3-14-2014

Validation of the Air Force Weather Agency Ensemble Prediction Systems

William B. Clements

VALIDATION OF THE AIR FORCE WEATHER AGENCY ENSEMBLE PREDICTION SYSTEMS

THESIS

William B. Clements, Captain, USAF

AFIT-ENP-14-M-04

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENP-14-M-04

VALIDATION OF THE AIR FORCE WEATHER AGENCY ENSEMBLE PREDICTION SYSTEMS

THESIS

Presented to the Faculty

Department of Engineering Physics

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Master of Science in Applied Physics

William B. Clements, BS

Captain, USAF

March 2014

AFIT-ENP-14-M-04

VALIDATION OF THE AIR FORCE WEATHER AGENCY ENSEMBLE PREDICTION SYSTEMS

William B. Clements, BS
Captain, USAF

Approved:

//signed//	10 Mar 14
Kevin S. Bartlett. Lt Col, USAF (Chairman)	Date
//signed//	10 Mar 14
Robert S. Wacker, Lt Col, USAF (Member)	Date
//signed//	10 Mar 14
Dr. Steven T. Fiorino (Member)	Date

Abstract

Air Force Weather Agency's (AFWA) Ensemble Prediction Systems (EPS), Global Ensemble Prediction System (GEPS), 20km Mesoscale Ensemble Prediction System (MEPS20) and 4km Mesoscale Prediction System (MEPS4), were evaluated from April to October 2013 for 10 locations around the world to determine how accurately forecast probabilities for wind and precipitation thresholds and lightning occurrence match observed frequencies using Aerodrome Routine Meteorological Reports (METARs) and Aerodrome Special Meteorological Reports (SPECIs). Reliability diagrams were created for each forecast hour detailing the Brier skill score (BSS) to depict EPS performance compared to climatology for each site and score composition through reliability, resolution and uncertainty. To illustrate how the BSS changed, the score and its composition were plotted for all forecast hours. This study showed that all three EPS suffered from a lightning overforecasting bias at all locations and most forecast hours. For wind speeds, it was clear that decreased model grid spacing allowed better resolution of terrain features, producing a better BSS. Likewise, precipitation was better resolved with increased horizontal resolution as explicit resolution of precipitation processes outperformed cumulus parameterization schemes.

I dedicate this work to Ovelle Keith Clements who passed during my time at AFIT. While I could not be with her during her last days on earth, her words of wisdom and life will forever shape my remaining days.

Acknowledgments

I wish to express much appreciation for the gracious and timely assistance provided to me by my committee chair Lt Col Kevin S. Bartlett and committee member Lt Col Robert S. Wacker. Daily questions directed to them were always met with an answer or a viable means of realization. Their counsel was instrumental in the focus and development of my research. I would also like to thank external committee member Dr. Steven T. Fiorino for engaging discussions of how to optimally present and defend my thesis.

This research would have not been possible without AFWA ensemble data and expert ensemble knowledge provided by Mr. Evan L. Kuchera. Also, I would like to express my gratitude to Mr. Jeff H. Zaunter for painstakingly working with me to provided station observations and climatology in a format compatible with my validation processes. His help singlehandedly saved me countless hours.

I am indebted to my fellow AFIT classmates, Capt Jeremy J. Hromsco, Capt Haley A. Homan, Capt Kyle R. Thurmond and 2Lt Coy C. Fischer for their support and assistance during my AFIT experience.

And, I would like to thank my parents and my fiancée who provided daily support while at AFIT. My mother played an integral role in the editing process for which I am most grateful.

Lastly and most importantly, I would like to think the Lord for the continual grace and blessings He has bestowed on my life. Without Him I am nothing.

William B. Clements

Table of Contents

Page

List of Figures

List of Tables

VALIDATION OF THE AIR FORCE WEATHER AGENCY ENSEMBLE PREDICTION SYSTEMS

I. Introduction

1.1 Motivation

During November 2012, the Director of Air Force Weather was briefed on the current status of the Air Force Weather Agency's (AFWA) ensemble weather forecasting operations as well as a way forward to increasingly incorporate these stochastic outputs into daily Air Force and other Department of Defense (DoD) operations. The plan included AFWA providing timely and operationally significant ensemble modeling data to users. The plan also included a means for Air Force weather personnel to interpret the model output by incorporating quantifiable performance metrics of their ensemble prediction systems (EPS).

While ensemble weather forecasting has expanded over the past two decades, there still remains a disconnect between the research community and weather forecasters within the DoD. This disconnect arises from a lack of understanding among the research community of what information needs to be communicated to weather forecasters where as forecasters need to understand how EPS work and how they can be superior to deterministic models. Results from ensemble weather input into operational risk management (ORM) destruction of enemy air defense simulations clearly showed the applicability of ensembles over deterministic inputs for future DoD missions (Eckel et al, 2008). The motive of this research is to help bridge the gap between the researcher and the weather forecaster by evaluating and quantifying the value of AFWA's three EPS.

1.2 Technological and Numerical Weather Prediction Theory Advancement

As technology and numerical weather prediction (NWP) theory have continued to progress, so has the importance of NWP in being able to provide operational forces with accurate and actionable weather intelligence. With continued improvements in computer processing speed, physical parameterization schemes, estimates of the initial state of the atmosphere and data assimilation techniques, ensemble forecasting has come to the forefront of NWP. AFWA runs EPS that contain multiple deterministic models with perturbed initial conditions and different parameterization schemes (AFW-WEBS, 2013). These prediction systems produce operationally useful modeled weather forecasts in a timely manner that, unlike a single deterministic model, provide a sense of forecast uncertainty by indicating the range of solutions forecasted by the ensemble members.

1.3 Human Element

Operational risk management is defined as balancing a mission's objective against its risk. Weather is a significant factor in a mission's risk assessment. Effective communication of this risk to operators remains the responsibility of Air Force weather forecasters who can use ensemble output to offer a better assessment of mission critical weather limiting factors to the warfighter. Knowledge of the forecast probability provides the operator with additional information to develop the correct operational risk management for successful mission execution.

1.4 Research Topic and Objective

With the quantification of uncertainty enabled by ensemble NWP, the Air Force and other DoD organizations are currently transitioning from deterministic to stochastic weather information for mission planning. This allows for a more comprehensive understanding of how weather uncertainty might potentially affect the mission. Verification of ensemble predictions is not as straightforward as verifying deterministic predictions. Each probability needs to closely match the observed frequency of the parameter forecasted for the EPS to be deemed valuable, while few such studies have been undertaken to validate many EPS (Ehrendorfer, 1997). Therefore the main purpose of this study is to verify how well AFWA's EPS - one-degree Global Ensemble Prediction System (GEPS) and 20km and 4km Mesoscale Ensemble Prediction Systems (MEPS20 and MEPS4) - perform by relating ensemble member agreement to probability of occurrence using station observations as well as defining EPS skill over climatology. This initial performance information will allow AFWA to fine-tune their EPS and provide useful metrics to forecasters in the field.

1.5 Preview

In this chapter the remit and general application of ensemble weather forecasting in the Air Force is introduced. Chapter 2 covers a more extensive overview of ensemble weather prediction in general and at AFWA. The methodology for this research will be covered in Chapter 3. The subsequent chapter covers all research findings followed by a conclusion of the findings, recommendations and future research possibilities.

II. Background

2.1 Numerical Weather Prediction

2.1.1 Chaotic Atmosphere and Model Error

The earth's atmosphere is a dynamic system of interconnected processes that must be modeled correctly to determine its future state. These processes range from solar radiation entering the top of the atmosphere to sensible heat fluxes at the earth's surface. Lorenz (1963) discovered that small variances in the initial state of the atmosphere lead to dramatically differing results as a numerical forecast progresses in time. He suggested that error in correctly resolving the initial state of the atmosphere is the major factor in numerical forecast error and the ultimate limiting factor in atmospheric predictability (Lorenz, 1963). Theoretically, given a perfect set of initial conditions, the atmosphere can be precisely modeled. However, the initial conditions used for the data assimilation process in NWP have some degree of uncertainty due to instrument error and data sparsity, thus numerical forecasts always maintain some uncertainty that grows over time (Kalnay, 2003).

2.1.2 Deterministic vs. Stochastic Prediction

Since the first successful one-day numerical weather forecast in 1947 by Charney, Fjortoft and von Neumann, NWP for the majority of the past half-century has been deterministic forecasting (Charney et al, 1950). Today a deterministic forecast is comprised of one model solution based on a single set of initial conditions and a set of fixed parameterization schemes for processes that cannot be resolved by the model due to their horizontal and vertical grid scales. Even with computational advancements, improved resolution down to 1.67km, fewer required parameterizations, increased availability of data, and improved data assimilation, deterministic models can still deviate significantly from observations in the early forecast hours (Kalnay, 2003). Stochastic forecasts provide a way to account for these errors. The lineage of stochastic forecasting methods can be traced back to Epstein's (1969) concept of stochastic-dynamic forecasting, Leith's (1974) Monte Carlo method and Hoffman and Kalnay's (1983) lagged average method. Today operational ensembles use breeding during data assimilation to create perturbations in initial conditions (Wei et al, 2008). Breeding, the basis for all perturbations to initial conditions in operational use today, consists of: (1) adding a small arbitrary perturbation to the initial state of the atmospheric analysis at a given time t_0; (2) integrating the model from both the perturbed and unperturbed initial conditions for a short period $t_1 - t_0$; (3) subtracting one forecast from the other; (4) scaling down the difference field so that it has the same norm (e.g., root mean square amplitude) as the initial perturbation; (5) adding this perturbation to the analysis corresponding to the following time t_1, and then repeating steps 2 through 5 forward in time to simulate error growth during the analysis period (Toth and Kalnay, 1993; Toth and Kalnay, 1997). From this framework the ensemble transform bred vector, ensemble transform technique and ensemble transform Kalman technique were developed (Wei et al, 2008). Incorporating these perturbations methods into ensemble members provides an opportunity for each member in an EPS to represent the initial state as well as the future state of the atmosphere.

2.1.3 Characterizing Ensemble Uncertainty

By employing a set of perturbed initial conditions to account for observational uncertainty and various parameterization schemes to represent convection, turbulence, surface features and other phenomena, an EPS helps quantify uncertainty in the forecast. This quantification is based on the spread of model solutions which can be portrayed using a probability density function (PDF). The PDF indicates whether the ensemble members' solutions are closely grouped or widely spread, and whether the solutions cluster into distinct groups of closely-related solutions (Eckel and Mass, 2005). A PDF is shown in Figure 1 depicting how modeled solutions can vary over time.

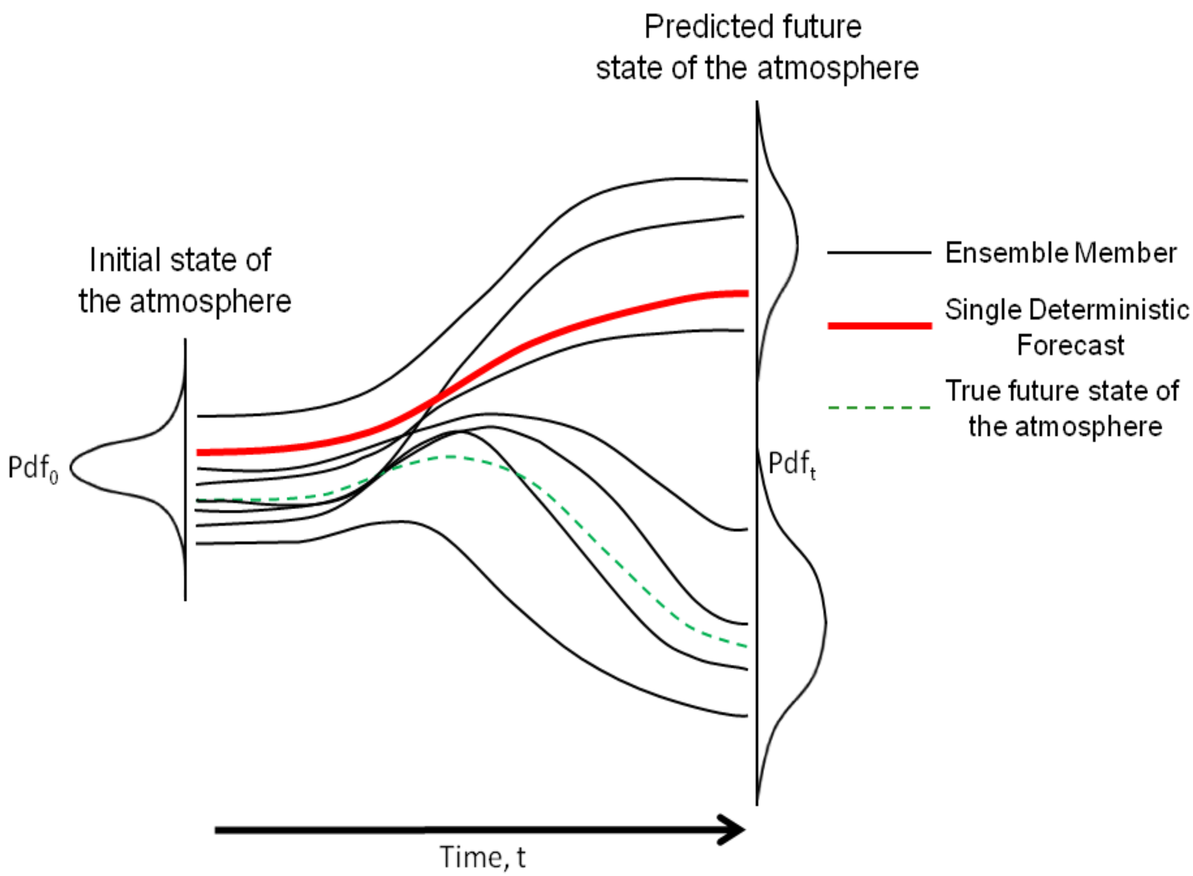

Figure 1: Probability density functions associated with ensemble prediction. The initial probability density function, pdf_0, characterizes the uncertainty in the atmosphere's initial state. The collection of the bold line and non-bold lines represents individual deterministic forecasts produced from different initial conditions while the dashed line is the actual state of the atmosphere. The resulting forecast probability distribution at time t, pdf_t, characterizes the uncertainty in the forecasts, and in this case, is bimodal.

A single point forecast represented by the bold red line fails to resolve the future state of the atmosphere depicted by the dashed green line. Each ensemble forecast, represented by solid non-bolded lines, are used to sample the uncertainty in the initial state of the atmosphere with two results close to the actual future state. Note that this example shows a bimodal result meaning two subsets of modeled solutions deviated. In a perfect ensemble the perturbed initial conditions would incorporate all sources of uncertainty; however, in reality an ensemble member can only include uncertainty to a limited degree based on the uncertainty in the initial PDF.

2.2 Previous Research

EPS are continually updated in efforts to optimally resolve the atmosphere. These updates include varying numbers of members, boundary conditions, vertical levels, horizontal resolution, perturbation methods, and physics schemes, leading to a constant need for testing and evaluating EPS performance. Wei et al (2008) tested four main perturbation methods in National Centers for Environmental Prediction’s (NCEP) Global Forecast System (GFS): breeding, ensemble transform, ensemble transform with rescaling, and the ensemble transform Kalman filter. They used the Brier score (BS), Brier skill score (BSS), and ranked probability skill score to show that the rescaled ensemble transform outperformed the other methods and that increasing the number of ensemble members generally increased these skill scores. During a three month study, Buizza et al (2004) used outlier statistics, BSS, root mean square error, and pattern anomaly correlation to compare three widely used operational global spectral ensemble models: the European Centre for Medium-Range Weather Forecasts (ECMWF), the Canadian Meteorological Centre’s (CMC) Global Ensemble Model (GEM), and the

NCEP's Global Ensemble Forecast System (GEFS). The majority of the verification metrics employed showed that the ECMWF performed best overall, with GEFS being competitive during the first few days and GEM being competitive in the last few days of the forecast. Hamill et al (2007) discussed the utility of reliability diagrams and BSS in the calibration of EPS. A recent study by Wang et al (2012) evaluated and compared Aire Limitée Adaptation Dynamique Développement International-Limited Area Ensemble Forecasting (ALADIN-LAEF) EPS to ECMWF EPS to investigate whether any value is added by their regional EPS. In this study ALADIN-LAEF EPS was comprised of 16 perturbed members of the ECMWF with a horizontal resolution of 18km while the ECMWF EPS was compromised of 50 members at a 50km resolution. Results were compared over a two-month period in the summer of 2007 for Central Europe. ALADIN-LAEF EPS proved to be more skillful in forecasting surface weather phenomena including 10-meter winds, 12-hour accumulated precipitation and mean sea level pressure, despite fewer ensemble members, while the ECMWF EPS performed better for upper air weather variables. While none of these studies directly tested AFWA's EPS, they do provide some insight as to how some of the models used by AFWA's EPS perform and ways to provide useful performance metrics. Also, the ALADIN-LAEF and the ECMWF EPS comparison provides preliminary support for possible differences between AFWA global and regional EPS.

2.3 Air Force Weather Agency Ensembles

2.3.1 Probability Generation

The EPS used by AFWA do not have enough members to explicitly resolve a PDF therefore other methods must be employed to estimate forecast probability (AFW-

WEBS, 2013). These probabilities are generated by a technique called uniform ranks. This method first takes into account democratic voting - how many of the ensemble members that make up the EPS are forecasting the selected threshold. For example, if 7 out of 10 members forecast winds greater than 25kts, the probability of that event occurring is 70% as shown in the top portion of Figure 2.

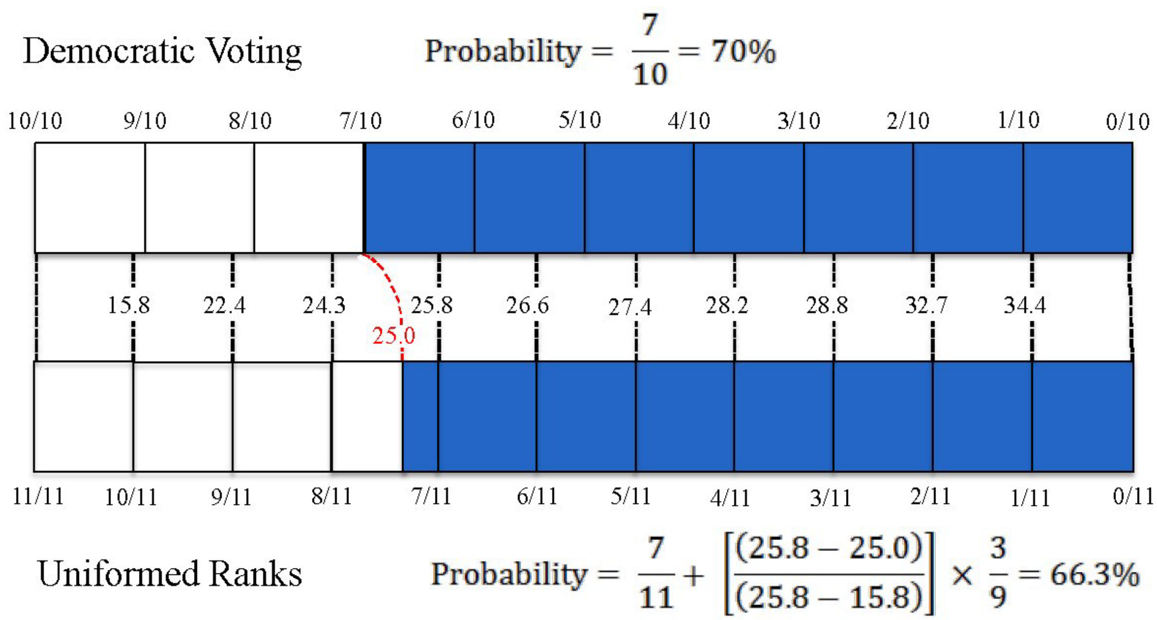

Figure 2: Graphic of probability generation methods used by AFWA. The top half depicts a basic democratic method of generating a probability based on how many ensemble members forecast the event. The bottom half, uniformed ranks, is a more robust method that also uses the values that did not exceed the threshold desired to generate a more realistic probability. (Adapted from AFW-WEBS, 2013.)

Democratic voting probability generation, however, disregards some portions of the PDF leading to more extreme forecast probabilities. A more robust approach lies in uniformed ranks which involves adding a probability rank, using democratic voting, and using linear interpolation to account for how close the forecasts that did not exceed the forecast threshold desired are. This is shown in the bottom portion of Figure 2. By doing so, portions of the PDF that the democratic voting method ignored are now accounted for, producing a more realistic probability of 66.3%. If all the ensemble members forecast a value below or above the forecasted threshold, this probability falls on the tail of the PDF and in an extreme rank as shown in Figure 3.

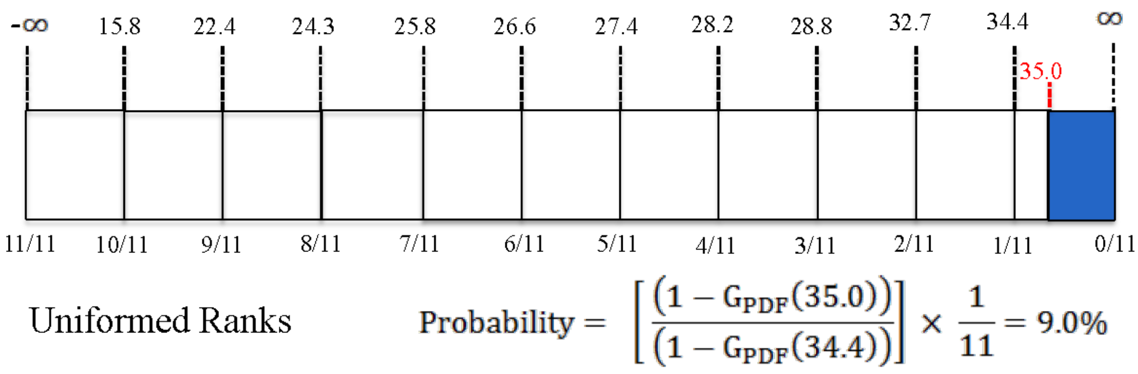

Figure 3: Graphic of probability generation when forecasted threshold falls in an extreme rank. A Gumbel distribution function is used to find the numerical distance between the highest forecasted value and the desired threshold. (Adapted from AFW-WEBS, 2013.)

