AMERICAN
METEOROLOGICAL
SOCIETY

Weather and Forecasting

EARLY ONLINE RELEASE

This is a preliminary PDF of the author-produced manuscript that has been peer-reviewed and accepted for publication. Since it is being posted so soon after acceptance, it has not yet been copyedited, formatted, or processed by AMS Publications. This preliminary version of the manuscript may be downloaded, distributed, and cited, but please be aware that there will be visual differences and possibly some content differences between this version and the final published version.

The DOI for this manuscript is doi: 10.1175/2010WAF2222410.1

The final published version of this manuscript will replace the preliminary version at the above DOI once it is available.

© 2010 American Meteorological Society

GPCE-AX: An Anisotropic Extension to the Goerss Predicted Consensus Error

James A Hansen

James S Goerss

Charles Sampson

Naval Research Laboratory, Monterey

Date of Submission: February 16, 2010

Revised: June 25, 2010

Corresponding author address

James A. Hansen

Naval Research Laboratory

7 Grace Hopper Ave MS 2

Monterey, CA 93943, USA

E-mail

jim.hansen@nlrmry.navy.mil

PRELIMINARY ACCEPTED VERSION

Abstract

A method to predict an anisotropic expected forecast error distribution for consensus forecasts of tropical cyclone (TC) tracks is presented. The method builds upon the Goerss Predicted Consensus Error (GPCE), which predicts the isotropic radius of the 70% isopleth of expected TC track error. Consensus TC track forecasts are computed as the mean of a collection of TC track forecasts from different models and are basin dependent. A novel aspect of GPCE is that it uses not only the uncertainty in the collection of constituent models to predict expected error, but also other features of the predicted storm, including initial intensity, forecast intensity, and storm speed. The new method, called GPCE along/across (or GPCE-AX) takes a similar approach but separates the predicted error into across-track and along-track components.

GPCE-AX has been applied to consensus TC track forecasts in the Atlantic (CONU/TVCN) and in the western North Pacific (CONW). The results for both basins indicate that GPCE-AX outperforms or is equal in quality to GPCE in terms of reliability (the fraction of time verification is bound by the 70% uncertainty isopleths) and sharpness (the area bound by the 70% isopleths).

GPCE-AX has been implemented at both the National Hurricane Center and at the Joint Typhoon Warning Center for real-time testing and evaluation.

1. Introduction

The multiple sources of objective guidance available for tropical cyclone (TC) track prediction plus the ever-growing database of historical TC forecast/verification pairs enable the production of statistically-based real-time guidance products based on available NWP model output (guidance on guidance). These products can produce improved deterministic estimates of the future state of the atmosphere and can produce estimates of forecast uncertainty. Such products can be used by forecasters to help improve their official forecasts, and by decision makers to help them manage risk associated with the approaching TC. Operational examples include TC track consensus forecasts (Goerss 2000, Goerss et al. 2004; Sampson et al. 2005) based on averaging the track forecasts from several different numerical models, the wind-speed probability product (DeMaria et al. 2009) based on historical track, intensity and size forecast errors, and the Goerss Predicted Consensus Error (GPCE; Goerss 2007) based on the collection of track forecasts that go into the consensus product along with other aspects of the forecast TC.

The work presented here is guidance on guidance aimed at quantifying the state-dependent, across-track and along-track uncertainty associated with consensus TC track forecasts. It is a natural extension of the GPCE product, which aims to quantify the state-dependent, *isotropic* uncertainty associated with consensus TC track forecasts, and so we denote the new product as GPCE along/across, or GPCE-AX.

GPCE-AX is applied to multi-model TC ensemble mean track forecasts in the Atlantic (CONU/TVCN[1]) and western North Pacific (CONW[2]) basins. Results are dependent upon the choice of training and testing sets. In the Atlantic basin the training period was 2002-2007 with a test period over 2008. For 2008 in the Atlantic, GPCE-AX is found to be more reliable than GPCE at all forecast leads, and to be sharper than GPCE at leads greater than 24hrs. In the western North Pacific basin the training period was 2004-2006 with a test period over 2007. For 2007 in the western North Pacific, GPCE-AX is more reliable than GPCE for all leads except 120hrs (where the two methods are identical), and is sharper than GPCE for leads greater than 48hrs. In this work we choose to focus on 2007 for the western North Pacific. 2008 was a year with relatively few storms and a year where the objective aids did not perform well. Since GPCE and GPCE-AX are both based on objective aids, the reliability of the predicted uncertainties was poor. Note that all testing was performed on independent datasets. More sophisticated forms of cross-validation (e.g. K-fold or leave-one-out) are not appropriate because of the non-stationarity of the dataset; e.g. using the improved NWP guidance from 2007 as part of the training set and testing over the 2006 data would lead overly-optimistic results. As the goal of probabilistic forecasting is to produce as sharp a forecast as possible subject to the constraint of reliability (Murphy and Winkler 1987; Gneiting et al. 2007), the unreliable forecasts in the 2008 western North Pacific render a comparison of the methods moot. Another aspect of the choice of training set is that if

[1] CONU: Consensus version U. TVCN: Track Variable Consensus. Consensus forecasts are straight multi-model ensemble mean TC track forecasts.

