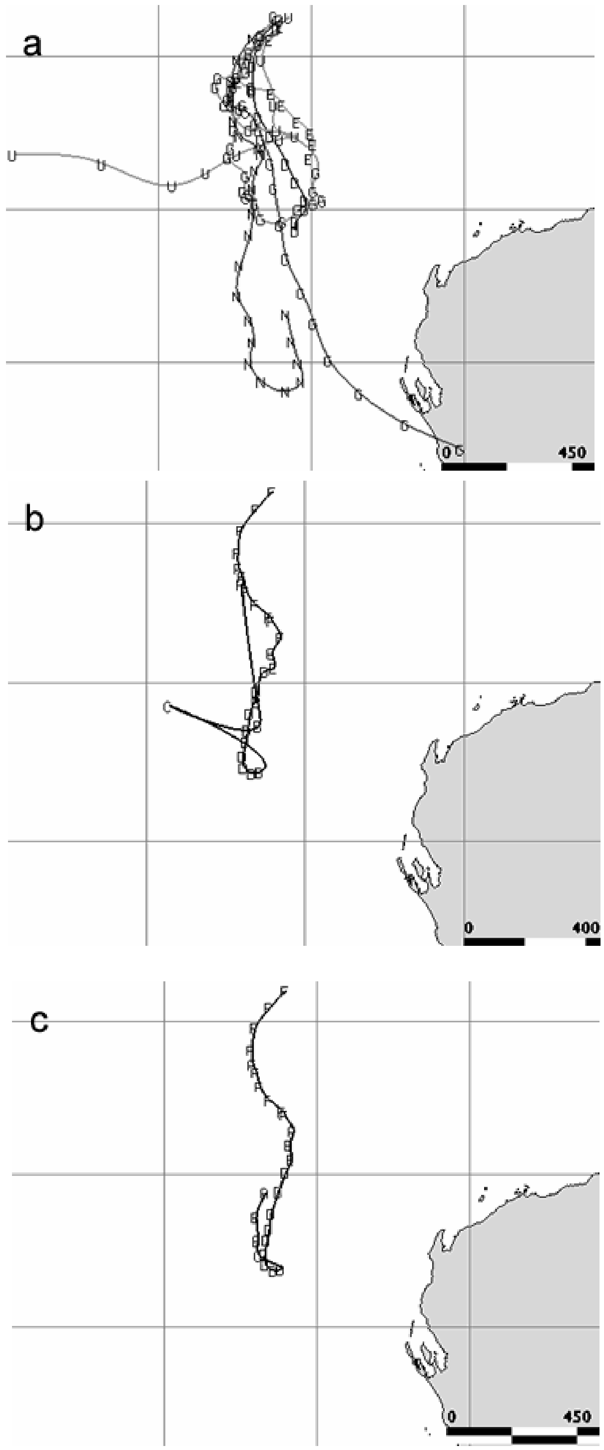


Figure 1.3 Consensus forecast tracks generated from the tracks shown in (a) (with initial position correction applied) using (b) average geographical positions and (c) average vector motions (From Burton 2006).

In addition to smoother track forecasts, the motion vector consensus could yield smaller errors for long-range TC forecasts. Large errors might be associated with erratic track changes due to a position consensus with varying numbers of models with time as in Figure 1.3b, whereas a smoother track forecast that is more consistent with the consensus before the number of models in the consensus is reduced potentially could reduce track errors.

C. OBJECTIVES OF THESIS

As introduced in the previous section, a potential exists to improve over a simple unweighted position consensus by weighting the consensus members. This study tests a simple weighted position consensus method that does not require a training period to develop statistical weights. Instead, only the model track positions are needed for this proposed weighted consensus technique. The weighted consensus technique is proposed to improve long-range track forecasts through weighting the consensus members according to their consistency with the consensus track at earlier forecast times. The justification for basing the weighting on the consensus positions at earlier forecast times is that up to 11 models are available in the consensus at JTWC from 60 h to 72 h, but only up to 6 (5) model tracks are available for the consensus at 96 h (108 h and 120 h). The additional guidance available at 60 h – 72 h relative to later times should result in a more skillful consensus at 60 h – 72 h than at 96 h – 120 h. Thus, those members that remain at 96 h – 120 h will be weighted inversely proportional to their distance from the consensus at earlier forecast times (60 h – 72 h). The hypothesis is that the 96 h – 120 h track errors will be reduced since the 96 – 120 h model track positions are weighted according to an earlier, generally more skillful, consensus track. Additionally, a consensus weighted toward a 60 – 72 h consensus should result in a smoother track through 120 h as models begin to drop out of the consensus, since those remaining that were closest to the earlier consensus receive the largest weighting factors.

Thus, the first objective of the thesis is to evaluate the proposed weighted position consensus technique to determine if it reduces long-range track errors and produces smoother long-range track forecasts.

Burton (2006) demonstrated the potential improvement in track forecasts from using a motion vector consensus instead of a position consensus. Although the technique clearly results in smoother track forecasts, Burton (2006) does not indicate how much it reduced the track errors. Since the reduction in the number of model tracks used in the JTWC consensus generally occurs after 72 h, the motion vector consensus will be evaluated at 96 h, 108 h, and 120 h to determine its performance relative to a simple unweighted position consensus. The weighted motion vector consensus will be implemented in a similar fashion as for the weighted position consensus with the weighting factors applied to motion vectors instead of track positions for each model track. Thus, the second objective of the thesis is to evaluate a weighted motion vector consensus to determine if it adds value over a weighted position consensus in terms of reduced track errors and smoother long-range track forecasts.

D. OVERVIEW

Chapter II describes the methodology for the weighted position consensus including the data used for the study, the weighting procedure, and the case selection criteria. Presented in Chapter III are results from a first validation study of the weighted position consensus followed by case studies to demonstrate the performance of the technique. Sensitivity studies are then presented to evaluate modifications to the first validation study. Chapter IV introduces the weighted motion vector methodology, and describes the results, and some case studies to demonstrate the weighted motion vector performance. Conclusions and recommendations are presented in Chapter V.

II. METHODOLOGY FOR WEIGHTED POSITION CONSENSUS

A. DATA

The western North Pacific (WPAC) was chosen for this study because it is a primary area of responsibility of the Joint Typhoon Warning Center (JTWC). The 2006 tropical cyclone season was selected to evaluate the weighting scheme as a timely application to current operations at JTWC. That is, applying the weighting technique with 2006 cases ensures that the most recent model configurations are used. However, the weighting technique should be applicable regardless of which models are used and the biases of those models.

Sampson and Schrader (2000) describe the Automated Tropical Cyclone Forecasting (ATCF) system that is used at JTWC to display graphics to aid in forecasting, track TCs, transmit warnings, etc. Model track positions, consensus track positions, and best-track (BT) positions for this study were extracted from the ATCF data. The so-called interpolated tracks are used for this study because that is the track guidance that is used operationally at JTWC. Interpolated tracks must be used at JTWC because the dynamical model output only becomes available 6 or 12 hours after run time (Goerss et al. 2003). Thus, the model track position at 6 h is compared with the current warning position, and the entire track is translated so that the 6-h forecast position agrees with the initial position as determined by the JTWC forecaster. The first 24 storms of 2006 are examined for this study since the ATCF data were available for these storms at the inception of the study.

B. WEIGHTING PROCEDURE

1. Assumptions

Initially it is assumed only four models are available after 72 h: Navy Operational Global Atmospheric Prediction System (NOGAPS), National Centers for Environmental Prediction (NCEP) Global Forecast System (GFS), the U.K. Meteorological Office (UKMO) global model, and the Geophysical Fluid

Dynamics Laboratory – Navy model (GFDN). These four models are used for the weighting scheme. Another assumption is that the operational consensus (CONW) reflects the average of the available model guidance at each forecast time. It was discovered that both of these assumptions are not always valid. Studying deviations from these assumptions provide insights that will be addressed in the sensitivity studies in Chapter III.C.

2. Precedence of Model Guidance

When the primary interpolated tracks are missing, JTWC can use tracks derived from their in-house tracker called the Enhanced FNMOC TC Tracking Scheme (hereafter Fiorino tracker) (Mike Fiorino, personal communication, February 2007). The Fiorino tracker improves on the original FNMOC tracker by using 850 mb vorticity when the surface wind tracker fails. The model fields are needed to run the Fiorino tracker, so these back-up Fiorino tracks are only available for those models that JTWC receives GRIB data from: NOGAPS, GFS, UKMO, and the Japanese Meteorological Agency global spectral model JGSM (Ryan Kehoe, personal communication, February 2007). In the case that the primary and Fiorino tracks are both unavailable, JTWC can interpolate from a 12-h old track as well as a 12-h old Fiorino track. An order of precedence (Table 2.1) is established to mimic operations at JTWC as closely as possible and to maintain consistency throughout the analysis. When the primary or other track is no longer available, the next track in the order of precedence in Table 2.1 is used.

Table 2.1 Order of precedence of model track guidance

Precedence	Track Description	Example (NOGAPS)
I	Primary interpolated track	NGPI
II	Fiorino interpolated track	JNGI
III	12-h old interpolated track	NGP2
IV	12-h old Fiorino interpolated track	JNG2

This technique can lead to some inconsistencies when applying the weighting technique. For example, if the primary track is only available through 72 h, but the Fiorino track is available through 120 h, the primary track will be used for the weighting while the later Fiorino track positions will be used to apply the weighted consensus. A significant inconsistency will only occur when a large difference exists between the primary and Fiorino tracks. The decision was made to use, whenever possible, the guidance used at the time the forecast was made. Although this technique may not lead to the clearest interpretation of the results, it is aimed at replicating the operational application.

3. Weighting Technique

a. Times Used for Weighting

Three times (60 h, 66 h, and 72 h) are chosen to calculate the weighting factors. If a model track is close to the consensus track at 60 h, 66 h, and 72 h, the implication is that the model 72-h position is not only close to the consensus 72-h position but is also heading in a similar direction as the consensus. If only the 72-h positions were considered for the weighting, the model 72-h position may be close to the consensus position at that time, but have a divergent motion vector that quickly steers the model track away from consensus. Using three times for the weighting is a better indication of whether the model track is consistent with the consensus track.

b. Weighting Scheme

The weighting scheme is designed to weight the model tracks after 72 h proportional to their distance from the CONW track prior to 72 h. The distance formula used for this study is described in the Appendix. That is, the un-normalized weighting factors (w) for the number (n) of models weighted are

$$w_1 = \frac{1}{d_1}, \quad w_2 = \frac{1}{d_2}, \quad w_3 = \frac{1}{d_3}, \quad \ldots, \quad w_n = \frac{1}{d_n}. \tag{2.1}$$

The weights are then normalized to sum to one with a normalization factor (x) such that

$$x(w_1) + x(w_2) + x(w_3) + \ldots + x(w_n) = 1. \tag{2.2}$$

Solving for x yields

$$x = \frac{1}{\left(w_1 + w_2 + w_3 + ... + w_n\right)}. \tag{2.3}$$

Substituting Equation (2.1) into Equation (2.3) gives

$$x = \frac{1}{\left(\frac{1}{d_1} + \frac{1}{d_2} + \frac{1}{d_3} + ... + \frac{1}{d_n}\right)}. \tag{2.4}$$

Applying the normalization factor leads to the normalized weights

$$W_1 = x\left(w_1\right), \quad W_2 = x\left(w_2\right), \quad W_3 = x\left(w_3\right), \quad ..., \quad W_n = x\left(w_n\right). \tag{2.5}$$

Equation (2.5) is used to calculate the weights at 60 h, 66 h, and 72 h, and these weights are averaged for application at subsequent times. The weights are not applied at 84 h since the JGSM, the Japan Meteorological Agency typhoon model (JTYM), and the Coupled Ocean-Atmosphere Mesoscale Prediction System (COAMPS) are often available at 84 h and weights are not computed for these models. The weights are only applied at 96 h, 108 h, and 120 h.

The weighting scheme described above will fail when one or more model position is equal to the consensus track position, which will result in a division by zero, and thus an infinite weighting. Thus, a maximum weighting of 0.9 is assigned to that model. In case of multiple models that have the same position as the consensus, this weight of 0.9 is divided evenly among the multiple models. The remaining weighting of 0.1 is divided among the remaining models in a similar manner as in Equations (2.1) – (2.5). However, in Equation (2.2) the 1 on the right side is replaced with 0.1. Likewise in Equations (2.3) - (2.4) the 1 in the numerator is replaced with 0.1. If all model track positions are the same as the consensus position an unweighted consensus will be used.

c. *Applying the Weighting*

The model weighting factors are applied at 96 h, 108 h, and 120 h when in this first example only four models (NOGAPS, GFS, UKMO, and GFDN) are assumed to exist after 72 h. At 96 h, 108 h, and 120 h each of the model latitude and longitude positions is multiplied by their corresponding weights calculated in Equation (2.5). The weighted latitude and longitude positions from each model are then summed to yield the weighted consensus latitude and longitude. When the distances from the operational consensus (CONW) are used to calculate the weights, the weighted consensus is defined as WCOW.

C. SELECTION CRITERIA

Testing as many cases as possible is desirable to validate the weighted consensus technique. However, not all cases may be conducive to using a weighted consensus technique. Therefore, the following criteria were used to select the cases to test the weighting. First, at least two primary model tracks (NOGAPS, GFS, UKMO, and GFDN) at 96 h, 108 h, and 120 h must be available to form the weighted consensus. This first criterion is obvious since at least two models must be available to form a consensus. Second, at least four model tracks must be available at 72 h other than the primary four models tracks. In addition to the JGSM, JTYM, and COAMPS, the other models commonly included in CONW are the Weber barotropic model (WBAR), the Air Force Weather Agency MM5 (MM5), and the Australian TC-Local Area Prediction System (TC-LAPS). The European Center for Medium-range Weather Forecasts (ECMWF) model also became available during the 2006 season, but it is not considered as one of the models to meet the four model criteria since it is not subsequently used for the weighted consensus in this first validation study. The second criterion ensures that there is sufficient guidance contributing to the consensus from 60 h – 72 h to produce a skillful consensus on which to base the weighting factors.

D. AN ALTERNATE POSITION CONSENSUS

While applying the weighted consensus, it was discovered that the CONW latitudes and longitudes did not always match the average of the available model guidance in the ATCF archive. Such differences between the operational CONW and the consensus of available models may be due to operational procedures that have not been fully documented. Whereas the CONW can not be reproduced in all cases, an alternate weighted consensus can be created from a reproducible unweighted consensus of the available models (UAVE). Analogous to the application to CONW, the distance from the UAVE track at 60 h, 66 h, and 72 h is used to calculate the weights as in Equations (2.1) to (2.5). At 96 h, 108 h, and 120 h, a weighted average (WAVE) is computed using these weights and is hypothesized to be an improvement on average over UAVE at these times. Since CONW and UAVE occasionally differ, the weighting scheme will be validated using both WCOW and WAVE.

III. RESULTS FOR WEIGHTED POSITION CONSENSUS

A. VALIDATION OF WEIGHTING CONCEPT

A first validation study using only the four primary models will indicate the usefulness of a weighted consensus technique. Further refinements will be described in the sensitivity studies in Chapter III.C.

1. Weighted Consensus Impact

This first validation test will demonstrate that a weighted consensus can improve over an unweighted consensus when averaged over a large number of cases during the 2006 WPAC season. The weighted consensus improvement is measured by the difference between the unweighted consensus position error and the weighted consensus position error

WCOW Improvement = CONW(error) – WCOW(error), or (3.1)

WAVE Improvement = UAVE(error) – WAVE(error), (3.2)

such that smaller weighted consensus errors will result in a positive value. Improvements are averaged over all cases to yield the mean improvement. The mean percent improvement is calculated as the percentage of the mean weighted consensus improvement relative to the mean unweighted consensus error.

The WCOW and WAVE improvements averaged over all cases resulted in positive mean improvements for both WCOW and WAVE at 96 h, 108 h, and 120 h (Table 3.1). The weighted consensus WAVE of the available model tracks from the primary four models (NOGAPS, GFS, UKMO, and GFDN) in the UAVE sample has the greatest improvement over UAVE at 96 h in terms of both distance improvement and percent improvement. The improvements decrease by 108 h for WAVE, and the improvements again decrease through 120 h but only in terms of percent improvement. Since the average consensus error is increasing with time, a larger distance improvement is needed to maintain the same percent improvement. The consistent decrease in improvement is expected since the weights are determined at 60 h, 66 h, and 72 h. At longer forecast intervals, the weights should become less valid as model tracks change.

Table 3.1 Mean improvement of the weighted consensus WCOW or WAVE relative to unweighted consensus CONW or UAVE for the 2006 season (storms 1 through 24) validation cases (SS: Sample Size).

	Mean Imp	**SS**
96 h WCOW Imp (n mi):	6.0	222
96 h WCOW Imp (%):	2.7	
96 h WAVE Imp (n mi):	6.2	
96 h WAVE Imp (%):	2.8	
108 h WCOW Imp (n mi):	1.7	193
108 h WCOW Imp (%):	0.7	
108 h WAVE Imp (n mi):	5.7	
108 h WAVE Imp (%):	2.2	
120 h WCOW Imp (n mi):	5.3	168
120 h WCOW Imp (%):	1.8	
120 h WAVE Imp (n mi):	5.7	
120 h WAVE Imp (%):	1.9	

The weighted consensus WCOW of the available model tracks from the primary four models in the CONW has a consistent decrease in improvement with time as was the case with WAVE. Whereas the WCOW improvements at 96 h and 120 h are consistent with those of the WAVE, the WCOW improvement at 108 h is inconsistent with that of the WAVE. It might be expected that the 108-h WCOW improvement should lie somewhere between the 96-h and 120-h improvement, but it is much lower than even the 120-h improvement.

The inconsistency can be traced to a couple of outliers at 108 h since WCOW performs particularly poorly for two consecutive cases at 108 h: Storm 21 at 1800 UTC 9 October 2006 and six hours later at 0000 UTC 10 October 2006. The UAVE performs comparably to the WAVE for the 1800 UTC case (Figure 3.1). The WAVE has greater errors than UAVE at 96 h, but the WAVE improves slightly on the UAVE at 108 and 120 h. Neither the UAVE nor the WAVE predicts the observed recurvature and thus they both have large position errors.

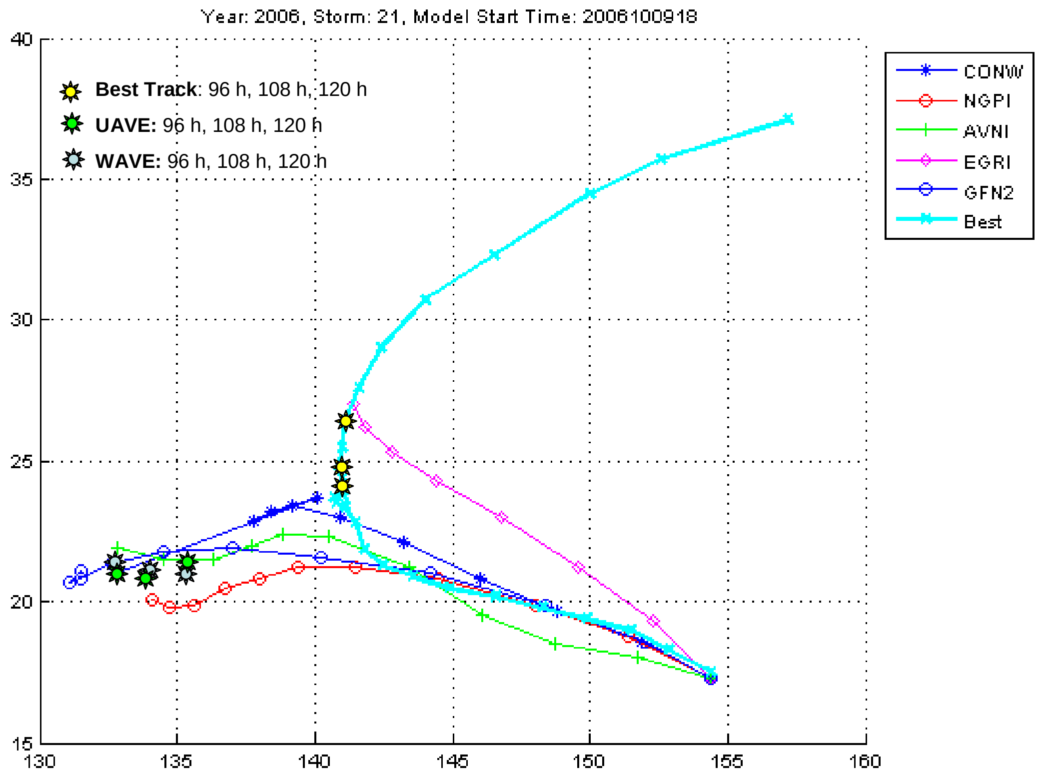


Figure 3.1 Forecast track positions each 12 h for Storm 21 at 1800 UTC 9 October 2006 with CONW track positions (blue asterisks), NOGAPS interpolated track positions (NGPI, red circles), GFS interpolated track positions (AVNI, green crosses), UKMO interpolated track positions (EGRI, purple diamonds), GFDN 12-h interpolated track positions (GFN2, blue circles), and Best Track positions each 6 h (Best, light blue crosses). The 96-, 108-, and 120-h positions are highlighted by stars for the Best-Track (yellow), UAVE (green), and WAVE (grey).

In contrast, the CONW dramatically outperforms the WCOW at 96 h and 108 h while performing comparably at 120 h (Figure 3.2). The 96-h and 108-h CONW positions evidently include additional guidance that indicates recurvature. The interpolated UKMO track (EGRI) indicates recurvature but is only available through 84 h. The UKMO 96-h and 108-h forecast positions were likely removed from the ATCF database, but not until after CONW had been calculated. Because the UKMO model is the only model correctly forecasting recurvature, the UKMO contribution to CONW in this case greatly improves the CONW relative to the UAVE. This additional guidance evidently drops out by 120 h and leads to the same position error for the CONW as for the UAVE. As a result, the WCOW slightly improves on the CONW at 120 h. While the WCOW and the WAVE had somewhat similar errors at 108 h in both cases, the CONW outperformed UAVE by a wide margin in both cases (Table 3.2).