When this takes place the approach used is similar; however, the probability is found by taking a portion of the probability in the last rank based on the numerical distance between the highest forecasted value and the desired threshold using a Gumbel distribution as shown in Equation 1 (Wilks, 2011).

$$G_{CDF}(x) = exp\left[-exp\left(\frac{\xi - x}{\beta}\right)\right] \tag{1}$$

Here ξ and β are Gumbel parameters defined in Equations 2 and 3 (Wilks, 2011), respectively, and x is the variable.

$$\beta = \frac{s\sqrt{6}}{\pi} \tag{2}$$

$$\xi = \bar{x} - \gamma\beta \tag{3}$$

Here s is the standard deviation, $\bar{x}$ is the sample mean, and γ is Euler's constant - 0.57721.

Many meteorological phenomena that Air Force weather forecasters try to forecast such as winds greater than a certain threshold or lightning occurrence within a specific radius are not directly resolved by AFWA's EPS. Thus, algorithms must be employed to generate probabilities from existing model output.

Specifically for wind threshold probabilities, a continuous distribution function must be generated to capture the surface wind gust for each ensemble member. To create this continuous distribution function, a generalized extreme value distribution is used as defined in Equation 4 (Wilks, 2011).

$$f(x) = \ exp\left\{-\left[1 + \frac{\kappa(x - \ \zeta)}{\beta}\right]^{\frac{-1}{\kappa}}\right\} \tag{4}$$

Each ensemble member's sustained wind speed is used as the shift parameter, ζ while the shape parameter, κ, is three over land and one over water (AFW-WEBS, 2013). The scale parameter, β, for over land is each ensemble members sustained wind speed in meters per second raised to the 0.75 power and a value of 1.25 for over water (AFW-WEBS, 2013).

For lightning, the algorithms used by AFWA to create a probability of at least one lightning strike within the forecast radius of the location is based on regression equations developed from Rapid Update Cycle (RUC) model analysis and observed lightning and precipitation over CONUS (AFW-WEBS, 2013). These algorithms use known model output to include convective potential available energy, lifted index, precipitable water, convective inhibition and accumulated precipitation. Each individual member's probability is calculated and then averaged with the other members to create the EPS probability forecast. For a mathematical description of the lightning algorithms employed by AFWA, reference Appendix A.

2.3.2 Global Ensemble Prediction System (GEPS)

The GEPS is produced at a one-degree resolution and produces output at a 6-hour forecast interval out to 240 hours. It is comprised of 21 members each from the NCEP GFS and the CMC's GEM, with 20 additional members from the Fleet Numerical Meteorology and Oceanography Center (FNMOC) Navy Operational Global Atmospheric Prediction System (NOGAPS), totaling 62 ensemble members (AFW-WEBS, 2013). Because GEPS is comprised of multiple EPS it is considered a super ensemble. Other than amalgamating the members to create the super ensemble, no further physics configuration changes are made outside of what is done at each respective modeling center. Two of the models used in GEPS, the GFS and NOGAPS are global spectral models, which represent atmospheric parameters as a series sum of spherical harmonic functions. As harmonics of higher wavenumbers are added to the series, the atmosphere can be modeled with higher resolution. The GEM is a global grid model that uses finite differencing to solve the atmosphere's governing equations.

The GFS used by AFWA utilizes stochastic physics parameterizations and is at a spectral resolution of 254 wavenumbers (T254) with 64 vertical levels out to 192 hours (AFW-WEBS, 2013). Beyond this point and out to 384 hours the resolution is reduced to 190 wavenumbers (T190) (AFW-WEBS, 2013). For initial conditions, the GFS uses an ensemble transform bred vector with rescaling. A detail description of model characteristics can be found at: http://www.emc.ncep.noaa.gov/GEFS/mconf.php.

The NOGAPS is characterized by T159 with 42 vertical levels (AFW-WEBS, 2013). It uses an ensemble transform scheme for its initial condition and no perturbed parameterizations for its physics schemes. For further model characteristics please reference: http://www.nrlmry.navy.mil/metoc/nogaps/nogaps_char.html.

The GEM is characterized by a horizontal resolution of 66km with 74 vertical levels and uses an ensemble transform Kalman filter for its initial conditions (AFW-WEBS, 2013). Houtekamer and Mitchell (2009) explain that Kalman filters are used to maintain a representative spread between the ensemble members, avoiding the problem of inbreeding by using one ensemble of short-range forecasts as background fields in data assimilation while employing the weights calculated from another ensemble of short-range forecasts. For further model characteristics please reference: http://weather.gc.ca/ensemble/verifs/model_e.html.

Having different wavenumbers for each EPS results in differing horizontal resolutions. To standardize the resolution, all members of the GEPS are re-gridded to a one-degree grid (Kuchera, 2013). Therefore all the resulting probabilities have a one-degree horizontal resolution regardless of the wavenumbers for each EPS.

2.3.3 Mesoscale Ensemble Prediction System (MEPS)

The MEPS is a finer resolution model than GEPS, created to better resolve mesoscale meteorological features such as larger scale convection features. Each of its 10 members is run independently using different configurations in the framework of the Weather Research and Forecasting (WRF) Model version 3.4 from April to September and version 3.5 from September to October (AFW-WEBS, 2013). The 10 member suite of WRF members changed configurations four times during the course of this study and is detailed in Appendix D. For further information on AFWA's choice of operational configuration and physics options, refer to the User's Guide for the NMM Core of the Weather Research and Forecast Modeling System Version 3 Handbook. WRF is a fixed-domain model that uses finite differences to represent the primitive equations. The MEPS obtains its initial and lateral boundary conditions from deterministic global models. These deterministic models include the GFS from NCEP, the GEM from CMC and the Unified Model (UM) from the United Kingdom Met Office (UKMO). The ensemble members are created by varying the global model chosen for the initial and boundary values and the physics parameterizations of mesoscale and microscale processes - surface fluxes, the planetary boundary layer (PBL), cloud microphysics, cumulus parameterization, etc. The MEPS is run at horizontal grid resolutions of 20km and 4km. The 20km model is comprised of a hemispheric domain and tropical swath covering the majority of the tropics and is run every 12 hours at 3-hour time steps out to 144 hours producing forecasts from 6 to144 hours. Table 1 details its characteristics.

Table 1: 20km MEPS domains, cycle times, completion times and forecast hours (AFW-WEBS, 2013).

MEPS Domain	Cycle Time	Run Complete	Forecast Hours
Northern Hemisphere	06Z/18Z	0830Z/2030Z	144
Tropical Swath	06Z/18Z	0830Z/2030Z	144

AFWA's 4km MEPS, as depicted in Table 2, covers various operationally significant fixed domains in addition to seven relocatable 4km domains for tropical cyclones and other contingencies. Its members are comprised of the same 10 ensemble members as MEPS20, with forecast output for every hour out to 72 or 84 hours depending on location, while all locations in this study have output out to 72 hours. The cumulus parameterization is turned off in MEPS4 (AFW-WEBS, 2013). Weisman et al (1997) showed that a horizontal resolution of 4km is small enough to explicitly represent most convective scenarios.

Table 2: 4km MEPS domains, cycle times, completion times and forecast hours (AFW-WEBS, 2013).

MEPS Domain	Cycle Time	Run Complete	Forecast Hours
CONUS	00Z/12Z	0230Z/1420Z	72
East Asia	00Z/12Z	02Z/14Z	72
Alaska	00Z	03Z/15Z	72
South West Asia	06Z	10Z	72
Europe	06Z	12Z	72
Afghanistan	18Z	20Z	72
Colombia	18Z	21Z	72
JTWC	00Z/12Z	05Z/17Z	84
28 OWS	00Z	04Z	72
Contingency	00Z	06Z	72
17 OWS	12Z	15Z	84
1 WXG	18Z	22Z	72
21 OWS	18Z	23Z	72

2.4 Research Question and Objective

While AFWA's EPS Point Ensemble Probability (PEP) bulletins are understood to be useful for characterizing forecast uncertainty for point locations, none of the three EPS point probabilities have undergone a site specific rigorous validation process. This research intends to begin that validation by comparing GEPS, MEPS20, and MEPS4 ensemble forecast probabilities with the actual probability of occurrence using Aerodrome Routine Meteorological Reports (METAR) and Aerodrome Special Meteorological Reports (SPECI) for various forecast parameters and geographical locations. Reliability and model skill diagrams with respect to forecast duration and were used to determine how well forecast probability compares to the observed frequency of occurrence. The desire is for this validation to enable operational weather forecasters to translate ensemble probability of occurrence to the actual probability that the threshold will be exceeded and to determine how long each EPS forecast is useful.

III. Methodology

3.1 Location and Time Period Selection

Ten geographically diverse locations were chosen for this study comprised of Air Force, Army, and Navy bases. These sites are listed with their respective International Civil Aviation Organization airport code in parentheses. Five locations are within the United States: Cape Canaveral AFS, Florida (KXMR); Little Rock AFB, Arkansas (KLRF); Tinker AFB, Oklahoma (KTIK); Dyess AFB, Texas (KDYS); and Fort Greely, Alaska (PABI) (Figure 4). Five locations are overseas: Kadena AB, Japan (RODN); Kunsan AB, South Korea (RKJK); Camp Lemonnier, Djibouti (HDAM); Ramstein AB, Germany (ETAR); and Sigonella NAS, Italy (LICZ) (Figure 5). These locations were selected based on forecast availability of the three EPS coupled with a high frequency of severe weather for their respective latitudes. The forecast parameters of interest for this study include thunderstorms, appreciable precipitation, and strong winds. These types of events are typically the most damaging to DoD resources. A time period ranging from April through October 2013 was selected for this study providing an ample data set for phenomena of interest. A larger sample would have been tested; however, due to the data storage limitations that ensemble output currently presents, AFWA does not archive ensemble output.

3.2 Data Sources

3.2.1 Point Ensemble Probability Bulletins (PEP Bulletins)

PEP bulletins (Figure 6) were provided through collaboration with Evan Kuchera, AFWA’s 16/WS Deputy Chief, Numerical Models Flight, Fine Scale Models and Ensembles Team Lead.

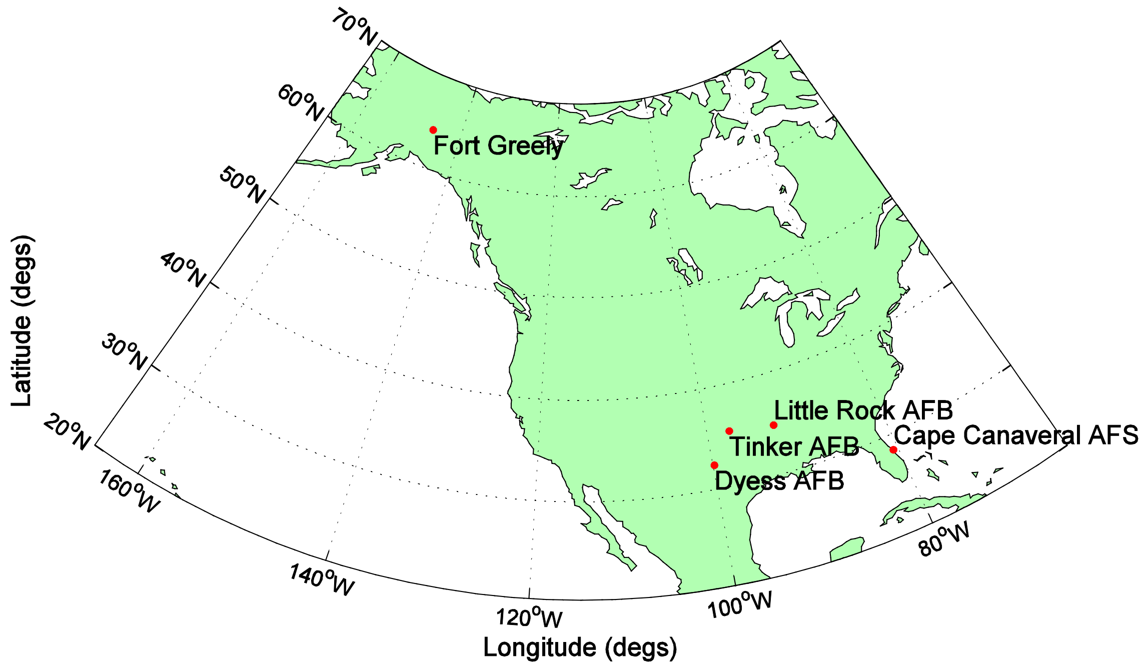

Figure 4: Map of locations selected in the United States based on frequency of severe weather events for the respective latitude.

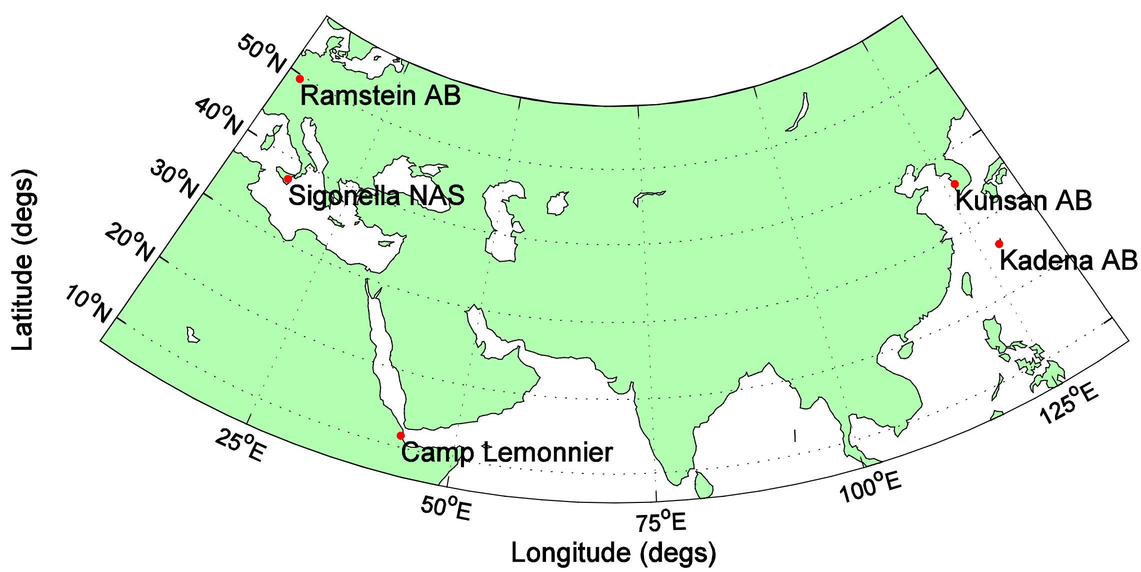

Figure 5: Map of locations selected overseas based on frequency of severe weather events for the respective latitude.

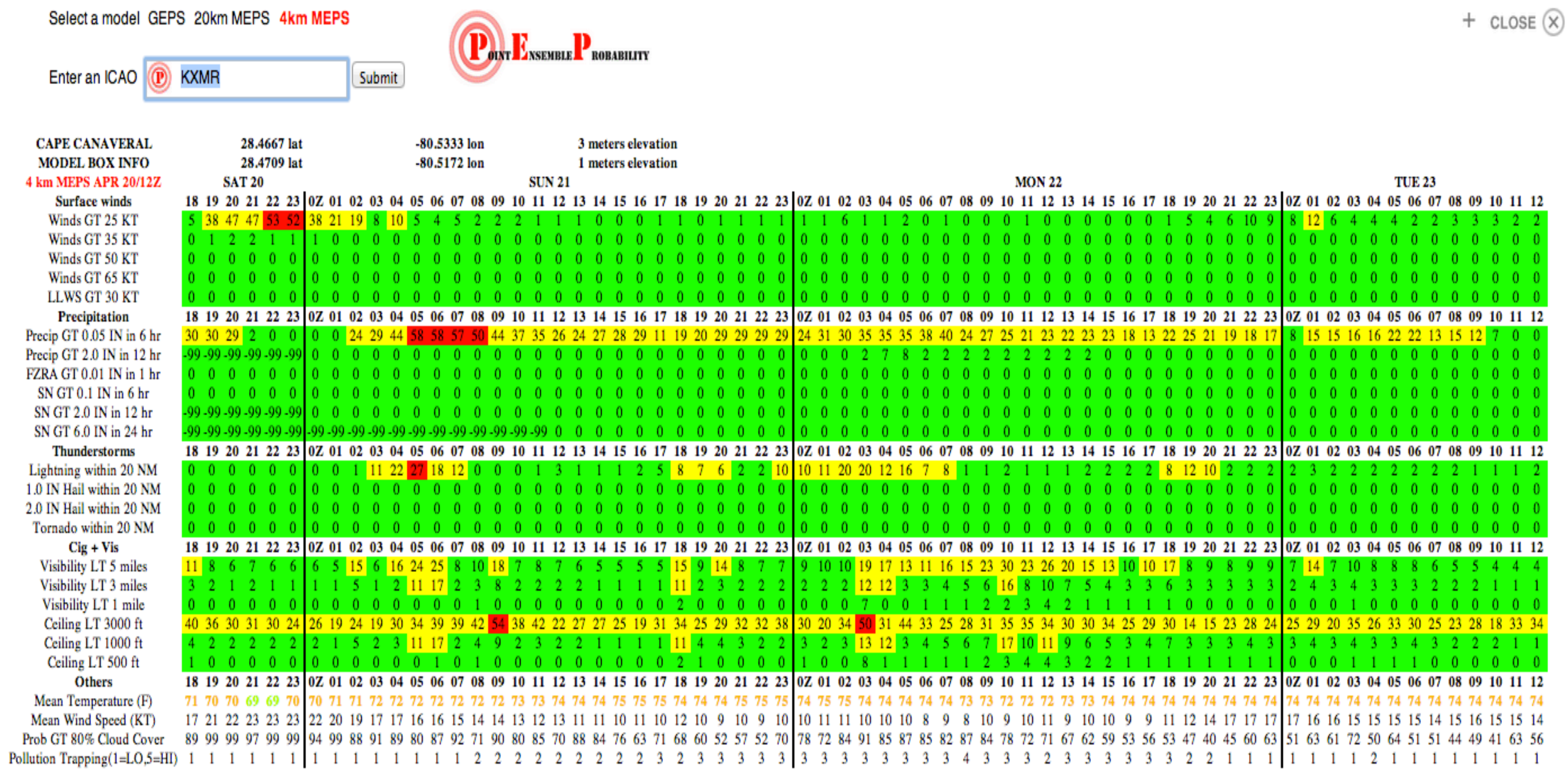

Figure 6: Example of a PEP bulletin showing forecast location, and for each forecast parameter, the probability for the respective forecast hours. Each probability is provided in percent and color coded based on an ORM threat level (AFW-WEBS, 2013)

Six different parameters from the PEP bulletins for each of AFWA's EPS were used in this research. For the GEPS and MEPS20, the parameters are: winds > 25kts ($11ms^{-1}$), > 35kts ($15ms^{-1}$), and > 50kts ($22ms^{-1}$); precipitation > 0.10in (2.5mm) in 6 hours, > 2.0in (51mm) in 12 hours; and, lightning within 20km. For the 4km MEPS the parameters are: winds > 25kts ($11ms^{-1}$), > 35kts ($15ms^{-1}$), > 50kts ($22ms^{-1}$); precipitation > 0.05in (1.27mm) in 6 hours, > 2.0in (51mm) in 12 hours; and, lightning within 20nm (37.04km). For each forecast interval, the probability is valid from the minute after the previous forecast hour to the current forecast hour. The colors overlaid on the forecast probabilities are based on a criteria developed at AFWA to highlight low risk (green), moderate risk (yellow) and high risk (red) to a warfighter's ORM process.

3.2.2 Aerodrome Routine Meteorological Report (METAR) and Aerodrome Special Meteorological Report (SPECI)

To compare EPS PEP bulletins to actual occurrences, this study used METARs and SPECIs archived by the 14th Weather Squadron, the Air Force's Combat Climatology Center, for the 10 selected locations. The METAR and SPECI format is prescribed by World Meteorological Office Publication 306 – Manual on Codes. The raw METARs and SPECIs were decoded and provided for this research by Mr. Jeff Zautner, 14/WS Meteorologist, Tailored Product Analyst. METARs are taken as a routine observation by an automated weather sensor once per hour within five minutes before the top of the hour for which the observation is valid. Anytime prescribed change thresholds were met, a SPECI observation was taken between routine top of the hour METARs. The worst-case condition within a particular forecast hour, whether from a METAR or a SPECI observation, was used to compare to the PEP bulletin probabilities.

3.2.3 Climatology

To evaluate the skill and utility of AFWA's EPS, the forecasts were analyzed using several different metrics. Most metrics require a reference, climatology, to compare with the forecasts. Climatology for each location was also provided by Mr. Jeff Zautner at the 14/WS. Maintaining consistency with the forecast intervals for each EPS, a 6-hour, 3-hour and 1-hour climatology for each month over a span of 10 years, January 2003 to December 2012, was used for the respective locations. This data set took into account up to 310 observations for each hour. The exception to this is Cape Canaveral (KXMR), which did not start taking METARs until 2006, thus totaling up to 212 observations used for each hour. For each wind parameter, the wind value and the peak wind remark were both considered to provide the highest wind recorded. For thunderstorms, on station, vicinity and lightning distant remarks were all used. Lastly, for precipitation parameters, routine METARs did not begin reporting 1-hour precipitation sums until sometime in 2007. Prior to 2007 precipitation was only required to be summed every 6 and 24 hours starting at 00Z for the respective day. To maintain consistency with all precipitation climatology, a smaller sample of METARs was used for each location running from January 2008 to December 2012 totaling up to a possible 155 observations for each respective hour.

3.3 Validation

3.3.1 Software Tools

Computer code was created to extract AFWA's PEP bulletins, METAR/SPECI and climatology spreadsheets, and to perform the statistical analysis used for this study. Code was also created to construct all map figures using MATLAB ® mmap toolbox.

3.3.1 Extracting PEP Bulletin Probabilities

Each PEP bulletin was in HTML format. These files were extracted based on month, day and forecast hour and placed in columns for each parameter. Once extraction was complete, each PEP bulletin was placed into a parsed text file as shown in Table 3.

Table 3: Example of extracted GEPS PEP bulletin for each parameter by month, day and hour. Probabilities are provided in percent.

