NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

USING BAYESIAN STATISTICAL POST-PROCESSING TECHNIQUES TO IMPROVE TROPICAL CYCLONE TRACK AND INTENSITY FORECASTS

by

Sabrina L. Cummings

June 2018

Thesis Advisor:	Wendell A. Nuss
Second Reader:	Eric A. Hendricks

Approved for public release. Distribution is unlimited.

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			*Form Approved OMB No. 0704-0188*

1. AGENCY USE ONLY *(Leave blank)*	**2. REPORT DATE** June 2018	**3. REPORT TYPE AND DATES COVERED** Master's thesis

4. TITLE AND SUBTITLE USING BAYESIAN STATISTICAL POST-PROCESSING TECHNIQUES TO IMPROVE TROPICAL CYCLONE TRACK AND INTENSITY FORECASTS	**5. FUNDING NUMBERS**
6. AUTHOR(S) Sabrina L. Cummings	
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)

This thesis examines the use of statistical post-processing techniques involving Bayesian estimation and Markov Chain Monte Carlo methods to aid in the reduction or elimination of tropical cyclone track and intensity forecast errors. The results of this research showed an improvement in the forecasts for intensity and total track error over the European Centre for Medium-Range Weather Forecasts (ECMWF) ensemble mean for all forecast times. These findings indicate that applying Bayesian statistical post-processing to forecasts made by the ECMWF ensemble can reduce the overall track and intensity error and result in more accurate forecasts. The most significant forecast improvement resulted from larger sample sizes and creative grouping schemes. By increasing the number of storms used and altering the manner in which the data is grouped, a more accurate forecast can be obtained. Future research using a larger sample size that spans several decades is indicated, but any significant physics alterations to the models over time, as well as more specific ways of grouping the data, must be taken into consideration.

14. SUBJECT TERMS Bayes, Bayesian, statistics, statistical, tropical cyclone, hurricane, track, intensity, forecast, weather, post-processing.	**15. NUMBER OF PAGES** 63
	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** UU

NSN 7540-01-280-5500 Standard Form 298 (Rev. 2-89) Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release. Distribution is unlimited.

USING BAYESIAN STATISTICAL POST-PROCESSING TECHNIQUES TO IMPROVE TROPICAL CYCLONE TRACK AND INTENSITY FORECASTS

Sabrina L. Cummings
Lieutenant Commander, United States Navy
BS, Old Dominion University, 2008

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN METEOROLOGY AND PHYSICAL OCEANOGRAPHY

from the

NAVAL POSTGRADUATE SCHOOL
June 2018

Approved by: Wendell A. Nuss
Advisor

Eric A. Hendricks
Second Reader

Wendell A. Nuss
Chair, Department of Meteorology

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

This thesis examines the use of statistical post-processing techniques involving Bayesian estimation and Markov Chain Monte Carlo methods to aid in the reduction or elimination of tropical cyclone track and intensity forecast errors. The results of this research showed an improvement in the forecasts for intensity and total track error over the European Centre for Medium-Range Weather Forecasts (ECMWF) ensemble mean for all forecast times. These findings indicate that applying Bayesian statistical post-processing to forecasts made by the ECMWF ensemble can reduce the overall track and intensity error and result in more accurate forecasts. The most significant forecast improvement resulted from larger sample sizes and creative grouping schemes. By increasing the number of storms used and altering the manner in which the data is grouped, a more accurate forecast can be obtained. Future research using a larger sample size that spans several decades is indicated, but any significant physics alterations to the models over time, as well as more specific ways of grouping the data, must be taken into consideration.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF FIGURES

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS AND ABBREVIATIONS

ATCF	automated tropical cyclone forecast system
BEMOS	Bayesian ensemble model output statistics
DSCA	defense support of civil authorities
ECMWF	European Centre for Medium-Range Weather Forecasts
EMOS	ensemble model output statistics
GFS	Global Forecast System
GPCE	Goerss predicted consensus error
JTWC	Joint Typhoon Warning Center
MC	Monte Carlo
MCMC	Markov chain Monte Carlo
MOS	model output statistics
MVN	multivariate normal distribution
NHC	National Hurricane Center
NWS	National Weather Service
NWP	numerical weather prediction
PPD	posterior predictive distribution
STE	intensity error
TTE	total track error
THORPEX	The Observing system Research and Predictability Experiment
TIGGE	the international grand global ensemble
WPC	Weather Prediction Center

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

A. OVERVIEW

Tropical Cyclone forecast accuracy has improved over the last several decades. Despite significant improvements, current tropical forecast models are limited in the skill needed to accurately forecast tropical systems, especially at extended time periods. Figures 1–4 show that although there is definite improvement in track (figures 1 and 3) and a smaller improvement in intensity (figures 2 and 4) over the last few decades, there is still a significant amount of error at all forecast hours and the errors are quite large at the extended forecast hours. Although the damage from a tropical cyclone cannot be avoided, the amount of damage and number of lives lost can be greatly reduced with proper preparation and, when needed, early evacuation and/or Sortie.

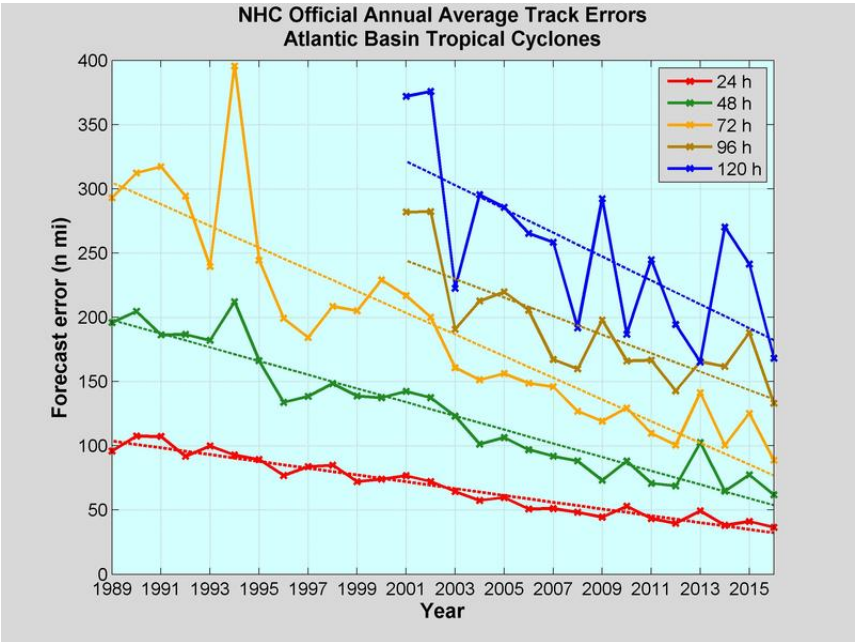


Figure 1. Average track errors. Source: National Hurricane Center (2017).

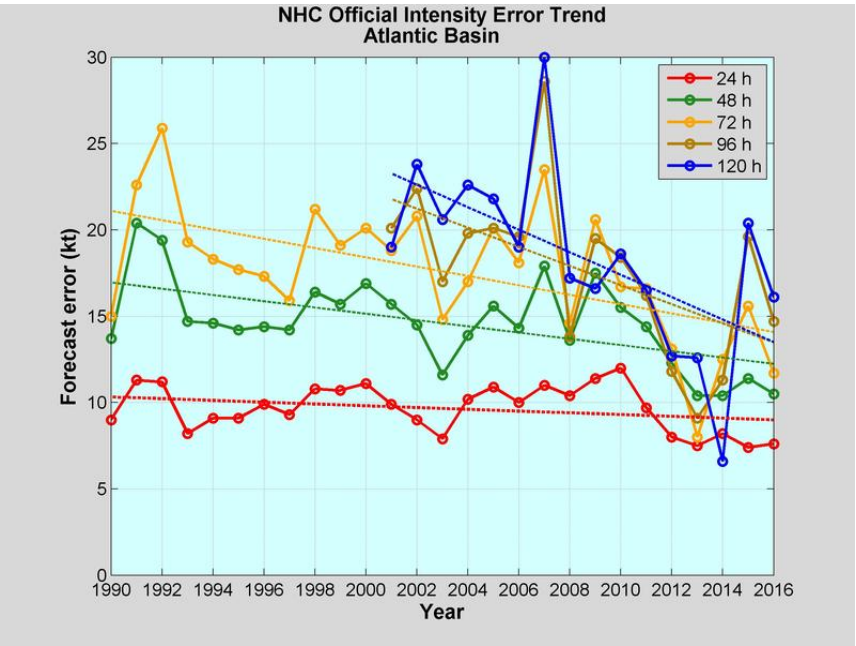


Figure 2. Intensity error trends. Source: National Hurricane Center (2017).

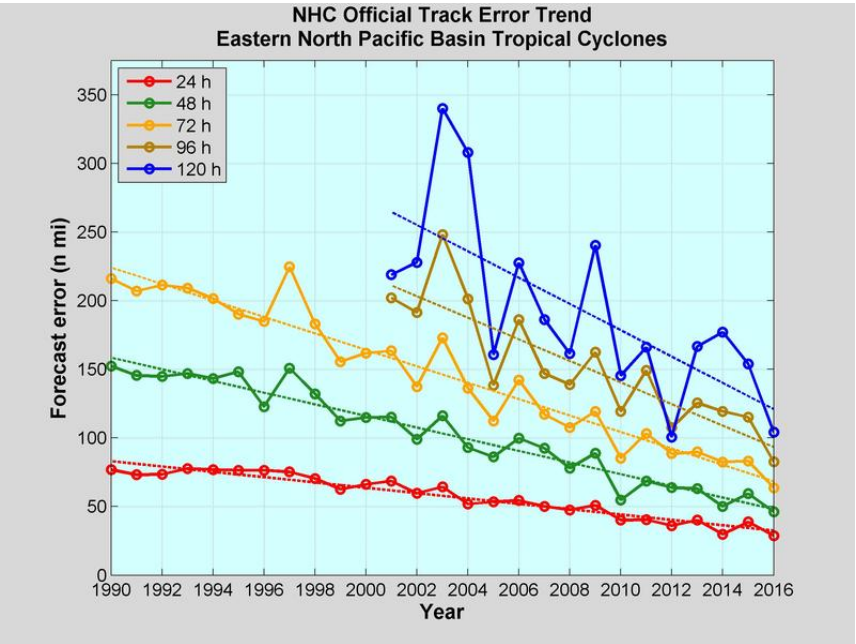


Figure 3. Average track errors. Source: National Hurricane Center (2017).

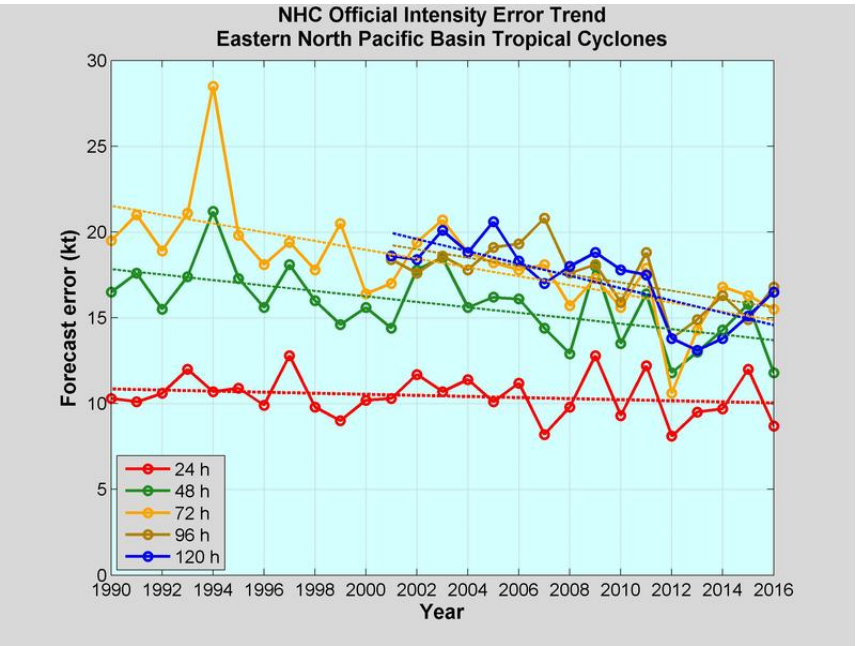


Figure 4. Intensity error trends. Source: National Hurricane Center (2017).