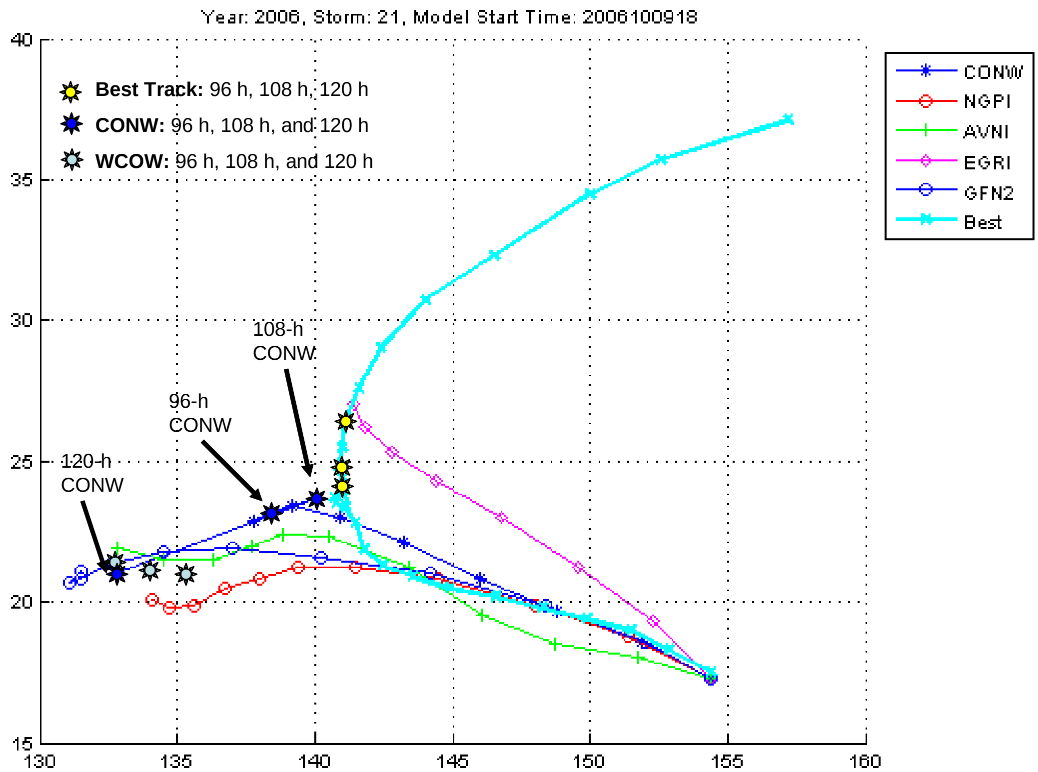


Figure 3.2 As in Figure 3.1, except with CONW 96-, 108-, and 120-h positions (blue stars), and WCOW 96-, 108-, and 120-h positions (grey stars).

Table 3.2 108-h error statistics for two poorly performing WCOW cases.

	09 Oct 18Z	10 Oct 00Z
CONW Error (n mi)	79	103
WCOW Error (n mi)	441	364
UAVE Error (n mi)	482	422
WAVE Error (n mi)	441	373
WCOW Imp (n mi)	-362	-261
WAVE Imp (n mi)	41	49

Because the CONW 108-h error was so small, the WCOW is highly degraded. By contrast, the WAVE resulted in a small improvement on the poor UAVE forecast. In this case, CONW clearly includes guidance that is not used in the UAVE. Since all of the available tracks in the ATCF were used to calculate the UAVE, the JTWC must be using additional guidance in the CONW that is available operationally but is not included in the ATCF file. Alternatively, there may have been a data entry error. Either way, these two cases severely degrade the results at 108 h averaged over all cases.

When these cases are removed, the improvement trends for CONW are consistent with the trends for UAVE (Table 3.3). For these two cases, major differences also exist between UAVE and CONW at 96 h, but little difference exists at 120 h. Apparently the additional guidance contributing to CONW dropped out by 120 h. After removing the two cases, the biggest impact to the mean improvement is for WCOW at 96 h and 108 h, with the biggest improvement at 108 h. As mentioned above, the percent mean improvement in Table 3.3 decreases in time for both WCOW and WAVE. In the two cases that were removed, the large difference between CONW and UAVE at 108 h caused highly negative WCOW improvements that impacted the average improvements at 108 h. It is reassuring that the WAVE improvements are not highly impacted by the removal of these two cases. The results using WAVE are more robust than WCOW because WAVE avoids the uncertainty of the occasional unaccounted-for guidance contributing to CONW.

Table 3.3 As in Table 3.1, except mean improvement for validation cases after removing the two cases in Table 3.2 (SS: Sample Size).

	Mean Imp	SS
96 h WCOW Imp (n mi):	7.8	220
96 h WCOW Imp (%):	3.5	
96 h WAVE Imp (n mi):	6.1	
96 h WAVE Imp (%):	2.8	
108 h WCOW Imp (n mi):	5.0	191
108 h WCOW Imp (%):	1.9	
108 h WAVE Imp (n mi):	5.3	
108 h WAVE Imp (%):	2.0	
120 h WCOW Imp (n mi):	4.9	166
120 h WCOW Imp (%):	1.7	
120 h WAVE Imp (n mi):	5.3	
120 h WAVE Imp (%):	1.8	

2. Performance Graphs

It is useful to evaluate the distribution of weighted consensus errors versus unweighted consensus errors. Insights into the performance of the weighted consensus can be draw from the distribution of weighted consensus errors. All points to the right and below the reference line in Figure 3.3 are cases in which the weighted consensus error is less than the unweighted consensus error. Conversely, the points above and to the left of the line are cases in which the unweighted consensus error was smaller than the weighted consensus error.

At 96 h (Figure 3.3), the WCOW best-fit line has a slope less than one, which indicates that the weighted consensus performs better for larger error cases. Although the WCOW errors are smaller than the CONW errors in only 50% of the cases, the average improvement of 2.7% is still encouraging (Table 3.1). Larger magnitudes of WCOW improvement below the reference line indicate that in many cases the WCOW significantly outperformed CONW. By contrast, not as many significant outliers above the reference line indicates that when CONW outperformed WCOW the differences tended to be modest. Thus, the 2.7% improvement is due to the larger number of cases in which WCOW outperforms CONW by a large margin.

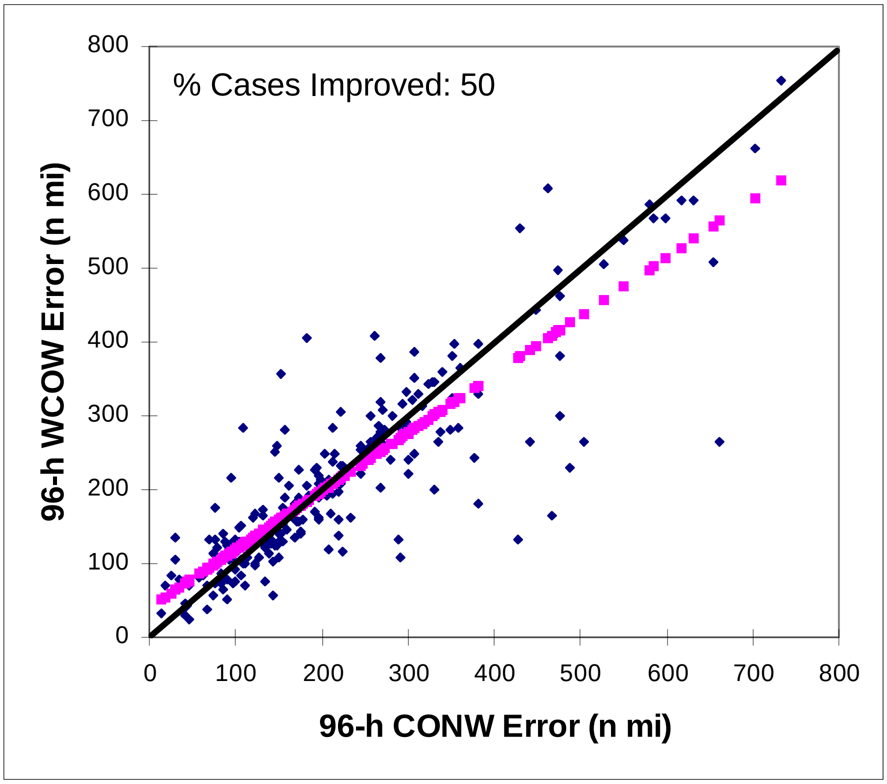


Figure 3.3 96-h CONW versus WCOW errors (blue diamonds) with 1:1 reference line (solid) and predicted best fit line for WCOW (pink boxes).

For the 96-h WAVE errors (Figure 3.4), the best-fit line slope is again less than one, which indicates that the weighted consensus performs better for cases with larger UAVE errors. Generally there is less spread from the reference line for WAVE than for WCOW, which indicates that many UAVE errors tend to be comparable to WAVE errors. Although the numbers of outliers below and above the reference line are smaller with the WAVE than the WCOW, the outliers below the reference line where WAVE dramatically outperformed UAVE contributed to the 2.8% WAVE mean improvement over the UAVE (Table 3.1). It is noteworthy that even with fewer cases of dramatic improvement with WAVE than with WCOW, a higher percent of the WAVE cases (55%) are improved.

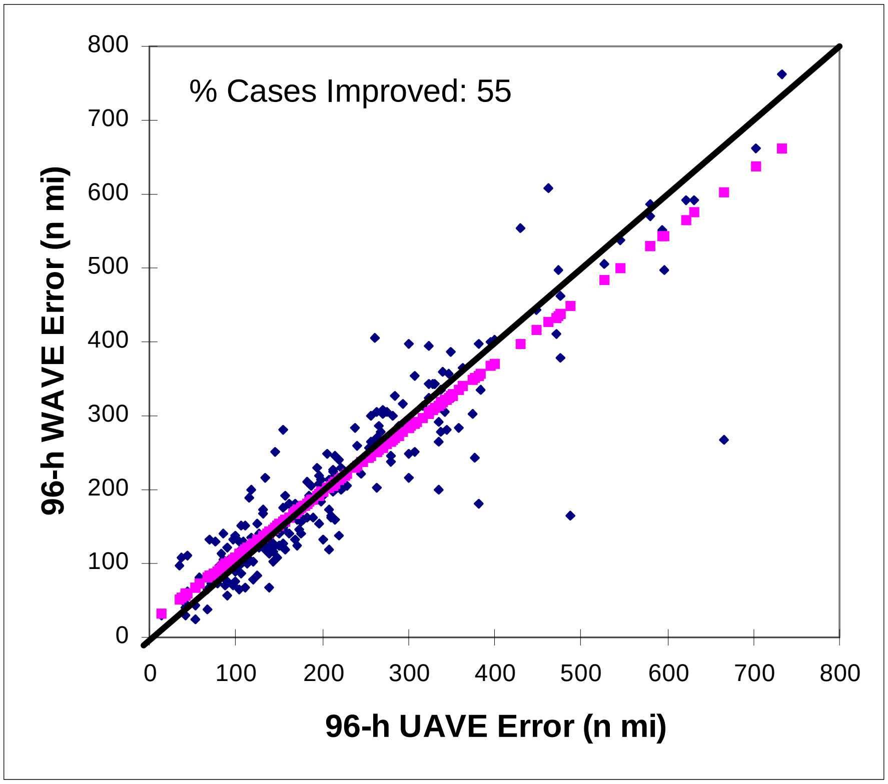


Figure 3.4 As in Figure 3.3, except for 96-h UAVE versus WAVE errors.

As was the case at 96 h, a large number of 108-h WCOW cases have a large improvement relative to CONW (Figure 3.5). However, less than half of the 108-h WCOW cases show improvement. Two noteworthy cases in which CONW significantly outperforms WCOW were discussed in Chapter III.A.1 above that had small CONW errors but large WCOW errors (Table 3.2). Two other CONW errors of about 250 n mi and 350 n mi with WCOW errors about 200 n mi larger are revealed in Figure 3.5. Due to these four WCOW errors (and perhaps the case of a WCOW 900 n mi error with a CONW error of about 750 n mi), the 108-h WCOW has less mean improvement (0.7%) than at 96 h.

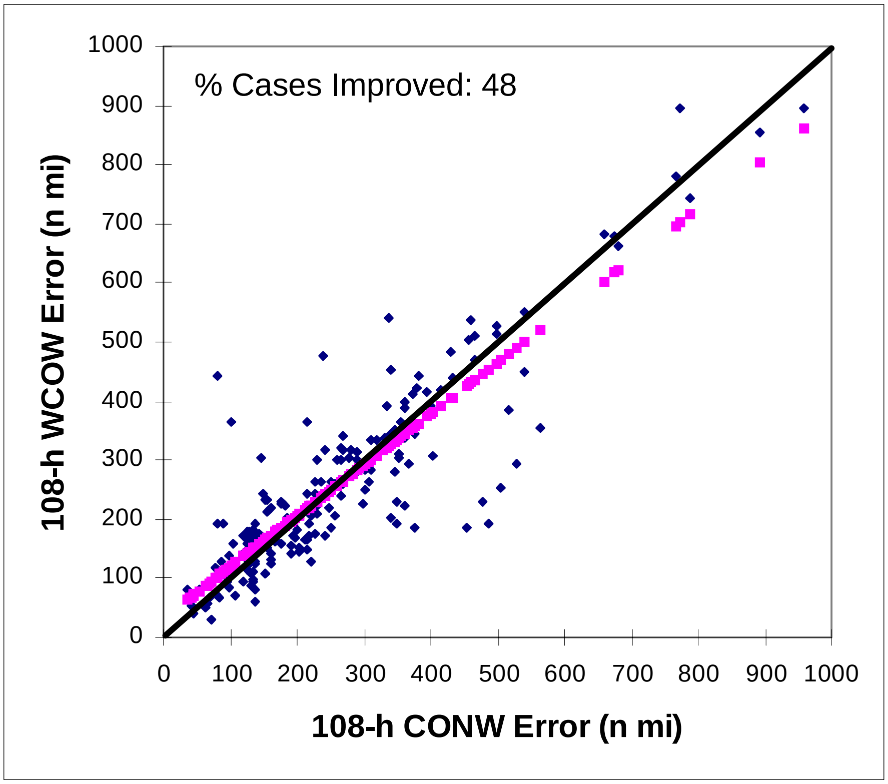


Figure 3.5 As in Figure 3.3, except for 108-h CONW versus WCOW errors.

By comparison, the 108-h WAVE errors relative to the UAVE consensus errors (Figure 3.6) have smaller deviations than the WCOW errors relative to CONW errors in Figure 3.5. Fewer cases of a dramatic improvement or degradation from the weighted consensus WAVE are found than for the WCOW, which indicates that the operational application with CONW discussed in Chapter III.A.1 leads to a large spread in weighted consensus improvements. It is noteworthy that the WAVE cases are biased below the reference line with 53% of the cases showing an improvement compared to only 48% of the WCOW cases.

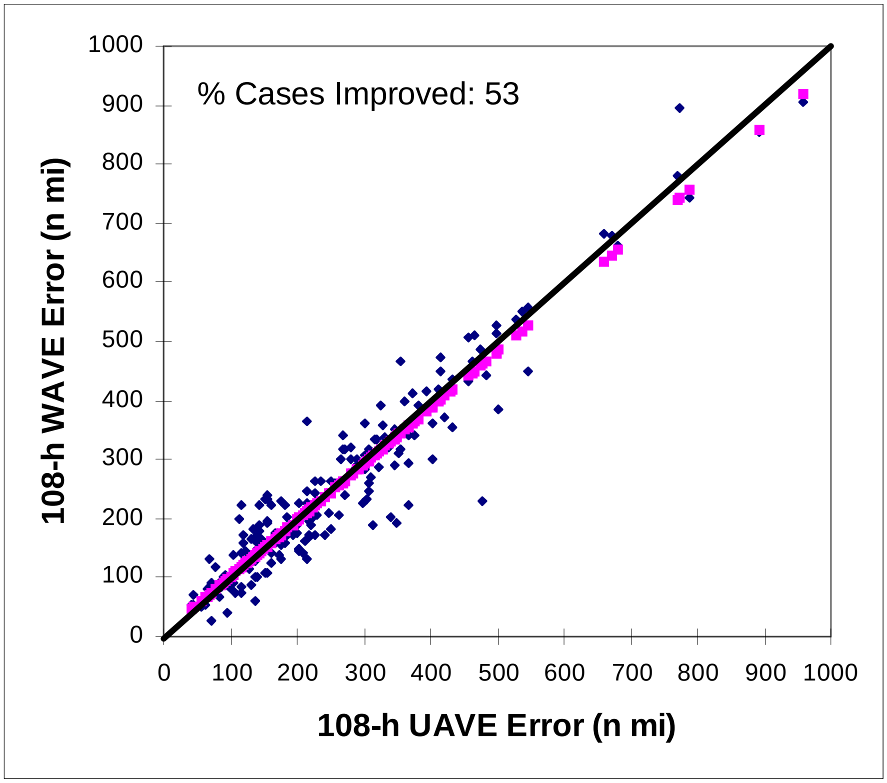


Figure 3.6 As in Figure 3.4, except for 108-h UAVE versus WAVE errors.

As was the case at 96 h, a large number of the 120-h WCOW cases fall below the reference line (Figure 3.7), and these cases of large WCOW improvement relative to the CONW contribute to a 1.8% mean improvement (Table 3.1). Despite the positive mean improvement, only 46% of the cases show improvement. Whereas a large fraction of the 120-h WCOW cases degraded relative to the CONW are tightly clustered just above the reference line in Figure 3.7, three outliers with WCOW errors 150 – 200 n mi larger than the CONW errors are also noted. Nevertheless, the outliers below the reference line lead to a mean improvement despite less than half of the WCOW cases being an improvement over the CONW.

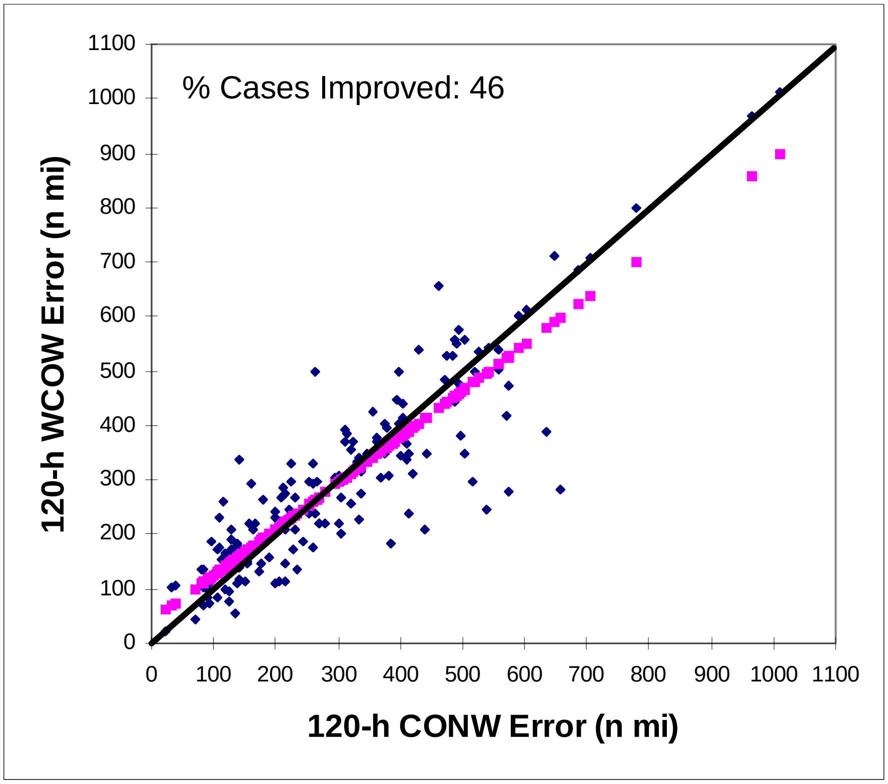


Figure 3.7 As in Figure 3.3, except for 120-h CONW versus WCOW errors.

The 120-h WAVE cases (Figure 3.8) are analogous to the 108-h comparison in Figure 3.6, with the WAVE cases more closely packed to the reference line than the 120-h WCOW cases (Figure 3.7). More outliers are found below the reference line, which again leads to an average improvement (1.9%) even though only 49% of the cases are improved. It is also noteworthy that the WAVE best-fit line nearly corresponds to the reference line, which indicates that the cases are fairly balanced above and below the reference line and that the weighted consensus technique performs similarly for low and high UAVE error cases.

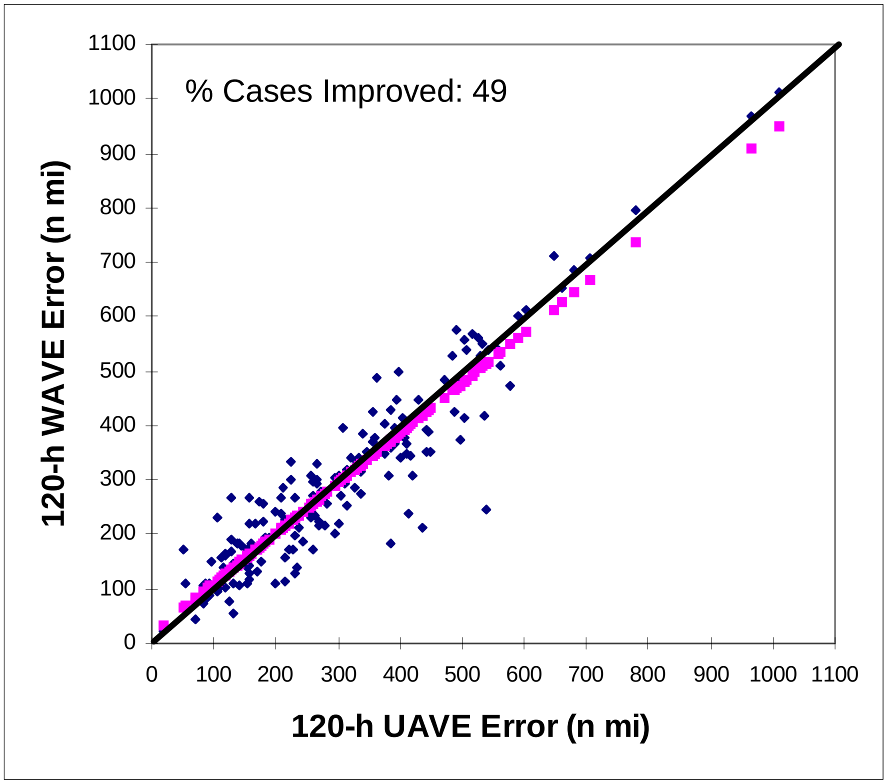


Figure 3.8 As in Figure 3.6, except for 120-h UAVE versus WAVE errors.

B. CASE STUDIES

It is insightful to analyze cases in which the weighted consensus performs well to understand in what situations the technique leads to improvement. It is also useful to analyze cases in which the unweighted consensus outperforms the weighted consensus. If the scenarios in which the weighted consensus technique performs poorly can be identified, a selective weighted consensus might be applied. By using the unweighted consensus instead of the weighted consensus in the scenarios that are unfavorable for the weighted consensus, the overall error statistics could be dramatically improved. If only a few cases with a large degradation from the weighted consensus are eliminated, it could significantly improve the overall error statistics.

1. Favorable Cases

The weighted consensus improved over the unweighted consensus by more than 200 n mi at 96, 108, and 120 h for storm 4 from 1200 UTC 3 July 2006

(Table 3.4). This dramatic improvement is evident in Figure 3.9 with the WCOW track beginning to turn poleward more slowly than the CONW track, and thus is much closer to the Best Track (BT). In addition, the CONW track has a distinct kink after 72 h, whereas the 96-, 108-, and 120-h WCOW track positions provide a smooth transition from the pre-72 h CONW track. The kink in the CONW track is due to a loss of model guidance after 72 h: nine models are contributing to the CONW through 72 h, seven models at 84 h, and only four models by 96 h. In particular, the loss of two west-northwestward oriented tracks (TC-LAPS after 72 h and the JTYM after 84 h) (Figure 3.10) resulted in the CONW track recurving more rapidly after 72 h. These two model tracks contributed to a CONW track that was shifted to the southwest compared to contributions from recurving GFS and UKMO tracks. As a result, the CONW positions at 60 h, 66 h, and 72 h are closer to the more slowly poleward turning NOGAPS and GFDN tracks, and thus these two models are weighted by 0.75 and 0.15, respectively (inset in Figure 3.9). Since the BT is also turning poleward slowly, the weighted consensus WCOW produces much smaller errors than the unweighted consensus CONW when these weights are applied at 96 h, 108 h, and 120 h. This case demonstrates the ability of the weighted consensus to utilize the valuable earlier model guidance (TC-LAPS and JTYM) to improve the consensus after these models are no longer available.