[2] CONW: Consensus version W.

the consensus forecasts are especially good (bad) over the independent test period both GPCE and GPCE-AX will over (under) estimate the uncertainty area.

In section 2 the GPCE-AX methodology is described, followed by the presentation of GPCE-AX results in section 3 where they are compared with results produced by the isotropic GPCE approach. Summarizing conclusions are presented in section 4.

2. GPCE-AX Description

GPCE-AX is constructed by employing multivariate linear regression (MVLR) to independently predict across-track and along-track TC track error of the objective consensus. These predicted errors are then scaled to define an ellipse that represents the 70% probability isopleth of bounding the true location of the TC. For each basin of interest, a training period is defined over which the predictands and potential predictors for all available storms are extracted and MVLR applied to identify the best few (typically two or three) predictors. Once the predictors have been identified, a scaling factor is determined that, when added to the predicted error, results in an ellipse that bounds the actual forecast error 70% of the time over the training period. Testing of the resulting MVLR coefficients and scaling is carried out over an independent time period.

a. Predictors and predictands

A selection of potential predictors are extracted or derived from the Automated Tropical Cyclone Forecasting System (ATCF; Sampson and Schrader 2000) objective aid data files (see http://ww.nrlmry.navy.mil/atcf_web) which are available in real time. They provide the same pool of potential predictors utilized by GPCE with the addition of predictors related to the across-track and along-track spread of the ensemble members making up the consensus forecast. The complete pool of potential predictors is given in Table 1. The predictands are the magnitudes of the across-track and along-track consensus track errors and are derived from the storm best track files produced by the National Hurricane Center (NHC) and the Joint Typhoon Warning Center (JTWC).

b. Determining the across-track and along-track directions

The across-track and along-track directions utilized for the spread and error calculations are determined by fitting the consensus forecast track piecewise with cubic Hermite splines, interpolating to hourly locations, and then calculating the track tangent direction using finite differencing. Cubic Hermite spline interpolation was chosen over the more traditional 3rd order spline interpolation because it was found that the continuous 2nd derivative constraint required by traditional 3rd order spline fitting would occasionally result in spurious along-track directions. Piecewise cubic Hermite splines only require continuous 1st derivatives at each data point. An example is shown in figure 1 for a 2004 storm in the western North Pacific. The blue circles are the consensus track forecast for hours 6, 12, 24, 36, 48, and 72. Points between the official forecast times are interpolated using either 3rd order splines (black curve) or Hermite polynomials (magenta curve). Across-track and along-track directions are determined by numerically calculating the slope at each of the official forecast times. The cyan axes show the across-track and along-track directions as determined by the Hermite polynomial interpolation. Note that different directions would have been obtained from the spline interpolation, especially at hours 6, 12, 48, and 72.

c. Determining across-track and along-track bias

GPCE-AX predictands are the *magnitude* of the across-track and along-track errors, but the calculation of the *signed* across-track and along-track errors (positive for verification falling to the right of and in front of verification, respective) over the training dataset enables the across-track and along-track bias to be calculated and for the consensus forecasts using out of sample data to be bias corrected. The bias correction values over the training periods for the two basins are given in table 2. The GPCE-AX results presented in this work predict the uncertainty associated with the across-track and along-track *bias corrected* consensus for two reasons: 1) the resulting product produces a display that is slightly different from the GPCE product, and 2) it supports a product under development that predicts the probability of falling to the left or right of the bias corrected track. The sensitivity of this choice is discussed below, but in brief, the out of sample results obtained when using bias correction or not using bias correction are statistically indistinguishable for the basins and periods tested.

d. Predictor selection

The GPCE/GPCE-AX approach is to utilize as few predictors as possible. To select the best 2 or 3 predictors from the 11 possible predictors listed in Table 1, a step-wise regression approach is employed. First, linear regression is applied to each predictor independently in an effort to find the one that has the strongest relationship with the predictands. Next, a two-predictor MVLR is applied where one of the predictors is the one that performed best on its own, and the second is one of the remaining 10 predictors. Again, the set of predictors that perform the best is retained and the process is repeated once more to find the top three predictors. Experiments were performed where MVLR was performed on all possible combinations of two or three predictors, and results identical to the step-wise approach were obtained.

To choose the number of predictors, the quality of the models for different numbers of predictors were compared in-sample. It was found that for all basins and forecast leads considered, adding a third predictor gave only a 1% or 2% improvement. Out of sample results showed that the three predictor models were worse than the two predictor models.