B. MOTIVATION

Improvements to tropical cyclone forecasts have obvious applications to both military and civilian operations. One clear advantage would be lowering the cost associated with hurricane damage. Although the path of a hurricane cannot be altered, knowing more precisely where a hurricane will make landfall would save a large amount of money. It takes a significant amount of time to evacuate a city, sortie an air wing or sortie a naval port. The decision to do these things must therefore be made far in advance. Due to poor forecasts at extended periods, it is not uncommon to sortie an air wing or naval port and consequently have the hurricane take a track in another direction. A good example of this is Hurricane Joaquin in 2015. The 5-day forecast for Joaquin on the evening of Wednesday September 30 predicted Joaquin to make landfall in Norfolk as a category 4 hurricane in the early afternoon on Sunday, October 4. Based on this and other early forecasts, naval leadership chose to sortie the fleet in Norfolk to avoid the increased winds and seas forecasted. With more accurate forecasts unnecessary evacuations and/or sorties could be avoided. Too often in recent decades the weather community warns of landfall in a particular location and encourages citizens to prepare for landfall but the

storm shifts its track or weakens in intensity leading to much less severe conditions than were forecast. This error leads to a false sense of security for civilians and military when faced with subsequent storms. It also leads to overall lower confidence in the forecaster's ability. More accurate models with accurate forecasts will lead to a higher level of confidence in the forecast when told to evacuate and citizens will therefore trust the forecast and evacuate. Knowing the track and intensity of the hurricane with more certainty could reduce hurricane related deaths. With enough advance notice cities in the path of dangerous hurricanes can complete an evacuation. Advance notice also allows cities to prepare for the impact of the storm resulting in less property damage as well.

More accurate hurricane forecasts can also have implications for military operations. As already stated, not having to waste money on unnecessary sorties would be beneficial. Ensuring that ships and aircraft always sortie when in the path of a destructive storm would save a large amount of money as well. Another advantage would be better planning for DSCA (Defense Support of Civil Authorities) operations, especially in situations with more than one storm in a basin. At the height of the Atlantic hurricane season it is not at all uncommon to have three or more storms in the basin at one time. If a storm has already impacted an area and naval forces are attempting to assist the affected location during DSCA operations, then it becomes vital to know the path of other storms in the basin which could affect those forces.

To be sufficiently prepared for destructive tropical systems we need to improve forecasting skill for tropical cyclone track and intensity. This thesis hypothesizes that applying statistical post-processing with Bayesian inference and Markov chain Monte Carlo sampling methods to ensemble numerical weather prediction (NWP) model forecasts (Wendt 2017) will yield more accurate forecasts at all forecast hours.

C. OBJECTIVE

This thesis aims to improve tropical cyclone forecasts for both track and intensity by applying Bayesian statistical post-processing techniques with Bayesian inference and Markov chain Monte Carlo (MCMC) sampling methods (Wendt 2017) to a given dataset. The approach will be to take data from the ECMWF (European Centre for Medium-

Range Weather Forecasts) 50-member ensemble model from 2010–2016 hurricane seasons and utilize that data as the "learning period" from which different statistical distributions can be created. Several predictor variables will be utilized in this process to determine correlations between forecasted and actual conditions. The correlations realized during the learning period will then be applied to the 2017 Atlantic hurricane season to assess the accuracy of this method. The goal is to use statistical postprocessing of ensemble NWP output to more accurately forecast tropical cyclone track and intensity.

THIS PAGE INTENTIONALLY LEFT BLANK

II. BACKGROUND

A. DETERMINISTIC MODELS

Deterministic tropical models offer valuable information about the behavior of tropical cyclones, yet there is still a large amount of error for both track and intensity. There are several sources of error in numerical weather prediction, the causes come from four main areas: "(i) imperfect initial conditions, (ii) insufficient model resolution, (iii) limits of the representation of physical processes, and (iv) limits of predictability" (Hendricks 2011). The main sources of error come from the limits of the representation of physical processes and imperfect initial conditions. Deterministic weather models forecast the future state of the atmosphere by solving mathematical equations of motion given some initial state. Since we do not know the physics of a tropical cyclone exactly, these equations of motion must work off of assumptions and simplifications. The equations of the atmospheric model are nontrivial and the solutions of the physical processes are therefore approximated, which lead to deficient forecasts (Richter 2012). Any assumption in a mathematical model will inevitably result in some amount of error in the final solution. In weather models, this error increases as the forecast time increases as seen in Figures 1–4.

One obvious approach to improve forecasts would be to develop more accurate model physical parameterizations. This has been the focus of the meteorological research community for decades, and although significant improvements have been made with each model update, error has only been decreased but not eliminated. Error can never be completely eliminated in deterministic models due to chaotic behavior of the primitive equations (Lorenz 1963). Also to note, each deterministic model update requires more computing power and therefore more money and resources

The other main source of error associated with deterministic weather models comes from incorrect initial conditions. Even a small error in the initial conditions will grow to produce large errors at extended time frames. One major advancement that addresses this in recent decades is the creation of weather ensembles. Ensembles work off

the notion of perturbing the initial conditions in an attempt to get a more accurate forecast. Ensemble systems have been proven to improve deterministic forecasts and predict forecast skill (Gneiting et al. 2004). Ensembles are useful due to the limits of predictability and they help quantify the uncertainty in the forecast, which is something deterministic models cannot do. Although weather model ensembles have proved to increase accuracy, they are still unable to forecast tropical systems accurately enough and often contain biases.

B. STOCHASTIC MODELS

Stochastic models operate differently than deterministic ones. They do not solve equations of motion to forecast the future state of the atmosphere but instead utilize probabilities of future conditions based on errors inherent in a chosen model. A stochastic model will utilize a given dataset which is referred to as the "learning data" to calibrate itself and identify relationships and then apply what has been learned to predict the future state. These types of models are substantially less expensive to run and do not possess the same types of error that exist with deterministic models. One way to increase forecast accuracy could be to utilize stochastic methods on ensemble models.

C. PREVIOUS RELEVANT RESEARCH

There have been numerous methods explored in recent decades that aimed to improve the uncertainty and inaccuracy in tropical cyclone forecasts. Chisler (2016), Neese (2010), and Hauke (2006) all looked at different divisions of tropical data that could address inaccurate forecasts. Chisler's work focused on grouping forecasts based on ranges of uncertainty estimated from ensemble spread and he found that by utilizing a larger number of groupings tropical cyclone forecasts could be improved. Neese examined the National Hurricane Center's (NHC) wind speed probabilities product and looked specifically at the possibility of improving the product using different distributions of track errors. He examined grouping the data by geographic location of storms to determine if that would result in a smaller variance. Neese did find that the product could possibly be improved if the Monte Carlo (MC) method was used with track error distributions based on storm location. Hauke examined NHC's Monte Carlo (MC)

model which uses historic track error distributions. He found that the NHC probability model could be improved by utilization of the Goerss Predicted Consensus Error (GPCE).

Another relevant area of research that this thesis draws from has to do with Bayesian ensemble model output statistics and specifically how those are applied to atmospheric predictions. Wendt (2017) explored the effectiveness of statistical post-processing methods using a hierarchical multivariate Bayesian approach to ensemble model output statistics. This thesis will follow the work of Wendt but apply the approach specifically to forecasting for tropical cyclones. This thesis will use Wendt's method of Bayesian inference using Markov Chain Monte Carlo (MCMC) methods to produce calibrated multivariate posterior predictive distributions (PPD) for 12, 24, 36, 48, 72, 96 and 120-hour forecasts of TC track and intensity and will then group the results based on key meteorological parameters of interest.

THIS PAGE INTENTIONALLY LEFT BLANK

III. METHODOLOGY

A. THEORY

British statistician George Box is famous for saying “All models are wrong, but some are useful” (Clear 2017). Box’s quotation is the basis for the theory of this thesis. As already discussed in Chapter I, deterministic models are wrong no matter how sophisticated the model physics may be or how good the resolution is. That does not mean those forecasts are not useful. A good example of this is operational weather forecasting. For the most part forecasters have access to all the same model data, yet no two forecasters will ever forecast the exact same thing. More experienced forecasters traditionally do much better than novice forecasters. This is because experience teaches forecasters which models do well in different situations and what the consistent model performance characteristics are. These model performance characteristics are often referred to as “model biases” or “model tendencies.” The majority of numerical weather prediction (NWP) models publish known model biases on their operational websites and update them regularly. Table 1 is an example of a model bias chart for GFS, taken from the National Weather Service’s (NWS) Weather Prediction Center (WPC) website.

As shown in Table 1, model tendencies can be identified over decades of utilizing the same model products. Human forecasters who look at these products daily can recognize these biases and create their own thumb rules to account for the biases in their operational forecasts. For example, from row 1 of Table 1, GFS tends to over forecast low-pressure systems crossing the Sierras, therefore an experienced forecaster would alter their forecast to show a weaker system with weaker winds than may be indicated by the model and would also advect the system more slowly than the model indicates while also taking into account what that slower movement will do to modify the air mass. Some forecasters even assign specific numbers to their thumb rules like advecting the surface low pressure system at 50% of the 500-mb winds or 75% of the 700- mb winds. Just as human forecasters can learn these tendencies and apply them to raw model data, the idea of this thesis is to explore if statistically-based weather models can do the same thing. This thesis is essentially seeking to formalize useful rules of thumb.

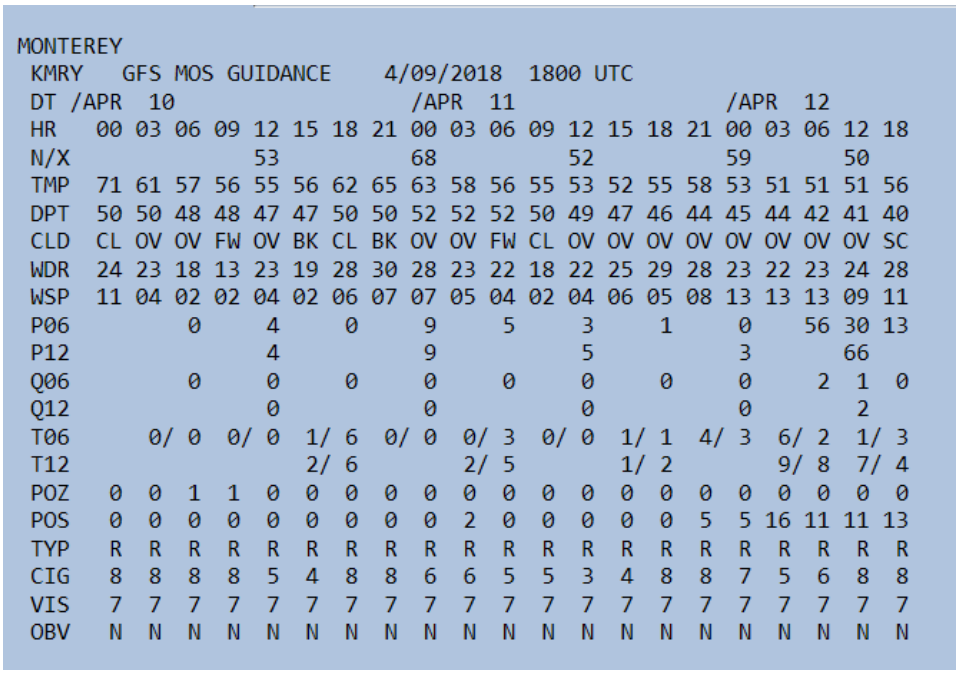
Table 1. GFS operational model bias chart. Source: GFS (2013)

Subjectively Observed Bias	***Geographical location of bias***	*Annual/ Diurnal attribute*	***Submitted by***	***Date Submitted***	***Operational Implication***	***Suspected Cause***
GFS too ambitious with strength and speed of systems crossing Sierras after fhr 36	SW US	Cool season	USGS	Dec 2005	Too progressive and strong with systems crossing Sierras	Model resolution of topography
Dry bias north of areas where over 2" of QPF has been produced in a 6hr period	Primarily east of front range and west of Appalachians	Warm season, any time of model day	NCEP WPC	Spring of 1998	QPF produced from convective feedback blocks northward advection of moisture	Result of GFS Convective Parameterization Scheme
QPF verification historically better than Eta	CONUS	Cool Season only	NCEP WPC	1999	Rely more heavily on QPF from GFS - especially beyond 36 hours	GDAS better than EDAS ?
Aerial coverage of QPF and mass fields over done (QPF at low thresholds .01" and .10")	CONUS	Anytime	NCEP WPC	Since mid 1990s	Over forecast of aerial coverage of precip can lead to high bias in PoPs	Model resolution (the lower the resolution the more geography a QPF pattern can get spread over)
Slightly ambitious with magnitude of high amplitude patterns	North America	Cool season so far	NCEP WPC	Since fall 2002	Prediction of southward progression of cold air over done. Model a bit too extreme in temp patterns beyond 84 hours. Precip Type Algorithm off of GFS too eager to depict snow	?
Ambitious to phase northern and southern stream systems in fast and spit flow patterns beyond fhr 84	North America	Cool season	NCEP WPC	Cool season 2001 (not noticed yet in 2002)	Over forecast of cyclogenesis east of Rockies	Suspect related to model resolution and lack of dense obs data where associated systems originate in forecast cycle
Major difference in QPF forecast than ETA	CONUS	Warm season	NCEP WPC	Since mid-1990s	Lack of run to run continuity in QPF	Different convective parameterization s between models

The goal of this thesis is to determine if it is possible to teach a model to "learn" from model tendencies as a human forecaster would and then apply those learned tendencies to the raw model data to create a forecast that is more accurate. The idea behind this is fairly simple, but the math needed to accomplish this task is not as straightforward. This thesis will look at using statistical post-processing to aid the model in learning, and then to use that learning to provide improved tropical cyclone intensity and track forecasts. The specific type of post-processing utilized will be Bayesian statistical post-processing using Markov Chain Monte Carlo methods, which will be discussed in the next section.