Table 3.4 CONW and WCOW errors (n mi) and improvement (n mi and percent) of the weighted consensus WCOW relative to the unweighted consensus CONW for storm 4 from 1200 UTC 3 July 2006.

Time	CONW error (n mi)	WCOW error (n mi)	WCOW Imp (n mi)	WCOW % Imp
96 h	381.6	181.0	200.6	52.6
108 h	477.9	230.5	247.5	51.8
120 h	537.5	245.4	292.1	54.4

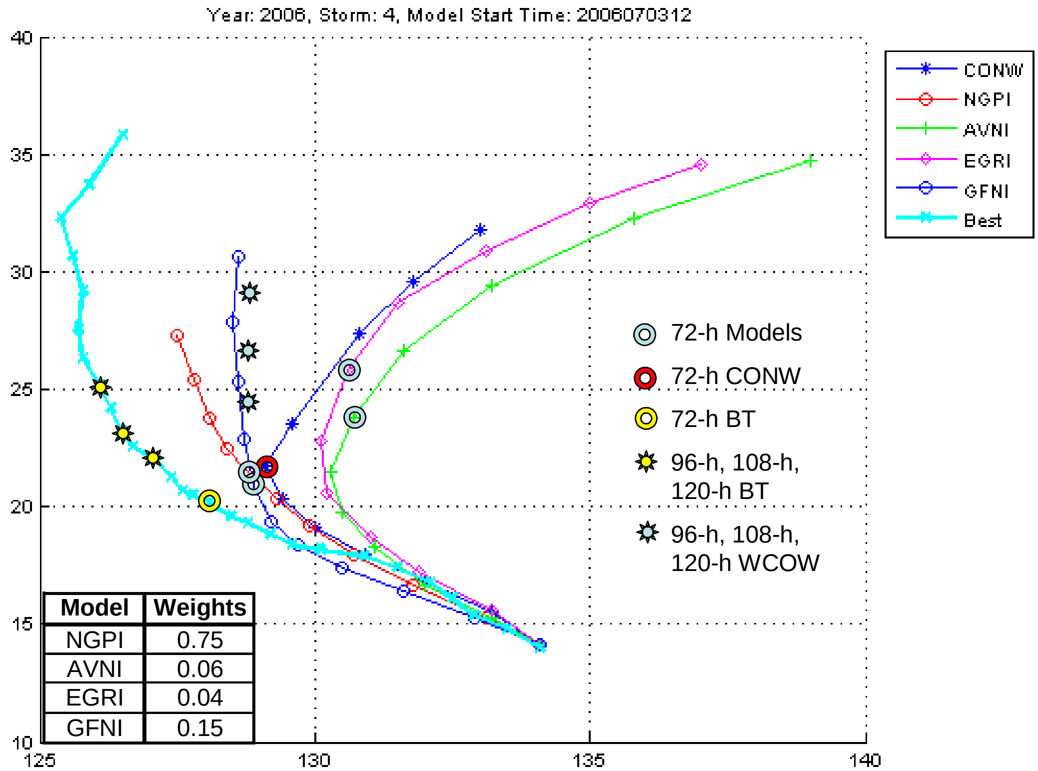


Figure 3.9 Forecast track positions each 12 h for Storm 4 from 1200 UTC 03 July 2006 with CONW track positions (blue asterisks), NOGAPS interpolated track positions (NGPI, red circles), GFS interpolated track positions (AVNI, green crosses), UKMO interpolated track positions (EGRI, purple diamonds), GFDN 12-h interpolated track positions (GFNI, blue circles), and Best-Track positions every 6 h (Best, light blue crosses). The 72-h positions are highlighted by rings for the Best-Track (yellow), and those for the models (grey) and the CONW (red) are used to assign weights (see inset). The 96-, 108-, and 120-h positions are highlighted by stars for the Best-Track (yellow) and WCOW (grey).

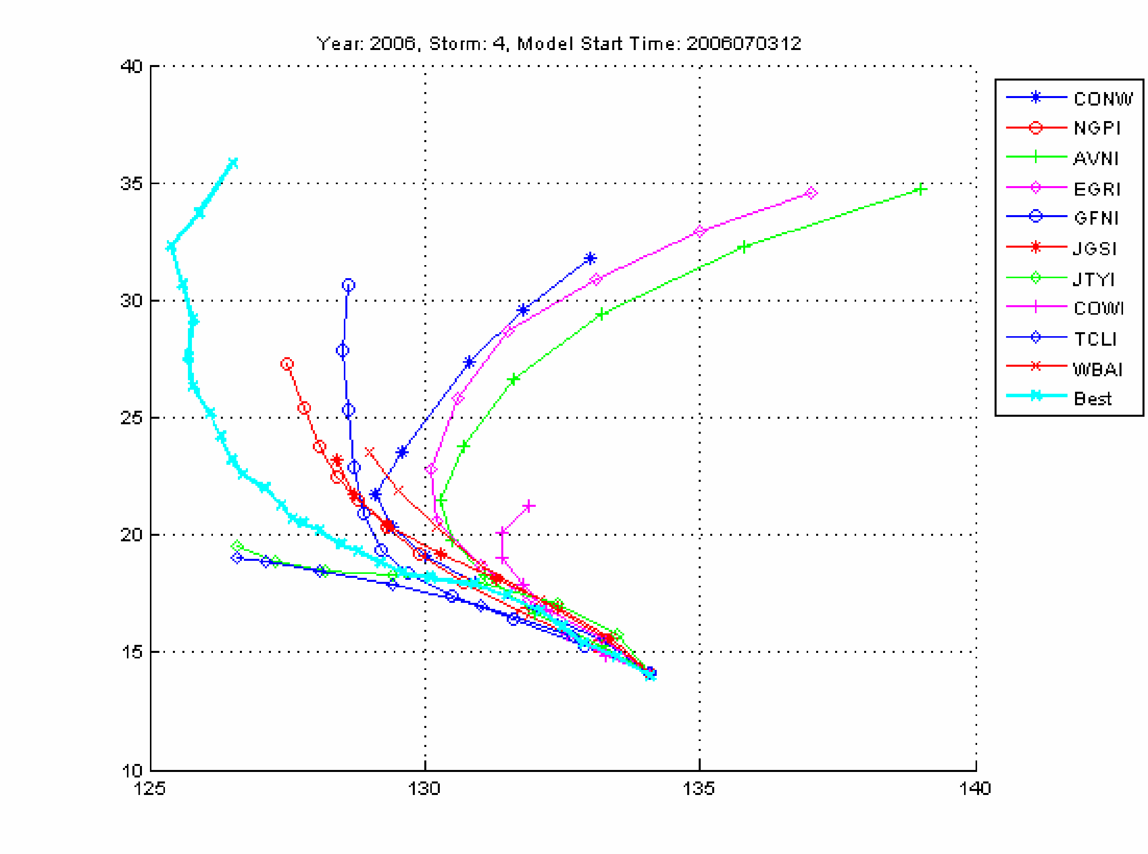


Figure 3.10 Forecast track positions each 12 h for Storm 4 from 1200 UTC 3 July 2006 as in Fig. 3.17, except with the addition of JGSM interpolated track positions (JGSI, red asterisks), JTYM interpolated track positions (JTYI, green crosses), COAMPS interpolated track positions (COWI, purple crosses), TC-LAPS interpolated track positions (TCLI, blue diamonds), WBAR interpolated track positions (WBAI, red crosses), and Best Track interpolated positions (light blue crosses).

The weighted consensus for storm 6 from 1200 UTC 20 July 2006 is able to improve on an already reasonable unweighted consensus forecast (Table 3.5). The weighted consensus WCOW track is shifted southwestward from the CONW track at 96 h, 108 h, and 120 h, which greatly reduces the cross-track error although the WCOW positions have a slow along-track bias (Figure 3.11). In this case, the success of the weighted consensus is due to the reduced weight given to an outlier (GFS; inset in Figure 3.11). That is, the GFS track has a northward heading that departs significantly from the CONW track that generally has a northwesterly heading, and thus the GFS receives the smallest weighting (0.24).

The other model tracks that go into the CONW through 72 h are also shown in Figure 3.12, which indicates that NOGAPS and GFDN are also outliers relative to the cluster of models. However, these two model tracks have a similar 60-72 h heading as the CONW track and receive greater weightings (0.40 and 0.36, respectively) due to their closer agreement with the CONW track than the GFS track. Although the GFS is a distinct outlier, it is still valuable in pulling the weighted consensus to the east. By weighting the GFS track less in the weighted consensus than in the unweighted consensus, the GFS has less influence in pulling the weighted consensus to the east, which results in a near alignment with the Best-Track despite a timing error. Thus, the weighting technique was able to successfully identify the GFS as an outlier, which yielded improvements over an unweighted consensus.

Table 3.5 As in Table 3.4, except for storm 6 from 1200 UTC 20 July 2006.

Time	CONW error (n mi)	WCOW error (n mi)	WCOW Imp (n mi)	WCOW % Imp
96 h	111.6	69.7	41.9	37.6
108 h	204.4	151.5	52.9	25.9
120 h	280.3	221.4	58.9	21.0

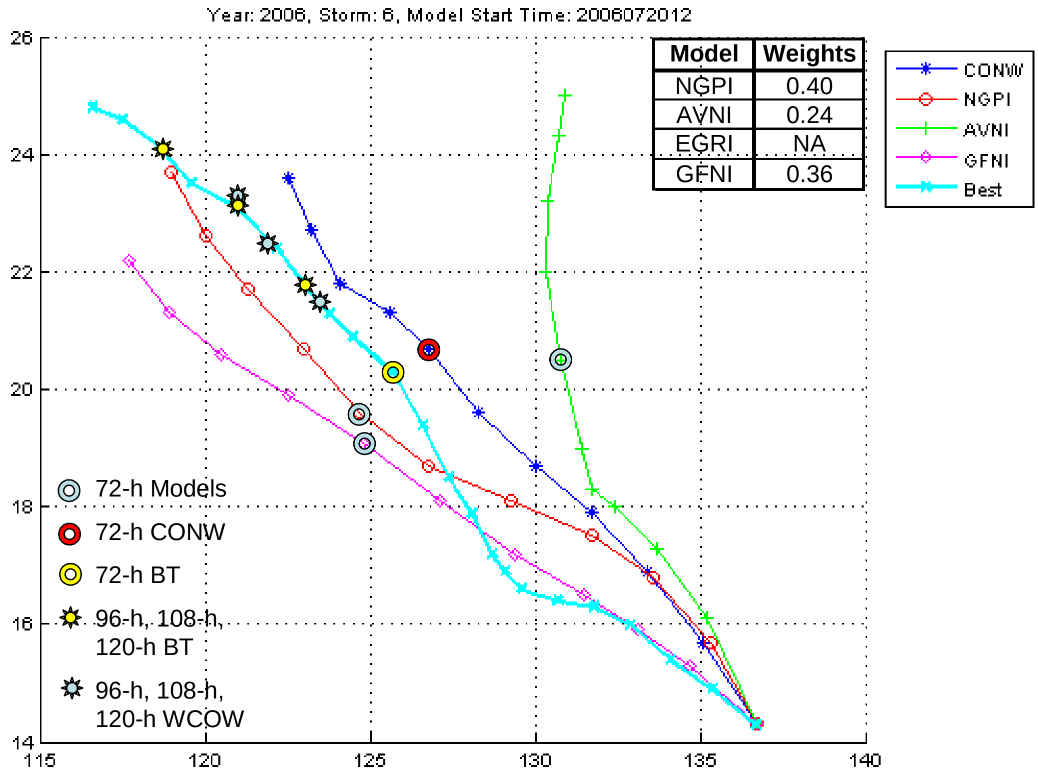


Figure 3.11 As in Figure 3.9, except for Storm 6 from 1200 UTC 20 July 2006 and without the UKMO interpolated track positions.

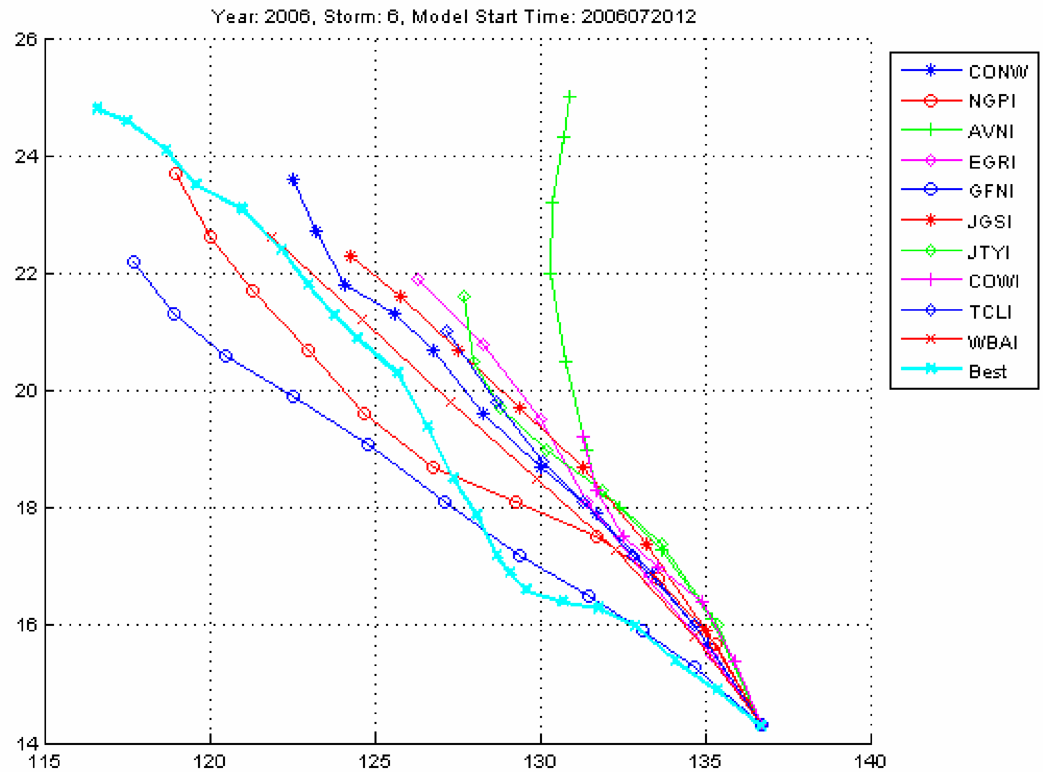


Figure 3.12 As in Figure 3.10, except for Storm 6 from 1200 UTC 20 July 2006.

The weighted consensus forecast for storm 14 from 0600 UTC 11 September 2006 is a case of dramatic improvement over the unweighted consensus forecast (Table 3.6) and in consensus track consistency (Figure 3.13). The CONW track has a physically unreasonable sharp turn to the northeast after 84 h due to the additional influence of the poor GFS forecast when only four model tracks are available (Figure 3.13). Continuing 96-, 108-, and 120-h positions from the 72-h CONW position through the WCOW gives a more realistic solution that is closer to the BT and closely resembles the curvature of the BT. In this case, it would be desirable to eliminate the GFS track from the consensus CONW since it clearly is unlikely to be correct. In this case, the weighted consensus has little influence from the GFS track by giving it a weight of only 0.09 (inset, Figure 3.13). By contrast, the NOGAPS and UKMO model tracks are given weighting factors of 0.44 and 0.32, respectively, since they are close to consensus CONW at 60 h, 66 h, and 72 h (Figure 3.13). This is advantageous since both models have a recurvature path similar to the BT. By contrast, the GFDN model track is given a small weighting factor since the GFDN track is south of the BT, and most of the other models (Figure 3.14), and thus the CONW track is north of the BT. In this case, the primary reason for dramatic improvement in WCOW relative to the CONW is again due to the assignment of a justifiably small weighting factor to an outlier (the GFS track).

Table 3.6 As in Table 3.4, except for storm 14 from 0600 UTC 11 September 2006.

Time	CONW error (n mi)	WCOW error (n mi)	WCOW Imp (n mi)	WCOW % Imp
96 h	143.8	102.9	40.8	28.4
108 h	219.5	129.7	89.8	40.9
120 h	234.2	135.9	98.3	42.0

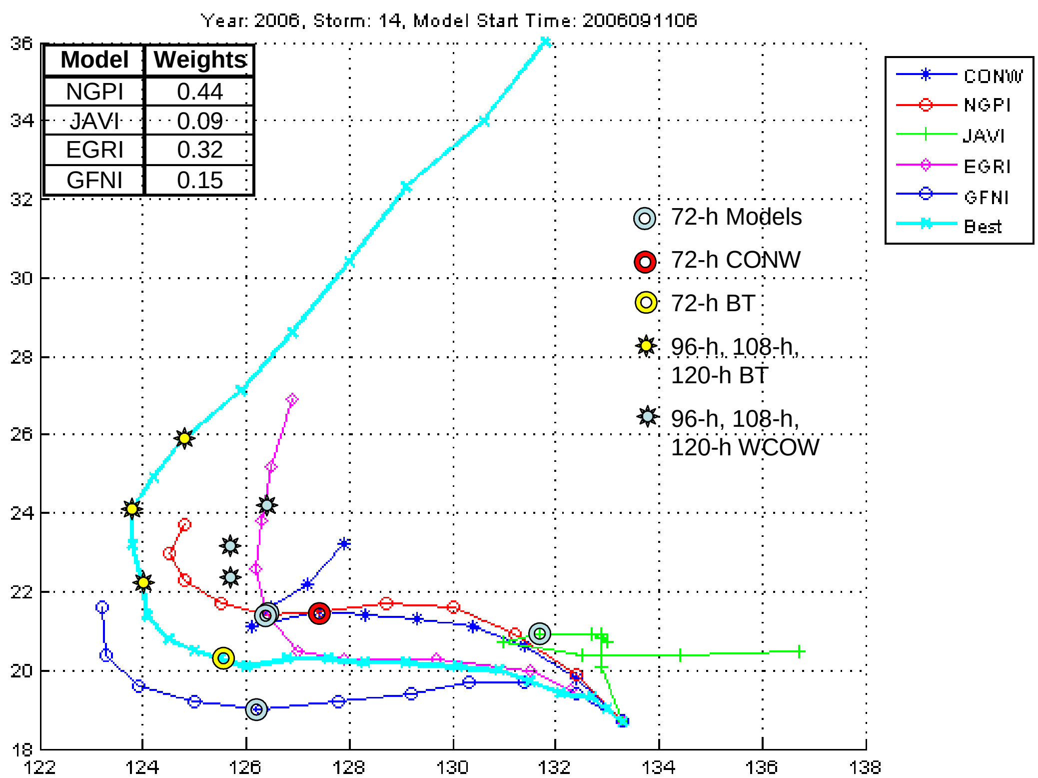


Figure 3.13 As in Figure 3.9, except for Storm 14 from 0600 UTC 11 September 2006 and with GFS interpolated track positions (JAVI, green crosses).

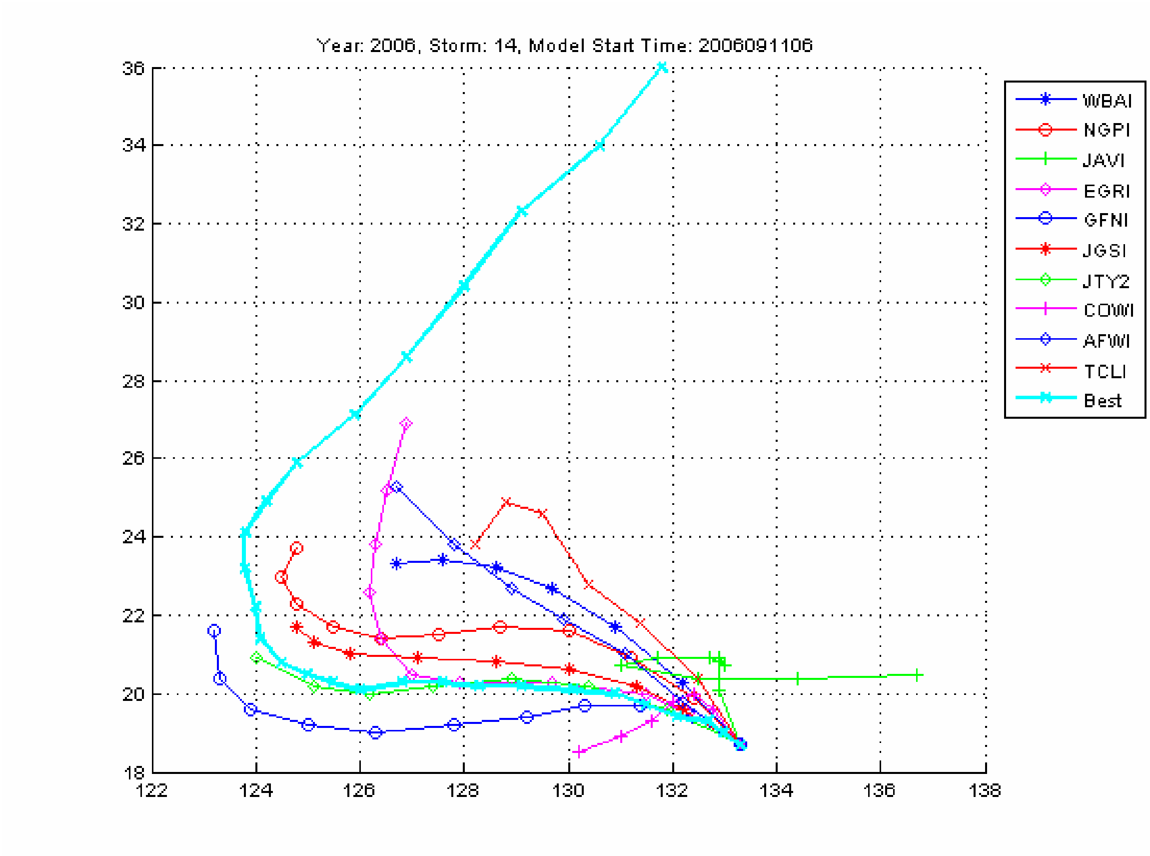


Figure 3.14 As in Figure 3.10, except for Storm 14 from 0600 UTC 11 September 2006, with WBAR interpolated track positions (blue asterisks), GFS interpolated track positions (JAVI, green crosses), JTYM 12-h interpolated track positions (JTY2, green diamonds), MM5 interpolated track positions (AFWI, blue diamonds), and TC-LAPS interpolated track positions (TCLI, red crosses).