The final step in predictor selection is subjective. Perhaps more important than the accuracy of the regression models employed is the consistency of the predicted error as a function of forecast lead. It is expected that forecasters will mistrust a product that tends to predict a smaller error at, say 96hrs than it does at 72hrs. Such things occasionally happen when predictors change as a function of forecast lead. To minimize this risk, the three or four predominant predictors that appear in the collection of forecast leads are isolated and the step-wise MVLR predictor selection is repeated using only those three or four predictors. Tables 3 and 4 provide the predictors utilized by GPCE-AX in this study. Note that for each forecast lead the dominant predictor is the ensemble spread.

e. Scaling for predicted uncertainty

The MVLR approach described above predicts the consensus forecast error in the along-track and the across-track directions. GPCE-AX predicts uncertainty, and a heuristic approach is employed to transform the predicted error at a given forecast lead to the predicted uncertainty. A forecast-lead dependent constant boost term is added to the predicted error with the aim of identifying the 70% probability isopleth of bounding the storm. The boost term is computed by starting with a boost of 1nmi and calculating the fraction of in-sample storms bounded by the resulting ellipse. The boost is increased until 70% of the in-sample storms are bound. The boost is added directly to the predicted across-track error, but is scaled by the eccentricity before being added to the predicted along-track error so that the eccentricity of the uncertainty ellipse is maintained as the boost increases. The boost values used in this work are given in Table 5. An additive boost was utilized rather than a multiplicative boost to mimic the approach used for the operational GPCE product. This enables code to be reused in the operational implementation of GPCE-AX. Experiments utilizing a multiplicative boost factor indicate that the factor is stable across forecast leads with a value of approximately 1.2 for GPCE and approximately 1.9 for GPCE-AX. Employing a multiplicative boost did not have a statistically significant impact on the results.

3. Results

This section produces GPCE-AX out-of-sample results for the Atlantic and western North Pacific basins. Example cases are presented to demonstrate what GPCE-

AX guidance looks like. In addition, basin-wide summary statistics are presented. The norms utilized measure reliability and sharpness. Since GPCE-AX predicts the isopleths of 70% uncertainty, the reliability norm calculates the fraction of times the storm actually falls within that bound. A perfectly reliable forecast system would find that verification falls within the bound 70% of the time. Sharpness measures the degree to which an uncertainty forecast is different from some baseline. Typically the baseline chosen is climatology (e.g. the uncertainty cone associated with the operational "Potential Day 1-5 Track Area", Franklin 2009), but in this work we utilize GPCE as the baseline. GPCE has already been shown to provide a better uncertainty forecast than the cone of uncertainty, and so is the more relevant null hypothesis. The GPCE results used here were developed using the same time periods and pool of predictors as were available to GPCE-AX.

a. Cases

An example of the GPCE and GPCE-AX guidance products for a 72hr forecast of Ike in 2008 is shown in figure 2. An example of GPCE and GPCE-AX for the western North Pacific is given in figure 3. In each figure the best track is given as the red curve, and the consensus forecast is given as the magenta curve. The 72hr consensus forecast position is plotted as a thick blue circle and the 72hr across-track/along-track bias-corrected consensus forecast is plotted as a thin blue circle. The verifying storm location is plotted as the thick red square. The 72hr forecasts of the individual members making up the consensus are plotted as green x's. The GPCE forecast of the 70% uncertainty isopleth is plotted as the thin magenta circle and the GPCE-AX forecast of the 70% uncertainty isopleths is plotted as the thin blue ellipse.

It is not possible to comment on the quality of a probabilistic forecast product based on individual realizations. Instead, the purpose of these figures is to demonstrate the typical differences one sees between GPCE and GPCE-AX guidance products. Because the leading GPCE-AX predictor is the across-track and along-track spread, the shape of the GPCE-AX ellipse is more consistent with the distribution of the individual objective aids.

b. Summary statistics

In brief, it is found that GPCE-AX produces reliable uncertainty forecasts that are sharper than GPCE, and that GPCE-AX improvements relative to GPCE are greater in the Atlantic basin than in the western North Pacific basin.

Out of sample reliability and sharpness results are presented in Tables 6 and 7 for the Atlantic and the western North Pacific, respectively. The reliability columns indicate the fraction of times verification is bound by the predicted 70% isopleth. A perfectly reliable system would have 70% for every forecast lead. Percentages greater than 70% indicate that verification is bound too frequently, percentages less than 70% indicate that verification is not bound frequently enough. In the Atlantic, GPCE-AX reliabilities are closer to 70% than GPCE reliabilities at all leads except 120hrs, where they are identical.

In the western North Pacific, GPCE-AX reliabilities are closer to 70% in 4 of the 7 forecast leads, and identical to GPCE-AX in 1 of the 7 forecast leads.