B. BAYESIAN ESTIMATION MODEL OUTPUT STATISTICS

As already discussed, deterministic models suffer from a number of error sources that contribute to inaccurate forecasts. The atmosphere is so dynamic that deterministic models are unable to fully account for all the various processes that are taking place. The use of model output statistics (MOS) has been around for decades and is one way to account for the incorrect model physics and correct biases deterministic models that result from all sources of error. The basic idea is to take archived forecasts and actual observed conditions and determine what correlations exist between the 2 datasets. MOS uses multiple linear regression schemes to determine relationships and then apply those relationships to the deterministic model predictions. MOS data has proven to output more accurate predictions than the raw model output. Figure 5 shows an example of MOS output for Monterey, CA.



```
MONTEREY
 KMRY   GFS MOS GUIDANCE    4/09/2018  1800 UTC
 DT /APR  10                  /APR  11                /APR  12
 HR   00 03 06 09 12 15 18 21 00 03 06 09 12 15 18 21 00 03 06 12 18
 N/X              53          68          52          59       50
 TMP  71 61 57 56 55 56 62 65 63 58 56 55 53 52 55 58 53 51 51 51 56
 DPT  50 50 48 48 47 47 50 50 52 52 52 50 49 47 46 44 45 44 42 41 40
 CLD  CL OV OV FW OV BK CL BK OV OV FW CL OV OV OV OV OV OV OV OV SC
 WDR  24 23 18 13 23 19 28 30 28 23 22 18 22 25 29 28 23 22 23 24 28
 WSP  11 04 02 02 04 02 06 07 07 05 04 02 04 06 05 08 13 13 13 09 11
 P06         0     4     0     9     5     3     1     0    56 30 13
 P12               4           9           5           3       66
 Q06         0     0     0     0     0     0     0     0     0     2  1  0
 Q12               0           0           0           0        2
 T06      0/ 0  0/ 0  1/ 6  0/ 0  0/ 3  0/ 0  1/ 1  4/ 3  6/ 2  1/ 3
 T12                  2/ 6        2/ 5        1/ 2        9/ 8  7/ 4
 POZ   0  0  1  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
 POS   0  0  0  0  0  0  0  0  0  2  0  0  0  0  0  5  5 16 11 11 13
 TYP   R  R  R  R  R  R  R  R  R  R  R  R  R  R  R  R  R  R  R  R  R
 CIG   8  8  8  8  5  4  8  8  6  6  5  5  3  4  8  8  7  5  6  8  8
 VIS   7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7  7
 OBV   N  N  N  N  N  N  N  N  N  N  N  N  N  N  N  N  N  N  N  N  N
```

Shows a numerical forecast of key meteorological parameters over time.

Figure 5. GFS MOS guidance for Monterey, CA. Source: GFS (2018).

In 2004, Gneiting et al. explored the use of ensemble model output statistics (EMOS) and proved that the EMOS predictions were sharper and better calibrated than the raw ensemble (Gneiting et al. 2004). In 2012 Richter et al. applied a Bayesian model to the idea of EMOS and cited calibrated forecasts with good performance when applied to the University of Washington's mesoscale ensemble over the North American Pacific Northwest (Richter et al. 2012). This thesis will continue the work done using MOS and utilize the NPS Bayesian ensemble model output statistics (BEMOS) model adapted by Wendt in 2017. Wendt found that by directly parameterizing meteorological phenomena with probability distributions describing the structure of the data, calibrated forecasts can be produced which can outperform the parent ensemble (Wendt 2017). The goal of this thesis is to apply the NPS BEMOS model to raw ECMWF ensemble output for tropical systems to determine if it is possible to improve upon the ECMWF ensemble forecasts.

C. ECMWF ENSEMBLE

This thesis will utilize archived data from the ECMWF (European Centre for Medium-Range Weather Forecasts) ensemble via the TIGGE (The International Grand Global Ensemble) database. The TIGGE database was originally developed as an integral component of the THORPEX (The Observing system Research and Predictability Experiment) research project but is now utilized mostly for scientific research. TIGGE has been utilized for a wide range of research projects, including research on ensemble weather forecasting and prediction of severe weather (ECMWF 2006). The database consists of the ECMWF global ensemble model datasets as far back as 2006 and has 50 members.

The data utilized in this research will be the 2010–2017 tropical season datasets for the Atlantic Basin. The data for years 2010–2016 will be used as the "learning period" for the model. However, it is important to note that a larger training dataset would be better to capture the interannual variability of TC track and intensity. In this case the datasets prior to 2010 did not contain all the data needed to conduct testing. The model will be able to "learn" from this data and will become calibrated for the test data. The calibrated model will then be tested on the 2017 tropical season data. For the purpose of this research only hurricanes (maximum sustained surface winds of 64 kt or greater) will be considered, all other storms will be thrown out. The reason behind this is that tropical depressions and tropical storms typically have larger track and intensity mean error distributions that may not reflect hurricane errors and we would therefore not be able to extract meaningful information from their distributions. If a tropical system was not forecasted to be at least hurricane strength (64 kt) then it will not be utilized. Also to note, in order for a specific forecast to be included it must have been carried by at least 10 of the 50 ensemble members as a hurricane for that forecasting hour. The datasets were filtered based on initial intensity, if a hurricane-strength storm subsequently weakened to a tropical storm or depression it is still included in the dataset. Since model error increases as the forecast hour increases, the model was tested by considering each forecast hour separately: 12, 24, 36, 48, 72, 96 and 120 hours, respectively, to determine

if forecast performance is forecast length dependent. The ensemble mean, spread and error were calculated for each forecast hour and used to calibrate the model for the 2017 data.

D. VERIFICATION DATA

1. A-Decks

The "A-Decks" contain the official NHC track and intensity forecasts as well as other guidance information. This file contains the complete listing of all available forecast products for the storm. The A-Deck data can be found in the Automated Tropical Cyclone Forecast System (ATCF) database on NHC's website. ATCF is an "IBM-AT compatible software package developed for the Joint Typhoon Warning Center (JTWC) and designed to assist forecasters with the process of making tropical cyclone forecasts" (Sampson, 1990). The A-Deck files are utilized to determine what the official forecast was for a particular storm.

2. B-Decks

The "B-Decks" are commonly referred to as the "best track data." This file contains the history of the storm's location, intensity and all other parameters at the synoptic forecast times (00, 06, 12 and 18 UTC). During the hurricane season this file contains the best estimate of storm parameters but after the season is complete the files get updated with revised information that is more accurate.

E. MODEL STRUCTURE

As stated in section A, in order to "teach" the ECMWF model this thesis will utilize statistical post processing of the ECMWF model data. The specific method will be Bayesian estimation using a Markov chain Monte Carlo sampling scheme. Bayesian estimation is an application of Bayes' rule, which is shown in Equation 1. Many machine learning algorithms rely on Bayes' rule because it can easily be applied to a wide range of problems. Bayes' rule is based upon the idea of conditional probability, which gives the probability of one event occurring given that another event has occurred. This is given in Equation 1

$$P(A|B) = \frac{P(B|A)}{P(B)} P(A), \quad (1)$$

where P(A) is the probability of event A occurring, P(B) is the probability of event B occurring, $P(A|B)$ is the probability of event A given that event B occurred and $P(B|A)$ is the probability of event B given that event A occurred. The NPS BEMOS model will utilize predictor variables from the ECMWF forecasts and compare those to the observed data from the ATCF B-decks. The model will assess the relationship between the forecasted ensemble spread in track and intensity forecasts and observed track and intensity errors and create output that accounts for the differences. Figure 6 shows a diagram explaining the model processes.

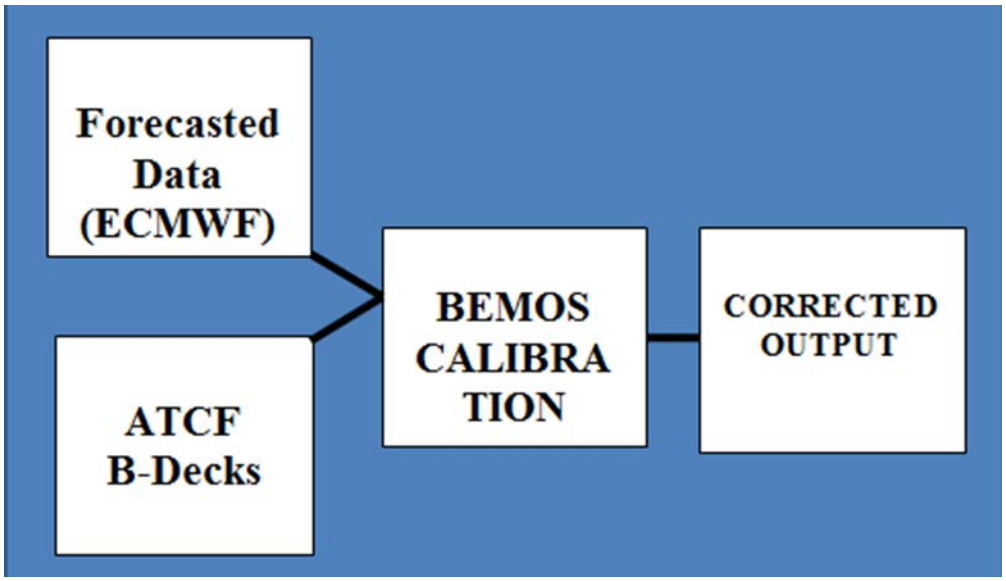


Figure 6. NPS BEMOS Model process flow diagram

This first step of the model is to read in the master file. This master file consists of the ECMWF track, cross-track, along-track and intensity ensemble means and spread, the observed forecast track and intensity errors of the ensemble mean derived from the B-decks, and storm source region data. Next the model selects the columns that correspond to the needed predictor variables for the multivariate predictions. The model then

calculates the log-likelihood function to evaluate the normal log priors with the fixed μ and σ. The MVN (multivariant normal distribution) log-likelihood function can be written as

$$l(B, \sum) = -\frac{n}{2} ln|\sum| - \frac{1}{2} \sum_{i=1}^{n} (y_i^T - x_i^T B) \sum^{-1} (y_i^T - x_i^T B)^T + C, \quad (2)$$

where "l is the log-liklihood of data with n forecast instances, y_i^T is a 1xM vector of observations, x_i^T is a 1xK vector of independent predictors, B is a KxM parameter matrix of regression coefficients, (including the intercepts) and $\sum$ is an MxM covariance matrix that replaces the univariant variance" (Wendt 2017). The derivation of this equation for the purpose of Bayesian multivariate linear regression is beyond the scope of this thesis but is available in "A Hierarchical Multivariate Bayesian Approach to Ensemble Model Output statistics in Atmospheric Prediction" (Wendt 2017). The log-likelihood function is used to describe the probability of a model parameter given an observed dataset. In Bayesian Inference you can describe the likelihood of any random variable given observed data. Using Bayes rule, we can multiply the calculated likelihood by the prior probability to get the Bayesian posterior probability as shown in Equation 3.

$$p(\theta | y) \propto p(y | \theta) p(\theta) \quad (3)$$

Where y represents the observed data and θ represents the set of statistical parameters being solved for. Next, the NPS BEMOS model utilizes an adaptive Metropolis sampler with a MVN kernel function and block updates. The basic idea of the Metropolis sampler is that it takes random samples from the posterior, performs Monte Carlo integration and then calculates parameters of interest from the sample. The Metropolis algorithm can be applied to almost any data. "An adaptive variant of the random-walk Metropolis algorithm" was used in the NPS BEMOS model to "complete the inference scheme with block-wise multiparameter updates, to produce calibrated posterior predictive distributions (PPD)" (Wendt 2017).

F. EVALUATION OF RESULTS

It is important to test whether the training dataset and resultant testing dataset are statistically different from one another. The test data should show lower error than the training data if improvement in the forecast results. The Student's t-test was utilized to accomplish this. The Student's t-test compares two means and determines if they are different from one another, by accounting for the variance in each sample. It can also tell you how different the two are. The Student's t-test will output two values, the t-value and the p-value.