Month	Day	Hour	Lightning within 20km	Winds > 25kts	Winds > 35kts	Winds > 50kts	Precip > 0.1in in 6hrs	Precip > 2in in 12hrs
6	7	0	10	16	0	0	42	0
6	7	6	6	9	0	0	24	0
6	7	12	0	2	0	0	17	0

3.3.3 Extracting METAR and SPECI Occurrences

All METARs and SPECIs were parceled out into smaller spreadsheets for each location and month. Rolling hourly sums were used for precipitation amounts; therefore, each month also included the last day of the previous month. Also, there is always a roll over into the next month due to the forecast length for each EPS. The GEPS has the longest forecast duration at 240 hours thus 10 days of the following month were tacked on to the end of each month’s spreadsheet. Since the shortest EPS forecast frequency is 1 hour, all the extracted METAR and SPECI for a given hour were checked to see if any of the six parameters tested occurred. If a particular event occurred between hours, it is always marked as occurring at the latter hour since each forecast probability includes the hour of forecast minus the previous 59 minutes. Once an event occurs, either at the routine METAR time or within the hour as a SPECI, it is flagged as occurring with a value of 1. If it did not occur, the value remains 0. Each SPECI occurring prior to the next routine METAR was flagged for each meteorological parameter that occurred. A text file was created for each location and month plus 10 days. In the text file, each row corresponds to a month, day and hour while each column corresponds to one of the six meteorological parameters indicated in Table 4.

Table 4: Example of METAR and SPECI verification for each parameter by month, day and hour. A value of 1 indicates that the parameter occurred during the previous hour.

Month	Day	Hour	Lightning within 20km	Winds > 25kts	Winds > 35kts	Winds > 50kts	Precip > 0.1in in 6hrs	Precip > 2in in 12hrs
6	7	4	1	1	0	0	1	0
6	7	5	1	0	0	0	1	0
6	7	6	0	0	0	0	1	0

3.3.4 Verification

For all three EPS, the model grid point varies in location and distance from the forecast site. Table 5 details the latitude and longitude for each site along with the distance from each site to the three EPSs' closest model grid points in nautical miles and kilometers.

For all forecast sites, MEPS4 is within approximately 1nm/1.85km, MEPS20 is within approximately 4.5nm/8.33km, and the GEPS is the nearest degree in latitude and half degree in longitude to the forecast sites which range from approximately 8nm/14.82km to 29nm/53.71km. An example of model grid point variability is evident in the difference between Figure 7 and Figure 8. The probability generated at these closest grid points was used for all wind and precipitation thresholds.

Table 5: Location and distance from the closest model grid point for each EPS in nm and km (AFW-WEBS, 2013).

Site	Lat (°)	Lon (°)	GEPS (nm/km)	MEPS20 (nm/km)	MEPS4 (nm/km)
ETAR	49.42	7.58	25.21/46.69	1.71/3.16	0.88/1.64
HDAM	11.55	43.17	28.77/53.28	1.73/3.21	1.03/1.91
KDYS	32.42	-99.83	26.42/48.93	4.29/7.94	1.00/1.86
KLRF	34.92	-92.15	8.92/16.52	4.08/7.55	0.69/1.27
KTIK	35.42	-97.37	25.86/47.89	3.93/7.28	0.79/1.47
KXMR	28.47	-80.53	28.09/52.02	3.53/6.53	1.04/1.93
LICZ	37.38	14.92	23.35/43.24	2.28/4.21	0.92/1.71
PABI	64.00	-145.73	6.12/11.34	4.48/8.29	0.99/1.83
RKJK	35.90	126.62	8.27/15.32	1.09/2.02	0.47/0.88
RODN	26.35	127.77	25.48/47.19	1.33/2.47	0.17/0.32

Lightning, on the other hand, represents the probability at the forecast site and for its surrounding area up to either within 20km or 20nm (37.04km) depending on which EPS is used. Both the GEPS and MEPS20 forecast lightning for a range of 20km for the forecast site, which has an area of 314.16 km^2. The closest METAR verification radius is vicinity thunderstorms (10nm/18.52km), which has an area of 269.38km^2. The 4km MEPS forecasts lightning for a range of 20nm from the forecast site, which has an area of 1077.54km^2. This area falls between the vicinity thunderstorm area already mentioned and the lightning distance verification radius (30nm/55.56km) which is an area of 2424.46km^2. For comparison, the respective forecast range rings of 20km and 20nm (37.04km) are plotted in Figure 7 and Figure 8 along with the METAR verification range rings of 5, 10 and 30nm (9.26, 18.52 and 55.56km). To bolster the sample size of lightning events, lightning occurrence on station in the predominate grouping of the METAR (within 5nm of the observation point), vicinity (lightning within 5-10nm), and distant lightning (lightning out to 30nm in the remarks section of the observation) were used. Consequently, the forecast probabilities for the GEPS and MEPS20 have to be scaled by a factor of eight and the MEPS4 by a factor of two to approximate the verification area of 2424.46km^2. By scaling, the assumption is made that all areas used to make up the validation areas have the same forecast probability. Please reference Appendix B for mathematical detailing of how this is accomplished while keeping the forecast probabilities less than 100%.

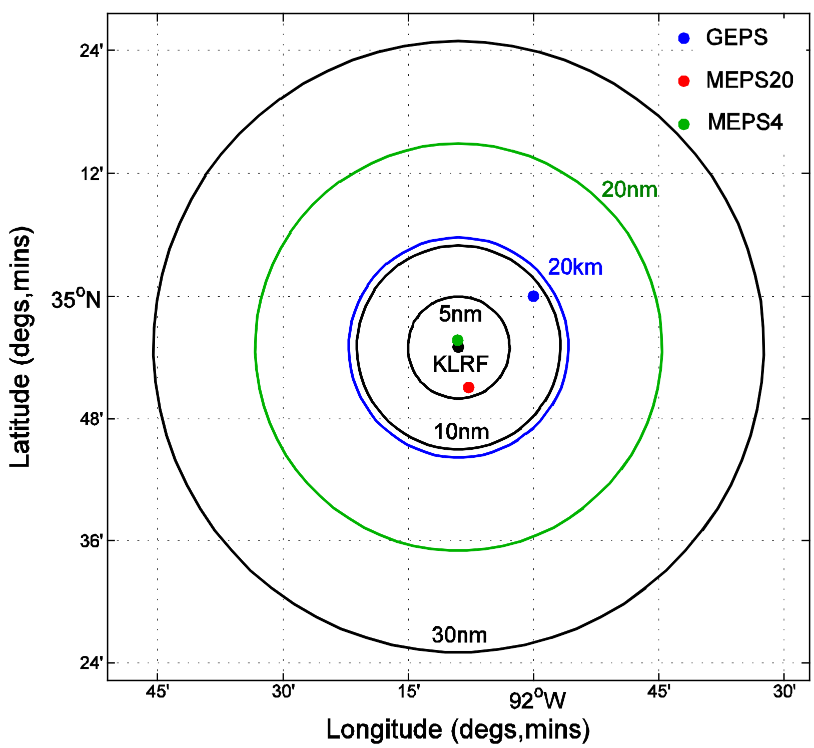

Figure 7: Little Rock AFB with range ring distances of 5, 10 and 30nm shown in black, GEPS and MEPS20 lightning within 20km range ring shown in blue, and MEPS4 lightning within 20nm/37.04km range ring shown in green. Also, each model grid point is displayed; GEPS in blue, MEPS20 in red and MEPS4 in green.

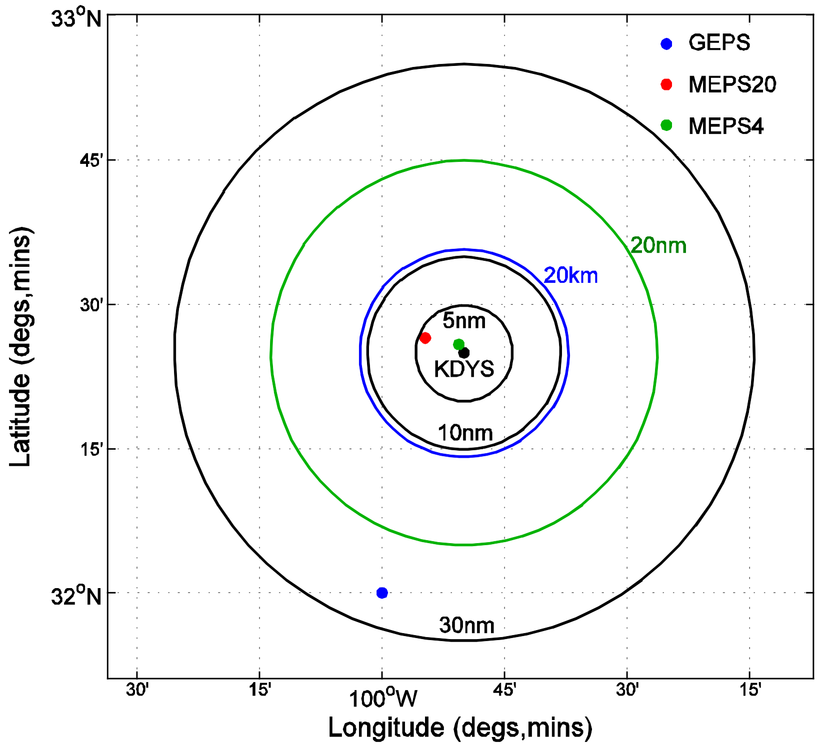

Figure 8: Dyess AFB with range ring distances of 5, 10 and 30nm shown in black, GEPS and MEPS20 lightning within 20km range ring shown in blue, and MEPS4 lightning within 20nm/37.04km range ring shown in green. Also, each model grid point is displayed; GEPS in blue, MEPS20 in red and MEPS4 in green.

3.3.5 Frequency of Occurrence vs. Ensemble Probability

Frequency of occurrence is the ratio of the number of actual occurrences of an event to the number of possible occurrences (Devore, 2004). This study measured the frequency of occurrence from METARs and SPECIs of the selected weather parameters as given by,

$$P(y_i) = \frac{N_i}{n}, n = \sum_{i=1}^{l} N_i \tag{5}$$

where $P(y_i)$ is the observed frequency of a particular event y_i, N_i is the number of actual occurrences of event y_i, and n is the total number of forecasted occurrences (Wilks, 2011). These observed frequencies were plotted in reliability diagrams for the ensemble forecast probabilities from the PEP bulletins. The goal is for the forecast probabilities to optimally match the probability of occurrence allowing for a skillful EPS.

3.3.6 Brier Score

The Brier score (BS) expresses how well a probability forecast verifies in relation to occurrence and non-occurrence for a specific forecast parameter (Brier, 1950). The BS averages the squared differences between the groupings of forecast probabilities and the corresponding binary representation of whether or not the forecasted event occurred (Wilks, 2011). The most widely used form of the BS is shown in Equation 6 where *n* is the number of occurrences, *y* is the forecast probability from 0 - 1.0, and o indicates whether the event occurred, with 1 signifying occurrence and 0 non-occurrence.

$$BS = \frac{1}{n}\sum_{k=1}^{n}(y_k - o_k)^2 \qquad (6)$$

For this study, the BS was calculated using ensemble probabilities, y_k, from AFWA's PEP bulletins and actual occurrences, o_k, from the decoded METAR and SPECI observations. Probabilistic forecasts that perfectly match reality (i.e. 100% forecast probability for every occurrence and 0% forecast probability for every non-occurrence) will produce a BS of 0, while forecasts that are universally incorrect (i.e. 100% forecast probability for every non-occurrence and 0% forecast probability for every occurrence) will result in a BS of 1; therefore, a lower BS indicates more reliable probabilistic forecasts.

To provide further utility of the BS, Murphy (1973) suggested that the BS can be decomposed into three terms – reliability, resolution and uncertainty as indicated in Equation 7 (Wilks, 2011).

$$BS = \frac{1}{n}\sum_{i=1}^{I} N_i(y_i - \bar{o_\iota})^2 - \frac{1}{n}\sum_{i=1}^{I} N_i(\bar{o_\iota} - \bar{o})^2 + \bar{o}(1-\bar{o}) \tag{7}$$
$$\quad Reliability \qquad\qquad Resolution \qquad Uncertainty$$

The first term, reliability, consists of the weighted average of the squared differences between binned forecast probabilities, y_i, and the subsample relative frequency of occurrences for the parameter in question, $\bar{o_\iota}$. Equation 8 from Wilks (2011) shows how $\bar{o_\iota}$ is calculated.

$$\bar{o_\iota} = \frac{1}{N_i}\sum_{k\in N_i} o_k \tag{8}$$

A reliability value of 0 indicates that the forecast exhibits perfect reliability meaning that the forecast probability perfectly matches the observed frequency while a score of 1 indicates no correlation between the forecast probability and the observed frequency (Wilks, 2011). The second term, resolution, consists of the weighted average of the squared differences between the subsample relative frequency of the parameter in question, $\bar{o_\iota}$, and the overall relative frequency (climatology), $\bar{o}$, for the parameter. The overall relative frequency as shown in Equation 9 is the sum of all the occurrences divided by the sample size.

$$\bar{o} = \frac{1}{n}\sum_{k=1}^{n} o_k \tag{9}$$

Resolution values range from 0 to 1. The higher the resolution value, the easier it is to obtain a good BS and BSS. A high resolution value indicates the EPS ability to forecast higher probabilities that occur. The third term, uncertainty, is independent of the probability forecast and is a function of the climatology used. Events that occur rarely or frequently will possess a low uncertainty while an event that never occurs or always will have a value of 0. The most difficult events to forecast are those that have climatology of exactly 50% probability of occurrence, thus leading to the highest obtainable uncertainty value of 25%. These examples as well as all other scenarios are shown in Figure 9.

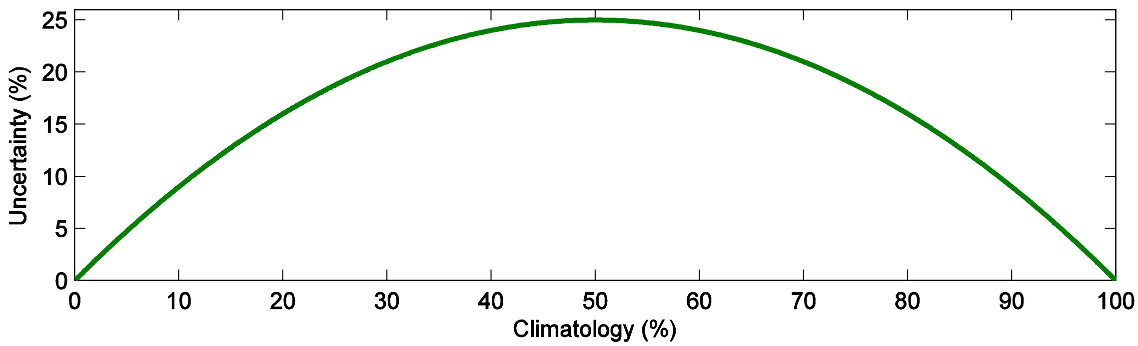

Figure 9: Graph of the relationship between uncertainty and climatology.

The BS decomposition shown in Equation 7 will never exactly equal the BS from Equation 6 due to multiple factors. First, the decomposition requires binning the EPS probabilities to solve, leading to variance and covariance within the bins used (Stephenson, 2008). Stephenson showed that with the addition of two terms in the decomposition, one accounting for the variance and the other the covariance, the impact of the truncation errors from binning is less severe. Secondly, if an enormous sample of forecast probabilities and corresponding observations are tested, allowing for each possible probability from 0-1.0 to have its own bin, the three terms in the decomposition will not equal Equation 6 due to bias in each term (Bröcker, 2012). Bröcker showed that even if the sample size is increased to infinity, the reliability is systematically overestimated and the uncertainty is systemically underestimated while resolution can be either. To account for these biases, Ferro and Fricker (2012) developed a new decomposition where each term is less sensitive to their respective biases. These two additional decompositions were not used in this research as the results showed that, overall, the reliability, resolution and uncertainty display correct trends in producing BSS values.

3.3.7 Brier Skill Score

The more rare an event, the easier it is to obtain a good BS without the EPS possessing any real skill over climatology. For this reason, the BSS was used to determine the relative skill of the EPS over that of climatology forecasting whether or not an event will occur. BSS is defined in Equation 10 as the ratio of the BS minus the climatological BS (BS_{ref}) to a perfect BS of 0, minus the climatological BS (BS_{ref}) (Wilks, 2011).

$$BSS = \frac{BS - BS_{ref}}{0 - BS_{ref}} = 1 - \frac{BS}{BS_{ref}} \qquad (10)$$

Using the decomposition provided in Equation 7 and some algebra, the BSS can also be solved for in terms of reliability, resolution and uncertainty as shown in Equation 11 (Wilks, 2011).

$$BSS = \frac{Resolution - Reliability}{Uncertainty} \qquad (11)$$

Because of the truncation error due to binning and the biases in the three terms already mentioned, the BSS was calculated and plotted using Equation 10. While the BSS from Equation 11 is not plotted, it is important to understand how the decomposition values can be used to solve for the BSS.

3.3.8 Reliability Diagrams

Although the numerical values for reliability, resolution, uncertainty, BS and BSS provide a sense of how well an EPS performs, a more comprehensive approach lies in the conceptual understanding and graphical depiction of a reliability diagram as shown in Figure 10. Shaded in red is the area of skill. This area of skill encompasses the region between the vertical line created from the intersection of the climatology and the zero reliability line to the diagonal line that splits the area between the climatology and the zero reliability line into equal areas. For this example, the 1-10% probability bin falls on the skill line thus being marginally skillful. The 41-50% bin falls outside the area of skill while the remaining bins fall within the area of skill making the BSS positive. When resolution is greater than the reliability, positive skill will exist. However, if binning and bias errors are greater than the difference between the two, it is possible for the reliability to be greater than the resolution while Equation 10 still gives a positive BSS. This was very rarely observed in the approximately 5,000 figures investigated. Reliability (how close the observed frequencies of occurrences match the zero reliability line), resolution (how far away the is the observed frequency away from climatology) and the skill of the EPS (majority of the observed frequency weight in the in area skill) are clearly apparent and aided by the value of each metric (REL = reliability, RES = resolution and UNC = uncertainty) in Figure 10. To create reliability diagrams, the EPS forecast probabilities were binned to get the total number of forecasts that occurred in each respective bin. The bin width chosen was 10% with the exception of having a 0% bin when the EPS forecasts no chance of occurrence. Next, the number of times the event occurred in each bin was calculated. These two quantities, forecast probabilities and number of times the event occurred, allowed for the calculation of the frequency of occurrence as detailed in Equation 5.

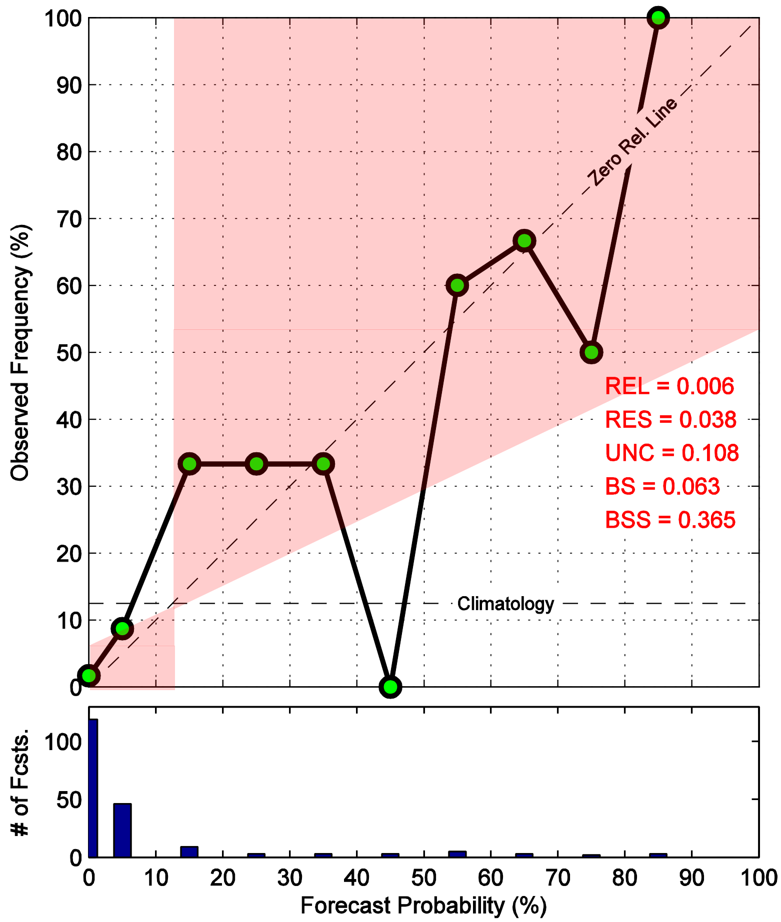

Figure 10: Reliability diagram example. The observed frequency for each probability bin is depicted as the black line with green points representing the center of each bin. The area of skill is highlighted in red. The dashed diagonal line represents the line of zero (perfect) reliability. The climatology (no resolution) is shown as a horizontal dashed line. BS and BSS are provided along with the components that make up the score. The subplot indicates the number of forecasts in each bin.

Each of these observed frequency values was plotted as a green dot at the center of each bin with a line connecting each point. Also, the climatology and zero reliability lines were plotted for each figure. The more the frequency of occurrence line correlates with the zero reliability line, the lower the reliability value, thus achieving a better score.

A good score can be achieved regardless of how frequent the event occurs at the location. For resolution, the further the forecast probabilities verify away from climatology, the higher and better the score. If the EPS struggles to forecast away from climatology, the resolution values will remain small. Uncertainty will fluctuate solely due to the climatology. Lastly, to show the sample size within each bin, all of the reliability diagrams include a subplot detailing how many forecasts exist for each bin at the bottom of the plot as shown in Figure 10.