The weighted consensus achieved significant improvements relative to the unweighted consensus for Storm 16 from 0000 UTC 19 September 2006 (Table 3.7) by shifting the WCOW track consensus towards the BT and advancing it in the along-track direction to closer match the BT (Figure 3.15). The weighted consensus works so well in this case because the tracks are smoothly varying and the model distances from the consensus CONW at 60 h, 66 h, and 72 h are a good indication of the model distances from consensus at 96 h, 108 h, and 120 h. It is also advantageous that most of the model guidance that goes into CONW is distributed fairly evenly about the BT through 72 h (Figure 3.16), and therefore the CONW track is close to the BT and is a good basis for the weighting. By 96 h, the four remaining models (NOGAPS, GFS, UKMO, and GFDN) do not form an even distribution about the BT, which results in a westward CONW track jog from 72 h to 96 h and then an overall westward shift in the CONW track due to the outlying GFS. Although the GFDN path is also close to the BT, it is lagging behind as indicated by the 72-h GFDN position in Figure 3.15, and thus it receives a smaller weight (0.12; inset in Figure 3.15). Since the WCOW gives the GFDN little weight, the WCOW track is advanced farther in the along-track direction than the CONW. The WCOW performs so well relative to the CONW because it gives small weighting factors to the GFS and GFDN and gives the NOGAPS and UKMO tracks significant weights because those two tracks are close to the CONW track at 60 h, 66 h, and 72 h. These consistent positions relative to the consensus are why the weighted consensus works so well in this case.

Table 3.7 As in Table 3.4, except for Storm 16 from 0000 UTC 19 September 2006.

Time	CONW error (n mi)	WCOW error (n mi)	WCOW Imp (n mi)	WCOW % Imp
96 h	232.7	161.9	70.8	30.4
108 h	348.9	230.6	118.3	33.9
120 h	503.7	346.8	156.9	31.2

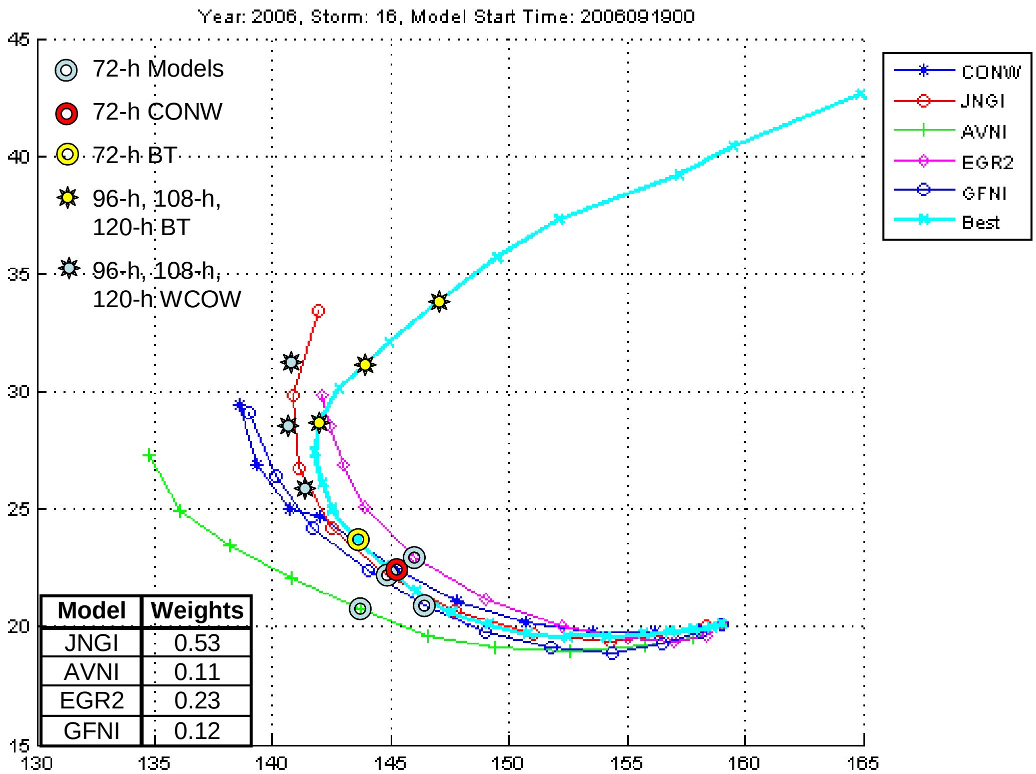


Figure 3.15 As in Figure 3.9, except for Storm 16 from 0000 UTC 19 September 2006 and with NOGAPS interpolated track positions (JNGI, red circles), and UKMO 12-h interpolated track positions (EGR2, purple diamonds).

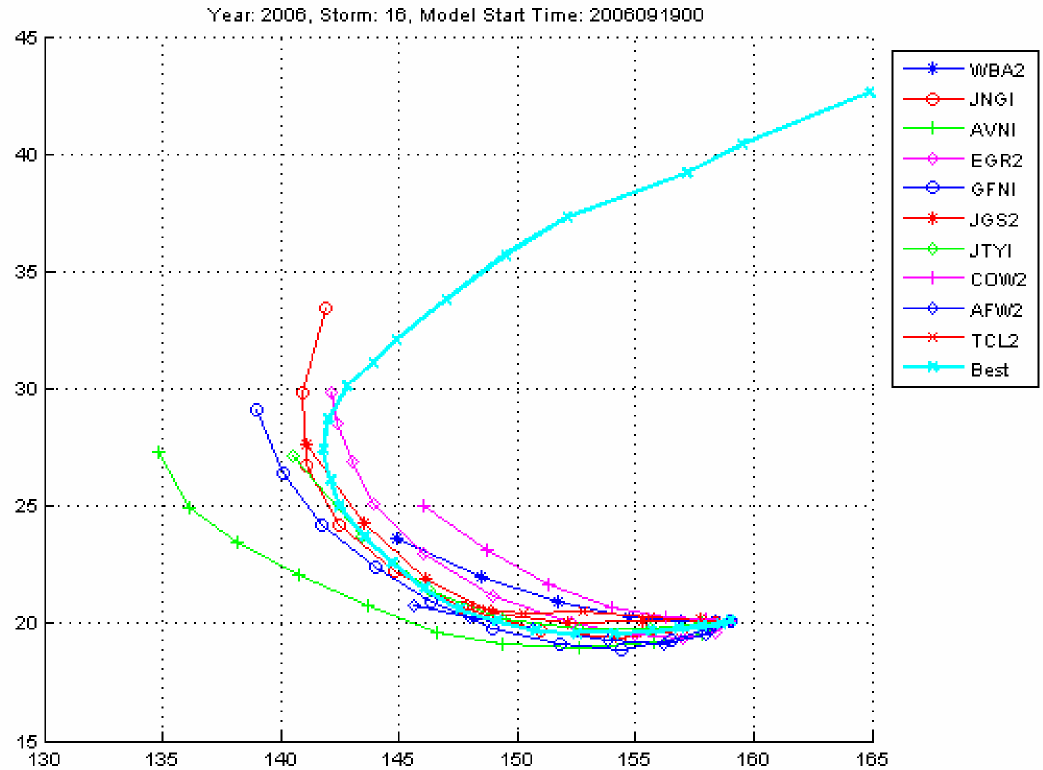


Figure 3.16 As in Figure 3.14, except for Storm 16 from 0000 UTC 19 September 2006 and with WBAR 12-h interpolated track positions (WBA2, blue asterisks), NOGAPS interpolated track positions (JNGI, red circles), GFS interpolated track positions (AVNI, green crosses), UKMO 12-h interpolated track positions (EGR2, purple diamonds), JGSM 12-h interpolated track positions (JGS2, red asterisks), COAMPS 12-h interpolated track positions (COW2, purple crosses), MM5 12-h interpolated track positions (AFW2, blue diamonds), and TC-LAPS 12-h interpolated track positions (TCL2, red crosses).

2. Unfavorable Cases

The weighted consensus is degraded for Storm 2 from 1200 UTC 12 May 2006 (Table 3.8). The NOGAPS track is given the largest weighting factor (0.53; inset in Figure 3.17) because it closely parallels the CONW track from 60 h through 72 h. However, the NOGAPS track does not turn northward as fast as the storm or the other consensus members, which results in a weighted consensus track that is progressively farther to the west relative to the unweighted consensus track (Figure 3.17). It is noteworthy that the various tracks that are included in the CONW consensus reasonably encompass the BT through 72 h (Figure 3.18) and thus the CONW track is close to the BT, and the 72-h CONW position in Figure 3.17 is close to the 72-h BT position. Although the CONW should be a good basis for the weighting, the weighting values through 72 h could not account for the subsequent northward turn in most of the guidance through 120 h. The remaining three models have more accurate recurving tracks than the NOGAPS, but they are too far displaced from the CONW during 60 – 72 h to receive comparable weights. Coincidently, the four consensus member tracks at 96 h encompass the BT fairly well, which leads to a successful unweighted consensus at that time.

Table 3.8 As in Table 3.4, except for Storm 2 from 1200 UTC 12 May 2006.

Time	CONW error (n mi)	WCOW error (n mi)	WCOW Imp (n mi)	WCOW % Imp
96 h	12.8	31.6	-18.7	-145.9
108 h	45.0	75.6	-30.7	-68.1
120 h	118.7	164.1	-45.4	-38.3

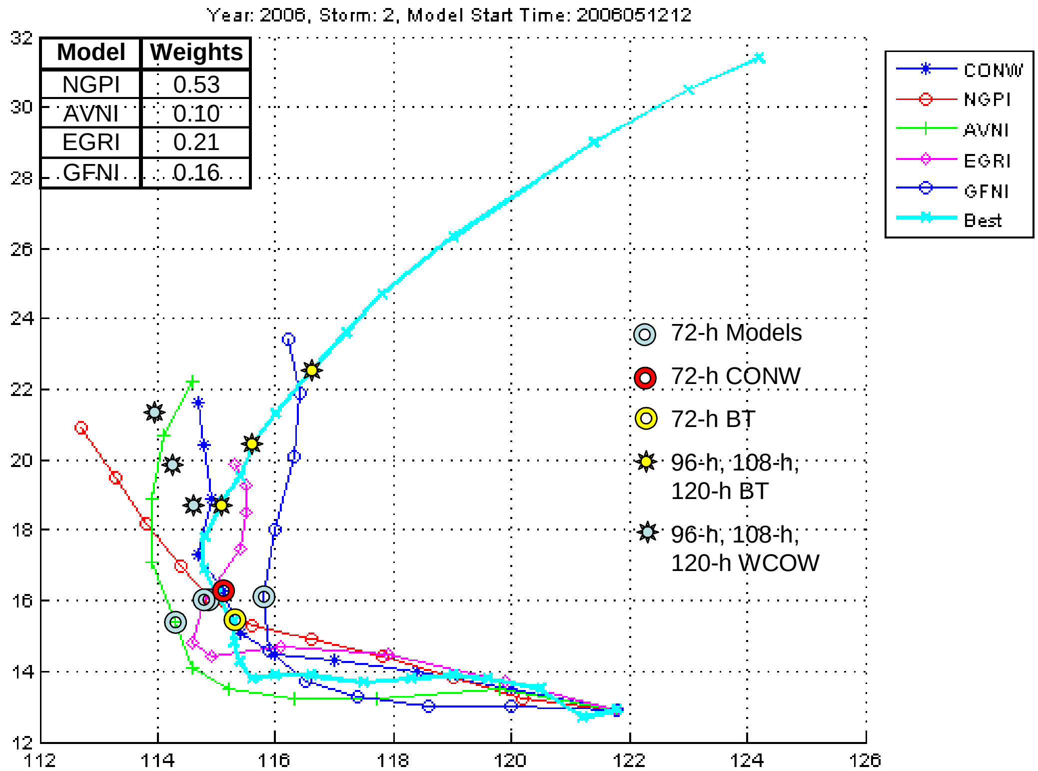


Figure 3.17 As in Figure 3.9, except for Storm 2 from 1200 UTC 12 May 2006.

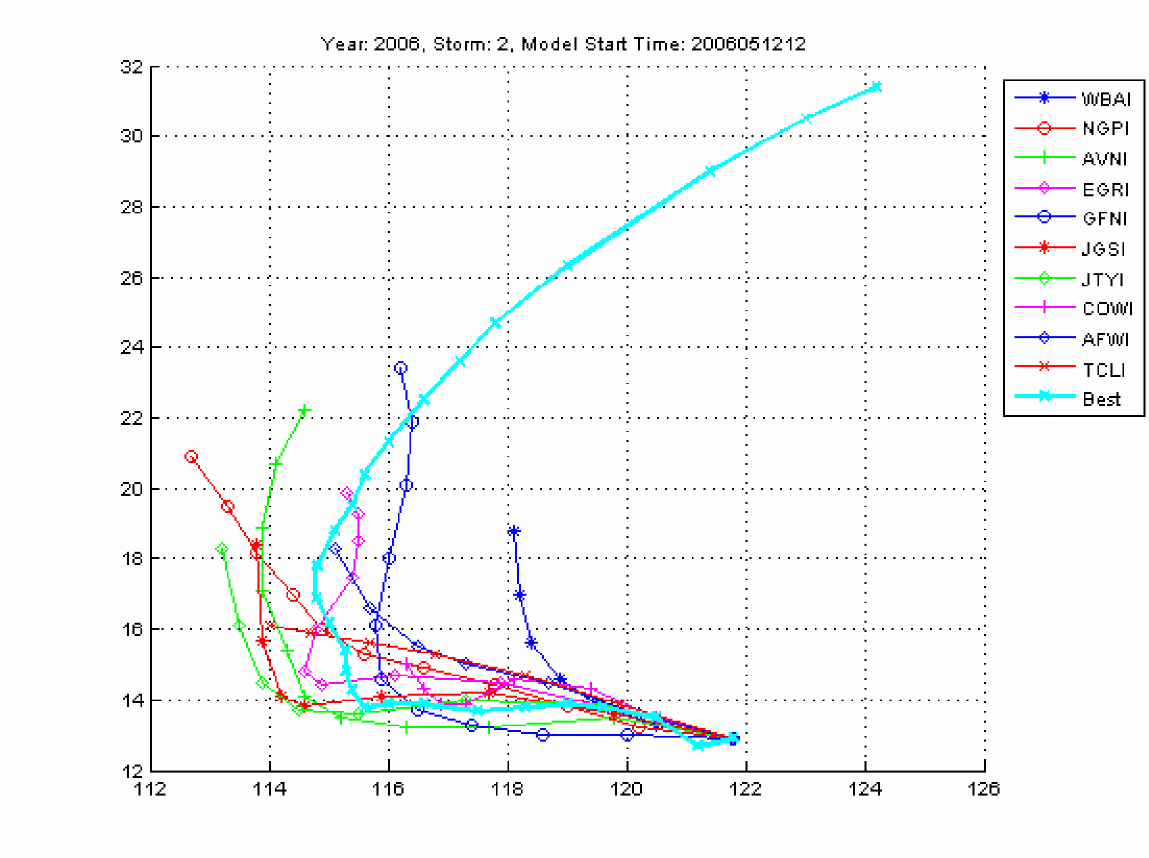


Figure 3.18 As in Figure 3.14, except for Storm 2 from 1200 UTC 12 May 2006 and with GFS interpolated track positions (AVNI, green crosses), and JTYM interpolated track positions (JTYI, green diamonds).

Another case in which the weighted consensus performed poorly is Storm 4 from 0600 UTC 1 July 2006 (Table 3.9). In this case, the weighting scheme assigned the largest weighting factor to what turned out to be the worst track (the GFDN) and assigned the smallest weighting factor to the UKMO model with the track that would have verified best. Because three of the model tracks that had agreed more closely with the 60-h – 72-h CONW positions subsequently recurved, the weighted consensus WCOW track recurves much too fast (Figure 3.19). The weighted consensus fails because the weighting factors are calculated before any of the members have begun any significant turn poleward, which began after 72 h. Since most of the additional models (those not used in the weighted consensus) are north of the BT (Figure 3.20), the CONW track is on the north edge of the four weighted consensus members (Figure 3.19) and is nearly aligned with the GFDN track. As in the previous case, one might have assumed that the CONW track should be a good basis for the weighting factor calculation since the CONW track is closely aligned with the BT through 72 h. Unfortunately, the GFDN track diverges from CONW shortly after 72 h as the GFDN track begins to recurve, which results in a degraded weighted consensus. Even though the UKMO is the closest model from the BT at 96 h, 108 h, and 120 h, it is the farthest model from the CONW track and the BT at 72 h. The low weight assigned to the UKMO track, as well as the subsequent recurvature of the other three models, lead to a degraded weighted consensus track forecast. Although the WCOW technique provides a reasonable representation of the model guidance after 72 h, the straight west-northwest track of Storm 4 is not predicted in three of the four primary models (NOGAPS, GFS, UKMO, and GFDN) beyond 72 h.

Table 3.9 As in Table 3.4, except for Storm 4 from 0600 UTC 1 July 2006.

Time	CONW error (n mi)	WCOW error (n mi)	WCOW Imp (n mi)	WCOW % Imp
96 h	85.1	140.5	-55.4	-65.1
108 h	153.7	232.3	-78.6	-51.2
120 h	225.3	330.9	-105.6	-46.9

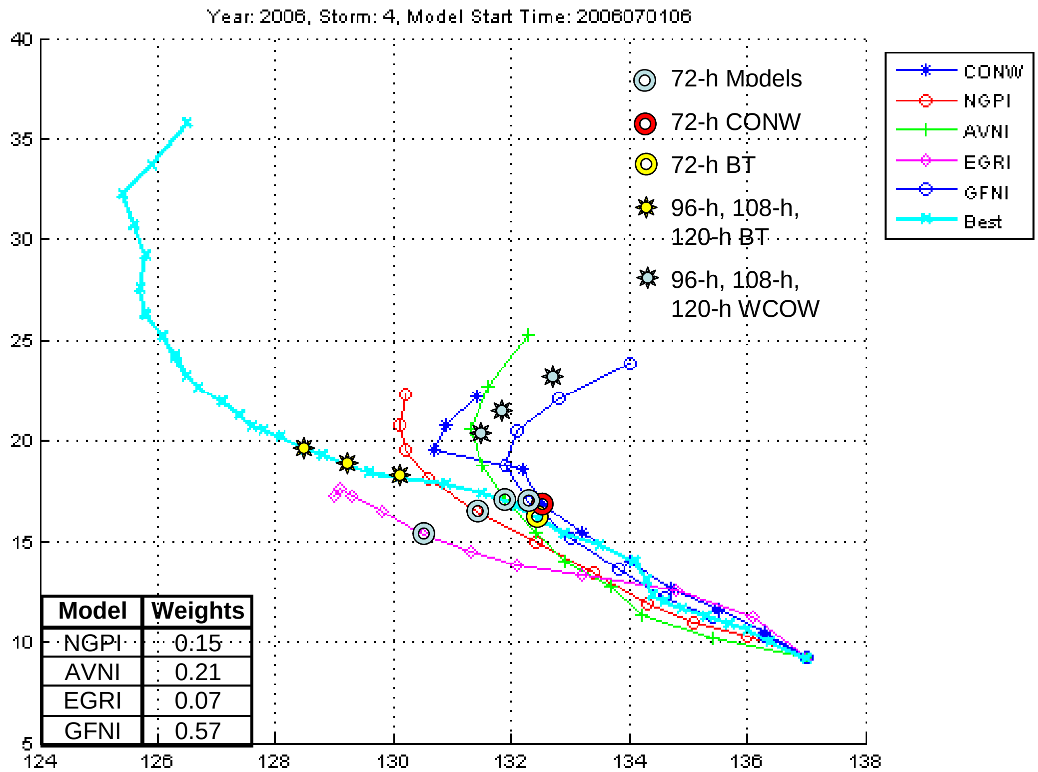


Figure 3.19 As in Figure 3.9, except for Storm 4 from 0600 UTC 1 July 2006.

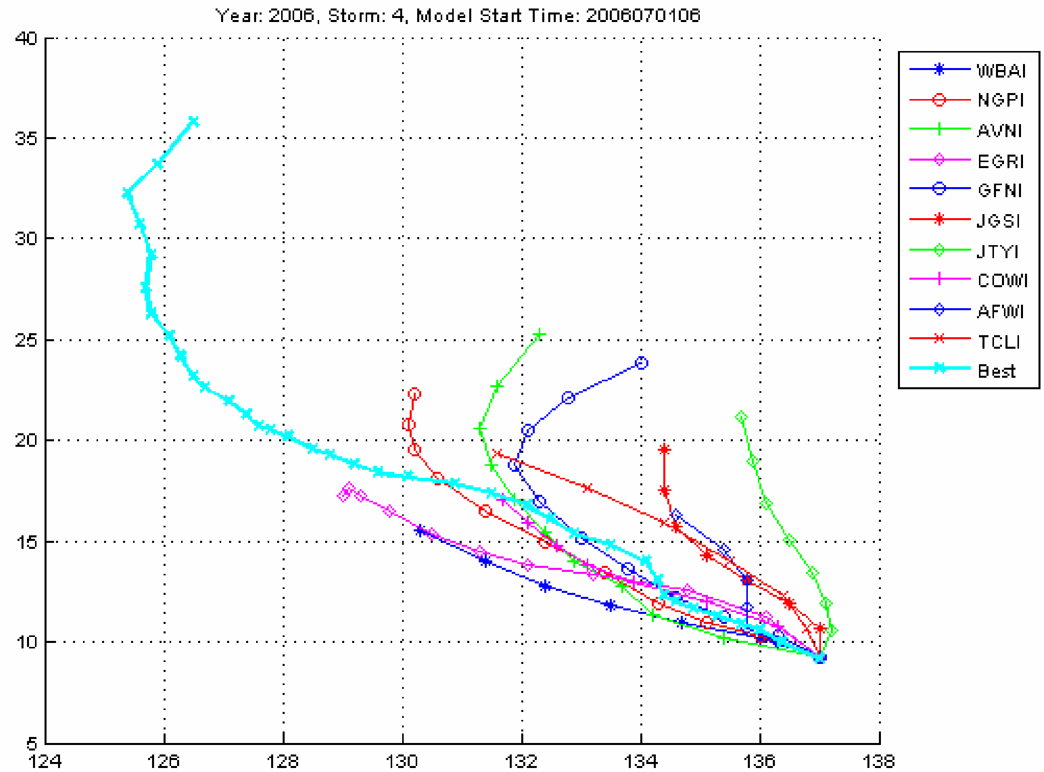
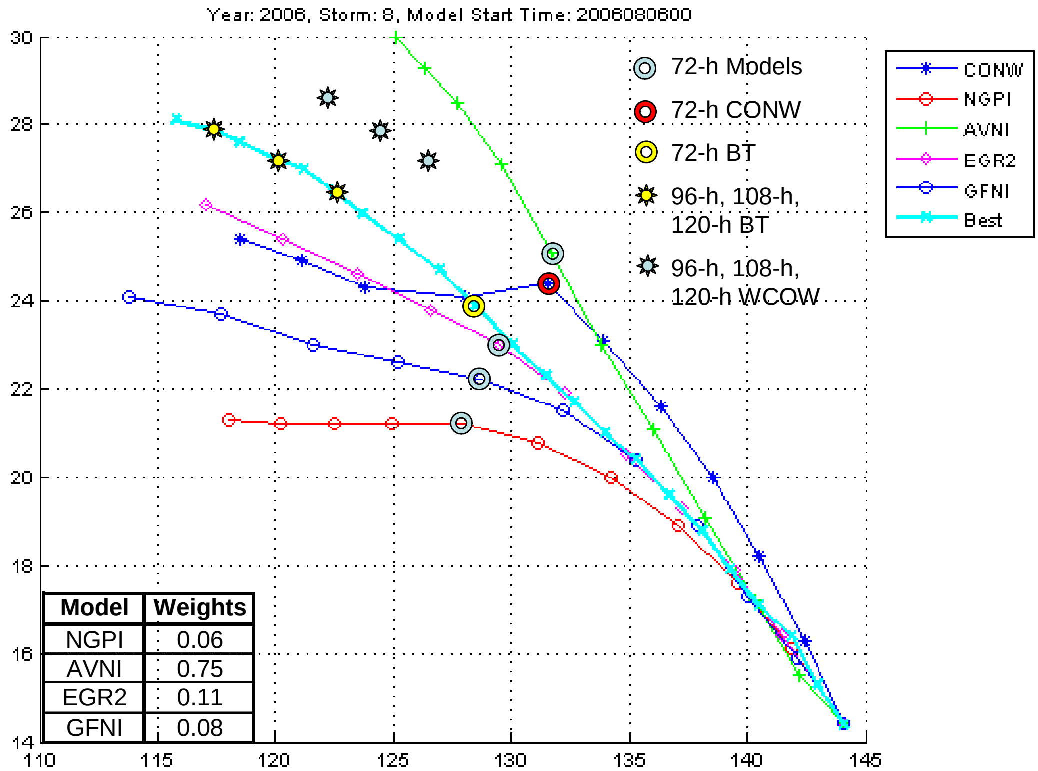


Figure 3.20 As in Figure 3.14, except for Storm 4 from 0600 UTC 01 July 2006 and with GFS interpolated track positions (AVNI, green crosses), JTYM interpolated track positions (JTYI, green diamonds).