The sharpness measure is presented in the "Mean area difference" column. For each forecast over the test period the areas bound by the GPCE and the GPCE-AX 70% isopleths are calculated and the fractional difference between them (normalized by the GPCE area) is computed. Fractional differences greater than 0 indicate that the GPCE area is greater than the GPCE-AX area. The results reported in Tables 6 and 7 are the means of those fractional differences. In the Atlantic basin the GPCE and GPCE-AX areas are statistically identical (one-sided T-test, 95% confidence) for 12hr and 24hr leads. GPCE-AX areas are smaller than GPCE areas for all leads greater then 24hrs; GPCE-AX forecasts are both more reliable and sharper than GPCE forecasts. In the western North Pacific basin the sharpness results are mixed. GPCE-AX areas are never larger than GPCE areas, and are statistically smaller for 12hr forecasts, and for all forecasts beyond 48hrs.

Summary statistics can be misleading because they lump together all available cases. To gain a bit more insight into the performance of GPCE-AX relative to GPCE the out of sample cases were conditioned on the size of the predicted error; the smallest third, the middle third, and the largest third. For both basins, GPCE-AX tends to bound too many verifying observations and has larger areas than GPCE for the forecasts with the smallest actual errors. The reliability of GPCE-AX improves as the actual error increases, and the sharpness of GPCE-AX forecasts relative to GPCE is best for the cases with the largest actual error. These results indicate that the value of the anisotropic GPCE-AX approach lies primarily in situations where there is significant spread amongst the objective aids; the area bound by the 70% isopleth is reduced relative to GPCE areas because GPCE-AX has the freedom to partition the large uncertainty into two directions.

Due to the poor performance of the objective aids in the 2008 western North Pacific season, both GPCE and GPCE-AX provided unreliable uncertainty estimates. There is little point in comparing the reliability in two unreliable forecast systems, and there is no point in discussing relative sharpness between unreliable forecast systems. Forecasters must keep in mind that because GPCE and GPCE-AX use information from the objective aids as predictors, when the objectives aids perform systematically poorly so too will the GPCE and GPCE-AX uncertainty estimates. This is an example of a general problem faced by statistical approaches applied to non-stationary datasets. Forecast systems change each year, and unless forecast centers are willing to re-run all of their guidance on historical cases prior to each storm season the historical performance of statistically-based guidance products will only be a weak predictor of the quality of the products during the season.

4. Conclusions

GPCE-AX provides a small, but consistent improvement over GPCE as an uncertainty guidance product, especially in the Atlantic basin. Like GPCE it predicts the 70% isopleth of bounding the verifying TC location by regressing objective aid predictors onto actual forecast error, but instead of assuming an isotropic uncertainty distribution, GPCE-AX splits the uncertainty into across-track and along-track components. GPCE-AX is found to be more reliable than GPCE and is found to produce sharper uncertainty forecasts. GPCE-AX performs best for storms with large error because it is able to partition the uncertainty into across-track and along-track directions rather than using an isotropic distribution.

GPCE-AX has been incorporated into the ATCF system and is undergoing operational assessment at both JTWC and NHC. It utilizes the same infrastructure as GPCE and so that any value-added products that use GPCE could also make use of GPCE-AX.

Acknowledgements

This work was support by the NOAA Joint Hurricane Testbed project titled "TC Dressing: A Probabilistic Approach to Providing State Dependent, Non-Isotropic Forecast Track Error Guidance". The authors would like to thank Mark DeMaria and two anonymous reviewers whose comments greatly improved the manuscript.

References

DeMaria M. , J.A. Knaff, R. Knabb, C. Lauer, C.R. Sampson, and R.T. DeMaria, 2009: A New Method for Estimating Tropical Cyclone Wind Speed Probabilities. *Wea. Forecasting*, **24**, 1573-1591.

Franklin, J., 2009: 2008 National Hurricane Center Forecast Verification Report. http://www.nhc.noaa.gov/verification/pdfs/Verification_2008.pdf

Gneiting, T., F. Balabdaoui, and A.E. Raftery, 2007: Probabilistic Forecasts, Calibration and Sharpness. *J. Royal Stat. Soc., Series B*, **69**, 243-268.

Goerss, J., 2007: Prediction of consensus tropical cyclone track forecast error. *Mon. Wea. Rev.,* **135**, 1985-1993.

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

Goerss, J., 2000: Tropical cyclone track forecasts using an ensemble of dynamical models. *Mon. Wea. Rev.,* **128**, 1187-1193.

Murphy, A.H., and R.L. Winkler 1987: A General Framework for Forecast Verification. *Mon. Wea. Rev.*, **115**, 1330-1338.

Sampson, C.R., and A.J. Schrader, 2000: The automated tropical cyclone forecasting system (Version 3.2). *Bull. of the AMS*, **81**, 3148-3158.

Sampson, C., J. Goerss, and A. Schrader, 2005: A consensus track forecast for southern hemisphere tropical cyclones. *Aust. Met. Mag*., **54**, 115-119.