The t-value is the ratio between the difference of the two groups and the difference within the two groups. A higher t-value means there is a large difference between the two groups, a lower t-value indicates that there is similarity between the two groups. For this test, it was important to see a larger t-value to determine that the two groups were statistically different. The p-value tells you the probability that the results from your sample could occur by chance. P-values range from 0% to 100%. If the data has a low p-value that indicates the data did not occur by chance. For the purposes of this thesis, it was important to see low p-values. For the purposes of this thesis, the confidence interval was set at 95%, so any p-value greater than .05 would mean the null hypothesis should be rejected. Tables 2-5 show all p-values. For the larger datasets in which data was broken into forecast hours, the p-values show high confidence levels and the data passed the test with no issues. For the datasets that were divided into regions and forecast hour within the region the datasets were too small and they failed the test. The p-values for those cases were larger and the null hypothesis was therefore rejected in those cases.

Tests were run on the data using various grouping schemes. Data was grouped by forecast hour (12, 24, 36, 48, 72, 96, 120 hours), by source region, and also by forecast hour within each source region. The entire data set was also tested without being grouped by forecast hour. For the source regions the Atlantic Basin was divided into four main regions: the Atlantic (AT), the Caribbean (CA), the Gulf of Mexico (GO) and the region near West Africa (WA). Each source region is labeled on the map in Figure 7. The GO and CA regions are easily identifiable on the map. For the AT and WA regions, they were

separated by 25°N latitude and 45°W longitude. A Student's t-test was conducted for each of the forecast hour runs, the source region runs, the forecast hours within each source region runs and the entire data set. The results are summarized in Chapter IV, Table 2. These results will be discussed in depth in Chapter IV.

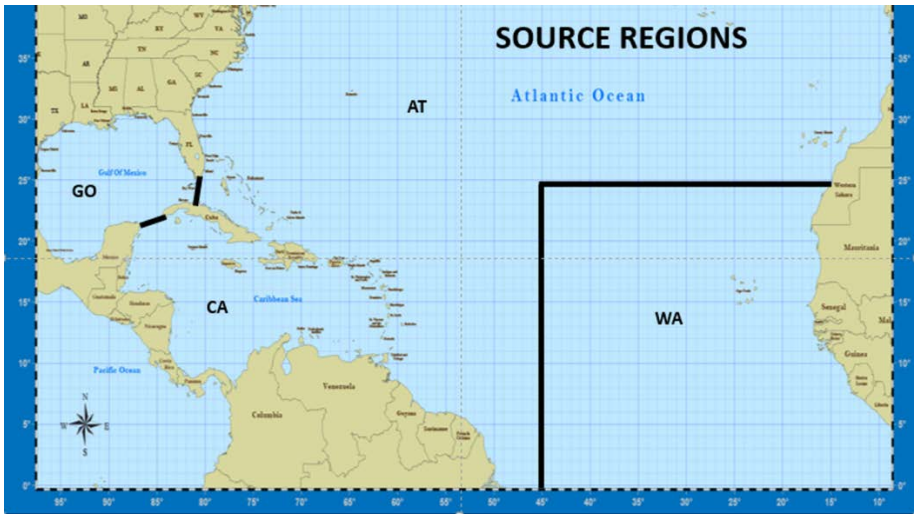


Figure 7. Map of source regions utilized when grouping data by source region

IV. RESULTS

A. OVERVIEW

Bayesian statistical post-processing techniques that utilize MCMC methods were shown to remove the forecast error bias in most cases and reduce the spread in some resulting in overall lower error percentages. There were some marked advantages to how the data was grouped. Another item to note was the length of the "learning period." The longer learning periods resulted in better results with less error. When the model only had a limited amount of data to include in the learning period the results were not as successful and in some cases there was no improvement at all.

B. DATA GROUPING

1. Forecast Hour

The data was grouped in several different ways in order to determine what statistical relationships exist and may be utilized to improve forecast results. Data was first grouped by forecast hour in order to determine if there was a dependence on forecast length. Figures 8 and 9 show total track error (TTE) for each forecast hour and figures 10 and 11 show intensity error (STE) for each forecast hour. For TTE, the error in nautical miles is along the x-axis and for STE, the intensity error in millibars is along the x-axis. In both figures the red represents the unadjusted error and the blue represents the corrected error. The data was fit to a Gaussian distribution as seen in the figures. Some of the plots show a Gaussian centered at a negative value, which can be misleading. In the instances where the Gaussian is centered at a negative value it means that the model over corrected in that case. The model applies the correction that it believes will remove the bias. In cases where the model over estimates the required correction, the result will be a Gaussian centered at a negative value due to over correction of the raw model data. This happens in several cases. Results are summarized in Tables 2 and 3.

The unadjusted error is from the raw model output and is represented in the

graphs by the red lines. The blue lines represent the error once the BEMOS model created correction has been applied to the raw forecast. In order for there to be error reduction two things would be expected. The first would be for the forecast error spread to be less for the corrected error than for the unadjusted error. More simply stated, when the blue lines are narrower than the red lines, or more "peaked," there is less spread in the data and therefore improvement in the track error. The second indication that there has been improvement would be for the peak to be centered at zero on the graphs. A Gaussian distribution centered at zero indicated the mean is zero and the model bias has been removed.

When grouping by forecast hour there was improvement for both track and intensity. For TTE, the bias was removed or reduced for all forecast hours and the forecast error spread was reduced for some forecast hours. A reduced forecast error spread means less error. For the 12-, 24-, 72-, and 96-hour forecasts there was not a significant decrease in the ensemble spread. When analyzing deterministic models it is expected that the ensemble spread increases as forecast hour increases. In this case, the expectation would be quite different since the NPS BEMOS model 'does not necessarily care if the deterministic forecast was accurate, it works by determining how inaccurate it is on average and applying a correction to account for that inaccuracy. Therefore, if the BEMOS model is working as expected then there would not necessarily be any increase in forecast spread as forecast hour increases, Figure 8 (a) – (d) illustrate this point well. Figure 8 (a) is the 12-hour forecast and shows a much larger spread than the 24-, 36-, or 48-hour forecasts; this is a drastically different outcome from what would be expected with a deterministic forecast. This is one way stochastic models could outperform deterministic models. As already noted, deterministic models struggle to forecast accurately at the extended forecast hours due to less-than-perfect initial conditions and model physics, and because those errors carry forward with each forecast hour. A stochastic model like the NPS BEMOS model would not suffer from the same drawbacks.

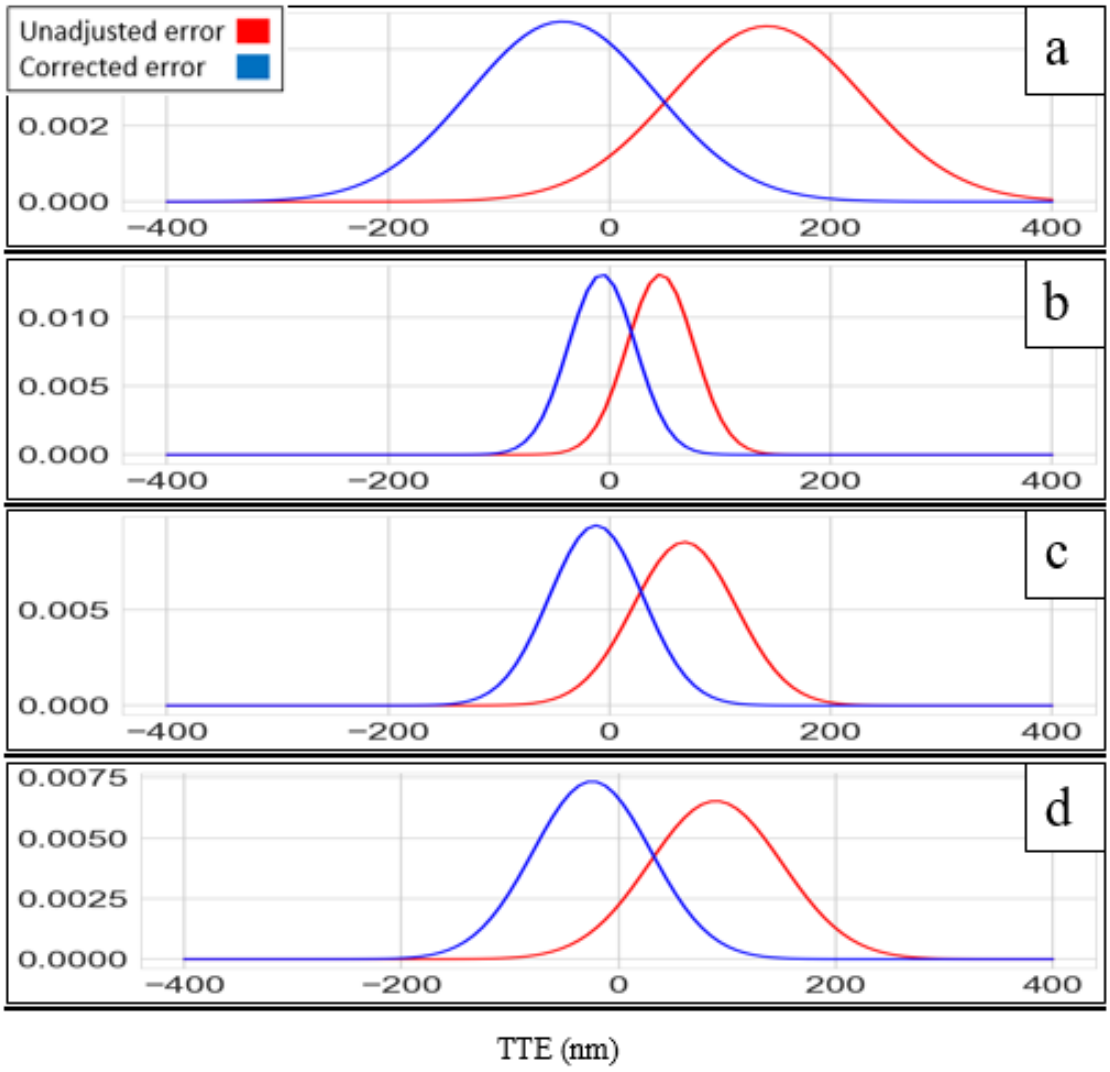


Figure 8. Total track error in nm for (a) 12-hour forecasts, (b) 24-hour forecasts, (c) 36-hour forecasts and (d) 48-hour forecasts

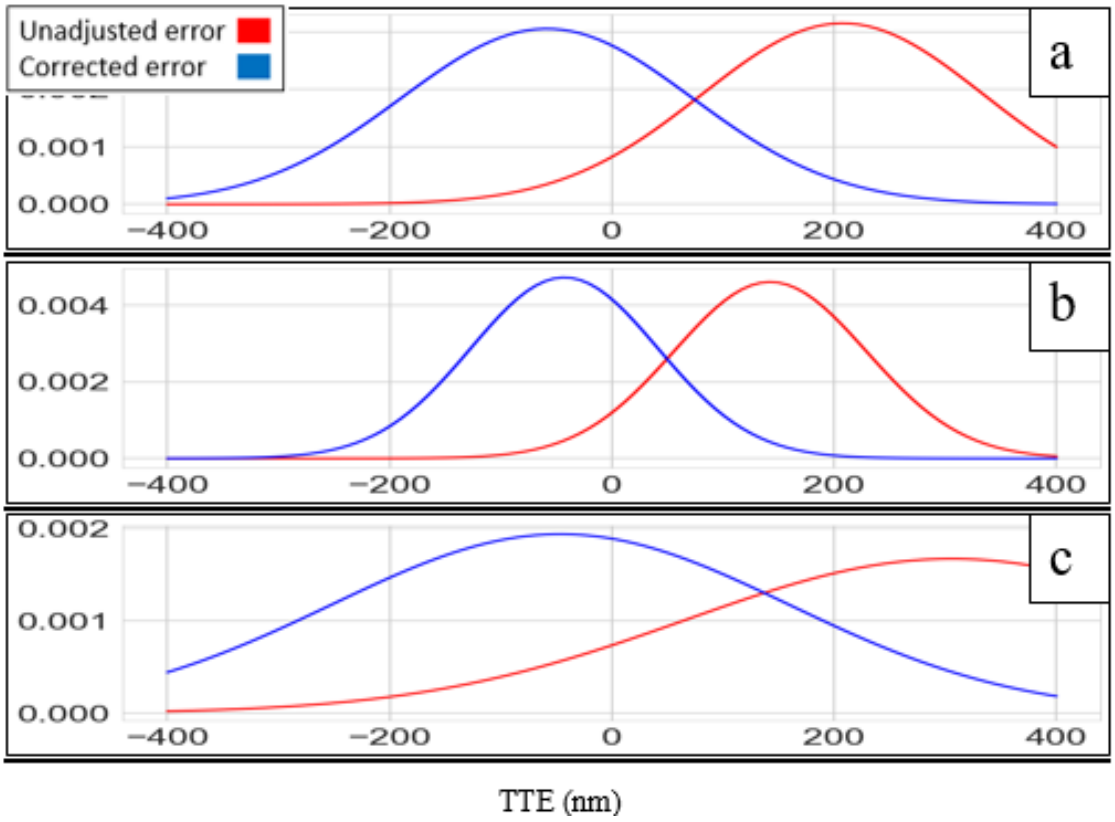


Figure 9. Total track error in nm for (a) 72-hour forecasts, (b) 96-hour forecasts and (c) 120-hour forecasts

For STE there was improvement in ensemble mean for all forecast hours but there was no improvement in forecast error spread. Figures 10 and 11 show no increase in error as forecast hour increases. Again, with a deterministic model it is expected to see the intensity error increase as forecast hour increases; that is not the case for a Bayesian model. The results therefore indicate that the Bayesian model corrections when grouping the data by forecast hour can offer a significant improvement over raw model output.

