2007

# Mesoscale Superensemble Forecasts with a Suite of Models over the Continental United States and North America

Donald F. Van Dyke III

THE FLORIDA STATE UNIVERSITY

COLLEGE OF ARTS AND SCIENCES

# MESOSCALE SUPERENSEMBLE FORECASTS WITH A SUITE OF MODELS OVER THE CONTINENTAL UNITED STATES AND NORTH AMERICA

By

Donald F. Van Dyke III

A Thesis submitted to the
Department of Meteorology
in partial fulfillment of the
requirements for the degree of
Master of Science

Degree Awarded:
Summer Semester, 2007

The members of the Committee approve the Thesis of Donald Van Dyke defended on Tuesday, June 12th, 2007.

_____

T.N. Krishnamurti
Professor Directing Thesis

_____

Robert Hart
Committee Member

_____

Paul Ruscher
Committee Member

The Office of Graduate Studies has verified and approved the above named committee members.

TABLE OF CONTENTS

LIST OF FIGURES

ABSTRACT

Using a suite of high resolution models, the forecast skills of the superensemble for precipitation and 2-meter temperature over the continental United States and North America are shown. In this study, models and/or gridded fields such as the ETA, WRF-ARW, MM5, NDFD, several global models, and their ensemble mean are used. The final resolution for the multimodel superensemble is at 32 km at 3-hourly temporal intervals for temperature and ¼ degree daily intervals for precipitation. The forecast length is 60 hours for temperature and 5 days for precipitation.

This study utilizes an optimization for the training length (number of days) to arrive at the best results. The period of the study includes March-September 2006. The metrics for the forecast evaluation include the mean absolute error, rms error, bias, and equitable threat scores. The results show a significant improvement of the multimodel superensemble compared to its member models and their ensemble mean.

# CHAPTER 1
# INTRODUCTION

## 1.1 Background and Thesis Objectives

Temperature and precipitation patterns affect everyone on the planet. In today's era, computer models are used to aid in forecasting temperature and precipitation on a daily basis across the globe. Accurate temperature forecasts play a significant role in the energy industry across the United States. According to an American Meteorological Society forum on weather, climate, and energy, an improvement of temperature forecasts by one degree Fahrenheit could decrease the annual cost of electricity by one billion dollars in the United States (Rogers 2001). From a societal standpoint, extreme temperatures and precipitation can have life or death consequences. The 1995 heat wave in the United States killed well over 500 people in Chicago alone (Karl and Knight 1997). Meanwhile, floods have been rated the number one natural disaster in the United States in terms of property damage and lives lost (USGS Facts Sheet 2000). The purpose of this study is to apply the FSU superensemble technique to regional, high resolution temperature and precipitation forecasts (32 km resolution for temperature, ¼ degree resolution for precipitation) in order to gain greater accuracy in these areas over the standard numerical models.

## 1.2 Previous Work

Much work has been done with the superensemble with many atmospheric variables since it was first developed. An experimental numerical weather prediction system was developed to forecast several variables such as geopotential heights, winds, surface pressure, and precipitation. Forecasts for the mean sea-level pressure, 500 hPa heights, surface winds, 850 hPa winds, and 200 hPa winds are made on a real-time basis. Krishnamurti et al (2003, 2001) showed generally higher skills of the superensemble for these variables over its global member models. These global models included the GFS (NCEP), NOGAPS (U.S. Navy), GGEM (Canadian Weather Service), JMA (Japan

Weather Service), BMRC (Australian Weather Service), and different versions of the FSU global spectral model. Cartwright (2004) applied the superensemble method to mesoscale models for summertime precipitation over the southeastern quadrant of the United States with increased skill. These included models such as the Operational ETA (now discontinued), ETA with Kain-Fritsch convection, NCAR MM5, FSL RUC, COAMPS, and the FSU Nested Regional Spectral Model. However, little to no effort has been given to temperature forecasting at any resolution using the superensemble.

### 1.3 Organization of Thesis

This thesis is divided into five chapters. Chapter 2 will discuss the history and give a description of the superensemble. Chapter 3 will discuss the datasets used for this study. Chapter 4 will discuss the methodology, observed temperature and precipitation patterns during the study, and the forecast results that were obtained. Finally, chapter 5 will discuss conclusions and possible future work based on the results.

# CHAPTER 2
# SUPERENSEMBLE HISTORY AND DESCRIPTION

## 2.1 Superensemble History

The multimodel superensemble forecast method was developed in 1998 by Prof. T.N. Krishnamurti as a way to improve Atlantic tropical cyclone forecasts. It utilizes past forecasts from a suite of member models in an effort to correct their biases and rank their relative strength. This method has been demonstrated to produce tropical cyclone forecasts that are generally better than any member model for a given time (Krishnamurti et al. 1999).

The superensemble is a way to improve upon ensemble mean forecasting. It is important to note that ensemble mean forecasting has been proven to be an alternative to forecasts from a single model. In the area of tropical cyclone forecasting, Goerss (2000) combined three dynamical models into an ensemble mean and demonstrated increased skill over the majority of single models. An ensemble means involves a straight average of all member models, whereby all models are assigned an equal weight. In the case of the superensemble, member models are optimally, unequally weighted according to their relative skills.

With the success of the superensemble in tropical cyclone forecasting, the superensemble technique was expanded to include global numerical weather prediction and climate forecasting (Krishnamurti et al. 2000a), as well as global precipitation forecasting (Krishnamurti et al. 2000b). The results demonstrated high skill in these areas as well. However, this thesis will only investigate regional, high resolution forecasting of precipitation as well as temperature.

## 2.2 Superensemble Description

Unlike a normal ensemble mean, the superensemble also accounts for the member model biases as well as their relative strengths based on their good and bad past

performances. The superensemble then assigns these models higher and lower weights accordingly utilizing a multiple linear regression technique. Such a technique attempts to reward more skillful models with higher weights and less skillful models with lower weights (Williford 2002).

In order to accomplish this, the superensemble methodology divides multimodel datasets into a training phase and a forecast phase. In the training phase, the forecast fields are regressed against the observed fields using multiple linear regression at all grid locations along the horizontal and vertical coordinates. The linear regression technique uses a minimization function to limit the spread between the member model forecasts and the observed state. This minimization equation is given by

$$O' = \sum_{i=1}^{N} \sum_{j=1}^{train} a_i (F_{ij} - \overline{F_{ij}}) + \varepsilon_i , \qquad (2.1)$$

where $F_{ij}$ is the $i^{th}$ model forecast out of N total models, $\overline{F_{ij}}$ is the mean of the $i^{th}$ forecast over the training period "train", O' is the observed anomaly relative to the observed mean over the training period, $a_i$ is the $i^{th}$ regression coefficient, and $e_i$ is an error term. The $(F_{ij} - \overline{F_{ij}})$ term represents the forecast anomaly for model i at time j. The $a_i$ coefficients are determined by minimizing the summed squared error over the training period $E = \sum_{i=1}^{N} \varepsilon_i^2$. In the global numerical weather prediction forecasts, Krishnamurti (2000a) has noted that the member models performances vary from location to location. This can arise from several factors, including treatment of physics, orography, lakes, land surfaces, boundary conditions, etc. Inaccurate modeling of these factors can lead to systematic errors in the member models, which the superensemble attempts to correct through the regression coefficients $a_i$. The obtained regression weights are then used in the forecast phase to calculate the superensemble forecast. These regression weights, which are calculated for every gridpoint in the domain, provide a collective bias correction of the member models. This allows a superensemble forecast to have somewhat higher skill compared to the ensemble mean and the best member model.

The length and selection of a given training dataset is an important part of achieving a highly skillful forecast. For tropical cyclones, previous work indicates

between 50 and 75 forecast cases are required. For Atlantic tropical cyclones, approximately two years of previous forecast cases are used. For the realtime, operational, global numerical weather prediction, including precipitation, approximately 65 days of previous forecasts are used as training. However, these global numerical weather prediction forecasts are on the order of 1 degree resolution. Krishnamurti et al. (2001) demonstrated that selecting a training dataset with poor quality can degrade the superensemble forecasts. For the high resolution forecasts of this thesis, a range of training days was utilized in order to provide a sensitivity study on training required for high resolution datasets, as well as the training required for the temperature variable.

In the forecast phase of the superensemble, the superensemble forecast S(t) is determined utilizing data gathered in the training phase as well as current member model forecasts. S(t) is given by

$$S(t)=\overline{O}+\sum_{i=1}^{N}\sum_{j-1}^{train} a_i(F_{ij}-\overline{F_{ij}}), \qquad (2.2)$$

where $\overline{O}$ is the observed mean over the training period and the rest of the variable are defined the same as above. The determination of the $a_i$ coefficients comes through an error covariance matrix that is solved using singular value decomposition (SVD) as described in Yun and Krishnamurti (2002). Figure 2.1 shows a schematic of the superensemble technique.

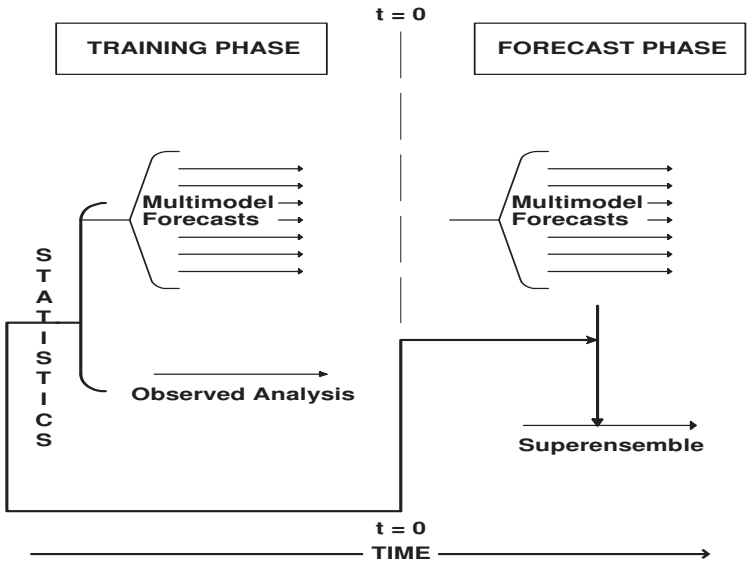


**Figure 2.1 Schematic showing the training and forecast phases of the superensemble**

The issue of how many member models are required for optimal skill of the superensemble was addressed in Krishnamurti (2000a). Here, superensemble forecasts were performed using between 1 and 7 models. The best performing member model was added first, followed sequentially by the next best performing member model until finally the worst model was added last. Figure 2.2 shows the results of the ensemble mean and superensemble for the 850 hPa mean global wind RMS error. The errors of the ensemble mean increase beyond 3 models due to less skillful models being added into the ensemble. However, the errors of the superensemble do not increase in this manner due to the bias correction and assignment of lower weights to these models.

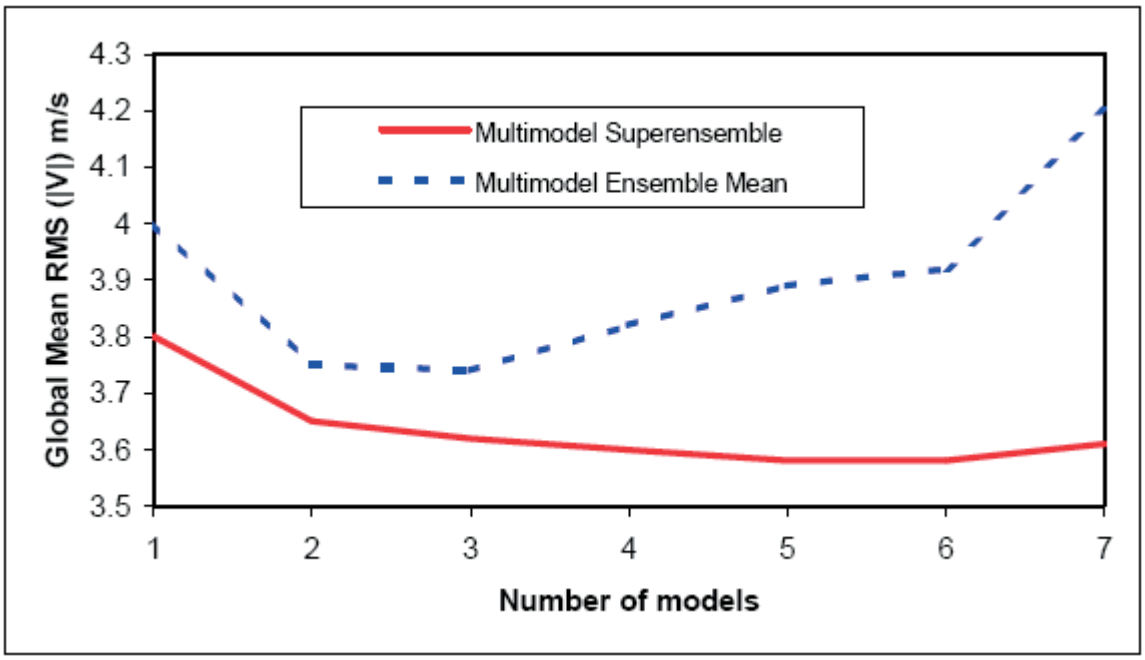


**Figure 2.2 August 1998 global mean RMS error (m/s) of 850 hPa wind for day 3 forecast (Krishnamurti et al. 2000a)**

# CHAPTER 3
# DATASETS

## 3.1 Analysis Dataset for Temperature

A good training dataset begins with a good analysis dataset. For the temperature portion of this study, the North American Regional Reanalysis (NARR) dataset was employed in the superensemble training phase as well as forecast verification. The NARR represents a major improvement to the earlier NCEP global reanalysis datasets for both resolution and accuracy. Data is available from January 1979 to near present time at 3-hourly time steps. Its domain and topography are shown below in figure 3.1.

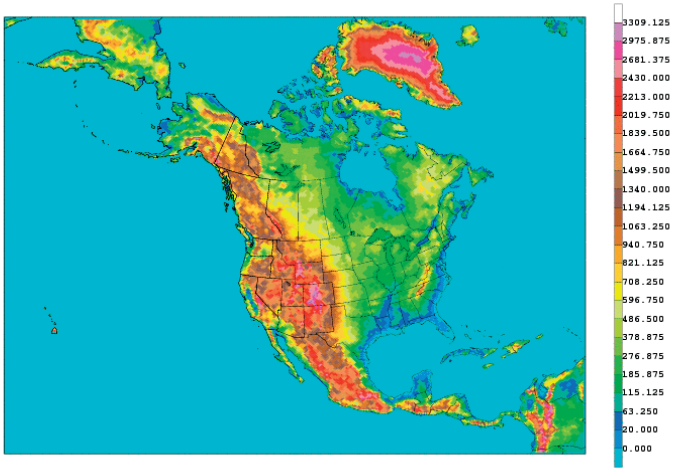


**Figure 3.1 NARR domain and topography (scale in meters)**

It employs an analysis system similar to the ETA model's 3D-Var Data Assimilation System (EDAS) that was operational in April 2003, although it utilizes additional datasets for assimilation as described in Mesinger et al. (2005). It has a 32 km/45 layer resolution. The 2-meter temperature fields from this dataset were utilized for this study to represent the "observed" field. Mesinger et al (2005) also shows the gridded 2-meter temperature fields to be superior to the gridded global reanalysis 2-meter temperature fields in terms of RMS error, as shown in figure 3.2, and bias, as shown in figure 3.3.

This fact, combined with its high resolution, made the NARR dataset an ideal choice for this study.

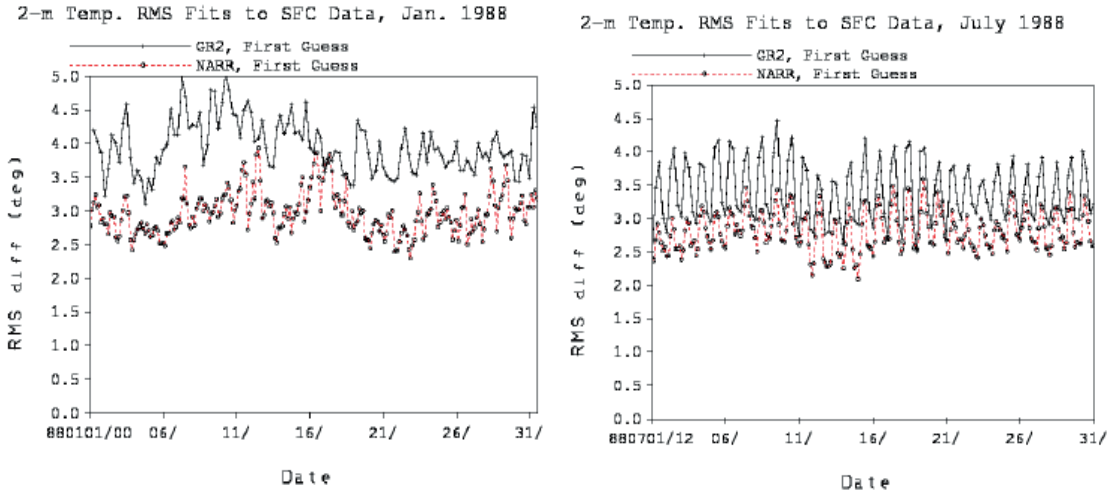


**Figure 3.2 Comparison of the RMS error of the first guess 2-m temperatures for the NARR and Global Reanalysis for January 1988 (left) and July 1998 (right). (Mesinger et al. 2005)**

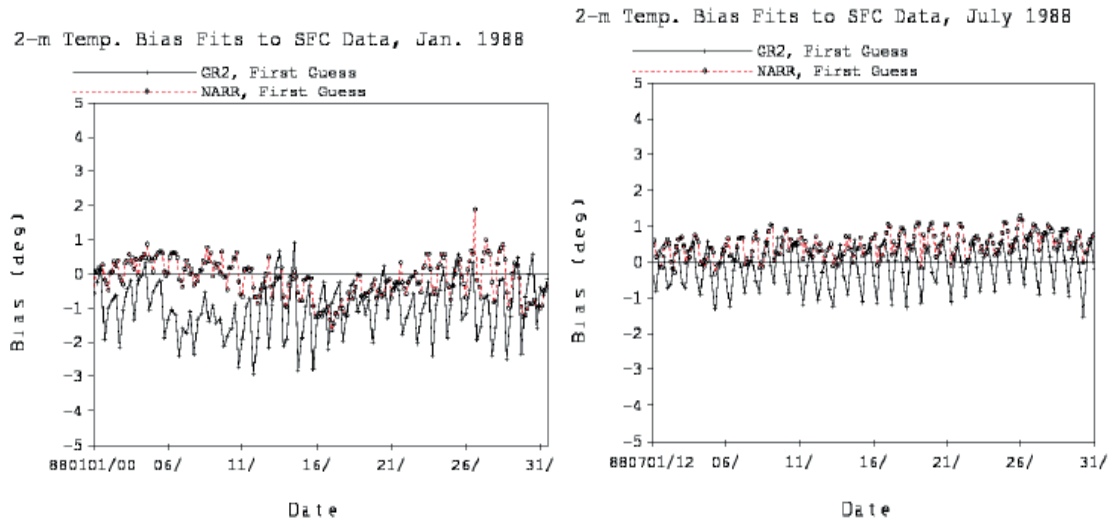


**Figure 3.3 Comparison of the bias of the first guess 2-m temperatures for the NARR and Global Reanalysis for January 1988 (left) and July 1988 (right). (Mesinger et al. 2005)**

### 3.2 Analysis Dataset for Precipitation

For the precipitation portion of this study, the CMORPH daily 0.25 x 0.25 degree lat/lon resolution dataset was employed in the superensemble training phase as well as forecast verification. CMORPH represents a major improvement over previous techniques that used microwave and infrared information. The improvement comes through the use assembling, propagating, and morphing rainfall estimates from satellites using passive microwave sensors. More specific details on this technique can be found in Joyce et al (2004). Over the United States, Joyce et al. (2004) show the skill of CMORPH, radar, and other satellite techniques that use microwave and infrared information when compared to a rain gauge analysis as seen in figure 3.4.

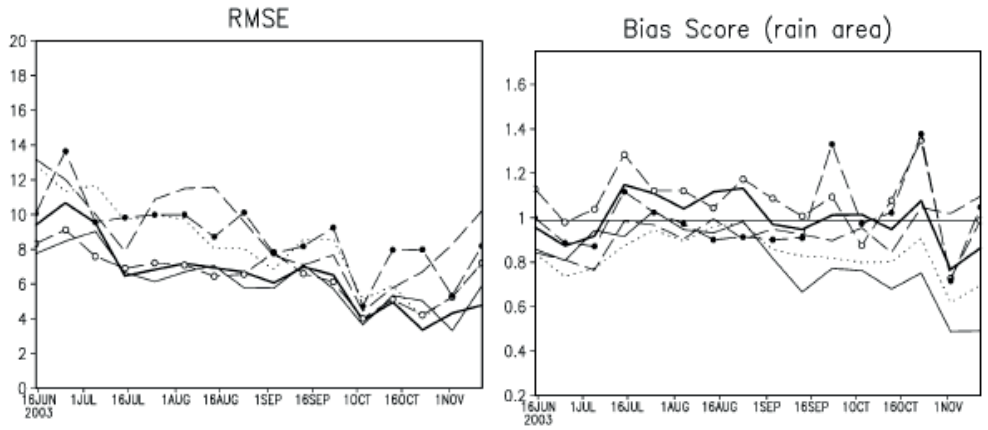


**Figure 3.4 RMS error and bias of rainfall estimation over the United States from June 15 through November of 2003. The solid thick line is CMORPH, solid thin line is radar, dotted line is MWCOMB (an ensemble mean of other satellite estimation techniques), and other lines are blended microwave-IR techniques. From Joyce et al. (2004)**

Over the United States, radar and CMORPH analysis perform the best when compared with the actual rain gauge measurements. However, the CMORPH analysis has superior bias characteristics than radar over the second half of the period, which may be due to beam blocking of the radar by the terrain in the western U.S. CMORPH outperforms the other satellite estimation methods in almost every validation statistic except for bias (Joyce et al. 2004). Figure 3.5 shows the distribution of daily rain rates among the rain gauge, CMORPH, and MWCOMB.

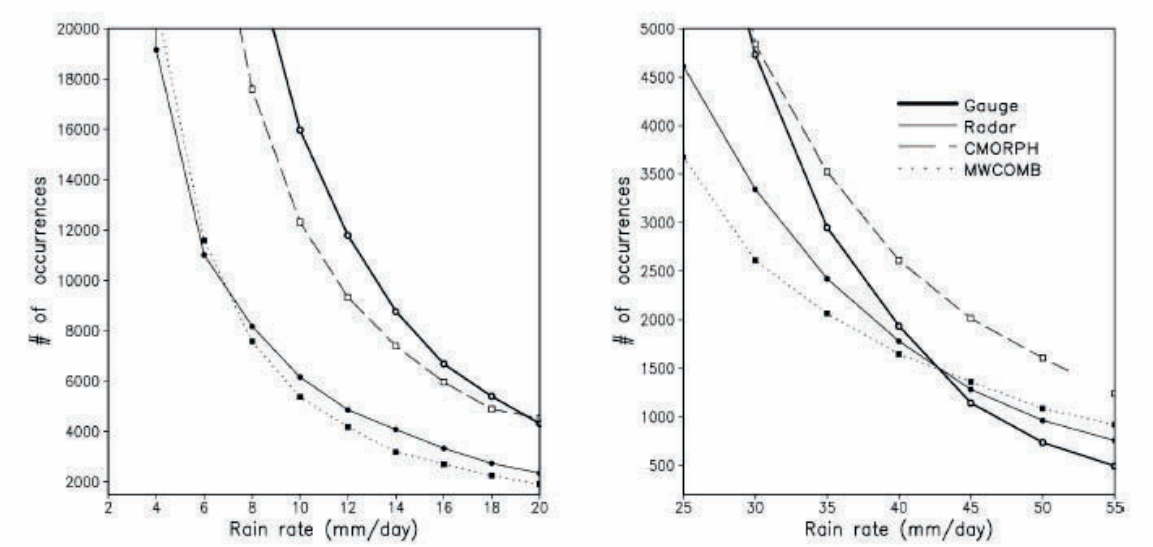


**Figure 3.5 Daily rain rates over the United States from April-May of 2003.**

At rainfall levels between 2 and 20 mm/day, all measurements detect less rainfall than observed by the gauges. However, CMORPH has the closest estimate. Between 25 and 35 mm/day, CMORPH has a high bias while the other techniques have a low bias. However, CMORPH is still the closest to the observed number of rainfall occurrences at those thresholds. Beyond 45 mm/day, Joyce et al. (2004) states that the comparisons should be viewed judiciously due to the well known tendency of objective analyses to suppress high rainfall amounts because of the smoothing employed.

## 3.3 Model Datasets for Temperature

### 3.3.1 ETA Model

The ETA model is a hydrostatic mesoscale model that was run at NCEP with 12 km horizontal grid spacing and 50 vertical levels. The forecast length was 84 hours run four times daily initialized at 00 UTC, 06 UTC, 12 UTC, and 18 UTC daily. The domain is shown in figure 3.6.