Figures

1. Piece-wise cubic Hermite spline vs. traditional 3rd order spline interpolation. The red squares are the actual track of the TC, the circles are the consensus forecast (CONW). A traditional 3rd order spline interpolation is given by the black curve while a piece-wise cubic Hermite spline interpolation is given by the magenta curve. Across-track and along-track directions are plotted as cyan axes and were determined by numerically differentiating the interpolated forecast track obtained through the use of Hermite splines. Note that the constraint imposed by a continuous second derivative in 3rd order splines results in an interpolated track (black curve) that is physically unrealistic and would result in incorrect across-track and along-track directions.
2. This is a 72hr forecast of Ike in 2008. The black curve shows the best track trajectory and the red curve shows the CONU forecast (hours 12, 24, 36, 48, and 72). The 72hr track forecasts from the five models that make up CONU are given as green x's. The 72hr CONU forecast is given by the solid blue circle, the 72hr across-track/along-track bias corrected forecast is given by the open blue circle, and the 72hr verifying location is given by the black square. The GPCE forecast of expected error is given by the red circle centered on the CONU forecast, and the GPCE-AX forecast of expected error is given by the blue ellipse centered on the bias corrected CONU forecast.
3. As for figure 2 but for storm 4 in the western North Pacific in 2004.

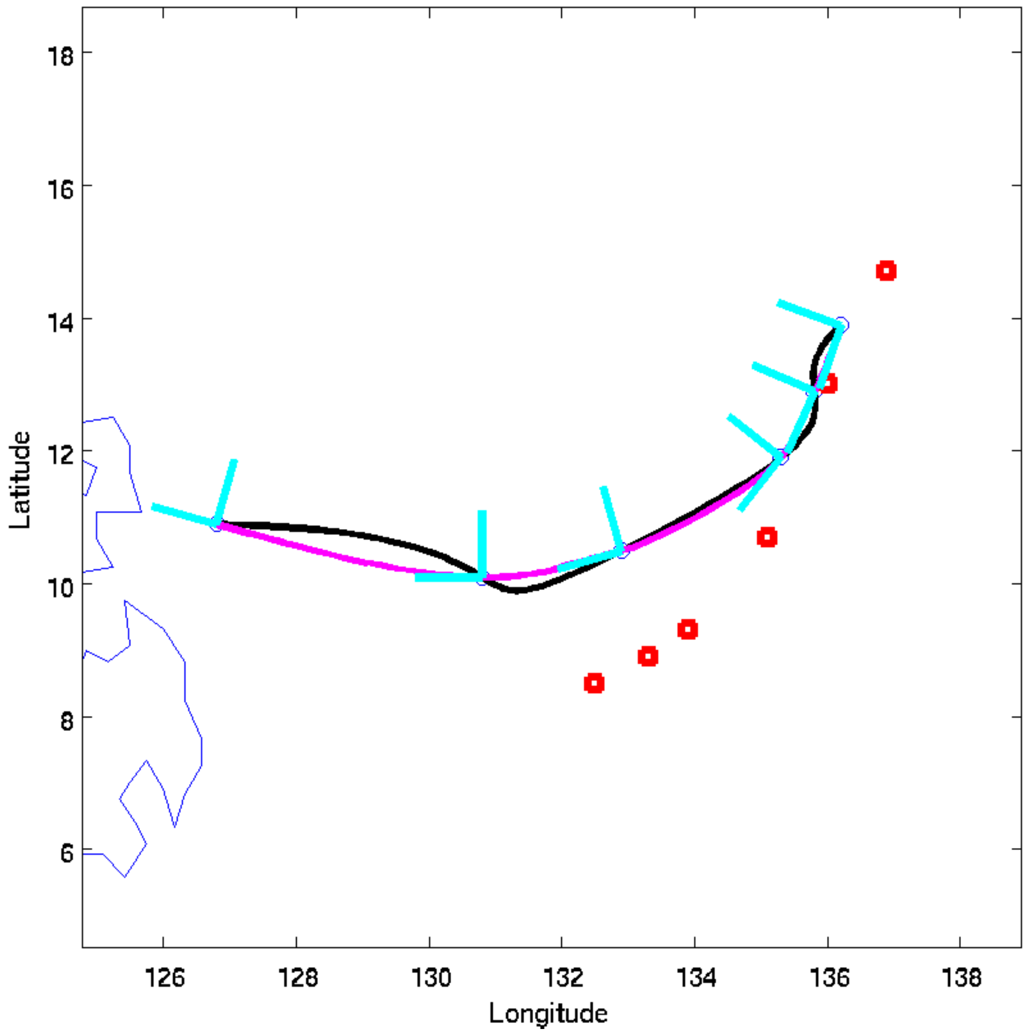

Figure 1: Piece-wise cubic Hermite spline vs. traditional 3rd order spline interpolation. The red squares are the actual track of the TC, the circles are the consensus forecast (CONW). A traditional 3rd order spline interpolation is given by the black curve while a piece-wise cubic Hermite spline interpolation is given by the magenta curve. Across-track and along-track directions are plotted as cyan axes and were determined by numerically differentiating the interpolated forecast track obtained through the use of Hermite splines. Note that the constraint imposed by a continuous second derivative in 3rd order splines results in an interpolated track (black curve) that is physically unrealistic and would result in incorrect across-track and along-track directions.