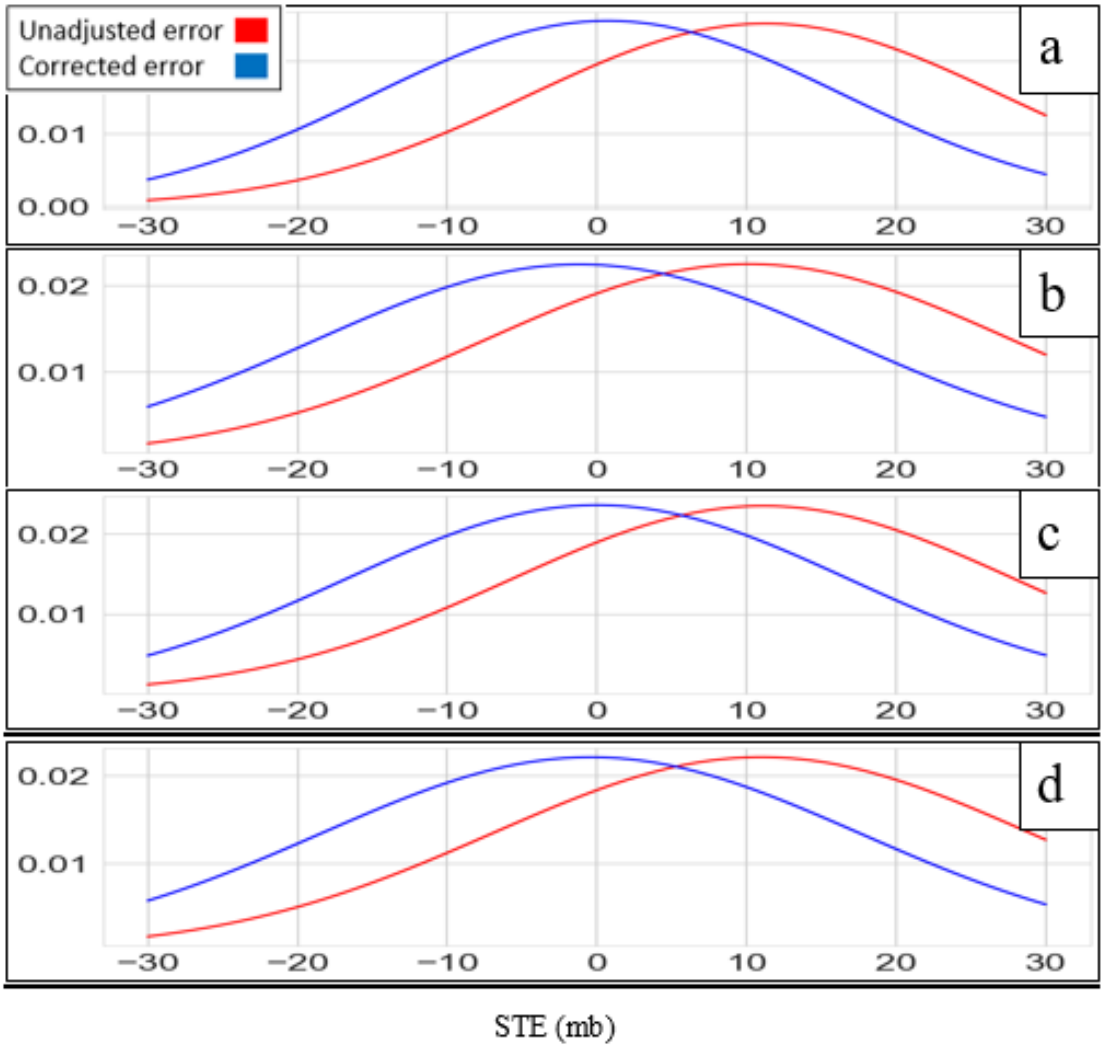


Figure 10. Intensity error in mb for (a) 12-hour forecasts, (b) 24-hour forecasts, (c) 36-hour forecasts and (d) 48-hour forecasts

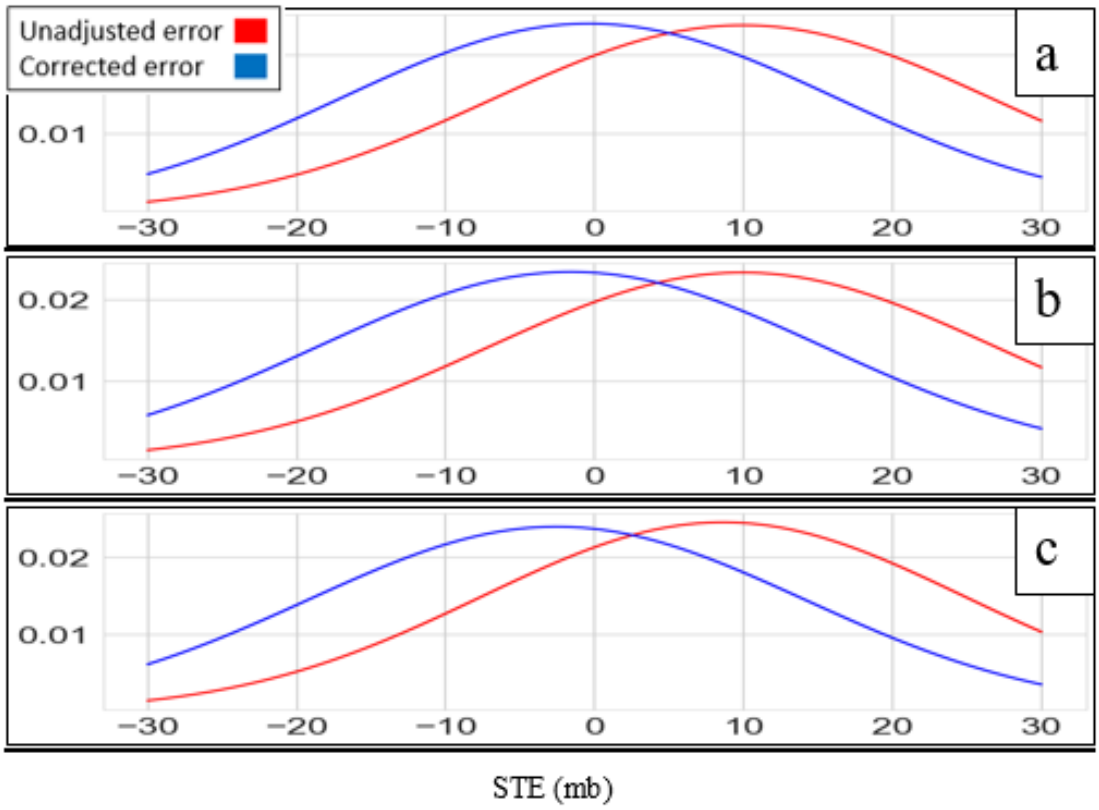


Figure 11. Intensity error in mb for (a) 72-hour forecasts, (b) 96-hour forecasts and (c) 120-hour forecasts

2. Source Region

The data was also grouped by source region to determine if there was any dependence on location of the TC formation. TCs were divided into the four source regions shown in Figure 7. Figures 12 and 13 show TTE and STE, respectively, for all data together and by each region. Results are summarized in tables 4 and 5. There was significant improvement when you utilize all storms, but when you break the storms into four separate regions there was extremely little improvement and therefore almost no change in both TTE and STE.

There are a few possibilities why there was no improvement with this manner of grouping. One possibility is that as shown in section 1, there is a dependence on forecast hour and since the data was not broken down into forecast hour in this case, we would see no improvement. A second possibility is that there is simply not enough data to create

usable statistics for training the model. Once the data was broken down into 4 regions the learning periods for the model became much smaller. For the GO and CA regions specifically, due to their proximity to the continental U.S. landmass, there was even less data for extended forecast periods.

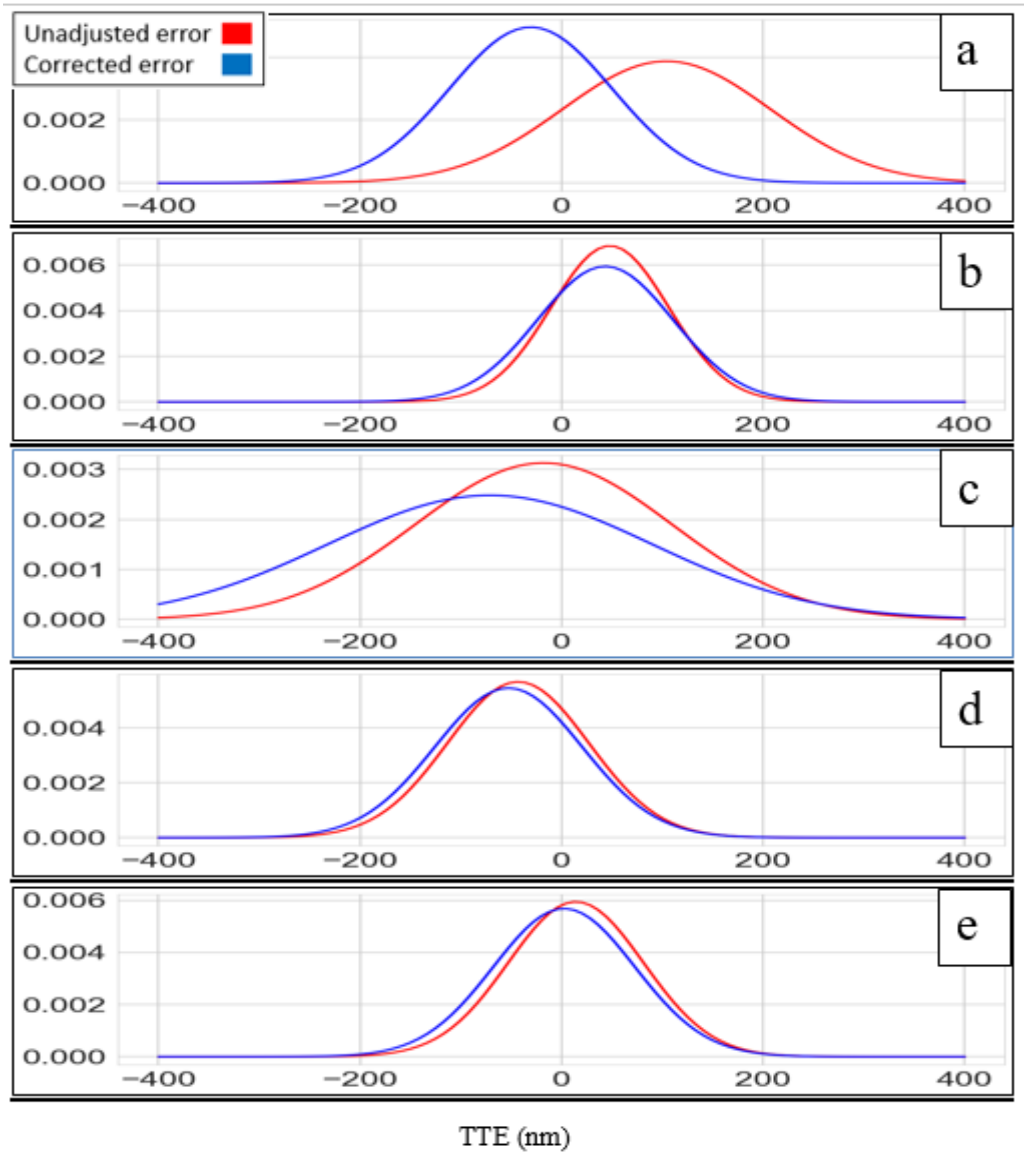


Figure 12. Total track error in nm for all forecast hours within (a) The entire Atlantic Basin, (b) the CA region, (c) the GO region, (d) the AT region and (e) the WA region