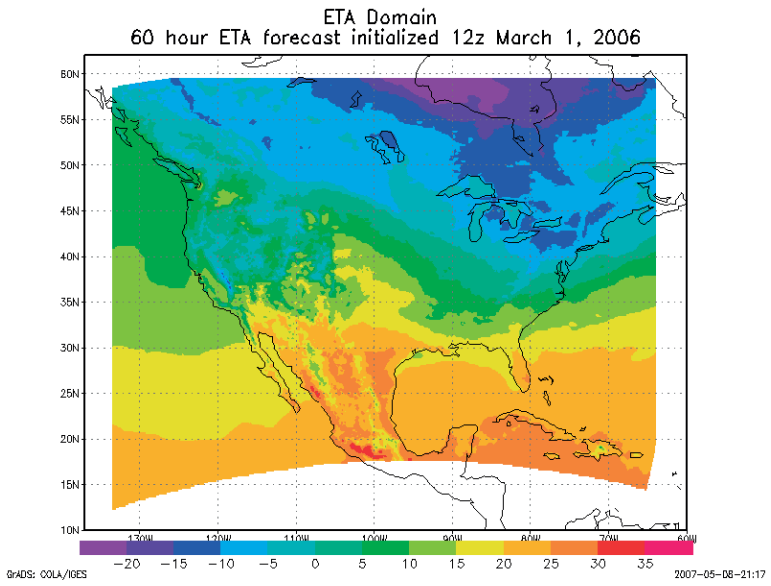


**Figure 3.6 ETA domain. The plot shown is the 60-hour 2-meter temperature forecast in degrees C from the 12 UTC March 1, 2006 run.**

It uses a semi-staggered Arakawa E grid as the basis of its horizontal structure. The topography is represented as discrete steps whose tops coincide with one of the model's 50 vertical levels. The eta coordinate is used, which is defined according to equation 3.1a and 3.1b.

$$\eta = \frac{p - p_t}{p_s - p_t} \times \eta_k \tag{3.1a}$$

$$\eta_k = \frac{p_{rf}(z_s) - p_t}{p_{rf}(0) - p_t} \tag{3.1b}$$

In these equations, p is pressure, and the subscripts rf, s, and t refer to the reference pressure, the model surface, and the model top, which is 50 hPa. The eta coordinate's advantage over the commonly used sigma coordinate becomes apparent over steep terrain where problems arise in the pressure gradient force computation using the sigma coordinate. These result in advection and diffusion errors, which are reduced using the eta coordinate (Chuang and Manikin 2001). Figure 3.7 shows a schematic of the eta coordinate system and topography representation in the model.

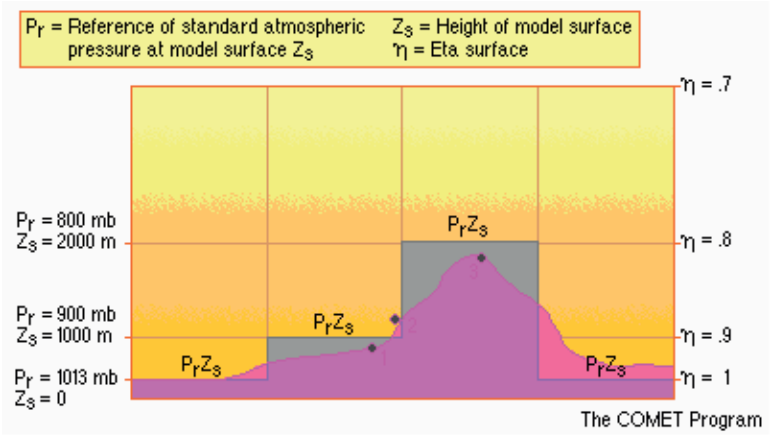


**Figure 3.7 Eta coordinate system and topography. Points 1 and 2 are raised and lowered respectively to the .9 eta surface and point 3 is raised to the .8 eta surface.**

The model uses the Eta Data Assimilation Scheme (EDAS), which is a 3DVAR technique. Boundary conditions for the model are provided by 6 hour old GFS model forecasts. The model also uses the Betts and Miller (1986) convective parameterization scheme further modified by Janjic (1994). The 2-meter temperatures for the model are calculated using the ETA post processor. A bulk layer parameterization of the surface-layer consistent with the Mellor-Yamada Level 2.0 model is utilized in the model. Using the Mellor-Yamada Level 2.0 model, Lobocki (1993) derived an equation for the surface layer relating the potential temperature, , between two levels, z1 and z2 as given by equation 3.2

$$\Theta(z_2) - \Theta(z_1) = \frac{\Theta^*}{x}\Phi_U(z_1, z_2, L) \tag{3.2}$$

where L = Monin-Obukhov scale, = constant coefficients, and x = von Karman constant. The functions and are integrated forms of similarity functions for dimensionless differences of the quantity U or across the layer z1 to z2. Specifically,

$$\Phi_s(z_1, z_2, L) = \phi_s(0) \times \left[\ln\left(\frac{z_2}{z_1}\right) + \psi_s(\zeta_2) + \psi_s(\zeta_1)\right] \tag{3.3}$$

where is a constant, , and . For 2-meter temperatures, the height z1 refers to the 2-meter level, and the height z2 refers to values in the first eta layer above ground.

### 3.3.2 WRF-ARW Model

This study also used the WRF-ARW model that is regularly run at NCAR. It was configured to run at 22 km resolution with initialization of boundary conditions from a 40 km resolution ETA. The forecast length is 72 hours run twice daily at 00 UTC and 12 UTC. It uses WRF-var data assimilation, which is a 3DVAR data assimilation scheme. The domain is shown in figure 3.8.

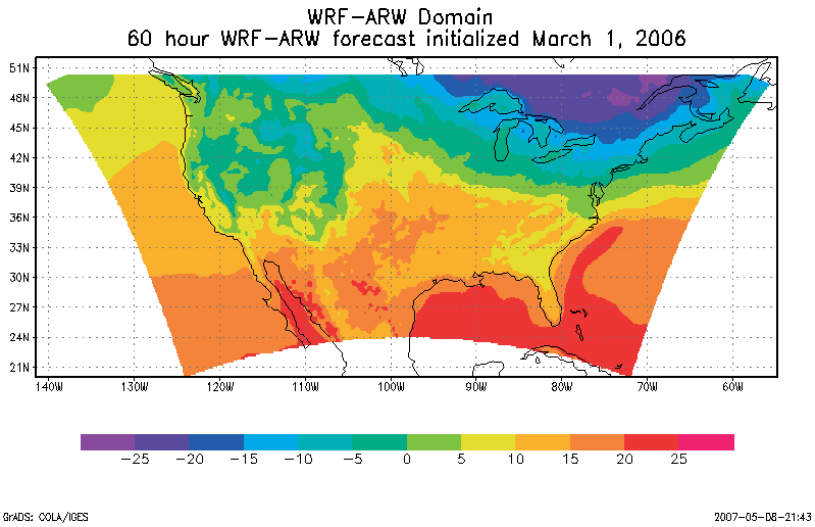


**Figure 3.8 WRF-ARW domain. The plot shown is the 60-hour 2-meter temperature forecast in degrees C from the 12 UTC March 1, 2006 run.**

It uses a terrain following, hydrostatic-pressure vertical coordinate defined by equation 2.1a and 2.1b.

$$\eta = (p_h - p_{ht})/\mu \qquad (2.1a)$$

$$\mu = p_{hs} - p_{ht} \qquad (2.1b)$$

where $p_h$ is the hydrostatic component of pressure, and $p_{hs}$ and $p_{ht}$ are values along the surface and top boundaries, respectively (Skamarock et al. 2005). Unlike the eta coordinate system used in the ETA model, this coordinate system is the traditional sigma coordinate system used in many hydrostatic models. Figure 3.9 shows a schematic of the sigma coordinate system.

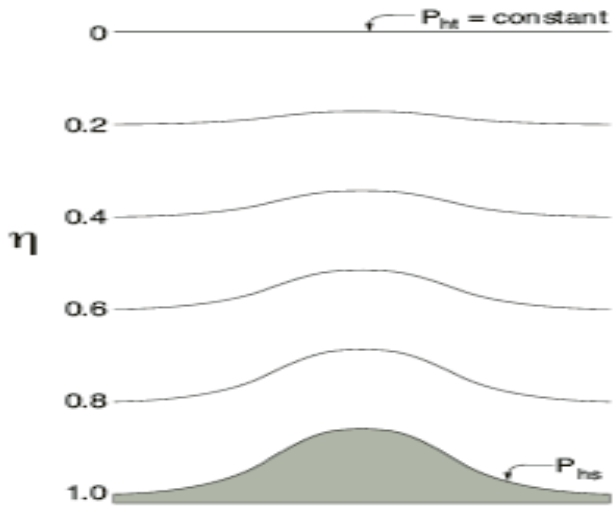


**Figure 3.9 Sigma coordinate system. Unlike the eta coordinate, the sigma coordinate is terrain following.**

The 2-meter temperatures for the model are calculated using the WRF-RIP post processor. The post processor uses a rough approximation of similarity theory. If the air temperature at the lowest model layer is warmer than the ground temperature, then the 2-meter temperature is set to be the lowest model layer temperature. If the air temperature at the lowest model layer is colder than the ground temperature, then the 2-meter temperature is equal to the average of the lowest model layer temperature and ground temperature. This is significantly different and more simplistic than the ETA post processor, and the skill scores suffer as will be shown later.

### 3.3.3 National Digital Forecast Database

This study also used the National Digital Forecast Database (NDFD) produced by the National Weather Service. Although this is not automated forecast model, the output is produced on a very high resolution 5 km grid that is compatible for use in the superensemble. The forecast length for 2-meter temperature is 60 hours at 3-hourly time steps and 156 hours at 6-hourly time steps. The NDFD is a compilation of forecasts from all of the National Weather Service Weather Forecast Offices across the United States. For each weather element at each forecast time step, the individual grids from the 117 WFOs across the continental United States are entered into the NDFD. These grids do not overlap except for at a narrow boundary (Glahn and Ruth 2003). Figure 3.10 shows

the domain of the NDFD, while figure 3.11 shows the locations and shapes of the 117 WFOs.

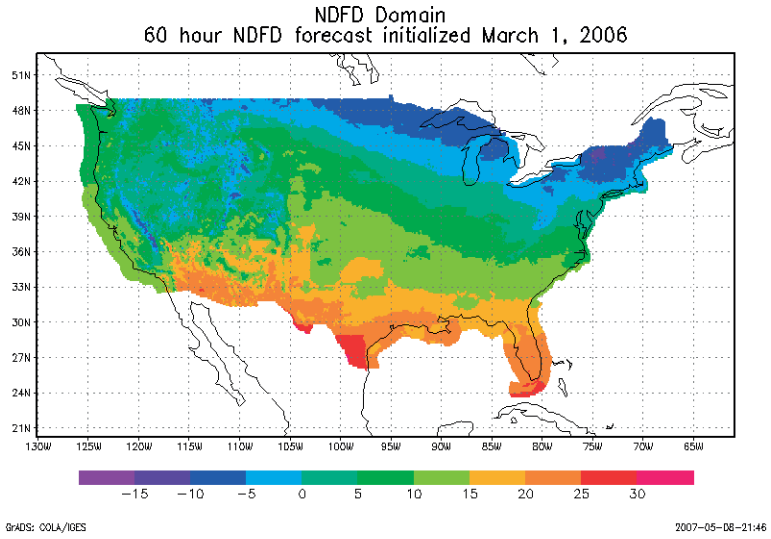


**Figure 3.10 NDFD domain. The plot shown is the 60-hour 2-meter temperature forecast in degrees C from a starting point of 12 UTC March 1, 2006.**

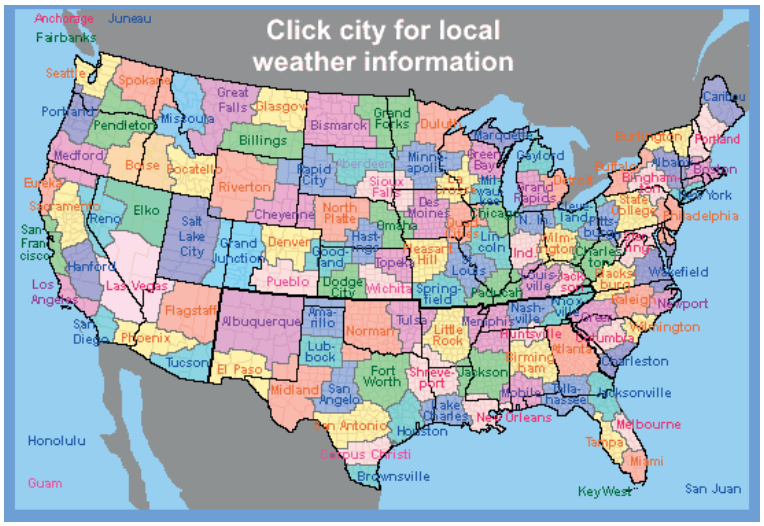


**Figure 3.11 The locations and shapes of the 117 WFOs across the continental United States that make up the NDFD.**

Careful attention is given to minimizing any discontinuities of individual WFO forecasts with each other at their boundaries. Collaboration on the forecasts is carried out on three levels: NCEP and WFO, WFO and WFO, and a final automated quality control check of the mosaic NDFD (Glahn and Ruth 2003).

NCEP's Hydrometeorological Prediction Center (HPC) provides digital guidance in space and time to individual WFOs. HPC forecasters also analyze the NDFD to see if any forecasts may need to be changed. If they feel a forecast may need updating, then they can initiate collaboration with the affected WFO(s). Collaboration with HPC can also be initiated by the WFOs. This collaboration helps to ensure a more accurate NDFD product.

WFOs also collaborate with each other. Forecast grids at a WFO are sent to neighboring WFOs for simultaneous viewing. This allows for the discussion of the forecast amongst neighboring WFOs to ensure boundary consistency.

Finally, the forecasts are automatically checked by software at the NDFD central server. If a discontinuity greater than an agreed upon threshold is found, then the affected WFOs are automatically notified and have the chance to modify their forecasts in order to make them more compatible with the surrounding WFOs. HPC also has the opportunity to view the forecasts and make suggestions at this stage (Glahn and Ruth 2003). This new forecasting process of preparing the NDFD is shown in figure 3.12.

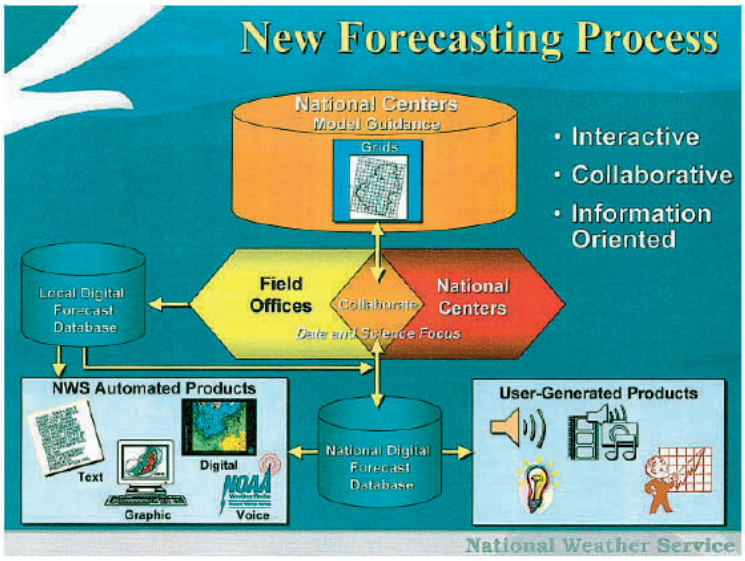


**Figure 3.12 The process of creating the NDFD. From Glahn and Ruth (2003).**

### 3.4 Statistical Datasets for Temperature

Superensemble temperature forecasts were also compared to NCEP's operational Model Output Statistics (MOS) product at 6 stations across the country. MOS is an objective technique which determines a statistical relationship between variables forecast

by a numerical model and the predictand at a given forecast hour (Glahn and Lowry 1972). There are currently three versions of MOS in operation: NGM MOS, ETA MOS, and GFS MOS. In the GFS and ETA versions of MOS, the statistical equations are developed using multiple linear regression through forward selection. Equations are developed for two phases of the year: warm and cold. The warm phase is defined to run from April 1 to September 30, and the cold phase runs from October 1 to March 31. The NGM and ETA MOS are available two times per day at 00 UTC and 12 UTC, while the GFS MOS is available four times per day at 00 UTC, 06 UTC, 12 UTC, and 18 UTC. NGM MOS provides a 60 hour forecast of several variables, including 2-meter temperature, at 3 hourly time steps, while ETA and GFS MOS provide forecasts out to 72 hours at 3 hourly time steps. When compared to raw model output of 2-meter temperature, MOS exhibits a lower mean absolute error (MAE), as shown in figure 3.13.

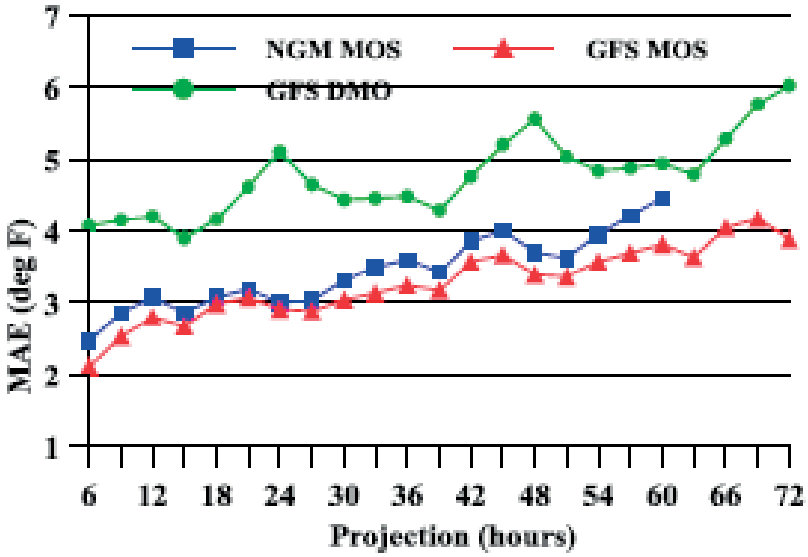


**Figure 3.13 Comparison of NGM MOS, GFS MOS, and GFS Direct Model Output of 2-meter temperature from the 00 UTC cycle of the 2002-2003 cool season. Note the MOS improvement over the raw model output. From Dallavalle (2004).**

## 3.5 Model Datasets for Precipitation

### 3.5.1 Mesoscale Models

Two mesoscale models were included in the precipitation portion of this study: the WRF-ARW and the NCAR MM5. The WRF-ARW model was already described

above. The NCAR MM5 is a non-hydrostatic model run twice daily at 00 UTC and 12 UTC at NCAR. The model grid spacing is 30 km, and it produces forecasts out to 48 hours. The domain is shown below in figure 3.14.

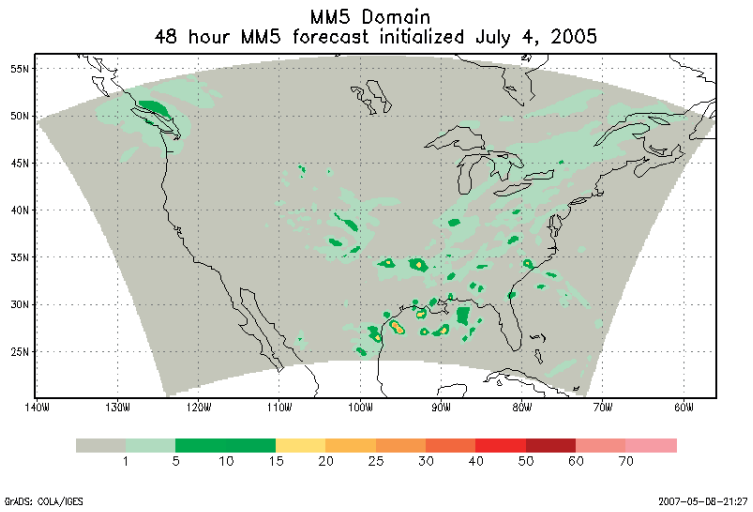


**Figure 3.14 NCAR MM5 domain. The plot shown is the 48 hour accumulated precipitation (mm) from the 12 UTC run initialized on July 4, 2005.**

The model was initialized with the early ETA model fields from NCEP. The ETA model also provides boundary conditions at 6-hour intervals. This MM5 used a Kain-Fritsch convective parameterization scheme that was developed specifically for mesoscale modeling. This scheme includes the effects of downdrafts and has been shown to organize convection owing to downdraft processes.

### 3.5.2 Global Models

Five global models were included as well for this study. Table 3.1 lists these models as well as some of their characteristics, including horizontal and vertical resolutions, various physical parameterization schemes, and the land surface schemes. Overall, there is a strong diversity in these operational models in terms of these characteristics. Some of the global operational centers are using more advanced versions of these models presently. However, table 3.1 lists the model configurations that were available at the time of this study.

**Table 3.1 Description of the global models used for precipitation.**

| Model | Resolution | | Cumulus Param. | PBL and Surface Physics | Radiation | Land Surface Processes |
|---|---|---|---|---|---|---|
| | Horizontal | Vertical | | | | |
| BMRC | T239 | 29 Levels upto 10hPa | Kuo 1974 | Stability dependent constatnt flux layer (Monin-Obukhov similarity theory) | Radiative transfer (Schwarzkopf and Fels 1991) (Edwards and Slingo, 1996) | 1.Gravity Wave drag parameterization 2.Interactive Soil Moisture (bucket Type) 3. Land Albedo climatology |
| JMA | T319 | 40 levels up to 0.4 hPa | Prognostic Arakawa Schubert Cum. Param. (1974) | Mellor Yamada Level 2, Monin-Obukhov similarity theory | short wave (every hour ) infrared (3 hourly), | Simple Biosphere Model (SiB) Gravity Wave drag |
| NCEP | T255 | 42 sigma levels up to 0.27 hPa | Pan and Wu (1994), based on Arakawa Schubert as simplified by Grell (1993) | PBL (Troen and Mahrt 1986), roughness length by (Charnock, 1955) and Monin Obukhov similarity theory | Long wave computation every 3 hours, short wave every hour Schwarzkopf and Fels (1991), Harshvardhan et al (1989) | Gravity wave drag SST; 5 day running mean Analysis Snow cover from NESDIS |
| NRL | T239 | 24 sigma levels up to 1 hPa, 5 sigma levels below 850 hPa | Relaxed Arakawa Schubert (RAS) (Moorthi and Suarez, 1992) | After ECMWF's vertical mixing parameterization (Louis et al 1982) Sfc flux param. (Louis 1979) | Harshvardhan et al (1989) | Gravity wave drag (Palmer et al 1986) SST :US Navy Ice cover %age : NAVICE center weekly analysis |
| RPN | T199 | 28 $\eta$ levels up to 10 hPa | Kuo type deep convection scheme | Sfc layer based on Monin-Obukhov similarity theory | Solar and IR radiation interactive with water vapor, $CO_2$, $O_3$ and clouds | Gravity wave drag, Prediction of sfc temperature over land (force-restore method) |

In addition to the diverse resolutions and model physics, these global models also have diverse data assimilation schemes. While all the models use roughly the same surface based observations, the satellite observations used are different. NCEP and NRL use the satellite based datasets in their 3DVAR data assimilation system. JMA implemented 4DVAR in their global model starting in March 2005. RPN also has 4DVAR. On October 31, 2006, RPN made major changes to their model which included increased horizontal and vertical resolution, improved physical parameterization, improved condensation and precipitation, improved model physics, and a more sophisticated ISBA (Interactions Soil Biosphere Atmosphere) surface scheme. They also made an improvement of 40% in the efficiency of their 4DVAR data assimilation. BMRC is still using a 1DVAR in their global model, but plans are being made to change to 3DVAR.

# CHAPTER 4
# METHODOLOGY AND RESULTS

## 4.1 Methodology

### 4.1.1 Temperature

For temperature, the three gridded temperature forecast datasets (mesoETA, WRF-ARW, NDFD) were re-gridded from their original horizontal resolutions to a resolution of 32 km to match the analysis field resolution of the NARR. The re-gridding was done using bi-linear interpolation in the GrADS software package. This allows for a one to one comparison of the error results. The NDFD domain was utilized since that domain represented the least common denominator amongst the datasets. (See Figure 3.10 shown earlier.) Once the three forecast datasets were re-gridded, an ensemble mean of those was calculated and also included as a member model to the superensemble, making a total of 4 member models used. The time period studied for temperature was the spring months of 2006, March, April, and May. Three separate superensemble runs were done for each day available during this period with each run utilizing a different number of training days. Either 80, 110, or 140 days was used in order to help determine the optimum number of training days. Each run consisted of a 2-meter temperature forecast out to 60 hours at 3-hourly time steps at a 32 km resolution. For March 2006, a training period ending on February 28 was used. For April 2006, a training period ending March 31 was used. For May 2006, a training period ending April 30 was used. It is also important to note that the training period was only updated once per month at the end of each month and not on a daily basis. However, this did not appear to affect the quality of the results in the latter half of each month, as will be shown later. Each forecast hour was trained with member model data from that same forecast hour only. For instance, the hour 3 forecast was trained with hour 3 member forecasts only, hour 24 with hour 24 member model data, hour 60 with hour 60, etc. Also, any day in the training phase or forecast phase that did not have all member models present was excluded.

In addition to the three models and their ensemble mean that were used, the NGM, ETA, and GFS MOS datasets were obtained for six stations across the U.S.:

Tallahassee, FL, Baltimore, MD, Hibbing, MN, Lincoln, NE, Fresno, CA, and Lovelock, NV. These stations were chosen for their differing geographical locations across the country, as well as for an attempt to study the skill of the superensemble for different regions using actual station data versus the model re-analysis data of the NARR.