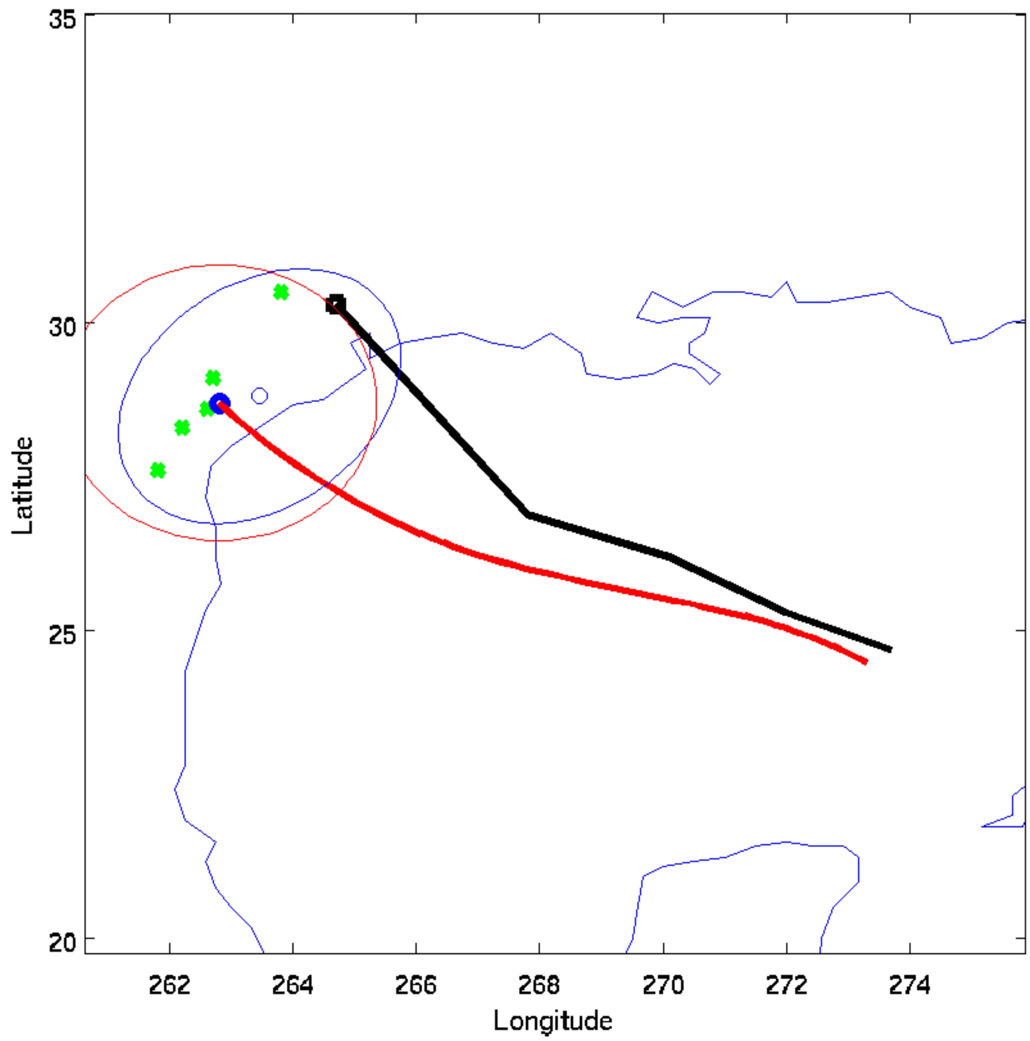

Figure 2: This is a 72hr forecast of Ike in 2008. The black curve shows the best track trajectory and the red curve shows the CONU forecast (hours 12, 24, 36, 48, and 72). The 72hr track forecasts from the five models that make up CONU are given as green x's. The 72hr CONU forecast is given by the solid blue circle, the 72hr across-track/along-track bias corrected forecast is given by the open blue circle, and the 72hr verifying location is given by the black square. The GPCE forecast of expected error is given by the red circle centered on the CONU forecast, and the GPCE-AX forecast of expected error is given by the blue ellipse centered on the bias corrected CONU forecast.