IV. Results

4.1 EPS Skill

For each location and the six parameters tested, the utility of each EPS with respect to forecast hour is calculated using the BSS as defined in Equation 10 along with the decomposition of the BS from Equation 7. Two parameters, winds > 50kts and precipitation > 2.0in in 12 hours, occur too infrequently to obtain any useful results thus are not included. Because it would be impractical to include all the figures generated, tables are used to convey forecast skill for each parameter. These tables list each site and EPS with the corresponding forecast hours of positive skill, skillful percentage of the forecast time, and the average positive skill for sites that had a sufficient number of occurrences, approximately 15 events or more, to obtain meaningful results. Typically with less than 15 events the BSS behaves erratically and little value is gleaned from the results. Due to diurnal variations in the uncertainty, some periodicity is evident in the BSS as shown in Figure 11. The BSS is shown in black while the subplot in the lower portion shows the composition of the BSS – uncertainty in green, reliability in red and resolution in blue. To get a better sense of model skill trends, the BSS trend is smoothed by averaging with the two closest BSS values to its left and right taking into account five BSS values total. For the BSS values next to the endpoints, they are averaged with the first and last BSS values, respectively, while the first and last BSS values are unaltered. These values are used to create the duration of forecast hours with positive skill shown in Tables 6 though 9. The BSS with respect to all forecast hours as illustrated in Figure 11 will continue to show unaltered BSS values.

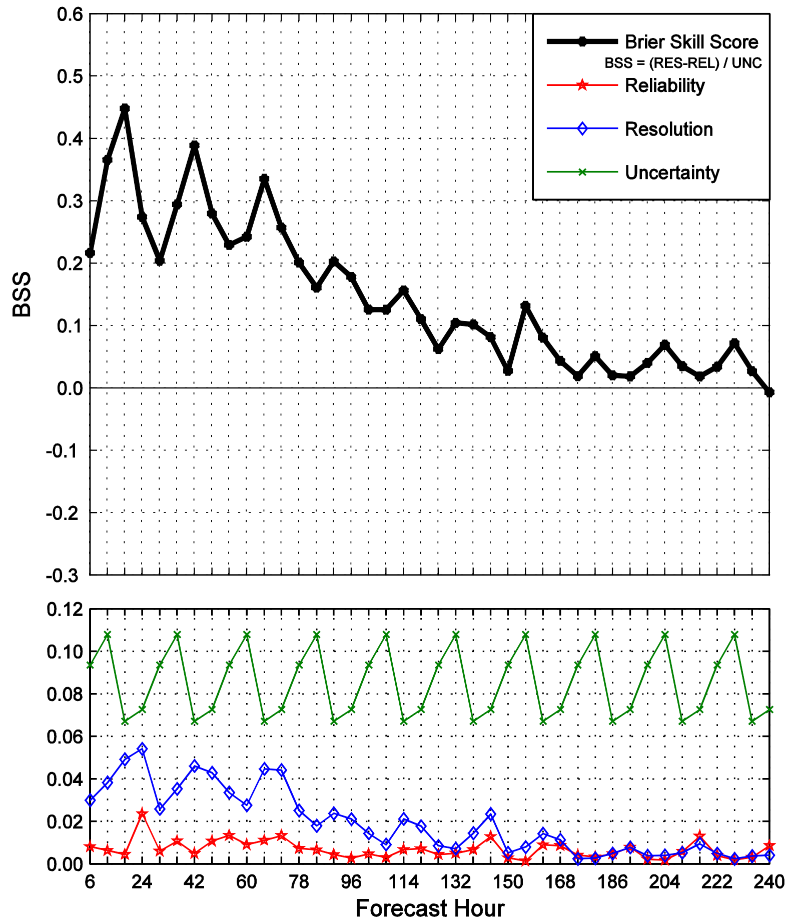

Figure 11: GEPS Precipitation > 0.1in BSS for Ramstein AB from April-October 2013 with reliability, resolution and uncertainty data shown in the subplot.

All three EPS overforecast lightning. As horizontal resolution increases the BSS is positive for more forecast hours. In general, for precipitation, the GEPS provides the longest duration of positive skill; however, both of AFWA's regional EPS (MEPS20 and MEPS4) typically provide a greater BSS during their respective hours of positive skill. A potential reason for the GEPS having a longer period of positive skill lies in its composition of 62 ensemble members from three different model systems as compared to the MEPS20 and MEPS4 only being comprised of 10 members from one model system. Having 52 additional members allows the GEPS to account for more forecast uncertainty whereas the spread of model solutions should provide a more realistic resemblance of the future state of the atmosphere. However, because of the one-degree, approximately 111km resolution, the parameters tested are resolved with less accuracy than with the MEPS20 and MEPS4. In most cases the MEPS20 average skill for precipitation is close to the GEPS average skill and in some cases less. The tradeoff for having less ensemble members and an increased horizontal resolution does not pay off in all cases for the MEPS20 while its does for the MEPS4. The opposite is true for winds where both the MEPS20 and MEPS4 prove to have a significant increase in average positive skill versus GEPS. Tables 6 through 9 highlight differences that arise due to geographic location, model resolution, and convective parameterization vs. explicit resolving convection. Additionally, to take a closer look at possible conditional biases, reliability diagrams must be analyzed to see what trends exist in the EPS.

4.2 Lightning

Table 6 indicates there is no real correlation between geographic region and positive skill duration. However, for the four sites where the three EPS have a sufficient sample size of occurrence, the MEPS4 produces the longest duration of positive forecast skill while the MEPS20 average positive BSS are slightly higher (< 0.05) than the MEPS4. This can be attributed to the MEPS4 having more positive forecast hours of positive skill than the MEPS20. For MEPS20, the BSS is only positive for a few hours and has a steeper slope towards negative values. Since the average positive BSS are similar and the MEPS4 has considerably more forecast hours of positive skill, the MEPS4 performed the best. One reason that the MEPS4 outperformed the other two EPS is that the 4km horizontal grid spacing allows for resolution of smaller convective processes with improved precision while the other two EPS rely on cumulus parameterization schemes for generation of thunderstorms.

For most locations there were more hours that possessed a positive BSS than indicated in Table 6. These hours do not show up in the table because, typically, the hours surrounding these positive BSS have larger negative values and when using the smoothing technique already mentioned, these averaged hours are negative.

Table 6: Lightning Positive Skill Duration, Skillful Percentage of Forecast and Average Positive Skill. Blanks indicate insufficient occurrence sample size.

Site	EPS	Forecast Hours of Positive Skill	Skillful % of Forecast	Avg Positive Skill
ETAR	GEPS	0	0	0
	MEPS20	0	0	0
	MEPS4	—	—	—
KDYS	GEPS	0	0	0
	MEPS20	6-9	4.2	0.190
	MEPS4	6, 13-16, 21-23, 37-45, 63-68	35.8	0.145
KLRF	GEPS	0	0	0
	MEPS20	6-9	4.2	0.127
	MEPS4	7, 9, 11-16, 27-32	55.2	0.089
KTIK	GEPS	24-36	7.5	0.072
	MEPS20	6-42, 60-66, 84-90	40.4	0.159
	MEPS4	11-27, 35-51, 59-61, 65-72	67.2	0.137
KXMR	GEPS	0	0	0
	MEPS20	0	0	0
	MEPS4	35-39, 61-62	10.4	0.051
LICZ	GEPS	6-12	5	0.159
	MEPS20	6-51, 69-75, 99, 141-144	89.5	0.112
	MEPS4	—	—	—
RKJK	GEPS	0	0	0
	MEPS20	0	0	0
	MEPS4	6-7, 31-35	10.4	0.086
RODN	GEPS	0	0	0
	MEPS20	0	0	0
	MEPS4	0	0	0

4.2.1 Lightning Overforecasting

Lightning reliability diagrams for most locations and forecast hours depict the BSS as less than 0 (worse than climatology) which indicates that lightning is overforecast. Little Rock AFB GEPS forecast hour 24 (Figure 12) serves as an example of this overforecasting.

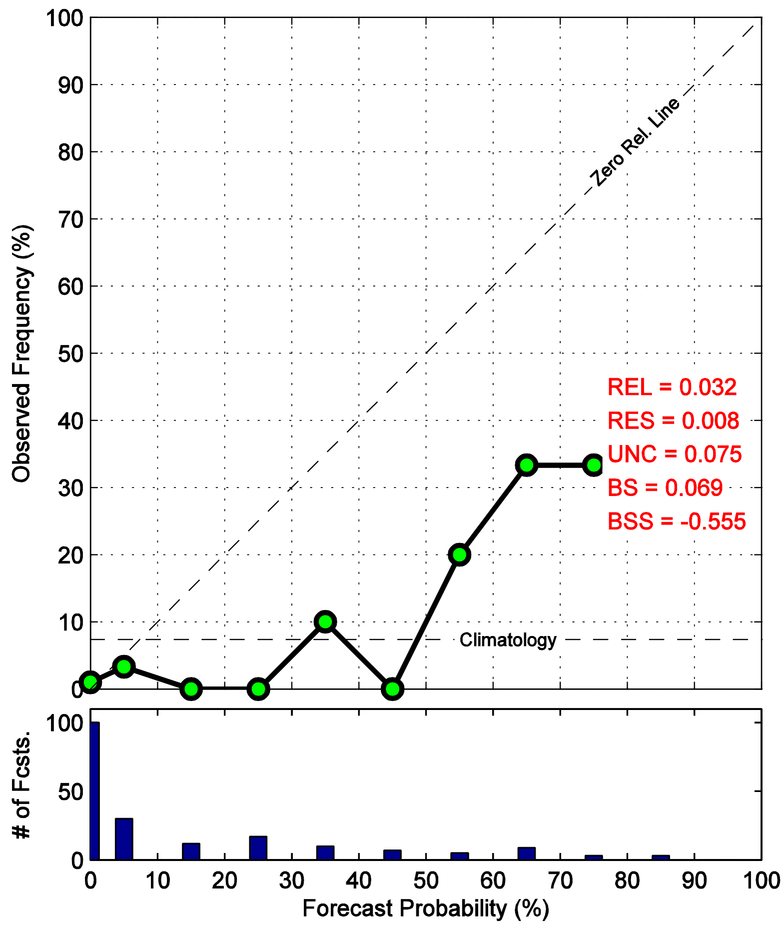

Figure 12: GEPS 24hr Lightning within 20km reliability diagram for Little Rock AFB from April-October 2013 indicating that lightning is overforecast.

The heaviest weighted probability bin, 0%, and the second heaviest weighted probability bin, 1-10%, closely match the observed frequency of occurrence while the remaining forecast probability bins are severely overforecast. For example, probability bin 51-60% has an observed frequency of 20% in only one out of the five forecasts verified. Also, the forecast probabilities greater than 10% total 66 forecasts which is more than double the 30 forecasts in the 1-10% probability bin. Since all of these observed frequencies comprised double the weight of the 1-10% probability, are well below the zero reliability line, and most observed frequencies do not deviate far from climatology, the observed frequency line falls outside the area of skill leading to a BSS of -0.55. To demonstrate that lightning is overforecast for the majority of forecast hours, an average observed frequency value is calculated by totaling the observed frequencies for all forecast hours for each probability bin and dividing by the total number of forecasts at every forecast hour for each probability bin. This calculation yields the following eleven averaged observed frequencies in order of probability bins from 0% to 91-100%, respectively; 0.57%, 2.39%, 4.46%, 10.21%, 13.15%, 17.94%, 21.65%, 27.20%, 39.13%, 46.99% and 62.50%. For example the 41-50% probability bin there is an observed frequency of occurrence of 17.94% which is too low by at least 23%. The other forecast probability bins show that the averaged frequency of occurrence values are remarkably less, clearly indicating the overforecasting bias that persists for the entire forecast period.

Upon review of the BSS trends for the GEPS forecast period at Little Rock AFB (Figure 12), it is evident that the majority of the overforecasting takes place during overnight hours. These hours are typically not favorable for lightning as surface heating has subsided and the atmosphere has used up its available energy for convection. A clear indication of less convection occurring overnight is the dip in uncertainty values from approximately 0.15 during the afternoon to approximately 0.08 overnight.

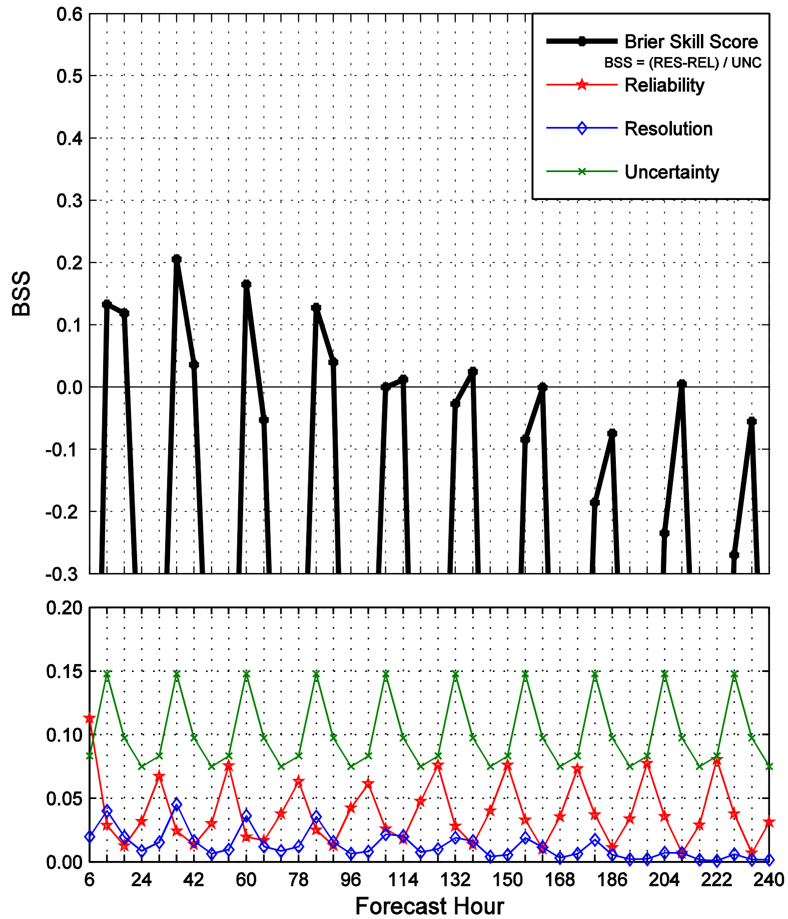

Figure 13: GEPS Lightning within 20km BSS for all forecast hours at Little Rock AFB from April-October 2013 indicating most scores near 0.

Likewise, the same overforecasting bias is observed for the MEPS20 reliability plots for most locations and forecast hours; however, it is less pronounced. Figure 14 for Tinker AFB forecast hour 21 illustrates this as more of the observed frequencies of occurrence are closer to the zero reliability line than the GEPS example allowing for a weak positive BSS. Calculating average observed frequencies as defined previously yields the following eleven averaged observed frequencies in order of probability bins from 0% to 91-100%, respectively; 3.69%, 11.06%, 14.07%, 20.46%, 19.50%, 33.95%, 32.73%, 40.54%, 42.03%, 51.43% and 100%. These values clearly indicate an overforecasting bias; however, this bias is less severe than the GEPS example.

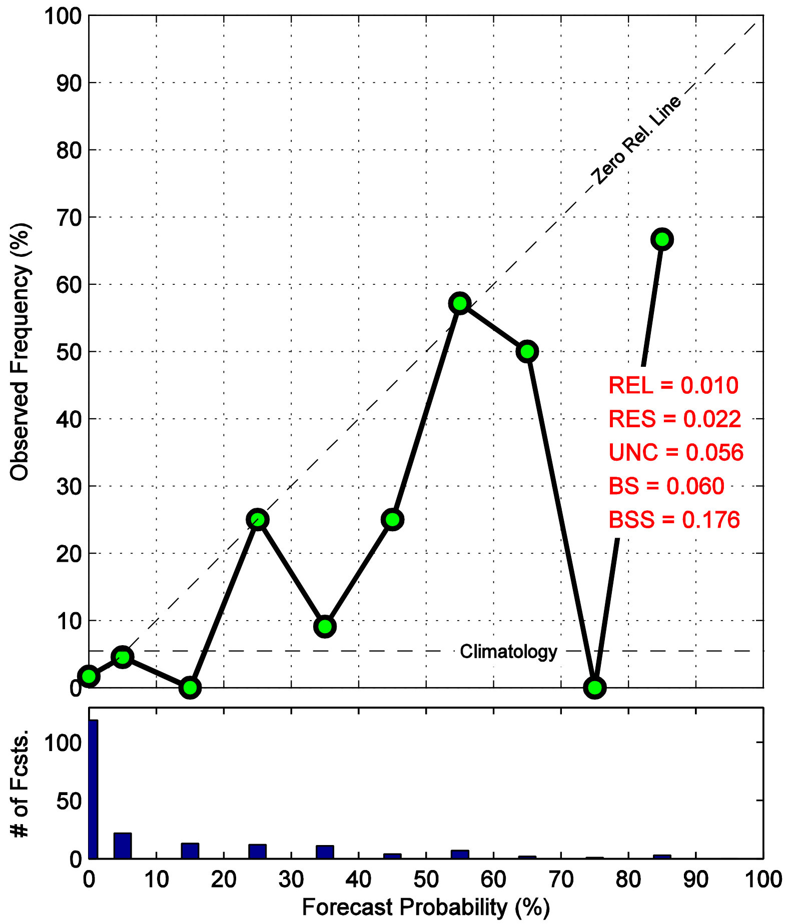

Figure 14: MEPS20 21hr Lightning within 20km reliability diagram for Tinker AFB from April-October 2013 indicating that lightning is overforecast.

Considering the BSS for the forecast period (Figure 15), it is evident that fewer hours are below 0 than in the GEPS example (Figure 12) and BSSs are higher when above 0. Similar to the GEPS example, sharp BSS dips can be seen when the MEPS20 forecasts lightning overnight when it typically does not occur.

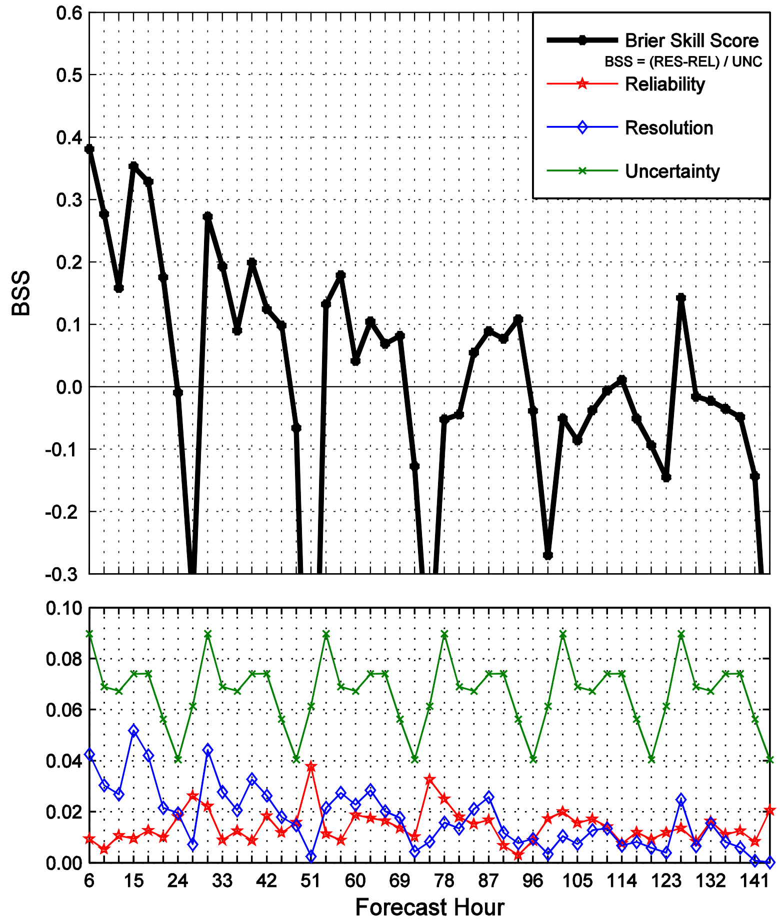

Figure 15: MEPS20 Lightning within 20km BSS for all forecast hours at Tinker AFB from April-October 2013 indicating most scores above 0 during the day and below 0 at night.

The MEPS4 is adversely impacted by the overforecasting bias more so than the MEPS20 for all locations. Calculating an average observed frequency for Tinker AFB yields the following eleven averaged observed frequencies in order of probability bins from 0% to 91-100%, respectively; 1.66%, 3.78%, 11.74%, 17.25%, 25.61%, 35.38%, 34.19%, 36.11%, 39.02%, 61.67% and 52.63%. There are more hours of positive skill, as indicated in Table 6, as MEPS4 does not forecast high probabilities thus not populating many of the bins where overforecasting bias is most prevalent. The BSS trend for the entire forecast period (Figure 16) is similar to the other two EPS, better than climatology during the day and worse than climatology late at night when uncertainty values dip.

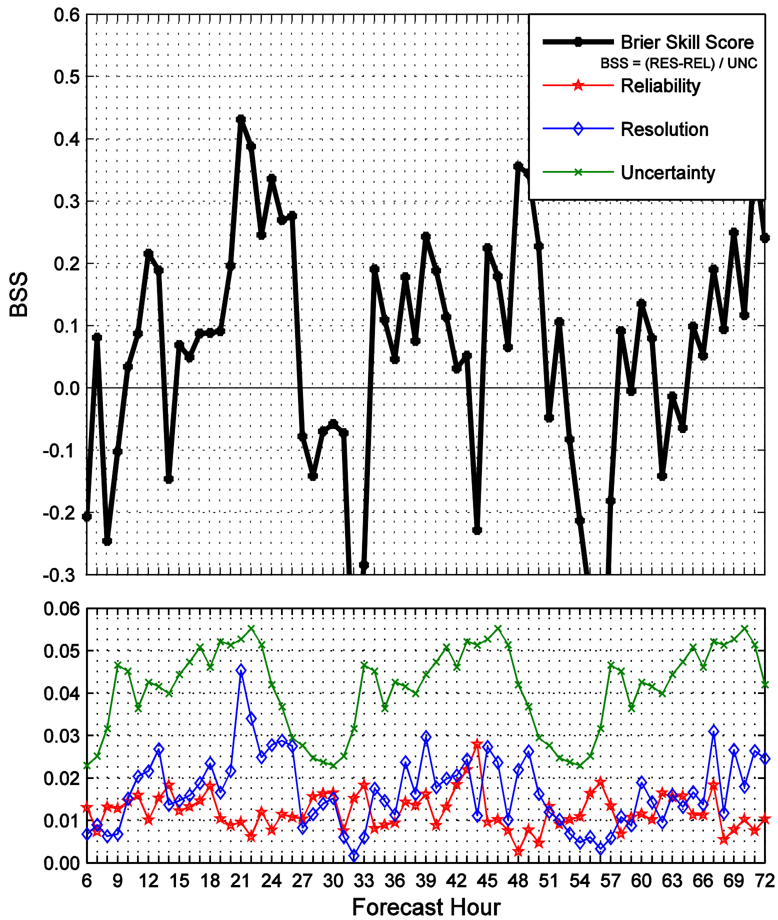

Figure 16: MEPS4 Lightning within 20km BSS for all forecast hours at Tinker AFB from April-October 2013 indicating most scores above 0 during the day and below 0 at night.