Three skill statistics were calculated, including the mean error (bias), mean absolute error (MAE), root mean square error (RMSE). A frequency check of how many forecasts from each model had the lowest MAE was also completed. Equations for the bias, MAE, and RMSE are given below in equations 4.1, 4.2, and 4.3.

$$Mean\ Error = \frac{1}{N}\sum_{i=1}^{N}(F_i - O_i) \tag{4.1}$$

(4.2)

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(F_i - O_i)^2} \tag{4.3}$$

In these equations, N is the number of gridpoints, F is the forecast value, and O is the observed value. The mean error statistic answers the question, "What is the average forecast error?" However, it does not measure the magnitude of the errors, and it is possible to get a perfect score and still have a poor forecast if there are compensating errors in either direction. The MAE answers the question, "What is the average magnitude of the forecast errors?" However, it does not indicate the direction of the errors. The RMS error also answers the question, "What is the average magnitude of the forecast errors?" However, unlike MAE, it puts greater emphasize on larger errors.

**4.2.2 Precipitation**

For precipitation, the five gridded global model precipitation forecasts shown in table 3.1 were re-gridded from their original horizontal resolutions to a resolution of ¼ degrees to match the analysis field resolution of the CMORPH. Similar to temperature, this allows for a one to one comparison of the error results. A domain centered over the United States and Mexico was utilized, as shown below in figure 4.1.

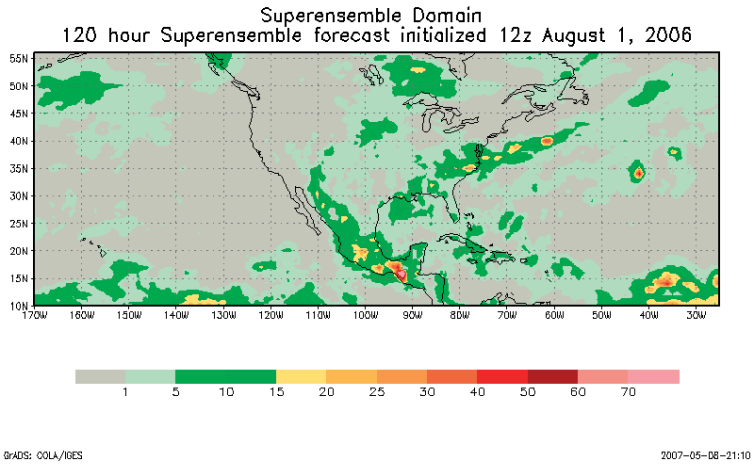


**Figure 4.1 FSU Superensemble domain. The plot shown is the day 5 precipitation (mm) forecast from the 12 UTC run initialized on August 1, 2006.**

Once the five forecast datasets were re-gridded, an ensemble mean was calculated for skill score comparisons later. However, unlike temperature, it was not included as a member model for the superensemble. In addition to the five global models, two mesoscale models were also used for comparison of skill, the NCAR MM5 and NCAR WRF-ARW. These models were re-gridded to ¼ degree horizontal resolution from their original horizontal resolutions as well. The time period studied for precipitation was June-September of 2006. The number of training days was set to 65 days during this period. However, during the month of September, seven runs per day were done, utilizing 65, 80, 95, 110, 125, 140, and 150 days of training in order to help determine the optimum number of training days to use. Each run consisted of a 24 hour accumulated precipitation forecast out to 5 days at ¼ degree resolution. For June 2006, a training period ending on May 31 was used. For July 2006, a training period ending on June 30 was used. For August 2006, a training period ending on July 31 was used. For September 2006, a training period ending on August 31 was used. Like temperature, it is important to note that the training period was only updated once per month at the end of each month and not on a daily basis. However, this did not appear to affect the quality of the results in the latter half of each month, as will be shown later. Also, any day in the training phase or forecast phase that did not have all member models present was

excluded. The skill score statistic used to validate the results was the equitable threat score (ETS), which is given below in equation 4.4.

(4.4)

The ETS statistic answers the question, “How well did the forecast "yes" events correspond to the observed "yes" events accounting for hits due to chance?” It measures the fraction of observed and/or forecast events that were correctly predicted, and also adjusts for any hits that may have been random chance. The ETS is often used in the verification of rainfall in NWP models because its "equitability" allows scores to be compared more fairly across different regimes.

## 4.2 Results

### 4.2.1 Spring 2006 Weather Pattern

The March 2006 weather pattern was dominated by troughing centered along the west coast, ridging centered in the Mississippi Valley, and troughing just off the east coast, as shown below in figure 4.2. As expected with this kind of upper level pattern, temperatures were generally near or below the 30 year climatological normals in the western U.S. and above the 30 year climatological normals in the central part of the country (figure 4.3). Along the east coast, temperatures were at near the normals. According to the National Climatic Data Center, March 2006 was the 37$^{th}$ warmest on record out for the U.S. as a whole dating back to 1895. The preliminary national average temperature was 6.7°C, or 0.8°C above the long term mean. Temperatures ranked warmer than average for 23 states and colder than average for only California (figure 4.4).

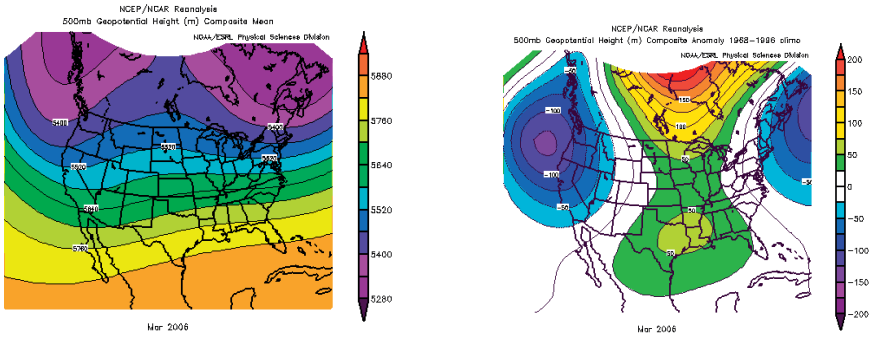


**Figure 4.2 March 2006 500 mb heights and height anomalies. Note the negative anomalies near the west coast and just off the east coast, indicating troughing, and the positive anomalies in the center part of the country, indicating ridging.**

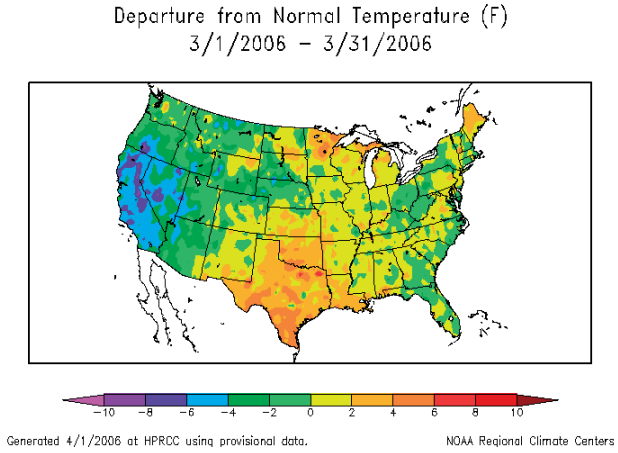


**Figure 4.3 March 2006 temperature anomalies across the U.S. compared to the 1961-1990 normals, courtesy of the High Plains Regional Climate Center**

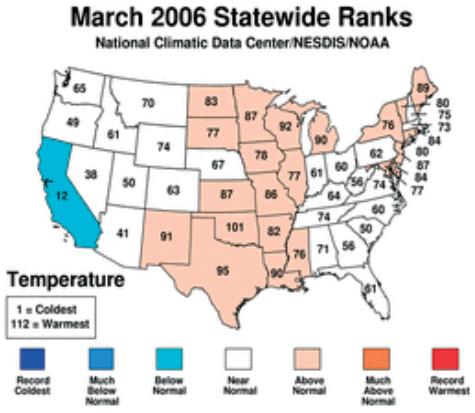


**Figure 4.4 March 2006 statewide temperature ranks using 1971-2000 normals, courtesy of NCDC**

Precipitation anomalies across the country were highly variable in March 2006, except for along the east coast where generally below normal precipitation was dominant (figure 4.5). According to the National Climatic Data Center, March had near average precipitation on the national level, ranking 47th driest since 1895. However, March was the driest on record for five eastern states, and twelve other eastern states were well below normal in precipitation. Eighteen states were wetter than average, including one state, Utah, with well above normal precipitation (figure 4.6).

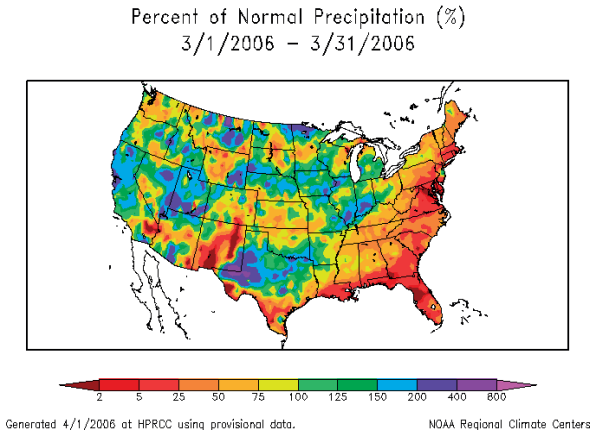


**Figure 4.5 March 2006 precipitation anomalies across the U.S. compared to the 1961-1990 normals, courtesy of the High Plains Regional Climate Center**

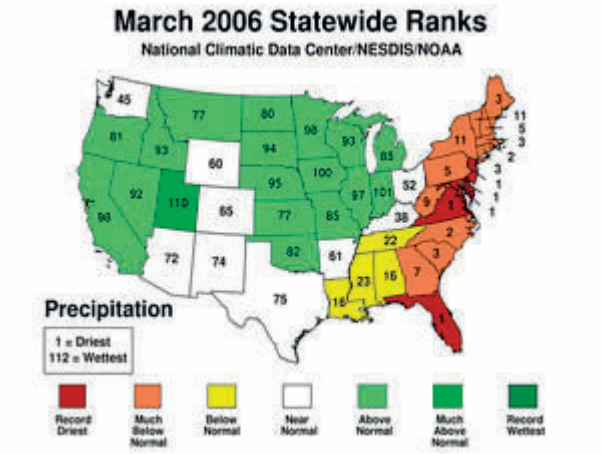


**Figure 4.6 March 2006 statewide precipitation ranks using 1971-2000 normals, courtesy of NCDC**

The April 2006 weather pattern was dominated by weak troughing along both coasts of the U.S. and ridging in between (figure 4.7). Height anomalies were generally positive for much of the country, and temperatures were at record levels (figure 4.8).

According to the National Climatic Data Center, April 2006 was the warmest April on record for the U.S. dating back to 1895. The preliminary nationally averaged temperature was 13.6°C, which was 2.5°C above the 1901-2000 mean. All states were near or above their long-term means. Temperatures were the warmest on record for two states, Texas and Oklahoma, and above to much above normal for nearly all other states (figure 4.9).

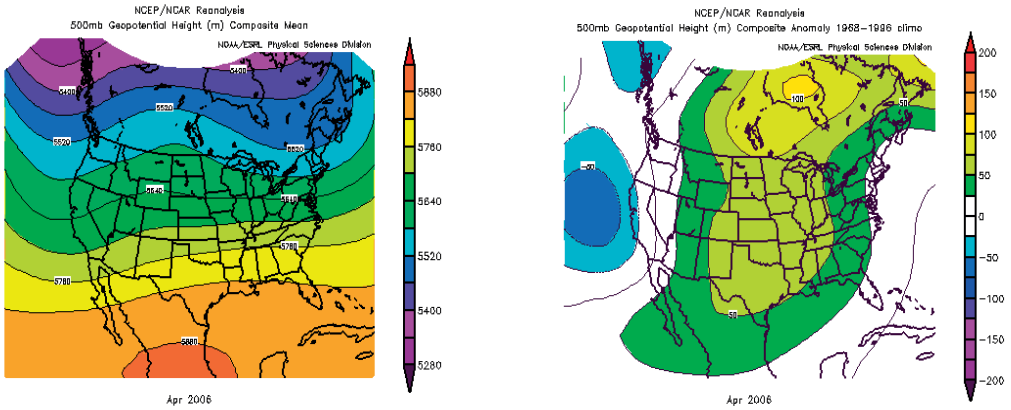


**Figure 4.7 As in figure 4.2 but for April.**

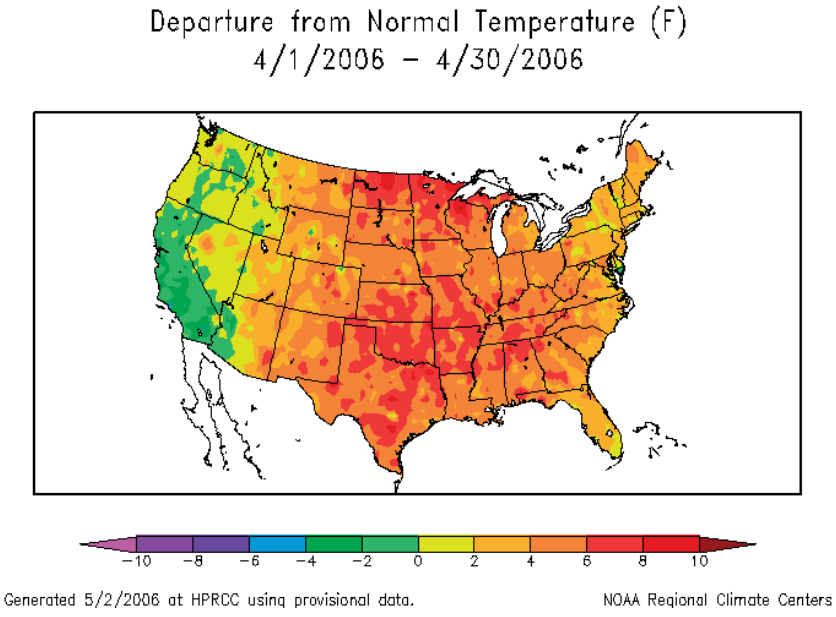


**Figure 4.8 As in figure 4.3 but for April.**

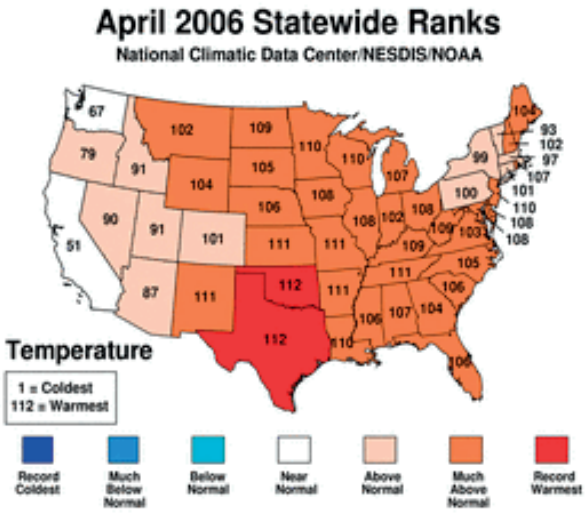


**Figure 4.9 As in figure 4.4 but for April**

The precipitation anomaly map for April is shown below in figure 4.10. The west-central U.S. was generally above normal, while the northwestern, southwestern, and eastern Rockies were generally below normal. According to the National Climatic Data Center, April 2006 had near average precipitation nationally, ranking 45th wettest since 1895. Four western states had much above normal rainfall, while Colorado saw much below normal rainfall (figure 4.11).

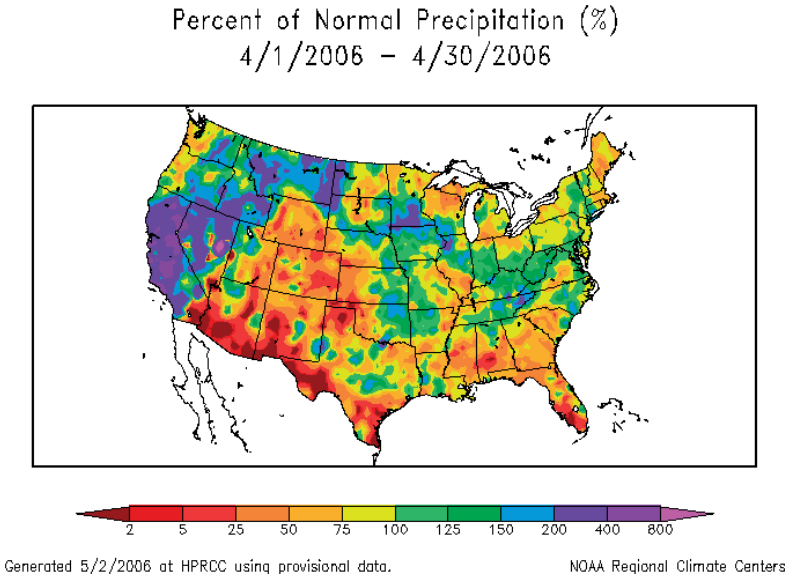


**Figure 4.10 As in figure 4.5 but for April**

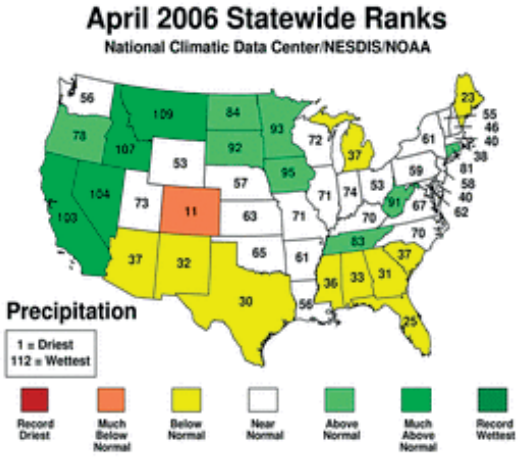


**Figure 4.11 As in figure 4.6 but for April**

The May 2006 weather pattern was dominated by ridging in the western U.S. and troughing in the eastern U.S., as shown below in figure 4.12. As expected with this kind of upper level pattern, temperature were generally at or above the 30 year climatological normals in the western U.S. and at or below the 30 year climatological normals in the eastern U.S (figure 4.13). According to the National Climatic Data Center, May 2006 was the fifth warmest May on record nationally since 1895. The preliminary nationally averaged temperature was 17.6°C, which was 1.5°C above the 1901-2000 mean for the month of May. Statewide temperatures ranked below normal in four states in the eastern U.S. and much above normal throughout all of the southwestern U.S. (figure 4.14).

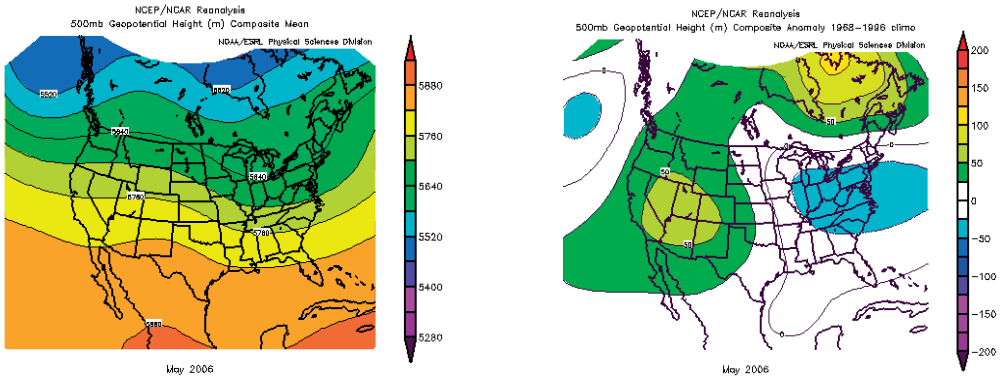


**Figure 4.12 As in figure 4.2 but for May.**

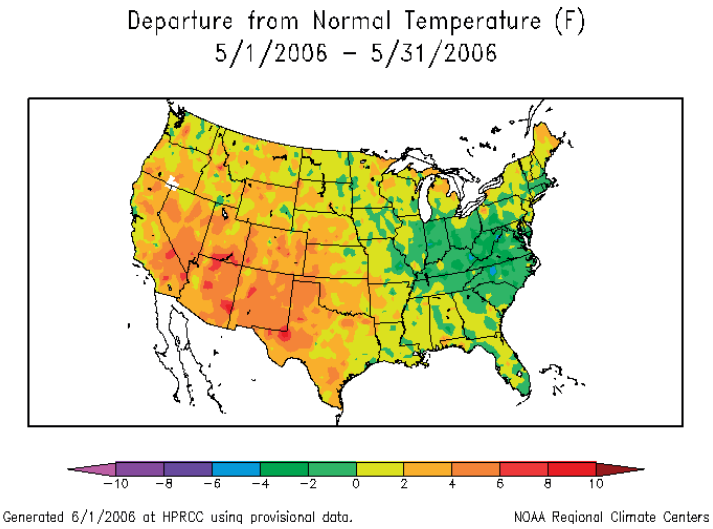


**Figure 4.13 As in figure 4.3 but for May**

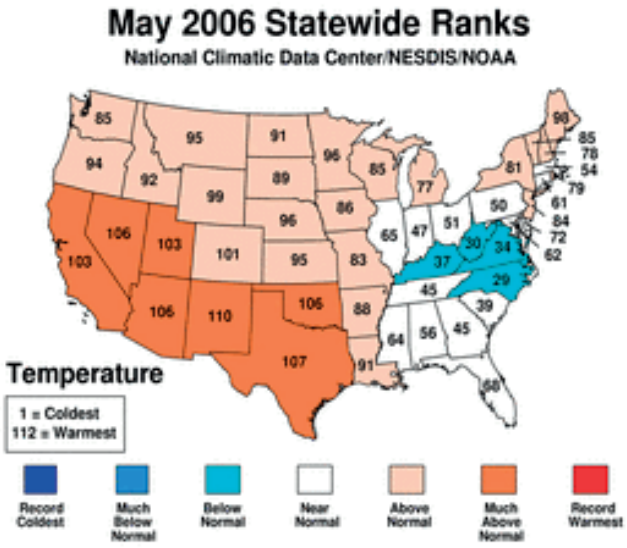


**Figure 4.14 As in figure 4.4 but for May**

The precipitation anomaly map for May is shown below in figure 4.15. Much of the country featured below normal precipitation, with some notable exceptions in New England, the Great Lakes region, and the California coast. According to the National Climatic Data Center, May 2006 ranked 17th driest in the 1895-2006 record. Nebraska and Iowa received well below normal precipitation, while several New England states received well above normal precipitation (figure 4.16).

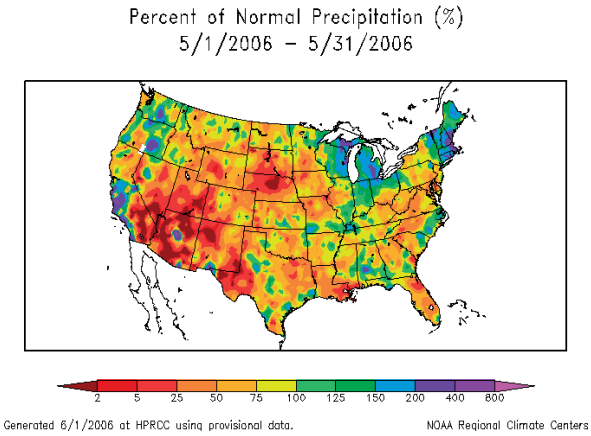


**Figure 4.15 As in figure 4.5 but for May**

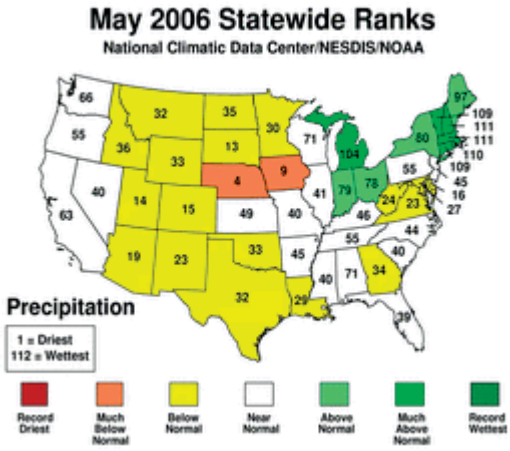


**Figure 4.16 As in figure 4.6 but for May**

### 4.2.2 Spring 2006 Temperature Forecast Results

In this section, the superensemble temperature forecast results are displayed. Daily temperature forecasts were made out to 60 hours at 3-hourly time steps during the spring months of March, April, and May of 2006. The figures that follow below show the mean absolute error (mae), root mean square error (rmse), and mean error (bias) of the member models as well as the superensemble for all days available. There were 29 days missing during this three month period, either due to corrupted member model data or a lack of member model availability for that particular day, leaving 63 days available.