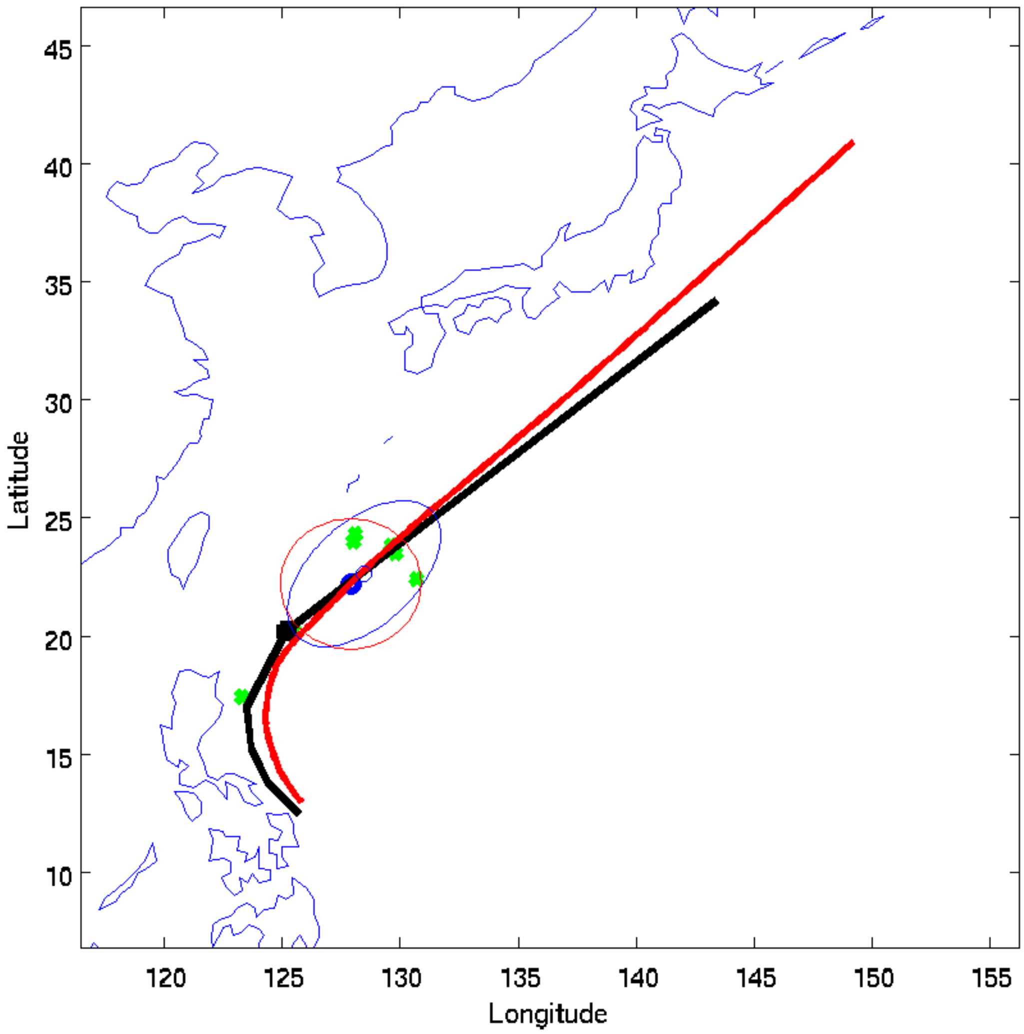

Figure 3: As for figure 2 but for storm 4 in the western North Pacific in 2004.

Tables

1. Available GPCE-AX predictors. Predictors are extracted or derived from ATCF data files and are available to the GPCE-AX system in real time.

2. Across-track and along-track bias (in NMI) over the training periods for the Atlantic and West Pacific basins. GPCE-AX is trained to predict the bias corrected consensus forecasts.

3. Atlantic basin GPCE-AX predictors based on a 2002-2007 training set.

4. West Pacific basin GPCE-AX predictors based on a 2004-2006 training set.

5. Boost values added to the predicted error of the consensus to transform the predicted error value to the predicted 70% uncertainty isopleths. Boost values are directly added to the across-track predicted error, but are scaled before being added to the along-track directions in order to maintain eccentricity.

6. Out of sample summary statistics for the Atlantic basin in 2008. GPCE-AX has greater reliability than GPCE, and the area bound by the 70% isopleth is systematically smaller for all leads greater than 24hrs.

7. Out of sample summary statistics for the western North Pacific basin in 2007. While not as clear-cut at the Atlantic basin, GPCE-AX forecasts are more reliable

than GPCE forecasts for a majority of forecast leads, and the area bound by the 70% isopleth is smaller than the GPCE areas for 12hr forecasts and all forecasts greater then 48hrs.