If the GEPS, MEPS20, and MEPS4 can be tuned to bring the observed frequencies closer to the zero reliability line, the EPS would become either skillful or more skillful correcting the overforecasting bias. One way to potentially achieve this is by forecasting less probabilities of lightning occurrence during the late night hours when lightning rarely occurs

4.3 Winds

4.3.1 Winds > 25kts

The average positive skill in Table 7 shows that for winds > 25kts, increasing horizontal resolution equates to a more positive and better BSS regardless of location.

Table 7: Winds > 25kts Positive Skill Duration, Skillful Percentage of Forecast and Average Positive Skill. Blanks indicate insufficient occurrence sample size.

Site	EPS	Forecast Hours of Positive Skill	Skillful % of Forecast	Avg Positive Skill
ETAR	GEPS	6-228	95	0.083
	MEPS20	—	—	—
	MEPS4	—	—	—
KDYS	GEPS	6-126	52.5	0.086
	MEPS20	6-144	100	0.206
	MEPS4	6-72	100	0.267
KLRF	GEPS	6-18	7.5	0.020
	MEPS20	6-9	4.2	0.235
	MEPS4	—	—	—
KTIK	GEPS	6-132	55	0.108
	MEPS20	6-144	100	0.224
	MEPS4	6-8, 18-35, 42-58, 66-72	53.7	0.286
KXMR	GEPS	6-36, 72-132, 198-240	55	.068
	MEPS20	—	—	—
	MEPS4	—	—	—
LICZ	GEPS	6-18	7.5	0.039
	MEPS20	—	—	—
	MEPS4	—	—	—
PABI	GEPS	0	0	0
	MEPS20	6-45, 60-72	34	0.024
	MEPS4	6-72	100	0.421
RKJK	GEPS	6-144	60	0.169
	MEPS20	6-18, 27-42, 51-57, 78-84	25.5	0.178
	MEPS4	—	—	—
RODN	GEPS	6-240	100	0.132
	MEPS20	6-144	100	0.298
	MEPS4	6-72	100	0.527

This is especially true for areas where terrain effects play a large role in wind speed and direction. For example, Fort Greely, AK (PABI) is located on the edge of the Tanana Valley and bordered by three extensive mountain ranges – the White Mountains to the North, the Yukon Tanana Uplands to the Northeast, and the Alaska Range to the South as shown in Figure 17. These mountain ranges cause winds to funnel through mountain passes and valleys. The coarser the resolution the harder it is for the EPS to resolve these terrain effects. A comparison of AFWA's three EPS is shown in Figure 18. For the GEPS, all forecast hours have a negative BSS as indicated by the blue line with values ranging from approximately -0.32 to -0.12.

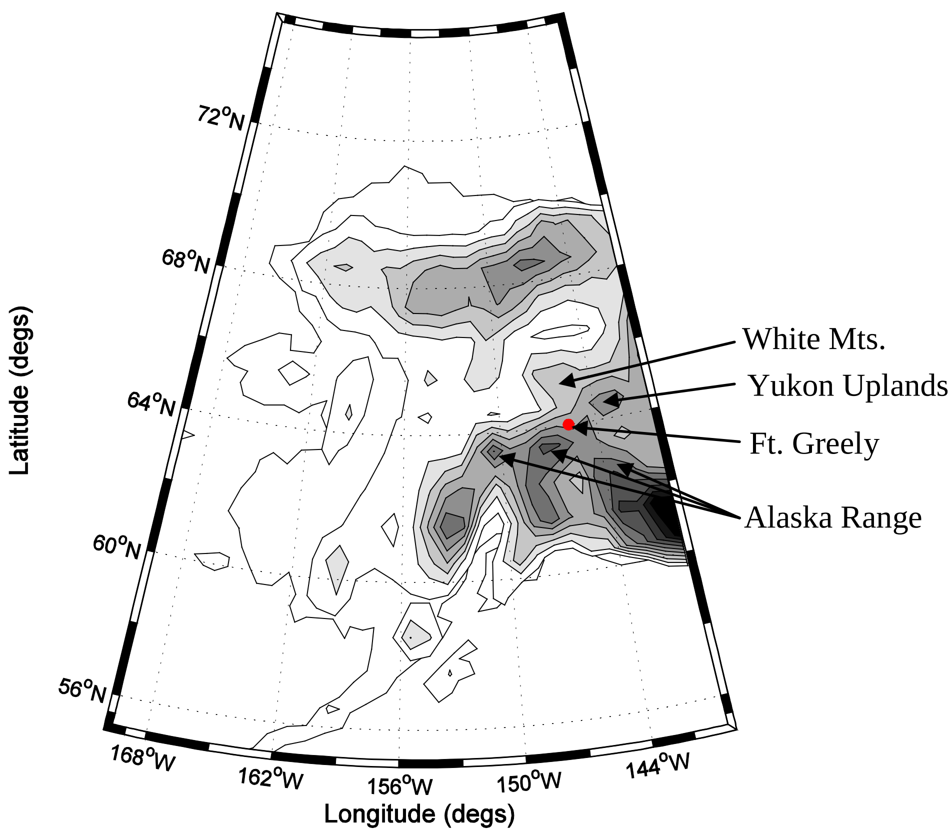

Figure 17: Map of one-degree resolution terrain around Fort Greely (red point). Darker filled contours represent increasing terrain heights.

Scores begin to improve with the MEPS20 as 34% of the forecast times show a weak positive BSS as indicated by the red line. These values oscillate around 0 from approximately -0.15 to 0.1 adding or detracting little from climatology. For the MEPS4, the 4km resolution indicated by the black line substantially affects all forecast hours resulting in positive BSS and an average BSS increase of two orders of magnitude over MEPS20. Values range from approximately 0.2 to 0.7 with substantial skill over climatology for all forecast hours.

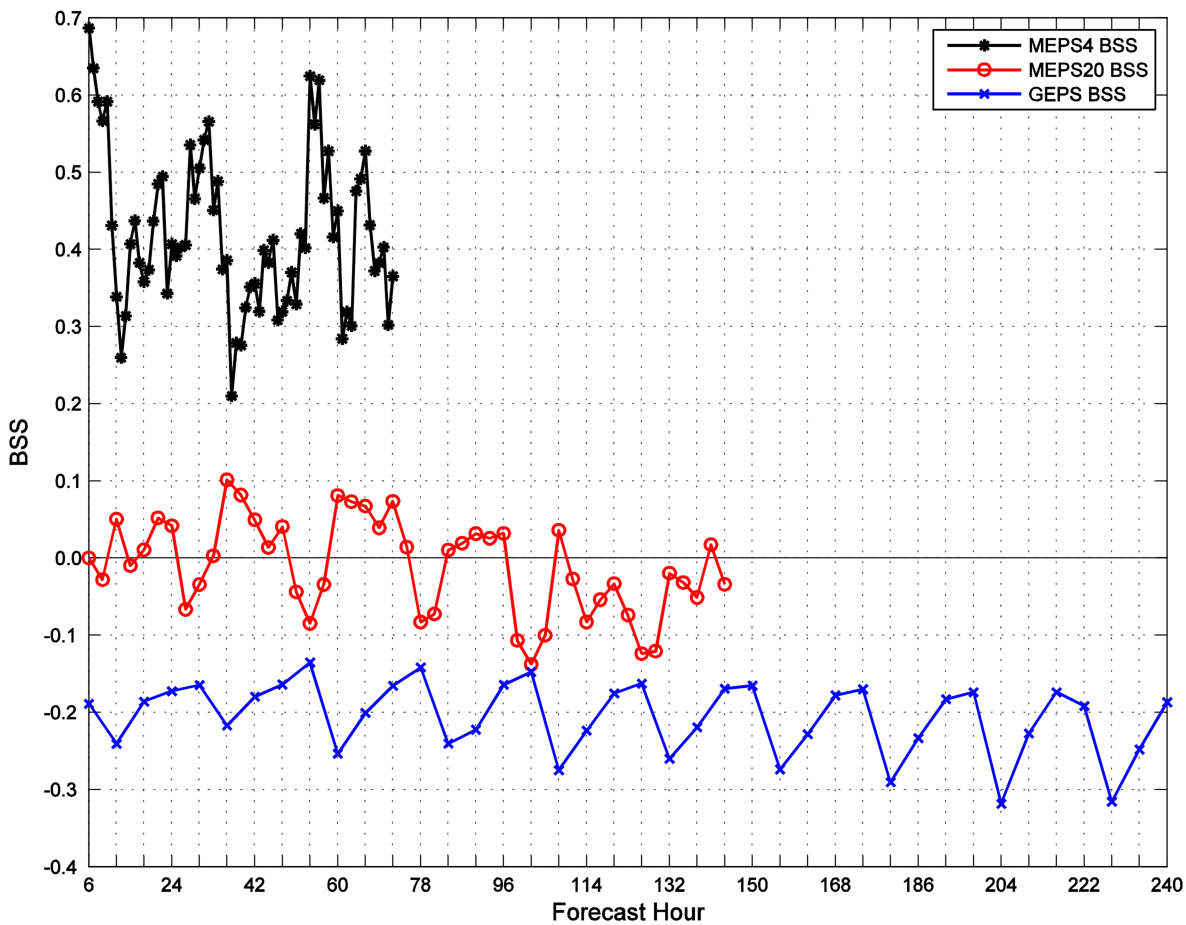

Figure 18: Comparison of MEPS4, MEPS20 and GEPS BSS from Apr-Oct 13 for Fort Greely Winds > 25kts. MEPS4 is shown in black, MEPS20 is shown in red, and GEPS is shown in blue.

Reliability diagrams for all three of the EPS depict reasons for these results. For all GEPS forecast hours, wind events are missed meaning that when the EPS forecasts a probability of 0% there are instances where winds greater than 25kts occur. Overall, 18 events are missed for each forecast hour when the average is taken for all forecast hours. Also, all the wind speeds are severely underforecast. For example, the 30-hour forecast depicted in Figure 19 shows that the GEPS missed 11 events out of 180, producing a 6% observed frequency when the forecast probability is 0%.

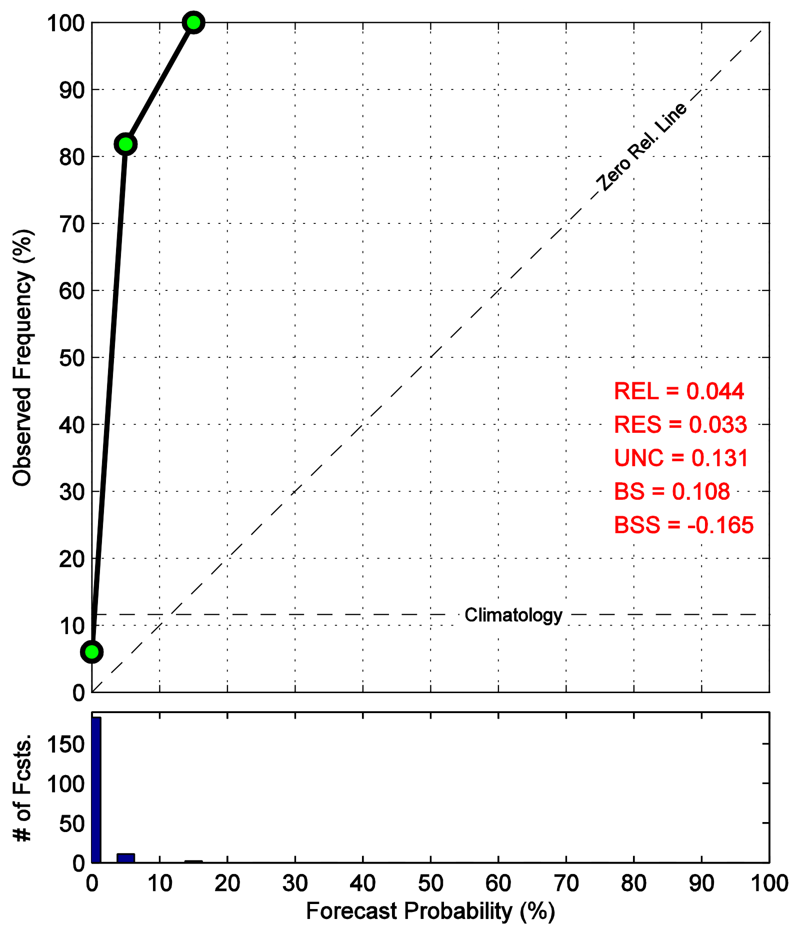

Figure 19: GEPS 30hr Winds > 25kts reliability diagram for Fort Greely from April-October 2013 indicating that occurrences are missed and wind speeds are underforecast.

This 0% probability bin falls on the skill line thus not adding any significant skill to the forecast. The next probability bin, 1-10%, has 9 occurrences out of 11 forecasts leading to an 82% observed frequency. Calculating the mean of the observed frequency for this particular bin and all forecast hours results in a 58.0% observed frequency. The last bin, 11-20%, has a 100% observed frequency as the two forecasts for this bin verified; however, the sample size is small, only adding a small positive contribution to the BSS. Due to missed events and severe underforecasting bias, the reliability for most forecast hours is relatively high while the resolution is relatively low because the forecast probabilities do not deviate much for climatology. Consequently, the GEPS's BSS stays negative for the entire forecast duration.

The MEPS20 with its increased grid resolution shows some improvement by missing less events and possessing a less severe underforecasting bias as displayed in Figure 20. When all forecast hours are averaged eight events are missed per forecast hour which is 10 less than the GEPS. Forecast hour 21, as shown in Figure 20, confirms this result with five events missed out of 173 producing a 2.8% observed frequency when the forecast probability is 0%. Also, the 1-10% probability bin contains only eight occurrences out of 15 forecasts thus the observed frequency is 53.3%, 5% less than the GEPS example. The mean for all forecast hours for this bin results in a 42.8% observed frequency, approximately 13% less than the GEPS. The next four probability bins where probabilities exist are underforecast but show skill and the relative sample sizes range from one in the 31-40% bin, two in both the 11-20% and 41-50% bins, and three in the 21-30% bin. This is roughly half the size of the 1-10% bin compensating for some of the skill lost by that bin's contribution. The continued but less drastic trend of missing events and underforecasting the wind speed produces a better reliability while more forecast samples that verify away from climatology produce an increased resolution. However, the MEPS20 resolution does not increase enough to overcome the underforecasting bias. This is why the MEPS20 performs better than the GEPS but does not have a BSS that deviates far from 0.

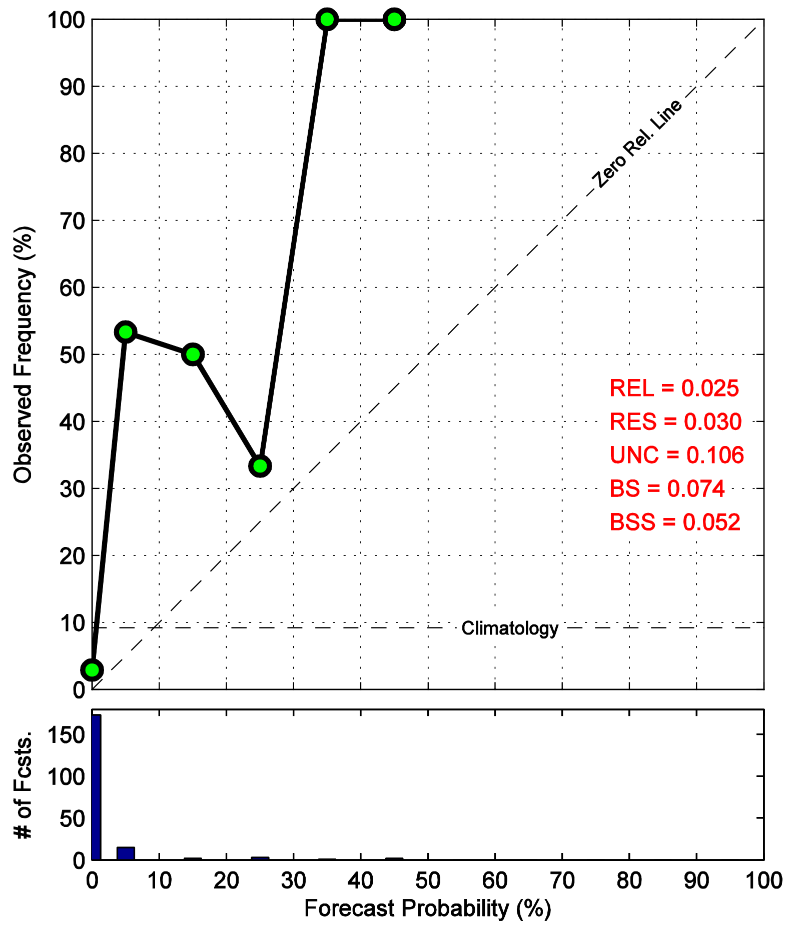

Figure 20: MEPS20 21hr Winds > 25kts reliability diagram for Fort Greely from April-October 2013 indicating that occurrences are missed and wind speeds are underforecast.

Considering the increased grid resolution of MEPS4, it is noted that this ensemble rarely misses any events, as the average misses for all the forecast hours is 1.2 per forecast hour. Also, winds are underforecast but less severely than the other two EPS. For the 1-10% bin, the average observed frequency is 14.8% which is 43.2% less than GEPS and 28% less than MEPS20. Also, the average observed frequency is only 5% over its bin probability max resulting in only a slight underforecasting bias. Figure 21 provides an illustration of these trends for forecast hour 9. In this example no events are missed. For the second heaviest weighted bin, 1-10%, only three out of the 20 forecasts verified thus the observed frequency is 15% as depicted in Figure 21.

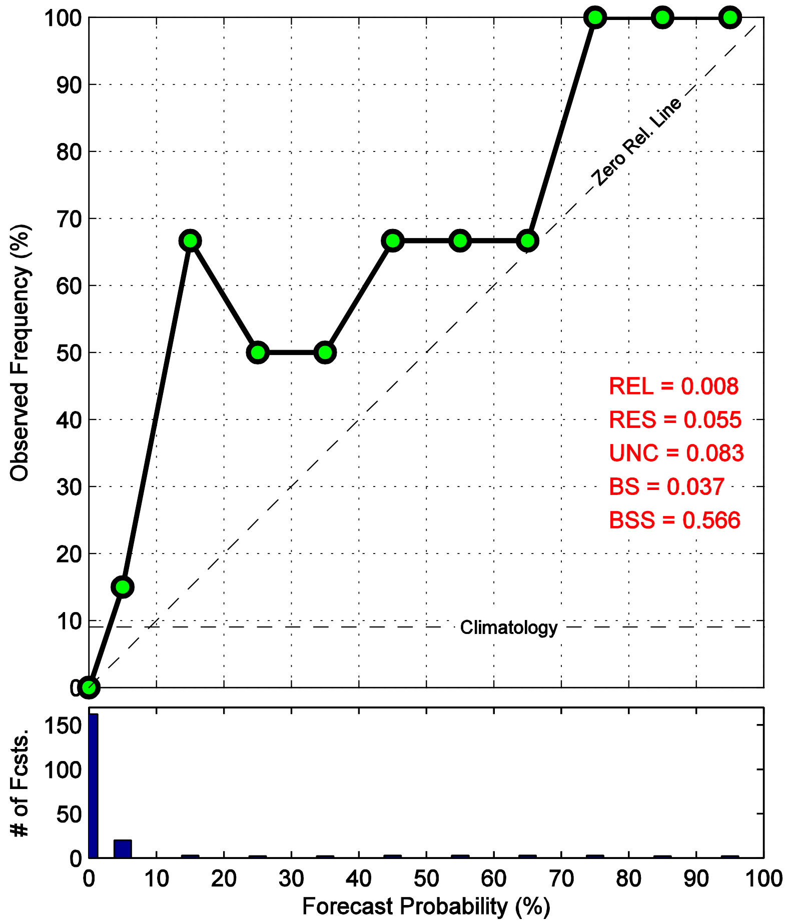

Figure 21: MEPS4 9hr Winds > 25kts reliability diagram for Fort Greely from April-October 2013 indicating that occurrences are not missed and wind speeds are only slightly underforecast.

Also, bins that previously did not have any samples are now populated and verify providing forecasts that strongly deviate from climatology. This allows for increased resolution while the reliability is fairly low due to less of an underforecasting bias. Although the other bin forecast sample sizes are small, all with only two or three forecasts, they total 23. This is larger than the 1-10% bin adding an appreciable amount of skill. Due to these factors, MEPS4 produces all positive BSSs.

4.3.2 Winds > 35kts

Table 8 details the results for winds > 35kts. Only three out of the ten sites had enough occurrences to evaluate and two of the sites, Dyess and Kadena AB, did not have enough hourly occurrences to evaluate MEPS4. The average positive skill shows that for winds > 35kts, increasing horizontal resolution equates to a more positive and better BSS regardless of location, as is the case for winds > 25kts.

Table 8: Winds > 35kts Positive Skill Duration, Skillful Percentage of Forecast and Average Positive Skill. Blanks indicate insufficient occurrence sample size.

Site	EPS	Forecast Hours of Positive Skill	Skillful % of Forecast	Avg Positive Skill
KDYS	GEPS	6-18	7.5	0.115
	MEPS20	6-72	50	0.037
	MEPS4	—	—	—
PABI	GEPS	0	0	0
	MEPS20	0	0	0
	MEPS4	6-72	100	0.255
RODN	GEPS	6-204	85	0.078
	MEPS20	6-144	100	0.165
	MEPS4	—	—	—

The only exception is the results for Dyess AFB (KDYS). Because the GEPS was only skillful for forecasts at 6, 12 and 18 hours, the average is based on only three numbers and results in a higher average than the MEPS20. The MEPS20 is a skillful forecast out to 72 hours; however, BSSs are close to 0 thus adding very little skill over climatology. None of the sites tested have a large enough occurrence sample size to obtain useful results for all three EPS. Since Fort Greely (PABI) winds > 25kts have already been investigated it seems appropriate to assess an alternate site Kadena AB (RODN). Looking at the reliability diagrams for GEPS and MEPS20 and all the forecast hours there are some similarities to the Fort Greely winds > 25kts results. Figure 22 for forecast hour 48 demonstrates missing events and underforecasting.