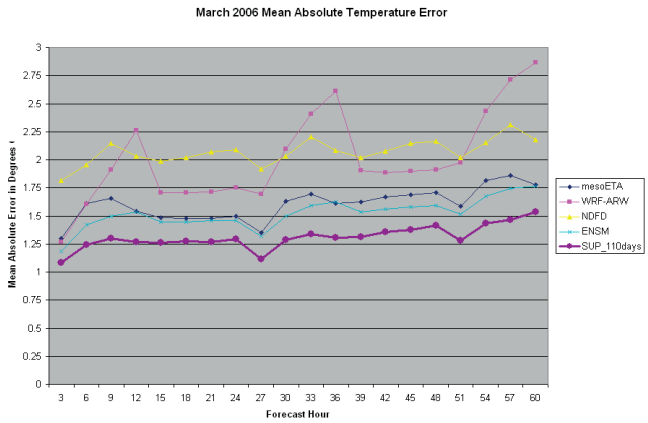


**Figure 4.17 MAE for temperature in degrees Celsius for the member models and superensemble for March 2006**

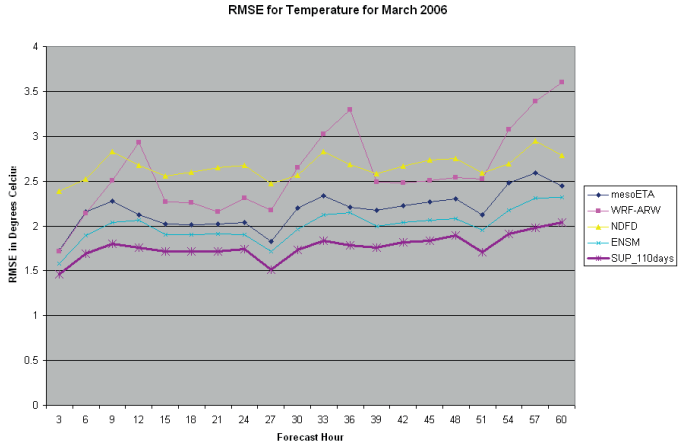


**Figure 4.18 As in figure 4.17 but for RMSE**

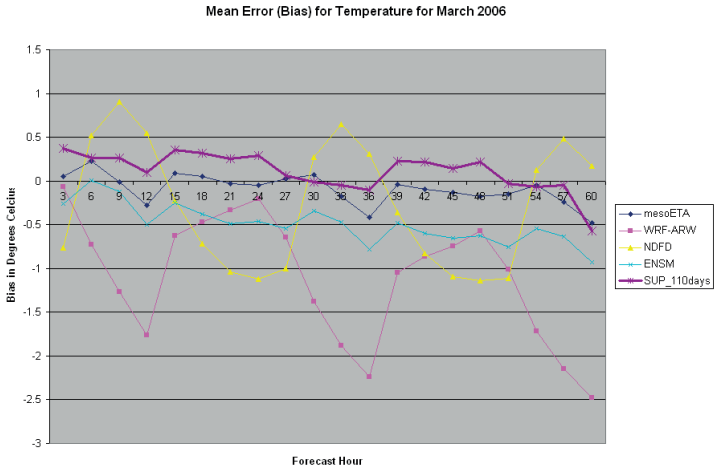


**Figure 4.19 As in figure 4.17 but for bias**

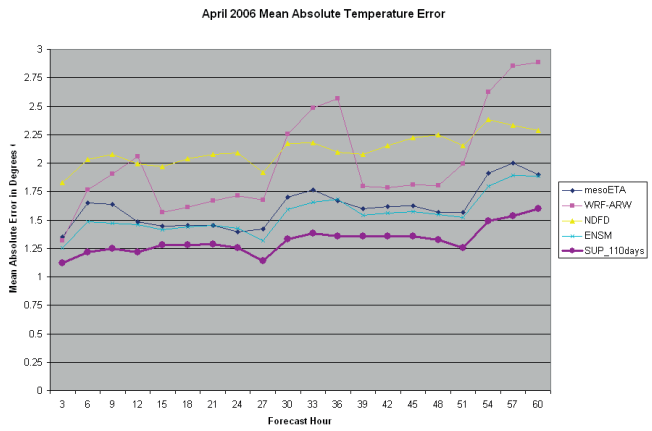


**Figure 4.20 As in figure 4.17 but for April**

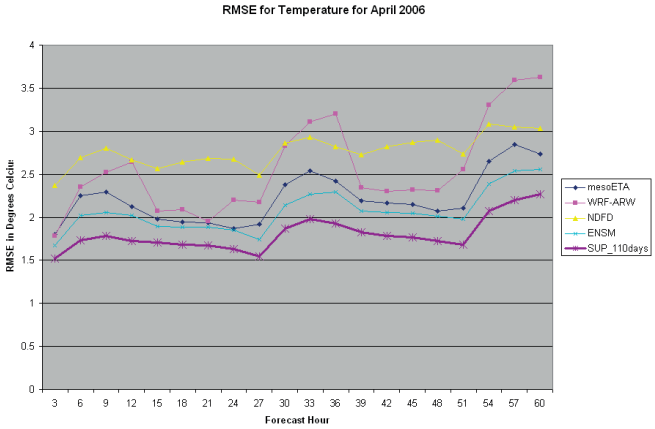


**Figure 4.21 As in figure 4.20 but for RMSE**

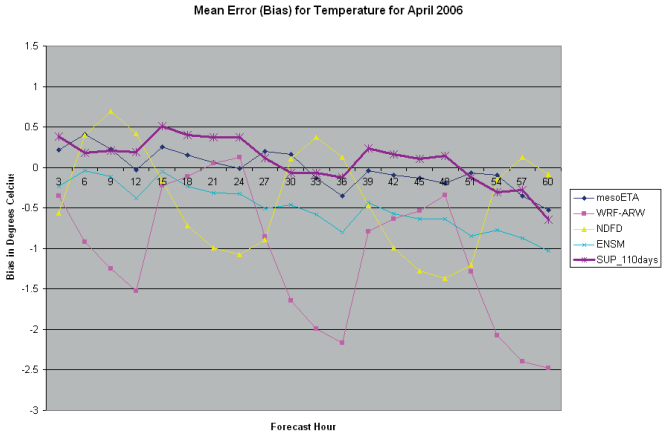


**Figure 4.22 As in figure 4.20 but for bias**

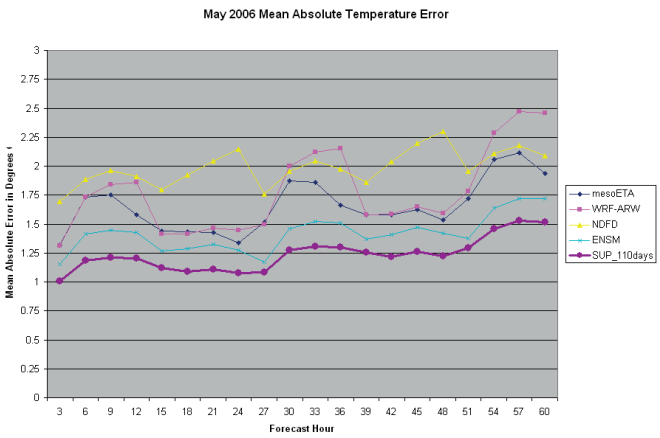


**Figure 4.23 As in figure 4.17 but for May**

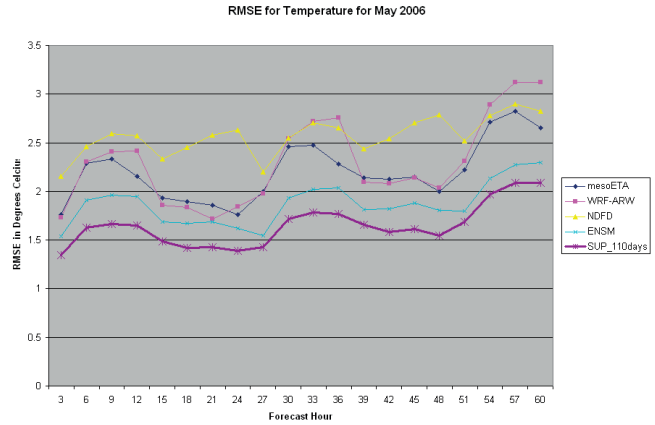


**Figure 4.24 As in figure 4.23 but for RMSE**

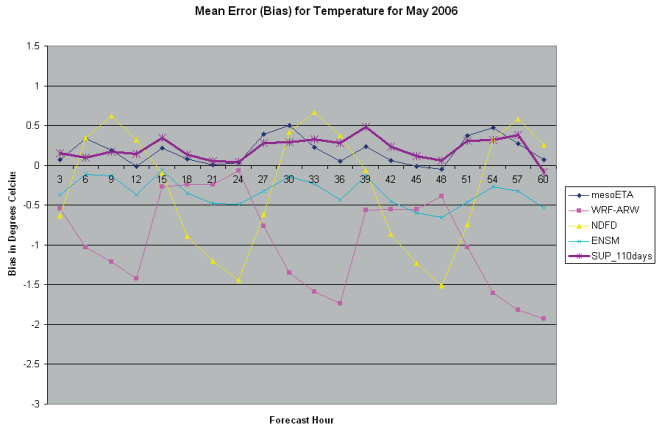


**Figure 4.25 As in figure 4.23 but for bias**

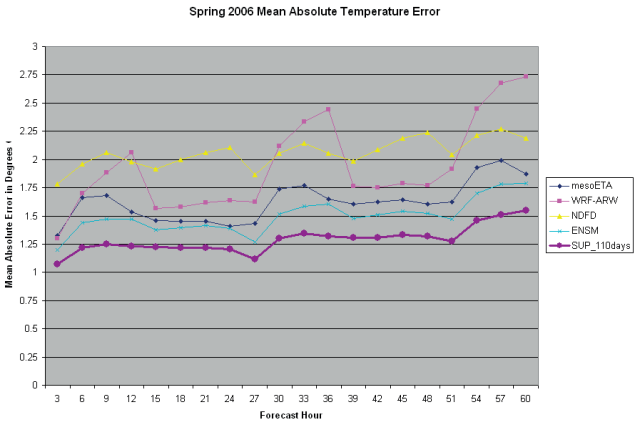


**Figure 4.26 As in figure 4.17 but for Spring**

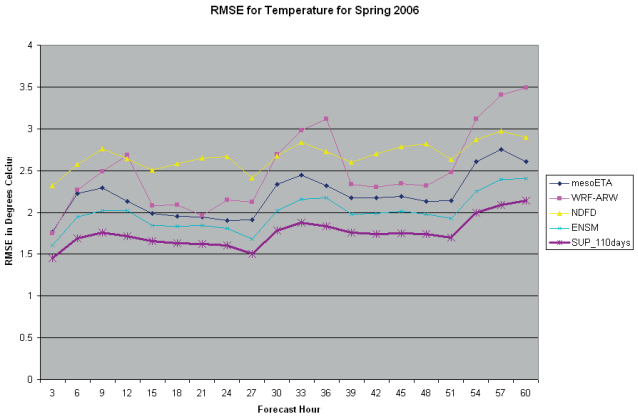


**Figure 4.27 As in figure 4.26 but for RMSE**

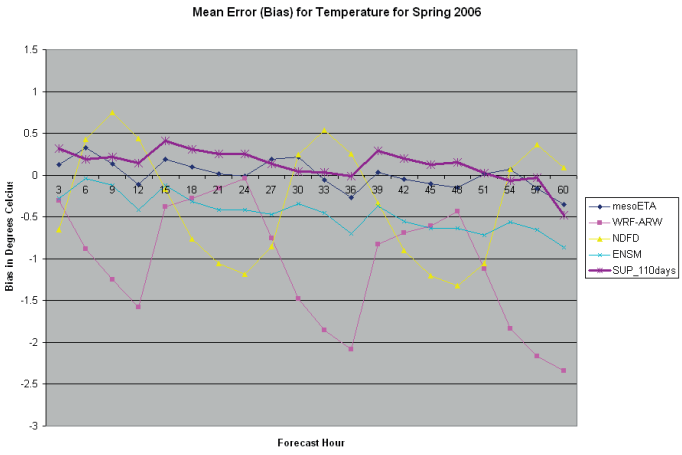


**Figure 4.28 As in figure 4.26 but for bias**

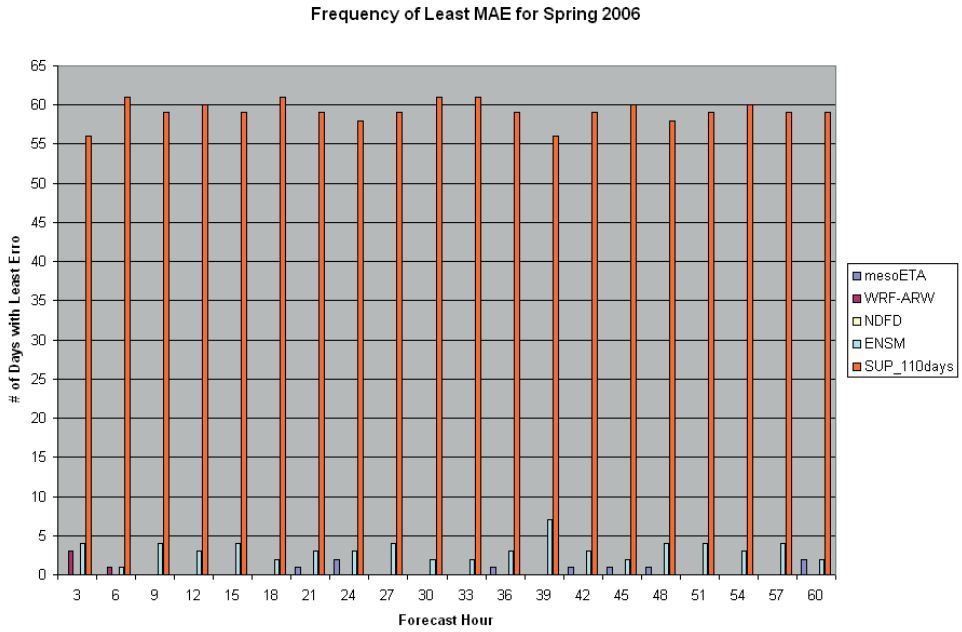


**Figure 4.29 The number of days during March, April, and May for each forecast hour that each member model and the superensemble had the lowest MAE. The superensemble consistently had the lowest MAE during every forecast hour for the time period studied. There were 63 days available for study.**

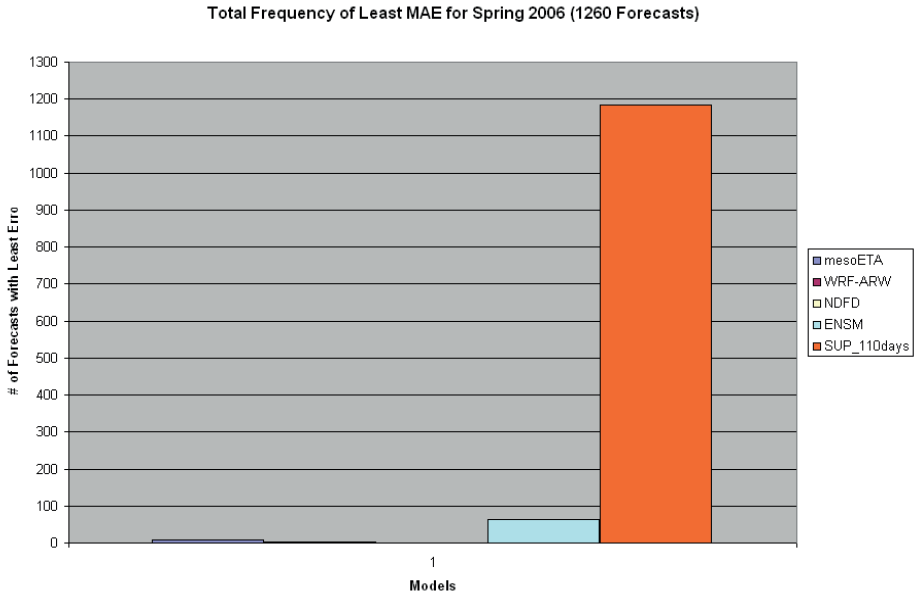


**Figure 4.30 The total number of forecasts during March, April, and May that each member model and the superensemble had the lowest MAE. The time step was not considered here, which means there were 63 days * 20 forecasts/day = 1260 forecasts available. The superensemble had the lowest MAE in 1183 forecasts, or 93.9% of the time.**

These figures clearly show the superensemble having superior MAE and RMSE statistics as well as very strong bias scores, both from month to month and overall. The WRF-ARW experienced a consistent cold bias throughout the entire forecast period, which was most prevalent during the afternoon and evening hours of the forecast, i.e. hours 6, 9, 12, 30, 33, 36, 54, 57, and 60. The ETA exhibited very little bias. However, its other error MAE and RMSE scores exhibited considerably higher error than those of the superensemble. The NDFD exhibited a very systematic, sinusoidal bias of being too warm overall during the daylight hours and too cold overall during the nighttime hours. This is consistent with over predicting the magnitude of the regular diurnal cycle. The ENSM was a blend of these biases with a slightly better MAE and RMSE score overall than the other member models, although the superensemble was still lower. The NDFD MAE and RMS errors were surprisingly high, considering that the human forecasters who create the NDFD use the ETA model as part of their guidance, which showed lesser errors than that of the NDFD. This was not an expected result, and it prompted further investigation into possible causes for this result.

The primary cause of this result is likely due to the verification method employed in this study. This study used the NARR 2-meter temperatures as the "observed" for both the training and verification phases. While the NARR has the advantage of being a fully gridded dataset with temperature values at every gridpoint in the domain, the disadvantage is that the NARR is still a computer re-analysis, and as such it does not always represent the real temperature data from individual observing stations across the United States. The National Weather Service publishes verification of the NDFD using 1222 observing sites across the U.S. While one disadvantage over the NARR is that the entire grid cannot be verified using that method, the advantage is that the verification method does employ actual station observations. MAE and bias of the NDFD were compared using the NWS verification method and verification using the NARR. Results are shown below for each forecast hour in figures 4.30 through 4.33. In addition, six stations were chosen across the country to study the verification patterns of all the models plus the NWS Model Output Statistics (MOS) product discussed earlier in Chapter 3. Real data from those stations was taken for comparison to the NARR gridpoints where those stations are located.

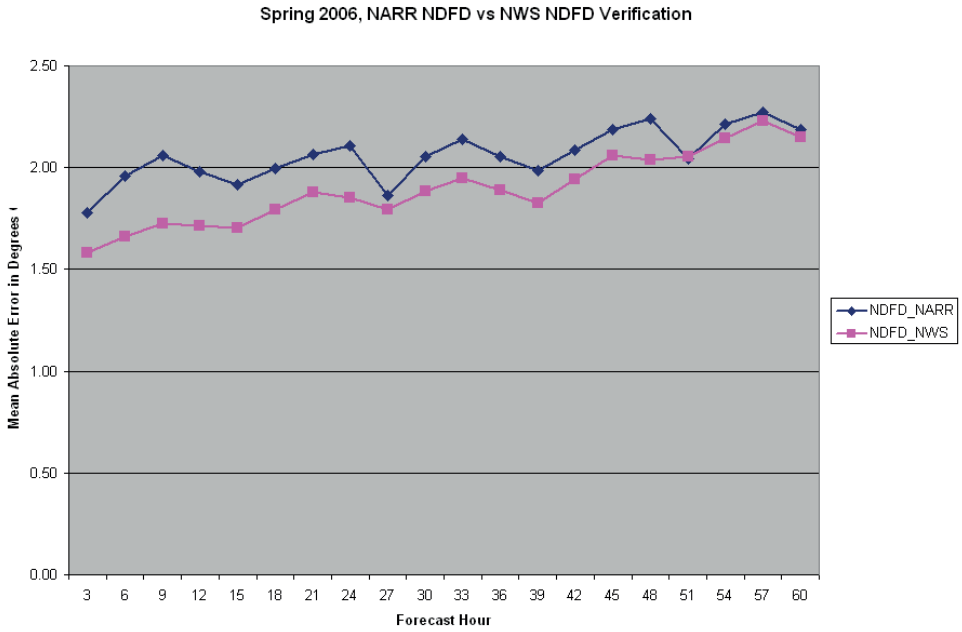


**Figure 4.31 Comparison of MAE of the NDFD using the NARR grid as the "observed" and using the NWS station data from 1222 sites as the "observed". NARR verification shows a higher MAE at all forecast hours except hour 51.**

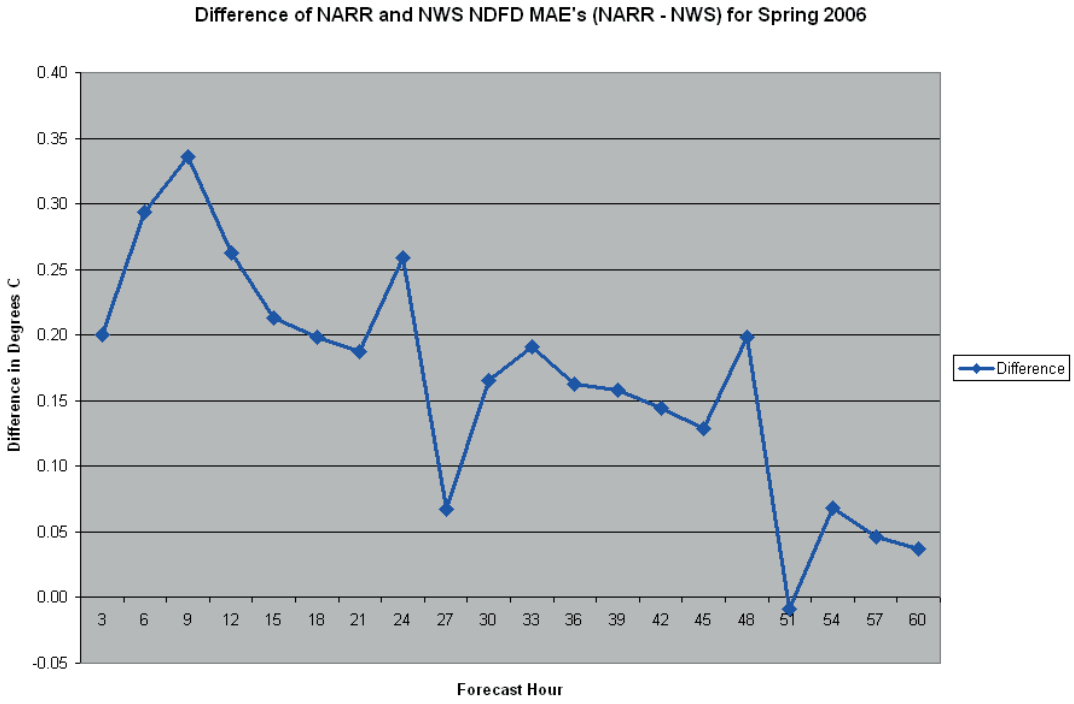


**Figure 4.32 Similar to figure 4.30, except the MAE's were subtracted (NARR-NWS). Note the steady decline in the difference between MAE's as the forecast hour increases.**

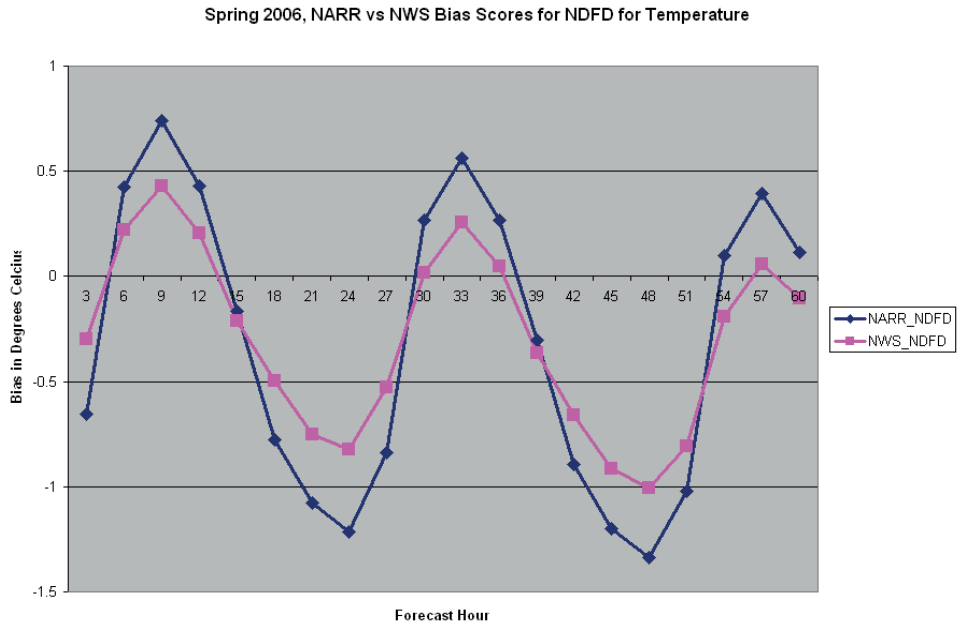


**Figure 4.33 Comparison of bias of the NDFD using the NARR grid as the "observed" and using the NWS station data from 1222 sites as the "observed". NARR verification shows a higher bias during all daytime forecast hours and shows a lower bias during all the nighttime forecast hours when compared to the NWS verification. Forecasts started at 12 UTC so hour 3 would be valid at 15 UTC, etc.**

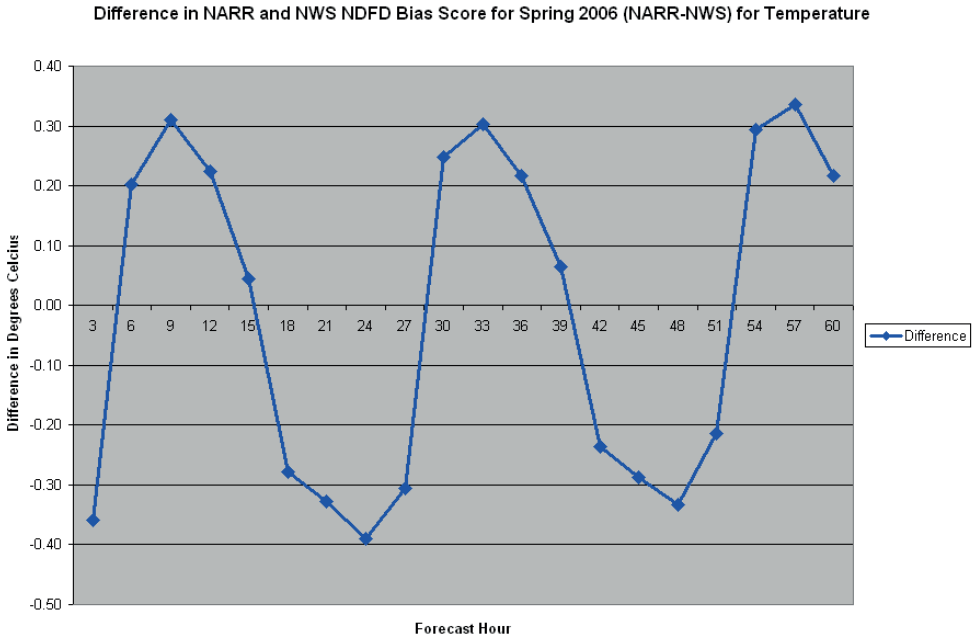


**Figure 4.34 Similar to figure 4.32, except the biases were subtracted (NARR-NWS). This further illustrates the sinusoidal, persistent daytime/nighttime differences in verification techniques.**

These charts show that while the MAE values do converge in the later forecast hours, the bias values exhibit very systematic differences. These differences in biases are likely caused by differences in the treatment of the diurnal cycle. The NARR re-analysis itself may have some systematic bias in the diurnal cycle, as evidenced by being too high during the daytime hours and too low during the overnight hours, suggesting an overall over prediction in magnitude when compared to actual station observations.