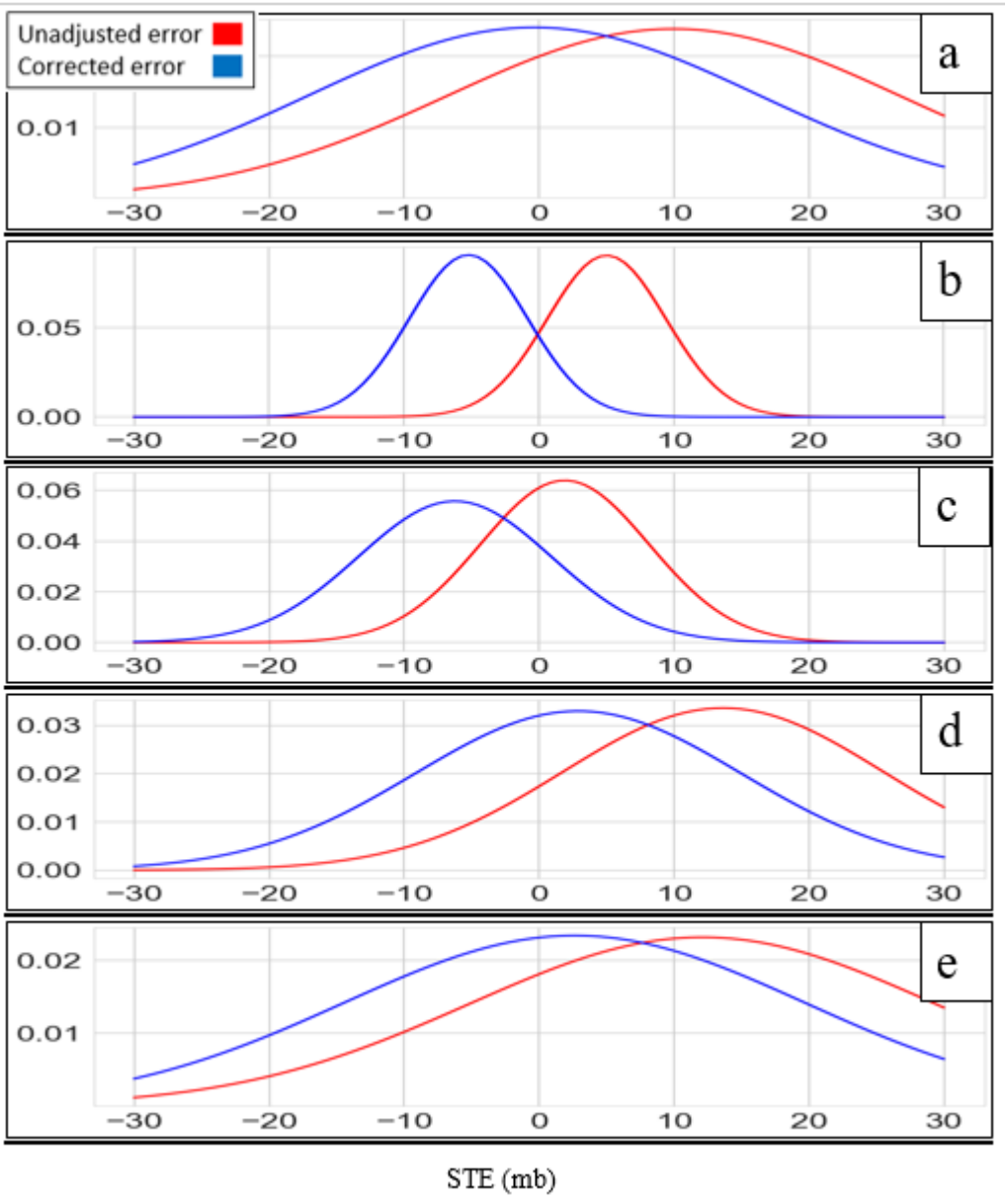


Figure 13. Intensity error in mb for all forecast hours within (a) The entire Atlantic Basin, (b) the CA region, (c) the GO region, (d) the AT region and (e) the WA region

3. Forecast Hour within Source Region

Data was also grouped by forecast hour within each source region. Figures 14–21 show TTE and STE for selected regions. Results are summarized again in Tables 2 and 3. For TTE, the bias was removed in 71% of the cases but there was only a decrease in forecast error spread 24% of the time. Take the AT region as an example, figures 14 and 15 show TTE for the AT region where although the bias was removed most of the time, there was very little improvement in forecast error spread.

For STE, there was a marked bias removal in 43% of the cases, but very little success in reducing the forecast error spread. Using the AT region as an example again, figures 16 and 17 show STE results with improvement in the bias but essentially the same exact spread as with the unadjusted data. The best results to illustrate this finding came from the WA region. There was a significantly larger amount of storms in the WA region for every forecast hour which could have contributed to the more successful results seen for that region. Figures 18–21 show the WA region results. These results further support the theory that a larger learning period will produce more successful results.

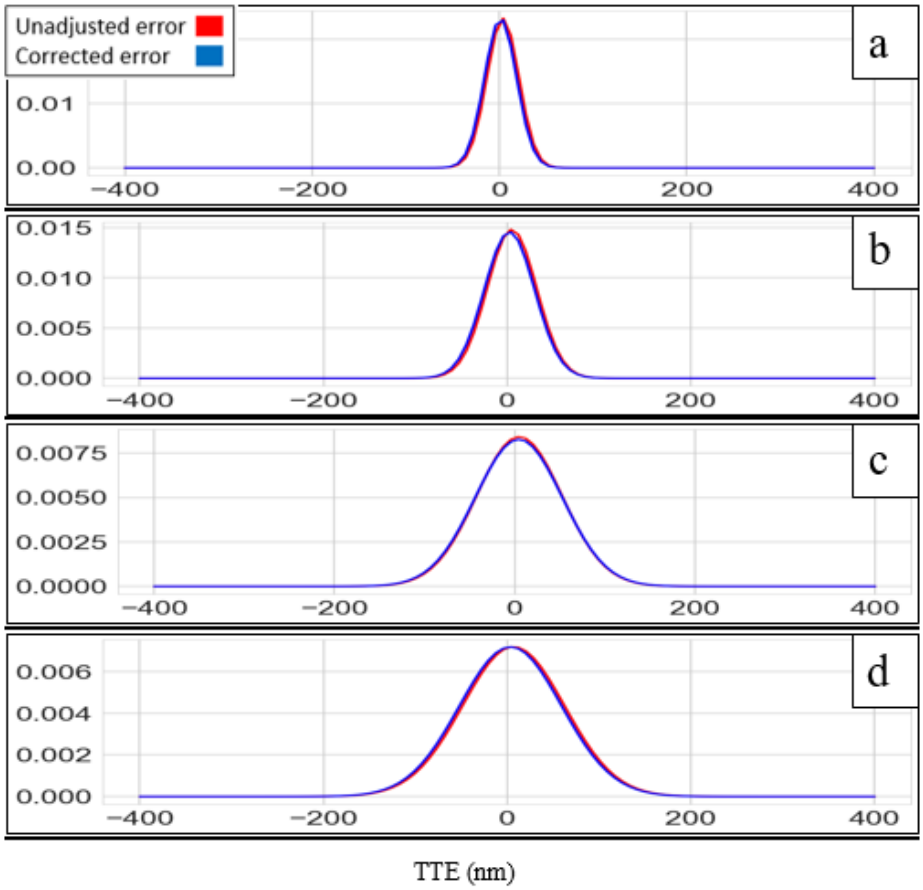


Figure 14. Total track error in nm for storms in the AT region for (a) 12 hour forecasts, (b) 24 hour forecasts, (c)36 hour forecasts, and (d) 48 hour forecasts

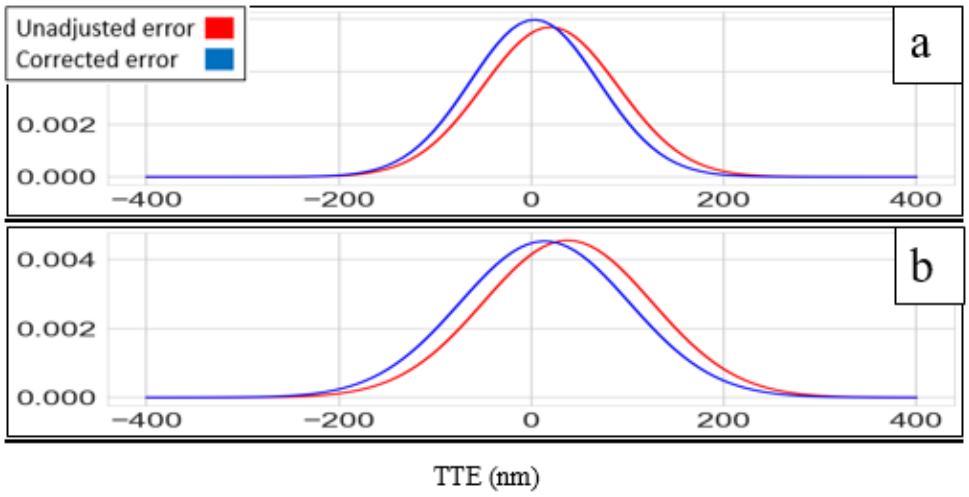


Figure 15. Total track error in nm for storms in the AT region for (a) 72-hour forecasts and (b) 96-hour forecasts

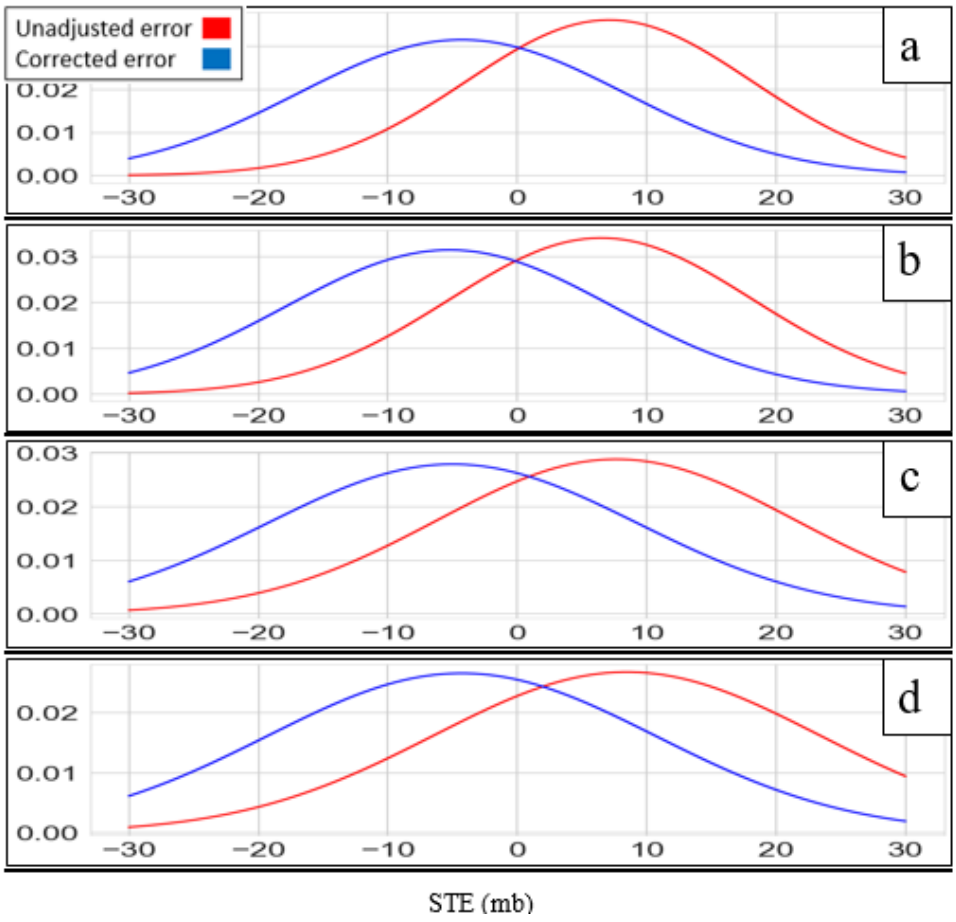


Figure 16. Graph of STE in mb for storms in the AT region for (a) 12-hour forecasts, (b) 24-hour forecasts, (c) 36-hour forecasts, and (d) 48-hour forecasts