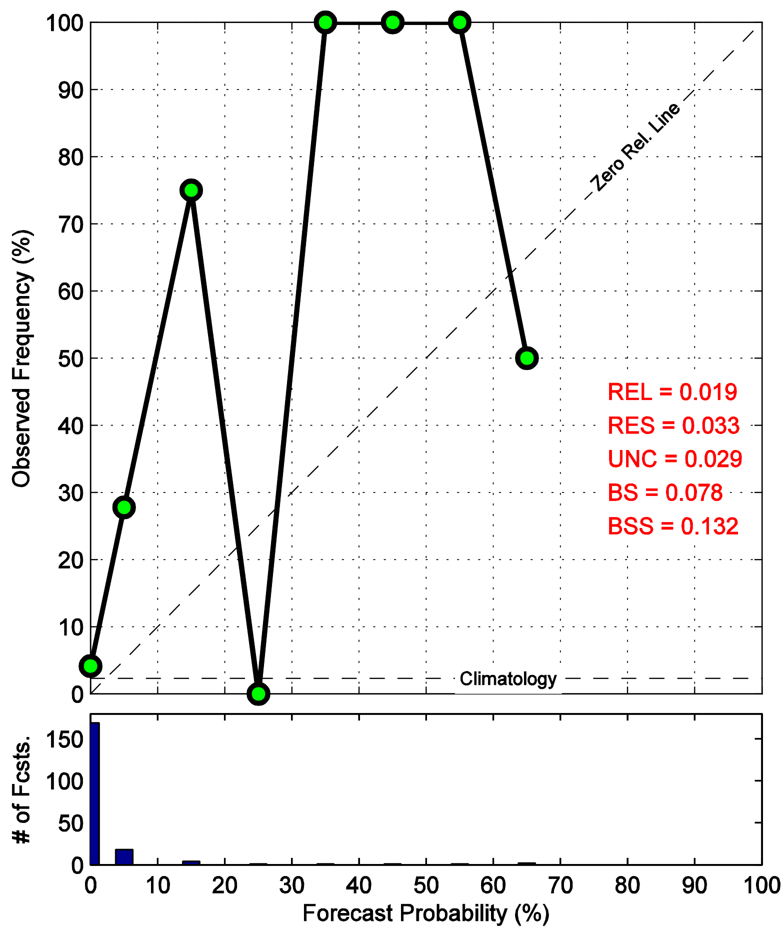

Figure 22: GEPS 48hr Winds > 35kts reliability diagram for Kadena AB from April-October 2013 indicating that occurrences are missed and wind speeds are underforecast.

For GEPS, averaging all missed events for all the forecast hours results in six events missed per forecast hour. Calculating an average of forecast hour observed frequencies for the 0% probability bin results in 4.3%. Also, events in other probability bins are slightly underforecast for most forecast hours. The main difference between this example (Figure 22) and the > 25kts winds investigated at Fort Greely (Figure 19) is that the climatologically probability for winds > 35kts at Kadena AB is substantially lower, less than 5% for all forecast hours. Consequently, the 1-10% probability bin falls into the area of skill. With this trend present in most forecast hours, the reliability values in Figure 23 will still be relatively high, however, enough events are forecasted from the low climatology values and verify to produce a relatively high resolution value leading to a positive BSS for 204 hours, 85% of the forecast duration.

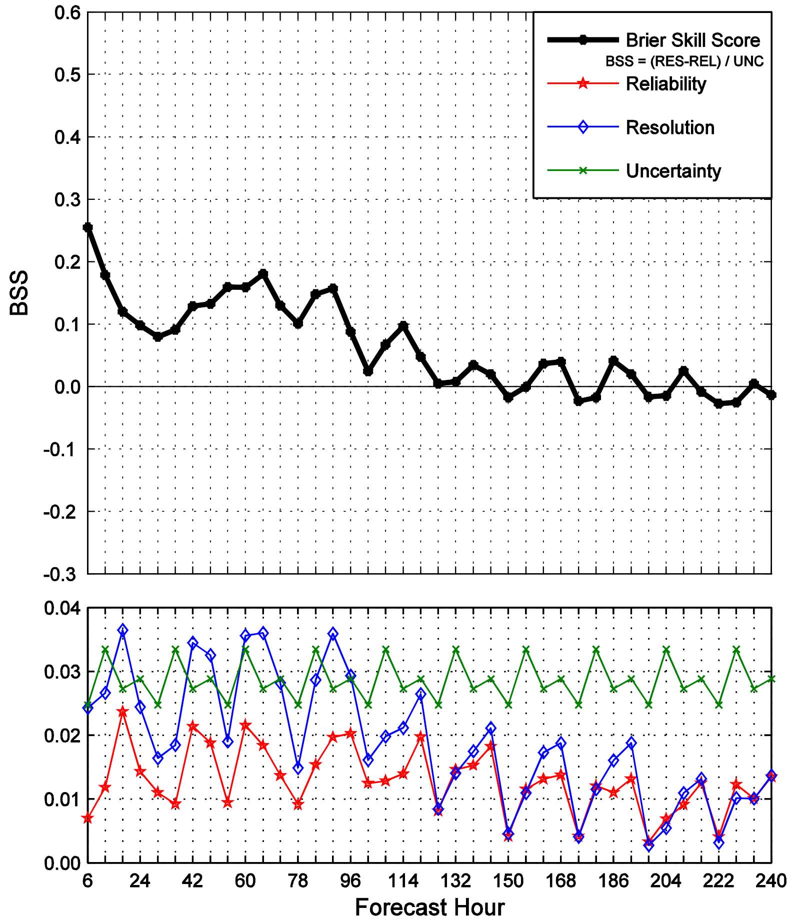

Figure 23: GEPS Winds > 35kts BSS for all forecast hours at Kadena AB from April-October 2013 indicating a positive BSS for most forecast hours.

With increased resolution, MEPS20 misses fewer events with an average of 4.7 misses per forecast hour. Also, the underforecasting is less prevalent in the second heaviest weighted bin with an average observed frequency of 10.8%. This forces the reliability to a lower value while resolution increases slightly with more forecasts away from climatology verifying, thus the MEPS20's BSS is higher for the majority of the forecast duration. Figure 24 illustrates the MEPS20's BSS, reliability and resolution to allow for visual comparison to previously mentioned GEPS results.

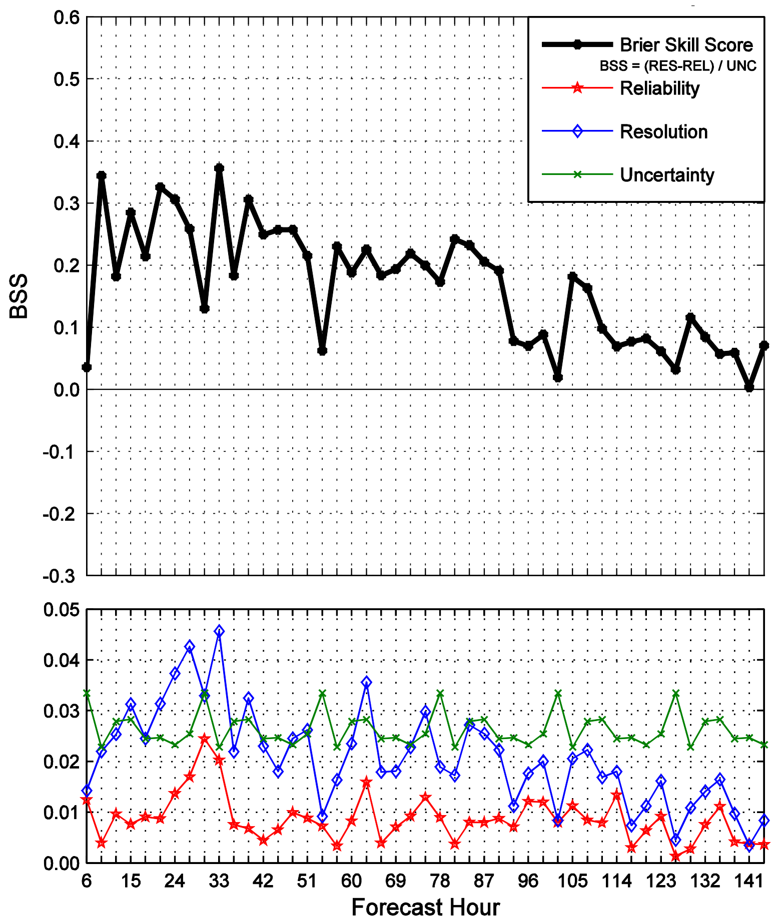

Figure 24: MEPS20 Winds > 35kts BSS for all forecast hours at Kadena AB from April-October 2013 indicating a positive BSS for most forecast hours.

4.4 Precipitation

4.4.1 Precipitation > 0.1in and > 0.05in in 6 hours

Similar to the wind results, the average positive skill detailed in Table 9 indicates that for all three EPS, increasing horizontal resolution yields a more positive and better BSS for six out of the eight sites. Also, for six out of the eight sites there is an increase in the number of forecast hours of positive BSS duration with increasing resolution from GEPS down to MEPS4.

Table 9: Precipitation Positive Skill Duration, Skillful Percentage of Forecast and Average Positive Skill. Blanks indicate insufficient occurrence sample size.

Site	EPS	Forecast Hours of Positive Skill	Skillful % of Forecast	Avg Positive Skill
ETAR	GEPS	6-234	95	0.145
	MEPS20	6-144	100	0.170
	MEPS4	6-72	100	0.346
KDYS	GEPS	6-204	52.5	0.089
	MEPS20	6-114	77.3	0.170
	MEPS4	6-72	100	0.250
KLRF	GEPS	6-216	90	0.139
	MEPS20	6-126	85.8	0.154
	MEPS4	6-72	100	0.346
KTIK	GEPS	6-240	100	0.164
	MEPS20	6-144	100	0.173
	MEPS4	6-72	100	0.346
KXMR	GEPS	0	0	0
	MEPS20	36	23.4	0.091
	MEPS4	10-29, 33-50, 65-72	56.7	0.121
PABI	GEPS	6-162	67.5	0.044
	MEPS20	6	2.1	0.018
	MEPS4	6-55, 63-72	79.1	0.123
RKJK	GEPS	6-222	92.5	0.157
	MEPS20	6-111	76.6	0.230
	MEPS4	6-72	100	0.374
RODN	GEPS	6-180	75	0.199
	MEPS20	6-123	85.1	0.110
	MEPS4	6-72	100	0.209

These increased BSS forecast durations and average values can be attributed to two factors. First, precipitation processes predominately occur at the microscale and mesoscale level, thus precipitation is better resolved by higher resolution models. Parameterization schemes are employed to mitigate a model's lack of resolution but these schemes suffer from their own pitfalls. The GEPS will explicitly miss many of these smaller scale processes only capturing larger synoptic features like frontal boundaries while relying on inherent schemes to compensate for smaller scale processes. MEPS20's 20km resolution will pick up on many of the mesoscale processes like dry lines, squall lines and others while the MEPS4's 4km resolution will resolve most mesoscale processes and some microscale processes without parameterization. The second reason for the increased positive BSS durations and values is that the GEPS and MEPS20 create probabilities for precipitation > 0.1in in 6 hours while the MEPS4 generates probabilities for precipitation > 0.05in in 6 hours. This makes it slightly easier for the MEPS4 precipitation probabilities to verify leading to a longer positive BSS duration and higher average positive BSS as detailed in Table 9. For all three EPS, precipitation is the easiest event to forecast and verify. While both precipitation forecast thresholds are more than a typical brief rain shower, the amounts are not considered significant. Also, both the MEPS20 and MEPS4 use 6 hours leading up to the forecast hour to verify the events while the other forecast parameters utilize only 1 or 3 hours, depending on the EPS.

4.4.2 Synoptic Forcing vs. Convective Heating

Based on the data represented in Table 9 it is evident that all three EPS perform better at resolving precipitation for locations that experience rainshowers and thunderstorms predominately associated with frontal lift versus rainshowers and thunderstorms that typically develop due to daytime heating and small scale lifting mechanisms. At Cape Canaveral AFS (KXMR) the majority of rain showers and thunderstorms develop due to lift associated with daytime convective heating and/or daily sea breezes. Tables 7 and 10 exhibit that each EPS does worse than the respective climatology for both lightning and precipitation occurrence. The poor BSS for lightning is due to the high lightning climatology percentages coupled with all three EPS overforecasting lightning as previously mentioned in the lightning results section. Similarly, precipitation is overforecast for the majority of the forecast hours yielding a negative BSS in all EPS. Figure 25, for forecast hour 30, highlights an example of this trend for the GEPS.

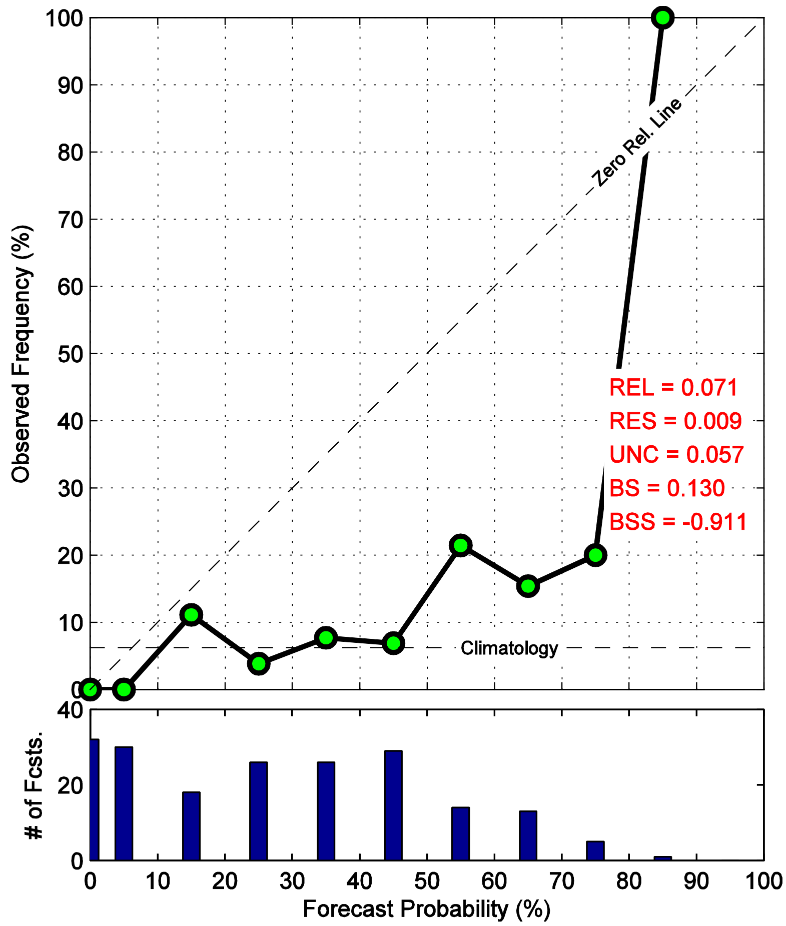

Figure 25: GEPS 30hr Precipitation > 0.1in reliability diagram for Cape Canaveral AFS from April-October 2013 indicating that precipitation is overforecast.

Different from the other parameters discussed thus far, the precipitation forecast contributions are not as heavily weighted towards the 0% and 1-10% bins. Forecasts are fairly evenly distributed into other higher probability forecast bins. If forecasts in these higher probability bins verify more frequently, the resolution would increase; however, most events do not verify thus the resolution is relatively low since most observed frequencies are near climatology. This overforecasting bias also causes the reliability to increase due to the observed frequency moving further away from the zero reliability line as the forecast probabilities increase. The resulting high reliability and low resolution cause the BSS to become negative for virtually all the forecast hours as indicated in Figure 26.

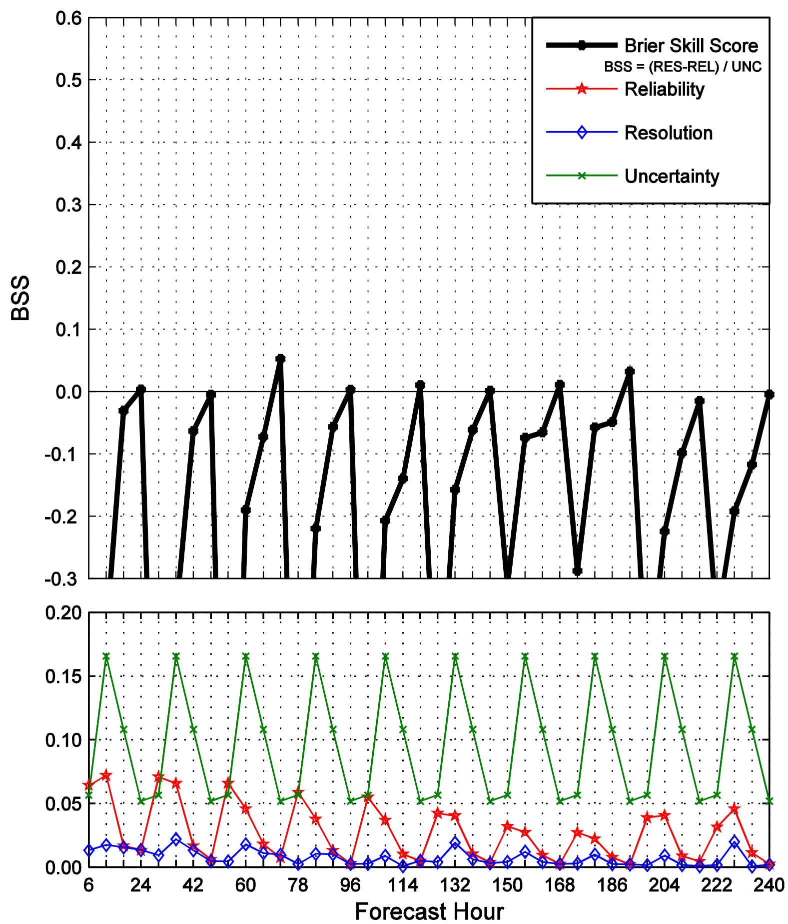

Figure 26: GEPS Precipitation > 0.1in BSS for all forecast hours at Cape Canaveral AFS from April-October 2013 indicating a negative BSS for most forecast hours.

MEPS20 suffers from a similar overforecast bias, but not as prevalent in all the forecast hours. Fewer events are forecast in the higher probability bins thus alleviating some of the contributions from these bins as shown in forecast hour 75, Figure 27.

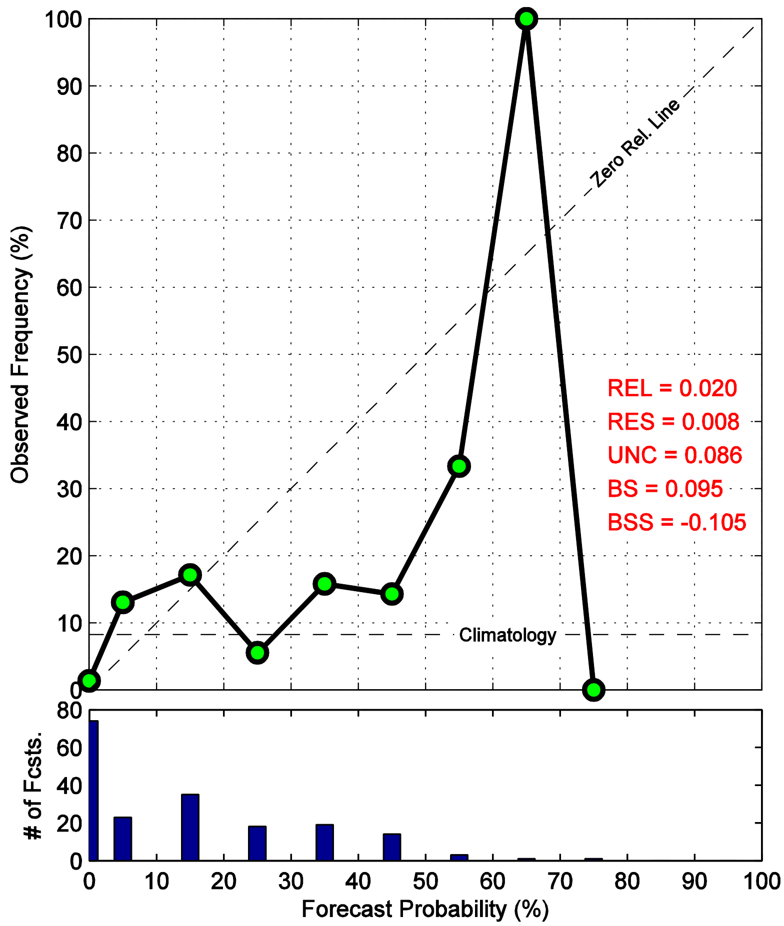

Figure 27: MEPS20 75hr Precipitation > 0.1in reliability diagram for Cape Canaveral AFS from April-October 2013 indicating that precipitation is overforecast.

Here, the 0% probability bin provides the largest contribution to the BSS and its components. The second largest contribution is from the 11-20% bin and falls close to the zero reliability line. Other significant weighted bins fall near the climatology line providing little to no resolution. Overall, compared to the GEPS, the reliability decreases and the resolution is similar for most hours as indicated in Figure 28. This translates to the first 36 forecast hours possessing a positive BSS duration using the previously discussed technique to smooth the trend.