Further investigation into the differences in the NARR analysis vs actual station data was done by utilizing temperature data from six different stations across the United States: Tallahassee, FL; Baltimore, MD; Hibbing, MN; Lincoln, NE; Fresno, CA; and Lovelock, NV. A single forecast hour was chosen at random for each of the stations. For this forecast hour, all member models, the superensemble, the NWS MOS products, and the NARR were all compared to the reported station temperatures during the period using both MAE and bias. Results for the six stations are shown in the following figures. In all of the plots, the superensemble is the middle, purple bar. All bars to the left of the superensemble were included as member models, while all bars to the right were not included in the superensemble as member models. Three different ensemble means were calculated and shown in addition to the superensemble: the ensemble mean of the original member models (also a member model, shown immediately left of the superensemble), the ensemble mean of the MOS guidance (shown as the third bar from the right in the left plot…fourth bar from the right in the right plot), and the ensemble mean of all models and MOS shown except superensemble and NARR (shown as the second bar from the right in the left plot…third bar from the right in the right plot).

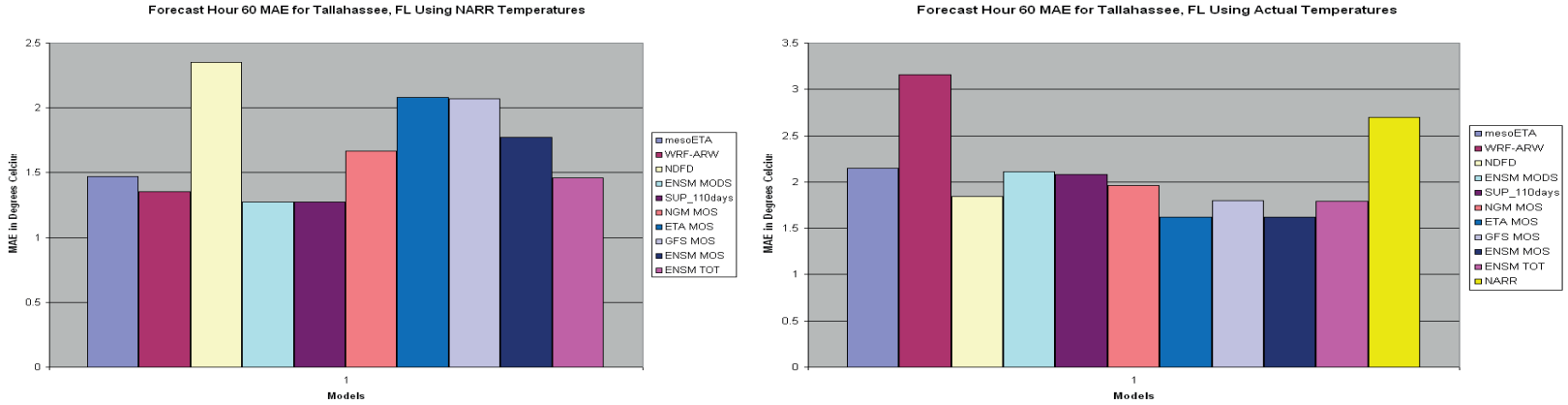


**Figure 4.35 The 60 hour forecast MAE results for Tallahassee, FL from both the NARR dataset (left graph) and the station data (right graph).**

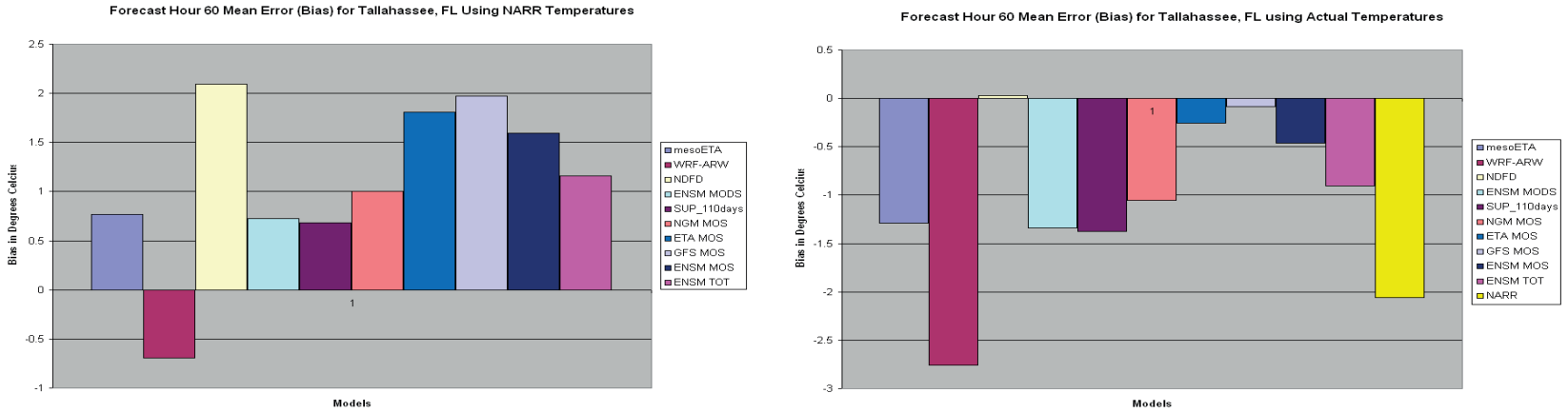


**Figure 4.36 The 60 hour forecast bias results for Tallahassee, FL from both the NARR dataset (left graph) and the station data (right graph).**

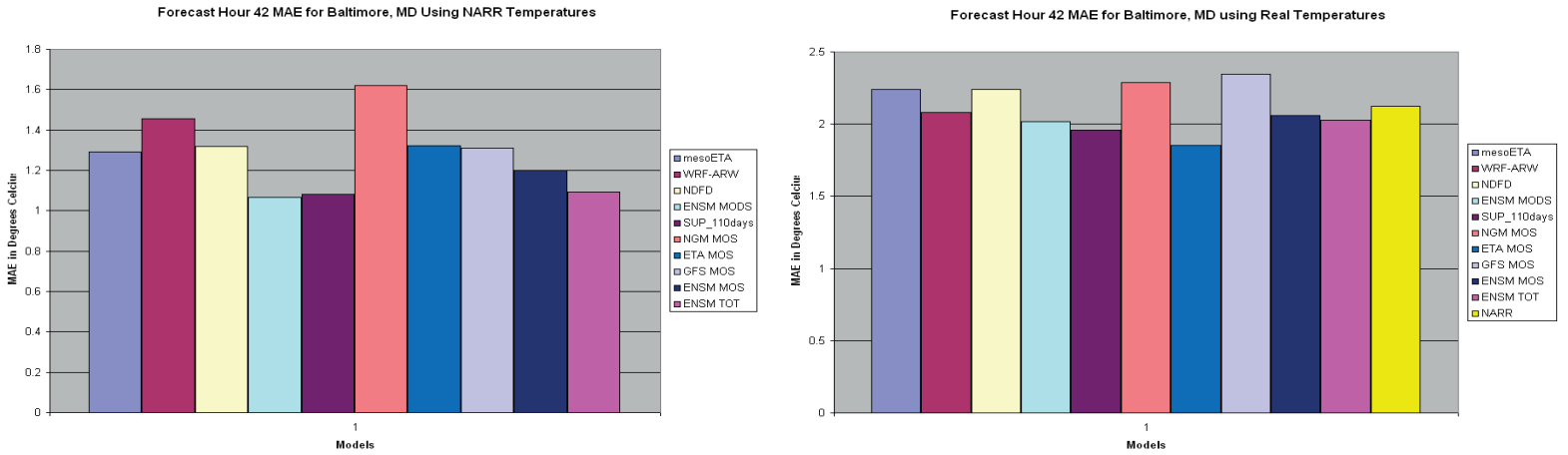


**Figure 4.37 The 42 hour forecast MAE results for Baltimore, MD from both the NARR dataset (left graph) and the station data (right graph).**

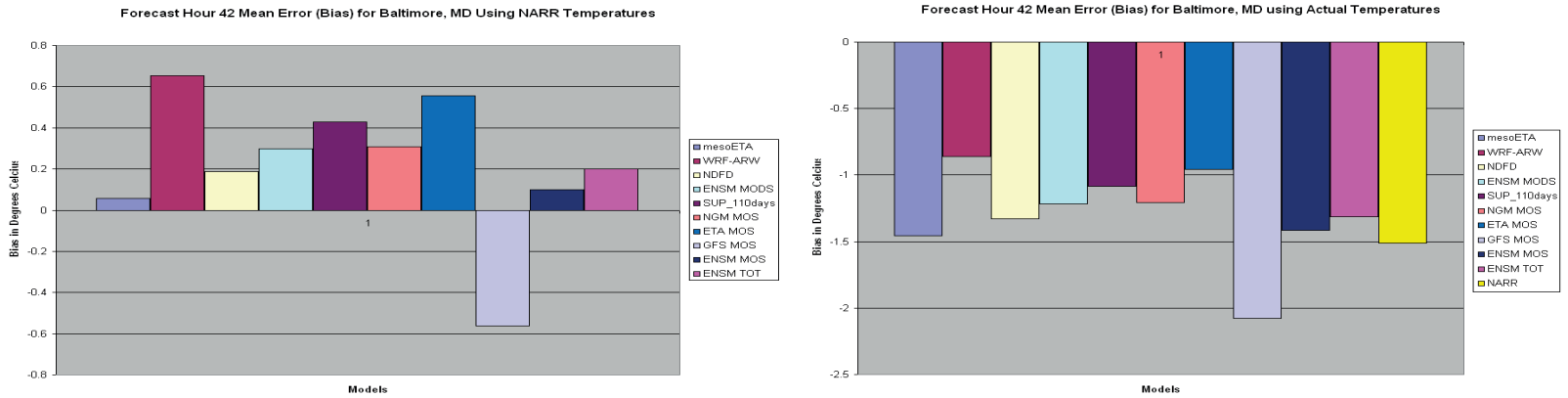


**Figure 4.38 The 42 hour forecast bias results for Baltimore, MD from both the NARR dataset (left graph) and the station data (right graph).**

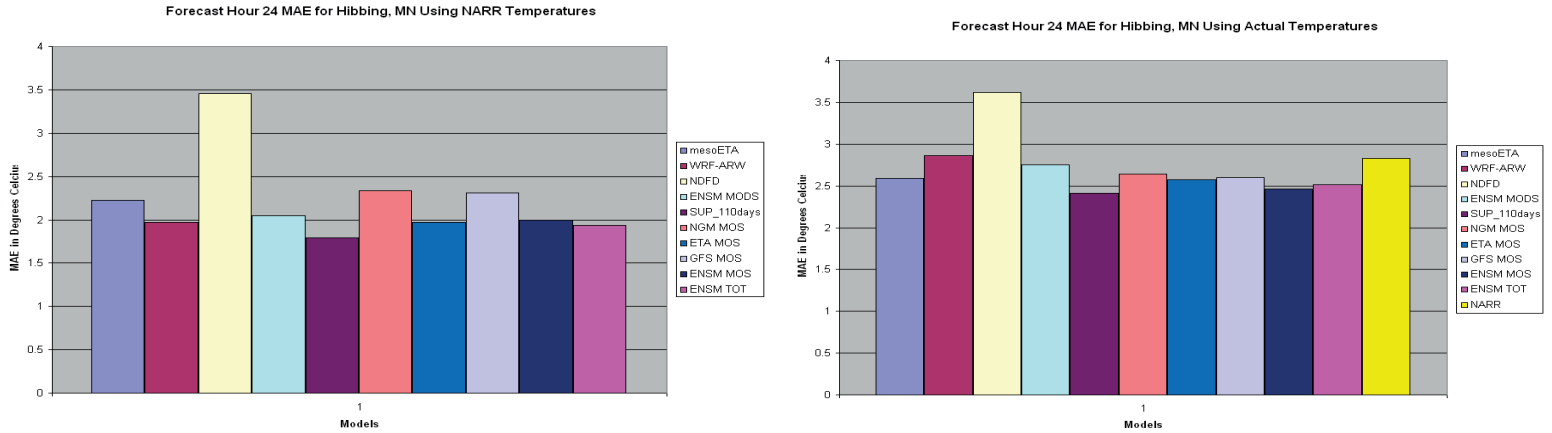


**Figure 4.39 The 24 hour forecast MAE for Hibbing, MN from both the NARR dataset (left graph) and the station data (right graph).**

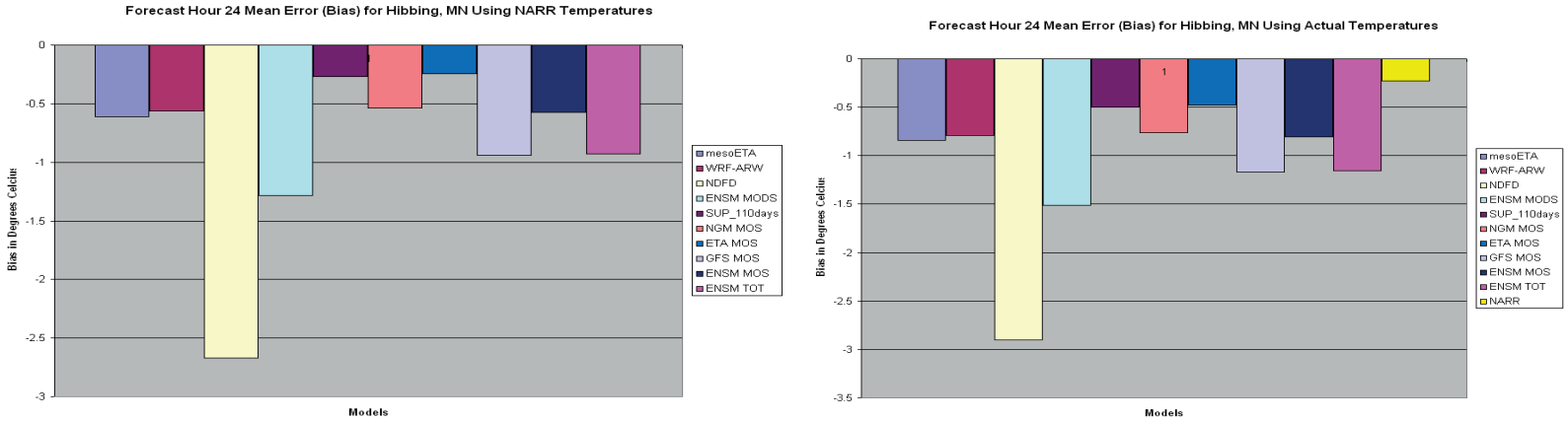


**Figure 4.40 The 24 hour forecast bias for Hibbing, MN from both the NARR dataset (left graph) and the station data (right graph).**

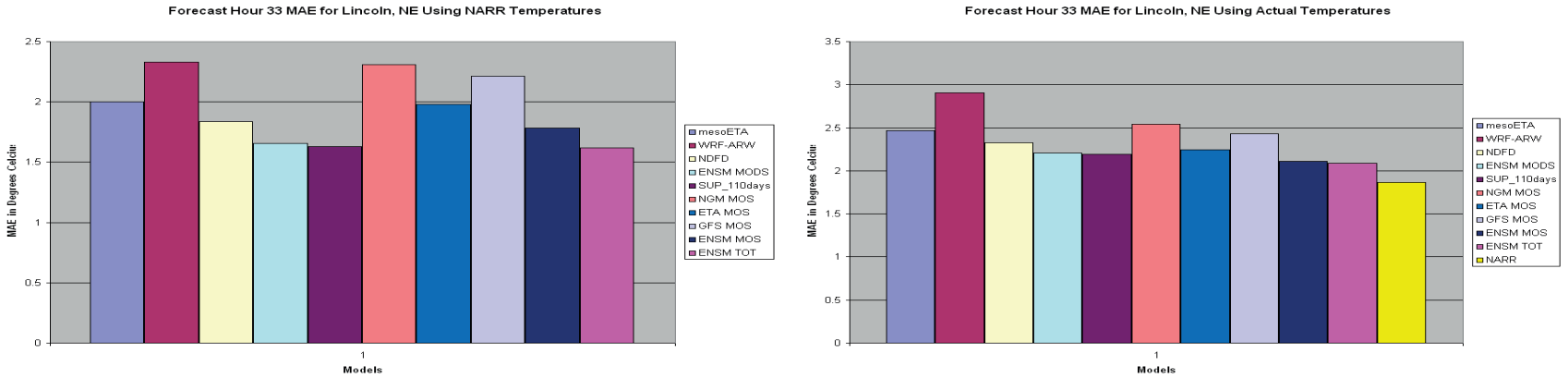


**Figure 4.41 The 33 hour forecast MAE for Lincoln, NE from both the NARR dataset (left graph) and the station data (right graph).**

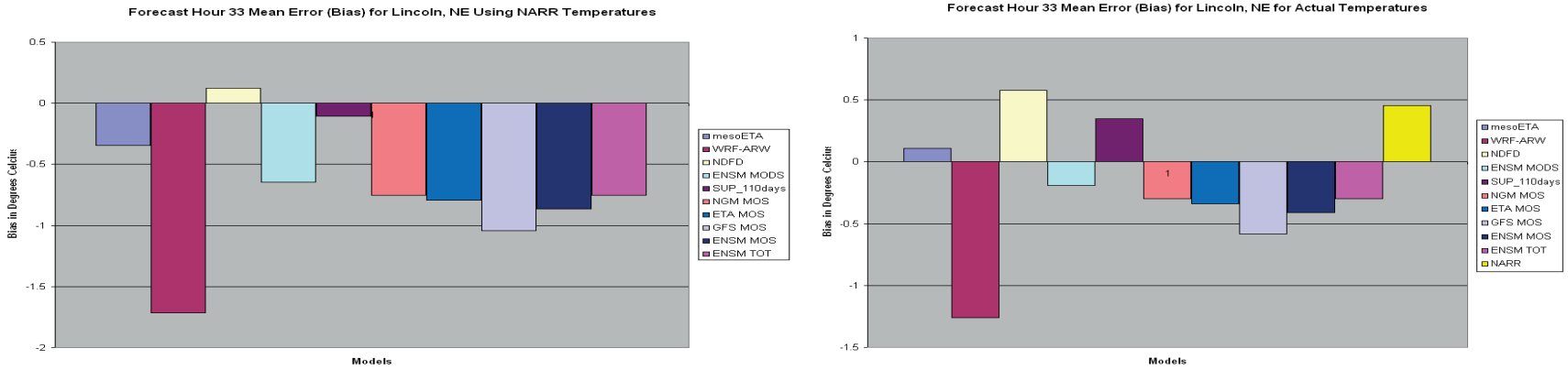


**Figure 4.42 The 33 hour forecast bias for Lincoln, NE from both the NARR dataset (left graph) and the station data (right graph).**

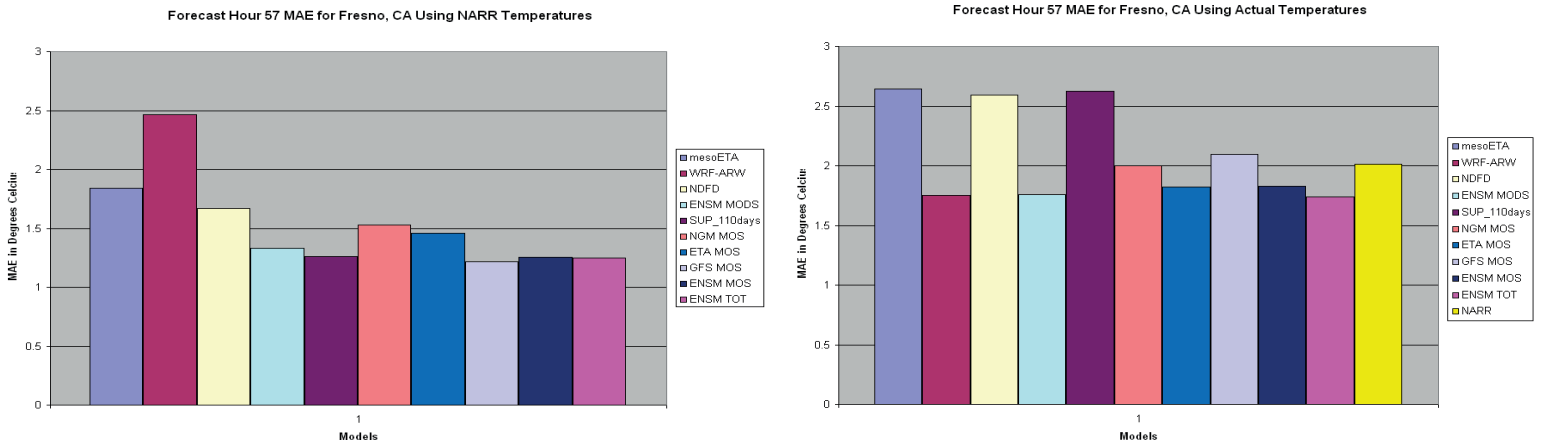


**Figure 4.43 The 57 hour forecast MAE for Fresno, CA from both the NARR dataset (left graph) and the station data (right graph).**

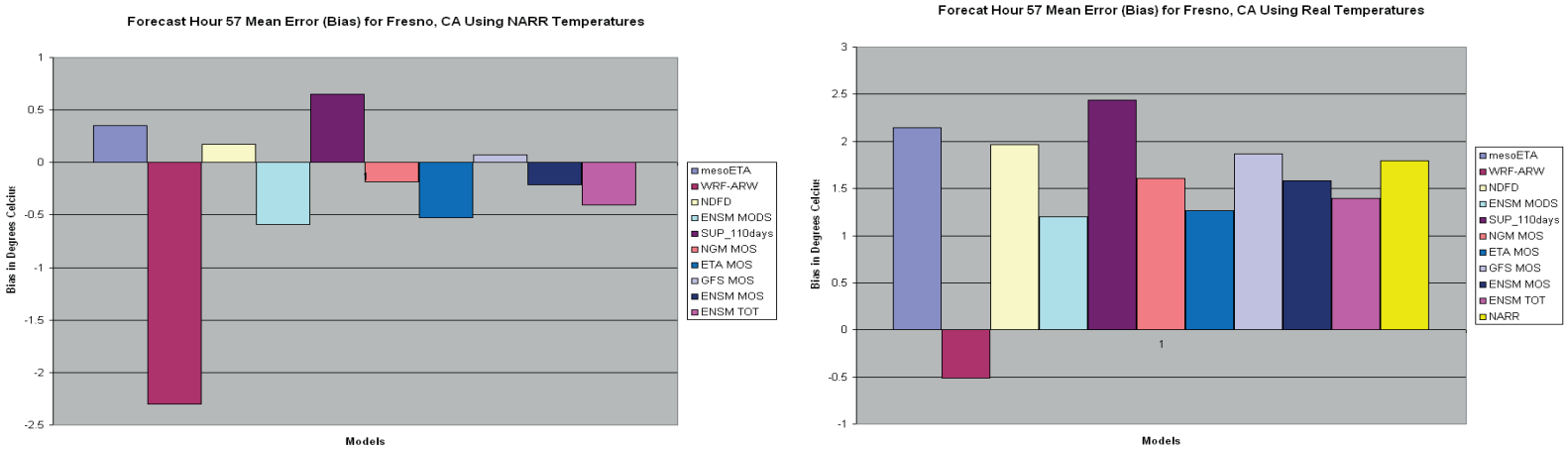


**Figure 4.44 The 57 hour bias for Fresno, CA from both the NARR dataset (left graph) and the station data (right graph).**