Across-track spread	The mean of the across-track distances between the consensus and each ensemble member.
Along-track spread	The mean of the along-track distances between the consensus and each ensemble member.
Isotropic spread	The mean of the distances between the consensus and each ensemble member.
Initial latitude	The latitude of the TC at the beginning of the forecast
Initial longitude	The longitude of the TC at the beginning of the forecast
Initial intensity	The TC intensity at the beginning of the forecast
Storm speed	The speed of the TC at the beginning of the forecast
Predicted longitudinal displacement	The difference between the longitude at the forecast lead of interest and the longitude at initial time.
Predicted latitudinal displacement	The difference between the latitude at the forecast lead of interest and the latitude at initial time.
Forecast intensity	The NHC or JTWC forecast intensity at the forecast lead of interest.
Ensemble size	The number of ensemble members making up the consensus.

Table 1: Available GPCE-AX predictors. Predictors are extracted or derived from ATCF data files and are available to the GPCE-AX system in real time.

Forecast Lead	Atlantic Basin (2002-2007)		Western North Pacific Basin (2004-2006)	
	Along-track	**Across-track**	**Along-track**	**Across-track**
12hr	0	2	4	-1
24hr	-2	5	10	1
36hr	-5	9	15	3
48hr	-9	15	22	4
72hr	-19	28	37	4
96hr	-41	27	-2	0
120hr	-76	14	0	-3

Table 2: Across-track and along-track bias (in NMI) over the training periods for the Atlantic and western North Pacific basins. GPCE-AX is trained to predict the bias corrected consensus forecasts.

Forecast Lead	Atlantic Basin Across Track		Atlantic Basin Along Track	
	Predictor 1	**Predictor 2**	**Predictor 1**	**Predictor 2**
12hr	Across spread	Initial intensity	Along spread	Initial intensity
24hr	Across spread	Initial intensity	Along spread	Initial intensity
36hr	Across spread	Initial longitude	Along spread	Initial intensity
48hr	Across spread	Initial longitude	Along spread	Initial intensity
72hr	Across spread	Initial longitude	Along spread	Predicted long. displacement
96hr	Across spread	Initial latitude	Along spread	Predicted long. displacement
120hr	Across spread	Initial latitude	Along spread	Predicted long. displacement

Table 3: Atlantic basin GPCE-AX predictors based on a 2002-2007 training set.

Forecast Lead	Western North Pacific Basin Across Track		Western North Pacific Basin Along Track	
	Predictor 1	**Predictor 2**	**Predictor 1**	**Predictor 2**
12hr	Across spread	Forecast intensity	Along spread	Initial intensity
24hr	Across spread	Forecast intensity	Along spread	Forecast intensity
36hr	Across spread	Forecast intensity	Along spread	Forecast intensity
48hr	Across spread	Forecast intensity	Along spread	Forecast intensity
72hr	Across spread	Forecast intensity	Along spread	Forecast intensity
96hr	Across spread	Initial latitude	Along spread	Forecast intensity
120hr	Across spread	Initial latitude	Along spread	Forecast intensity

Table 4: Western North Pacific basin GPCE-AX predictors based on a 2004-2006 training set.

Forecast Lead	Atlantic Basin boost values (2002-2007 training)	Western North Pacific Basin boost values (2004-2006 training)
12hr	17	20
24hr	29	31
36hr	37	44
48hr	48	55
72hr	71	81
96hr	104	104
120hr	143	130

Table 5: Boost values added to the predicted error of the consensus to transform the predicted error value to the predicted 70% uncertainty isopleths. Boost values are directly added to the across-track predicted error, but are scaled before being added to the along-track directions in order to maintain eccentricity.

Forecast Lead	**Atlantic Basin, 2008**			
	GPCE Reliability	**GPCE-AX Reliability**	**Mean area difference**	**Statistically significant?**
12hr	79%	78%	-2%	No
24hr	79%	78%	-2%	No
36hr	76%	75%	4%	Yes
48hr	78%	76%	2%	Yes
72hr	76%	68%	12%	Yes
96hr	81%	72%	9%	Yes
120hr	79%	79%	4%	Yes

Table 6: Out of sample summary statistics for the Atlantic basin in 2008. GPCE-AX has greater reliability than GPCE, and the area bound by the 70% isopleth is systematically smaller for all leads greater than 24hrs.

Forecast Lead	**Western North Pacific Basin, 2007**			
	GPCE Reliability	**GPCE-AX Reliability**	**Mean area difference**	**Statistically significant?**
12hr	72%	74%	8%	Yes
24hr	73%	71%	0%	No
36hr	75%	73%	0%	No
48hr	74%	73%	2%	No
72hr	79%	83%	3%	Yes
96hr	79%	79%	4%	Yes
120hr	91%	85%	14%	Yes

Table 7: Out of sample summary statistics for the western North Pacific basin in 2007. While not as clear-cut at the Atlantic basin, GPCE-AX forecasts are more reliable than GPCE forecasts for a majority of forecast leads, and the area bound by the 70% isopleth is smaller than the GPCE areas for 12hr forecasts and all forecasts greater then 48hrs.