The weighted consensus for Storm 8 from 0000 UTC 6 August 2006 resulted in large degradations from the unweighted consensus (Table 3.10). The outlier (the GFS) receives the greatest weighting, which shifted the weighted consensus to the opposite side of the BT from the CONW and much farther east of the 96-, 108-, and 120-h BT positions (note longitudinal scale) (Figure 3.21). Contrary to the previous two examples, the CONW performs poorly relative to the BT through 72 h. Thus, CONW is not a good basis for the weighting in this case. The model guidance through 72 h is biased toward the right side of the BT (Figure 3.22), which results in the CONW track being well to the right of the BT (Figure 3.21). As a result, the outlier (the GFS) is given the largest weighting factor because it is to the right of the BT near the CONW track. As a test, the CONW track was replaced by the BT positions at 60 h, 66 h, and 72 h, which resulted in weighted consensus positions that are close to the CONW track at 96, 108, and 120 h. This test indicates that the reason for such large degradations is a poor CONW track at 60 h, 66 h, and 72 h. The premise that the additional models available through 72 h add value to the 96 h, 108 h, and 120 h consensus by weighting the remaining members relative to the consensus is not valid in this case since the additional members are biased to the right of the BT (Figure 3.22) and thus degrade the consensus.

Table 3.10 As in Table 3.4, except for Storm 8 from 0000 UTC 6 August 2006.

Time	CONW error (n mi)	WCOW error (n mi)	WCOW Imp (n mi)	WCOW % Imp
96 h	149.6	215.7	-66.1	-44.2
108 h	148.2	243.5	-95.4	-64.4
120 h	161.2	293.4	-132.2	-82.0



Model	Weights
NGPI	0.06
AVNI	0.75
EGR2	0.11
GFNI	0.08

Figure 3.21 As in Figure 3.9, except for Storm 8 from 0000 UTC 6 August 2006 and with UKMO 12-h interpolated track positions (EGR2, pink diamonds). Note that the WCOW positions are farther from the best-track positions at 96 h, 108 h, and 120 h than the CONW positions since the longitudinal scale is different from the latitudinal scale.

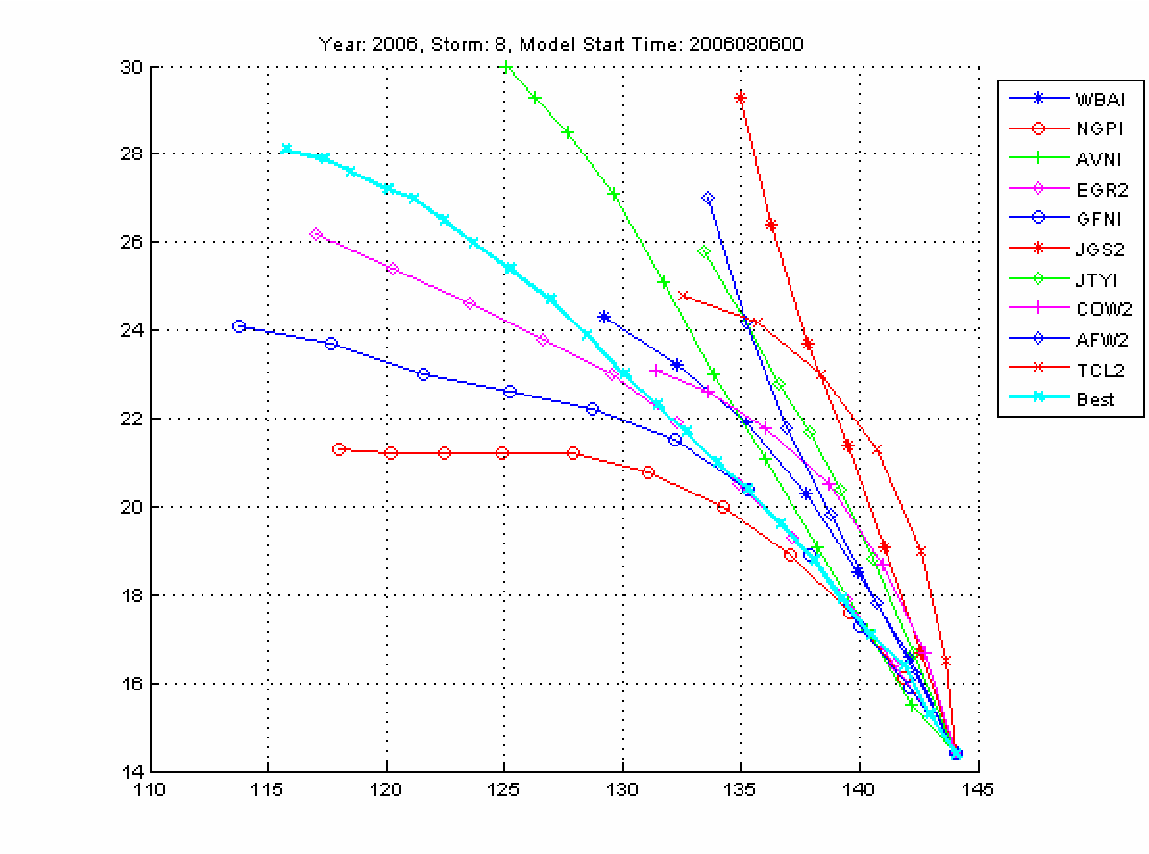


Figure 3.22 As in Figure 3.14, except for Storm 8 from 0000 UTC 6 August 2006 and with GFS interpolated track positions (AVNI, green crosses), UKMO 12-h interpolated track positions (EGR2, purple diamonds), JGSM 12-h interpolated track positions (JGS2, red asterisks), JTYM interpolated track positions (JTYI, green diamonds), COAMPS 12-h interpolated track positions (COW2, purple crosses), MM5 12-h interpolated track positions (AFW2, blue diamonds), and TC-LAPS 12-h interpolated track positions (TCL2, blue diamonds).

The track forecast of Storm 22 from 1800 UTC 2 November 2006 is another case of a degraded weighted consensus (Table 3.11). The CONW track is slow relative to the BT but the WCOW is even slower, which explains the degradations (Figure 3.23). As in the previous case, the CONW performs poorly relative to the BT through 72 h as indicated by the 72-h positions in Figure 3.23. The primary reason for the poor performance of CONW through 72 h is the contribution from the WBAR track, which is completely off course (Figure 3.24). The subsequent CONW track in Figure 3.23 reflects the influence of the WBAR with a dramatic jump forward after 72 h when the WBAR is no longer available. Without the WBAR track, the CONW would be improved and hence the weighted consensus technique should perform much better since the superior performing NOGAPS and GFS tracks would have received larger weights. In this case, the slowest and worst performing member of the weighted consensus (the GFDN) is given the largest weighting factor (0.42). In this case, it would be desirable to selectively remove the WBAR track. The unweighted consensus without WBAR should form a superior basis for the weighting factors at 60 h, 66 h, and 72 h. With the consensus shifted away from the WBAR track the GFDN would probably be assigned the smallest weighting factor, which would yield a weighted consensus that improves on the unweighted consensus since the NOGAPS and GFS tracks would receive greater weighing factors and they are superior forecasts.

Table 3.11 As in Table 3.4, except for Storm 22 from 1800 UTC 2 November 2006.

Time	CONW error (n mi)	WCOW error (n mi)	WCOW Imp (n mi)	WCOW % Imp
96 h	75.4	174.6	-99.2	-131.6
108 h	91.0	193.7	-102.7	-112.8
120 h	75.4	174.6	-99.2	-131.6

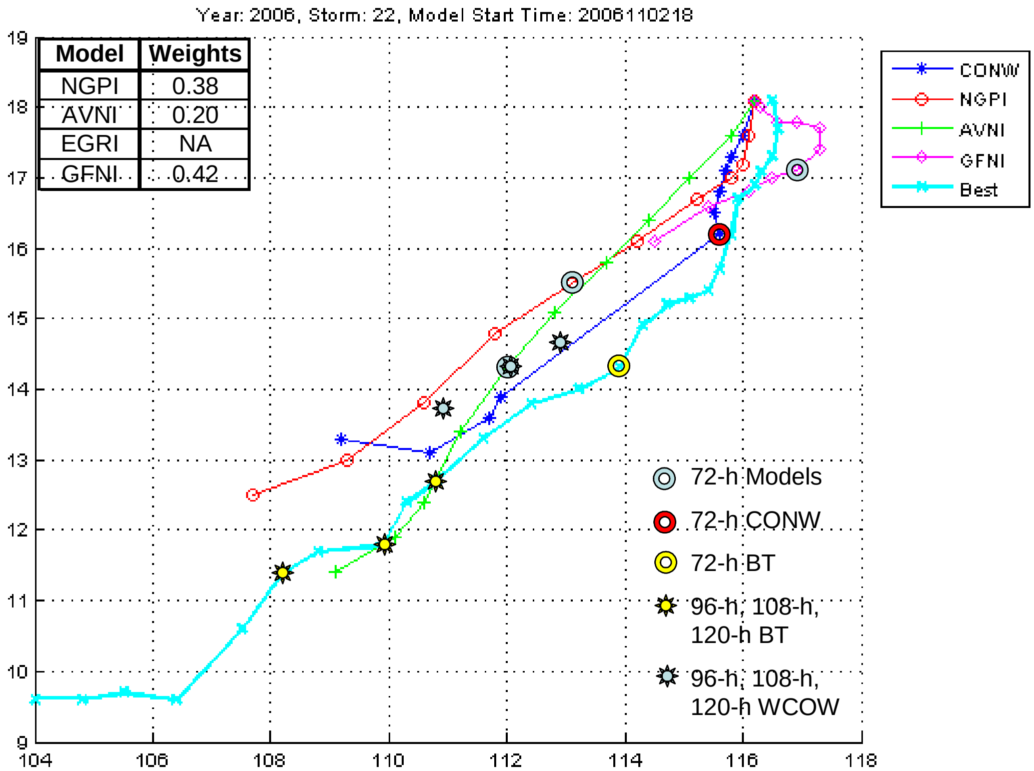


Figure 3.23 As in Figure 3.9, except for Storm 22 from 1800 UTC 2 November 2006 and without UKMO interpolated track positions.

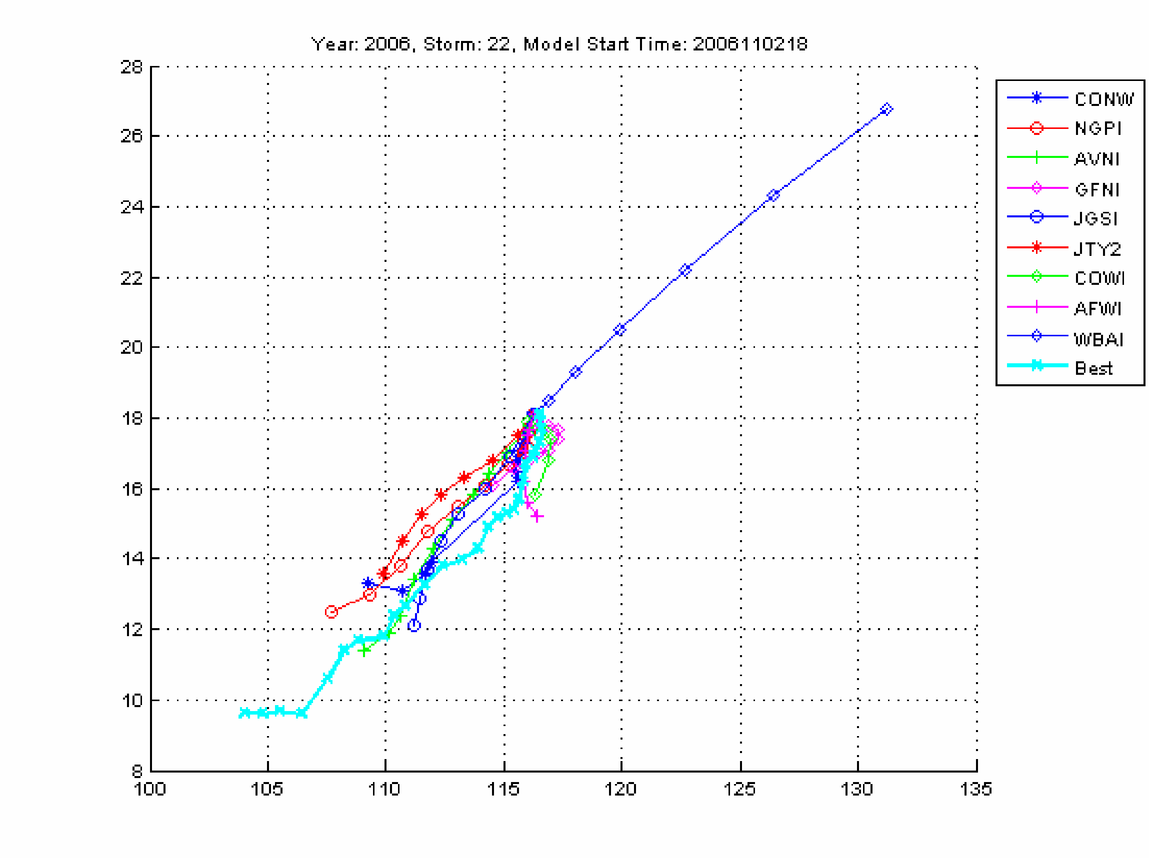


Figure 3.24 As in Figure 3.10, except for Storm 22 from 1800 UTC 2 November 2006 without UKMO interpolated track positions and TC-LAPS interpolated track positions and with GFDN interpolated track positions (GFNI, purple diamonds), JGSM interpolated track positions (JGSI, blue circles), JTYM 12-h interpolated track positions (JTY2, red asterisks), COAMPS interpolated track positions (COWI, green diamonds), MM5 interpolated track positions (AFWI, purple crosses), and WBAR interpolated track positions (WBAI, blue diamonds).

C. SENSITIVITY STUDIES

1. Removing COAMPS and MM5 from UAVE

a. Motivation and Description

The COAMPS and MM5 model tracks have recently been removed from the CONW consensus at JTWC because the inclusion of these models in the CONW has degraded the performance of CONW (B. Sampson, personal communication, February 2007). Both models are typically available at 60 h, 66 h, and 72 h when the weightings are computed, but are never available at 96 h, 108 h, and 120 h when the weightings are applied. Thus, including the MM5 and COAMPS models in the consensus (CONW or UAVE) is expected to degrade the weighting used for the weighted consensus. Removing these models from the consensus should provide improved weighting values for the remaining models, and improve the weighted consensus applied at 96 h, 108 h, and 120 h. The unweighted consensus will not change at 96 h, 108 h, and 120 h since MM5 and COAMPS are not available at these times. Thus, the weighted consensus should also improve relative to the unweighted consensus. The weighted consensus UAVE is used for this test since the input to the consensus can be easily controlled, whereas the operational consensus CONW without COAMPS and MM5 can not be reproduced exactly after the fact. COAMPS and MM5 are removed from UAVE at 60 h, 66 h, and 72 h, and the weighted consensus is re-computed.

b. Impact on 60 h, 66 h, and 72 h Error

Before looking at mean improvement, it is useful to compare the impact of removing COAMPS and MM5 on the errors at the times the weightings are computed (60 h, 66 h, and 72 h). If the errors are decreased, then an improved weighted consensus will be expected. Indeed, the average error for all cases is decreased by 4 n mi at 60 h, 66 h, and 72 h (Table 3.12). This comparison of the consensus when COAMPS and MM5 are removed re-confirms the decision that these models should be removed from the operational consensus CONW.

Table 3.12 Average UAVE errors at 60 h, 66 h, and 72 h with (Control) and without the COAMPS and MM5 (W/O C&M) models.

	UAVE Error (n mi)			**Sample**
Time	Control	W/O C & M	Difference	**Size**
60 h	111.4	107.9	3.5	222
66 h	124.3	120.6	3.7	222
72 h	138.9	135.2	3.7	222

c. *Weighted Consensus Impact*

The improved basis for calculating the weighting values from CONW without COAMPS and MM5 leads to an improved performance of the weighted consensus (Table 3.13). Improvements in the weighted consensus in terms of error reduction increased at 96 h, 108 h and 120 h. In terms of percent improvement, the increase was greater at 108 h (1.1%) and 120 h (1.2%) than at 96 h (0.7%). This leads to decreasing improvement from 96 h to 108 h (0.2%) and from 108 h to 120 h (0.2%). It is clear that the skill of the consensus used for the weighting has a large impact when such significant improvements in the weighted consensus are realized after removing COAMPS and MM5 from the consensus at 60 h, 66 h, and 72 h. These positive results also give further justification to the JTWC decision to remove MM5 and COAMPS from the operational consensus CONW.

Table 3.13 As in Table 3.1, except with (Control) and without COAMPS and MM5 (W/O C&M) models (SS: Sample Size).

	Mean Improvement		
	Control	W/O C & M	**SS**
96 h WAVE Imp (n mi):	6.2	7.8	222
96 h WAVE Imp (%):	2.8	3.5	
108 h WAVE Imp (n mi):	5.7	8.7	193
108 h WAVE Imp (%):	2.2	3.3	
120 h WAVE Imp (n mi):	5.7	9.4	168
120 h WAVE Imp (%):	1.9	3.1	

d. *Performance Graphs*

The 96-h scatter plot without the COAMPS and MM5 included in the UAVE (Figure 3.25) is quite similar to the control case (see Figure 3.4). The most noticeable difference between the two graphs is the smaller degradations (closer to reference line) of poorly performing cases (cases above the reference line) for UAVE errors less than 400 n mi. Having these cases with smaller degradations contributed to an increase in mean improvement from 2.8% to 3.5%

(Table 3.13). Removing MM5 and COAMPS from the consensus UAVE also led to an increase from 55% to 58% of the cases improved.

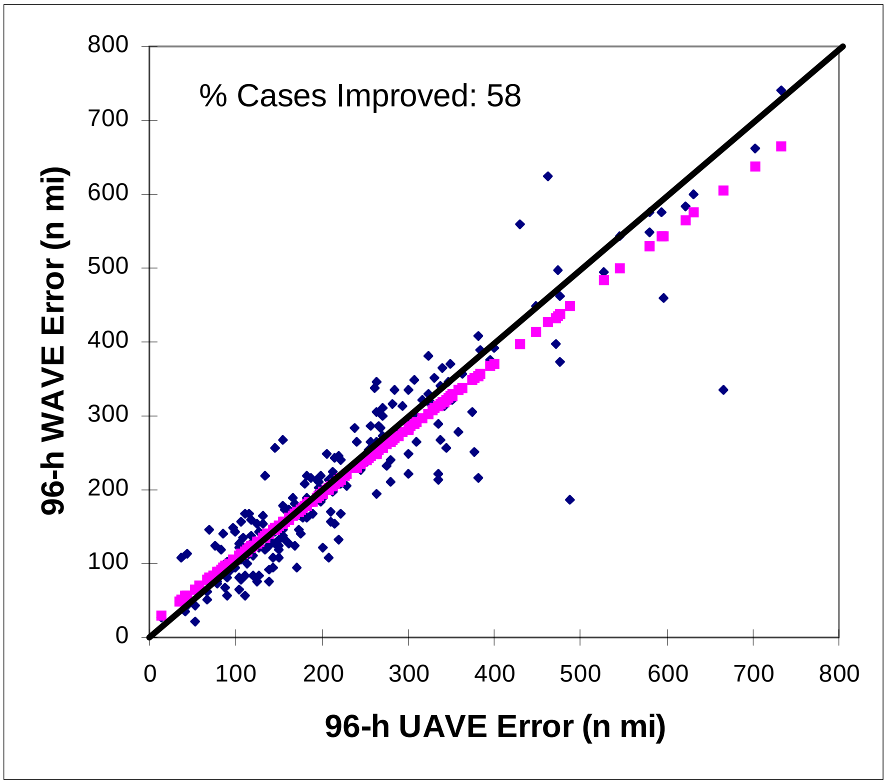


Figure 3.25 As in Figure 3.4, except for the 96-h UAVE versus WAVE errors for cases with the MM5 and COAMPS models removed from the consensus UAVE.

Again, the differences at 108 h between the control case (Figure 3.6) and the case with the removal of the COAMPS and MM5 models (Figure 3.26) are not easily detected. As at 96 h, it appears that the poorly performing cases are not degraded as much for UAVE errors less than 600 n mi. This smaller number of poorly performing cases contributed to an increase in mean improvement from 2.2% to 3.3% (Table 3.13). The percent of cases improved also increased from 53% to 56% with the removal of the MM5 and COAMPS models from the consensus. Small improvements in the weighted consensus UAVE are evidently enough to cause a shift in some cases from a degradation to an improvement.