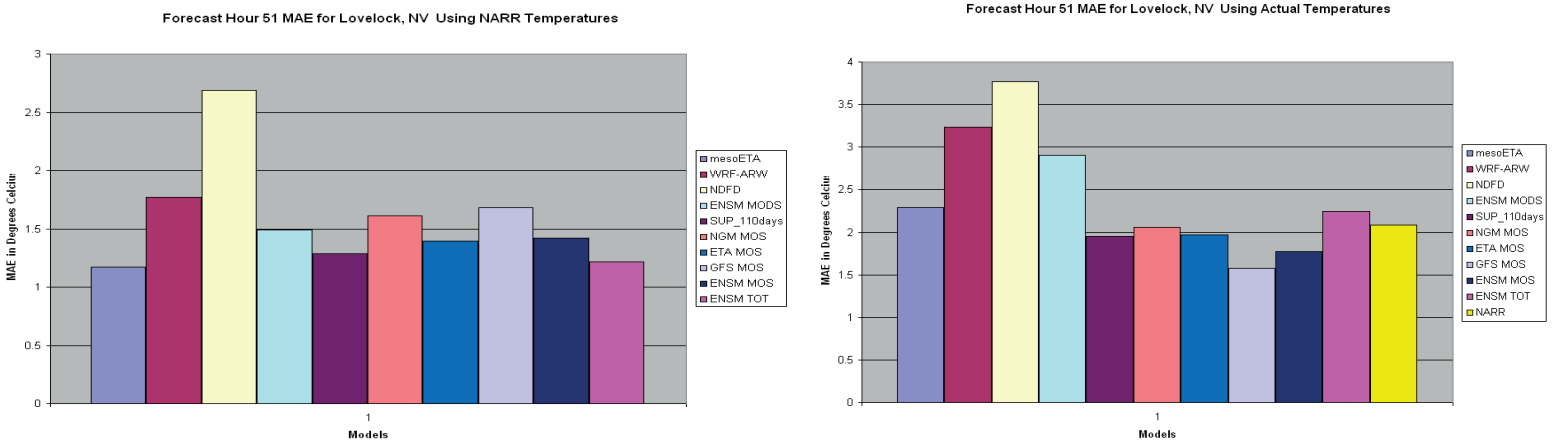


**Figure 4.45 The 51 hour MAE for Lovelock, NV from both the NARR dataset (left graph) and the station data (right graph).**

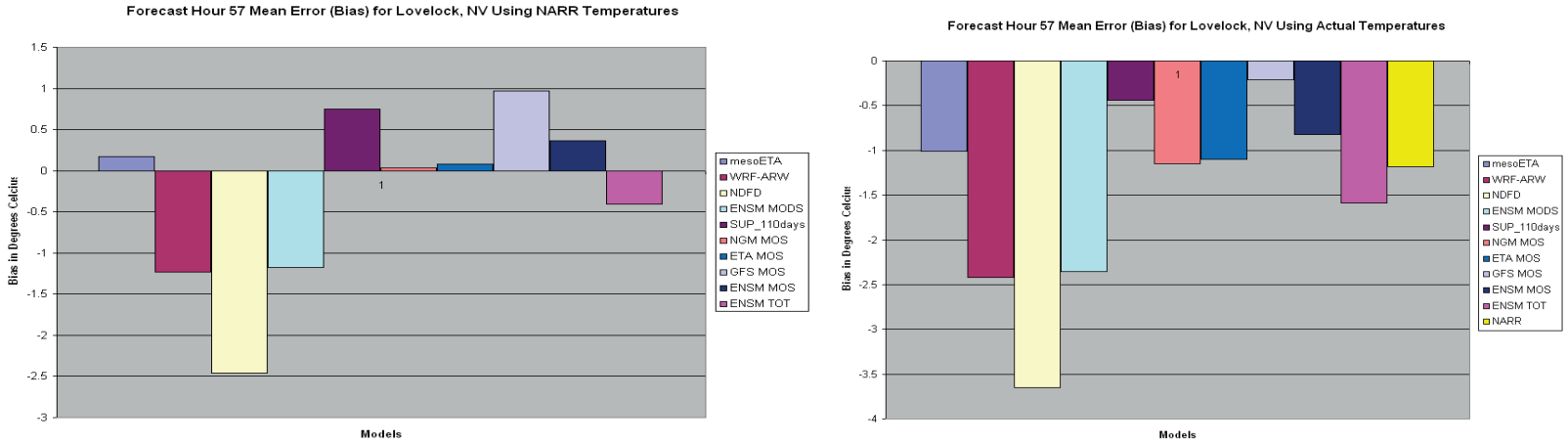


**Figure 4.46 The 51 hour bias for Lovelock, NV from both the NARR dataset (left graph) and the station data (right graph).**

There are several things to note from each of these cases. In every case, the MAE in general was lower for all models using the NARR as the observed dataset. For the MOS, the scores were also lower in every case using the NARR, except for Tallahassee, where using the actual station temperature produced a lower MAE for the MOS. Since the MOS uses a form of bias correction to the raw model output using actual station information, it is interesting that in most cases studied, the actual station temperature data produced a higher MAE than the model derived re-analysis data of the NARR. It was expected that the MOS would exhibit a lower MAE using the actual station data since that data is used for its bias correction. Instead, a lower MAE was obtained using NARR data. The model error statistics performed as expected. Using the NARR produced lower a MAE for the models, and this was expected since the NARR is a re-analysis that uses some model information. With the exception of Tallahassee, the NDFD also exhibited a lower MAE using the NARR. For Tallahassee, the NDFD MAE was significantly lower using the actual station data, which could indicate that forecasters there have a good handle on local conditions at TLH. The superensemble skill is tied heavily to the bias of the NARR when compared to the real station data. While the superensemble performed very well for the individual stations studied using the NARR analysis (lowest or second lowest MAE amongst its member models in every case), its performance was mixed when using the station data. When the NARR bias is closer to zero such as in the case of Hibbing, Lincoln, and Lovelock, then the superensemble performance using the station data is very good. When the NARR has a significant bias such as in the case of Fresno, then the superensemble does rather poorly when compared to the station data. However, it is important to note that the superensemble did very well for Fresno when using the NARR to verify. This behavior in verification scores is likely due to the fact that the superensemble used the NARR data as the "observed" during its training phase, and thus it indirectly inherits whatever biases that the NARR may contain.

A final superensemble experiment was done with temperature to look at how sensitive the results are to the number of training days used. In all of the above figures, 110 days of training were used in the superensemble. In the following three figures of MAE, RMSE, and bias, the number of training days was varied between 80 and 140 days

to see if there was any noticeable difference in skill. Very little difference was found, as illustrated in the following figures.

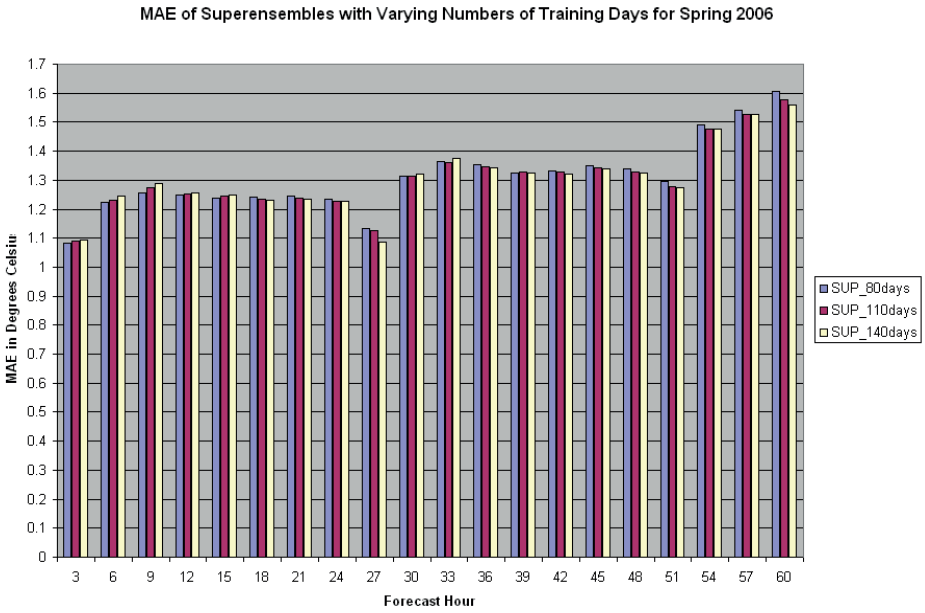


**Figure 4.47 MAE for temperature in degrees Celsius for the superensemble at 80, 110, and 140 days of training for March, April, and May 2006**

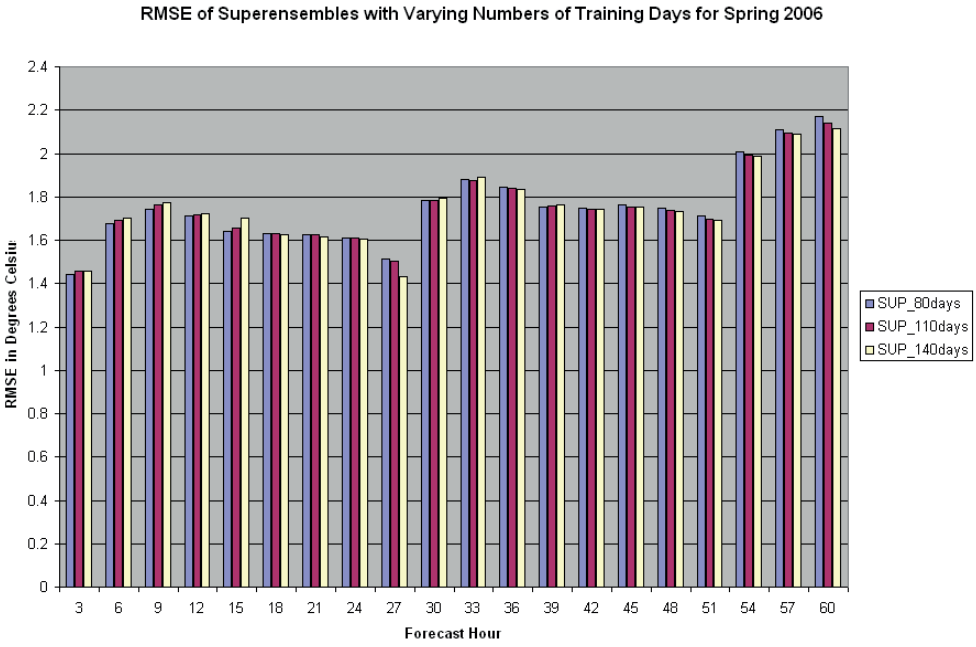


**Figure 4.48 As in figure 4.47 but for RMSE**

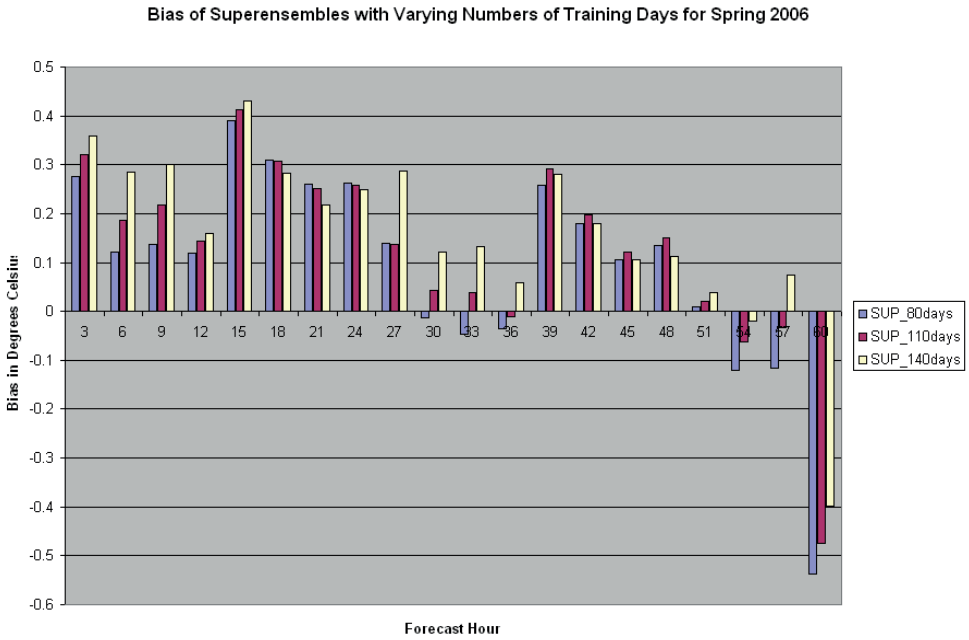


**Figure 4.49 As in figure 4.47 but for bias**

There is a slight tendency for the 140 day superensemble to perform best at later forecast hours, and a slight tendency for the 80 day superensemble to perform best at the earlier forecast hours. However, the improvement is on the order of hundredths of a degree in either direction, which makes it seem rather insignificant.

### 4.2.3 June, July, August, September Weather Pattern

The June 2006 weather pattern was dominated by ridging in the western U.S. and near Bermuda and weak troughing on the east coast and out into the Pacific Ocean (figure 4.49). As expected with this kind of upper level pattern, temperatures were generally above to much above the 30 year climatological normals in the western half of the U.S. and near the 30 year climatological normals in the eastern U.S (figure 4.50). According to the National Climatic Data Center, June 2006 was the second warmest June in the 1895-2006 record. The preliminary nationally averaged temperature was 22.1°C, which was 1.4°C above the 1901-2000 June mean. Temperatures ranked much above average for 13 states and below average for 5 states (figure 4.51).

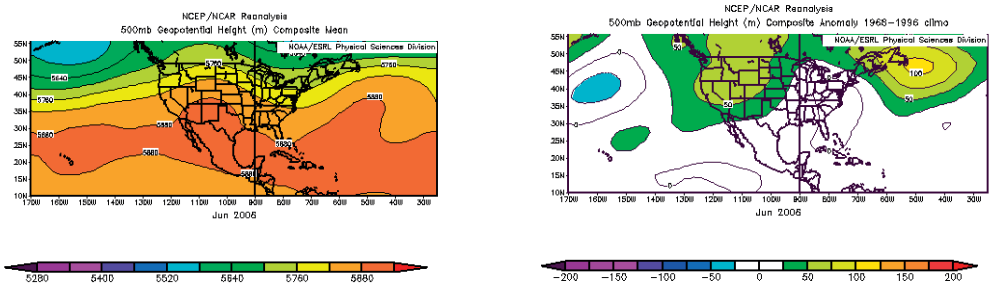


**Figure 4.50 As in figure 4.2 but for June**

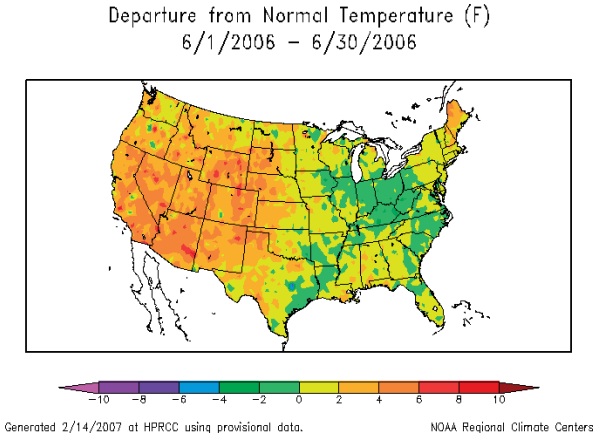


**Figure 4.51 As in figure 4.3 but for June**

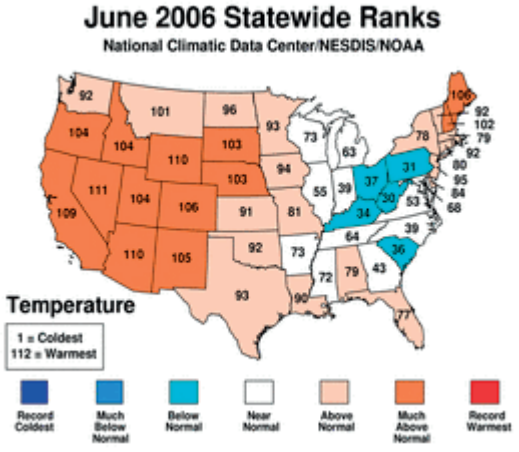


**Figure 4.52 As in figure 4.4 but for June**

The precipitation anomaly map for June is shown below in figure 4.52. Precipitation across the country was generally above normal in the eastern, southwestern, and northwestern U.S. and near to below normal elsewhere. According to the National Climatic Data Center, June had well below normal precipitation for the country as a whole, ranking 25th driest sense 1895. However, by region, the northeastern U.S. had its second wettest June on record (figure 4.53).

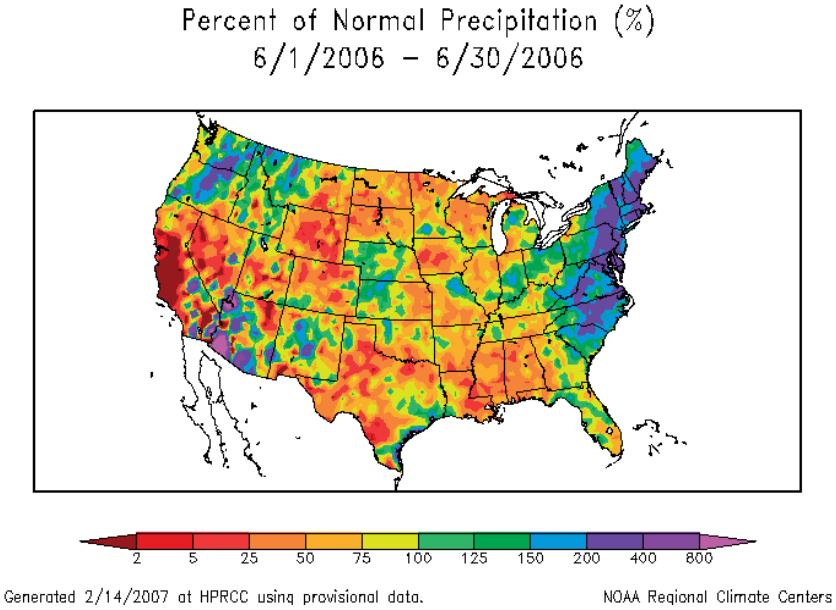


**Figure 4.53 As in figure 4.5 but for June**

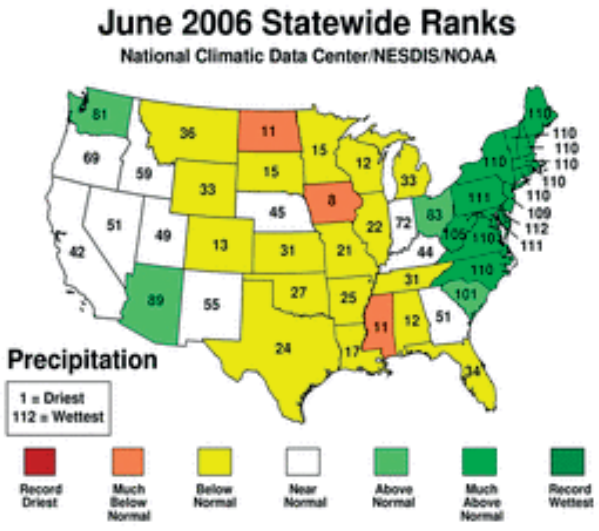


**Figure 4.54 As in figure 4.6 but for June**

The July 2006 weather pattern was dominated by a large ridge centered in the Rocky Mountains. Heights at 500 mb were at or above normal throughout the entire United States as shown below in figure 4.54. This produced record or near record warmth over much of the country during the month (figure 4.55). According to the National Climatic Data Center, July 2006 was the second warmest July in the 1895-2006 record. The preliminary nationally averaged temperature was 25.1°C. The record warmest July was set in 1936, with an average temperature of 25.3°C for the nation. Temperatures ranked above normal in the contiguous U.S. in all states except four, and Wyoming experience its warmest July on record (figure 4.56).

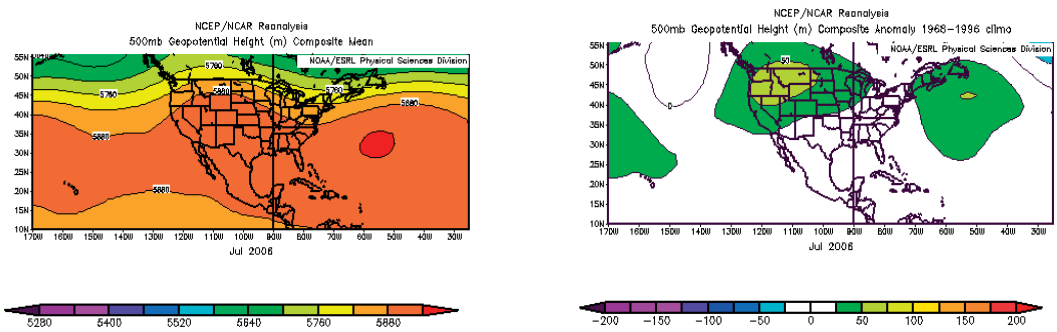


**Figure 4.55 As in figure 4.2 but for July**

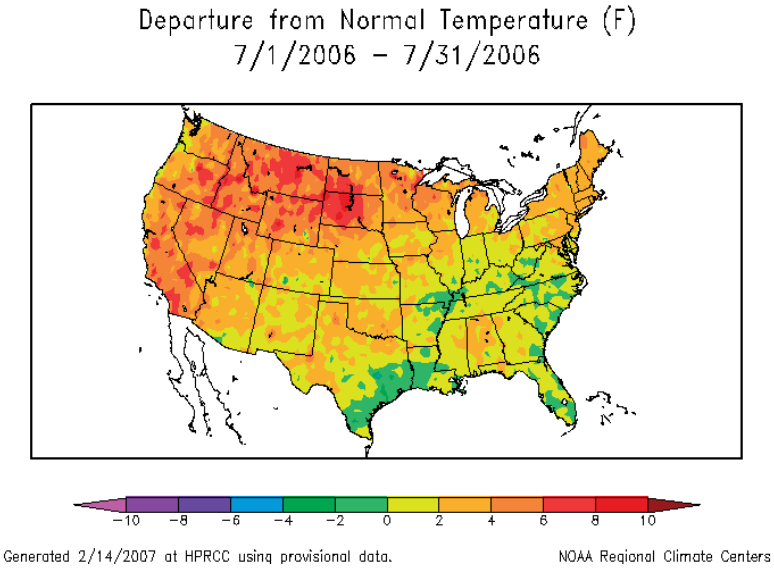


**Figure 4.56 As in figure 4.3 but for July**

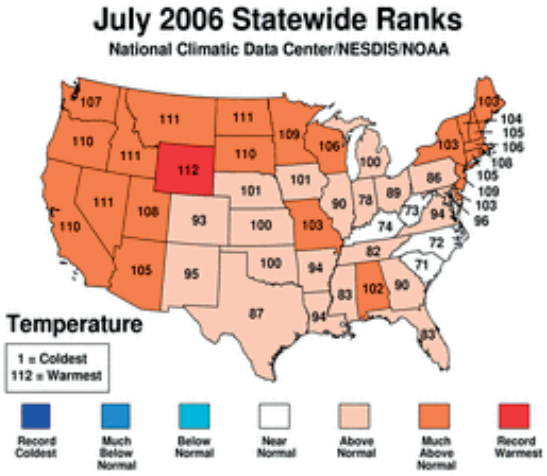


**Figure 4.57 As in figure 4.4 but for July**

The precipitation anomaly map for July is shown below in figure 4.58. Precipitation anomalies across the country were variable. The southwestern and northeastern quadrants of the country were generally above normal, while the northern tier states were generally below to much below normal. According to the National Climatic Data Center, July had below-average precipitation nationally, ranking as the 26th driest July in the 1895-2006 record. An average of 66 mm fell over the contiguous U.S. in July, 8 mm below the 20th century mean for the month (figure 4.59).

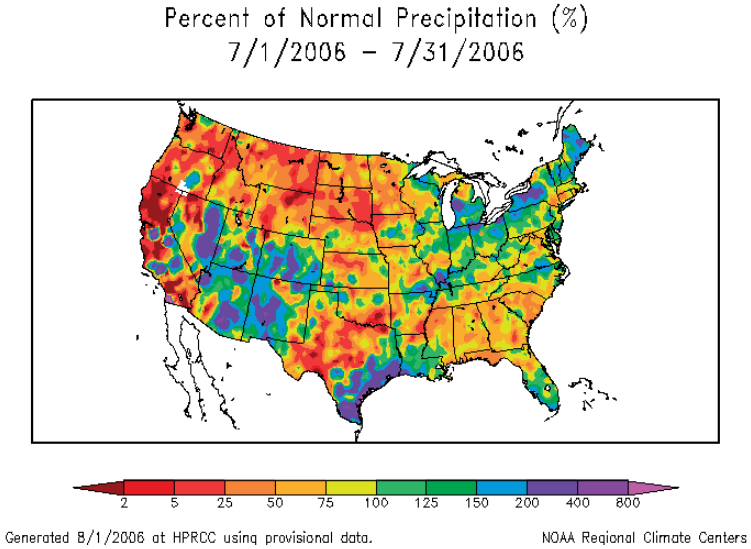


**Figure 4.58 As in figure 4.5 but for July**

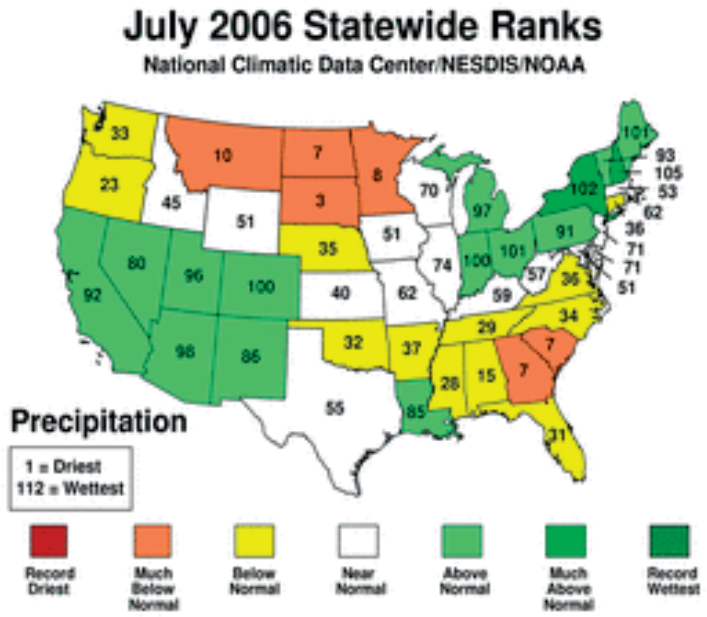


**Figure 4.59 As in figure 4.6 but for July**

The August 2006 weather pattern was dominated by ridging in the central part of the country and troughing off the east coast near Newfoundland, as shown in figure 4.60. As expected with this kind of pattern, above normal temperatures dominated in the middle and eastern part of the country, with below normal temperatures in the extreme northeastern part of the country in association with the mean trough near there. The western states generally had near normal temperatures. The temperature anomaly map is shown in figure 4.61. According to the National Climatic Data Center, August 2006 was the 11th warmest August in the 1895-2006 record. The preliminary nationally averaged temperature was 23.6°C. The record warmest August was set in 1983, with an average temperature of 24.3°C for the nation. Fourteen states were much-warmer-than-normal. North Carolina experienced its second warmest August on record, while Maine and New Mexico both were the only two states that were cooler-than-normal for the month. The statewide rankings for August 2006 are shown in figure 4.62.