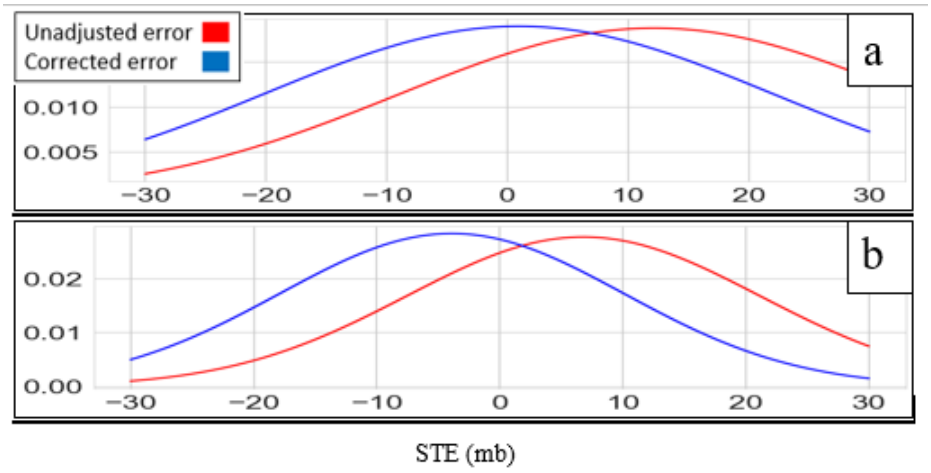


Figure 17. Graph of STE in mb for storms in the AT region for (a) 72-hour forecasts, and (b) 96-hour forecasts

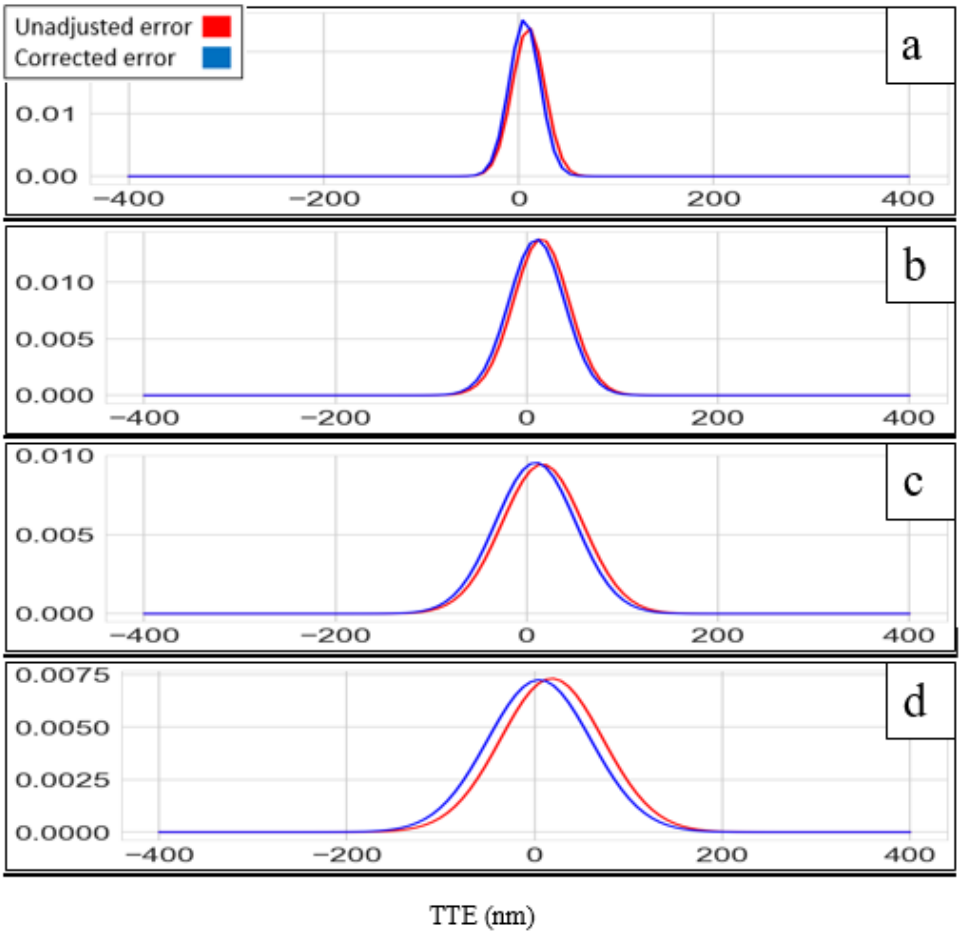


Figure 18. Graph of TTE in nm for storms in the WA region for (a) 12-hour forecasts, (b) 24-hour forecasts, (c) 36-hour forecasts and (d) 48-hour forecasts

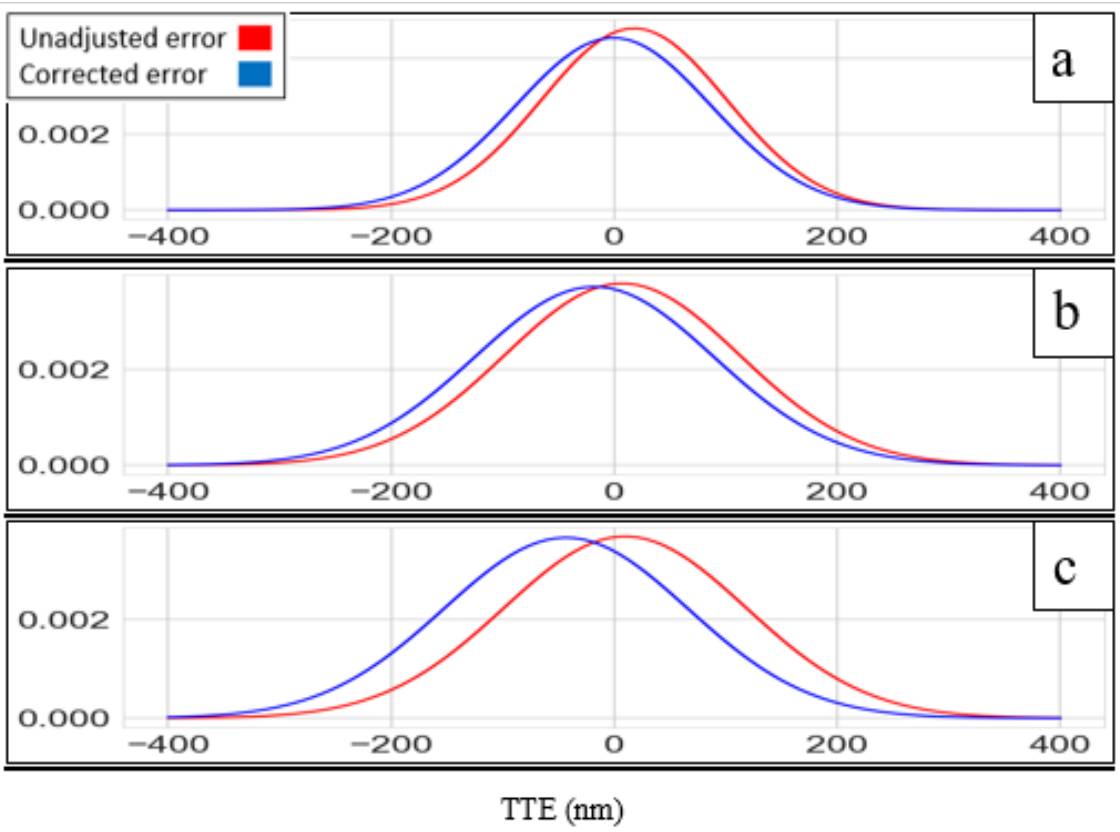


Figure 19. Graph of TTE in nm for storms in the WA region for (a) 72-hour forecasts, (b) 96-hour forecasts and (c) 120-hour forecasts

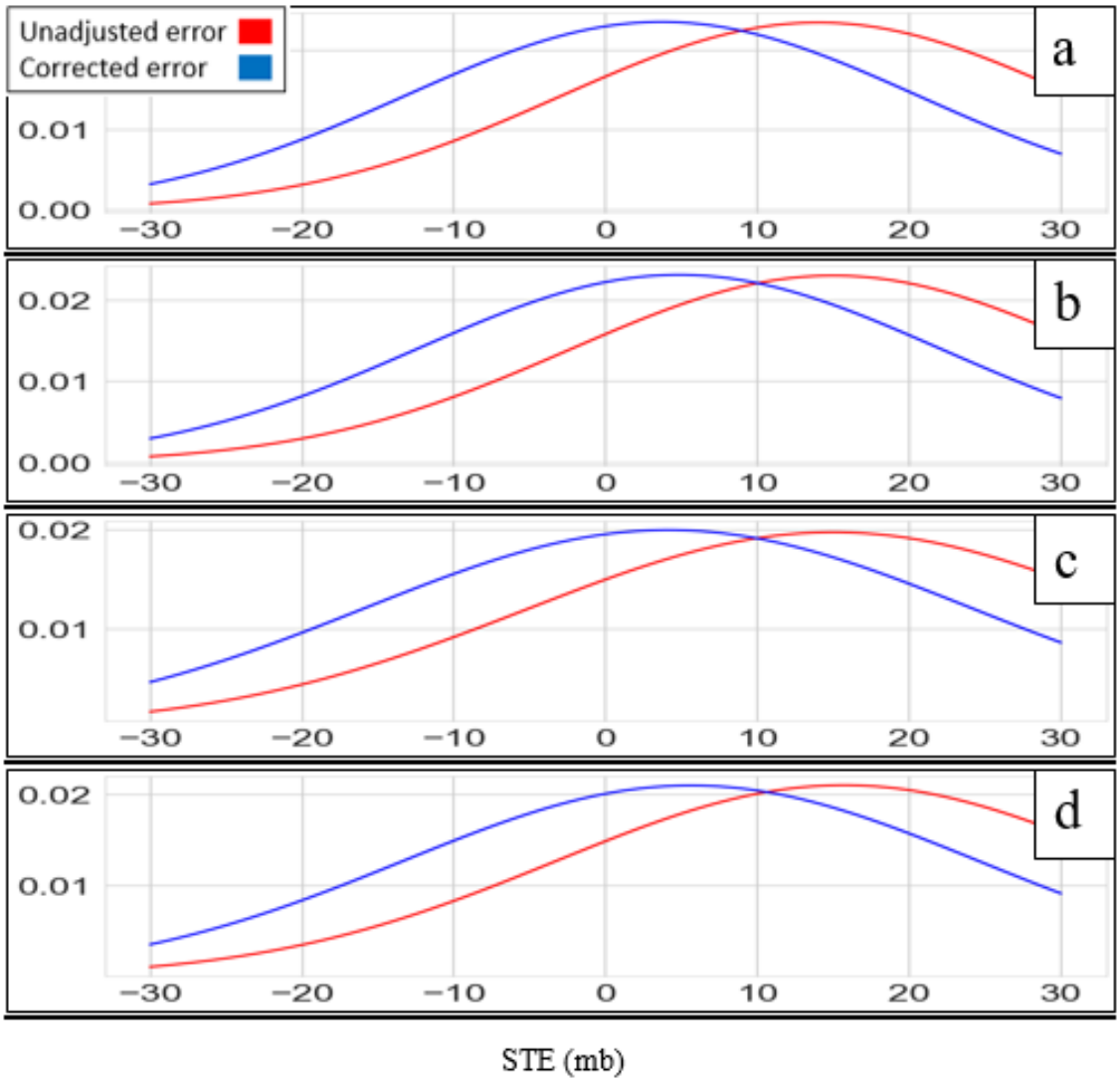


Figure 20. Figure 20. Graph of STE in mb for storms in the WA region for (a) 12-hour forecasts, (b) 24-hour forecasts, (c) 36-hour forecasts, and (d) 48-hour forecasts

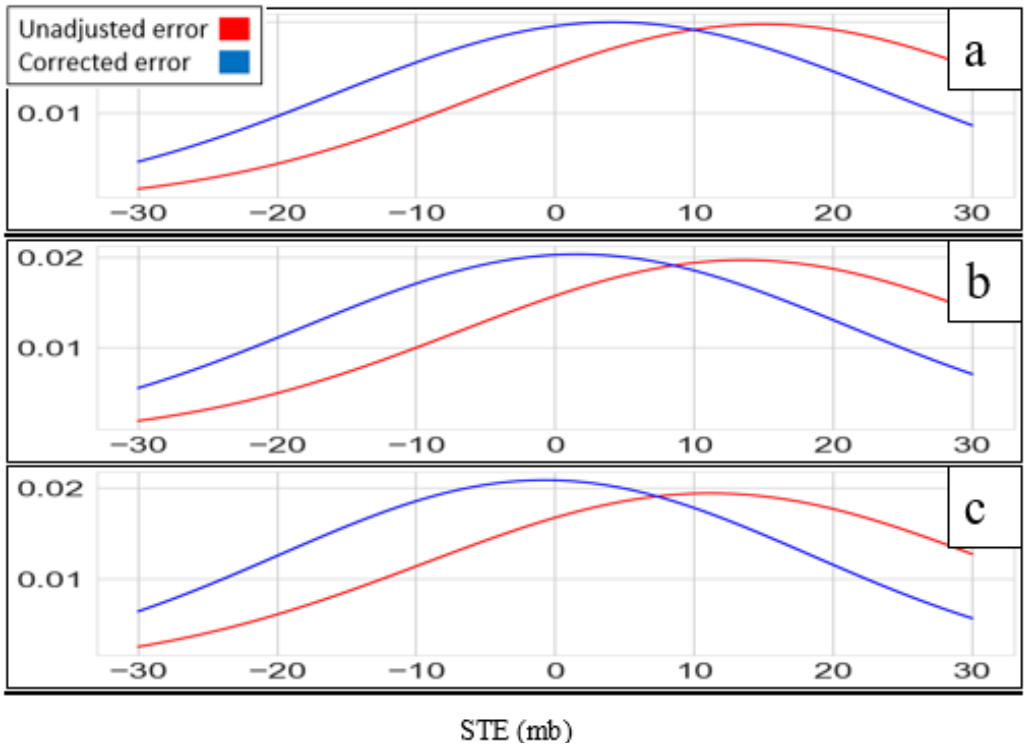


Figure 21. Graph of STE in mb for storms in the WA region for (a) 72-hour forecasts, (b) 96-hour forecasts, and (c) 120-hour forecasts