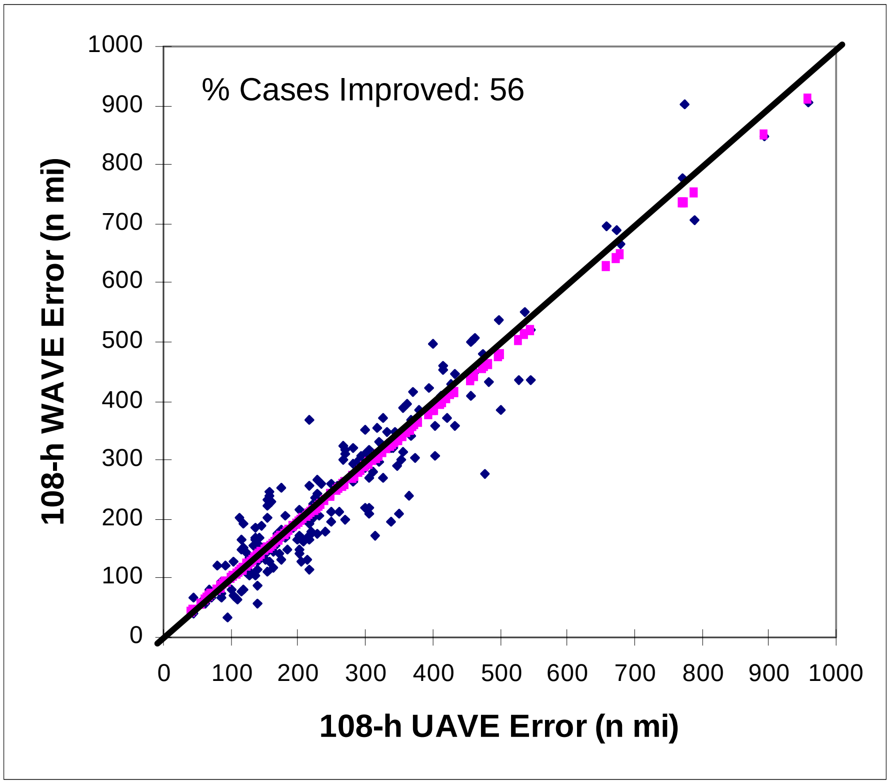


Figure 3.26 As in Figure 3.6, except for the 108-h UAVE versus WAVE errors for cases with the MM5 and COAMPS models removed from the consensus UAVE.

Although the 120-h scatter plot (Figure 3.27) without the COAMPS and MM5 models in the consensus UAVE is again similar to the control (Figure 3.8), an increase in mean WAVE improvement at 120 h from 1.9% to 3.1% is achieved (Table 3.13). It is evident that the degraded cases for UAVE errors less than 600 n mi are less frequent. Now more than half (51%) of the cases are improved, whereas in the control only 49% of the cases are improved.

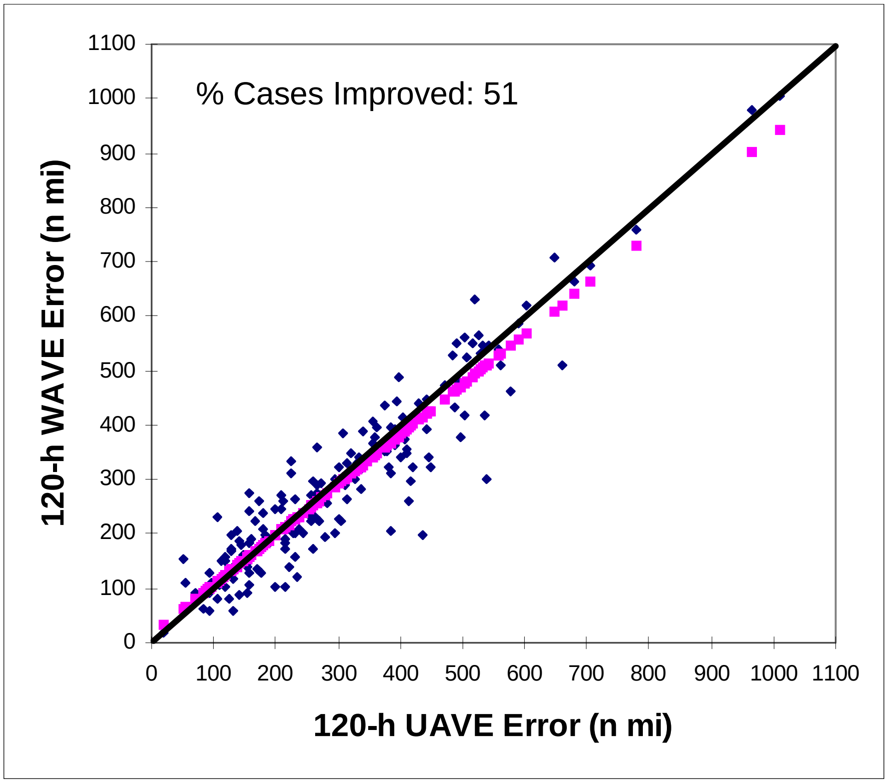


Figure 3.27 As in Figure 3.8, except for the 120-h UAVE versus WAVE errors for cases with the MM5 and COAMPS models removed from the consensus UAVE.

2. Weighting Optimization – Adding JGSM to weighted consensus

a. Motivation and Description

The initial validation study in Chapter III.A above considered that only four models (NOGAPS, GFS, UKMO, GFDN) were available for the weighted consensus at 96 h, 108 h, and 120 h. However, it was discovered that the JGSM occasionally was still available at 96 h. Since the JGSM is skillful, its inclusion in the unweighted consensus CONW (or UAVE) should improve it on average. On the 38 occasions when the JGSM was available at 96 h, the unweighted consensus CONW or UAVE gained the benefit of its skill. According to Buck Sampson (personal communication 2007), who is the developer of CONW, if a 96-h JGSM position is available, it will automatically be used in

CONW. The objective of this section is to demonstrate that inclusion of another skillful model (JGSM) at 96 h also contributes to an improved weighted consensus WCOW or WAVE.

b. Weighted Consensus Impact

In this test, the only modification is to include JGSM in the weighting scheme at 96 h. Both WCOW and WAVE are evaluated for this test (Table 3.14). The impact should be similar for WCOW and WAVE since neither included JGSM in the weighting scheme in the control case. In the control, the improvement in the WCOW is relatively small because it is being compared with the CONW that has the JGSM included (now five models) because only four models are used.

Cntrl: CONW Error (**5** models) – WCOW Error (orig. **4** models) = WCOW Imp.

In the test case, WCOW and CONW both include five models so it is expected that the errors should decrease and the WCOW improvement should increase.

Test: CONW Error (**5** models) – WCOW Error (**5** models) = WCOW Imp.

Since similar dramatic improvements are achieved in the 38 test cases using both WCOW and WAVE (Table 3.14, left side), the weighted consensus technique is still able to improve on the unweighted consensus in the control cases despite the weighted consensus technique having one less model for guidance. When the improvements are spread over all 222 cases (Table 3.14, right side), the improvements are not as great, but are still positive considering only 17 percent of the cases were updated to include JGSM.

It is important to realize that the dramatic improvements here are not simply a result of adding a skillful model JGSM. Including the same model guidance in both the unweighted and weighted consensus allows the weighted consensus to perform much better relative to the unweighted consensus. It is true that any additional model needs to be skillful for this improvement to be achieved by the weighted consensus. If addition of a model instead degraded the unweighted consensus, including it in the weighted consensus should not improve it. This sensitivity study clearly shows the importance of including all available skillful guidance in the weighting scheme, even if it is only available a fraction of the time.

Table 3.14 Mean improvements in the weighted consensus techniques WCOW and WAVE without (Control) and with the JGSM included for just those cases that JGSM was available at 96 h (left side) and all cases (right side) (SS: Sample Size).

	Cases weighted with JGSM			**All Cases**		
Mean Improvement	**Control**	**W/ JGSM**	**SS**	**Control**	**W/ JGSM**	**SS**
96 h WCOW Imp (n mi):	1.2	8.7	38	6.0	7.3	222
96 h WCOW Imp (%):	0.5	3.6		2.7	3.3	
96 h WAVE Imp (n mi):	0.5	7.5	38	6.2	7.4	222
96 h WAVE Imp (%):	0.2	3.1		2.8	3.4	

c. Performance Graphs

The 96-h scatter plot after including JGSM in the WCOW weighting (Figure 3.28 right side) is similar to the plot for the control case (Figure 3.28 left side). Despite the similarity, there are a few notable cases where adding the JGSM to the weighting yields significant improvements. One poorly performing case with a CONW error of about 150 n mi reduces the WCOW error from near 350 n mi to 300 n mi after the JGSM is included in the weighting. Another outlier with a CONW error of about 450 n mi has a WCOW error near 530 n mi that is reduced to around 500 n mi. A case with an CONW error of around 300 n mi has a WCOW error that is reduced by over 100 n mi to about 30 n mi. These better performing cases contributed to an increase in mean improvement from 0.5% to 3.6% for those cases weighted with JGSM (Table 3.14, left side). The percent of cases improved also increased from 58% to 61% for the cases weighted with JGSM (Figure 3.28).

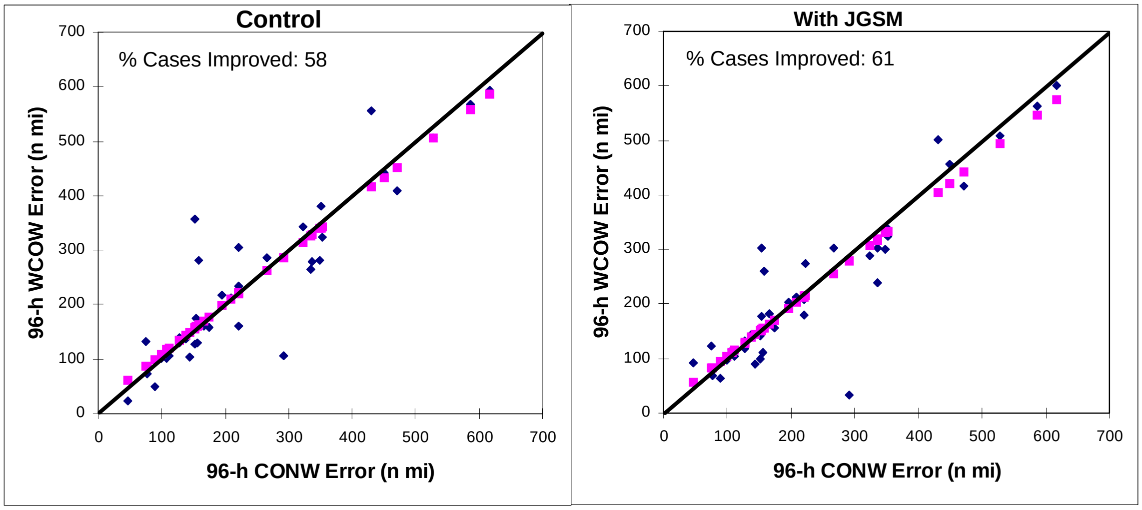


Figure 3.28 As in Figure 3.3, except without JGSM included in the weighting (control) on the left side and with JGSM included in the weighting on the right side, and just for those cases that JGSM was available at 96 h.

A similar increase in mean improvement was achieved with weighted consensus WAVE (Figure 3.29) as for WCOW (Figure 3.28) for the subset of cases weighted with the JGSM from 0.2% in the control to 3.1% (Table 3.14) when JGSM was included in the weighting. A large 6% increase was found in the number of cases improved (Figure 3.29). In contrast to the CONW cases in Figure 3.28, the UAVE cases are closely aligned with the reference line, which indicates less spread in the weighted consensus performance for UAVE than for CONW. Thus, more cases are near the reference line with the UAVE control cases in Figure 3.29 (left side) than in the CONW control cases in Figure 3.28 (left side). Even small improvements are more likely to displace cases below the reference line and change them from being degraded cases to being improved cases, which explains the greater increase in the percent of cases improved for UAVE (6%) than for CONW (3%).

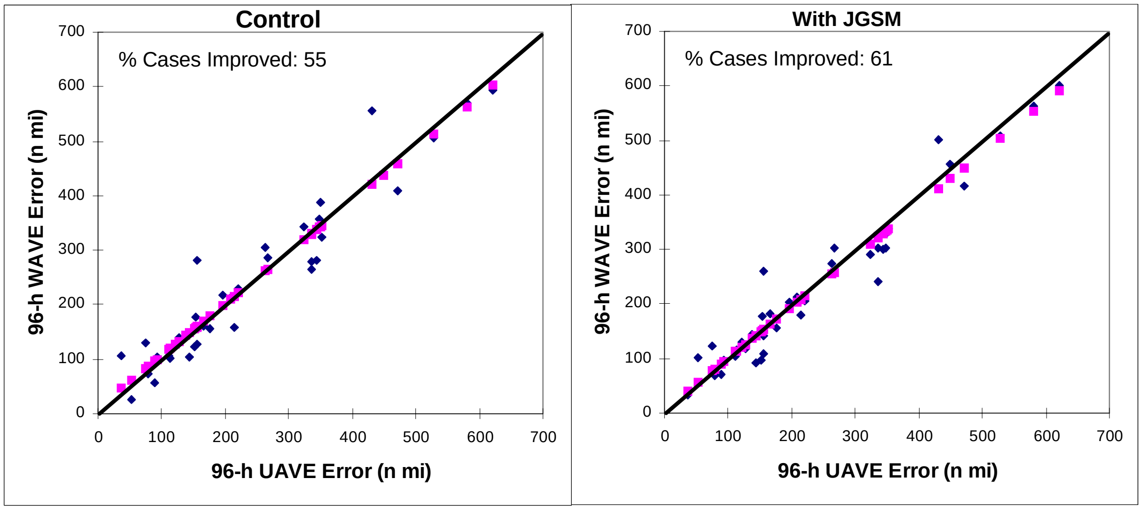


Figure 3.29 As in Figure 3.28, except for 96-h UAVE versus WAVE errors.

3. Impact of Skillful Model – ECMWF

a. Motivation and Description

In the validation study in Chapter III.A, the ECMWF tracks were not included in the simple consensus UAVE and were not included in calculating the weighting factors for either the CONW or the UAVE. Since the ECMWF became available during the 2006 season, an opportunity was available for a sensitivity study on the impact of including these ECMWF tracks. As a first demonstration, B. Sampson of the Naval Research Lab - Monterey provided a comparison of the JTWC official forecast, ECMWF, and a revised consensus (TESB) that included the ECMWF and excluded MM5 and COAMPS (Table 3.15). Note that B. Sampson assumed the ECMWF tracks would be available at JTWC with a time delay of less than 6 h so that the interpolated track was matched with the 6-h warning position. Since it is likely that the ECMWF tracks will not be available for use until 12 h after synoptic times, the ECMWF errors in Table 3.15 are probably smaller than will be achieved operationally.

Table 3.15 Western North Pacific 2006 season average track errors (n mi) for a homogeneous sample of JTWC forecast (JTWC), 6-hour interpolated ECMWF tracks (ECMI), and an experimental consensus (TESB) that is CONW with the ECMWF tracks and without the MM5 and COAMPS tracks (provided by B. Sampson, personal communication, 2007).

Average Track Errors (n mi) for Homogeneous Sample (2006)								
	00 h	12 h	24 h	36 h	48 h	72 h	96 h	120 h
JTWC	10.5	39.7	67.7	93.8	118.3	160.7	225.5	320.8
ECMI	10.6	54.0	77.2	101.0	120.9	164.8	195.3	240.2
TESB	10.6	36.2	58.9	79.6	97.2	134.0	188.2	277.2
# Cases	289	277	261	244	221	169	120	90

The ECMWF track errors were larger than both the JTWC and TESB errors through 72 h. Except for the 12-h ECMWF error, the errors at other times through 72 h are not unreasonably large considering ECMWF is only one model versus the multiple models used for TESB and the guidance used for the JTWC forecasts. By 96 h, the ECMWF errors are smaller than for JTWC, and by 120 h the ECMWF errors are smaller than the JTWC and TESB errors. It is concluded that ECMWF will add value to the long-range forecasts when it is added to the consensus.

It is also expected that the ECMWF will add value to the weighted consensus when it is included in the consensus and used for the weighting. The simple unweighted consensus UAVE is used to test this impact by adding ECMWF to the consensus at 60 h, 66 h, 72 h, 96 h, 108 h, and 120 h and including ECMWF in the calculation of the weights to be applied at 96 h, 108 h, and 120 h. Including ECMWF in UAVE at 60 h, 66 h, and 72 h should improve the weighting factors because ECMWF should improve the consensus used for the weighting. Based on Table 3.15, including ECMWF in UAVE at 96 h, 108 h, and 120 h should improve the unweighted consensus UAVE even though the interpolated ECMWF track (ECMI) is operationally compared with the 12-h warning position rather than the 6-h warning position as assumed by B. Sampson in creating Table 3.15.

Including the ECMWF in the simple consensus UAVE reduced the 96-, 108-, and 120-h errors significantly in the 59, 55, and 52 cases, respectively (left side of Table 3.16). This result was expected from Table 3.15, since the

ECMWF performs quite well for the extended forecasts (96 h and 120 h). A smaller but positive impact is achieved over the entire sample of cases considered in this study (right side of Table 3.16), which again verifies that adding ECMWF to the consensus does improve it. Thus, the ECMWF tracks should definitely be included in the unweighted consensus even with a 12-h time delay.

Table 3.16 UAVE errors (n mi) without (Control) and with ECMWF included for just those cases that ECMWF tracks were available (left side) and all cases (right side) (SS: Sample Size).

	UAVE Error (n mi)					
	ECMWF Cases			**All Cases**		
	Control	ECMWF	SS	Control	ECMWF	SS
96 h	205	181	59	221	214	222
108 h	251	221	55	262	253	193
120 h	293	262	52	302	293	168

b. Weighted Consensus Impact

For the particular set of 59, 55, and 52 cases at 96, 108, and 120 h, respectively, for which the ECMWF tracks were available, the weighted consensus WAVE performed particularly well in the control with mean error improvements of 9.2 n mi, 10.4 n mi, and 13.6 n mi, respectively (left side of Table 3.17). The corresponding percentage improvements of 4.5%, 4.2%, and 4.7%, respectively, were larger than were achieved by excluding the MM5 and COAMPS (Table 3.13) or including JGSM at 96 h (Table 3.14), which reflects that the weighted consensus will perform better for some sets of tracks than other sets. The unexpected result was that the inclusion of the ECMWF tracks in the WAVE did not further improve on the WAVE without the ECMWF (Table 3.17, left side) for this particular set of cases. However, the WAVE with the ECMWF included does result in significant improvements of 7.8 n mi, 8.7 n mi, and 11.2 n mi at 96 h, 108 h, and 120 h, respectively, relative to the unweighted average used operationally. When the entire sample of cases was considered (Table 3.17, right side), the inclusion of the ECMWF tracks in the WAVE slightly degraded the performance relative to WAVE control case, although the percentage differences are not significant.

Table 3.17 Mean improvements in the weighted consensus WAVE without (Control) and with ECMWF included for just those cases that ECMWF tracks were available (left side) and all cases (right side) (SS: Sample Size).

	ECMWF Cases			All Cases		
	Control	ECMWF	SS	Control	ECMWF	SS
96 h WAVE Imp (n mi):	9.2	7.8	59	6.2	5.8	222
96 h WAVE Imp (%):	4.5	4.3		2.8	2.7	
108 h WAVE Imp (n mi):	10.4	8.7	55	5.7	5.2	193
108 h WAVE Imp (%):	4.2	3.9		2.2	2.1	
120 h WAVE Imp (n mi):	13.6	11.2	52	5.7	4.5	168
120 h WAVE Imp (%):	4.7	4.3		1.9	1.5	

The explanation for why the inclusion of the ECMWF tracks did not further improve the WAVE over the unweighted consensus UAVE is sought in Tables 3.15 and 3.16. Note in Table 3.15 that the average 48-h and 72-h ECMWF errors are large compared to the errors for the consensus TESB, which is the experimental CONW without the MM5 and COAMPS but does include the ECMWF. Although these 48-h and 72-h ECMWF errors do represent skill in track forecast relative to a Climatology and Persistence track, the ECMWF tracks are less skillful than the consensus TESB. However, the ECMWF tracks are particularly skillful at 96 h and 120 h (Table 3.15), and are even more skillful than the TESB at 120 h, for this particular sample of cases. Thus, the inclusion of these highly skillful ECMWF forecasts in the unweighted consensus UAVE results in dramatic reductions in errors at 96 h, 108 h, and 120 h (Table 3.16). Thus, the explanation for the lack of further improvement in WAVE in Table 3.17 is that the inclusion of the relatively less skillful 60-h through 72-h ECMWF forecasts in the consensus WAVE that is used to calculate the weighting factors does not give as good weighting factors as an equal weighting would give (i.e., the control UAVE in Table 3.17). That is, the highly skillful 96 h, 108 h, and 120 h ECMWF tracks are being given less weight in this sample of cases because the weighting factors are dependent on the 60 h, 66 h, and 72 h positions that are not as skillful as the experimental consensus TESB. This deficiency needs to be examined with a larger sample of cases. If found to be a general result, then the weighting factors will need to be re-calculated at later times rather than be fixed at the values derived for the 60 h, 66 h, and 72 h positions.

c. Performance Graphs

For just the set of cases in which ECMWF tracks were available, there is not a large difference in spread normal to the reference line before the ECMWF is included (Figure 3.30, left side) and after the ECMWF is included (Figure 3.30, right side). That is, the improved and degraded cases are distributed similarly in both cases. The major difference is a shift towards the origin along the reference line after the ECMWF is included (Figure 3.30, right side), which indicates a general reduction in UAVE and WAVE errors. This is consistent with the dramatic reductions in UAVE error for the ECMWF cases from 205 n mi to 181 n mi (Table 3.16). A less than one percent reduction in the percent mean improvement is found when ECMWF is included, which indicates that the WAVE errors are not reduced as much as the UAVE errors when the ECMWF is included. For the cases for which the ECMWF is available, the percent of cases improved does not change from the inclusion of the ECMWF.

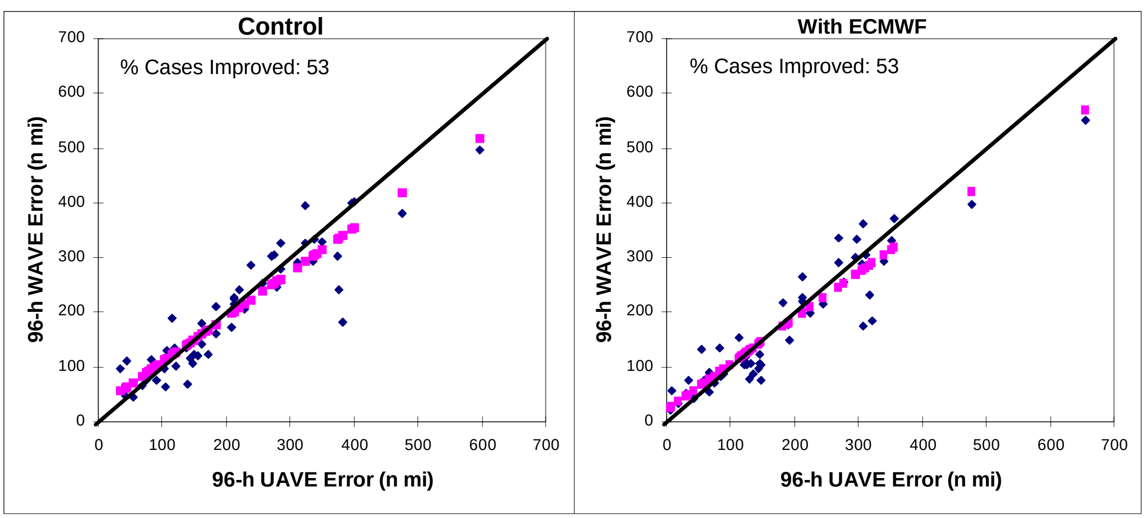


Figure 3.30 As in Figure 3.4, except without ECMWF (control) on the left side and with ECMWF on the right side, and just for those cases that ECMWF was available.