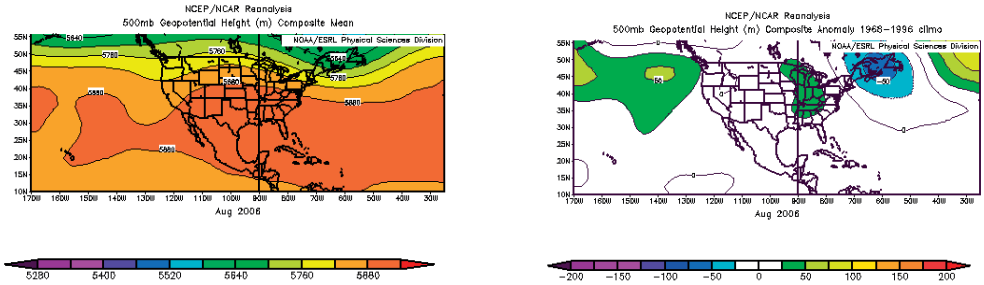


**Figure 4.60 As in figure 4.2 but for August**

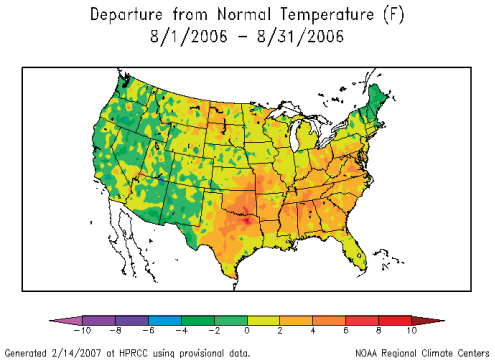


**Figure 4.61 As in figure 4.3 but for August**

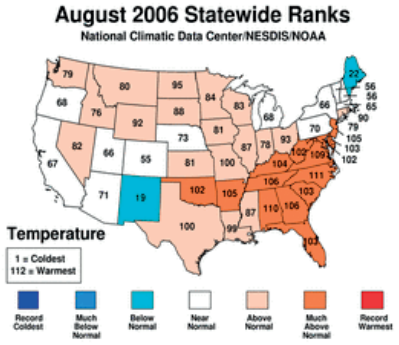


**Figure 4.62 As in figure 4.4 but for August**

The precipitation anomaly map for August is shown below in figure 4.63. The far western part of the country was below normal, while the central part of the country was above normal. New Mexico experienced its wettest August on record since 1895. The eastern part of the country was generally near normal, with the exception of Maryland. According to the National Climatic Data Center, August 2006 ranked as the 17th wettest

August in the 1895-2006 record. An average 75 mm fell over the contiguous U.S. in August, which is 10 mm above the 20th century mean for the month. The statewide precipitation rankings are shown in figure 4.64.

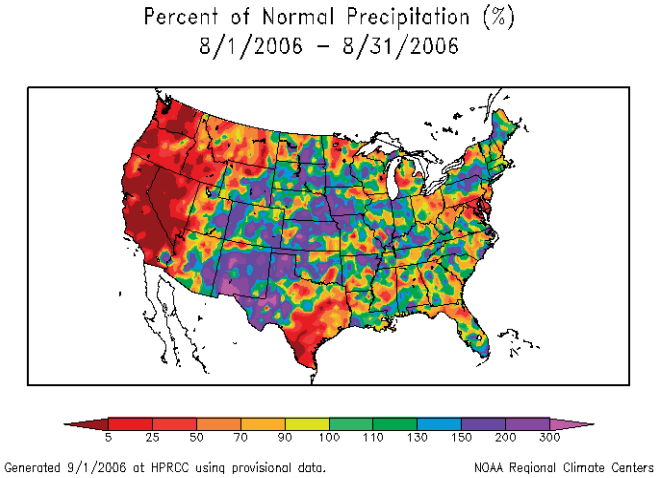


**Figure 4.63 As in figure 4.5 but for August**

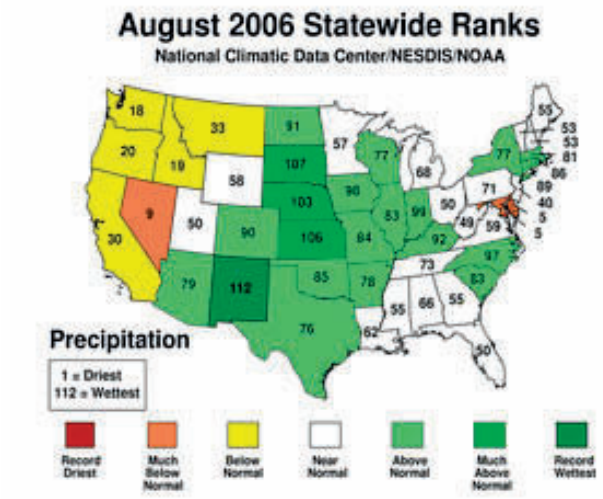


**Figure 4.64 As in figure 4.6 but for August**

The September 2006 weather pattern was dominated by ridging off the west coast of the U.S. and troughing centered in the Mississippi Valley as shown in figure 4.65. Temperatures were generally below normal across most of the country, with the exception of the Pacific Northwest states, where slightly above normal temperatures were present. The temperature anomaly map for September 2006 is shown in figure 4.66. According to the National Climatic Data Center, September 2006 was the 31st coolest

September in the 1895-2006 record. The preliminary nationally averaged temperature was 18.2°C. Temperatures ranked above normal in only five states, with the remaining states in the contiguous U.S. near or below normal for the month. The statewide rankings are shown in figure 4.66.

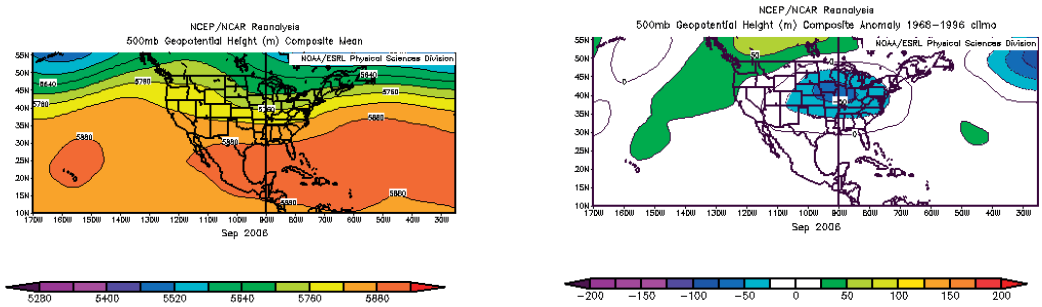


**Figure 4.65 As in figure 4.2 but for September**

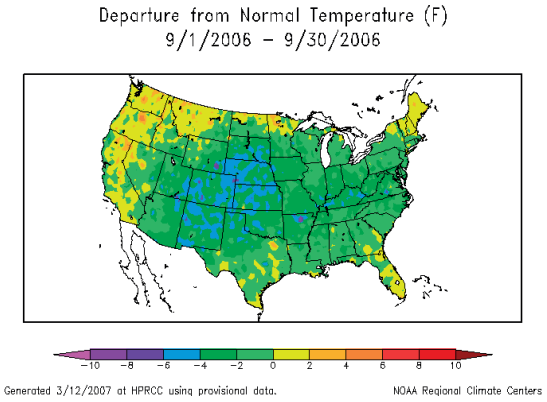


**Figure 4.66 As in figure 4.3 but for September**

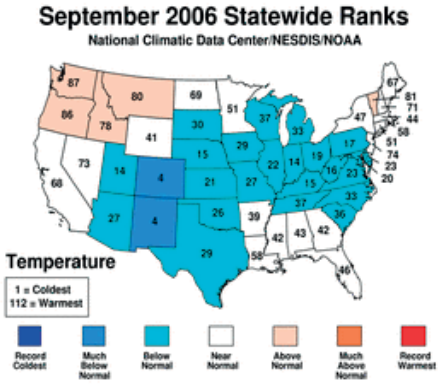


**Figure 4.67 As in figure 4.4 but for September**

The precipitation anomaly map for September 2006 is shown in figure 4.68. Precipitation anomalies across the country were highly variable. The west coast generally had below normal precipitation, while the Ohio Valley generally had above normal precipitation. The rest of the country saw a mix of below normal, near normal, and above normal precipitation. According to the National Climatic Data Center, September 2006 had above-average precipitation nationally, ranking as the 37th wettest September in the 1895-2006 record. An average of 68 mm fell over the contiguous U.S. in September, 5 mm above the 20th century mean for the month. One state, Kentucky, had its wettest September on record. Figure 4.69 shows the statewide ranking of precipitation for the month.

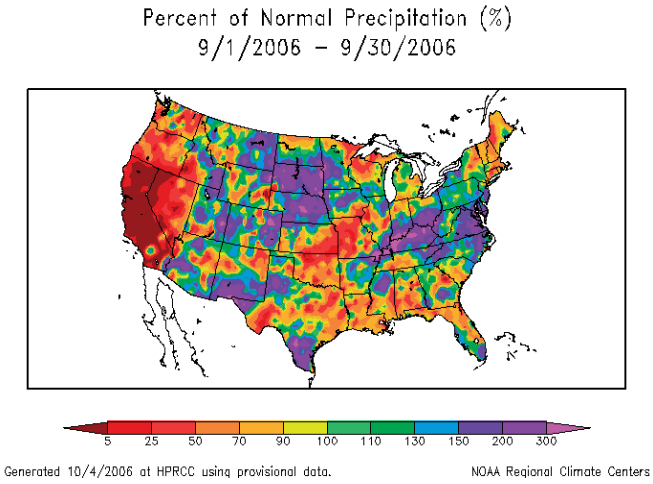


**Figure 4.68 As in figure 4.5 but for September**

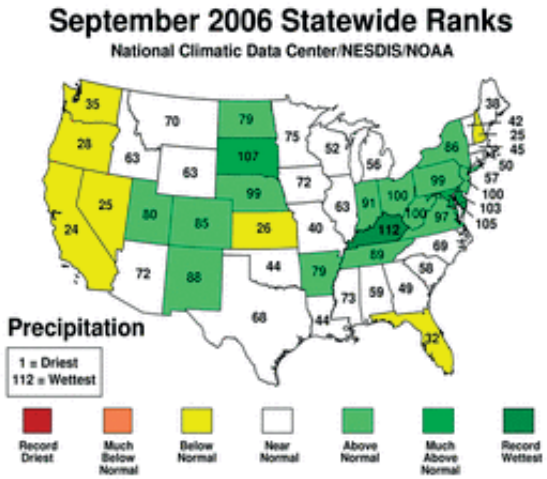


**Figure 4.69 As in figure 4.6 but for September**

### 4.2.4 June, July, August, September 2006 Precipitation Forecast Results

In this section, the superensemble precipitation forecast results are displayed. Daily precipitation forecasts were made out to 5 days at 24-hour time steps during the months of June, July, August, and September of 2006. There were 9 days missing during the 4 month period, either due to corrupted member model data or a lack of member model availability for that particular day, leaving 113 days available for this study. The figures that follow show the equitable threat score (ETS) of the member models as well as the ensemble mean and superensemble for each forecast period for all days available using the domain shown previously in figure 4.1.

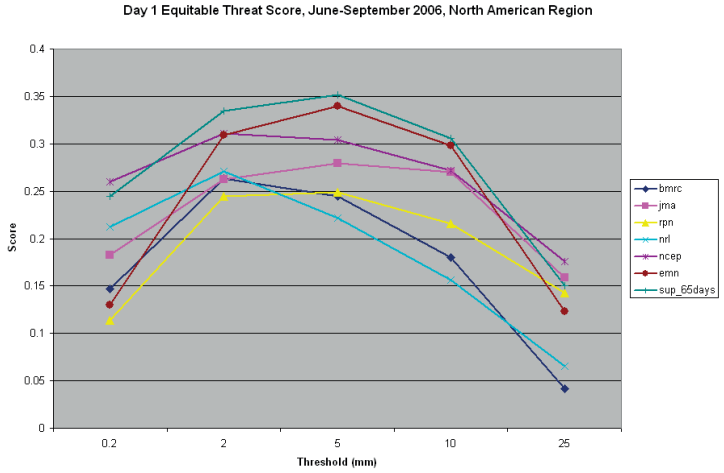


**Figure 4.70 Equitable Threat Score for the Day 1 precipitation forecast during the June-September period**

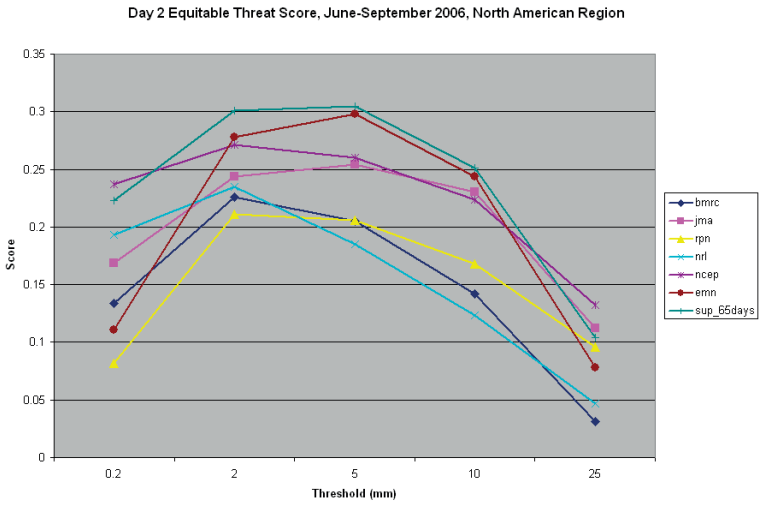


**Figure 4.71 Equitable Threat Score for the Day 2 precipitation forecast during the June-September period**

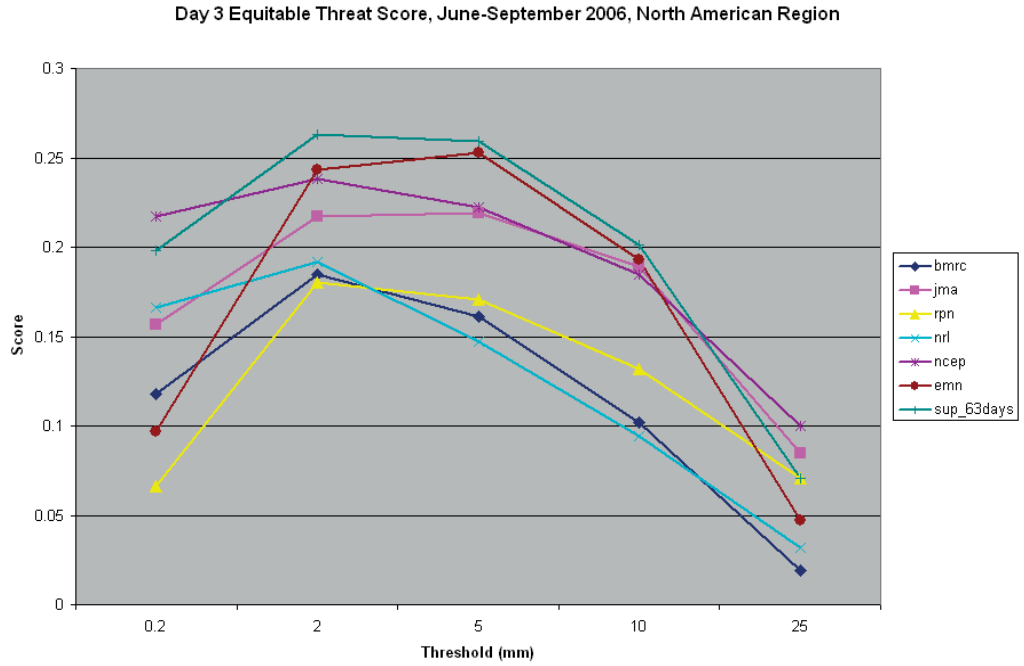


**Figure 4.72 Equitable Threat Score for the Day 3 precipitation forecast during the June-September period**

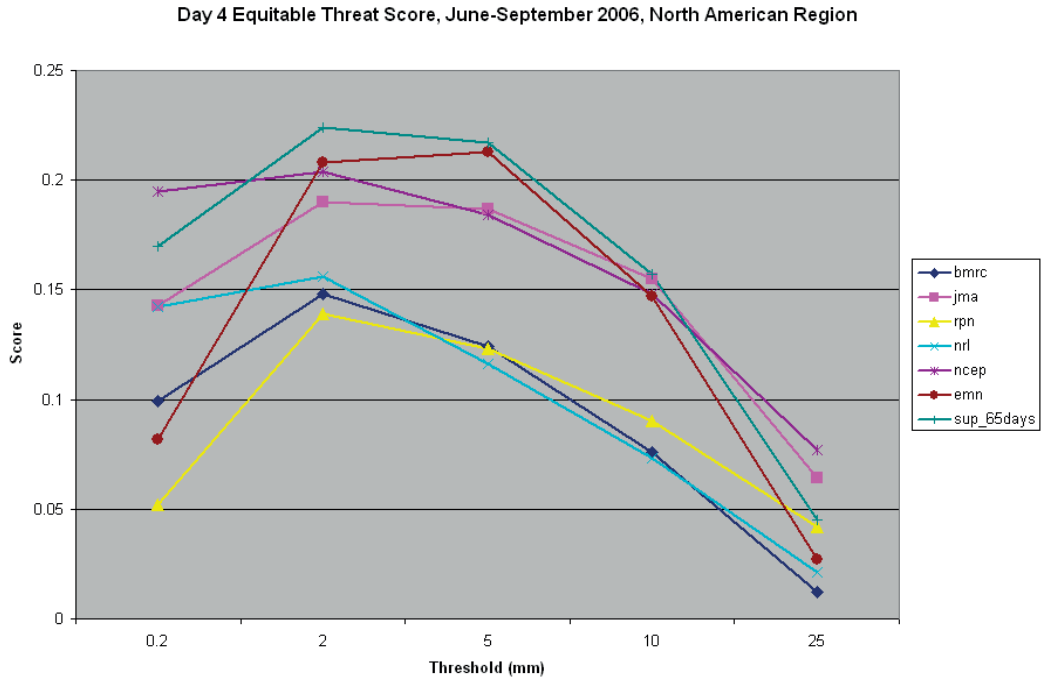


**Figure 4.73 Equitable Threat Score for the Day 4 precipitation forecast during the June-September period**

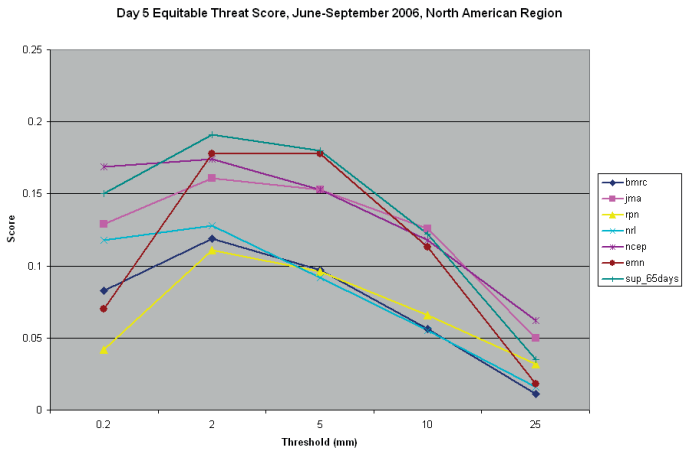


**Figure 4.74 Equitable Threat Score for the Day 5 precipitation forecast during the June-September period**

At the 2 mm/day threshold (includes all precipitation 2 mm/day and greater), the superensemble shows the greatest skill throughout all five days. The same goes for the 5 mm/day threshold. At heavier thresholds of 10 mm/day and 25 mm/day, the superensemble results are also good, although all models have very low skill beyond the 25 mm/day threshold. The superensemble also exhibits superior skill over the ensemble mean at the trace amount threshold of 0.2 mm/day. As an example, figures 4.74 and 4.75 below show the remarkable consistency of the improvement of the superensemble over the member models at the 2 mm/day threshold for the day 1 forecast.

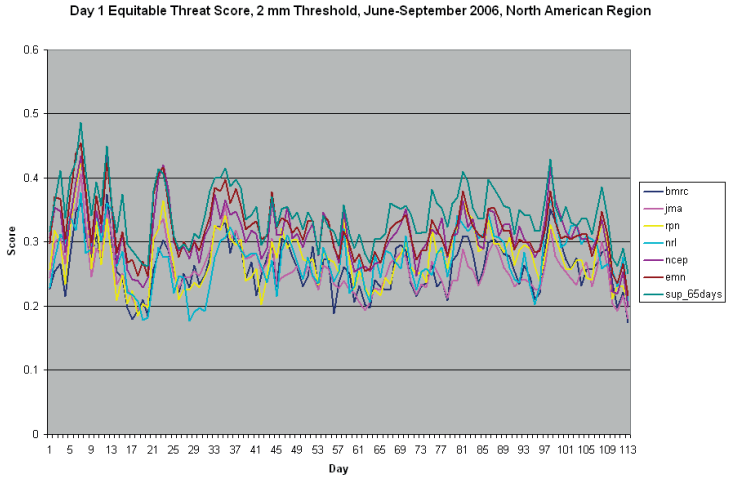


**Figure 4.75 Equitable Threat Score for the 2 mm/day threshold for Day 1 precipitation forecasts during the June-September period**

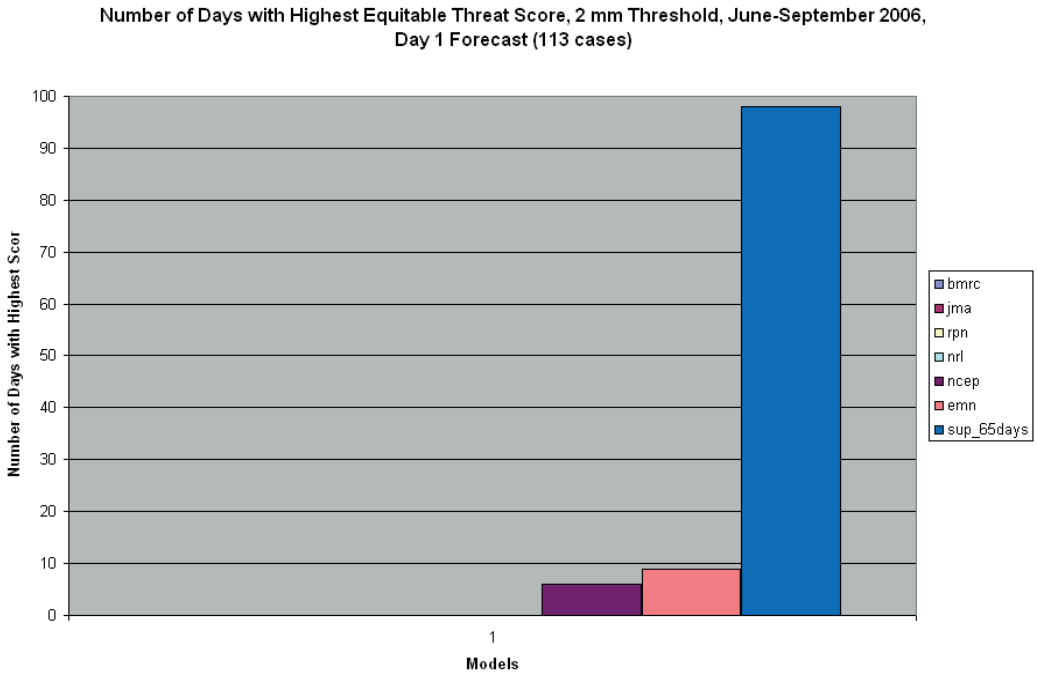


**Figure 4.76 Number of times all models had the highest equitable threat score for the 2 mm/day threshold for Day 1 precipitation forecasts during the June-September period**

It is important to note that all of the member models in this superensemble were downscaled global models. Their resolution was downscaled to ¼ x ¼ degree so that the superensemble also output forecasts at the same ¼ x ¼ degree resolution. The downscaling occurred from an original resolution of 1 degree for each model. The skill of these downscaled, high resolution global models was compared with two mesoscale models: the MM5 and WRF-ARW, as described in chapter 3. The MM5 mesoscale model only produces a 2 day forecast, so only forecasts up to 2 days could be compared. One month of mesoscale model data (August) was obtained for the comparison. The domain of these mesoscale models was also smaller than the original domain used for the superensemble runs, so the domain had to be altered to match the WRF-ARW domain shown in figure 3.8, the smallest domain. ETS scores for days 1 and 2 are shown below.