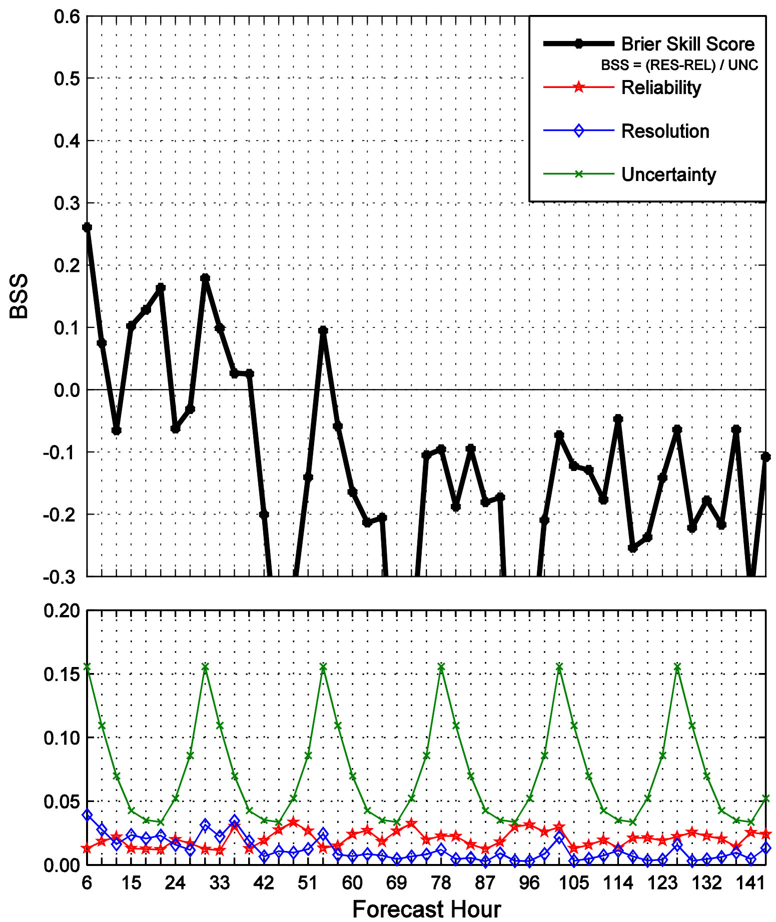

Figure 28: MEPS20 Precipitation > 0.1in BSS for all forecast hours at Cape Canaveral AFS from April-October 2013 indicating a positive BSS for the initial portion of the forecast.

For MEPS4, no biases are noted when reviewing the reliability diagrams. For most forecast hours the majority of the probability bins closely parallel the zero reliability line as shown in Figure 29 for forecast hour 17. This trend allows the EPS to remain positive for the majority (56.7%) of the forecast hours with a better average positive skill than the other two EPS tested.

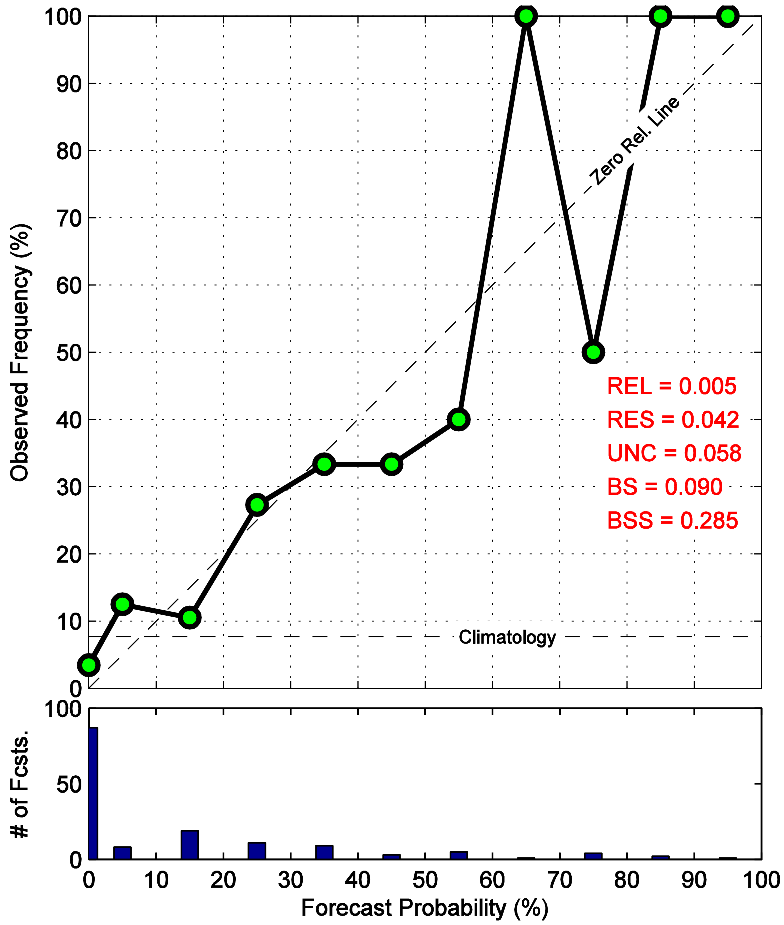

Figure 29: MEPS4 17hr Precipitation > 0.05in reliability diagram for Cape Canaveral AFS from April-October 2013 indicating that probabilities closely match the zero reliability line.

Kunsan AB located on the western side of the South Korean peninsula, bordered by the West Sea, experiences precipitation events predominately from migratory low pressure systems that traverse to the north over Manchuria or across the West Sea. Because these events are predominately frontal in nature, the GEPS is able to resolve a considerable amount of the precipitation correctly producing positive skill 92.5% of the time. MEPS20 and MEPS4 still produce a better average positive skill due to increased resolution but are fairly comparable showing that all three EPS resolve frontal precipitation well as exhibited in Figure 30. Reliability diagrams indicate no significant biases for the three EPS at this location.

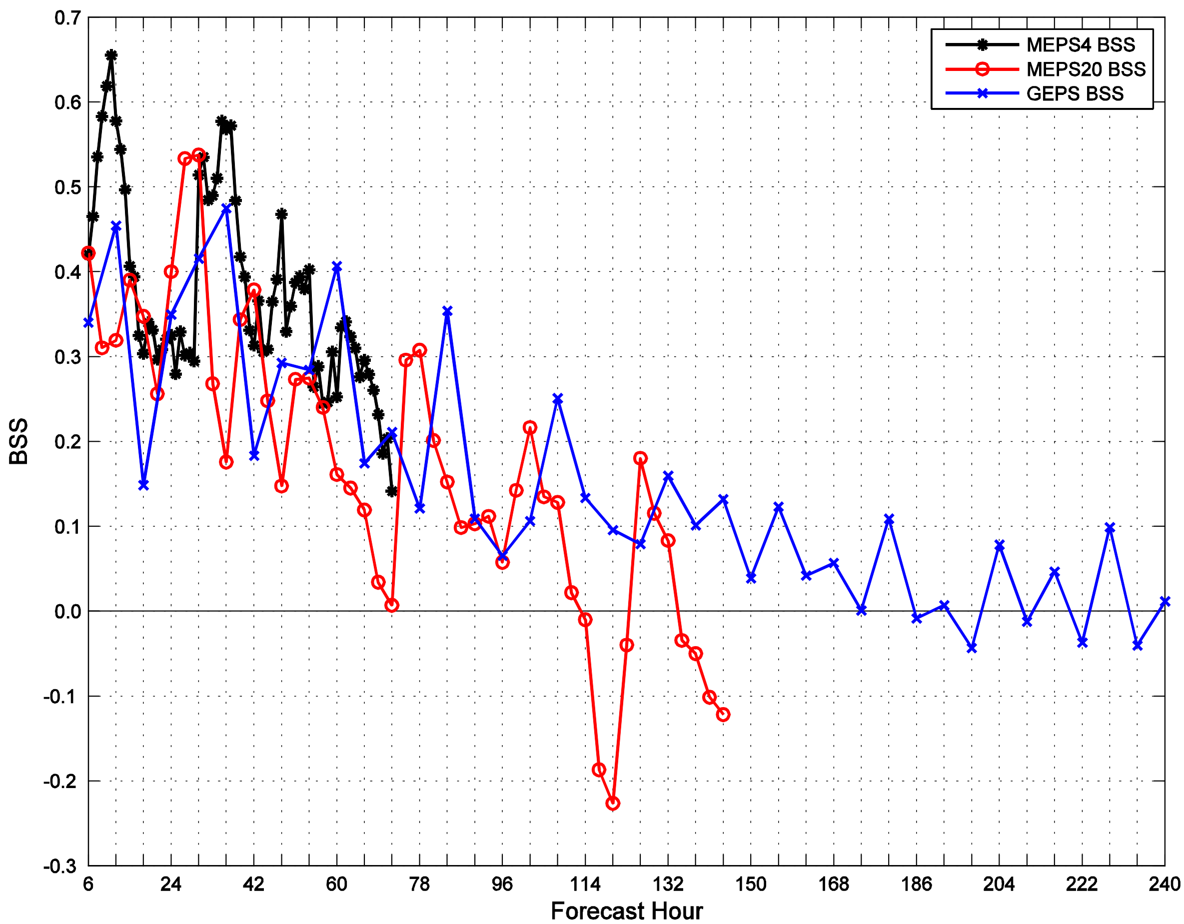

Figure 30: Comparison of MEPS4, MEPS20 and GEPS BSS for Kunsan AB Precipitation. MEPS4 shown in black, MEPS20 is shown in red and GEPS is shown in blue.

Based on results from Cape Canaveral AFS it is apparent that the parameterization schemes in the GEPS and MEP20 struggle with resolving precipitation from daytime, convective heating. Frontal precipitation, on the other hand, at Kunsan AB is resolved well by the parameterization schemes used in all three EPS.

4.5 Tropical Cyclone EPS Skill

During October 2013, three tropical cyclones passed within approximately 222km of Kadena AB, as shown in Figure 31, providing the opportunity to investigate EPS performance for winds and precipitation during tropical cyclone impacts.

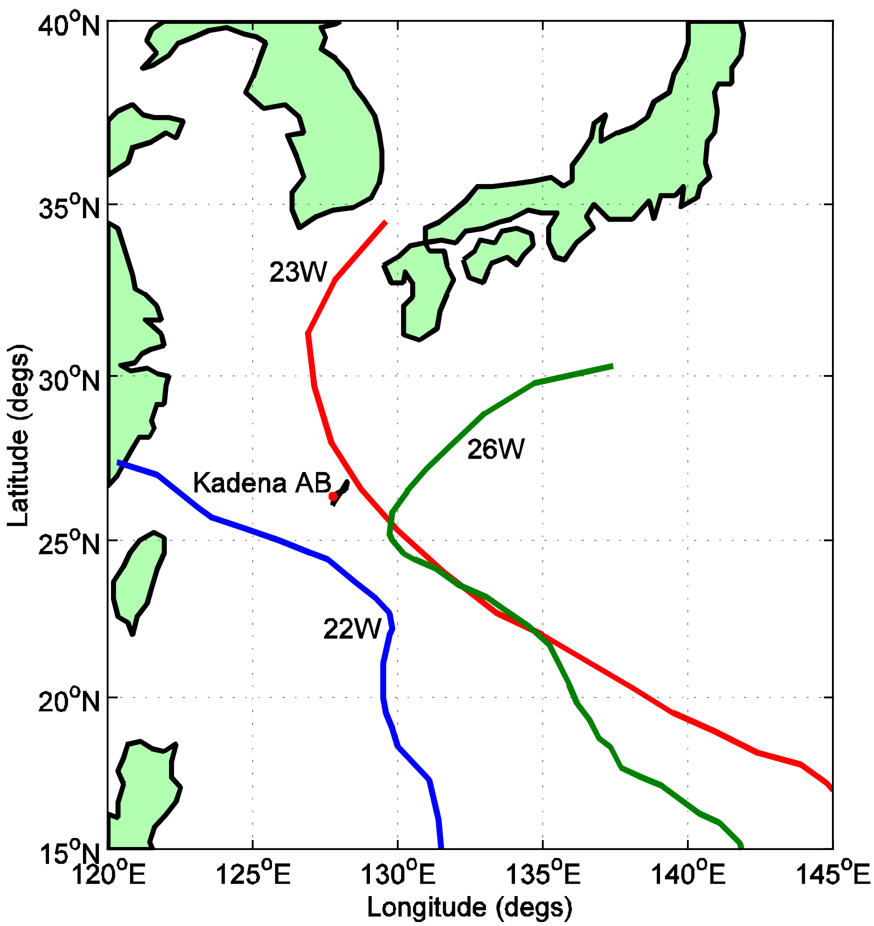

Figure 31: Graphic of tropical cyclones 22W, 23W and 26W passing within 222km of Kadena AB during the month of October.

Sample sizes for tropical cyclone events are small allowing for any bin to considerably affect the BSS. Also, due to these small samples, only two parameters possess a sufficient sample size to compare all three EPS - winds > 25kt, precipitation > 0.1in in 6 hours for GEPS and MEPS20, and precipitation > 0.05in in 6 hours for MEPS4.

4.5.1 Winds > 25kts

A comparison of all three EPS is provided in Figure 32. For MEPS4, it is evident that the BSS remains highly positive for the entire forecast duration with the exception of the two outliers at hours 29 and 53. Without these outliers, values range from approximately 0.5 to 0.9.

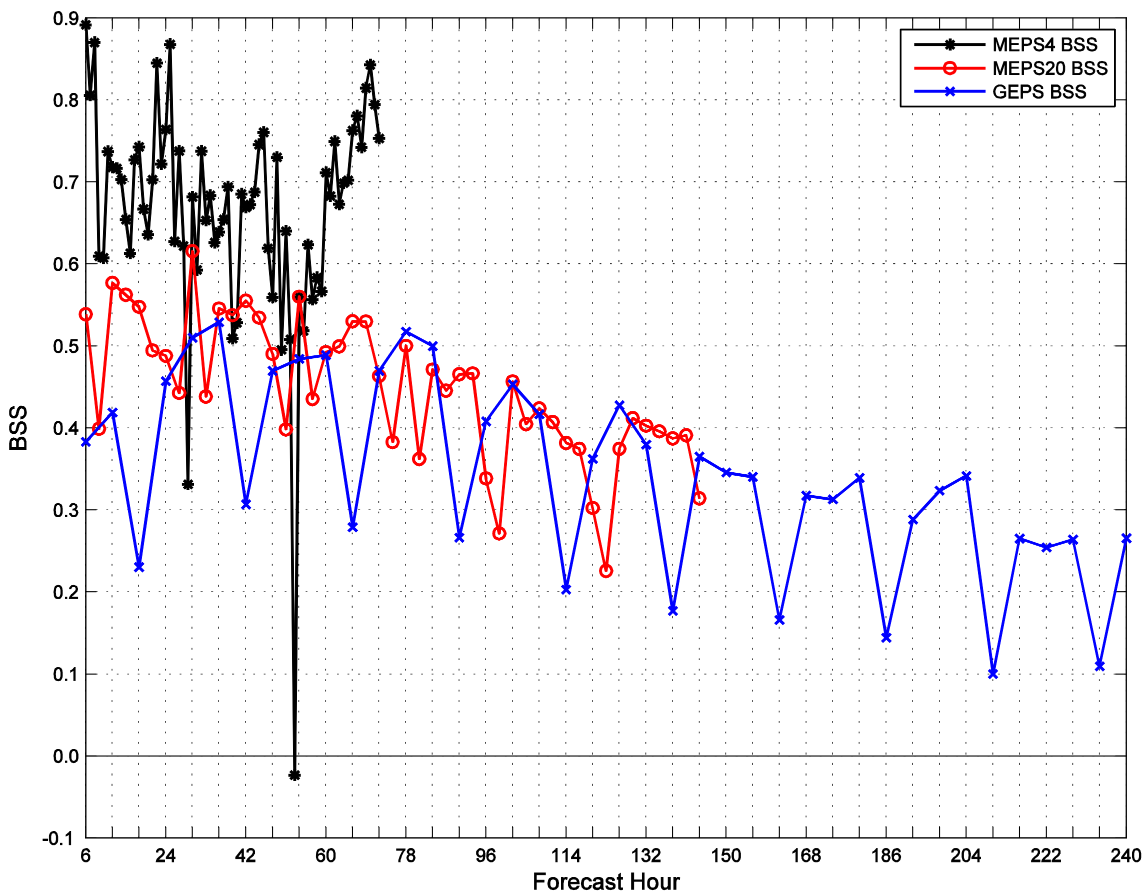

Figure 32: Comparison of MEPS4, MEPS20 and GEPS BSS for Kadena AB Winds > 25kts. MEPS4 is shown in black, MEPS20 is shown in red, and GEPS is shown in blue.

MEPS20's BSS trends downward and stays positive during the forecast with values ranging from approximately 0.22 to 0.62. Lastly, GEPS's BSS trends downward as well remaining positive for the forecast duration with values ranging from approximately 0.1 to 0.53. Upon review of the wind test data for the three tropical cyclone passes, it is evident that horizontal resolution differences play a significant role in increasing the BSS for winds and that all three EPS perform well.

4.5.2 Precipitation > 0.1in and > 0.05in in 6 hours

A comparison of all three EPS is provided in Figure 33.

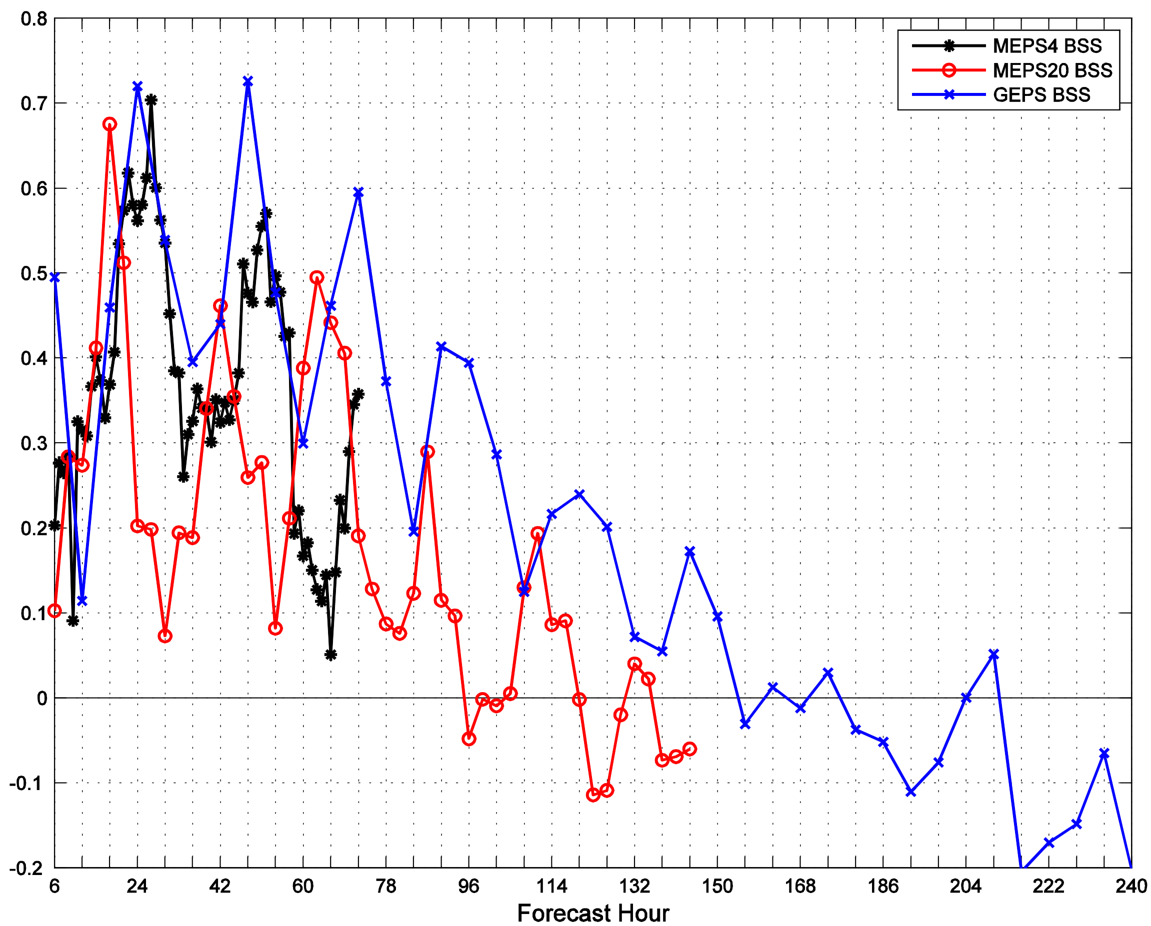

Figure 33: Comparison of MEPS4, MEPS20 and GEPS BSS for Kadena AB Precipitation. MEPS4 is shown in black, MEPS20 is shown in red, and GEPS is shown in blue.

Unlike the whole sample results for Kadena AB showing a clear increase in average positive skill with increasing resolution for winds, no noticeable changes in skill are noted in Figure 33 for precipitation. In fact, the GEPS performed slightly better than both regional EPS. It is possible that both of these EPS, while able to resolve typically small-scale convection at Kadena AB during the year, they poorly resolve large-scale forcing from tropical cyclones.

IV. Conclusion

5.1 Summary of Results

Ensemble modeling has begun to revolutionize weather forecasting. By characterizing uncertainty through a group of ensemble members, the probability of an event occurring is generated providing users an understanding as to how well models are in agreement when forecasting a particular parameter. Probabilities of parameter occurrence provide more information than a simple “yes” or “no” deterministic result. While more descriptive than a deterministic model, ensembles still possess pitfalls as the atmosphere is not absolutely resolved regardless of model configuration.

This study exploits how each of AFWA’s EPS - GEPS, MEPS20 and MEPS4 - performs over one convective season ranging from April to October 2013. For the six parameters tested, reliability diagrams and BSS time series were constructed. Two parameters, 50kts winds and precipitation > 2in in 12 hours, proved too infrequent of an event to test, while others generated useful metrics as detailed in Chapter 4.

Lightning for GEPS, MEPS20 and MEPS4 for all locations and most forecast hours suffered from substantial overforecasting bias leading to poor reliability and resolution outcomes which yielded marginally positive BSSs during the day and negative BSSs at night.

For both winds > 25kts and > 35kts, horizontal resolution plays a significant role in resolving terrain effects which helps resolve wind speed. MEPS4 was found to provide the best average BSS for winds > 25kts and > 35kts for all locations and the best skillful percent of the forecast for 4 of the 5 sites where each EPS had a sufficient sample of forecasts. Where MEPS4 sample sizes are insufficient, MEPS20 outperformed the GEPS.