At 108 h (Figure 3.31) as for the 96-h UAVE, a discernable shift in the cases toward the origin along the reference line is found when the ECMWF is included, which again indicates a reduction in both the UAVE and WAVE errors. A comparable decrease in mean improvement (0.3%) is found as at 96 h. In contrast to 96 h the percent of cases improved at 108 h is increased from 56% to 60%.

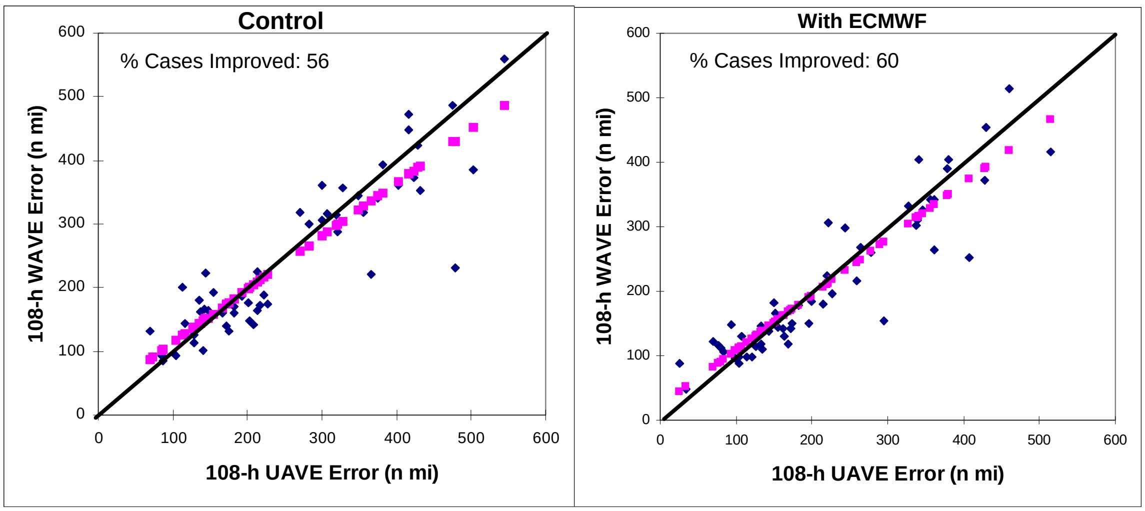


Figure 3.31 As in Figure 3.30, except for 108-h UAVE versus WAVE errors.

As at 96 h and 108 h, a general reduction in UAVE and WAVE error is found at 120 h (Figure 3.32), and a general reduction in spread from the reference line occurs when the ECMWF is included. Even though the percent of cases improved is increased, the weighted consensus improves over the unweighted consensus in the control by 0.4% more than after the ECMWF is included (Table 3.17). Perhaps the control case with an UAVE error around 540 n mi and WAVE error around 240 n mi (Figure 3.32 left side) contributes to the superior performance of the control.

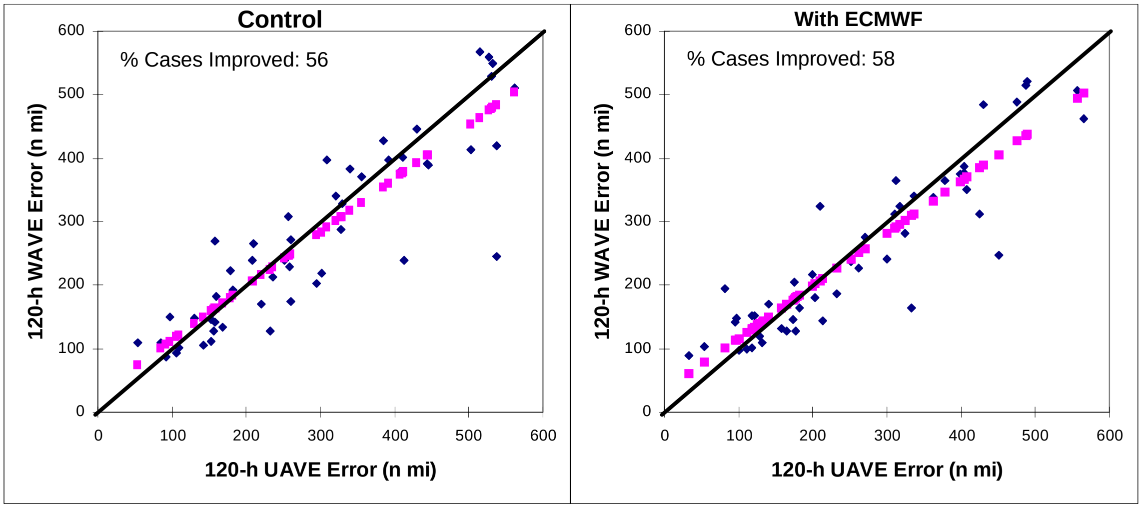


Figure 3.32 As in Figure 3.30, except for 120-h UAVE versus WAVE errors.

IV. WEIGHTED MOTION VECTOR CONSENSUS

A. METHODOLOGY

The same data set, assumptions, and weighting factors are used for the weighted motion vector consensus study as for the weighted consensus validation study in Chapter III.A. The unweighted consensus UAVE is used to validate the weighted motion vector consensus, which avoids the irregularities from the occasional unexplained guidance in CONW. The UAVE includes tracks from all available models except the ECMWF in the priority defined in Table 2.1. Additionally, the weights for this first validation study do not include the ECMWF or JGSM tracks, which is consistent with the formulation of WAVE in Chapter III.A.

A weighted motion vector consensus WVAE is determined at 96 h, 108 h, and 120 h starting from the UAVE position at 84 h. The 96-h WVAE latitude is calculated as

96-h WVAE lat = 84-h UAVE lat + NOGAPS weight * (96-h NOGAPS lat – 84-h NOGAPS lat) + GFS weight * (96-h GFS lat – 84-h GFS lat) + ……..

The 108-h WVAE latitude is calculated in a similar manner except starting with the 96-h WVAE as

108-h WVAE lat = 96-h WVAE lat + NOGAPS weight * (108-h NOGAPS lat – 96-h NOGAPS lat) + GFS weight * (108-h GFS lat – 96-h GFS lat) + ……..

The 120-h WVAE latitude is calculated analogous to the 108-h WVAE latitude. Similarly, the WVAE longitude is calculated in the same manner as latitude.

An unweighted motion vector consensus UVAE is also calculated to determine the value of the weighting in the weighted motion vector consensus. The UVAE is calculated in the same manner as the WVAE except that equal weights are used to multiply each model motion vector. The weighted motion vector consensus WVAE and the unweighted motion vector consensus UVAE are compared to the unweighted position consensus UAVE to assess the improvement gained from each consensus.

B. RESULTS

1. Weighted Motion Vector Consensus Impact

The weighted motion vector consensus WVAE achieves remarkable improvements over the unweighted position consensus UAVE at 96 h, 108 h, and 120 h (Table 4.1). The unweighted motion vector consensus UVAE improvements are also large, which indicates the value of the motion vector approach over the weighted position approach. The WVAE further improves on the UVAE at 96 h (1.5%), 108 h (1.3%), and 120 h (1.5%), which justifies the use of a weighted motion vector consensus. However, most of the improvements are achieved by simply using a motion vector consensus instead of a position consensus.

Table 4.1 Mean improvement of the weighted position consensus (WAVE) and the weighted (WVAE) and unweighted (UVAE) motion vector consensus over unweighted position consensus UAVE for the 2006 season (Storms 1 through 24) validation cases (SS: Sample Size).

	WAVE	WVAE	UVAE	SS
96 h Imp (n mi):	6.2	21.9	18.5	222
96 h Imp (%):	2.8	9.9	8.4	
108 h Imp (n mi):	5.7	19.7	16.2	193
108 h Imp (%):	2.2	7.5	6.2	
120 h Imp (n mi):	5.7	17.0	12.5	168
120 h Imp (%):	1.9	5.6	4.1	

The WVAE performs best at earlier forecast times with a near 10% improvement over the UAVE at 96 h, which is then reduced to a still significant 5.6% improvement at 120 h. Since the weighting factors are determined at 60 h, 66 h, and 72 h, this decline in improvement might be expected. Notice the unweighted motion vector consensus UVAE also decreases in performance with increasing forecast interval with an 8.4% improvement at 96 h and a 4.1% improvement at 120 h.

This decrease in improvement could be due to the nature of the motion vector consensus or simply because it is not the identical data set with 222 cases at 96 h and only 168 cases at 120 h. When only the 168 cases that are available through 120 h are evaluated, a decrease in the UVAE skill still remains in terms of percent improvement from 8.2% at 96 h to 4.1% at 120 h (Table 4.2), which suggests that there is an inherent decrease in the skill of the motion vector consensus with increasing forecast time. Surprisingly the improvement for the weighted position consensus WAVE increases from 1.9% at 96 h to 3.0% at 108 h for the 168 cases, which is contrary to the decreasing trend for the data set in Table 4.1.

Table 4.2 As in Table 4.1 except for only the 168 cases available through 120 h (SS: Sample Size).

	WAVE	WVAE	UVAE	SS
96 h Imp (n mi):	3.7	19.1	16.0	168
96 h Imp (%):	1.9	9.8	8.2	
108 h Imp (n mi):	7.4	21.7	17.2	168
108 h Imp (%):	3.0	8.9	7.0	
120 h Imp (n mi):	5.7	17.0	12.5	168
120 h Imp (%):	1.9	5.6	4.1	

2. Performance Graphs

Comparing Figure 4.1 with Figure 3.4, it is evident that the weighted motion vector consensus WVAE markedly improves over the WAVE at 96 h. An overall downward shift of the WVAE errors relative to the WAVE errors indicates the further improvement in performance relative to UAVE. It is noteworthy that an increase the percent of cases improved increases from 55% for the WAVE to 67% for the WVAE. The best-fit line has smaller slope for WVAE (Figure 4.1) than for WAVE (Figure 3.4), which again indicates an improvement for WVAE relative to WAVE. The WVAE performs particularly well for large UAVE error cases, with no WVAE degradations for UAVE errors greater than about 450 n mi as occurred with the WAVE (Figure 3.4). Despite the overall improvement of the WVAE over the WAVE, one case of note with an UAVE error of about 500 n mi performs much better with WAVE (about a 160 n mi error) than for WVAE (about a 300 n mi error), which indicates that on a case-by-case basis the WVAE is not always superior to the WAVE.

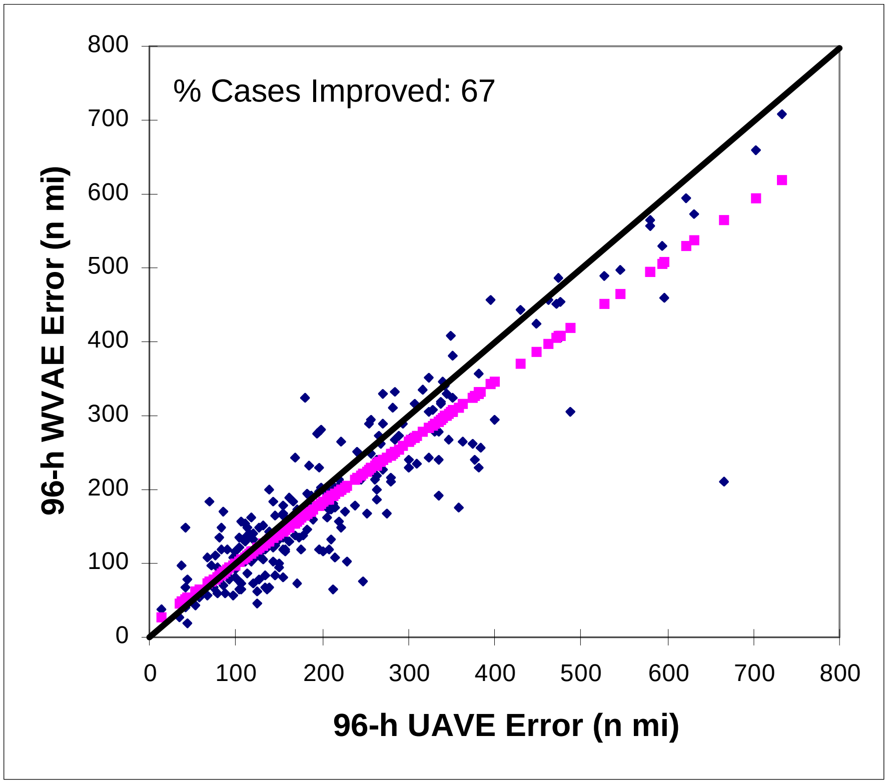


Figure 4.1 As in Figure 3.3, except for 96-h UAVE versus WVAE errors.

Again at 108 h, the WVAE (Figure 4.2) has dramatic improvements relative to the WAVE (Figure 3.6). The associated increase in the percent of cases is improved from 53% for the WAVE to 65% for the WVAE. One noteworthy degradation occurs for a UAVE error of about 150 n mi from less than 220 n mi for the WAVE (Figure 3.6) to about 375 n mi for the WVAE.

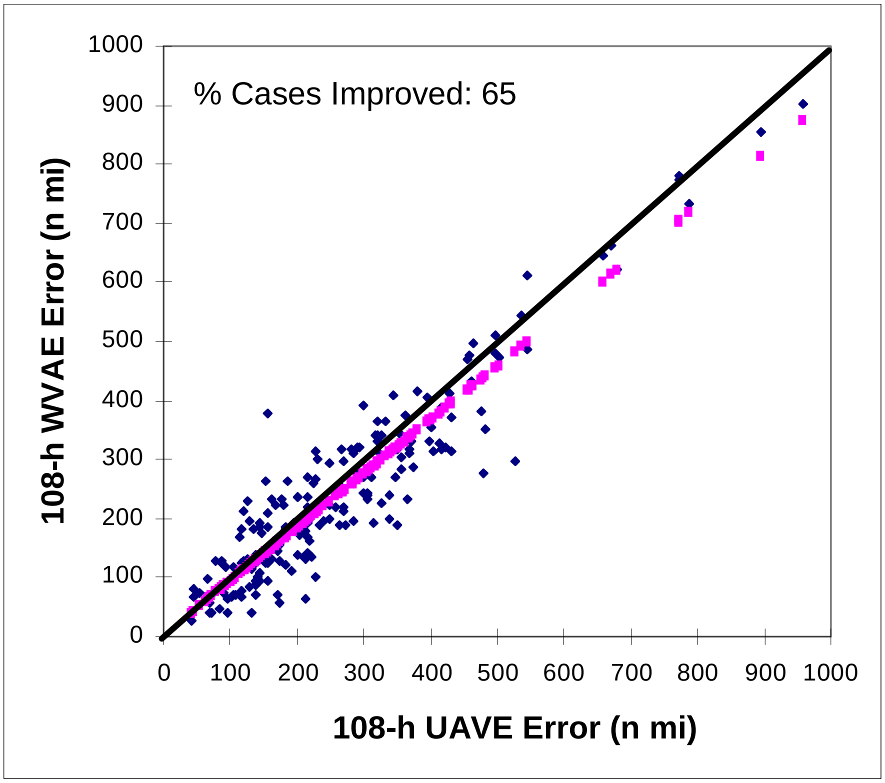


Figure 4.2 As in Figure 4.1, except for 108-h UAVE versus WVAE errors.

The 120-h WVAE (Figure 4.3) performs comparably to the 96-h and 108-h WVAE with dramatic improvements over the 120-h UAVE (Figure 3.8). The WVAE has an average improvement of 5.6% compared to a 1.9% improvement for the UAVE (Table 4.1). It is noteworthy that the percent of cases improved increases from 49% for WAVE to 63% for WVAE.

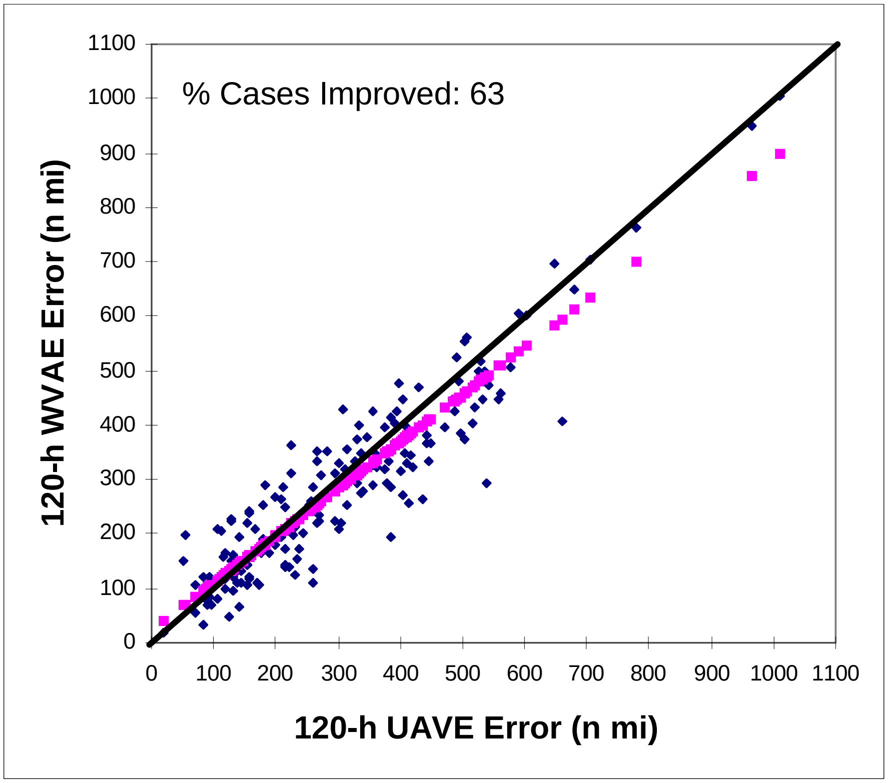


Figure 4.3 As in Figure 4.1, except for 120-h UAVE versus WVAE errors.

3. Case Studies

Some case studies are examined to help understand why the weighted motion vector consensus WVAE can yield much greater improvements over UAVE than the weighted position consensus WAVE. Favorable cases are examined, which are cases in which the WVAE performs better than the WAVE. Although on average the WVAE performs better than the WAVE, a few unfavorable cases are examined, which are cases in which the WAVE performs better than the WVAE.

a. Favorable Cases

Storm 6 from 0000 UTC 20 July 2000 is a case in which the WVAE track improves significantly on the UAVE track while the WAVE track only slightly improves on the UAVE track (Table 4.3). Both the WAVE and WVAE tracks are shifted north of the UAVE 96-h, 108-h, and 120-h positions moving them closer to the BT latitudes (Figure 4.4). The WAVE (WVAE) track is shifted to the east (west) of the UAVE 96-h, 108-h, and 120-h positions, which moves the WAVE

(WVAE) track farther from (closer to) the BT longitude. This case demonstrates the ability of the motion vector consensus to increase the along-track component in the WVAE consensus relative to a WAVE position consensus. In the position consensus WAVE (and the UAVE), the GFDN track slows the consensus track since the GFDN position are so far to the southwest. For the motion vector consensus WVAE, the northward component of the GFDN track vector contributed to a favorable northward translation. In the case of the position consensus WAVE (and the UAVE), the GFDN track hindered a northward translation since the GFDN track is the farthest south from 96 h – 120 h. The westward component of the GFDN track vector has a favorable contribution to the westward translation from 84 h – 96 h, but it has an unfavorable eastward component from 96 h – 108 h and 108 h – 120 h. The WVAE is an improvement on the UAVE despite the unfavorable eastward translations in the GFDN track since the GFDN has the smallest weighting factor (0.15; inset of Figure 4.4). The GFS (which has the greatest weighting factor of 0.53) track vector component contributes favorably to the westward translation of the WVAE track from 84 to 120 h. In the position consensus WAVE, the GFS track contributes unfavorably to a desired westward translation since the GFS track positions are farther east than the other two consensus members (NOGAPS and GFDN) from 96 h – 120 h, with the exception of the GFDN at 120 h.

Table 4.3 UAVE, WAVE, and WVAE errors (n mi) and improvement (n mi and percent) of the weighted consensus WAVE and the weighted motion vector consensus WVAE relative to the unweighted consensus UAVE for Storm 6 from 0000 UTC 20 July 2006.

Time	UAVE error	WAVE error	WAVE Imp	WVAE error	WVAE Imp
96 h	252.4	245.3	7.1	167.4	84.9
108 h	283.1	275.8	7.4	196.2	87.0
120 h	378.0	370.9	7.1	293.5	84.6

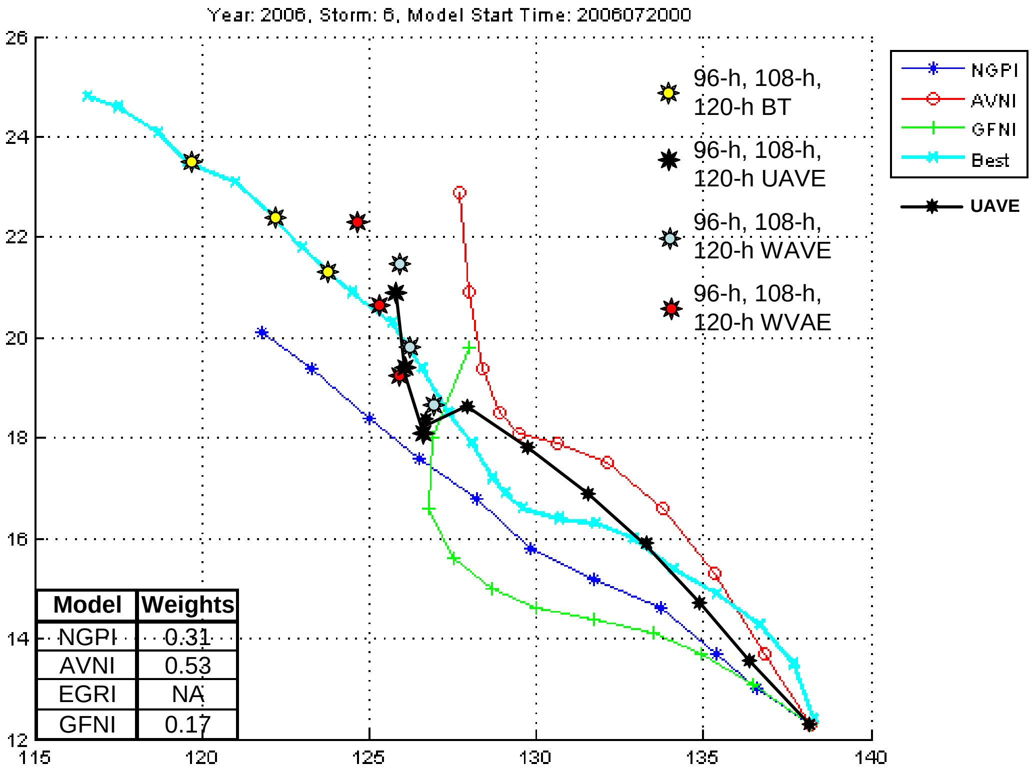


Figure 4.4 Forecast track positions each 12 h for Storm 6 from 0000 UTC 20 July 2006 with NOGAPS interpolated track positions (NGPI, blue asterisks), GFS interpolated track positions (AVNI, red circles), GFDN interpolated track positions (GFNI, green crosses), UAVE track positions (black stars), and Best-Track positions every 6 h (Best, light blue crosses). The 96-, 108-, and 120-h positions are highlighted by stars for the Best-Track (yellow), UAVE (black), WAVE (grey), and WVAE (red).