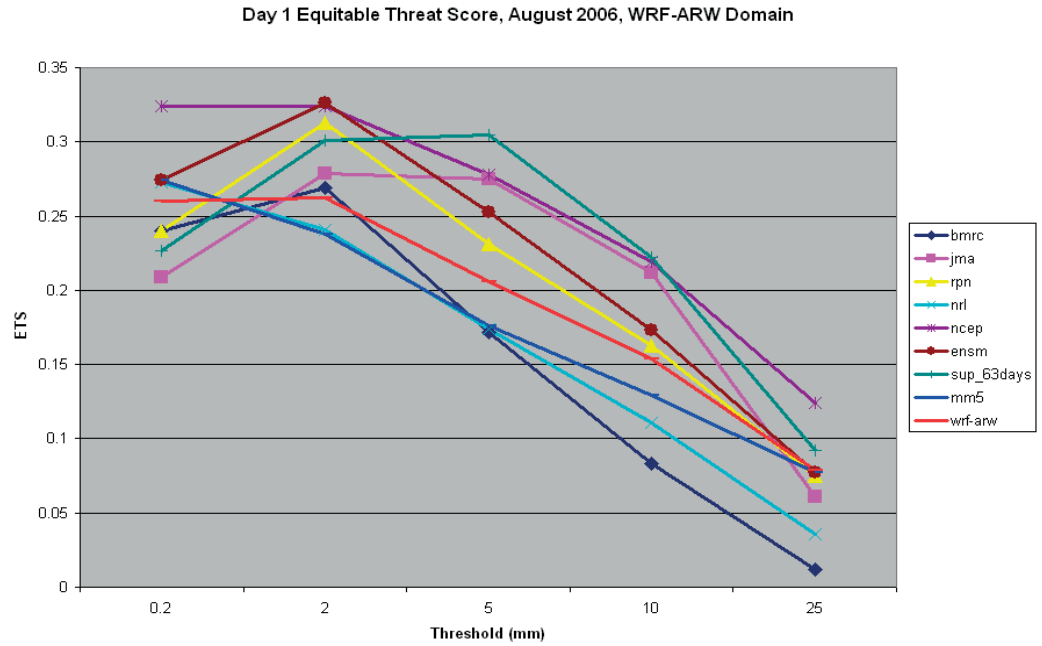


**Figure 4.77 Equitable Threat Score for the Day 1 precipitation forecast during the August period for WRF-ARW domain with mesoscale models added to chart**

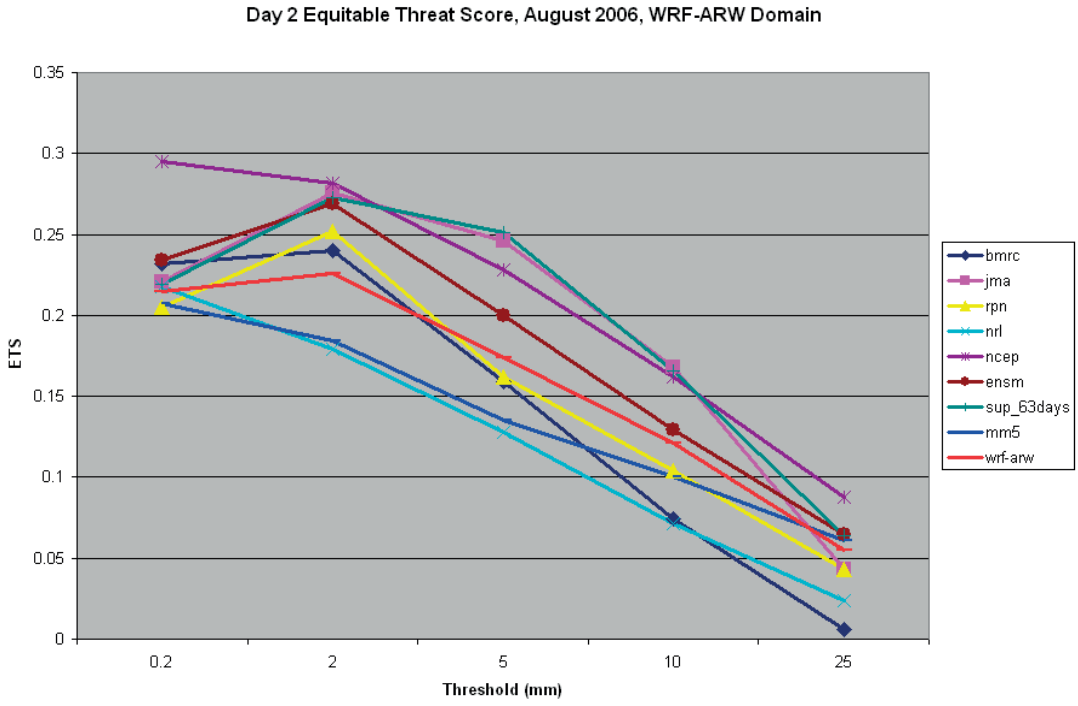


**Figure 4.78 Equitable Threat Score for the Day 2 precipitation forecast during the August period for WRF-ARW domain with mesoscale models added to chart**

These charts would suggest that the skill of these two mesoscale models at high resolutions is actually less than the skill of most of the global model precipitation output downscaled to the same high resolution. These mesoscale models were not added to the superensemble as member models, but rather they are shown for comparison purposes. With the domain changing somewhat from the larger domain used before, as well as the time period of the forecast changing (August vs June-September previously), the skills of all models changed somewhat as compared to the other ETS plots for days 1 and 2. However, the superensemble still shows high skill, with a peak in skill relative to the member models and ensemble mean now occurring at the 5 mm/day threshold rather than the 2 mm/day threshold seen in the larger domain.

Finally, a sensitivity experiment on the number of training days was run for high resolution precipitation. In all of the figures above, the superensemble output shown was run using 65 days of training. In order to see what affect the number of training days might have on high resolution runs, a range of runs was done for September, utilizing 65, 80, 95, 110, 125, 140, and 150 days of training. The figures below show the equitable threat score results. Very little difference in skill was noted between 65 and 150 days of training. The small changes between runs were negligible in either direction.

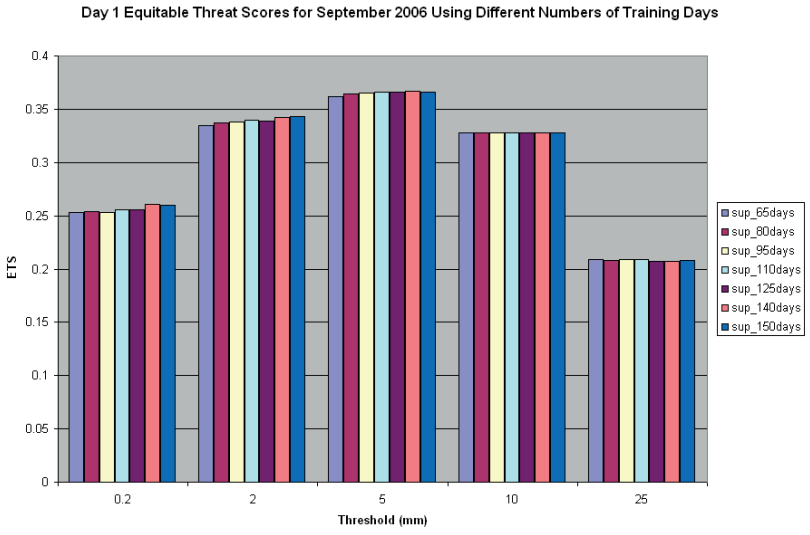


**Figure 4.79 Day 1 Equitable Threat Score for the superensemble at 65, 80, 95, 110, 125, 140 and 150 days of training for September 2006**

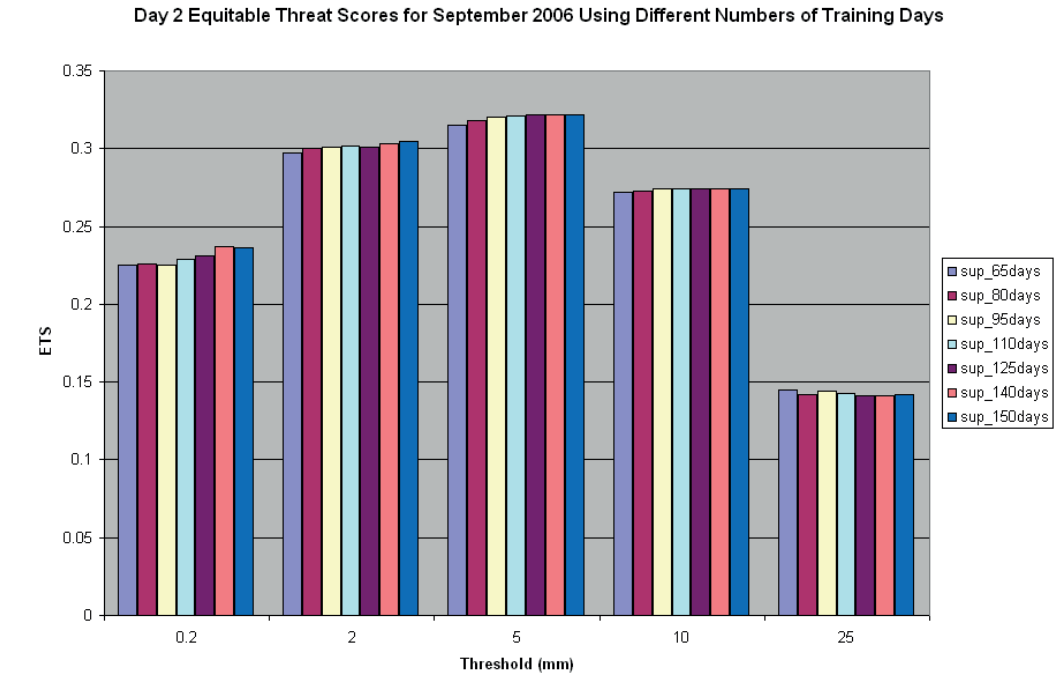


**Figure 4.80 Day 2 Equitable Threat Score for the superensemble at 65, 80, 95, 110, 125, 140 and 150 days of training for September 2006**

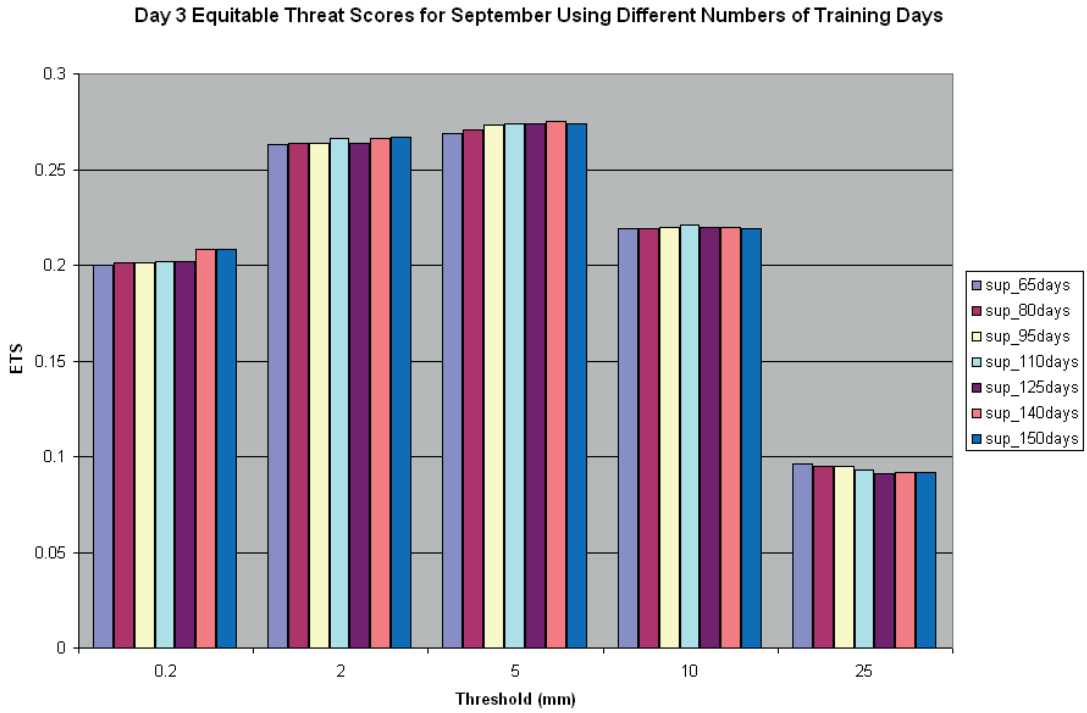


**Figure 4.81 Day 3 Equitable Threat Score for the superensemble at 65, 80, 95, 110, 125, 140 and 150 days of training for September 2006**

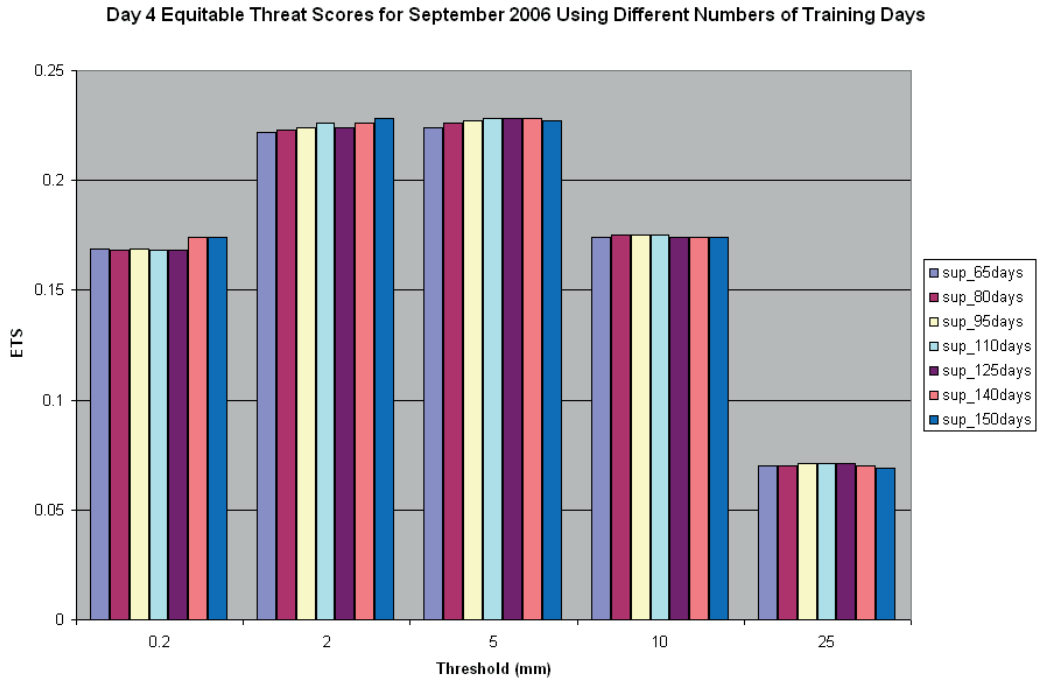


**Figure 4.82 Day 4 Equitable Threat Score for the superensemble at 65, 80, 95, 110, 125, 140 and 150 days of training for September 2006**

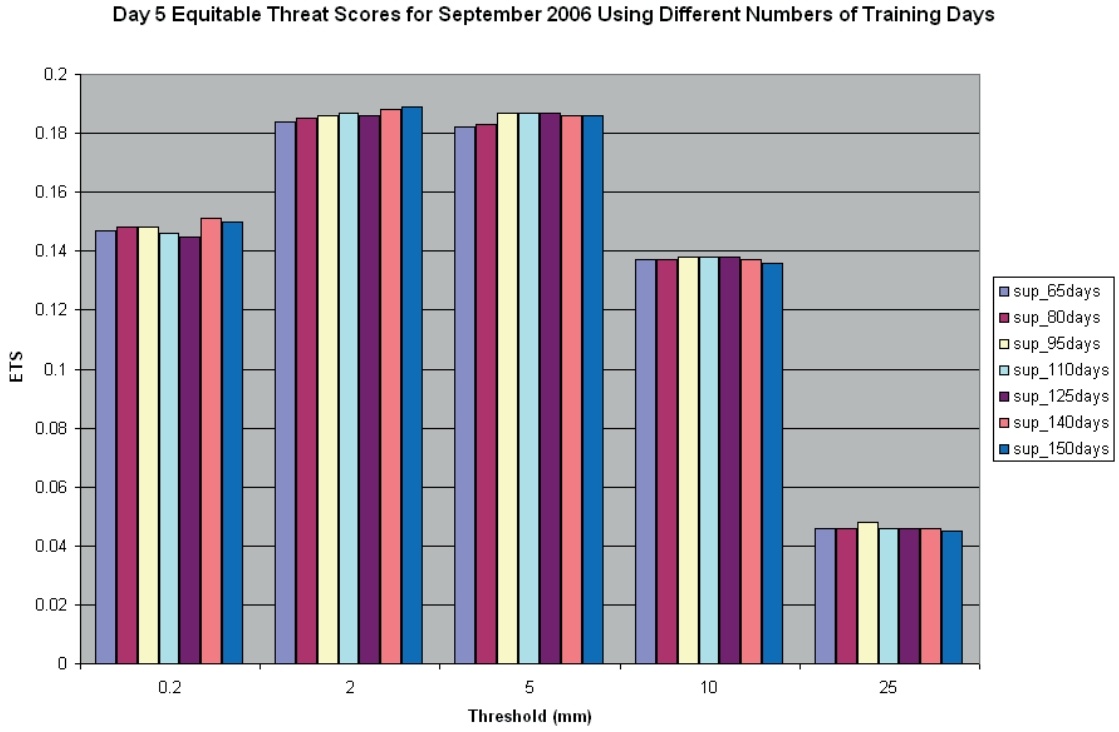


**Figure 4.83 Day 5 Equitable Threat Score for the superensemble at 65, 80, 95, 110, 125, 140 and 150 days of training for September 2006**

# CHAPTER 5
# CONCLUSIONS AND FUTURE WORK

## 5.1 Conclusions

This study has evaluated mesoscale superensemble forecasts of temperature and precipitation across the United States and North America. Real time operational and research mesoscale model forecasts were used to create the mesoscale superensemble forecasts for temperature in an effort to better predict daily temperatures at 3-hourly time spans. The average improvement in mean absolute error over all 60 hours was 0.20 degrees Celsius, or 0.36°F. Based on the information researched in Chapter 1 that stated a 1 billion dollar savings in electricity costs per year in the United States for every 1°F improvement in temperature forecasts, this could theoretically save 360 million dollars per year in electricity costs if the improvement was consistent throughout the entire year *and* it was shown to improve as much on the actual station temperatures as it does on the NARR dataset temperatures. Precipitation forecasts also showed consistently high equitable threat scores throughout the forecast period. With floods rated as the number one natural disaster in the United States in terms of property damage and lives lost, these improved precipitation forecasts would help to provided better warning of impending danger.

## 5.2 Future Work

Future work should include expanding the runs to include an entire year to see if the same consistency in improvement can be demonstrated from season to season. Also, a database of stations and actual station temperatures should be collected and used as training for specific station by station superensemble forecasts instead of superensemble forecasts on a full grid using computer re-analysis data. If the temperature improvement can be shown to be valid on actual station temperatures using station temperature training, then significant economic savings in the weather-sensitive industries could be gained using the FSU superensemble.

REFERENCES

Betts, A. K., and M. J. Miller, 1986: A new convective adjustment scheme. Part II: Single column tests using GATE wave, BOMEX, and arctic air-mass data sets. *Quart. J. Roy. Meteor. Soc.*, **112**, 693–709.

Cartwright, Tina J: Warm Season Mesoscale Superensemble Precipitation Forecasts. PhD. dissertation, The Florida State University, Tallahassee, FL 32306, 105 pp.

Chuang, Hui-Ya, and Geoff Manikin, 2001: The NCEP Meso ETA Model Post Processor: A Documentation. National Centers for Environmental Prediction Office Note 438. http://wwwt.emc.ncep.noaa.gov/mmb/papers/chuang/1/OF438.html

Dallavalle, J. Paul, Mary C. Erikson, and Joseph C. Maloney III, 2004: Model Output Statistics (MOS) Guidance for Short Range Projections. *Preprints*, AMS 20th Conference on Weather Analysis and Forecasting/16th Conference on Numerical Weather Prediction, Seattle, WA, paper #6.1. http://www.nws.noaa.gov/mdl/synop/amspapers/dallava_seattleams.pdf

Glahn, H. R., and D. A. Lowry, 1972: The use of Model Output Statistics (MOS) in objective weather forecasting. *J. Appl. Meteor.,* **11**, 1202-1211.

Glahn, H.R., and D. P. Ruth: 2003: The new digital forecast database of the National Weather Service. *Bull. Amer. Meteor. Soc.*, **84**, 195-201.

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

Janjic, Z. I., 1994: The step-mountain Eta coordinate model: Further developments of the convection, viscous sublayer, and turbulence closure schemes. *Mon. Wea. Rev.,* **122,** 927–945.

Joyce, R. J., J. E. Janowiak, P. A. Arkin, and P. Xie, 2004: CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution.. *J. Hydromet*., **5**, 487-503.

Karl, Thomas R., and Richard W. Knight, 1997: The 1995 Chicago heat wave: How likely is a recurrence? *Bull. Amer. Meteor. Soc.*, **78**, 1107-1119.

Krishnamurti, T.N., C.M. Kishtawal, D.W. Shin, and C.E. Williford, 2000b: Improving tropical cyclone precipitation forecasts from a multianalysis superensemble. *J. Climate*, **13**, 4217-4227.

Krishnamurti, T.N., C.M. Kishtawal, T. LaRow, D. Bachiochi, Z. Zhang, C.E. Williford, S. Gadgil, and S. Surendran, 1999: Improved skills for weather and seasonal climate forecasts from multimodel superensemble. *Science*, **285**, 1548-1550.

Krishnamurti, T.N., C.M. Kishtawal, Z. Zhang, T. LaRow, D. Bachiochi, E. Williford, S. Gadgil, and S. Surendran, 2000a: Multimodel ensemble forecasts for weather and seasonal climate. *J. Climate*, **13**, 4196-4216.

Krishnamurti, T.N., S. Surendran, D. Shin, R. Ricardo-Torres, V. Kumar, E. Williford, C. Kummerow, R. Adler, J. Simpson, R. Kakar, W. Olson, and F. Turk, 2001: Real-Time multianalysis–multimodel superensemble forecasts of precipitation using TRMM and SSM/I Products. *Mon. Wea. Rev.* **129**, 2861-2883.

Krishnamurti, T.N., K. Rajendran, T. S. V. Vijaya Kumar, Stephen Lord, Zoltan Toth, Xiaolei Zou, Steven Cocke, Jon E. Ahlquist, and I. Michael Navon, 2003: Improved skill for the anomaly correlation of geopotential heights at 500 hPa, *Mon Wea Rev* **131***,* 1082–1102.

Lobocki, L., 1993: A procedure for the derivation of surface-layer bulk relationships from simplified second-order closure models. *J. Appl. Meteor*, **32**, 126-138.

Mesinger, F., inc., 2006: North American Regional Reanalysis. *BAMS*, **87**, 343-360.

Perry, Charles A., 2000: Significant floods in the United States during the 20th century - USGS measures a century of Floods. USGS Fact Sheet 024-00. http://ks.water.usgs.gov/Kansas/pubs/fact-sheets/fs.024-00.html

Rogers, David P., 2001: Government Development of National Weather Products and Services. Weather, Climate, and Energy: A policy forum developed by the Atmospheric Policy Program, American Meteorological Society in collaboration with The University of Oklahoma. http://www.ametsoc.org/atmospolicy/climenergyforum/advanceswxclim_sciserv.pdf

Skamarock, William C., J. B. Klemp, J. Dudhia, David O. Gill, Dale M. Barker, Wei Wang, and Jordan G. Powers, 2005: A Description of the Advanced Research WRF Version 2. NCAR Technical Note TN–468+STR, Boulder, CO, http://www.mmm.ucar.edu/wrf/users/docs/arw_v2.pdf

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

## BIOGRAPHICAL SKETCH

Donald F. Van Dyke III was born on February 11, 1982 in Lincoln, Nebraska. In May of 2000 he graduated from Starr's Mill High School in Peachtree City, GA in the top 5% of his class academically. In August of 2000, he started college at the University of North Carolina-Asheville, where he graduated in May of 2004 as Summa Cum Laude with Distinction as a University Scholar from the University Honors Program with a Bachelor of Science degree in Atmospheric Science and a minor in Mathematics. He was also awarded the Certificate of Achievement for Academic Excellence by the Department of Atmospheric Science there. He started at Florida State University in the Fall of 2004 and has since been inducted into Chi Epsilon Pi as well as the Chancellor's List, which is a national program that recognizes the academic achievements of graduate students in America. He will complete his Master's Degree in meteorology under the direction of Dr. T.N. Krishnamurti in the summer of 2007 and continue for a PhD under Dr. Carol Anne Clayson and Dr. Eric Chassignet.