Table 2. Results of TTE data for all forecast hours and source regions with green indicating improvement over raw model forecasts and red indicating no improvement

FORECAST HOUR	REMOVE BIAS	REDUCE SPREAD	STUDENT'S T-TEST	
			t-value	p-value
12	SOME	NO	15.9	<0.01
24	YES	NO	17.9	<0.01
36	YES	YES	18.3	<0.01
48	SOME	YES	20.1	<0.01
72	SOME	YES	22.1	<0.01
96	SOME	NO	21.1	<0.01
120	SOME	YES	16.2	<0.01
12 CA	YES	NO	.6	.6
24 CA	SOME	YES	1.2	.2
36 CA	SOME	YES	.9	.4
48 CA	SOME	YES	.2	.9
72 CA				
96 CA				
120 CA				
12 GO	YES	NO	.6	.6
24 GO	NO	NO	.3	.8
36 GO	SOME	NO	1.2	.3
48 GO	NO	NO	.5	.6
72 GO				
96 GO				
120 GO				
12 AT	SOME	NO	.5	.6
24 AT	SOME	NO	.5	.7
36 AT	NO	NO	.6	.6
48 AT	NO	NO	.2	.8
72 AT	YES	YES	1	.4
96 AT	YES	NO	1	.3
120 AT				
12 WA	YES	YES	1.3	.2
24 WA	SOME	NO	1	.4
36 WA	YES	NO	1	.4
48 WA	SOME	NO	1.3	.2
72 WA	YES	NO	1.2	.2
96 WA	NO	NO	1.3	.2
120 WA	NO	NO	2.7	<0.01

Table 3. Results of STE data for all forecast hours and source regions with green indicating improvement over raw model forecasts and red indicating no improvement

FORECAST HOUR	REMOVE BIAS	REDUCE SPREAD	STUDENT'S T-TEST	
			t-value	p-value
12	YES	NO	7	<0.01
24	YES	NO	7	<0.01
36	YES	NO	7	<0.01
48	YES	NO	6	<0.01
72	YES	NO	6	<0.01
96	YES	NO	7	<0.01
120	YES	NO	7	<0.01
12 CA	NO	NO	7	<0.01
24 CA	NO	NO	1	.2
36 CA	NO	NO	5	<0.01
48 CA	NO	YES	3	<0.01
72 CA				
96 CA				
120 CA				
12 GO	NO	NO	.6	.5
24 GO	NO	NO	2	.1
36 GO	NO	NO	2	.2
48 GO	YES	NO	.9	.4
72 GO				
96 GO				
120 GO				
12 AT	NO	NO	4	<0.01
24 AT	NO	NO	3	<0.01
36 AT	NO	NO	6	<0.01
48 AT	NO	NO	3	<0.01
72 AT	YES	NO	2	<0.01
96 AT	NO	NO	1	.3
120 AT				
12 WA	SOME	NO	3	<0.01
24 WA	SOME	NO	1	.4
36 WA	YES	NO	3	<0.01
48 WA	YES	NO	1	.2
72 WA	SOME	NO	1	.2
96 WA	YES	NO	3	<0.01
120 WA	YES	YES	3	<0.01

Table 4. Results of TTE data for all forecast hours for each source region with green indicating improvement over raw model forecasts and red indicating no improvement

FORECAST	REMOVE BIAS	REDUCE SPREAD	STUDENT'S T-TEST	
			t-value	p-value
ALL HOURS	SOME	YES	16	<0.01
ALL HOURS CA	NO	NO	.3	.8
ALL HOURS GO	NO	NO	5	<0.01
ALL HOURS AT	NO	NO	.6	.6
ALL HOURS WA	YES	NO	4	<0.01

Table 5. Results of STE data for all forecast hours for each source region with green indicating improvement over raw model forecasts and red indicating no improvement

FORECAST	REMOVE BIAS	REDUCE SPREAD	STUDENT'S T-TEST	
			t-value	p-value
ALL HOURS	YES	NO	7	<0.01
ALL HOURS CA	NO	NO	9	<0.01
ALL HOURS GO	NO	NO	<0.01	.9
ALL HOURS AT	SOME	NO	4	<0.01
ALL HOURS WA	SOME	NO	12	<0.01

4. Student's t-Test

As discussed in Chapter III, the Student's t-test compares two means and determines if they are different from one another and how different the two are. Results of the Student's t-test are shown in tables 2–5. The t-value is the ratio between the difference of the two groups and the difference within the two groups. The p-value tells you the probability that the results from your sample could occur by chance. For this test, it was important to see a larger t-value to determine that the two groups were statistically different and low p-values to ensure the results did not happen by chance.

For the TTE results in Table 2 you can see that when you take the entire data set and break it apart into forecast hours, there are large t-values and small p-values. Therefore, it is certain that the samples are statistically different and did not occur by chance. Looking at how the t and p values change over various forecast hours, there was little variation for the most part therefore the BEMOS model is just as good at correcting the 120-hour forecasts as it is at correcting the 12-hour forecasts.

For the tests where the data set was broken down into 4 regions, the t-values and p-values indicate that the samples are not statistically different. The t-values were relatively small and the p-values were relatively large. The results were similar for STE. Again, this could be due to the small size of the training data in these cases. In this research we were using ensemble spread to predict error and then adjust track and intensity for those errors. The lack of reduction of spread indicated that the inherent forecast uncertainty has not been reduced.

THIS PAGE INTENTIONALLY LEFT BLANK

V. SUMMARY

A. CONCLUSIONS

Based on the results of this thesis, it is certain that applying statistical post-processing to ensemble mean and spread data using Bayesian estimation and MCMC methods can decrease there error inherent in the ensemble models. The error inherent in all ensemble models can be estimated and characterized in order to train a model to do better. Although it is certain that these statistical methods can decrease error it is also clear that the error cannot be completely eliminated, only reduced. Although this research indicates that there is an advantage to statistical post-processing of raw model data, it is quite apparent that there are still several issues to be addressed with the method before it can be utilized operationally.

One major issue that was realized during this research is the dependence on the NPS BEMOS model's test dataset or "learning period." In order to produce statistically relevant and usable output the model needs a rather lengthy learning dataset. In this case a learning dataset of 6 years was used but in most cases it was not long enough to reduce forecast error spread, even though the bias was typically reduced.

Also to note, it is clear that there is a significant dependence on forecast length. The most successful output resulted from breaking the data into forecast hours as long as there was still a sufficiently large enough dataset to produce statistically relevant results.

B. RECOMMENDATIONS

There are several issues that need to be addressed in order to make this method of statistical post-processing operationally useful. This first issue is with the learning dataset. In order to get more successful results a longer learning period is needed. Since TCs are somewhat rare phenomena it would be best to utilize the longest learning set possible. A learning set that goes back as far as possible taking into account the last model update would be the best case. The issue that arises with this is that all deterministic models undergo updates. If a model is updated then the model tendencies will change and therefore you cannot utilize data from those datasets. For example, if the

GFS underwent a major upgrade in 2010 then you may only compare 2010 years and later for use in this type of statistical post-processing method. Since operational deterministic weather models are constantly being upgraded this becomes a challenge. One possible solution could be to continue running older models to increase the length of the dataset from which the NPS BEMOS model can draw its learning data from. However, as costly as it is to run global deterministic models, this solution is probably not cost efficient.

Another possible solution to increase the size of the dataset without increasing the number of years the data is drawn from could be to include all storms in the dataset vice just using those that were hurricane strength or greater as this research did. Including all tropical systems in the model's learning dataset would significantly increase the amount of data the model has to learn from.

Another option could be to utilize alternate grouping methods for storms such as seasons. Grouping storms in the first half of the Atlantic Basin tropical season (June-August) and storms in the second half of the season (September-November) could be useful. Finding different ways to slice the data so that you get unique statistics could be advantageous as long as the data grouped is still large enough to produce statistically relevant results. By increasing the size of the dataset and allowing the model to gain a better understanding of the inherent error, it will be able to determine precisely how to correct the raw model forecasts. When you look at ensemble model forecasts there is less error at earlier forecast hours than at extended hours. This is because even a small amount of error in initial conditions or model mathematics carries forward and increases until the error is so great that the long-term forecasts are not useful. This is why it is helpful to group forecasts by forecast hour in this research. Characterization of the 12-hr forecast error will always be different than the 120-hr error. Also, both increasing the dataset size and/or length while also grouping the data in other relevant ways could produce better results.

Another thing that needs to be done that this thesis did not address is breaking the corrections down into along and cross track components. In this research we did not extract along/cross track spread to use it as a predictor in the model. Future research

should examine this to potentially identify tendencies for left/right, fast/slow track errors. Overall it is clear that statistical post-processing of raw deterministic model data using Bayesian estimation and MCMC methods is the way of the future and most certainly the most effective method of reducing model error but more work to refine the method is needed before it could be operationally utilized.

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

Chisler, N., 2016: Relating tropical cyclone track forecast error distributions with measurements of forecast uncertainty. M.S. thesis, Dept. of Meteorology, Naval Postgraduate School, 57 pp, https://calhoun.nps.edu/handle/10945/48503

Clear, J., 2017: All models are wrong, some are useful. Accessed 03 February 2018, https://www.huffingtonpost.com/james-clear/all-models-are-wrong-some_b_11196880.html

ECMWF, 2006: TIGGE—Global ensemble forecast data. Accessed 20 March 2018. https://www.ecmwf.int/en/research/projects/tigge

ECMWF, 2017: TIGGE—Advancing global NWP through international collaboration. Assessed on 05January 2018. https://www.ecmwf.int/

GFS, 2013: National Weather Service Weather Prediction Center. Assessed on 25 March 2018. http://www.wpc.ncep.noaa.gov/mdlbias/biastext.html

GFS, 2018: MAV MOS bulletins for CA. Assessed on 9 April 2018. www.nws.noaa.gov/mdl/forecast/text/state/CA.AVN.htm

Gneiting, T., A. E. Raferty, A. Westveld, and T. Goldman, 2004: Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Monthly Weather Review*, **133,** 1098–1118.

Hauke, M., 2006: Evaluating Atlantic tropical cyclone track error distributions based on forecast confidence. M.S. thesis, Dept. of Meteorology, Naval Postgraduate School, 84 pp, https://calhoun.nps.edu/handle/10945/2840

Hendricks, E., 2011: Performance of a dynamic initialization scheme in the Couple Ocean-Atmosphere Mesoscale Prediction System for Tropical Cyclones (COAMPS-TC), *American Meteorological Society*, **26**(4), 650–663. doi:10.1175/WAF-D-10-05051.1

Lorenz, E. N., 1963: Deterministic nonperiodic flow, *World Meteorological Organization*, **218**(115), 161 pp.

National Hurricane Center, 2017: National Hurricane Center forecast verification. Assessed 01 August 2017, http://www.nhc.noaa.gov/verification/verify5.shtml.

Neese, J., 2010: Evaluating Atlantic tropical cyclone track error distributions for use in probabilistic forecasts of wind distribution. M.S. thesis, Dept. of Meteorology, Naval Postgraduate School, 87 pp, https://calhoun.nps.edu/handle/10945/5150

Richter, D., 2012: Bayesian ensemble model output statistics for temperature. Diploma thesis, Heidelberg University, Germany.

Sampson, C.R., 1990: The Automated Tropical Cyclone Forecasting System (ATCF) *American Meteorological Society*, **5**(12), 653–660). doi:10.1175/1520-0434(1990)005<0653:TATCFS>2.0.CO;2

Wendt, T., 2017: A Hierarchical Multivariate Bayesian approach to Ensemble Model Output Statistics in atmospheric prediction. PHD, Dept. of Meteorology, Naval Postgraduate School, 199pp, https://calhoun.nps.edu/handle/10945/56188

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California