Storm 24 from 1200 UTC 29 November is a case in which the WVAE track significantly improves on the UAVE track while the WAVE track is significantly degraded relative to the UAVE track (Table 4.4). As in the previous case, the WVAE track is advanced in the along-track direction relative to the UAVE and WAVE tracks, which moves the WVAE track closer to the BT (Figure 4.5). The WAVE track is degraded relative to the UAVE track since the GFDN track receives the lowest weighting factor of 0.10. Although the northward translation of the GFDN is unfavorable, the GFDN is the closest model track to the BT longitude at 96 h, 108 h, and 120 h. The low weighting factor for the

GFDN causes the WAVE track to shift to the east relative to the UAVE track, which is away from the BT and thus causes degradation. The WVAE is able to improve on the UAVE by translating the favorable westward progression of all the models from a single position at 84 h, 96 h, and 108 h, which results in more westward translation of the WVAE track from 84 h to 120 h and a track that is closer to the BT than the UAVE at 96 h, 108 h, and 120 h.

Table 4.4 As in Table 4.3, except for storm 24 from 1200 UTC 29 November 2006.

Time	UAVE error	WAVE error	WAVE Imp	WVAE error	WVAE Imp
96 h	238.7	285.1	-46.4	179.4	59.3
108 h	270.3	317.3	-47.0	211.3	59.0
120 h	340.4	383.6	-43.1	278.6	61.8

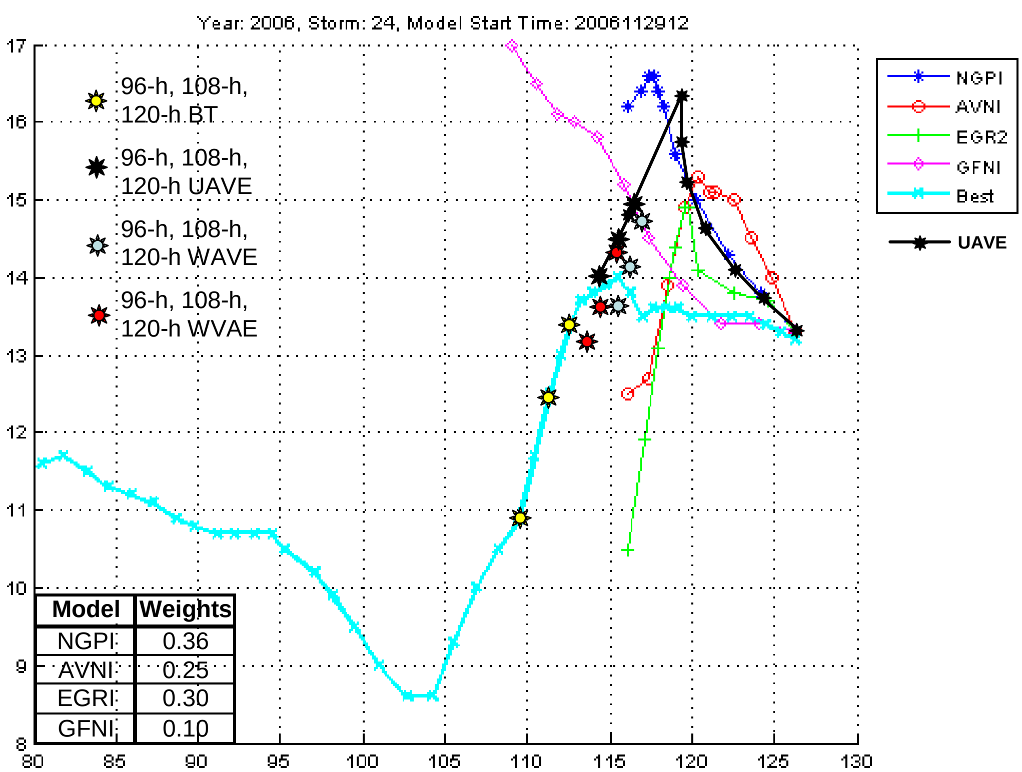


Figure 4.5 As in Figure 4.4, except for Storm 24 from 1200 UTC 29 November 2006 and with UKMO 12-h interpolated track positions (UKM2, green crosses), and GFDN interpolated track positions (GFNI, purple diamonds).

b. Unfavorable Cases

Whereas the weighted motion vector consensus WVAE performs unfavorably relative to the UAVE for Storm 8 from 1800 UTC 5 August 2006, the weighted position consensus WAVE improved slightly on the UAVE (Table 4.5). Although the WAVE track is farther from the BT than the unweighted position consensus UAVE in the cross-track direction, it is much closer in the along-track direction, which results in an improvement over the UAVE (Figure 4.6). Notice the jog in the UAVE track after 84 h that results from the loss of model guidance other than the models in Figure 4.6. The WVAE track is comparable to the UAVE track in the along-track direction, but is much farther from the BT in the cross-track direction. In this case, the WVAE performs so poorly relative to the UAVE and the WAVE because the initial 84-h UAVE position used for the WVAE is so far from the BT. The remaining model tracks (NOGAPS, GFS, and GFDN) all primarily advance in the along-track direction after 84 h, and thus the weighted motion vector consensus WVAE likewise mostly advances in the along-track direction starting from the 84-h UAVE position. The WVAE is not able to account for the shift in the position consensus UAVE and WAVE towards the BT in the along track direction since the WVAE only accounts for the motion vector components of the model tracks.

Table 4.5 As in Table 4.3, except for storm 8 from 1800 UTC 5 August 2006.

Time	UAVE error	WAVE error	WAVE Imp	WVAE error	WVAE Imp
96 h	199.5	184.1	15.4	281.8	-82.3
108 h	230.1	206.6	23.4	301.4	-71.4
120 h	282.0	257.6	24.4	350.3	-68.3

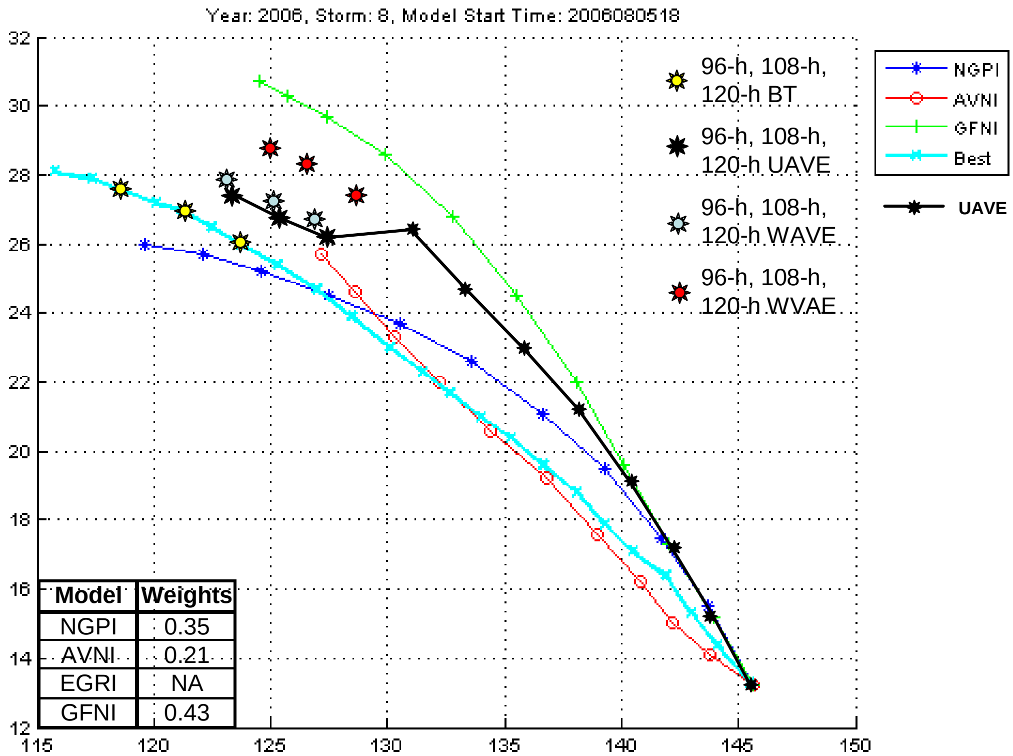


Figure 4.6 As in Figure 4.4, except for Storm 8 from 1800 UTC 5 August 2006.

Storm 14 from 1200 UTC 12 September 2006 is another case in which the weighted position consensus WAVE has smaller errors than the weighted motion vector consensus WVAE. However, the WVAE still improves on the UAVE at 96 h and 108 h (Table 4.6). Both the weighted WAVE and WVAE tracks are closer to the BT in the cross-track direction than the unweighted UAVE track at 96 h, 108 h, and 120 h (Figure 4.7). For the WAVE and WVAE, the improved cross-track forecast is due to the low weighting factor of 0.06 for the GFS track. The GFS track gains greater weight as additional models drop out after 84 h, which causes the unweighted consensus UAVE to shift away from the BT toward the GFS track after 84 h. Analogous to the previous case, all of the model tracks after 84 h have a similar track direction including the erroneous GFS track, which results in a smooth translation from the 84-h UAVE position to the WVAE 96 h, 108 h, and 120 h positions. In contrast to the previous case, the unweighted consensus UAVE is shifted away from the BT in the cross-track direction after 84 h, which allows the weighted vector consensus WVAE to slightly improve on the UAVE at 96 h and 108 h. The WVAE is not able to improve as much as the WAVE because it does not account for the more northwesterly positions of the three best-performing models (NOGAPS, UKMO, and GFDN). Since all of the model guidance does not predict the rapid acceleration, all of the unweighted and weighted consensus 120-h track errors are large.

Table 4.6 As in Table 4.3, except for Storm 14 from 1200 UTC 12 September 2006.

Time	UAVE error	WAVE error	WAVE Imp	WVAE error	WVAE Imp
96 h	207.3	118.2	89.1	174.3	33.0
108 h	299.7	227.8	72.0	272.4	27.3
120 h	390.0	365.7	24.3	402.2	-12.2

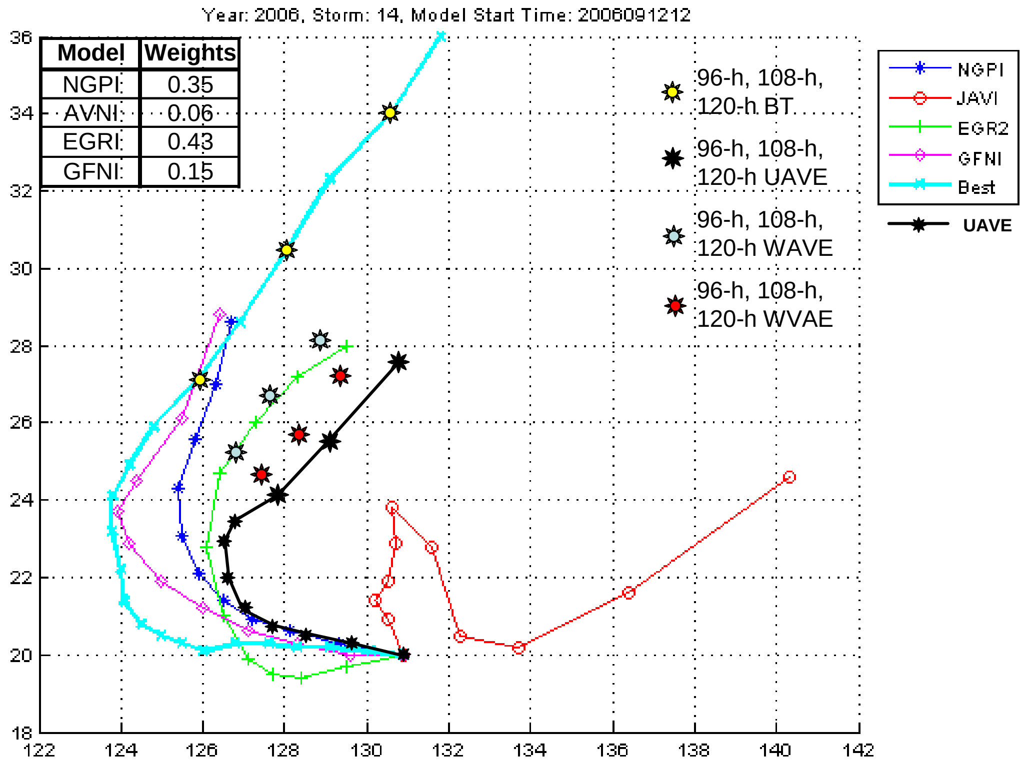


Figure 4.7 As in Figure 4.4, except for Storm 14 from 1200 UTC 12 September 2006 and with GFS interpolated track positions (JAVI, red circles), UKMO 12-h interpolated track positions (UKM2, green crosses), and GFDN interpolated track positions (GFNI, purple diamonds).

THIS PAGE INTENTIONALLY LEFT BLANK

V. CONCLUSIONS AND RECOMMENDATIONS

A. CONCLUSIONS

This study demonstrates that a weighted position consensus at 96 h, 108 h, and 120 h can improve on a simple unweighted position consensus, as is used operationally at JTWC. The average error statistics for storms 1-24 during the 2006 western North Pacific season are reduced through using the weighted position consensus. In addition, the weighted position consensus tends to reduce erratic track changes associated with the loss of model track guidance with forecast time. These improvements are achieved without any training period to determine the weightings. All that is needed to apply the weighted position consensus technique are the model track positions, which are already used for an unweighted consensus.

The removal of the MM5 and COAMPS model tracks from the consensus resulted in greater improvements with the weighted position consensus. An improved 60-h, 66-h, and 72-h consensus track on which to base the weighting factors is the reason for the reduced errors. If the consensus positions used to determine the weighting factors are improved, the weighted consensus improves. In addition, the improved weighted position consensus at 96 h, 108 h, and 120 h further justifies the decision to remove the MM5 and COAMPS model tracks from the operational consensus at JTWC.

Adding the 96-h JGSM forecasts (when available) to the weighted position consensus yields dramatic improvements over the unweighted position consensus. This increased improvement is because the JGSM is a skillful model, and all skillful model guidance that is available should be included with the weighted position consensus. In this context, a skillful model is one that reduces the unweighted consensus track errors when the model is included in a sufficient number of cases to test model skill.

The availability of the ECMWF tracks late in the 2006 season gave an opportunity to evaluate the impact that the ECMWF model has on the weighted position consensus technique. For the weighted position consensus technique to be successful, the 60-h, 66-h, and 72-h model performance relative to the unweighted consensus needs to be a good indication of future performance relative to the best-track (BT) at 96 h, 108 h, and 120 h. First, the unweighted consensus should provide good guidance on which to base the weighting factors (close to the BT from 60 h – 72 h). Second, the model performance at 60 h, 66 h, and 72 h should be consistent with the model skill at 96 h, 108 h, and 120 h when the weighted consensus is computed. In the case of the ECMWF, it was found that the model performance was skillful relative to an unweighted test consensus called TESB by B. Sampson only after 72 h. Adding the ECMWF tracks greatly improved the unweighted consensus at 96 h, 108 h, and 120 h, but did not further improve the performance of the weighted position consensus. Thus, the weighted position consensus technique likely did not assign optimal weights to the ECMWF model tracks due to the poorer performance of the ECMWF model during 60 h, 66 h, and 72 h. Although the performance of the weighted consensus is not necessarily directly proportional to the performance of the models used for the weighting, the weighted consensus is simply a way to better use the available skillful model guidance.

The most significant results of the study were the improvements gained by a motion vector consensus over a position consensus. The use of a motion vector consensus in this sample dramatically improved long-range TC track prediction. A near 10% improvement over an unweighted position consensus at 96 h and a 5% improvement at 120 h should be translated into guidance for military commanders and civilian leaders in preparing for TC impacts. From the case studies in this study, it appears that the success of the motion vector consensus is due to an enhanced translation in the along-track direction. A position consensus track tends to have a slow bias when the spread of the model tracks increases, which generally occurs with increased forecast time. Using a motion vector consensus at long lead times can enhance the along-track component by the average of the motion vector magnitudes. An even larger improvement over an unweighted position consensus is achieved by weighting the motion vectors with the same weighting factors as used previously. It appears that the weighted motion vector consensus also results in better track continuity than both an unweighted position consensus and a weighted position consensus when the number of model tracks in the consensus is reduced. This better continuity is due to the translation of motion vectors from disparate tracks to a single point to determine a consensus track.

Although these results are for the western North Pacific, the weighted consensus method could be easily applied to other basins and other sets of model tracks since the weighting factors are not statistically based and are only dependent on the model guidance for each forecast time.

B. RECOMMENDATIONS

The validation study of a weighted position consensus shows promising results with average improvements over an unweighted consensus at 96 h, 108 h, and 120 h. The sensitivity studies demonstrate ways to improve on the initial validation study. Combining the results from the separate validation studies should yield a more robust weighted consensus. That is, a new weighted position consensus should be evaluated with the MM5 and COAMPS removed, the JGSM added to the weighting scheme at 96 h, and the ECMWF added to the unweighted and weighted consensus. This configuration should produce reduced errors from the validation study and best mimic current operations at JTWC. One may ask why the ECMWF should be included since it degraded the weighted position consensus relative to an unweighted position consensus. The ECMWF should be included in a weighted position consensus since the JTWC is using an unweighted consensus that includes the ECMWF and because the ECMWF is such a skillful model at longer forecast intervals (Figure 3.15). Perhaps to further improve on the unweighted consensus, the ECMWF should be systematically assigned greater weighting factors at longer forecast intervals when its skill is greatest.

Similar tests described above for the weighted position consensus should be applied to the weighted motion vector consensus. In addition, the motion vector consensus should be initiated from 72 h – 84 h instead of from 84 h – 96 h. To implement this successfully, the JGSM and JTYM need to be included in the weighting scheme at 84 h. Although the COAMPS model is also sometimes available at 84 h, it should not be included since it has been removed from the JTWC unweighted position consensus CONW. Once the JGSM and JTYM are included in the weighting at 84 h, the position consensus starting from a 72 h unweighted position consensus should yield similar results at 84 h as for the weighted motion vector consensus validation study (Table 4.2) at 96 h. If the 84 h position is improved relative to an unweighted position consensus, it will form a superior basis for the subsequent translation to 96 h, etc. Thus, starting the motion vector consensus at 72 h should yield improvements at 84 h, 96 h, 108 h, and 120 h.

In this study, an attempt was made to replicate the guidance that JTWC uses at each forecast time. As mentioned in Chapter II.B.2, this could result in the use of one set of model tracks (i.e., primary interpolated track) to compute the weighting factors and another set of tracks (i.e., interpolated Fiorino track) to apply the weighting. Even though the Fiorino tracks are from the same model, there could be differences in the tracks and thus in the skill. It would be useful to compute the weighting factors again using the same model tracks to form the weightings as to apply the weightings. For example, if the GFS interpolated track AVNI is only available through 72 h, but the GFS Fiorino interpolated track JAVI is available through 120 h, the weighting factors could be computed using the 60-h, 66-h, and 72-h JAVI track positions, even though the AVNI was available at those times. Although the AVNI may be a superior track at 60 h, 66 h, and 72 h, it is desirable to use the JAVI track to compute the weights since this should better reflect the future performance of the JAVI track at 96 h, 108 h, and 120 h. Although this change may only yield modest improvements, it is a more consistent way of applying the weighting factors.

Once further refinements are completed, a weighted motion vector consensus could be implemented at JTWC. The weighted motion vector technique is systematic and could easily be automated. The ATCF could potentially be updated to compute the weighted motion vector consensus since the only input needed is the model tracks, which are already included in the ATCF. The results from the motion vector consensus validation study are promising and, as mentioned above, opportunities are available to improve on these first results. Even a 9.9% improvement over an unweighted position consensus at 96 h could significantly raise the bar for forecasters at JTWC since an unweighted position consensus is currently the primary guidance at JTWC. Raising the standard for forecasters should lead to improved TC track forecasts and superior support to military operations and planning in the western North Pacific.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX: HAVERSINE DISTANCE FORMULA

The Haversine formula was used to compute the great circle distances between the forecast and best-track positions. This formula was adapted from the Haversine distance formula available at http://www.movable-type.co.uk/scripts/GIS-FAQ-5.1.html derived from Sinnott (1984).

$$\Delta lon = lonA - lonB \qquad \text{(A.1)}$$

$$\Delta lat = latA - latB \qquad \text{(A.2)}$$

$$a = \sin^2\left(\frac{\Delta lat}{2}\right) + \cos(latA) * \cos\ (latB) * \sin^2\left(\frac{\Delta lon}{2}\right) \qquad \text{(A.3)}$$

$$c = 2\left[\sin^{-1}\left(\sqrt{a}\right)\right] \qquad \text{(A.4)}$$

$$dist = c\left(\frac{21600}{2(\pi)}\right) \qquad \text{(A.5)}$$

Equation (A.4) is the angular distance between points A and B in radians. Equation (A.5) is the distance between points A and B in nautical miles.

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

Burton, A., 2006: Sharing experiences in operational consensus forecasting. Topic 3a, Proceedings of the Sixth WMO International Workshop on Tropical Cyclones, San Jose, Costa Rica, 21-30 November 2006.

Carr, L. E., III, and R.L. Elsberry, 2001: Beta test of the Systematic Approach expert system prototype as a tropical cyclone track forecasting aid. *Wea. Forecasting*, **16**, 355-368.

Goerss, J. S., C. R. Sampson, and J. M. Gross, 2004: A history of western north Pacific tropical cyclone track forecast skill. *Wea. Forecasting*, **19**, 633-638.

Jeffries, R.A., and E.J. Fukada, 2002: Consensus approach to track forecasting. Paper TP3.2, Extended Abstracts, Fifth International Workshop on Tropical Cyclones, Cairns, Australia, World Meteorological Organization (Geneva).

Kumar, T.S.V., T. N. Krishnamurti, M. Fiorino, and M. Nagata, 2003: Multimodel superensemble forecasting of tropical cyclones in the Pacific. *Mon. Wea. Rev*., **131**, 574-583.

Sampson, C.R., and A. J. Schrader, 2000: The automated tropical cyclone forecasting system (Version 3.2). *Bull. Amer. Meteor. Soc.,* **81**, 1231-1240.

Sinnott, R.W., 1984: Virtues of the Haversine. *Sky and Telescope*, **68**, 159.

Weber, H.C., 2003: Hurricane track prediction using a statistical ensemble of numerical models. *Mon. Wea. Rev*., **131**, 749-770.

THIS PAGE INTENTIONALLY LEFT BLANK

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California

3. Air Force Weather Technical Library
   Air Force Combat Climatology Center
   Asheville, North Carolina

4. Air Force Institute of Technology
   Wright-Patterson Air Force Base, Ohio

5. Professor Philip Durkee
   Naval Postgraduate School
   Monterey, California

6. Professor Russell Elsberry
   Naval Postgraduate School
   Monterey, California

7. Mark Boothe
   Naval Postgraduate School
   Monterey, California

8. Charles Sampson
   Naval Research Lab Marine Meteorology Division
   Monterey, California

9. Joint Typhoon Weather Center
   Pearl Harbor, Hawaii