Precipitation is also better resolved as horizontal resolution increases with a smaller grid size. The MEPS4's superior performance in forecasting precipitation is most likely due to its explicit resolution outperforming the cumulus parameterization schemes used to resolve convection in the GEPS and MEPS20. Likewise, the MEPS20 performs slightly better than the GEPS due to its better explicit resolution of mesoscale convective processes greater that or equal to 20km in horizontal extent. BSS performance differences were noted in Chapter 4 when investigating the results for Cape Canaveral AFS and Kunsan AB. At Cape Canaveral the majority of convection and resulting precipitation is generated by localized heating and small scale circulations like land and sea breezes. These results show that the cumulus parameterizations in both the GEPS and MEPS20 fail to resolve the small scale convection sufficiently enough to provide skillful results, while the explicit resolution of convection in MEPS4 is slightly better with 7 forecast hours of positive skill. At Kunsan AB much of the precipitation that occurs is a result of large scale lift from transient fronts. This mechanism for precipitation is resolved well by the cumulus parameterizations schemes in the GEPS and MEPS20 as well as the explicit resolution in MEPS4. Based on these differences it is hypothesized that until model grid spacing is reduced to the actual size of most convective precipitation cells, small scale events like the ones observed at Cape Canaveral will continue to be poorly resolved at larger grid scales while large scale events like frontal induced precipitation are resolved well by cumulus parameterization and explicit resolution.

For the three tropical cyclones that passed near Kadena AB, wind results continued to show that increased model resolution leads to a better forecast; however, tropical precipitation is resolved slightly better by the GEPS. This is potentially due to both regional EPS, MEPS20 and MEPS4, struggling to resolve precipitation processes associated with large scale forcing from tropical cyclones.

Lastly, it is worth noting that a diurnal trend in the BSS was evident in many of the figures presented. GEPS showed the strongest diurnal trend for the parameters tested with MEPS20 showing less of a trend and MEPS4 showing the least. One potential reason for the more obvious GEPS diurnal trending lies in the 6 hour forecast probability interval. For example, if thunderstorms rarely occur at a location overnight, GEPS may have a 2% probability for lightning occurrence for each hour. Combining those probabilities to create a 6 hour forecast probability equates to a much higher probability for an event rarely occurring, causing poor overnight BSSs. The probabilities for each of the MEPS20 and MEPS4 forecast intervals, 3 hours and 1 hour respectively are overall less, thus the decrease in BSS in either case is less pronounced or nonexistent.

5.2 Recommendations and Future Research

Since this research only tested one convective season, six months over spring and summer, it would be beneficial to bolster the forecast sample size to include multiple years and all seasons. Doing so could further validate these findings along with the potential to discover other EPS trends. With AFWA currently producing PEP bulletins for roughly 10,000 locations worldwide, these EPS probabilities are not achieved due to data storage limitations (Kuchera, 2013). Future validation of point locations using

AFWA’s EPS will require that locations be selected and a daily routine of archiving this data be implemented.

Because many of the meteorological events that the Air Force forecasts do not occur often, it should be noted that the majority of the results showed more forecast samples in the smaller forecast probability bins e.g. 1-10% and 10-20%. To provide further detail on EPS performance it would be beneficial to create smaller bin widths for these lower probability bins.

Because ensemble modeling is becoming more prevalent in the civilian sector and DoD, EPS are being updated at a higher cadence as discoveries are being realized relative to ensemble performance. In just seven months, the 10 ensemble member MEPS suite changed member parameters four times as noted in Appendix D. To truly understand how well these four different suites perform, each would need to be tested over a longer time period than existed during this study. This could result in a greater understanding of which ensemble model suite performs best for different regions of the world.

Appendix A: AFWA Lightning Algorithms

The basic regression equation is applied to MEPS20 when convective available potential energy (CAPE) and accumulated precipitation (AP) are greater than 0.

$$LTG\ prob = 0.13\ \times log[(CAPE \times AP) + 0.7)] + 0.05$$

AP is adjusted using precipitable water (PW) values because models often produce showery precipitation that does not use the instability in very moist environments.

$$AP = AP - \left(\frac{PW}{1000}\right)$$

If AP is less than 0.01, then

$$LTG\ prob = 0.025\ \times\ log\left[\left(\frac{CAPE}{CIN + 100}\right) + 0.31\right] + 0.03$$

If there is no CAPE but the model atmosphere is on the verge of becoming unstable, lightning can occur. This typically happens when heavy precipitation stabilizes the model atmosphere. Therefore, to be unstable there must be a positive lifted index (LI).

$$LI = LI + 4$$

If the LI is less than 0, then the LI is set to 0. If the CAPE is less than 0, then

$$LTG\ prob = 0.2\ \times\ (\mathrm{LI}\ \times\ \mathrm{AP})^{0.5}$$

If the PW value is low, then graupel can not form, which starts the charging process thus the probability of lightning is reduced. If PW is less than 20,

$$LTG\ prob = LTG\ prob\ \times\ \left(\frac{PW}{20}\right)$$

The regression equation can only be as skillful as 95% thus the probabilities that are above 95% are set to 95%.

For MEPS4, the cumulus parameterization is turned off and thunderstorms are resolved explicitly. Petersen et al (2005) and McCaul et al (2009) showed that the incorporation of a graupel flux is a more accurate way to predict lightning. Because of the computational expense involved in predicting graupel amounts, most of AFWA's MEPS ensemble members do not predict graupel, but instead used total cloud ice. MEPS20 convective parameterization schemes incorporate total ice content, however, the explicit method used by the MEPS4 does not, therefore the following equation is used to incorporate total cloud ice content in the MEPS4.

$$LTG\ prob = 0.076\ \times\ (total\ cloud\ ice - 7.5\ )$$

Appendix B: Combined Region Lightning Probability

To find the probability of lightning, *P*, in a combined region of *n* smaller regions, each with a probability of lightning, *p*, use:

$$P(p,n) = \sum_{r=1}^{n} \frac{n!}{r!\,(n-r)!} p^r (1-p)^{(n-r)}$$

In this expression, the $\frac{n!}{r!(n-r)!}$ term is the number of combinations of *n* objects taken *r* at a time. This term gives us the number of combinations of *r* out of *n* areas containing lightning. The $p^r(1-p)^{(n-r)}$ term gives the probability of occurrence of a particular combination, with p^r the contribution to the probability of areas with lightning and $(1-p)^{(n-r)}$ the contribution of areas without lightning. The summation accumulates the probabilities of outcomes with 1, 2, ..., *n* areas having lightning.

Examples:

1. PEP bulletin probability of 20% (*p* = 0.20) for two areas combined (*n = 2*):

$$\begin{aligned} P(0.20,2) &= \sum_{r=1}^{2} \frac{2!}{r!\,(2-r)!} (0.20)^r\,(1-0.20)^{(2-r)} \\ &= \frac{2!}{1!\,(2-1)!}(0.20)^1(1-0.20)^{(2-1)} + \frac{2!}{2!\,(2-2)!}(0.20)^2(1-0.20)^{(2-2)} \\ &= 2(0.20)(0.80) + 1(0.04)(1) \\ &= 0.36 \end{aligned}$$

2. PEP bulletin probability of 10% (*p* = 0.10) for eight areas combined (*n* = 8):

$$P(0.10,8) = \sum_{r=1}^{8} \frac{8!}{r!\,(8-r)!} (0.10)^r\,(1-0.10)^{(8-r)}$$

$$= \frac{8!}{1!\,(8-1)!}(0.10)^1(1-0.10)^{(8-1)} + \cdots$$

$$+ \frac{8!}{8!\,(8-8)!}(0.10)^8(1-0.10)^{(8-8)}$$

$$= 8(0.10)(0.90)^7 + \cdots + 1(0.10)^8(1)$$

$$= 0.57$$

Appendix C: Acronym List

AB – Air Base

AFB – Air Force Base

AFS – Air Force Station

AFWA – Air Force Weather Agency

AFW-WEBS – Air Force Weather Web Services

ALADIN-LAEF - Aire Limitée Adaptation Dynamique Développement International-Limited Area Ensemble Forecasting

BS – Brier Score

BSS – Brier Skill Score

CMC – Canadian Meteorological Centre

DoD – Department of Defense

ECMWF – European Centre for Medium-Range Weather Forecasts

EPS – Ensemble Prediction Center

FNMOC – Fleet Numerical Meteorology and Oceanography Center

GEFS – Global Ensemble Forecast System

GFS – Global Forecast System

GEM – Global Ensemble Model

GEPS – Global Ensemble Prediction System

MATLAB – Matrix Laboratory

MEPS20 – 20km Mesoscale Ensemble Prediction System

MEPS4 – 4km Mesoscale Ensemble Prediction System

METAR – Aerodome Routine Meteorological Report

NAS – Naval Air Station

NCEP – National Centers for Environmental Prediction

NOGAPS – Navy Operational Global Atmospheric Prediction System

NWP – Numerical Weather Prediction

ORM – Operational Risk Management

OWS – Operational Weather Squadron

PDF – Probability Density Function

PEP – Point Ensemble Probability

SPECI – Aerodrome Special Meteorological Report

UKMO – United Kingdom Met Office

UM – Unified Model

WRF – Weather Research and Forecasting

WXG – Weather Group

Appendix D: AFWA MEPS Member Configuration

The ensemble members for both MEPS20 and MEPS4 are listed below in Tables B1-B4. Note that for MEPS4 the convective parameterization, C, is turned off as the model explicitly resolves convection. All WRF-NMM dynamics and physics options can be found in the User's Guide for the NMM Core of the Weather Research and Forecast (WRF) Modeling System Version 3 Chapter 5.

Table B1: MEPS first configuration during research sample (Kuchera, 2013).

M	LIC	LUT	IC	LBC	SW	LW	LSM	MP	H	CCN	PBL	SL	C
1	LIS	n/a	UM	UM	n/a	n/a	2	4	n/a	n/a	1	1	1
2	LIS	n/a	GFS	GFS	n/a	n/a	2	10	n/a	n/a	8	2	2
3	LIS	n/a	GEM	GEM	n/a	n/a	2	16	n/a	n/a	1	1	5
4	LIS	n/a	GEM	GEM	n/a	n/a	2	5	n/a	n/a	8	1	1
5	LIS	n/a	UM	UM	n/a	n/a	2	16	n/a	n/a	7	1	2
6	LIS	n/a	GFS	GFS	n/a	n/a	2	8	n/a	n/a	7	1	5
7	LIS	n/a	GEM	GEM	n/a	n/a	2	10	n/a	n/a	1	1	2
8	LIS	n/a	GFS	GFS	n/a	n/a	2	5	n/a	n/a	1	1	6
9	LIS	n/a	UM	UM	n/a	n/a	2	8	n/a	n/a	7	1	5
10	LIS	n/a	GFS	GFS	n/a	n/a	2	4	n/a	n/a	7	1	6

Table B2: MEPS second configuration during research sample (Kuchera, 2013).

M	LIC	LUT	IC	LBC	SW	LW	LSM	MP	H	CCN	PBL	SL	C
1	LIS	10	UM	UM	1	1	2	4	1	n/a	1	1	1
2	LIS	2	GFS	GFS	1	1	2	10	1	1E+9	8	2	2
3	LIS	2	GEM	GEM	1	1	2	16	0	1E+9	1	1	5
4	LIS	AFWA	GEM	GEM	1	1	2	5	n/a	n/a	8	1	1
5	LIS	5	UM	UM	1	1	2	16	1	1E+8	7	1	2
6	LIS	6	GFS	GFS	1	1	2	8	n/a	n/a	7	1	5
7	LIS	7	GEM	GEM	1	1	2	10	0	1E+8	1	1	2
8	LIS	8	GFS	GFS	1	1	2	5	n/a	n/a	1	1	6
9	LIS	8	UM	UM	1	1	2	8	n/a	n/a	7	1	5
10	LIS	1	GFS	GFS	1	1	2	4	n/a	n/a	7	1	6

Table B3: MEPS third configuration during research sample (Kuchera, 2013).

M	LIC	LUT	IC	LBC	SW	LW	LSM	MP	H	CCN	PBL	SL	C
1	LIS	10	UM	UM	1	1	2	16	1	5E+8	1	1	2
2	LIS	2	GFS	GFS	1	1	2	10	1	1E+8	8	2	6
3	LIS	2	GEM	GEM	5	5	2	16	0	1E+9	1	1	14
4	LIS	AFWA	GEM	GEM	5	5	2	10	1	1E+9	8	1	2
5	LIS	5	UM	UM	3	3	2	8	n/a	n/a	7	1	14
6	LIS	6	GFS	GFS	1	1	2	16	1	1E+8	7	1	6
7	LIS	7	GEM	GEM	1	1	2	8	n/a	n/a	1	1	2
8	LIS	8	GFS	GFS	3	3	2	10	0	1E+8	1	1	6
9	LIS	8	UM	UM	3	3	2	16	0	5E+8	7	1	14
10	LIS	1	GFS	GFS	5	5	2	8	n/a	n/a	7	1	6

Table B4: MEPS fourth configuration during research sample (Kuchera, 2013).

M	LIC	LUT	IC	LBC	SW	LW	LSM	MP	H	CCN	PBL	SL	C
1	LIS	10	UM	UM	1	1	2	16	1	5E+8	1	1	2
2	LIS	2	GFS	GFS	1	1	2	10	1	1E+8	8	1	6
3	LIS	2	GEM	GEM	5	5	2	16	0	1E+9	1	1	14
4	UM	AFWA	GEM	GEM	5	5	2	10	1	1E+9	8	1	2
5	UM	5	UM	UM	3	3	2	8	n/a	n/a	7	1	6
6	LIS	6	GFS	GFS	1	1	7	16	1	1E+8	7	1	6
7	UM	7	GEM	GEM	1	1	7	8	n/a	n/a	1	1	14
8	LIS	8	GFS	GFS	3	3	7	10	0	1E+8	1	1	2
9	UM	8	UM	UM	3	3	7	16	0	5E+8	7	1	2
10	UM	1	GFS	GFS	5	5	7	8	n/a	n/a	7	1	14

References

Air Force Weather Web Services (AFW-WEBS), 2013: Operational AFWA Ensemble Information. URL https://weather.af.mil/confluence/display/AFPUBLIC/ Operational+AFWA+Ensemble+Information.

Brier, G. W., 1950: Verification of forecasts expressed in terms of probability. *Mon. Wea. Rev.*, **78**, 1–3.

Bröcker, J., 2012: Estimating reliability and resolution of probability forecasts through decomposition of the empirical score. *Clim. Dyn.*, **39**, 655-667.

Buizza, R., P. L. Houtekamer, G. Pellerin, Z. Toth, Y. Zhu, and M. Wei, 2005: A Comparison of the ECMWF, MSC, and NCEP Global Ensemble Prediction Systems. *Mon. Wea. Rev.*, **133**, 1076–1097.

Charney, J. G., R. Fjortoft, and J. von Neuman, 1950: Numerical integration of the barotropic vorticity equation. *Tellus* **2**, 237-254.

Devore, J. L., 2004: *Probability and Statistics for Engineers and the Sciences*. Thomson Brooks/Cole, 67 pp.

Eckel, F. A., and C. F. Mass, 2005: Aspects of effective mesoscale, short-range ensemble forecasting. *Wea. Forecasting*, **20**, 328–350.

Eckel, F. A., J. G. Cunningham, and D. E. Hetke, 2008: Weather and the calculated risk, exploiting forecast uncertainty for operational risk management. *Air & Space Power Journal*, **22**, 71–82.

Ehrendorfer, M., and J. J. Tribbia, 1997: Optimal prediction of forecast error covariances through singular vectors. *J. Atmos. Sci.*, **54**, 286–313.

Epstein, E. S., 1969: Stochastic-dynamic prediction. *Tellus* **21**, 739-759.

Ferro, C. A. T., and T. E. Fricker, 2012: A bias-corrected decomposition of the Brier score. *Quart. J. Roy. Meteor. Soc.*, **138**, 1954-1960.

Hamill, T. M., R. Hagedorn, and J. Whitaker, 2008: Probabilistic forecast calibration using ECMWF and GFS ensemble reforecasts. Part II: precipitation. *Mon. Wea. Rev.*, **136**, 2620–2632.

Hoffman, R. N., and E. Kalnay, 1983: Lagged average forecasting, an alternative to Monte Carlo forecasting. *Tellus* **35A,** 100-118.

Houtekamer, P. L., H. L. Mitchell, and X. Deng, 2009: Model error representation in an operational ensemble Kalman filter. *Mon. Wea. Rev.*,**137**, 2126–2143.

Kalnay, E., 2003: *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge University Press, 341 pp.

Kuchera, E. Deputy Chief, Numerical Models Flight, Fine Scale Models and Ensembles Team Lead, AFWA 16/WS, NB. Personal Correspondence. 9 March – 18 December 2013.

Leith, C. E., 1974: Theoretical skill of Monte Carlo forecasts. *Mon. Wea. Rev.*, **102**, 409–418.

Lorenz, E. N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20**, 130–141.

McCaul, E., Jr., S. J. Goodman, K. M. LaCasse, and D. J. Cecil, 2009: Forecasting lightning threat using cloud-resolving model simulations. *Wea. Forecasting*, **24**, 709–729.

Murphy, A. H., 1973: A new vector partition of the probability score. *J. Appl. Meteor.*, **12**, 595–600.

Petersen, W. A., H. J. Christian, and S. A. Rutledge, 2005: TRMM observations of the global relationship between ice water content and lightning. *Geophys. Res. Lett.*, **32**.

Stephenson, D. B., C. A. S. Coelho, and I. T. Jolliffe, 2008: Two extra components in the Brier score decomposition. *Wea. Forecasting*, **23**, 752–757.

Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Meteor. Soc.*, **74**, 2317–2330.

Toth, Z., and E. Kalnay, 1997: Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.*, **125**, 3297–3319.

Wang, Y., S. Tascu, F. Weidle, and K. Schmeisser, 2012: Evaluation of the Added Value of Regional Ensemble Forecasts on Global Ensemble Forecasts. *Wea. Forecasting*, **27**, 972–987.

Wei, M., Z. Toth, R. Wobus, and Y. Zhu, 2008: Initial perturbations based on the ensemble transform (ET) technique in the NCEP global operational forecast system. *Tellus A*, **60**, 62–79.

Weisman, M. L., J. B. Klemp, and W. C. Skamarock, 1997: The resolution dependence of explicitly modeled convective systems. *Mon. Wea. Rev.*, **125**, 527–548.

Wilks, D. S., 2001: *Statistical methods in the atmospheric sciences*. Academic Press, 331 pp.

WMO Publication NO. 306 – Manual on Codes

Vita

Captain William B. Clements was born in Durham, NC, where he graduated from North Raleigh Christian Academy in 2004. He attended North Carolina State University graduating with Magna Cum Laude honors earning a Bachelor of Science Degree in Meteorology with a minor in Environmental Science and Military Studies. Concurrently, on a 4-year Air Force Reserve Officer Training Corps scholarship, he was commissioned a Second Lieutenant and Distinguished Graduate of Detachment 595.

Captain Clements' first assignment was to the 17th Operational Weather Squadron, Joint Base Pearl Harbor-Hickam, HI serving in multiple positions including - Theater Meteorological Supervisor, Forecasting Element OIC and Assistant Operations Flight Commander where he supported operations for the largest DoD weather area of responsibility spanning 113M Sq miles. There he supported multiple multinational exercises and humanitarian efforts to include the Fukashima tsunami/nuclear disaster. In May 2012 he was selected to enter the Graduate School of Engineering and Management at the Air Force Institute of Technology at Wright-Patterson AFB, OH, to obtain a Master's Degree in Applied Physics with a concentration in Atmospheric and Space Science. Following graduation, he will be assigned to Holloman AFB, NM, 49th Operations Group as the Weather Flight Commander.

REPORT DOCUMENTATION PAGE						*Form Approved* *OMB No. 074-0188*

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From – To)*
27-03-2014	Master’s Thesis	June 2012 – March 2014

4. TITLE AND SUBTITLE	
Validation of the Air Force Weather Agency Ensemble Prediction Systems	**5a. CONTRACT NUMBER**
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	**5d. PROJECT NUMBER**
Clements, William B, Captain, USAF	**5e. TASK NUMBER**
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S)	**8. PERFORMING ORGANIZATION REPORT NUMBER**
Air Force Institute of Technology Graduate School of Engineering and Management (AFIT/EN) 2950 Hobson Way, Building 640 WPAFB, OH 45433	AFIT/ENP/14-M-04
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	**10. SPONSOR/MONITOR’S ACRONYM(S)** AFWA
Air Force Weather Agency Mr. Evan E Kuchera 101 Nelson Drive Offutt AFB, NE (402) 294-3724, (DSN 271) evan.kuchear@us.af.mil	**11. SPONSOR/MONITOR’S REPORT NUMBER(S)**

12. DISTRIBUTION/AVAILABILITY STATEMENT
DISTRIBUTION STATEMENT A. APPROVED FOR PUBLIC RELEASE; DISTRIBUTION IS UNLIMITED

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Air Force Weather Agency’s (AFWA) Ensemble Prediction Systems (EPS), Global Ensemble Prediction System (GEPS), 20km Mesoscale Ensemble Prediction System (MEPS20) and 4km Mesoscale Prediction System (MEPS4), were evaluated from April to October 2013 for 10 locations around the world to determine how accurately forecast probabilities for wind and precipitation thresholds and lightning occurrence match observed frequencies using Aerodrome Routine Meteorological Reports (METARs) and Aerodrome Special Meteorological Reports (SPECIs). Reliability diagrams were created for each forecast hour detailing the Brier skill score (BSS) to depict EPS performance compared to climatology for each site and score composition through reliability, resolution and uncertainty. To illustrate how the BSS changed, the score and its decomposition were plotted for all forecast hours. This study showed that all three EPS suffered from a lightning overforecasting bias at all locations and most forecast hours. For wind speeds, it was clear that decreased model grid spacing allowed better resolution of terrain features, producing a better BSS. Likewise, precipitation was better resolved with increased horizontal resolution as explicit resolution of precipitation processes outperformed cumulus parameterization schemes.

15. SUBJECT TERMS
Ensemble Prediction System, GEPS, MEPS, Reliability Diagram

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** Kevin S, Bartlett, Lt Col, Ph.D, USAF (ENP)
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER** *(Include area code)*
U	U	U	SAR	96	(937) 255-6565 x4520 kevin.bartlett@afit.edu